

**STRUCTURAL ECONOMETRIC MODELLING:
METHODOLOGY AND TOOLS WITH
APPLICATIONS UNDER EVIEWS**

CONTENTS

FOREWORD	10
INTRODUCTION	11
THE EXAMPLE: A VERY BASIC MODEL	14
	15
CHAPTER 1: NOTATIONS AND DEFINITIONS	15
1.1 THE MODEL AS A SET OF EQUATIONS	15
1.2 THE ELEMENTS IN A MODEL	15
1.2.1 VARIABLES: ENDOGENOUS AND EXOGENOUS	15
1.2.2 EQUATIONS: BEHAVIORAL AND IDENTITIES	17
1.2.3 PARAMETERS	20
1.2.4 THE RANDOM TERM	21
1.2.5 RESIDUALS VERSUS ERRORS	23
1.2.6 FORMULATIONS	23
1.2.7 THE TIME DIMENSION	23
1.2.8 LINEARITY	29
1.2.9 OTHER PROPERTIES	32
1.2.10 CONSTRAINTS THE MODEL MUST MEET	36
1.2.11 NORMALIZATION AND IDENTIFICATION	42
1.2.12 CONCLUSION	46
2 CHAPTER 2: MODEL APPLICATIONS	47
2.1 OPERATIONAL DIAGNOSES	47
2.1.1 DIFFERENT TYPES OF DIAGNOSES: SCENARIOS AND SHOCKS	47
2.1.2 ADVANTAGES OF MODELS	48
2.1.3 A CERTAIN REASSESSING	49

2.2	THEORETICAL MODELS	50
2.3	QUANTIFIED SMALL MODELS	50
2.3.1	WITH SCIENTIFIC PURPOSE	50
2.3.2	WITH AN EDUCATIONAL PURPOSE	51
3	CHAPTER 3: MODEL TYPES	52
3.1	THE FIELD	52
3.2	THE SIZE	53
3.2.1	DETERMINANTS OF THE SIZE	53
3.2.2	A CLASSIFICATION	54
3.3	THE HORIZON	54
3.3.1	FOR FORECASTING	54
3.3.2	FOR MODEL ANALYSIS	55
3.3.3	A CLASSIFICATION	55
3.4	THE PERIODICITY	55
3.5	OTHER MODELS	56
4	CHAPTER 4: GENERAL ELEMENTS	57
4.1	THE STAGES IN THE PROCESS	57
4.1.1	PREPARING THE MODEL	57
4.1.2	ESTIMATION	57
4.1.3	SOLVING AND TESTING OVER THE PAST.	57
4.1.4	SOLVING AND TESTING OVER THE FUTURE	58
4.1.5	USING THE MODEL FOR FORECASTS AND POLICY STUDIES	58
4.2	THE CHOICE OF SOFTWARE	ERROR! BOOKMARK NOT DEFINED.
4.3	HOW TO ORGANIZE THE DEVELOPMENT OF THE MODEL	58
5	CHAPTER 5: PREPARING THE MODEL	61
5.1	PREPARING THE MODEL: THE FRAMEWORK	61

5.2	PREPARING THE MODEL: SPECIFIC DATA ISSUES	63
5.2.1	TYPES OF DATA	63
5.2.2	THE ACCESS TO DATA	64
5.2.3	PREPARING THE DATA FOR THE TRANSFER	66
5.2.4	THE PRELIMINARY PROCESSING OF SERIES	67
5.2.5	UPDATES	70
5.2.6	SUPPRESSIONS	70
5.2.7	THE DOCUMENTATION	70
5.2.8	CONSEQUENCES ON WORK ORGANIZATION	71
5.2.9	THE PRACTICAL OPTIONS	73
5.3	BACK TO OUR EXAMPLE	73
5.3.1	APPLICATION TO OUR EXAMPLE	77
5.3.2	THE EIEWS PROGRAM	87
5.3.3	A FIRST TEST: CHECKING THE RESIDUALS IN THE IDENTITIES	97
5.4	USING LOOPS AND GROUPS IN EIEWS	105
5.4.1	GROUPS	105
5.4.2	LOOPS	106
6	CHAPTER 6 THE ESTIMATION OF EQUATIONS	109
6.1	THE PROCESS OF ESTIMATION	110
6.2	SPECIFIC ISSUES	110
6.2.1	THE R2 OR R-SQUARED	110
6.2.2	THE CONSTANT TERM	114
6.3	APPLICATIONS: OUR MODEL	115
6.3.1	CHANGE IN INVENTORIES	116
6.3.2	INVESTMENT: THE NECESSITY TO ESTABLISH A CONSISTENT THEORETICAL EQUATION PRIOR TO ESTIMATION	128
6.3.3	EMPLOYMENT: STATIONARITY, ERROR CORRECTION MODELS, BREAKPOINT TEST.	135
6.3.4	EXPORTS: AUTOREGRESSIVE PROCESS, COINTEGRATION, LONG TERM STABILITY.	149

6.3.5	IMPORTS: GOING FURTHER ON COINTEGRATION AND LONG TERM STABILITY.	164
6.3.6	BACK TO THE RESIDUAL CHECK	180
6.3.7	THE PRESENT MODEL	182
7	<u>CHAPTER 7: TESTING THE MODEL THROUGH SIMULATIONS OVER THE PAST</u>	184
7.1	THE SOLUTION	185
7.1.1	GAUSS-SEIDEL	185
7.1.2	RITZ-JORDAN	186
7.1.3	NEWTON AND ITS VARIANTS	186
7.1.4	ITERATIONS AND TEST OF CONVERGENCE	191
7.1.5	STUDY OF THE CONVERGENCE	194
7.1.6	SOLVING THE MODEL: BASIC EViews TECHNIQUES	217
7.2	A FIRST VALIDATION	231
7.2.1	EX POST SIMULATIONS	231
7.2.2	EX-POST FORECASTS	237
7.2.3	SOLVING THE MODEL: SCENARIOS	238
7.2.4	OUR EXAMPLE	241
7.2.5	ANALYTIC SHOCKS	243
7.2.6	OUR EXAMPLE	246
8	<u>CHAPTER 8: TESTING THE MODEL OVER THE FUTURE</u>	251
8.1	MAKING THE MODEL CONVERGE IN THE LONG RUN	252
8.1.1	THE ASSUMPTIONS	252
8.1.2	ADAPTING THE FORMULATIONS	253
8.1.3	IMPROVING THE CHANCE (AND SPEED) OF CONVERGENCE	255
8.1.4	SOLVING PARTIAL MODELS	255
8.1.5	CHECKING THE EXISTENCE OF A LONG TERM SOLUTION	255
8.2	CONVERGENCE PROBLEMS IN THE SHORT RUN	257
8.3	CONVERGENCE PROBLEMS IN THE MEDIUM RUN	258

8.4	TESTING THE RESULTS	259
8.4.1	OF SIMULATIONS	259
8.4.2	OF SHOCKS	259
8.5	EViews FEATURES	259
8.5.1	PRODUCING SIMULATIONS	260
8.5.2	PRODUCING A BASE SOLUTION	262
8.5.3	PRODUCING SHOCKS	263
8.5.4	CHANGING MODEL SPECIFICATIONS	263
8.6	RATIONAL EXPECTATIONS	267
8.6.1	THE FRAMEWORK	267
8.6.2	CONSEQUENCES FOR MODEL SIMULATIONS	267
8.6.3	TECHNICAL ELEMENTS	268
8.6.4	OUR EXAMPLE	268
8.6.5	THE TEST	270
8.6.6	THE RESULTS	270
8.7	STOCHASTIC SIMULATIONS	272
8.7.1	PURELY STATISTICAL ERRORS	272
8.7.2	FORMULATION AND ASSUMPTION ERRORS	273
8.7.3	CONSIDERING THE ERRORS	273
8.7.4	THE INTEREST OF THE TECHNIQUE	275
8.7.5	BACK TO OUR EXAMPLE	276
8.7.6	THE RESULTS	276
8.8	GOING FURTHER: STUDYING MODEL PROPERTIES	279
8.8.1	EIGENVALUE ANALYSIS	279
8.8.2	THE CASE OF ERROR CORRECTION MODELS: A SIMPLE EXAMPLE	290
9	CHAPTER 9: USING MODELS	293
9.1	OPERATIONAL DIAGNOSES	293

9.1.1	SCENARIOS AND THEIR DIFFERENT TYPES	293
9.1.2	MANAGING ACTUAL FORECASTS: THE TARGETING OF SIMULATIONS	296
9.1.3	DIFFERENT TYPES OF SHOCKS	297
9.1.4	THE FORECAST : TECHNICAL ASPECTS	299
9.1.5	CHANGING MODEL SPECIFICATIONS	300
9.1.6	OPTIMAL CONTROL	302
9.2	TEACHING WITH MODELS	303
9.3	PRESENTATION OF RESULTS	303
9.3.1	GENERAL ISSUES	303
9.3.2	TABLES	304
9.3.3	GRAPHS	305
10	CHAPTER 10: APPLYING THE ABOVE PRINCIPLES TO OPERATIONAL CASES	307
10.1	THE ACCOUNTING FRAMEWORK	308
10.1.1	THE AGENTS: A FIRST DEFINITION	309
10.1.2	THE OPERATIONS	309
10.1.3	THE INTEGRATED ECONOMIC ACCOUNTS	309
10.1.4	SECTORS, BRANCHES AND PRODUCTS,	310
10.1.5	AGENTS SUBDIVISIONS	311
10.1.6	A MULTI COUNTRY MODEL	311
10.2	A SINGLE COUNTRY, SINGLE PRODUCT MODEL	311
10.2.1	THE ECONOMIC ASPECTS	311
10.2.2	THE EIEWS PROGRAMS	330
10.3	A SINGLE COUNTRY, MULTI PRODUCT MODEL	428
10.3.1	THE MAIN REASONS FOR PRODUCT DECOMPOSITION	428
10.3.2	INTRODUCING INTERMEDIATE CONSUMPTION	431
10.3.3	SPECIFIC SECTORAL ISSUES	432
10.3.4	THE DATA	434

10.3.5	CONSEQUENCES FOR ESTIMATIONS	434
10.3.6	THE PRODUCTION FUNCTION	435
10.3.7	UNEMPLOYMENT	436
10.3.8	CHANGE IN INVENTORIES	436
10.3.9	HOUSEHOLD CONSUMPTION	437
10.3.10	EXTERNAL TRADE	437
10.3.11	WAGES	437
10.3.12	PRICES	437
10.4	A MULTI COUNTRY, SINGLE PRODUCT MODEL	448
10.4.1	FIRST ISSUE: THE MODELS	448
10.4.2	SECOND ISSUE: INTER-COUNTRY TRADE	449
10.4.3	A CONSISTENT METHOD: MACSIM	450
10.4.4	THE EViews PROGRAM	458
10.5	A REGIONAL MODEL	458
10.6	A MULTI COUNTRY, MULTI PRODUCT MODEL	459
10.6.1	THE EViews PROGRAM	459

ANNEX 1: A SIMPLE METHOD ALLOWING TO SWITCH THE ENDOGENOUS/EXOGENOUS STATUS OF MODEL

VARIABLES **460**

11.1	THE PROBLEM	460
11.2	THE LINEAR CASE	460
11.3	THE NON-LINEAR CASE	461
11.4	A FEW QUESTIONINGS	462
11.5	GENERAL CONCLUSION	464
11.6	A TEST ON A VERY SMALL MODEL.	464
11.6.1	THE PROGRAM	465
11.6.2	THE RESULTS	471
11.7	A TEST ON A LARGER MODEL	473

12 A LIST OF USEFUL SERIES FOR PRODUCING A SINGLE COUNTRY, SINGLE PRODUCT MODEL **475**

12.1	THE PERIODICITY	475
12.2	GENERAL ELEMENTS	475
12.3	THE SERIES	475
12.3.1	THE SUPPLY - DEMAND VARIABLES AT CONSTANT PRICES (WHOLE ECONOMY)	475
12.3.2	THE SAME ELEMENTS AT CURRENT PRICES	476
12.3.3	THE SAME ELEMENTS DECOMPOSED INTO PRODUCTS	476
12.4	PRODUCTIVE CAPACITY	476
12.5	EMPLOYMENT	476
12.6	PRICE DEFLATORS	476
12.7	HOUSEHOLDS ACCOUNT	477
12.8	FIRMS ACCOUNT (ALL TYPES OF FIRMS)	477
12.9	REST OF THE WORLD	478
12.10	GOVERNMENT ACCOUNT (ELEMENTS NOT DESCRIBED EARLIER)	478
12.11	FINANCIAL DOMAIN	478
13	<u>A LIST OF USEFUL SERIES FOR A THREE SECTOR MODEL</u>	479
13.1	THE PERIODICITY	479
13.2	GENERAL ELEMENTS	479
13.3	THE SERIES	479
13.3.1	THE SUPPLY - DEMAND VARIABLES AT CONSTANT PRICES (WHOLE ECONOMY)	480
13.3.2	* THE SAME ELEMENTS AT CURRENT PRICES	480
13.4	PRODUCTIVE CAPACITY	480
13.5	EMPLOYMENT	480
13.6	PRICE DEFLATORS	481
13.7	HOUSEHOLDS ACCOUNT	481
13.8	FIRMS ACCOUNT (ALL TYPES OF FIRMS)	481
13.9	REST OF THE WORLD	482
13.10	GOVERNMENT ACCOUNT (NOT DESCRIBED EARLIER)	482

13.11	FINANCIAL DOMAIN	482
-------	------------------	-----

BIBLIOGRAPHY	489
---------------------	------------

ECONOMETRICS AND STATISTICS	489
-----------------------------	-----

MACROECONOMICS	490
----------------	-----

MODELS AND MODELLING	491
----------------------	-----

14	INDEX	493
-----------	--------------	------------

FOREWORD

The purpose of this book is a little special.

First, of course, by its subject: we have to admit that structural econometric modelling is no longer so popular, having lost ground to Computable General Equilibrium models and in particular their Dynamic Stochastic versions.

We will contend that while this might be true in the academic field (you just have to look at the program of congresses and symposiums) there is still a lot of place for structural models. Indeed many institutions are still using them and even building new ones; both in developed and developing countries. We shall try to show that this position is quite justified, and that for a large part of the modelling applications, in particular the analysis and interpretation of macroeconomic interactions, the call for structural models remains a good strategy, arguably the best one.

But we shall not stop at proving the usefulness of these models. For the people we have convinced, or which were so already, we will provide a set of tools facilitating all the tasks in the modelling process. Starting from elementary elements, it will lead by stages the user to a level at which he should be able to build, manage and use his professional, operational model.

This means this book will, as its title says, focus essentially on applied and even technical features, which does not mean it will be so simplistic.

After a necessary description of the field, we shall use the largest part of the book to show the reader how to build his own model, from general strategies to technical details. For this we shall rely on a specific example, presented at the beginning, and which we will follow through all the steps of model development. When the situation becomes more complex (with the addition of product and international dimensions), we shall still find this model at the core of the cases.

Our examples will be based on that package, the most popular modeling package at present. This will allow us to be more helpful to EViews users, concentrating on its practice (including some tricks).

Finally, just as important if not more so, we shall provide a set of files allowing readers to practice modelling (either alone or as part of a course). And for more advanced users, we shall give access to files allowing to produce operational (if small) models, which they can adapt to their own ideas, with the tedious tasks: producing the data, defining the accounting framework and organizing simulations over the future, being already prepared.

All these elements are provided for free, and downloadable on a specific site. However, to keep track of people interested in our research, we will require the potential users to ask for them first. We will then allow unrestricted access.

One final remark: this book is based on version 8 of EViews. However, all the programs we are providing work also with version 7 (and almost all with version 6).

The new features (mostly improvements) will be pointed out at the time, and the same descriptive elements gathered in a special chapter at the end of the book.

INTRODUCTION

Since an early date in the twentieth century, economists have tried to produce mathematical tools which, applied to a given practical problem, formalized a given economic theory to produce a reliable numerical picture. The most natural application is of course to forecast the future, and indeed this goal was present from the first. But one can also consider learning the consequences of an unforeseen event, or measuring the efficiency of a change in the present policy, or even improving the understanding of a set of mechanisms too complex to be grasped by the human mind.

In the last decades, three kinds of tools of this type have emerged, which share the present modelling market.

- The “VAR” models. They try to give the most reliable image of the near future, using a complex estimated structure of lagged elements, based essentially on the statistical quality, although economic theory can be introduced, mostly through constraints on the specifications. The main use of this tool is to produce short term assessments.
- The Computable General Equilibrium models. They use a detailed structure with a priori formulations and calibrated coefficients to solve a generally local problem, through the application of one or several optimizing behaviors. The issues typically addressed are optimizing resource allocations, or describing the consequences of trade agreements. The mechanisms described contain generally little dynamics.

This is no longer true for the Dynamic Stochastic General Equilibrium models, which dominate the current field. They include dynamic behaviors and take into account the uncertainty in economic evolutions. Compared to the traditional models (see later) they formalize explicitly the optimizing equilibria, based on the aggregated behavior of individual agents. This means that they allow agents to adapt their behavior to changes in the rules governing the behaviors of others, including the State, in principle escaping the Lucas critique. As the model does not rely on traditional estimated equations, calibration is required for most parameters.

- The “structural” models. They start from a given economic framework, defining the behaviors of the individual agents according to some globally consistent economic theory. They use the available data to associate to these behaviors reliable formulas, which are linked by identities guaranteeing the consistency of the whole set. These models can be placed halfway between the two above categories: they do rely on statistics, and also on theory. To accept a formula, it must respect both types of criteria.

The use of this last kind of models, which occupied the whole field at the beginning, is now restricted to policy analysis and medium term forecasting. For the latter, they show huge advantages: the full theoretical formulations provide a clear and understandable picture, including the measurement of individual influences. They allow also to introduce stability constraints leading to identified long term equilibriums, and to separate this equilibrium from the dynamic fluctuations which lead to it.

Compared to CGEs and DSGEs, optimization behaviors are present (as we shall see later) and introduced in the estimated equations. But they are frozen there, in a state associated with a period, and the behavior of other agents at the time. If these conditions do not change, the statistical validation is an important advantage. But sensitivity to shocks is flawed, in a way which is difficult to measure.

A very good (and objective) description of the issue can be found in:

http://en.wikipedia.org/wiki/Dynamic_stochastic_general_equilibrium

http://en.wikipedia.org/wiki/Macroeconomic_model#Empirical_forecasting_models

It seems to us that the main criterion in the choice between DSGEs and traditional structural models lie in the tradeoff between statistical validation and adaptability of behaviors.

In the last years, popularity of structural econometric modelling seems to have stabilized. A personal hint for this (if not an actual proof) is the growing demand for assistance in structural modelling addressed to the INSEE International Cooperation Unit from which I have just retired.

Another issue is that being the first tool produced (in the thirties of the last century) it was applied immediately to the ambitious task of producing reliable forecasts. The complexity of the economy, and the presence of many random shocks makes this completely unrealistic (and this is even more true today). During the golden years of structural modelling, when economy was growing at a regular (and high) rate, forecasting was as easy as riding a tame horse on a straight path: anybody could do it. But when the horse turned into a wild one, the quality of the rider showed, and it did not stay in the saddle too long. Failing to succeed in a task too difficult for any tool (including VAR and CGE models, which do not have to forecast the medium term), gave discredit to structural models and all their uses, including policy analysis and even the understanding and interpretation of complex economic mechanisms, applications for which neither VAR nor CGE can compete in our opinion.

However, even with limited ambitions, producing a sound econometric structural model is not a simple task. Even a professional economist, having an excellent knowledge of both economic theory (but not necessarily a complete and consistent picture) and econometric techniques (but not necessarily of their practical application) will find it quite difficult producing a reliable and operational econometric model.

The purpose of this book is to shorten the learning process, in several ways.

After a global presentation of economic models:

- Notations, definitions, mathematical characteristics (dynamics, linearity, continuity, identifiability...).
- Applications: economic theory, forecast, education.
- Classification of existing models.

We shall describe how to organize the sequence of model building tasks, from data production and framework specification to actual operational studies.

For each task, we shall give all the necessary elements of methodology.

We shall present the main economic options available, with some theoretical explanations.

All these explanations will be based on a practical example, the production of a very small model of the French economy. The size will not forbid us to address most of the problems encountered in the process.

The methods, techniques and solutions proposed will be based on the EViews software. This will allow us to present some useful features and tricks, and to provide a sequence of complete programs, which the user can modify at will, but not necessarily too heavily, as all the models of this type share a number of common elements. The main issue is of course the estimation process, each case leading generally to an original version of each behavioral equation.

A set of documented programs will be provided, following the above principles

- For the small example,
- For a more detailed product, a model for a single country, not far from an operational version.

These programs will allow to:

- Import the original data
- Build the model framework
- Transform the data to conform to the elements in the model.
- Estimate a set of equations, starting with standard behaviors, possibly updated.
- Check the technical and theoretical consistency of the resulting model.
- Produce forecasts and policy studies.

In each case, we shall present programs which actually work. An econometric solution will be found, reliable both in statistical and economic terms. And the properties of the models will be rather satisfying, with a long term solution and reasonable dynamics leading to it.

Finally, we shall address the more complex problems: multi-sector and multi-country models (and both options combined). The specific issues will be described, and a framework for a three-product model will be provided, following the same lines as the previous example.

The goal of this book is therefore both limited and ambitious. Without getting into theoretically complex features, it should give readers all the elements required to construct their own model. Being relieved of the more technical (and tedious) tasks, they will be allowed to concentrate on the more intelligent (and interesting) ones.

Readers must be aware they will find here neither a full description of econometric and statistical methods, nor a course in economic theory. We shall give basic elements on these fields, and rather focus on their links with the modelling process itself. For more detailed information, one can refer to the list of references provided at the end of the volume.

THE EXAMPLE: A VERY BASIC MODEL

To present the elements and the framework of a structural econometric model, we shall use a specific example, which we shall address permanently during our presentation. In spite of its limited size, we think it remains quite representative of the class of models we are considering in this manual.

At the start of any model building process, one has to specify in a broad manner the logic of his model, and the behaviors he wants his model to describe. No equation needs to be established at this time. We shall place ourselves in this situation.

In our example, an economist has decided to build a very simple model of the French economy. As our tests will be based on actual data, a country had to be chosen, but the principles apply to any medium sized industrialized country.

Our model includes the following elements.

- Based on their production expectations and the productivity of factors, firms invest and hire workers to adapt their productive capacity. However, they exert some caution in this process, as they do not want to be stuck with unused elements.
- The levels reached in practice define potential production.
- Firms also build up inventories.
- Households obtain wages, based on total employment (including civil servants) but also a share of Gross Domestic Product. They consume part of this revenue.
- Final demand is defined as the sum of its components: consumption, productive investment, housing investment, the change in inventories, and government demand.
- Imports are a share of local demand («domestic demand»). But the less capacities remain available, the more an increase in demand will call for imports.
- Exports follow world demand, but producers are limited by available capacities, and their priority is satisfying local demand.
- Supply is equal to demand.
- Productive capital grows with investment, but is subject to depreciation.

The above framework looks rather straightforward, and certainly simplistic. Obviously, it lacks many elements, such as prices, financial concepts, and taxes. This will be addressed as later developments.

Let us not go further for the time being. One can observe that if we have not built a single equation yet, a few are already implicit from the above text.

CHAPTER 1: NOTATIONS AND DEFINITIONS

Before we start presenting the process of model building, we must define the concepts we shall use. They will be based on individual examples taken from our (future) model.

1.1 THE MODEL AS A SET OF EQUATIONS

In a general way, a model will be defined as a set of fully defined formulas describing the links between a set of concepts.

Formally, a model can be written as the vector function of variables.

$$f(\dots) = 0$$

We shall address in turn:

- The nature of elements appearing in the function.
- The nature of the functions themselves.

1.2 THE ELEMENTS IN A MODEL

1.2.1 VARIABLES: ENDOGENOUS AND EXOGENOUS

Obviously, a model will be used to measure economic concepts, depending on other elements.

Two variable types will appear in a model:

- Endogenous variables, or results, whose value will be obtained by solving the system of equations,
- Exogenous variables, or assumptions, whose value is known from outside considerations, and which obviously condition the solution.

If the model is solved over past periods, this value should be known. But in forecasting operations, it will have to be chosen by the model builder (or user).

For the system to be solved, the number of endogenous variables must correspond to the number of equations. Our formulation becomes:

$$f(x, y) = 0$$

with

x vector of exogenous variables

y vector of endogenous variable (with the same dimension as f).

For instance in our model:

- Imports will be endogenous, as they depend on local demand. Exports too, depending on world demand.
- World demand will be exogenous, as we are building a model for a single country, and we are going to neglect the impact of local variables on the world economy. Of course, this impact exists, as France is (still) an important country, and its growth has some influence on the world economy. But the relatively limited improvement can only be obtained at the very high cost of building a world model. This simplification would be less acceptable for a model of the USA, or China, or the European Union as a whole (we shall address this issue in detail later).

Technically, one can dispute the fact that exports are endogenous. As we make them depend only on an exogenous world demand, they are de facto predetermined, apart from an unforecastable error. They are logically exogenous and technically endogenous. We shall use the second criterion (the problem will disappear with the introduction of other explanatory elements, which will make exports truly endogenous).

As to Government demand, models of the present type will keep it also exogenous, but for different reasons:

- The goal of this model is to show its user (which can be the Government, or a Government advising agency, an independent economist playing the role of Government, or even a student answering a test on applied economics) the consequences of its decisions. So these decisions must be left free, and not forced on him.
- The behavior of the State is almost impossible to formalize, as it has few targets (mostly growth, inflation, unemployment, budget and trade balances) and a much larger number of instruments. If their base values are more or less fixed, it can deviate from them arbitrarily, without too much delay. To achieve the same goal, past French governments have used different global approaches, and calling for different panels of individual instruments.¹
- The State alone has enough individual power to influence significantly the national economy.

Each of the two exogenous elements is characteristic of a broader category:

- Variables considered as external to the modeled area, on which economic agents taken into account by the model have no or little influence. In addition to the situation in other countries, this can mean population², or

¹ For instance, to decrease unemployment, a government can increase demand or reduce firms' taxes, and the tax instrument can be social security contributions, or subsidies

² In long term models growth might affect the death and birthed rates thus population.

meteorological conditions³, or the area available for farming. The theoretical framework of the model can also suppose exogenous structural elements, such as the real interest rate, the evolution of factor productivity, to the depreciation rate of capital.

- Variables controlled by an agent, but whose decision process the model does not describe. Even if it was formally possible, the model builder wants to master their value, to measure their consequences on the economic balance. These will be referred to as decision variables or «instruments».

Changing the assumptions on these two types of variables, therefore, will relate to questions of very different spirit:

- What happens if perhaps...? (the price of oil increases abruptly).
- What happens if I (the State), decide...? (to decrease the VAT rate on CDs⁴).

The second type of question can be inverted: what decision do I have to take to obtain this particular result? (By how much should I decrease the rate of employers' social contributions to create 1000 jobs?). This means that the status (exogenous/endogenous) of some variables is changed: the answer calls for specific techniques, or solving a transformed model. We will deal with this later.

Sometimes the two approaches can also be combined; by considering first the consequences of an evolution of uncontrolled elements, and then supposing a reaction of the State, for instance a change in policy that would return the situation to normal. For instance, the State could use its own tools to compensate losses in external trade due to a drop in world demand.

From a model to another, the field described can change, but also the separation between endogenous and exogenous. The real interest rate can change its nature depending on the endogeneity of the financial sector, technical progress can be assumed as a trend or depend on growth, and the level of population can depend on revenue.

1.2.2 EQUATIONS: BEHAVIORAL AND IDENTITIES

1.2.2.1 Behaviors

The first role of the model is to describe "behaviors": the model builder, following most of the time an existing economic theory, will establish a functional form describing the behavior of a given agent, and will use econometrics to choose a precise formulation, with estimated parameters.

In describing consumption, one might suppose that its share in household income is determined by

- The level of income (a higher income will make consumption less attractive or necessary, compared to savings⁵).

³ Which can depend on growth (glasshouse effect).

⁴ Provided the EU commission will allow it.

⁵ Let us recall that investment in housing is considered as savings.

- Recent variations of income (consumers take time in adapting their habits to their new status).
- The evolution of unemployment: if it grows, the prospect of losing a job will lead households to increase reserves.
- Inflation: it defines the contribution required to maintain the purchasing power of financial savings.

Once identified, all these elements will be united in a formula, or rather a set of possible formulas (households can consider present inflation, or the average over the last year; the increase in unemployment can use its level or percentage change). These formulas will be confronted with the available data, to find a specification statistically acceptable on the whole, each element participating significantly in the explanation, and presenting coefficient values consistent with economic theory. Once parameters are estimated, each element of the resulting formulation will contribute to the logical behavior of the associated agent.

But the process is not always so straightforward. Two other cases can be considered.

- The behavior can be formalized, but not directly as estimation-ready formulas. A framework has first to be formalized, then processed through analytical transformations possibly including derivations and maximizations, leading finally to the equation (or set of equations) to estimate. This will be the case for our Cobb-Douglas production function (page 105) for which we compute the combination of labor and capital which maximize profits for a given production level. Or for the definition of the wage rate as the result of negotiations between workers unions and firm managers, based on their respective negotiating power.
- Often the model builder will not be able to formulate precisely the equation, but will consider a set of potential explanatory elements, waiting for econometric diagnoses to make a final choice between formulations (generally linear). For instance, the exchange rate might depend on the comparison of local and foreign inflation, and on the trade balance.

In any case, even if the exact intensity of influences is unknown to the model builder⁶, economic theory generally defines an interval of validity, and especially a sign. Whatever the significance of the statistical explanation, it will be rejected if its sign does not correspond to theory. In the example above, the increase of labor demand must generate gains in the purchasing power of the wage rate.

The formulation of these theoretical equations often makes use of specific operators, allowing alternative calculations: Boolean variables, maximum and minimum operators. For instance, in disequilibrium models, the theoretical equation can include a constraint. We can consider also the case of a function of production with complementary factors, where the level of each factor determines an individual constraint:

$$CAP = \min(pl.L, pk.K)$$

⁶ Otherwise he would not have estimated it.

with CAP production capacity, L employment, K capital, and p_l , p_k the associated productivities

1.2.2.2 Identities

A model composed only of behavioral equations is not generally usable as such. Additional equations will be needed, this time with undisputable forms.

Several cases can be identified, which can apply simultaneously:

- Some concepts are linked by an accounting formula, and we need to ensure their numerical coherence. For example, once the model has defined household revenue, it cannot estimate savings and consumption separately as the sum of the two is known⁷. A single element will be estimated: it can be savings, consumption, the savings ratio or the consumption ratio, and the other elements will follow, using identities.
- Some concepts are linked by a causal sequence of elements, and some elements in the chain are not defined by behaviors. For example, if we estimate firms employment and household consumption, we must formalize household revenue (as a sum including wages) to make job creation improve consumption. And in our example, defining final demand (as a sum of its components) ensures that imports will follow consumption.

Of course, one can consider eliminating these identities by replacing each element they compute by the corresponding formula. This is not always technically possible, but in any case it would:

- Lead to overly complex formulations, difficult to interpret and slower to compute.
- Discard potentially interesting information.

In addition, one will be led to introduce:

- Intermediate variables simplifying formulations (and speeding up computations). Even if the growth rate of the real wage rate, which uses a slightly complex expression, was not considered interesting as an economic quantity, it will be useful to define it, if it appears as an explanatory element in many equations.
- Purely descriptive elements: the ratio of Government Balance to GDP is a crucial element in evaluating the financial health of the state (and one of the « Maastricht » criteria for entering the European Monetary Union).
- Finally, economic theory is not always absent from this type of equation: the supply – demand equilibrium has to be enforced:

Q (supply from local producers) + M (foreign supply to the country) = FD (demand from local agents) + X (foreign demand to the country).

And the choice of the variable which balances it has a strong theoretical impact on model properties.

- If exports and imports come from behaviors, and demand from the sum of its components, we need to compute Q as:

⁷ This would also be absurd in terms of household behavior.

Q (local output) = $(FD-M)$ (local demand supplied by local producers) + X (foreign demand supplied by local producers)

This means that production will adapt to demand (which itself can depend on the availability of products).

- But we could also suppose that:

The producers chose to limit their output at a level actually lower than demand, because additional production would bring negative marginal profits. In this case Q will be fixed, and we could have:

$$Q=\text{fixed}, X=f(WD), FD = f(\text{economy}), M = FD - Q + X$$

- Or the country can only import in foreign currency, which it obtains through exports.

$$X=f(WD), M=f(X), Q=\text{fixed}, FD =Q+(M-X)$$

1.2.3 PARAMETERS

Parameters can be defined as scalars with a varying value. The only formal difference with exogenous variables is that they lack a time dimension⁸.

Two types of parameters can be considered, according to the way their value is established:

- Those estimated by reference to the past: starting from a theoretical but fully defined formula including unknown parameters, the model builder will seek the values which provide the formulation closest to observed reality, according to a certain distance. This means using "econometrics".
- Those decided by the model builder: economic theory or technical considerations can supply a priori assumptions concerning a particular behavior. For instance, if a Central Bank uses a standard Taylor rule to decide its interest rate, its sensitivity to the inflation level should be 0.5. A special case will be represented by a control variable, giving (without changing the formulation) a choice between several types of independent behaviors.

The distinction is not as clear as it may seem: in particular, if estimation fails to provide an economically coherent result, the model builder can be driven to decide on the values of some parameters.

With \mathbf{a} as a vector of parameters ($\hat{\mathbf{a}}$ estimated) the system becomes:

⁸ In EViews, modifying a parameter value applies to the current model, and taking it into account calls for a new compilation, making the new version official. This is both tedious and error-prone. One might consider replacing parameters by series with a constant value, which gives access to the much more manageable "scenario" feature.

$$f(x, y, \hat{a}, a) = 0$$

And in our example, one could estimate the influence of world demand on exports, for example by supposing that relative variations are proportional (or equivalently that the elasticity of exports to world demand is constant).

$$\Delta X / X = a \cdot \Delta WD / WD$$

where a should be close to unity, if the share of the country on the world market is stable⁹.

But if the estimated coefficient not significant, we can get back to:

$$\Delta X / X = \Delta WD / WD$$

This choice could also have been made from the start for theoretical reasons.

Clearly, to estimate a parameter it is necessary to define entirely the associated formula.

1.2.4 THE RANDOM TERM

In practice, the behavior of agents does not answer exactly to formalized functions, and the formulation obtained by estimation will not reproduce the reality. It will only approximate this behavior, using elements which conform to some economic theory, each of them providing a sizable contribution to the changes in the explained variable. The number of estimated parameters will then generally be much lower than the size of the sample, or the number of observed values. In practice, adding elements to the explanation can:

- In the good cases, improve the quality of the explanation given by the elements already present, which can now concentrate on their natural role, instead of trying to participate in the explanation of other mechanisms in which their efficiency is limited¹⁰.

⁹ In our model WD stands for world trade (including its expansion), not the aggregate demand of countries.

¹⁰ Just like a worker which has to use his time on two tasks, and is really qualified for one. For example, if an excellent musician but average lyricist is teamed with a good lyricist, the quality of songs (both music and lyrics) will improve.

- But the new element can compete with the others in explaining a mechanism in which they all have some competence, limiting the improvement and leaving the sharing of the explanation rather undetermined (and therefore limiting the significance of the coefficients).¹¹¹²

In practice, these correlation problems will always appear, sometimes very early, and generally before the fifth or sixth element. Beyond that figure, the precision of individual coefficients will decrease, and global quality will improve less and less.

This means that a typical econometric equation will contain a maximum of four parameters, while variables will be known on fifty to one hundred quarters.

It will be therefore necessary, to formulate an exact model, to accept the presence of non-zero additional terms (residuals). If one believes in the model, this residual should be interpreted as a random perturbation without economic meaning. But if the equation is badly specified, it will also come from other sources: omitting a relevant variable, replacing it by another less relevant, or choosing the wrong form for the equation¹³.

The fault will not always lie with the model builder, who might not have been able to apply his original ideas. The variables he needs may not be precisely measured, or only with a slightly different definition, or they may not be available at all, as in, for example, the goals or anticipations of a given agent.

Practically speaking, one will often suppose that this residual follows a random distribution, with a null average, a constant standard error, and residuals independent across periods.

Our formulation becomes therefore, in the general case, noting u the vector of residuals:

$$f(x, y, \hat{a}, a, u) = 0$$

In the example, if we want to represent changes in household consumption as a constant share of total production variations, we will write:

$$CO = a \cdot Q + b + u$$

or rather, if we want u to have a constant relative influence:

¹¹ This can be a problem for the model if the two competing elements have a different sensitivity to a particular variable. For instance, if one is sensitive to a tax rate, the other not: then the role of the tax rate will be undetermined.

¹² If two workers with the same profile complete a task together, it is difficult to evaluate their individual contribution. One might have rested the whole period.

¹³ Of course, as we have said before, one is never able to estimate the « true » equation. This remark should apply to a large conceptual error, leading to behaviors distinctly different from an acceptable approximation of reality.

$$CO/Q = a + u$$

1.2.5 RESIDUALS VERSUS ERRORS

It is probably the time to bring an important issue about the nature of econometrics.

When he considers a behavioral equation, the economist can have two extreme positions.

- He believes the behavior can be exactly specified according to a formula, which is affected by an error term with a given distribution (maybe a white noise, or a normal law). With an infinite number of observations we would get an exact measurement of the parameters, and therefore of the error and its distribution.
- He thinks that the concept he wants to describe is linked with some other economic elements, but the relation is only an application, of which any formula represents only an approximation. To this application a random term can also be added, if one believes that the replication of the same explanatory elements will bring a different result. Additional observations will only get a better mapping.

The debate is made more complex by several facts:

- The data on which he wants to base his estimation is not measured correctly. One cannot expect the statisticians to produce error free information, for many reasons: measurement errors, inappropriate sample, mistaken concepts...
- Even if measured correctly, the concepts he is going to use are not necessarily the right ones. For instance a given behavior should be applied to the only firms which do make profits, a separation which is not available at the macroeconomic level.
- The discrete lags which he will apply to these concepts are not the right ones either. For instance it might be known that an agent takes into account the price index of the last month, but only quarterly data is available.
- The estimation period is not homogenous, and this cannot be explained by the data. For instance the mood of consumers (and their consumption behavior) can evolve without any link to measurable economic elements.

From the above elements, the logical conclusion should be:

- The first position is illusory, and to a point which is impossible to measure (of course).
- But we have to take it if we want to apply econometric methods.

This means that in the following text we shall put ourselves in the first position, but we will always keep in mind the true situation, and give to the difference between the concept and its estimation the less ambitious name of "residual".

1.2.6 FORMULATIONS

We shall now consider the form of the equations. Let us first approach the time dimension.

1.2.7 THE TIME DIMENSION

Variables in economic models have generally a temporal dimension, which means they are known through discrete values, almost always with a constant periodicity: generally annual, quarterly or monthly series. This means we will consider models in discrete time.

There are exceptions, however. The most frequent applies to micro-economic models, describing the behavior of a panel of individual firms or households, and the dimension will correspond to items in a set. Sometimes they will be ordered, using the level of one variable, such as the income level for a set of households. Time can be introduced as an additional dimension, but possibly with a varying interval, either predetermined (phases of the moon) or unpredictable (periods of intense cold).

1.2.7.1 Consequences of the discretization

The time discretization of variables will be introduced in several ways, leading to:

- Really instantaneous variables, measured at a given point in time: the capital on the 31st of December at midnight, in an annual model (defined as a **stock** variable).
- **averages**: the average level of employment observed during a period.
- **flows**: the goods produced during a period.

The same economic concept might appear under several forms: inflation and price level, stock of debt and balance for the period, average and end-of-period employment levels and net job creations. For one household, we can consider the revenue, its yearly change, and the global revenue accumulated during its existence.

1.2.7.2 The seasonality

When models have a less than yearly periodicity, some series can present a specific distortion depending on the sub-period inside the year. This can come from changes in the climate: in winter the consumption of electricity will increase due to heating and lighting, but construction will be mostly stopped. It can be due to social issues: the concentration of holidays in the summer months can reduce production, and the coming of Christmas will increase consumption (in Christian countries). We are going here to provide a basic sketch of the problem, leaving a more serious description to specialized books like Ladiray and Quenneville (2001).

Using unprocessed data can lead to problems: for instance the level of production in summer will be lower than what we could expect from labor and capital levels. This will disturb estimations and make model solutions more difficult to interpret.

Two solutions can be considered:

- Introducing in the estimated equations “dummy” variables associated to each sub-period.
- Extracting from the series their seasonal component and producing a completely new set of values.

Of course one should not mix the two type of techniques in the same equation (or model).

The second method will be favored, as it solves also the interpretation problem.

Several techniques are available, the most well-known being Census-X13 ARIMA, developed by the US Census Bureau and Statistics Canada¹⁴. But TRAMO-SEATS¹⁵ is also a common choice. Both are available under EViews.

One must be aware that this process often reduces the statistical quality of estimations. For instance if demand is particularly high in the last quarter of each year, and imports follow, seasonally adjusting both series will make the link less clear, bringing less precise results. Even more obviously, the relation between demand for heating and temperature will lose power from seasonal adjustment.

These examples show the main issue: in a one-equation model, the transformation is essential if the explanation contains a seasonal component, in addition to truly economic features. For instance, agricultural production will be lower in winter, even if the same level of labor, land, fertilizer, machinery is available.... Truly, at the same time, the use of fertilizer will decrease, and probably of labor too, but in a lower way. This means that the whole set of variables, both dependent and explanatory variables, must be seasonally adjusted.

On the contrary, if all the seasonal explanation comes from the seasonality of explanatory elements, seasonally adjusting is not necessary, and even reduces the quality of estimations (with the variability of elements). One could use raw series to estimate an imports equation, using demand, rate of use of capacities and price competitiveness as explanatory elements.

But what is true for one equation does not apply to the whole model. One cannot mix the two types of series, and this means seasonally adjusting will prevail in practice.

1.2.7.3 Static and dynamic models

To determine the equilibrium for a given period, some models will use only variables from this period: we shall call them static models. They correspond to the formulation:

$$f_t(x_t, y_t, a, u_t) = 0$$

The most frequent case is that of input-output models, which use a matrix of "technical coefficients" to compute the detailed production associated to a given decomposition of demand into categories of goods, which itself depends only on instantaneous elements.

$$Q = A \cdot FD$$

(A representing an n by n square matrix)

¹⁴ <http://www.census.gov/srd/www/x12a/>

¹⁵ <http://www.bde.es/webbde/es/secciones/servicio/software/econom.html>

$$FD = f(Q)$$

On the contrary, dynamic models use variables from other periods.

The reasons are quite numerous. They can be:

- **theoretical:** some agents will be supposed to base their behavior on the observation of the past. Firms will increase their prices if the profits of the previous quarter have been too low. Or they will build their expectations of demand growth on the previous evolutions of the same variable. These two examples illustrate the main issues: using the past to create an image of the future (backward looking expectations), or to measure a previous gap between actual and target values, which the agent will try to close in the present period.
- **institutional:** the income tax paid by households can be based on their income of the previous period (this is the case in France, for the time being).
- **technical:** if a model takes into account a variable and its growth rate, computing one from the other takes into account the previous level.

One observes that each of these justifications supposes that influences come only from previous periods: one will speak of (negatively) lagged influences.

The formulation becomes therefore:

$$f_t(y_t, y_{t-1}, \dots, y_{t-k}, x_t, x_{t-1}, \dots, x_{t-l}, a, u_t) = 0$$

Let us go back to our model. We can observe already an undisputable lagged influence: most of present capital will come from the remaining part of its previous level. Any other case is still undecided. However, without going too deep into economic theory, one can think of several lagged influences:

- For household consumption, we have already considered that adapting to a new level of revenue takes some time. This means it will depend on previous levels. If we detailed it into products, the previous level can have a positive influence (some consumptions are habit - forming) or a negative one (generally, people do not buy a new car every quarter):

$$CO_t = f(CO_{t-1}, CO_{t-2}, \dots, HRI_t)$$

- Firms invest to adapt their productive capacities to the level of production needed in the future. We can suppose that they build their expectations on past values.

$$I_t = f(Q_t, Q_{t-1}, Q_{t-2}, \dots)$$

It is interesting to note that the previous formulation could be simplified, eliminating any lag larger than one by the addition of intermediate variables:

$$f_t(y_{i,t}, y_{j,t-k}) = 0$$

(where y_i and y_j represent variables, indexed by time t and $t-k$)

is equivalent to

$$f_t(y_{i,t}, z_{j,t}) = 0$$

$$z_{1,t} = y_{j,t-1}$$

$$z_{2,t} = z_{1,t-1} \quad (= y_{j,t-2})$$

.....

$$z_{k-1,t} = z_{k-2,t-1} \quad (= y_{j,t-k+1})$$

$$z_{k,t} = z_{k-1,t-1} \quad (= y_{j,t-k})$$

in which a lag of k periods on a single variable has been replaced by k one period lags on as many variables (including new ones).

The same method clearly allows eliminating lagged exogenous variables.

On the investment equation of the example, this would give:

$$I_t = f(Q_t, Q1_t, Q2_t, Q3_t)$$

$$Q1_t = Q_{t-1}$$

$$Q2_t = Q1_{t-1}$$

$$Q3_t = Q2_{t-1}$$

But if this method simplifies the theoretical formulation, it has the obvious disadvantage of artificially increasing the size of the model and reducing its readability, without producing additional information. Its interest is reserved to specific studies. For instance, assessing of model dynamics can call for the linearization of the model according to

present and lagged variables. The above transformation will limit the matrix system to two elements (with lags 0 and 1), which will make further formal computations easier, and independent from the number of lags.

It also allows us to use a simplified formulation in subsequent presentations:

$$f_t(y_t, y_{t-1}, x_t, a, u_t) = 0$$

1.2.7.4 Particular case: rational expectations

It has appeared natural, in previous examples, to consider only negative lags. This will happen if we suppose that the anticipation of agents relies only on the observation of the past (and the present)¹⁶.

To justify positive lag formulations, it is necessary to suppose:

- That agents have the possibility, by their present decisions, to determine the future values of some variables (and the associated behavior can be formalized).
- That agents anticipate perfectly the future (perfect expectations).
- That the expectation by agents of specific evolutions has for consequence the realization of these evolutions (self-fulfilling expectations).
- That agents build their expectations on the behaviors of the other agents¹⁷, for which they know the properties (rational expectations). Basically, this means that they are able to apply the model controlling the economy (but not necessarily know its formulas), and the decision process defining its assumptions. For instance, they can forecast the investment program of the Government (depending on economic conditions), they know how firms and households will react, and they know the links between these elements (they are able to take into account the supply-demand equilibrium).
- However, they do not necessarily know the unexplained part of the behaviors (which can be associated with the random term). If know only their distribution, we shall speak of stochastic rational expectations. EViews does not provide this feature at present (only one or the other). They also do not have to know the actual formulas, just be able to compute them.

You do not have to believe in rational expectations to apply them. Producing alternate simulations with different assumptions on expectations will improve greatly the insight in one particular model or on economic mechanisms in general. We shall present this later using a specific case.

¹⁶ This use of proxies is made necessary by the absence of direct measurement of anticipations. Exceptionally, they can be obtained by surveys, leading to a specific estimation.

¹⁷ Including the State.

1.2.7.5 Other case: continuous time models

This also is a very specific area: some theoretical models will be formulated as a system of equations where variables appear as a function of continuous time, and variations (or growth rates) become exact derivatives. One ends up then with a system of differential equations, which one can be led to integrate.

These models seldom evolve beyond a theoretical stage, if only for lack of statistical information.

But some operational models, describing for instance the stock exchange, can reduce their periodicity to a daily or even shorter value.

1.2.8 LINEARITY

We will consider here the linearity relative to variables. The linearity relative to coefficients will appear in the chapter on estimation.

The potential linearity of a model represents a very important property for its analysis as well as its solution. But first we must define the notion of linearity, which can be more or less strict.

The most restrictive will be:

$$A \cdot y_t + B \cdot y_{t-1} + C \cdot x_t + b + u_t = 0$$

but one can let matrix elements change with time:

$$A_t \cdot y_t + B_t \cdot y_{t-1} + C_t \cdot x_t + b_t + u_t = 0$$

a definition again less restrictive will suppose linearity relative to the sole endogenous variables:

$$G(x_t, a)y_t + H(x_t, a)y_{t-1} + J(x_t, a) + u_t = 0$$

or even relative to the endogenous of the period:

$$G(y_{t-1}, x_t, a) \cdot y_t + H(y_{t-1}, x_t, a) + u_t = 0$$

Using the multiplier as an example, we can already show that these properties affect the computation of derivatives of model solutions. We will detail later the consequences on convergence properties.

The first property tells that it does not depend on the initial equilibrium, or the period considered. Multiplying the shock by a given factor will have a proportional effect. It is enough to compute it once to know it once and for all.

In the second case, the multiplier will depend only on the period. Starting from different base assumptions will not change the consequences of a given change.

In the third case, the multiplier will depend also on the exogenous values (and the coefficients). It has to be re-computed each time these elements change (or have changed in the past except for one period ahead solutions), but can be stored until the next time they do.

The last case is similar to the third one. But convergence will be affected (see later).

1.2.8.1 Practical cases of non-linearity

It is obvious enough that a single non-linear equation makes the model non-linear, according to one of the previous definitions. Reasons for non-linearity are multiple; one will find in particular:

- Expressions measured in growth rates (therefore possibly linear relative to the endogenous of the period). For example the growth rate of wages can depend on inflation.
- Expressions formulated as elasticities (generally integrated into logarithms). One will suppose for example that imports and domestic demand show proportional relative variations.
- Ratios entering in behavioral equations.
- Equations using elements at current prices, computed as the product of a quantity by a deflator (which shows the evolution of the price compared to a base year). For example, the trade balance will be obtained as the difference between the products of exports and imports at constant prices by their respective deflators.

Sometimes this distinction is purely formal, and an adequate variable change will allow the return to a linear formulation. However, if we take into account the whole model, replacing by its logarithm a variable computed in elasticities will only transfer the problem if the level appears also in the model.

Thus in our general example, if one uses for the exports equation the formulation:

$$\text{Log}(X) = a \cdot \text{Log}(WD) + b$$

one can very well introduce variables $LX = \text{Log}(X)$ and $LWD = \text{Log}(WD)$, which will make the equation linear:

$$LX = a \cdot LWD + b$$

But it will be necessary, to introduce exports in the supply - demand equilibrium:

$$Q + M = FD + X$$

to add the non - linear equation

$$X = \text{Exp}(LX)$$

Therefore, most economic models presenting a minimum of realism will not be linear. But numerical computations will generally show that even for models including many formal non - linearities, the approximation by a linearized form around a model solution (denoted by an asterisk):

$$(\partial_t / \partial y_t)(y_t - y_t^*) + (\partial_t / \partial y_{t-1})(y_{t-1} - y_{t-1}^*) + (\partial_t / \partial x_t)(x_t - x_t^*) = 0$$

is acceptable for general purposes.

On the other hand the stability of the derivatives with time is much more questionable.

Let us suppose the formulation for imports is:

$$\text{Log}(M_t) = a \cdot \text{Log}(FD_t) + b$$

Linearizing it around a particular solution (noted *), we get

$$(M_t - M_t^*) / M_t^* = a \cdot (FD_t - FD_t^*) / FD_t^*$$

Or

$$(M_t - M_t^*) = a \cdot M_t^* / FD_t^* \cdot (FD_t - FD_t^*)$$

which will represent an adequate linear approximation of the connection between M and FD, provided that M and FD do not move too far away from their base value¹⁸. This base value might represent a reference path, from which actual values differ due to a change in assumptions.

But, if we restrict further the expression to a constant influence (linearity to constant coefficients),

$$(M_t - M_t^*) = a \cdot (FD_t - FD_t^*)$$

the approximation can be accepted only if the ratio M / FD does not change too much with time. This is not generally true: the expansion of international trade has led, and still leads, to a sustained growth of the share of imports in domestic demand, for most countries. The ratio M * / FD * will grow strongly with time, and the last formulation will be quite inadequate for forecasts.

1.2.9 OTHER PROPERTIES

1.2.9.1 Continuity

We consider here the continuity of the whole set of endogenous variables relative to assumptions (exogenous variables, parameters). It is almost never verified formally, but should only be considered within the set of acceptable solutions (and assumptions).

For instance, most models use ratios, which is acceptable if the denominator can never become null (like the productivity of labor measured as the ratio of production to employment). Or using logarithms to link imports to demand requires (logically) that those elements are strictly positive. In other words, a fully linear model can produce a negative GDP, but this does not make it less operational if this value is associated with absurd assumptions or coefficients.

So even if all models show non-continuity potential, it should never occur in practice. We can think of only three cases:

- The model framework is correct but something is wrong with its elements: the numerical assumptions, the estimated coefficients.
- The algorithm used for solving the model leads to absurd values (more on this later).
- The behavioral equations are wrongly specified. As we also shall see later, it can be dangerous to put together elements without a previous assessment of the associated mechanisms (for instance using logarithms as a natural solution).

It is necessary, however, to distinguish these absurd cases from those where the discontinuity applies to the derivative of a variable differentiable by pieces, as we are going to see in the following paragraph.

¹⁸ In other words, if the terms of the derivative are negligible beyond the first order.

1.2.9.2 Differentiability

It is less necessary, but its absence can lead to problems in the system solving phase, as well as in the interpretation of results.

Separating from the previous criteria is not always straightforward, as the non-derivability of one variable can correspond to the discontinuity of another: a discontinuous marginal productivity can make the associated production non-differentiable at points of discontinuity.

Returning to the example, we could formalize household consumption in the following manner:

- They receive a constant share - a - of production Q .
- Under an income threshold - R - they consume a share c_0 .
- On the supplement they consume a share c_1 .

The consumption equation will become:

$$CO_t = c_0 \cdot a \cdot Q_t + (c_1 - c_0) \cdot \max(0, (a \cdot Q_t - R_t))$$

At the point $Q = R / a$, CO is not differentiable (the derivative to the left is $c_0 \cdot a$, to the right $c_1 \cdot a$). And the sensitivity of consumption to income is not continuous.

This derivative is not purely formal: it defines the marginal propensity to consume (consumption associated to a unitary income increase), which can appear itself in the model, at least as a descriptive element.

At the household level, the evolution of income tax as a function of revenue (with rates associated to brackets) would represent another example, determining disposable household income.

1.2.9.3 Existence of a solution

It is obviously necessary for the model to have a solution, at least when it is provided with acceptable assumptions¹⁹. But the potential absence of a solution is present in many formal systems, including linear models. This absence of solution is generally logically equivalent to the existence of an absurd solution, as one can illustrate on the following case.

Let us consider a model with $n+1$ endogenous variables: X (dimension n) and x (a single variable). We shall describe it as f , a vector of formulas (dimension $n+1$), in which x appears as an argument of a logarithm,

$$f(x, X, \log(x)) = 0$$

¹⁹ Refusing to provide a solution for absurd assumptions should rather be considered as a quality.

If none of the positive values of x ensures the solution of the complete model, it has no solution.

In other words, taking the argument of the logarithm as a parameter α

$$f(x, X, \log(\alpha)) = 0$$

and making it vary in \mathbb{R}^+ , solving the associated model on x and X never will provide a value of x equal to this parameter.

The model has obviously no solution.

But if the model builder has used a formulation in logarithms, he has probably not considered letting the argument take negative values. By replacing the logarithm by some other expression giving similar values, we would probably have obtained a solution. But if the variable remains negative, this solution would have been unacceptable.

To illustrate this case, we are going to reduce the usual model to a three equations version.

Production adapts to demand corrected by imports and exports, the last being exogenous:

$$[1] \quad Q + M = FD + X$$

as for demand, one supposes that its relative variations are proportional to those of production:

$$[2] \quad \text{Log}(FD) = a \cdot \text{Log}(Q) + b$$

And imports are a share of demand

$$[3] \quad M = c \cdot FD$$

Let us suppose that one has obtained by estimation in the past: $a = 1.05$ and $b > 0$, justified by a level and growth of demand generally superior to production, obviously associated to imports greater (and growing faster) than exports.

Now, let us produce a forecast.

The model can be reduced into:

$$[1'] \quad FD = Q/(1-c) - X$$

(from (1) and (3))

$$[2'] \quad FD = Q^a \cdot \exp(b)$$

from (2)

and

$$[3'] \quad X / Q = (1/(1-c) - Q^{a-1} \cdot \exp(b))$$

Obviously, if Q grows (as $a-1 = 0.05$), the negative element will become eventually higher than the positive one, which means that Q can only be negative, which is impossible as it enters in a logarithm in equation (2). The model has no solution.

Of course, these mathematical observations have an economic counterpart. In the long run, final demand cannot grow continuously faster than production, if imports are a share of demand and exports are fixed. Assumptions, therefore, are not consistent with the estimated formula.

One will notice that the absence of solution is due here to the implicit adoption of a condition verified numerically on the past, but not guaranteed in general. This will be in practice the most frequent case.

1.2.9.4 Uniqueness of the solution

The uniqueness of the solution, for given (and reasonable) values of parameters and assumptions, is also very important. Indeed we do not see how one could use a model which leaves the choice between several solutions, except maybe if this freedom has a precise economic meaning.

In practice, most models are highly nonlinear if you look at the equations, but the linear approximation is rather accurate within the domain of economically acceptable solutions. This limits the possibility of multiple equilibriums: if the system was fully linear, and the associated matrixes regular, there would be indeed a single solution. However, as we move away from this domain, the quasi-linearity disappears, and we cannot eliminate the possibility of alternate solutions, probably far enough from the reasonable solution to appear quite absurd. Fortunately, if we start computations inside the domain, an efficient algorithm will converge to the acceptable equilibrium, and we will never even know about any other.

The most significant exception will be that of optimization models, which look for values of variables giving the best result for a given objective (for example the set of tax decreases which will produce the highest decrease in

unemployment, given a certain cost): if several combinations of values give a result equal in quality²⁰, this lack of determination will not undermine the significance of the solution. The existence of several (or an infinity of) solutions will represent an economic diagnosis, which will have to be interpreted in economic terms²¹.

Another case appears when the formula represents the inversion of another formula giving a constant value, at least on a certain interval. For example, if over a certain threshold of income households save all of it:

$$CO = \min(f(Q), CO^*)$$

Then the income level associated with CO^* will represent the total set of values higher than the threshold.

In the general case, the main danger appears in sensitivity studies: if one wants to measure and interpret the economic effects of a modification of assumptions, the existence of a unique reference simulation is an absolute necessity.

Finally, finding several solutions very close to each other might come from purely numerical problems, due to the imprecision of the algorithm: any element of the set can then be accepted, if the difference is sufficiently low.

1.2.9.5 Convexity (or concavity)

The convexity of the system, that is the convexity of the evolution of each endogenous variable with each exogenous variable and parameter taken individually (or of a linear combination of them), can be requested by some algorithms, especially in optimization. In practice it is very difficult to establish, and even rarely verified. At any rate, this characteristic is linked to the definition of variables, and a single change of variables might make it disappear.

1.2.10 CONSTRAINTS THE MODEL MUST MEET

In addition to its theoretical validity, the model will have to meet a set of more technical constraints.

1.2.10.1 Global compatibility

Constraints of compatibility will bear in practice:

a - **on the endogenous between themselves**: one cannot let the model compute variables independently if they are linked by a logical relationship, accounting or theoretical. For example, if the consumer price enters in the determination of the wage rate, it also will have to be influenced directly by the (estimated) price of local production. Or the employment level has to affect household revenue and consumption through a sequence of links.

Accounting balances must be verified: once household revenue has been computed as a sum of elements, an increase in consumption must produce the associated decrease in savings.

²⁰ For instance if the model is too simple to differentiate the role of two taxes.

²¹ provided the algorithm used for solving the model is able to manage this indetermination.

Maybe the most important issue lies with the « supply = demand » identity, which will have to be enforced both in at constant and current prices. This can lead either to use one of its elements to balance the equation, or to distribute the residual over the global set of elements on one side. By formulating total supply and demand as:

$$O = \sum_{i=1}^n O_i$$

And

$$D = \sum_{j=1}^m D_j$$

One will use for instance, either

$$D_m = O - \sum_{j=1}^{m-1} D_j$$

Or one will correct the set of demand variables by multiplying each of them by the uncorrected ratio O / D .

In most cases the equilibrium at constant prices will be enforced automatically. It can be written as:

$$\text{Local production} + \text{Imports} = \text{Local demand} + \text{Exports}$$

Or identifying intermediate consumption:

$$\text{Local GDP} + \text{Intermediate consumption} + \text{Imports} = \text{Local final demand} + \text{Intermediate consumption} + \text{Exports}$$

- With only one product, intermediate consumption can be discarded, and one will generally use the equation to compute GDP, controlling that it does not get higher than productive capacity²².

²² This can be obtained by a share of imports growing with constraints on local productive capacity.

- With several products, we must consider as many equilibrium equations, in which the supply of intermediate consumption goods sums inputs needed for production of the good, and the demand for intermediate consumption goods sums the intermediate uses of the good itself.

$$Q_i + \sum_j IC_{j,i} + M_i = FD_i + \sum_j IC_{i,j} + X_i$$

If we suppose that returns to scale are constant, the vector of value added by good will come from a matrix transformation. The constraint on capacity will be achieved in the same way as above (provided a capacity equation can be obtained).

Defining $c_{i,j}$ as the quantity of good i needed to produce one unit of good j , we get:

$$Q_i + \sum_j c_{j,i} Q_j + M_i = FD_i + \sum_j c_{i,j} Q_j + X_i$$

Or in matrix terms

$$Q + C \cdot Q + M = FD + {}^t C \cdot Q + X$$

or

$$Q = (I - C + {}^t C)^{-1} (FD + X - M)$$

Using this framework will automatically enforce the supply-demand equilibrium for all goods.

In practice, most of the problem comes from the equilibrium at current prices. If demand prices are computed individually using behavioral equations, there is no chance the equilibrium will be met. The process described earlier will in practice correct the prices. With S and D as supply and demand elements at constant prices, ps and pd as the associated deflators, we can compute the global values as:

$$SV = \sum_{i=1}^n ps_i S_i$$

$$DV = \sum_{j=1}^m pd_j D_j$$

The first option will compute a specific price

$$pd_m = (SV - \sum_{j=1}^{m-1} DV_j) / D_m$$

and the second

$$pd'_j = (SV / \sum_{j=1}^m pd_j D_j) pd_j$$

where the “pd” elements are the independently computed demand prices, and the “pd'” elements the corrected values.

The correcting factor:

$$r = SV / \sum_{j=1}^m pd_j D_j$$

can also be written as

$$r = r (SV / \sum_{j=1}^m pd'_j D_j)$$

which with

$$pd'_j = r pd_j$$

gives a set of equations ensuring the equilibrium. As “r” measures the potential discrepancy between supply and demand, one must check that it is not too different from one.

The following issues appear:

- With the first method, which element should be used to balance the system? The choice is between
 - A small and unimportant variable, to reduce the consequences for model properties; perhaps even a variable which has absolutely no influence on the rest of the model.
 - A variable with large value, to reduce the correcting factor
- The second method represents an extreme application of the first one, where all variables on one side are affected in the same proportional way.

Actually none of the solutions dominates clearly, the worst being in our sense the very first, which is the same as accepting de facto an imbalance, hidden but with potentially damaging consequences. Also, the second could be associated with a converging economic process, while the first can have no economic interpretation whatsoever.

In fact, one should concentrate on limiting the size of the correction itself. One could represent the problem as eliminating toxic waste: instead of storing it in a specific location (hidden or not), or spreading it all over the place, the best solution is clearly to reduce its production as much as possible. This means that the initially computed prices should be designed to give naturally close values to global supply and demand.

b - **on exogenous -> endogenous connections**: Connections must be formulated with care. For example, if the social contributions rate is defined as an exogenous variable in the model, it has to enter in all computations of contribution levels. In particular, it cannot coexist with an exogenous representation of contributions, or one using an estimated coefficient.

To avoid this type of error, a systematic study of model properties must be undertaken before any operational application: in our example, this would mean checking that an increase of the social contribution rate has all expected effects on State revenues as well as on the accounts and behaviors of other agents.

Also, the true exogenous concept should be decided. Concerning contributions, the decision variable is clearly its rate, while the associated revenue is influenced by endogenous prices and employment.

c - **on the exogenous between themselves**: one should avoid defining two variables as exogenous if they are linked (in any direction) by a logical relationship. If possible, one should endogenize one of them by formalizing this connection.

Let us suppose for example that a model for France uses two exogenous measures of the prices established by its foreign competitors: in foreign currency and in Euros (with a fixed exchange rate). To take into account an increase of foreign inflation, these two variables will have to be modified simultaneously. This is as best more complex, and can lead to errors if one is not careful enough, while it can be avoided simply by endogenizing the price in Euros as the product of the price in foreign currency by the (exogenous) exchange rate.

However, establishing such links is not always possible. For instance, in a national model, foreign prices and foreign production are exogenous, but also clearly influenced by each other. But the nature and importance of the link are highly variable. For instance, a decrease in foreign production can produce world deflation,²³ while inflation can reduce exports and production. To describe them completely one should have to resort to a foreign or world model. An intermediary solution could be to establish a set of linear multipliers linking these elements, but generally the model builder himself will take care of the problem by producing a set of consistent assumptions (with perhaps some help from specialists of the world economy, or from a separate model).

d - **on endogenous->exogenous connections**: they are obviously proscribed, because contrary to the preceding links the model builder cannot master them. They will be found in some models, however, through the presence of the following exogenous:

- Elements measured in constant terms, while they should change with economic activity.
- Deflators, which should depend on other deflators.
- Elements measured in current terms, for both reasons.

²³ This is the case for the MacSim world model we shall present later.

If the associated model can possibly produce correct estimates and even forecasts, it runs the risk of showing abnormal sensitivity properties. Let us take an example:

Let us suppose household income HI is composed

- Of the wage revenue, computed as the product of employment by the wage rate: $LT \cdot W$.
- of other exogenous revenues

Salaries will be indexed perfectly on prices:

$$W = WR \cdot CPI$$

One will have therefore:

$$RHI = WR \cdot CPI \cdot LT + HIQ$$

This equation might perform well in forecasts. But if a change in the assumptions makes prices increase, the purchasing power of total wages will remain unchanged, but for the complement HIQ it will be reduced in the same proportion as the price rise:

$$\Delta(HIQ / CPI) = -(\Delta CPI / CPI) \cdot (HIQ / CPI)$$

One can question this assumption. Some elements in non-wage revenue (social benefits, rents, firm owner's profits, independent workers revenue) are more or less indexed, and can even be over indexed in the case of interests payments (the interest rate should increase with inflation). Others, associated to differed payments (dividends, income tax) will not change immediately. The global sensitivity to prices is not clear, but a null value is obviously not correct.

We will face the same problem with a change in GDP:

$$\Delta(HIQ / Q) = -(\Delta Q / Q) \cdot (HIQ / Q)$$

where we cannot suppose that revenue does not change (grow) with economic activity. Some elements do not, or show a limited sensitivity (pensions) but dividends and the revenue of owners of small firms certainly do.

In conclusion, even when a variable measured at current prices has no theoretical content, it should not be kept exogenous, especially if it can be supposed to grow at constant prices. It is general better to consider as exogenous its ratio to another variable, supposed to follow the same trend (in the absence of idea, one can use plain GDP). The model equation will compute the variable by applying the exogenous ratio. This is also can be valid for variables at constant prices (which generally increase with production), to the exception of decision variables identified as such.

In the case above, one could write:

$$HI = WR \cdot CPI \cdot LT + CPI \cdot Q \cdot r_{hiq}$$

in which the introduction of Q links additional revenue to the global growth of the economy.

1.2.10.2 Homogeneity

If some equations in a model do not meet homogeneity constraints, this endangers its properties, particularly its sensitivity to shocks. Let us quote some cases:

- Linear relationships between values and quantities. The equation:

CO (consumption at constant prices) = a HRI (current income) + b is not only absurd from a theoretical viewpoint, but will lead in the long term to a level of savings

$$S = HRI - CPI \cdot (a \cdot HRI + b)$$

that will become clearly negative over a certain price level.

- Mixing logarithms and levels. Similarly, the equation:

$$CO = a \cdot \text{Log}(HRI) + b$$

(this time the two elements will be measured in quantities) makes the ratio CO / HRI decrease to 0, and therefore the savings rate to 1, when HRI grows indefinitely.

This last example shows however a limit to the argument: on short periods the equation can present a satisfactory adjustment, as the consumption to income ratio (propensity to consume, complement to 1 of the savings rate) decreases effectively with income. It is the speed of the decrease, and its long-term evolution, that is questioned here.

1.2.10.3 Constants with dimension

The problem is identical to that of the exogenous with dimension. It invites a careful study of the theoretical content of the constant. Furthermore, as most variables grow with time, the influence of the constant will generally decrease or even disappear in practice. We shall address this issue later, on a practical case.

1.2.11 NORMALIZATION AND IDENTIFICATION

Once equations are estimated, the problem of normalization remains. We have seen that very often the estimated formula will not explain a variable, but an expression (logarithm, growth rate, ratio, or a more complex expression). But some simulation algorithms will request a model a specific form, called "identified", in which a single untransformed variables appears on the left hand side:

$$y_t = f_t(y_t, y_{t-1}, x_t, a, u_t)$$

This means the model builder might have, after estimation, to transform the formulation: this operation is called the normalization of the model.

The advantage is double:

- The application of some solution algorithms is made easier. In some cases (Gauss-Seidel), this form is actually requested.
- This type of formulation allows a better interpretation of the process determining the equilibrium, provided each equation can be interpreted as a causal relation. If the equation describes a behavior, the economist should have placed to the left the element it is supposed to determine, conditional on the elements on the right. This is what we can (and will) do naturally in our example. For instance, the equation describing the choice by households of their consumption level will place naturally the variable "consumption" to the left.

The vast majority of equations will take naturally an identified form. Sometimes, a simple transformation will be necessary, however. Perhaps the most frequent nonlinear operator is the logarithm, associated with the integration of a formula in elasticities.

$$dx/x = f(\dots)$$

represents

$$\text{Log}(x) = \int f(\dots) \cdot dx$$

In this case, one just needs to replace:

$$\text{Log}(x) = f(\dots)$$

by

$$x = \exp(f(\dots))$$

If you use EViews²⁴, the software will do it for you. You can write the equation using the first form, and the package will normalize the equation itself, computing x . This is also true if the left-hand element contains several variables, but allows straightforward normalization. The most frequent cases are:

A change in logarithm: $\text{Log}(x_t / x_{t-1}) = f(\dots)$

A growth rate: $(x_t - x_{t-1}) / x_{t-1} = f(\dots)$

A ratio: $x_t / y_t = f(\dots)$

To choose which variable to compute, EViews will take the first variable in the specification of the equation. This simple method will be applied even if the variable has been identified as computed by a previous equation. For instance in our model, if we introduce the estimation of imports M , then state:

$$M + Q = FD + X$$

EViews will give an error message, as M appears to be computed twice.

Moreover, when an equation is forecasted individually, one can choose between the computation of the left hand term and the element which determines it, for instance M or $\Delta \log(M)$ for our imports equation.

However, EViews does not solve analytically any equation for the variable. For instance:

$$M / (Q + M) = f(\dots)$$

will be translated into:

$$M = (Q + M) \cdot f(\dots)$$

introducing a non-recursive process over M .

In any event, normalizing the general equation

²⁴ Or most packages of the same type.

$$f(y, \dots) = 0$$

is possible by adding on both sides the same variable, which gives:

$$y = y + f(y, \dots)$$

However the convergence of a model defined in this manner is often difficult to obtain (for instance if “ f ” is positively linked to y). In that case, one can use (the value for “ a ” can be negative):

$$y = y + a \cdot f(y, \dots)$$

Stronger simplifications are sometimes possible and will be approached with the numerical solution process.

Identification is not always economically straightforward: in our example, when balancing demand and supply, we can observe that three last variables (Final demand, Exports and Imports) are going to be determined by their own equation (the sum of its elements for the first, estimated equations for the others). This means that balancing must be done through GDP, and we must write the equation as:

$$Q + M = FD + X$$

or

$$Q = (FD - M) + X$$

which makes its theoretical content clearer as: production must (and can) satisfy both exports and the non-imported part of domestic demand.

1.2.12 CONCLUSION

It must be clear by now that the formal definition of the whole set of equations represents with the estimation of behavioral equations an iterative and simultaneous process:

- Behavioral equations start from an initial theoretical formulation to evolve gradually to their final form by reconciling this theory with data and estimation results.
- Accounting equations have been defined as precisely as possible in the preliminary phase, to establish a coherent framework, but they often will have to adapt to the evolution of behavioral equations. Let us suppose for example that the first estimation results suggest excluding from the econometric explanation of exports their agricultural component, setting it as exogenous: a new equation and variable will appear, and the equation for total exports will become an identity.

CHAPTER 2: MODEL APPLICATIONS

We shall now give a panorama of applications using models. Comments will be centered on the example of economic models, and more particularly on the macro-economic ones. But most of the observations can be transposed to the general case.

For each of these applications, technical details shall be left to the "implementation" part (chapter 7). To understand these practical aspects of the use of models, one must first know about the way they are built, described later in chapters 4 to 8.

2.1 OPERATIONAL DIAGNOSES

The most natural use of a model seems to be the evaluation of the economic future, whether as its most probable evolution or as the consequences of some decisions. Assumptions concerning the future will be injected into the model, and its solution will produce the requested diagnosis. Thus one will seek to anticipate the evolution of the main aggregates of the French economy until the year 2020, taking into account assumptions on the evolution of international economy.

2.1.1 DIFFERENT TYPES OF DIAGNOSES: SCENARIOS AND SHOCKS

Two types of forecasts can be considered: **scenarios** and **shocks**.

- In a **scenario**, one is interested in absolute results, and associating to a full set of assumptions a future evolution of the economic equilibrium. One might seek to obtain
 - forecasts on the basis of most probable assumptions
 - forecasts associated to a given set (like a party's program)
 - an evaluation of the scope of potential evolutions
 - assumptions allowing to reach specific economic targets.
- On the contrary, with a **shock**, one starts from a base simulation (often called "reference forecast" or "baseline"), or a simulation on the historical period, and measures the sensitivity of the economic equilibrium to a change of assumptions. Two economic paths will then be compared (on the past, one of them can be the historical one).

These shocks can be more or less complex, from the modification of a single assumption to the testing of a new economic policy²⁵.

These two techniques, scenarios and shocks, before the production of any operational policy diagnosis, will play an important role in the model validation process.

²⁵ However, this new policy should stay within the economic framework of the original model.

2.1.2 ADVANTAGES OF MODELS

Now that we have described the characteristics of models and their basic use, we shall discuss the advantages they bring (and their failings too).

Relative to the diagnosis provided by a human expert, advantages common to all models will:

- Guarantee the accounting coherence of the resulting equilibrium.
- Take into account a practically unlimited number of interdependent influences.
- Provide an explicit formalization of behaviors, allowing an external user to interpret them.
- Produce an exact and instantaneous computation of associated formulas.
- Adapt immediately the full system to a local change of theoretical formulation.

but also

- Allow the stability of reasoning, for human users of an unchanged model.
- Provide the possibility of formal comparisons with other models.

This forecasting ambition was already the basis for the construction of the first models. But this type of use has benefited (since the 1970s) from some evolutions:

- The progress of economic theory, allowing the formalization of more sophisticated mechanisms, better adapted to the observed reality.
- The progress of econometrics, giving access to the statistical method that will produce the most reliable formulation associated with a given problem, and to test more complex assumptions.
- The improvement of numerical algorithms, both for computation speed, and solving more complex systems.
- The simultaneous improvement of computation hardware allowing to process problems of growing size, by increasingly complex methods.
- The progress of modelling science, in producing models better adapted to the original problem, facilitating the production of assumptions, and reducing the cost of reaching acceptable solutions.
- The production of computer software specialized in model building, increasingly efficient, user-friendly, and connected with other packages.

- The improvement of the reliability of data, and the growth of the available sample, regarding both the scope of series and the number of observations (years and periodicity)²⁶.
- The easier communication between modellers, through direct contact and forums, allowing to communicate ideas, programs and methods, and to get the solution to small and large problems already addressed by others.

2.1.3 A CERTAIN REASSESSING

However, the use of models has engendered criticism from the start, using often the term « black box », describing the difficulty in controlling and understanding a set of mechanisms often individually simple but globally very complex.

In recent decades criticism has mounted, to the point of calling for a global rejection of traditional (“structural”) models. Surprisingly, critics often find their arguments in the above improvements. One can find:

A utilitarian critique: models have proven unable to correctly anticipate the future. If this observation has appeared (in the beginning of the eighties), it is obviously not because the quality of models has declined. But information on model performance is more accessible (some systematic studies have been produced), and the fluctuations following the first oil shock have made forecasting more difficult. In periods of sustained and regular growth, extrapolating a tendency is very easy for experts as well as for models.²⁷ Paradoxically, the emergence of this criticism has followed, rather than preceded, the increasingly direct intervention of model builders and their partners in forecasting results.

An econometric critique: modern techniques require a quantity and a quality of observations that available samples have not followed. A gap has opened between estimation methods judged by econometrics theoreticians as the only ones acceptable, and methods really applicable to a model²⁸.

A theoretical critique: the development of economic theory often leads to sophisticated formulations that available information have difficulty to validate. And in any event many areas present several alternate theories, between which any choice runs the risk of being criticized by a majority of economists. Thus in the monetary area, going beyond a basic framework leads to rely on information unavailable in practice, or on formulations too complex to be estimated.

A mixed critique: users of models are no longer passive clients. They criticize formulations, as to their estimated specification, or their numerical properties. This evolution is paradoxically favored by the improvement of the logical interpretation of economic mechanisms, itself fathered essentially by economic knowledge (even the economic magazine articles use implicit macroeconomic relations) and modelling practice (the population of clients includes more

²⁶ However, the size of samples does not necessarily grow with time. In a system of national accounts, the base year has to be changed from time to time, and the old data is not necessarily converted.

²⁷ One could argue that recent years have presented a regular degradation of the activity. But apart from the fact that this observation is disputable, forecasters have often the temptation to anticipate an exit from a crisis, leading to a forecasting bias.

²⁸ Actually, the sample size required by present techniques (50 or better 100 observations) limits the possibility of estimating equations using deflators or variables at constant prices. Even using quarterly data, separating values into prices and volumes is quite questionable 15 years from the base period.

and more previous model builders or at least followers of courses on modelling). One could say that model users ask the tool to go beyond their own spontaneous diagnosis, and they want this additional information to be justified.

It is clear that these criticisms grow in relevance as the goal grows in ambition. Forecasts are more vulnerable than simple indicative projections, which seek to cover the field of the possible evolutions. As for policy shock studies, they are not prone to errors on the baseline assumptions, if we discount non-linearities²⁹.

This relevancy also will depend on credit granted to results. One can use figures as such, or be content with orders of magnitude, or even simply seek to better understand global economic mechanisms by locating the most influential interactions (possibly involving complex causal chains). In our sense, it is in this last aspect that the use of models is the most fruitful and the least disputable³⁰.

2.2 THEORETICAL MODELS

Contrary to previous models, theoretical models may be built for the single purpose of formalizing an economic theory. It may be sufficient to write their equations, associating to a theoretical behavior a coherent and complete system. Reproducing the observed reality is not the main goal of these models, and it is not mandatory to estimate parameters: one can choose an arbitrary value, often dictated by the theory itself. In fact, this estimation will often be technically impossible, when some of the variables used are not observed statistically (the goals or expectations of agents for example).

However, even based on an artificial series and arbitrary parameters, the numerical simulation of these models can be interesting. Actually, the formulas are often so complex that solving the model numerically will be necessary to observe its solutions as well as properties (such as the sensitivity of solutions to assumptions and to coefficients).

2.3 QUANTIFIED SMALL MODELS

2.3.1 WITH SCIENTIFIC PURPOSE

These models represent an intermediate case. One seeks a realistic representation of the economy, adapted to observed reality, but sufficiently simple to accept the application of complex analysis methods (and the interpretation of their results). In addition to scientific research, this study can be done to measure and to analyze properties of an operational model on a simplified representation (in the eighties MiniDMS, then MicroDMS have been used to characterize the Dynamic Multi Sectorial model of INSEE).

There are two categories of methods:

²⁹ With a linear model, the consequence of a shock depends only on its size, not on the simulation it starts from.

³⁰ One example is the impact of a decrease in local tariffs. Ex ante it increases imports (a negative demand shock). Ex post it decreases local factor costs (with cheaper investment and cheaper labor, indexed on a lower consumption price). This leads to more local capacity and competitiveness, both on the local scene (limiting the imports increase), and the foreign one. In most models, GDP decreases then grows.

The full interpretation of such a shock provides a lot of information, even if one remains at the non-quantitative level.

- “External” methods will use model simulations to observe its quantitative properties, and infer a descriptive comment, both statistical and economic.
- “Internal” methods seek to explain properties of the model by its structural characteristics, using mathematical tools. This does not necessarily call for actual simulations.

2.3.2 WITH AN EDUCATIONAL PURPOSE

Although often of the same type as the ones above, these models try to present economic mechanisms as complete as possible, based on real data, under an interpretable and concise form. If necessary, one will favor the message contained in the presentation over the respect of statistical criteria.

This is the case of the MacSim package, allowing students to interpret international mechanisms and interactions.

CHAPTER 3: MODEL TYPES

We shall now try to establish a classification of models, focusing on the link between the model's characteristics and the goal for which it has been built.

3.1 THE FIELD

The field described by a model is characterized by the variables it computes, but also by assumptions it takes into account.

In the economic model subset, we can consider:

- A geographical field: national models, multinational models, world models. These last can be built in two ways: by putting together preexisting national models, with potentially quite different structures, or by building simultaneously country models of identical structure, possibly with a single team of modellers. We shall deal with this later.
- A theoretical field: the theory used for the formalization of the model may or may not approach specific economic aspects. A Keynesian model might limit the treatment of monetary aspects. A short-term model will not formalize demographic evolutions.
- A field of units: a model might present only variables at constant prices, or physical quantities like barrels of oil or number of pigs.
- A field of agents: a model will describe the behavior of a single agent: households, the State, firms.
- A field of goods: a model might consider only the production and the consumption of one good, for example energy. An energy model can use physical units.

There are other types of fields. However, the distinction is not always easy: some models will describe summarily a global field, except for a certain aspect on which it will concentrate. An energy model, to take into account interactions with the rest of the economy, will have to model it also, but not in the same detail. And it can mix physical units (barrels of oil or gigawatts) with national accounts elements.

On the other hand, it always will be possible, and made easier by some modelling packages, to change (actually to restrict) at the time of simulation the scope of the model. The distinction is then no longer permanent: a multi-national model can be used to simulate a complete evolution of the world economy, but its user can also restrict calculations to the evolution of a group of countries or even a single one, the other elements being fixed. One can simulate a model of the real economy with or without additional monetary features. Or a model using normally rational expectation elements can drop them to become purely backward looking.

3.2 THE SIZE

The history of modelling shows that for a long period new models generally have seen their size grow, for the reasons cited earlier: the progress of model-building techniques, the increased availability of data, the faster computer computations. Additionally, for any given model, size increases regularly in the course of its existence, as new team members want to add their contribution.

However, the last decades have seen a trend favoring a return to models of limited size. Productivity improvements, requested from teams of model builders, are less and less compatible with the utilization of a large model. Despite the progress of model-building techniques, the desire to reduce costs and delays conflicts with the size, especially (but not only) regarding human operations: elaboration of assumptions and interpretation of results.

Also, the use of a very detailed model can make individual estimations and specifications look too expensive. The attractiveness of a calibrated and gemellar CGE model will increase.

Finally, the desire to reply to critics comparing models to "black boxes" leads model builders to look for more explicit and manageable instruments.

3.2.1 DETERMINANTS OF THE SIZE

Determinants of the size of the model will be:

- The size of the field covered (see above).
- The degree of aggregation, which can be
 - vertical: number of operations taken into account (for example one can distinguish several types of subsidies, or social benefits),
 - or horizontal: number of agents listed; one can distinguish more or less sectors of firms, or types of households.

The degree of aggregation will not be inevitably uniform: an energy model will use a particularly fine detail for energy products.

In fact the same model can appear under several versions of different size, depending especially on the degree of aggregation. Each version has then its proper area of utilization: detailed forecasts, quick simulations, mathematical analysis, and educational uses.

Thus at the end of the 1980s, the 3000 equation D.M.S model (Dynamics Multi Sectorial) used by INSEE for its medium-term forecasts had two companion versions of reduced size: Mini - DMS (200 equations), used for some operational projections and analysis which did not require detailed products, and Micro - DMS (45 equations), with an essentially educational purpose.

This distinction has lost most of its validity, however, following the reduction of the size of operational models.

3.2.2 A CLASSIFICATION

We propose the following classification, necessarily subjective:

- Small models: 1 to 50 equations

Examples: Klein-Goldberger, Micro - DMS (INSEE), D.M.M. (CEPREMAP), FAIR model.

- Average models: 150 to 400 equations

Examples: MULTIMOD (IMF), TESTUS (Federal Reserve Board), AMADEUS (INSEE), HERMES (Ecole Centrale de Paris), MESANGE (INSEE and French Ministry of Finance), MZE (INSEE and French Ministry of Finance)

- Large models: 800 to 2000 equations

Example: TESTMCM (Federal Reserve Board), METRIC (Direction de la Prévision), MOSAIQUE (OFCE), MEFISTO (Bank of France), MULTIMOD (IMF)

- Very large models: more than 4000 equations.

Example: NiGEM (NIESR), INTERLINK (OECD), the LINK model (United Nations + University of Toronto + partners), MacSim, MIMOSA (CEPII-OFCE). Today this last category should apply only to international modelling.

Gaps have been left for intermediary cases.

In addition, a number of models built by the author in the course of economic cooperation are listed in the study on algorithm efficiency, presented later. Current projects include China, Vietnam, Algeria, and Morocco. Past projects have concerned Poland, Slovakia, the Czech Republic, Lithuania, Ukraine, Tunisia, the Andean Community, Argentina, Kazakhstan, and Tajikistan.

3.3 THE HORIZON

3.3.1 FOR FORECASTING³¹

If a model is designed for forecasting, its horizon will be defined at the construction of the model. It will be strongly linked to its general philosophy and to the set of mechanisms it implements. A long-term model will be little interested in circumstantial phenomena (such as the lags in the adjustment of wages to prices), while a short-term one will not take into account the longest trends (such as the influence of the economic situation on demography).

³¹ One shall notice that we can use several words to characterize these exercises: forecasts, projections, scenarios, simulations. It all depends on the purpose for which the test was made, and perhaps the trust allowed to the results. We favor the last term, which unfortunately has to be completed into: « simulation over future periods ».

These differences seem to discard elaborating a model that can be used for both short - and long-term projections. But we shall see that strong reasons, in particular econometric, have made this option appear as the most natural in the present situation. We will develop them when we address periodicity, in paragraph 3.4.

In any case, one can find a certain asymmetry in the relevance of this observation. If long-term models can neglect intermediate periods if they do not show significant fluctuations, simulation of the periods beyond the operational horizon can evidence future problems, already present but not visible in the short term

3.3.2 FOR MODEL ANALYSIS

Here, the horizon depends on the type of analysis one wants to produce. Often, to analyze a model built with a given forecasting horizon, simulation over a longer period must be obtained. Even more than for forecasts, analytic shocks will show and explain anomalies that were not apparent in the normal projection period, but had already a significantly harmful influence. We shall stress these issues later.

3.3.3 A CLASSIFICATION

One could use the following classification:

- Short-term models: 1 quarter to 2 years.
- Medium-term models: 4 to 7 years.
- Long-term models: 10 years and more.

Obviously, for a dynamic simulation, the full path, including intermediate values, is of interest.

3.4 THE PERIODICITY

The periodicity of a model is linked to the mechanisms it seeks to study and therefore to its horizon.

Short-term models demand a short periodicity to take into account circumstantial phenomena: delays of the wage indexation on prices, progressive adjustment of the consumption level to an increase of income.

Long-term models can use a sparser periodicity, less for theoretical reasons (long-term behavior can be described by a short-periodicity model), than for technical ones: this choice will reduce constraints on the availability of series, facilitate the production of assumptions, and limit simulation costs.

However, we shall see that the use of “modern” econometrics methods calls for a short periodicity, for all kinds of models, as soon as estimations are considered.

This means that the main determinant of model periodicity comes from the data. Countries which produce quarterly national accounts use quarterly models, which allow them to apply modern techniques with some comfort, and produce both short and long term studies. When only yearly accounts are available, the techniques become more simplistic, and true short term applications are not possible. Unfortunately, this applies most often to countries with a short history of statistics, making the problem the harder.

3.5 OTHER MODELS

We have essentially concentrated on the macro-economic model case. One can also find:

- Micro-economic models: describing the behavior of firms, of households.

These models will sometimes be more theoretical, calling for optimization computations (cost minimization) or to elements of strategy (game theory). They will often be estimated on survey data.

- Non-economic models: they can apply to biology, physics, chemistry, astronomy, meteorology, ecology, process control, and so on.... and be used for evaluating the consequences of the building of a dam, controlling the functioning of a manufacturing process, looking for the best organization of a project, describing a biological process. These models will often be conceived not as a formalized equation system, but as the maximization of a criterion under some constraints, or as a system of propositions connected by logical operators.

CHAPTER 4: GENERAL ELEMENTS

This part of the book describes the process of development, utilization and update of a model, taking special interest in technical aspects and particularly computer-oriented features. Applications to EViews will be presented in detail, but most of the teachings can be applied to other packages, including those which are not dedicated to econometric structural modelling.

First, let us give a quick description of the organization of the model building process.

4.1 THE STAGES IN THE PROCESS

4.1.1 PREPARING THE MODEL

The first step in the building of any model is producing a draft which ensures some compatibility between available data (wherever it might come from) and the type of model its builder has in mind (goal, scope, nature of the variables, underlying theory).

Knowing the scope of available data, the economist will define a model framework for which values can be attributed to all variables, either using available elements or by computation. This means that a first decision has to be made as to the field described by the model, the variables used as assumptions, and the variables it shall compute. Moreover he must divide the equations into identities, which set indisputable links between variables, and equations describing the behavior of agents, for which the final formulation will be based on past evolutions of the associated elements.

The first task will be to gather, by reading from files and transforming the data, the full set of variables needed by the model, to define the form of the identities, and give a first assessment of the behaviors he intends to describe. He shall check for which periods the necessary data is known, and that on these periods identities hold true. If some elements are not available, he will use the best proxies he can get. And if this also fails, he will use his imagination.

He can also make a first economic analysis of the framework implied by model specifications (greatly helped by EViews).

4.1.2 ESTIMATION

The second phase will look for a satisfying description of the behavior of agents, by checking economic theory against available data. The modeller shall define alternate formulations with unknown parameters, compute for each formulation the values which give the best explanation of past evolutions, and make his selection, using as criteria both statistical tests and compliance to economic theory. This process can call for the introduction of new variables, or changes in some definitions, which will mean reformulating some identities.

4.1.3 SOLVING AND TESTING OVER THE PAST.

Once the full model is defined, one can try to solve it.

- One shall first check for consistency the set of equations, data and parameters, by applying each formula separately on the sample period. If the estimation residuals have been introduced as additional elements, the process should give the historical values in all cases.

- One shall then simulate the full model on the same period, setting (temporarily) the residuals to zero. This will show if taking into account current and lagged interactions does not amplify too much the estimation errors.
- Finally the reactions of the equilibrium to a change in assumptions, for instance the exogenous component of demand, will be measured. The results will be compared with the teachings of economic theory, and what is known of values given by other models. However, one should not spend too time here, as simulations over the future will provide a much better context.

Discovering discrepancies can lead to changes in some elements of the model, including the set of its variables. This means going back to step 1 or 2.

4.1.4 SOLVING AND TESTING OVER THE FUTURE

Once the model has passed all tests on the past, further tests will be conducted, under conditions more representative of its actual use: on the future. For this values will have to be established for future assumptions. Again, the sensitivity of the model to shocks will be studied, this time with a longer and smoother base. As to the reliability of baseline results, one shall rely this time on stochastic simulations.

4.1.5 USING THE MODEL FOR FORECASTS AND POLICY STUDIES

Finally, the model will be considered as fit for economic studies: forecasts and economic policy analysis.

We shall suppose we are using a dedicated package like EViews (even if some people still model through a spreadsheet).

4.2 HOW TO ORGANIZE THE DEVELOPMENT OF THE MODEL

Let us now consider the organization of the model production task.

To create a model, two extreme types of organization can be considered:

- Methodical option:

The model builder

- Specifies completely a coherent model (including accounting equations), precisely separating assumptions from results.
- Looks for the necessary series.
- Estimates behavioral equations.
- Uses the consequent model.

Applying such a framework is obviously illusory, as many backtrackings will be necessary in practice:

- Some series will show up as unavailable, and it will be necessary to replace them or to eliminate them from formulations. Thus, in the absence of series for interests paid by firms one will have to be content with profits before interests.

- Some estimations will give unsatisfactory results: it will be necessary to change formulations, to use additional or alternate series. Thus, a formulation in levels might have to be replaced by a formulation in logarithms (constant elasticities), or in growth rates. Or one will be led to explain the average monthly wage instead of the hourly wage, and to introduce in this last explanation the evolution of the minimal wage. For an oil producing country, it will appear necessary to identify oil (and non-oil products) in both production and exports.³²
- New ideas will appear during estimation. For example, a recent article on the role of Foreign Direct Investment might lead to test an original formulation.
- Formal errors are going to be identified. Thus, an element (a type of pension) might have been forgotten from households' income.
- Some variables defined as assumptions are going to appear sufficiently influenced by results to see their status modified.
- Improvisation

To the contrary, a model builder can

- establish general options for the model structure and theoretical framework,
- produce some formulations independent from each other,
- estimate them by accessing to separate series,
- And gradually connect selected elements by completing the model with linking identities and the data set with the necessary exogenous variables.

This framework will be even less effective, if only because the number of single operations on equations and series will present a prohibitive cost. Furthermore, enforcing the accounting and theoretical coherence of the model could prove difficult, and the modelling process might never converge at all to a satisfying version.

- The optimal solution is of course intermediate:
 - Define as precisely as possible the field and the classification of the model.
 - Define its general theoretical options and its goal.
 - Obtain, create and store the total set of presumably useful series.
 - Establish domains to estimate, specify associated variables and set formal connections, especially accounting ones.
 - Undertake estimations
 - And go through changes (hopefully limited) until an acceptable form is obtained.

³² Actually, this should have been evident from the start.

It is clear that this type of organization is all the more easy to implement if:

- The size of the model is small: it is possible to memorize the total set of variable names for a thirty equations model, but for a large model a formal documentation will be necessary, produced from the start and updated regularly.
- The number of concerned persons is small (the distinction comes essentially between one and several): for a team project, the role of each participant and his area of responsibility have to be clearly defined. Especially, physical changes (on both data and model specifications) should be the responsibility of one individual, who will centralize requests and apply them. And modifications must be clearly announced and documented.

Individual modifications of the model can be allowed, however, provided a base version is preserved. Thus several members of a team of model builders can test, one a new production function, another an extended description of the financial sector. But even in this case updates will often interfere, at the time modifications generated in separate test versions are applied to the base one. For instance, a new definition of the costs of wages and investment, which define the optimal shares of labor and capital in the production function, will influence the target in the price equation.

CHAPTER 5: PREPARING THE MODEL

We shall now start with the first (and probably most important task): preparing the production of the model.

One might be tempted to start model production as soon as possible. But it is extremely important to spend enough time at the start evaluating the options and choosing a strategy. Realizing much later that he has chosen the wrong options, the modeller is faced by two bad solutions: continuing a process leading to a subpar model, or backtracking to the point where the choice was made.

This can concern

- The organization of tasks, like producing at first single country models, for a world modelling project.
- Economic issues, like choosing the complexity of the production function, or the decomposition of products.
- Technical ones, like the number of letters identifying the country in a world model series names.

5.1 PREPARING THE MODEL: THE FRAMEWORK

At the start of the model building process, the modeler (or the team) has at least:

- General ideas about the logic of the model he wants to build.
- Information about the set of available data.

Actually, things can be more advanced:

- The data can be directly available, almost always as a computer file, but not necessarily in the format needed by the modelling package.
- Equations may have already been established, either as formulas or even estimated items, if the modeling is the continuation of an econometric study.

In any case, the first stage in the process should lead to:

- A fully defined set of equations, except for the actual estimated formulas.
- The corresponding set of data.

Obviously, these two tasks are linked, as equations are established on the basis of available data, and the data is produced to fit the model equations. This means that they are normally processed in parallel. However, it is quite possible:

- To produce most of the data before the equations are defined. Some concepts (the supply - demand equilibrium at constant and current prices, employment, the interest rates) will certainly appear in the model. But some model-specific variables will have to wait.
- To produce the model specification before any data is available. Of course, writing an identity, or stating the equation to be estimated, does not require data. It is only the application (checking the identity is consistent, or estimating the equation) which does. But one must be reasonably sure that the data will be available, or that there will be a reasonable technique to estimate it.

One can even produce a first version of the program transforming into model concepts the original data, once these concepts are completely defined, but before any data is technically available (just their definition).

One can compare the situation with the building of a house: one can draw the plans before the equipment is bought, but its eventual availability (at the right time) must be certain. And the goods can be bought before the plans are completely drawn (but the chance of having to use them must be reasonably high)³³.

These options are not optimal in the general case, but they can help to gain time. Most modelling projects have a deadline, and once the work force is available, the tasks should be processed as soon as possible, if one wants to have the best chance of meeting it.

One can question the feasibility of producing a full set of equations before any estimation. What we propose is to replace the future formulations by a “declaration of intent” which states only the variable to be explained, and the elements which will explain it. For each equation, the format should be as close as possible to:

Variable =f(list of variables)

The advantages of defining a full model are numerous:

- The modeller will be able to check by sight the logic of his model.
- The text can be given to other economists for advice.
- The full list of requested variables can be established, allowing to produce a complete transfer program.
- Processing the equations through EViews will give interesting advice on several elements:
 - The equations:
 - * The grammatical acceptability of equations will be checked: for instance the number of left and right parentheses.
 - * Also the fact that each endogenous variable is computed only once.
 - The variables.
 - * The most important information will come from the list of exogenous: one might find elements which should have been determined by the model, according to its logic. In general, this will mean one has forgotten to state the associated equation. Also, some elements might appear, which should not belong to the model. Normally this corresponds to typing errors.
 - The block structure:

It decomposes the set of equations into a sequence of blocks, either recursive (each variable depends only on preceding elements) or simultaneous (some variables are used before they are computed). If one is going to succeed in estimating equations which follow the same logic as intended in the preliminary version, the block structure described at this stage will be already fully representative of the future one. One can detect:

³³ As there is a cost to the goods. For free or quasi-free data, the chance can be lowered.

- * Abnormal simultaneities: a causal loop might appear, which is not supported by economic theory behind the model.
- * Abnormal recursive links: a block of equations containing a theoretical loop (the wage price loop, the Keynesian cross) can appear as recursive. This can come from a forgotten equation, a typing error...

In any case, observing the causal structure of the model will give some preliminary information about its general logic, and its potential properties.

5.2 PREPARING THE MODEL: SPECIFIC DATA ISSUES

Let us detail the process.

5.2.1 TYPES OF DATA

In the case of a national macroeconomic model, the needed data can be:

- National Accounts elements: operations on goods and services, transfers between agents, measured in value, at constant prices, or at the prices of the previous year. The producer will generally be the national statistical office. For France it would be INSEE (the National Institute for Statistics and Economic Studies).
- The corresponding deflators.
- Their foreign equivalents, using the accounting system and the corresponding base year of the particular country, or rather a synthesis produced by an international organism (OECD, International Monetary Fund, EuroStat....).
- Variables in a greater detail, possibly measured in physical quantities (oil barrels, tons of rice). They can come from a public or private agency, or from the producers themselves. In France energy elements would come from the Observatory of Energy.
- Monetary and financial data, coming mostly from the local National Bank (in France the Bank of France or the European Central Bank. ...), from an international bank (EBRD, ADB), or from the International Monetary Fund.
- Data on employment or unemployment. One can get detailed labor statistics (by age, qualification, sex...) from the US Bureau of Labor or the French Ministère du Travail.
- Demographic data: population, population in age of working, age classes (INSEE in France).
- Survey data: growth and investment prospects according to firm managers, productive capacity, living conditions of households (coming from public or private institutes).
- Qualitative elements: the fact of belonging to a specific set, meeting a specific constraint.
- Micro economic models will generally use survey data (households, firms) with sometimes a time dimension (panels, cohorts) and possibly include some of the above elements as global indicators.

As the area of application of models is unlimited, the field of potentially relevant data is also. A model on the economy of transportation would include technical data on the railway system and on distances between cities, an agricultural model meteorological data and information on varieties of species.

5.2.2 THE ACCESS TO DATA

The medium through which data can be obtained will play an important role. Accessing the necessary data takes into account several features:

- the mode of transmission
- the format used
- the institutional aspects

We shall treat them in turn, then present the most usual cases.

5.2.2.1 The mode of transmission

Several options are available for transferring the data to the model.

5.2.2.1.1 Physical transmission

Data can be obtained from a physical support, either commercially produced or created for the purpose. This can be either a CD or DVD-ROM, or another rewritable media such as an USB key, or a memory card. For instance, INSEE provides CD-ROMs containing the full National Accounts.

5.2.2.1.2 E-mail transmission

Files can be transferred from a user to another by e-mail, as an attachment to a message.

5.2.2.1.3 Internet transmission

Files can be downloaded from a website, commercial or not. The INSEE site www.insee.fr allows the access to a set of national account series, in Excel or HTML formats.

Most of the time, these files are available in Excel format, the most frequently used for data treatment, and one for which every package on the market provides a simple interface.

5.2.2.1.4 Other media

In less and less frequent cases, some data will not be available in magnetic form: series will be found in printed or faxed documents, or obtained directly from other experts, or fixed by the user (who then plays the role of expert). This data will have generally to be entered by hand, although a direct interpretation by the computer through optical character recognition (OCR) is quite operational (but this technique needs documents of good quality).

In this case it is essential not to enter figures directly into the model file, but to create a preliminary file (such as an Excel sheet or even an ASCII file) from which the information will be read. This separates the modelling process from the production of "official" information.

5.2.2.2 Change of format

As indicated above, the original data format is generally different from the one used by the model-building software.

In the worst cases, transfer from one software program to another will call for the creation of an intermediate file in a given format, which the model-building software can interpret. The Excel format is the most natural intermediary, as it is read and produced by all packages. In that case, it is not necessary to own a copy of the package to use its format.

In the very worst cases, it is always possible to ask the first program to produce a character file (in the ASCII standard) which can, with minimal editing, be interpreted by the second program as the sequence of statements allowing the creation of the transferred series, including data and definitions³⁴.

However, the situation has improved in the last years, as more and more packages provide a direct access to the formats used by the most common software. For instance, EViews will create a workfile automatically from a list of 12 formats:

Access .mdb
Aremos .tds
Dbase.dbf
Excel .xls
Gauss .dat
Givewin.im7
HTML .htm, .html
Lotus 123 .wks, .wk1, .wk3
MicroTSP .wf

³⁴ For instance, the sequence:

```
use 1970 to 2007
```

```
read x
```

```
----- values ----
```

```
end
```

```
can be translated easily by a word processor into
```

```
smpl 1970 2007
```

```
series x
```

```
----- values ----- ;
```

ODBC .dsn, .dqy
Rats .rat, .trl
SAS (various)
SPSS .sav, .por
Stata .dta
Text .txt, .csv, .prn, .dat
TSP .tsp

5.2.2.3 Institutional issues

Of course, one must also consider the relationship between the data producing and modelling institutions. The most technically complex transfers do not necessarily occur between separate institutions. A commercial contract might give the modelling institution direct access (through a modem or the access to a global network) to information managed by a data producing firm, under the same software format, while a large institution might still use CD-ROMs as a medium between separate units.

However, one must also consider the cost of establishing contracts, including perhaps some bartering between data producing and study producing institutions.

5.2.2.4 How to cope with several sources

As a general principle, one should favor using a single source. But this is not always possible. In that case, one should define a primary source, and take from the alternate ones the only additional series. The main problems might come from:

- Deflators and values at constant prices using a different base year.
- Financial and labor data bases sharing elements with national accounts.
- Variables measured in physical units (tons, square meters) having their counterparts in values.

In all these cases, the priority is the consistency of model equations, based on the data from the primary source. Additional elements must be adjusted to provide this consistency. This applies in particular to the balance of equilibrium equations (supply = demand), or sums (total demand= sum of its components).

5.2.3 PREPARING THE DATA FOR THE TRANSFER

Let us now define the best organization for transferring data from the original source to the package (we shall use EViews as an example).

We have to guarantee several things:

- The original data must remain available
- It must be updated easily.
- Transfer must be as easy as possible.

To achieve these goals, the best organization should be:

- Copying the original file under another name.
- In the file, creating a new page.
- Copying the original series into this page, using “copy with link”

We shall suppose the original data is organized as a matrix (or a set of matrixes) with series either in lines or columns. If not, an additional intermediary phase can be needed³⁵.

- Insert a line of series names above the first period data (or a column left of the first column).

It does not matter if the matrix does not start in cell B2. Just insert as asked.

- Read it in EViews using the extended matrix (using import or copy).

This guarantees that:

- The original data is not modified.
- Updates are easy: just copy the new page into the original one (and drag cells in the second page if new observations have appeared).

The only change in the EViews transfer programs concerns the sample period.

5.2.4 THE PRELIMINARY PROCESSING OF SERIES

Very often the nature of available series is not really adapted to the needs of the model. A preliminary processing is then necessary. This can apply to several features.

5.2.4.1 Time transformations

Most of the time the series the model builder will access have the right periodicity. Individual exceptions can occur. New series will have to be computed (inside the modelling package).

The change can be undertaken in two directions: aggregating and disaggregating.

5.2.4.1.1 Aggregation

The easiest case happens if the available periodicity is too short. The nature of the variable will lead naturally to a method of aggregation, giving the exact value of the series:

If we note X_t the aggregated variable in t , and $x_{t,i}$ the variable of the sub - period i in t , we can consider the following techniques:

- Sum, for a flow (such as the production of a branch).

$$X_t = \sum_{i=1}^n x_{t,i}$$

- Average, for a level (such as unemployment in a given period).

³⁵ For instance, quarterly data can appear in yearly lines of four columns.

$$X_t = 1/n \sum_{i=1}^n x_{t,i}$$

- First or last value, for a level at a given date (for example the capital on the first day of a year will come from the first day of the first quarter). This will apply to stock variables.

$$X_t = x_{t,1}$$

or

$$X_t = x_{t,n}$$

5.2.4.1.2 Disaggregation

When moving to a shorter periodicity, the computation will need an approximation. One can consider:

- Dividing the total value by the number of sub-periods (for a flow).

$$x_{t,i} = 1/n X_t$$

- Distributing on each sub-period the total evolution observed on the period (for a level).

$$x_{t,i} = X_{t-1} + i/n \cdot (X_t - X_{t-1})$$

- computing the series that associates global observed evolution with constant growth rate by sub-period, for example with:

$$x_{t,i}/x_{t-1,n} = (X_t / X_{t-1})^{i/n}$$

- Filtering the series over the period, so that the transformed series is a moving average of the original one.

However, one must be aware that the two last methods do not guarantee that the sum of disaggregated values equals the global one. Most packages (including EViews) can achieve this constraint if requested (see the User's Manual for the available options).

5.2.4.1.3 Smoothing

Smoothing represents a particular case: preserving the same periodicity as the original series, but with the constraint of a regular evolution, for example a constant growth rate. Instead of n free values, the choice is reduced to the value of one (or maybe two) parameters.

We will be more or less brought back to disaggregation, except that we will be able to use the additional information. For example, it is now possible to estimate the constant-rate formulation that presents the smallest distance to the observed series. The respect of coherence constraints, for instance on the average of smoothed and original variables, will often lead to specific estimation methods (a normal regression would only establish this constraint on the transformed variables, growth rates for example).

More complex smoothing techniques can be applied, the most popular being the Hodrick-Prescott (for simple cases) and Kalman (more complex) filters.

5.2.4.1.4 Seasonal adjustment

As we explained before, one method for dealing with variables presenting a seasonality is to eliminate it, and work with seasonally adjusted series.

Several algorithms can be considered, the best known being probably Census X-12 (or Census X-13) and TRAMO-SEATS, both available in EViews.

Obviously, one should not mix original and adjusted series in the same set of model variables.

5.2.4.2 Change of classification

We have already considered this problem when we addressed the fields of models.

Changing categories will usually correspond to aggregation. In the case of economic models, this will apply essentially to:

- Economic agents: one might separate more or less precisely households categories (following their income, their occupation, their size...), types of firms (according to their size, the nature of their production...), Government institutions (central and local, social security, agencies...).
- Products (production can be described in more or less detail).
- Operations (one can separate social benefits by risk, by controlling agency, or consider only the global value).
- Geographical units (a world model can aggregate countries into zones).

5.2.4.3 Formal transformations

Some variables needed by the model will not be available as such, but will have to be computed from existing series by a mathematical formula. For example, the rate of use of production capacity will be defined as the ratio between effective production and capacity, coming possibly from different sources. Or the relative cost of wages and capital (used for defining the optimal production process) will take into account the price of the two factors, but also the interest rate, the depreciation rate, the expected evolution of wages, and some tax rates.

5.2.5 UPDATES

Once adapted to needs of the model builder, series often will have to be modified.

Changing the values of existing series can have several purposes:

- Correcting a formal error, made by the model builder or the producer of series: typing errors, or errors of concept.
- Lengthening the available sample: new observations will keep showing up for the most recent periods.
- Improving information: for the last known years, series in the French National Accounts appear in succession as provisional, semi-final and final.
- Changing the definition of some variables.

One can also add to the bank a completely new series

- Which has appeared recently as useful to the model.
- Which has been made available by access to a new source of information, or the creation by data builders of a new, more interesting, concept.

This multiplicity of possible changes prohibits the global set of series used by the model to remain the same even for a short period. Adapting constantly model specifications (in particular the estimated equations) to this evolution would ask a lot from the model builder, to the detriment of more productive tasks. This means one should limit the frequency of reconstitutions, for the operational data set (for example once or twice per year for an annual model, or every quarter for a quarterly one), with few exceptions: correcting serious mistakes or introducing really important information.

Without doubt, the best solution is actually to manage two sets of data, one updated frequently enough with the last values, the other built at longer intervals (the periodicity of the model for example). This solution allows to study in advance, by estimations based on the first set, the consequences of the integration of new values on the specifications and properties of the next model version.

5.2.6 SUPPRESSIONS

It is beneficial to delete in the bank those series which have become useless:

- This allows to gain space.
- Searches will be faster.
- The bank will be more coherent with the model.
- The model builder will have less information to memorize, and the architecture of the bank will be easier to master (one will have guessed that this is the most important feature, in our sense).

Useless series that are preserved too long lead to forgetting what they represent, and their destruction will require a tedious identification process.

For EViews, this presents an additional interest: the elements in the workfile will be display in a single window, and it is essential for this window to contain as many interesting elements as possible.

5.2.7 THE DOCUMENTATION

Similarly, investment in the documentation of series produces quick returns. It can concern:

- The definition, possibly on two levels: a short one to display titles in tables or graphs, and a long one to fully describe the concept used.
- The source: original file (and sheet), producing institution and maybe how to contact the producer.
- The units in which the series is measured
- Additional remarks, such as the quality and status (final, provisory, estimated) of each observation.
- The date of production and last update (hours and even minutes also can be useful to determine exactly which set of values an application has used). This information is often recorded automatically by the software.
- If pertinent, the formula used to compute it.

Example: $\text{Wage rate} = \text{Wages paid} / (\text{employment} \times \text{Number of weeks of work} \times \text{weekly work duration})$.

EViews allows to specify the first four types, using the label command, and produces automatically the last two.

For example, a series called GDP can be defined through the sequence:

```
GDP.label(c)
GDP.label(d) Gross Domestic Product at constant prices
GDP.label(u) In 2005 Euros
GDP.label(s) from the Excel file accounts.xls produced by the Statistical Office
GDP.label(r) 2012Q4 is provisory
```

Which clears the contents, gives the definition, describes units, the source, and adds remarks.

- In addition, EViews 8 allows to introduce one's own labels, for instance the country for a multinational model, the agent for an accounting one, or the fact that a series belongs to a particular model.

For instance you can use:

```
HI.label(agent) Households
MARG.label(agent) Firms
```

- Moreover, if the workfile window screen is in "Display+" mode, you can sort the elements according to their characteristics. In addition to the name, the type and the time of last modification (or creation) you have access to the description.

And if you right click on one of the column headings, and choose "Edit Columns" you can display additional columns for any of the label types, including the ones you have created.

This can prove quite useful, as it allows you to filter and sort on any criterion, provided you have introduced it as a label.

This criterion can be for instance:

- The agent concerned
- The country
- The association with a given model
- The formula in the model
- The formula used to create the series (if any)³⁶
- The type within this model (exogenous, endogenous, identity, behavior...)
- The sub-type: for exogenous it can be policy, foreign, structural. For endogenous it can be behavior or identity.

Once the display is produced, it can be transferred to a table, which can be edited (lines, fonts...) and used for presentations.

For instance, one can produce a table for a model, with columns for type, agent, units, source, identity / behavior.... This table can be sorted using any of the criteria.

These new functions allow table production to be integrated in the modelling process, a very powerful information tool for both model development and documentation.

For instance you could use:

```
F_HDI.label(d) Disposable income
U_MARG.label(d) Margins
F_HDI.label(model) France small
U_MARG.label(model) USA small
F_HDI.label(agent) Households
U_MARG.label(agent!) Firms
```

and produce sorted tables according to any of the three criteria.

5.2.8 CONSEQUENCES ON WORK ORGANIZATION

Let us now give some specific considerations on data management.

In the general case, the model builder will be confronted with a large set of series of more or less various origins. Optimal management strategy might appear to vary with each case, but in fact it is unique in its main feature: one must produce a file, in the standard of the model building software, and containing the series having a chance to be useful for the model.

³⁶ You can also use the “source”

This is true even if the global set of necessary series is produced and managed on the same computer or computer network, using the same software (the task of transfer will be simply made easier): it is essential that the model builder has control over the series he uses, and especially that he manages changes (in particular updates of series in current use). Interpreting a change in model properties (simulations, estimations), one must be able to dismiss a change in the data as a source, except if this change has been introduced knowingly by the model builder himself³⁷.

Such an organization also makes the management of series easier. In particular, limiting the number of series in the bank, apart from the fact that it will save computer time and space, will make the set easier to handle intellectually.

Concerning the scope of the series, two extreme options can however be considered:

- Transferring in the model bank the whole set of series that have a chance (even if a small one) to become useful at one time to the development of the model³⁸.
- Transferring the minimum, then adding to the set according to needs.

If a median solution can be considered, the choice leans strongly in favor of the first solution. It might be more expensive initially, in human time, and in size of files, but it will prove generally a good investment, as it avoids often a costly number of limited transfers, and gives some stability to the bank as well as to its management procedures.

5.2.9 THE PRACTICAL OPTIONS

For models managed by institutions, the most frequently found organization is a team working on a local network or connected computers, where one can transfer a selection of data coming from distant sources (mainframe, data providers). One might in some cases access directly the banks of the provider from inside a model-building session. The producers of modelling packages are giving a high priority to this type of option.

For large teams, or teams working in different locations, communication is no longer an issue, provided Internet access is available. Internet connection is now even used by economists working in the same building.

One must however pay attention to format incompatibilities, especially if the operating system is different (Windows and its versions, Linux, UNIX, Macintosh...)³⁹.

5.3 BACK TO OUR EXAMPLE

Now that we know the principles, let us see how to apply them to the case we have defined earlier. To avoid switching between distant pages, we shall repeat its presentation.

³⁷ This remark is a particular application of the general principle « let us avoid potential problems which can prove expensive in thinking time ».

³⁸ Even if they are not considered for actual model variables. For instance, one can be interested in comparing the capital output ratio of the modelled country with those of other countries.

³⁹ Actually this problem appears quite rarely, as most modelling packages (except perhaps Troll) work only under Windows.

1. In the example, our economist has decided to build a very simple model of a country's economy, which includes the following elements: Based on their production expectations and productivity of factors, firms invest and hire workers to adapt productive capacity. However, they exert some caution in this process, as they do not want to be stuck with unused elements.
2. Productive capital grows with investment, but is subject to depreciation.
3. The levels actually reached for labor and capital define potential GDP.
4. They also need intermediate products (proportional to actual GDP), and build inventories, starting from the previous level.
5. Households obtain wages, based on total employment (including civil servants) and a share of Gross Domestic Product. They consume part of this revenue, and invest another (in housing).
6. Final demand is the sum of its components: consumption, productive investment, housing investment, inventories, and government demand. Total demand includes also intermediate consumption. Final and total demand are the sum of their components
7. Imports are a share of local total demand, final or intermediate. But the fewer capacities remain available, the more imports will be called for.
8. Exports follow world demand, but the priority of local firms is satisfying local demand. They are also affected by capacity constraints.
9. Supply is equal to demand.

We have voluntarily kept the framework simple (maybe not enough), as our purpose is only explanatory at this time. However the model we are building has some economic consistency, and can actually represent the nucleus for further extensions which we shall present later.

We shall also suppose that the following data is available in an Excel file called FRA.XLS, selected from OECD's Economic Perspectives data set. Series are available from the first semester of 1970 to the last of 2004.

The reason for the "FRA" prefix is to identify series for France in a large set of countries, representing all the OECD members as well as some groupings.

They use the following units:

Values: Euros

Deflators: base 100 in 1995.

Volumes (or quantities): Millions of 1995 Euros

Populations: persons

FRA_CGV Government Consumption, Volume

FRA_CPV	Private Consumption, Volume
FRA_EG	Employment, Government
FRA_ET	Total Employment
FRA_CGV	Government Consumption, Volume
FRA_CPV	Private Consumption, Volume
FRA_EG	Employment, Government
FRA_ET	Total Employment
FRA_FDDV	Final Domestic Demand, Volume
FRA_CGV	Government Consumption, Volume
FRA_CPV	Private Consumption, Volume
FRA_EG	Employment, Government
FRA_ET	Total Employment
FRA_FDDV	Final Domestic Demand, Volume
FRA_GAP	Output Gap
FRA_GDPTR	Potential Output, Total Economy at Current Prices
FRA_GDPV	Gross Domestic Product (Market prices), Volume
FRA_IBV	Gross Fixed Cap Form, Business Sector, Volume(Narrow Definition)
FRA_ICV	Intermediate consumption, Volume
FRA_IGV	Government Investment, Volume
FRA_IHV	Investment in Housing, Volume
FRA_ISKV	Increase in stocks. volume
FRA_KBV	Capital Stock, Business
FRA_MGSV	Imports Goods and Services, N.A. Basis, Volume
FRA_PCP	Deflator, Private Consumption
FRA_TDDV	Total Domestic Demand, Volume
FRA_WSSS	Compensation of Employees
FRA_XGSV	Exports Goods and Services, N.A. Basis, Volume
FRA_XGVMKT	Exports Goods and Services, Market Potential, Volume
FRA_YDRH	Real Household Disposable Income

Applying the principles we have defined above calls for:

- Creating the model specifications
- Identifying the variables in the model.
- Separating them into endogenous and exogenous.
- Writing down the full identities.
- Establishing each behavioral equation as an identity, presenting in the simplest way the variable it defines, and the explanatory elements.
- Creating the associated series, from the available data.
- Transferring the elements already available into model series using the names allocated to them.
- Specifying formulas computing the remaining elements

Now that we have obtained the data, we can move to the two tasks: transform it to fit the model needs, start specifying the model equations.

It should be clear that this will have to be done through a set of stored statements in a readable language (a program). This option will allow:

- Establishing an apparently consistent set of statements, which can be controlled visually.

- Locating errors and introducing corrections as simply and clearly as possible.
- Storing subsequent versions, including the last and most correct one, until a satisfying version is established.
- Replicating this process with the smallest amount of work.
- Displaying the steps in the process as clearly as possible, introducing comments.
- Once a satisfying stage has been reached, memorizing the actions for later use (especially if the modeling project faces breaks, short or long).
- Allowing external users to master the current state of operations, to evaluate the present stage of development of the project.

The program can be inserted with comments, making the sequence of tasks and the role of individual commands clearer, and allowing to warn of the presence of local problems and the way they have been processed. This is especially useful for a team project, for which the name of the author should also be included.

Under EViews, two other methods are available:

- Using a sequence of menu and sub-menu functions,
- Typing commands without saving them, directly from the command window.

These two methods fail on all criteria. The record of the tasks is not available, which means errors are difficult to detect. Reproducing the task, whether to correct errors or to update specifications or data, calls for a new sequence of menu selections or typed statements⁴⁰.

The obvious choice is even comforted by three features provided by EViews 8:

- You can run part of a program, by selecting it with the mouse (in the usual Windows way), clicking on the right button, and choosing "Run Selected".

This is generally more efficient than the previous method of copying the selected part into a blank program, and running it. However the new method does not allow editing, useful when one wants to run a selected AND modified set.

- Symmetrically one can exclude temporarily from execution part of a program, by "commenting it out". To do this, one should select the relevant part, click on the right button, and choose "Comment Selection". To reactivate the statements, one should select them again and use "Uncomment Selection".

This can be a little dangerous, especially if you (like myself) have the reflex of saving the program before each execution. To avoid destroying the original, one can save first the modified program under another name⁴¹.

Finally, one can ask a column of numbers to be displayed left of the program lines. This is particularly efficient if you use the "Go To Line" statement⁴².

⁴⁰ However, one can copy the sequence of statements entered in the command window into a program file.

⁴¹ Only once of course.

⁴² However, you have to be careful to update the numbers when the program changes.

So actually, the only option is the one we proposed above: defining a program producing all the necessary information, and the framework of equations which is going to use it. But the ordering of the tasks can be questioned, as we have started explaining earlier. Until both are completed, the job is not done, but they are technically independent: one does not need the physical model to create the data, or series filled with values to specify the equations. This means that one can consider two extreme methods:

- Creating all the data needed by the model, then specifying the model.
- Specifying all the model equations, and then producing the associated data.

The criterion is the intellectual feasibility of the ordered sequence of tasks.

Clearly the first option is not realistic, as writing down the equations will surely evidence the need for additional elements. The second is more feasible, as one does not need actual series to write an equation. But as the definition of the equation processes, one has to check that all the addressed elements are or will be available in the required form, either as actual concepts (GDP) or transformations of actual concepts (the budget deficit in GDP points calls for the deficit and GDP series). If a concept appears to be lacking, one will have to: use an alternate available element (a “proxy”), establish an assumption, look in alternate bases not yet accessed, or simply eliminate the element from the model.

This shows that if producing both sets can be done in any order, there is a preference for specifying the equations first. If the data is not ready, but its contents are known, it is possible to write down the equations and ask the software to proceed the text. The user will be told about possible syntax errors, about the nature of the variables (endogenous / exogenous), and the architecture of his model. This will lead to early model corrections, allowing to gain time and avoiding taking wrong directions later. And if the model specifications are still discussed, it is possible to build a first version of the associated data set, which will be updated when the model is complete.

In practice, especially in the simplest cases, one can also start defining the program with two blank paragraphs, and fill them with data and equation creating statements until both are complete. The eight original paragraphs in our model specifications can be treated one by one (not necessarily in the numerical order) filling separately the data and equation generating blocks with the associated elements.

Actually among the above proposals we favor two alternate techniques:

- **Model then data:** Specifying first the full model, checking that all elements used can be produced either directly or through a formula. Then producing the full set of data, preferably through a direct transfer or a transformation.
- **Model and data:** Producing the equations in sequence, or related block by related block, and establishing simultaneously the statements which create all the series they need.

5.3.1 APPLICATION TO OUR EXAMPLE

Let us now show on our example how the process can be conducted using the second method, probably more adapted to such a small model (one cannot expect to switch between the two processes to many times).

We shall first present the process in general (non EViews) terms, treating each case in sequence, and presenting both the equations and the statements generating the associated variables. To make things clearer, the equations will be numbered and the creation statements will start with “>>”

Also, the endogenous variable will use uppercase characters, the exogenous will use lowercase.

(1) *Based on their production expectations and productivity of factors, firms invest and hire workers.*

This defines two behavioral equations for factor demand, in which employment (let us call it LE) and Investment (called I) depend on GDP, called Q.

$$(1) LE=f(Q)$$

$$(2) I=f(Q)$$

We need:

$$\begin{aligned} >> IP=FRA_IBV \\ >> Q =FRA_GDPV \end{aligned}$$

But for LE, we face our first problem. Private employment is not directly available. However, we have supposed that total employment contained only public (government) and private. This means we can use:

$$>> LE=FRA_ET-FRA_EG$$

(2) *Productive capital grows with investment, but is subject to depreciation.*

Capital K, measured at the end of the period, is defined by an identity. Starting from the initial level, we apply a depreciation rate (called dr) and add investment. The equation is written as:

$$(3) K(t)= K(t-1).(1-dr(t)) + I(t)$$

Defining it at the end of the period would only change notations.

We need the data for K

$$>> K=FRA_KBV$$

And we get dr by inverting the formula:

$$\gg dr = ((K(-1) + IP) - K) / K(-1)$$

In other words, dr will be the ratio, to the initial capital level, of the difference between two levels of capital: the value we would have obtained without depreciation, and the actual one.

(3) The levels actually reached define potential production.

Capacity (called CAP) depends on factors LE and K

$$(4) CAP(t) = f(LE(t), K(t))$$

It can be computed directly as:

$$\gg CAP = FRA_GDPVTR$$

which rather represents a “normal” GDP value considering the present level of factors.

The direct availability of this concept as a series represents the best case, not often met in practice. Later in the text we shall address the alternate techniques available in less favorable situations.

(4) They need inputs, and also build inventories.

Intermediate consumption can be defined as proportional to GDP, using the actual value. This means that at any level of production, each unit produced will need the same amount of intermediary products.

$$(5) IC = r_icq \cdot Q$$

For inventories, we will estimate its change:

$$(6) CI = f(Q)$$

For this, we need to compute:

$$\begin{aligned} \gg IC &= FRA_ISKV \\ \gg r_icq &= IC/Q \text{ (or } FRA_ISKV/FRA_GDP) \\ \gg CI &= FRA_CIV \end{aligned}$$

- (5) Households obtain wages, based on total employment (including civil servants) and a share of Gross Domestic Product. They consume part of this revenue.

Now we need to define total employment, by adding government employment (called lg) to LE .

$$(7) LT = LE + lg$$

The new series are obtained by:

```
>> LT = FRA_ET  
>> lg = FRA_EG
```

Now we have to compute household revenue, which we shall call R_{HI} . We shall suppose that the same wage applies to all workers, and that the non-wage part of Household revenue is a given share of GDP, a series called r_{rhiq} . This gives:

$$(8) RHI = wr \cdot LT + r_{rhiq} \cdot Q$$

Actually the above assumption, while simplistic, is probably not too far from the truth. The sensitivity to GDP of the elements included in this heterogeneous concept can be low (such as pensions, or interests from long term bonds), high (the revenue of small firm owners, with fixed costs and variable output), or medium (self-employed working in the same capacity as wage earners).

Household consumption is given by applying to RHI the complement to 1 of a savings rate which we shall call sr . For the time being, the savings rate is exogenous:

$$(9) CO = RHI \cdot (1 - sr)$$

Housing investment is also a share of RHI , which we shall call r_{ih} .

$$(10) IH = r_{ih} \cdot RHI$$

The new variables are RHI , wr , r_{rhiq} , sr , IH and r_{ih} .

RHI is given simply by:

```
>> RHI = FRA_YDRH
```

Let now compute the real wage rate w_r . This is done through the following computation.

Dividing FRA_WSS by FRA_ET gives the individual nominal value, which we divide again by $FRA_CPI/100$ ⁴³ to get the real value⁴⁴.

```
>> wr = (FRA_WSS/FRA_ET)/(FRA_CPI/100)
```

(parenthesizes are added for clarity).

r_rhi will be obtained as the ratio to GDP of household revenue minus wages

```
>> r_rhi = (RHI - wr . LT) /Q
```

Consumption and housing investment will be obtained directly:

```
>> CO=FRA_CPV  
>> IH=FRA_IHV
```

Computing the savings rate and r_ih will use the inversion of the associated equation:

```
sr =(RHI-CO)/RHI
```

or

```
sr=(FRA_YDRH-FRA_CPV)/FRA_YDRH
```

(savings divided by revenue)

```
>> r_ih=IH/RHI
```

⁴³ The OECD deflators are measured as 100 in 1995.

⁴⁴ Considering the above list of available series, one can observe other options are possible.

Or

>> FRA_IHV/FRA_YDRH

(6) Final demand is the sum of its components: consumption, productive investment, housing investment, inventories, and government demand. Total demand includes also intermediate consumption.

(11) $FD=IP+CO+IH+gd+CI$

(12) $TD = FD + r_ic \cdot Q$

We need to compute gd as the sum of FRA_IGV and FRA_CGV .

>> $gd = FRA_IGV+FRA_CGV$

>> $FD = FRA_TDDV$

>> $r_ic = FRA_ICV/FRA_GDPV$

(7) Imports are a share of local demand («domestic demand»). But the less capacities are still available, the more an increase in demand will have to be imported.

This calls for:

(13) $UR=Q/CAP$

(14) $M=f(FD+IC,UR)$

We need to compute:

>> $UR=Q/CAP$ (its definition)

>> $M=FRA_MGSV$

(8) Exports will essentially depend on World demand. But we shall also suppose that if tensions appear (through UR) local firms will switch some of their output to local demand, and be less dynamic in their search for foreign contracts.

(15) $X=f(WD, UR)$

We need:

```
>> X=FRA_XGSV
>> WD=FRA_XGVTR
```

(9) Supply is equal to demand.

The supply-demand equation will for the moment use the following implicit form:

$$(16) Q + M = FD + X$$

(all variable values are obtained earlier)

We can now reorder the framework of our model into the following elements:

- [1] $LE = f(Q)$
- [2] $IP = f(Q)$
- [3] $K = K_{-1} (1 - \text{depr}) + IP$
- [4] $CAP = f(LE, K_{-1})$
- [5] $IC = r_{icq} \cdot Q$
- [6] $CI = f(Q)$
- [7] $LT = LE + lg$
- [8] $RHI = wr \cdot LT + r_{rhiq} \cdot Q$
- [9] $CO = (1 - sr) \cdot RHI$
- [10] $IH = r_{ih} \cdot RHI$
- [11] $FD = CO + IH + IP + CI + gd$
- [12] $TD = FD + r_{ic} \cdot Q$
- [13] $UR = Q / CAP$
- [14] $M = f(TD, UR)$
- [15] $X = f(wd, UR)$
- [16] $Q + M = FD + X$

Endogenous variables

I	Firms investment.
LE	Firms employment.
K	Firms (productive) capital
CAP	Potential output
LT	Total employment.
CI	Change in inventories
IC	Intermediate consumption
IH	Housing investment.
CO	Household consumption.
FD	French final demand
TD	French total demand
M	French Imports.
RHI	Household real income.
UR	Rate of use of capacities
X	French Exports.
Q	Gross Domestic Product

Exogenous variables

depr	Depreciation rate of capital
gd	State consumption and investment.
lg	Public employment
r_ih	Share of Housing investment in Household revenue.
r_rhiq	Share of GDP transferred to Households, in addition to wages
wd	World demand normally addressed to France.
r_icq	Ratio of intermediate consumption to GDP
wr	Real average wage rate

One observes:

- That we have indeed as many equations as variables to compute.
- That we have separated behavioral equations and identities.
- That accounting identities are completely defined.
- That on the other hand the form of behavioral equations is still indefinite, although the explanatory elements are known (at least as a first guess).

This distinction is normal. As we have already indicated, identities generally represent a mandatory formal connection, while conforming behavior equations to economic theory is not so restrictive.

- Computing formulas

By considering the formulas we have obtained, we can see that most of the data needed is available directly, so a simple transfer should be enough. We might even have considered using the original names. But as our model will apply only to France, there is no reason to keep the prefix, which helped to identify the French data inside a much larger multi-country file. And one might decide (rightly in our sense) that our names are clearer.

The correspondences are:

Q	= FRA_GDPV
CAP	= FRA_GDPVTR
CI	= FRA_ISKV
LT	= FRA_ET
LG	= FRA_EG
FD	= FRA_TDDV
CO	= FRA_CPV
RHI	= FRA_YDRH
I	= FRA_IBV
IH	= FRA_IHV
WD	= FRA_XGVMKT
X	= FRA_XGSV
M	= FRA_MGSV

Only eight elements are lacking, seven of them exogenous variables:

gd	=FRA_IGV+FRA_CGV	Government demand
UR	=Q/CAP	Rate of use of capacities
depr	$= ((K(t-1) + IP) - K(t))/K(t-1)$	Depreciation rate of capital
r_ic	=IC/Q	Ratio of intermediate consumption to GDP
r_ih	=IH/RHI	Share of Housing investment.
r_rhiq	$= (RHI - wr \cdot LT) / Q$	Non-wage households revenue: share of GDP
sr	$= (RHI - CO) / RHI$	Savings rate
wr	$= (FRA_WSS / FRA_ET) / (FRA_CPI / 100)$	Real average wage rate

In real cases, this kind of computation will be used often. One must be aware of **one important issue**:

The use of these formulas is separated from the definition of model equations. The only reason we need them is to produce the **historical** values of series not yet available. If the statisticians had made a comprehensive job (and if they knew the requirements of the model) they would have provided the full set, and no computation would have been necessary (just a changes of names).

So these two types of formulas have completely different purposes

- Applying the **computation statements** ensures that all the requested data is available. By associating formulas to missing elements, they allow to produce the set required for simulation and estimation. If the data was already available in the right format, and the names given to the variables were acceptable, no statement

would be necessary. And one can check that in our case, most of the computations are actually direct transfers, which allow to create a model element while retaining the original series.

Actually, one could question the necessity of having a full set of historical values for endogenous variables. These will be computed by the model, which will be simulated on the future anyway. The reasons for producing a full set are the following:

- Estimation will need all the elements in the associated equations.
- Controlling the consistence of identity equations with the data is a prerequisite before any simulation; otherwise we may start with a flawed model set.
- Checking that the model gives accurate simulations on the past will need all the historical elements.
- Many equations use lagged values. This requires actual values preceding the starting simulation date.

These formulas can include original data, transformed data computed earlier in the program, or simply assumptions. For instance:

- GDP has been drawn directly from the original set.
- The depreciation rate relates the sequence of capital values, and investment.
- In the absence of other information, the target for inflation can be set to 2%.
- The **model equations** establish a logical link between elements, which will be used by the model to produce a consistent equilibrium. This means that if the formula for computing variable A contains variable B, variable A is supposed to depend on B, in economic terms.

This is obviously true for estimated equations. For instance, the wage rate can depend on inflation, of exports on world demand. But this is also true for identities:

Household revenue is the sum of its elements. If one grows, revenue changes in the same way (ex-ante, of course). Basically, we suppose that some behaviors apply in the same way to every element of revenue, whatever its source. If household consumption is estimated, savings are the difference between revenue and consumption.

It is extremely important to understand this issue, at the start of any modeling project.

It is quite possible however that the same formula is present in both sets. For instance we might not have values for FD, and we believe that CO, I, IH and gd represent the whole set of its components. In this case the formula:

$$FD = CO + IP + IH + gd$$

will be used both for computing historical values of FD and to define FD in the model.

This introduces an obvious problem: if we make a mistake in the formula, or we use the wrong data, there is no way to detect it.

5.3.2 THE EIEWS PROGRAM

Let us now consider how the above task can be produced. We want to create:

- Workfile for all model elements
- An image of the model, with fully defined identities, indications as the intended estimated equations,
- The associated data.

5.3.2.1 The workfile

- First, we need a work file. In EViews, all tasks are conducted in memory, but they apply to the image of a file which will contain all the elements managed at a given moment.

We can create the file right now (as a memory image) or start from a pre-existing one, in which case the file will be transferred from its device into memory.

Some precautions have to be taken.

- First, only one version of the file must be open in memory. As we state elsewhere, EViews allows the user to open a second version (or even third, and fourth...) of a file already opened. Then changes can be applied separately to both memory versions, such as series generation and estimations.

This is obviously⁴⁵ very dangerous. At the least, one will lose one of the set of changes, as there is no way to transfer elements from an image to the other. Of course, each file can be saved under a different name, but this does not allow merging the changes⁴⁶. At the worst, one will forget the allocation of changes to the files, and one or both will become inconsistent, the best option being to avoid saving any of them, and to start afresh.

This means one should:

- * In command mode, check that no file of the same name is opened, and close it if necessary.
- * In program mode (the case here) make sure that no file is open at first. This calls for an initial "CLOSE" statement, which will not succeed most of the time⁴⁷ but will guarantee that we are in the required situation.
- Second, a new project must start from a clean (or empty) workfile. For an initial file to contain elements is at best confusing, at worst dangerous. For instance series with the same name as elements in our project can already be present with a different meaning (GDP for a different country?), and available for a larger period. Allowing EViews to estimate equations over the largest period available will introduce in the sample irrelevant values.

A simple way to solve the problem is to delete any existing element, through the statement:

⁴⁵ This is only a personal opinion.

⁴⁶ Providing this option does not look impossible.

⁴⁷ With fortunately no error message.

DELETE *

which will destroy any pre-existing item, except for the generic C (generic vector of coefficients) and RESID (generic series of residuals) which are created automatically with the work file, and cannot be deleted.

There is only one acceptable case for pre-existing elements: if the work file contains some original information, provided to the user by an external source. But even in this case the file has to be saved first, to allow tracing back the steps to the very beginning in which only this original information was present, in its original form.

In any case, in EViews, the possibility to define separate sheets inside the work file solves the problem. As we have seen earlier, one can just store the original data in one sheet, and start building the transformed data in a blank one, logically linked to the original.

First principle of modeling: always organize your work in such a way that if step n fails, you can always get back to the result of step n-1.

First principle of modeling (alternate version): Always organize you programs in such a way that you can produce again all the elements associated with the present situation.

This (long) discourse leads to the following statements:

```
CLOSE small
WFCREATE(page=model) small Q 1970Q1 2005Q4
DELETE *
```

Applying them guarantees:

- That the file small.wf1 is open in memory with the needed characteristics, for a page called “model”.
- That only one version is open (provided no more than one was open previously, of course, but we shall suppose you are going to follow our suggestions).
- That the page is empty (actually it contains only C and RESID).

5.3.2.2 The data

Now that we have a work file, we must fill it with the necessary information.

The original information is represented by the 20 series in the FRA.XLS⁴⁸ Excel file. We shall import them using the IMPORT statement. This statement is quite simple (see the User’s manual for detailed options):

⁴⁸ EViews allows also to read Excel 2010 .xlsx files (but not to produce them).

```
READ fra.xls 20
```

But beware: even if the Excel file contains dates (in the first column or line) this information is not taken into account. What is used is rather the current sample, defined by the last SMPL statement. Fortunately, in our case, the current sample, defined at workfile creation, is the same as the one in the Excel file. But this will not always be the case: better to state the SMPL before the READ.

```
SMPL 1970Q1 2004Q4  
READ fra.xls 20
```

Second principle of modelling: if introducing a (cheap) statement can be useful, even extremely seldom, do it.

One also has to be careful about the orientation of series: normally they appear as columns, and data starts from cell B2 (second line, second column). Any other case has to be specified, as well as the name of the sheet for a multi-sheet file.

An alternate (and probably better) option

If you follow the above method, all the data will be transferred to the “model” page. This makes things easier in a way, as all information will be immediately available. But

- The separation between original and model data will not be clear.
- The stability of the original data is not guaranteed.
- As the original series are probably more numerous, most of the screen will be occupied by elements no longer useful.

Of course, one can separate original and model data by using a prefix for the first type. But it is even better to separate the two sets physically. This can be done through the “link” EViews function.

Instead of loading the original series in the model page, a specific page is created (named for instance “oecd”) in which the data is imported.

Then in the model page the model variables are declared as “linked”, and a link is defined with the original series in the “OECD” page.

The associated syntax will be presented later.

5.3.2.3 The model

- Now, we need to define the model on which we shall work. Producing a model starts with the statement:

```
MODEL modelname
```

Let us call our model `_fra_1`.

A trick: starting the name of important elements by an underscore allows them to be displayed at the beginning of the workfile screen, avoiding a tedious scrolling if the number of elements is large. For very important elements (like the model itself) you can even use a double underscore.

The statement

```
MODEL _fra_1
```

defines `_fra_1` as the “active” model.

Two cases can be considered:

- The model does not exist. It is created (with no equations yet).
- The model exists. It is opened, with its present equations.

The second option is dangerous in our case, as we want to start from scratch. To make sure of that, the most efficient (and brutal) technique is to delete the model first, which puts us in the first case.

```
DELETE _fra_1  
MODEL _fra_1
```

This introduces a slight problem, however. In most cases (including right now) the model does not exist, and the DELETE statement will fail. No problem, as what we wanted is to make sure no model preexisted, and this is indeed the situation we obtain. But EViews will complain, as it could not perform the required task. And if the maximum number of accepted errors is 1 (the default option) the program will stop.

- We can use the “noerr” option, which accepts failure of the statement without error message.

```
DELETE(noerr) _fra_1  
MODEL _fra_1
```

- We can change the default number of accepted errors.

Another way to avoid this situation is obviously to set the maximum number of errors to more than 1. This is done by changing the number in the “Maximum errors before halting” box in the “Run program” menu.

Actually, if you have followed the principle above, there is no risk in proceeding in a program which produced error messages, even valid ones. You have saved the elements associated to the initial situation, and even if you forgot to do that, you can always repeat the steps which led to it.

The advantage of this option:

- The program will continue after irrelevant error messages.
- You can produce artificial errors, which can be quite useful as flags⁴⁹.
- The messages can be associated to several logically independent errors, which can be corrected simultaneously, leading faster to a correct version.

Now, which number should we specify? In my opinion, depending on the model size, from 1000 to 10000. The number has to be higher than the number of potential errors, as you want to get as close as possible to the end of the program. Of course, you will never make 10000 logical errors. But the count is made on the number of error messages. And in a 2000 equations model, if you have put all the endogenous to zero and you compute their growth rates, this single mistake will generate 2000 messages.

The only drawback is that if your program uses a loop on the number of elements of a group, and this group could not be created, the loop will run indefinitely with the message:

```
Syntax error in "FOR !I=1 TO G.@COUNT"
```

You will have to wait for the maximum number to be reached.

- Introducing the equations.

Now that we have a blank model, we can introduce the equations one by one. The text of these equations has already been defined, we just need to establish the EViews commands.

This is done through the APPEND statement.

The first one will define investment:

```
_fra_1.append IP=f*(Q)
```

Clearly the syntax

- Contains the statement “append”
- Adds the model name on the left, with a dot.
- Adds the text of the equation on the right, with a separating blank.

We must now explain the strange syntax of our equation.

⁴⁹ The message associated with a real error will locate it between the preceding and following artificial errors.

- At this moment, we expect the model to explain the decision on investment by the evolution of GDP. This seems quite logical, but we have not established the possible forms of the theoretical equation, and we have not checked that at least one of these equations is validated by all required econometric tests.
- But at the same time we want EViews to give us as much information as possible on the structure of our model: simultaneities, exogenous parts...
- The best compromise is clearly to produce a model which, although devoid of any estimated equation, nevertheless presents the same causal relationships as the (future) model we consider.

The simplest choice should be, as if we were writing model specifications in a document or on a blackboard, to state:

$$IP=f(Q)$$

Unfortunately, EViews does not accept an equation written in this way. It will consider we are using a function called f, with the argument Q. As this function does not exist, the equation will be rejected.

The trick we propose is to put an asterisk between “f “and the first parenthesis, which gives

$$IP=f^*(Q).$$

And state f as a scalar (to avoid confusion with an additional exogenous).

If more than one explanatory variable is used, such as in the productive capacity equation, we would like to write:

$$CAP=f^*(LE,K)$$

Again, this is not accepted by EViews, and we can write instead:

$$CAP=f^*(LE+K)$$

One just has to state his conventions, and you are welcome to use your own.

However, dropping the f is dangerous, such as in:

$$M=FD+TD$$

This will work too, but the equation can be confused with an actual identity, quite misleading in this case.

The complete set of equation statements is:

```
_fra_1.append LE =f*(Q)
_fra_1.append I=f*(Q)
_fra_1.append K = K(-1)*(1-depr) + I
_fra_1.append CAP=f*(LE+ K)
_fra_1.append IC=r_icq * Q
_fra_1.append CI=f*(Q)
_fra_1.append LT=LE+lg
_fra_1.append RHI = wr * LT + r_rhiq .*Q
_fra_1.append CO = (1-sr) * RHI
_fra_1.append] IH = r_ih * RHI
_fra_1.append FD = CO + IH + IP + CI + gd
_fra_1.append TD = FD + r_ic * Q
_fra_1.append UR = Q/CAP
_fra_1.append M=f*(TD+UR)
_fra_1.append X=f*(wd+UR)
_fra_1.append Q + M = FD + X
```

They produce a 16 equations model called `_fra_1`. After running the statements, an item will be created in the workfile, with the name “`_fra_1`” and the symbol “`M`” (in blue).

Double-clicking on this item will open a special window, with the list of equations:

- Text (with the icon “TXT” on the left),
- Number (in the order of introduction in the model).
- Economic dependencies: the dependent variable on the left, the explanatory on the right, using actually the syntax we could not apply earlier. Lags are not specified, as we shall see later. So `K` is presented as depending on `K`.

Actually three other display modes are available, using the “View” button:

- Variables: shows the variables (endogenous in blue with “En”, exogenous in yellow with “X”). For the endogenous, the number of the equation is given. This allows locating the equation in the model text, which is useful for large models.

The “dependencies” button gives access to a sub-menu, which allows to identify the variables depending on the current one (Up) and which influences it (Down).

For instance, for FD, “Up” will give TD and Q, “Down” will give CO,I,G,and IH

Of course, exogenous will only be allowed the “Up” button.

The “Filter” option will allow selecting variables using a specific “mask”. For instance in a multi-country model the French variables can be identified with FRA_*, provided one has used such a convention.

- Source text: this is basically the text of the model code. We shall see that this changes with estimated equations.
- Block structure: this gives information on the logical structure of the model (detailed later in Chapter 7).

We get:

- The number of equations.
- The number of blocks, separated into simultaneous and recursive.
- The contents of each block.

For the time being, let us only say that a simultaneous block contains interdependent elements. For any couple of elements in the block, a path can lead from the first to the second, and vice-versa. Of course this property does not depend on the ordering of equations inside the block.

EViews gives also number of feedback variables (this will be explained later too).

On the contrary, a recursive block can be ordered (and EViews can do it) in such a way that each variable depends only (for the present period) on previously defined variables.

This information is useful to improve the understanding of the model, to locate inconsistencies and to correct technical problems.

EViews can detect errors if:

- A variable is defined twice
- The syntax of an equation is wrong (a parenthesis is lacking for instance)

and allow the user to observe errors himself if:

- Normally endogenous elements appear as exogenous: the equation for the variable has been forgotten, or written incorrectly.
- Elements foreign to the model appear: variables have been misspelled.
- A loop appears where there should be none.

- Or (more likely) an expected loop does not appear: for instance a Keynesian model is described as recursive, or a model for two countries trading with each other can be solved as two independent blocks.

All these errors can be detected (and corrected) without calling for the data. This can speed up the building process, especially if the data is not yet produced.

For the production of series, there are two options.

If the original and model series share the same page, one will simply use the “genr” statement, in the sequence.

```
genr Q=FRA_GDPV
genr CAP= FRA_GDPVTR
genr CI =FRA_ISKV
genr IC=FRA_ICV
genr LT =FRA_ET
genr LG =FRA_EG
genr FD = FRA_TDDV
genr CO =FRA_CPV
genr RHI =FRA_YDRH
genr IP = FRA_IBV
genr IH = FRA_IHV
genr WD = FRA_XGVMKT
genr GD = FRA_IGV+FRA_CGV
genr X =FRA_XGSV
genr M =FRA_MGSV

genr r_icq= FRA_IC/FRA_Q
genr r_ih=IH/RHI (or FRA_IHV/FRA_YDRH)
genr r_rhiq=(RHI-WR*LT)/Q
genr sr=(RHI-CO)/RHI
genr UR=Q/CAP
genr wr=FRA_WSSS//FRA_LT (FRA_PCP/100)
genr rdep=((K(-1)+IP)-K)/K(-1)
```

If the original series are managed in their own page (a better option in our opinion), one will use:

```
for %1 Q CAP CI IC LT LG FD CO RHI I IH WD X M
link {%1}
next
Q.linkto oecd\FRA_GDPV
CAP.linkto oecd\FRA_GDPVTR
CI.linkto oecd\FRA_ISKV
IC.linkto oecd\FRA_ICV
LT.linkto oecd\FRA_ET
LG.linkto oecd\FRA_EG
FD.linkto oecd\FRA_TDDV
CO.linkto oecd\FRA_CPV
```



```
RHI. oecd\FRA_YDRH
I.linkto oecd\FRA_IBV
IH.linkto oecd\FRA_IHV
WD.linkto oecd\FRA_XGVMKT
X.linkto oecd\FRA_XGSV
M.linkto oecd\FRA_MGSV
GD .linkto oecd\FRA_IGV+FRA_CGV
```

However, a problem remains for GD, the sum of the two original variables FRA_IGV and FRA_CGV. The LINK function allows to refer to single variables and not functions (as Excel does). Until EViews 8 you had two options.

Creating links to the original elements in the model page.

```
LINK FRA_IGV
LINK FRA_CGV
FRA_IGV.linkto oecd\FRA_IGV
FRA_CGV.linkto oecd\FRA_CGV
genr gd=FRA_IGV+FRA_CGV
```

Or computing a FRA_GDV variable in the original page.

```
genr FRA_GDV=FRA_IGV+FRA_CGV
```

And linking it

```
LINK GD
GD.linkto oecd\FRA_GDV
```

But E Views 8 allows to refer to variables in a different page, as

```
page_name\variable name
```

This means you can use the much simpler method :

```
genr GD= oecd\FRA_IGV+oecd\FRA_CGV
```

Of course, the same method could have been used for single variables.

```
genr Q=oezd\FRA_GDPV
genr CAP=oezd\FRA_GDPVTR
genr CI=oezd\FRA_ISKV
genr IC=oezd\FRA_ICV
genr LT=oezd\FRA_ET
genr LG=oezd\FRA_EG
genr FD=oezd\FRA_TDDV
genr CO=oezd\FRA_CPV
genr RHI. oezd\FRA_YDRH
genr I=oezd\FRA_IBV
genr IH=oezd\FRA_IHV
genr WD=oezd\FRA_XGVMKT
genr X=oezd\FRA_XGSV
genr M=oezd\FRA_MGSV
genr GD =oezd\FRA_IGV+oezd\FRA_CGV
```

it all depends if you want changes in the original series to be applied automatically, or to control the process through GENR⁵⁰. But if the series is not present (like GD) in the original data, a GENR statement is called for anyway.

Now we have produced a first version of the model, and the associated data. At the behaviors have not been established, we obviously cannot solve it. But we can check two important things:

- The data required for estimation is present.
- The data is consistent with the identities.

These conditions are needed to start estimation, the next stage in the process. The first one is obvious, the second less so. But inconsistencies in identities can come from using a wrong concept for a variable, or computing it wrongly. If this variable is going to be used in estimation, whether as dependent or explanatory, the whole process will be based on wrong elements.

- The time spent in estimation will be lost.
- This time will probably be longer than usual, as it is generally more difficult (sometimes impossible) to find a good fit based on wrong data (fortunately?).
- If a good fit is found, the associated equation can remain in the model for a long time (if not indefinitely), and all the subsequent results will be invalidated. If one is honest, discovering the error later means that a lot of work will have to be done again, including possibly published results.

This test can be conducted through a very simple technique: the **residual check**

5.3.3 A FIRST TEST: CHECKING THE RESIDUALS IN THE IDENTITIES

⁵⁰ Of course, this will also increase the size of the workfile.

At this point, asking for a solution of the model cannot be considered. However, some controls can be conducted, which do call for a very specific “simulation”. This technique is called “residual check”.

This method will compute each formula in the model using the historical values of the variables. This can be done by creating for each equation a formula giving the value of the right-hand side expression (using the GENR statement in EViews). However, there is a much simpler method, provided by EViews.

If we consider a model written as:

$$y_t = f(y_t, y_{t-1}, x_t, \hat{\alpha})$$

with y and x the vectors of endogenous and exogenous variables.

We can perform a very specific “simulation”, in which each equation is computed separately using historical values.

Technically this means:

- Breaking down the model into single equation models, as many as there are equations.
- Solving each of these models separately, using as explanatory values the historical ones. If we call these historical values y_t^0

It means we shall compute:

$$y_t = f(y_t^0, y_{t-1}^0, x_t, \hat{\alpha}) + e_t$$

This method will control:

- For identities, the consistency between data and formulation.
- For the behavioral equations, the availability of the variables requested by the contemplated estimations. But one gets no numerical information (actually the method we are proposing will give a zero value).

Actually EViews allows the use of an expression on the left hand side. This applies also here, the comparison being made between the left and right expressions.

The interest of this method is obvious: if the residual in the equation is not zero, it means that there is *at least* one error in that *particular* equation. Of course the problem is not solved, but its location is identified. We shall see later that this method is even more efficient for a fully estimated model, and we shall extend our discussion at that time.

It would be illusory, however, to hope to obtain a correct model immediately: some error diagnoses might have been badly interpreted, and corrections badly performed. But even if the error has been corrected:

- There could be several errors in the same equation

- The correcting process can introduce an error in another equation that looked previously exact, but contained actually two balancing errors. Let us elaborate on this case.

Let us consider our example. If we had used for housing investment the value at current prices:

```
genr IH=FRA_IH
```

Then the equation for FD

```
_fra_1.append FD = CO + IH + IP + CI + gd
```

would not hold true, but the one for IH

```
_fra_1.append IH = r_ih . RHI
```

will, as the computation of r_{ih} as the ratio of IH to RHI will compensate the error by another error.

If we correct the error on IH without correcting r_{ih} , the IH equation will now appear as wrong, while its number of errors has decreased from 2 to 1.

This means achieving a set of all zero residuals might take a little time, and a few iterations, but should converge regularly until all errors have disappeared⁵¹.

5.3.3.1 The types of error met

The residual check allows diagnosing the following errors

- Failure to solve
 - syntax error (call to a non-existent function, unbalanced parentheses).
 - series with the right name, but unavailable, either completely (they have not been obtained), or partially (some periods are lacking).
 - bad spelling (call to a non-existent series)
- Non-zero residuals
 - bad spelling (call to the wrong series).
 - errors of logic. This can be more or less serious, as it can come from a purely technical error: forgetting a term for example, or from a conceptual error: stating an unverified theoretical identity.

⁵¹ Unless the modeler allows some identities to hold true only approximately.

- data error: badly entered information, badly computed series, information coming from non-coherent sources, or from different versions of the same bank.
- Non-verified behavioral equations (or with erroneous residual). This issue will be applicable (and addressed) later.

Observing error values can give clues as to their origin:

- If some periods give a correct result:
 - At the base year (where elements at constant and current prices are identical): the price indexes could be mistaken for one another, or values could be mistaken for volumes.
 - If a variable in the formula is null for these periods, it could be responsible.
 - Otherwise it could come from a typing error (made by the user or the data producer).
 - Or if it appears in the last periods, the provisory elements could be inconsistent.
- Observing the magnitude of the error also can be useful: a residual exceeding the normal economic magnitude (1000% for example) should come from a specification error: bad operator, inversion of coefficients, mistaking values and values per capita. A low residual will often come from confusion between two close concepts (the consumption price deflator excluding or including VAT).
- For additive equations, a missing or extra element may be identified by comparing the residual to the actual values of variables. For instance if the error on final demand for 2000Q1 is 56734 and this is the value of housing investment.
- If the sign of the error is constant (and especially if the order of magnitude is similar across periods), the error could come from the absence of an element, a multiplication by a wrong factor, or a missing positive influence.
- If several errors have identical values, they should have the same origin.
- If two variables show roughly identical errors with the opposite sign, this can come the fact that one of them has erroneous values and explains the other.

For instance if historical values for Q are overestimated, the relative error on UR and Q will be similar with different signs.

$$UR = Q/CAP$$

$$Q + M = FD + X$$

5.3.3.2 Processing errors

Diagnosing errors in the residual check phase can lead back to different phases of the modelling process:

- Data management: the data obtained from external producers is not consistent, for a full series or for specific observations (this happens!).
- Production of model data: using the wrong original series, using a wrong computation.

Example: using a variable at current prices instead of constant, or forgetting an element in a sum.

- Specification of the model (badly written equations).

Example: forgetting the term for Housing investment in the definition of demand.

- Estimation (modified series since estimation, bad coefficients).

Example: an error in the imports equation shows that the explanatory series for domestic demand has been changed since estimation.

Applying this process a number of times will be necessary to produce a coherent model.

5.3.3.3 Back to the example

Producing a residual check is quite easy in EViews: one just has to specify the option “d=f” in the SOLVE statement:

```
_fra_1.solve(d=f)
```

Of course, as all equations will be computed separately, all information must be available on the current sample period, including the endogenous variables (which should be exogenous somewhere else). Contrarily to computations and estimations, EViews does not adapt the simulation process to the feasible period (this seems rather logical).

As the model is recursive (super-recursive?) computation gives the result directly, and no element describing the solving method is needed (we shall see them later).

However:

- One should specify the name given to the computed variables.

Every time EViews has to solve a model, the name given to the results will be built from the original name of the variable, to which will be added a suffix (a prefix is also possible but less manageable in our opinion). This avoids destroying the original information, and allows comparing alternate solutions.

The prefix is specified using the statement:

```
modelname.append @all suffix
```

(remember: **append** adds text to the model, an identity equation is only a special case of text).

In our case, applying the suffix “_C” calls for:

```
_fra_1.append @all _C
```

The equation for FD will give FD_C, which we can compare with the actual values of FD.

Computing the differences between actual and computed values can be done in a loop, using the syntax described later. The elements in the loop can be defined “by hand” but it is more efficient to use the “**makegroup**” statement.

```
_fra_1.makegroup(a,n) groupname @endog  
_fra_1.makegroup(a,n) groupname @exog
```

In our case:

```
_fra_1.makegroup(a,n) g_vendo @endog  
_fra_1.makegroup(a,n) g_vexo @exog
```

Two remarks:

- You surely wonder about the reason for the (a,n). This modifies the default options of the “**makegroup**” statement, which would produce a group with the baseline names (in our case with _C added) and leave out the actual names. Stating (a,n):
 - Introduces the actual names (a for actual)
 - Eliminates the baseline ones (n for no baseline)

It would be best to restrict the computations to the identities. The residuals on the “estimated” have no meaning: as the “f” scalar is null, the right hand side will be computed as zero, and the percentage error as 100% as $100 \cdot (\text{value} - 0) / \text{value}$. But being able to compute the whole model proves that estimations can be conducted on that period.

One can create two sub-groups by

```
group g_vbeha CI I LE M X  
group g_viden CAP CO FD IC IH K LT Q RHI TD UR
```

Or

```
group g_vbeha CI I LE M X
_fra_1.makegroup(a,n) g_viden @endo
g_viden.drop CI I LE M X
```

This creates first a full group `g_viden`, then eliminates the estimated from it.

This last technique is clearly inefficient here, but will be much more with a 500 equations model with 50 estimated ones (a more usual situation).

However, both techniques call for a user-defined list, which will have to be updated each time the variable set is modified, something we want to avoid: we propose using a more complex, but automatic one.

A tip: A visual check is made difficult by the relative imprecision of EViews, which often produces small residuals for exact equations. In scientific format, these residuals appear as numbers with high negative exponents are hard to identify. One solution is to move to a fixed decimal presentation, by selecting a zone (in the “spreadsheet” view) then using the right mouse button to access “display format” then “fixed decimal”.

A simpler solution to observe if there is no error is to display all the residuals as a single graph, and look, not at the series (they should move around in Brownian motion) but at the scale: both maximum and minimum must be very small.

Another idea is to transfer the residuals to Excel and sort the sheet (unfortunately EViews does not sort a sheet across series on the values at a given period). The non-negligible elements should appear at the top and the bottom according to their sign and the sorting order. This technique takes more time but allows to identify immediately and fully the faulty elements.

5.3.3.4 A trick: generating the groups of identities and behavioral

You certainly have realized by now (and you knew it probably before anyway) that one should avoid as much as possible having to edit the text of modeling programs, each time changes have been made earlier in the process. This represents at best extra work, at worst a source of error. We have just violated this principle, by separating by ourselves the endogenous into behavioral and identity.

This will introduce problems, in particular in large models: the initial process will be tedious and error prone, and one will have to remember to update the list every time the model structure changes.

We propose a simple technique to avoid this, and make the initial separation and its updating automatic. It is based on the presence of the “*f*” scalar in the behavioral equations.

We just have to:

- Simulate the model with the option “*d=f*” and *f*=1, saving the results under a given suffix
- Set *f* to 2 and update the model (this is necessary to take into account the change).
- Simulate the model again with *f*=2 and another suffix.

- Create empty groups of estimated and identity variables.
- Produce a loop over the whole group of endogenous, and test each time if the results of the two simulations are different.
- * If they are, add the variable to the list of estimated elements.
- * If not, to the list of identity elements.

We can use the following program (for the period 2000 – 2002). We suppose that any percentage error higher than 0.00001 denotes an error.

```

_fra_1.makegroup(a,n) g_vendo @endog
_fra_1.makegroup(a,n) g_vexo @exog
group g_varia g_vendo g_vexo

group g_vbeha 'creates an empty group
group g_viden 'creates an empty group
smpl 2000S1 2002S2
_fra_1.append assign @all _c
scalar f=0
solve(d=f) _fra_1
scalar f=1
_fra_1.update
_fra_1.append assign @all _d
solve(d=f) _fra_1

for !i=1 to g_vendo.@count
%1=g_vendo.@seriesname(!i)
series pf_{%1}=100*({%1}_d-{%1}_c)/({%1}_c+({%1}_c=0))
if @max(@abs(pf_{%1}))>1e-5 then
g_vbeha.add {%1}
else
g_viden.add {%1}
endif
next

```

This sequence calls for some explanation.

- The loop (“for” to “next”) is reproduced for each variable in the list g_vendo. The number of these variables is g_vendo.@count. (For EViews, x.@count is an integer scalar containing the number of elements in group x).
- !i is the rank of the variable in the group g_vendo (from 1 to g_vendo.@count).

- %1 receives as a character string the contents of `g_vendo.@seriesname(!i)` , the name of the variable in group `g_vendo`, with rank `!i`.
- The subsequent formulas replace %1 by its string value, and brackets are dropped leaving the characters in the statement.

For regular users of EViews, or people familiar with programming, the above was probably clear. For others, this is the time to give very basic information about EViews programming (even if this is not the purpose of this book).

5.4 USING LOOPS AND GROUPS IN EIEWS

In the programs we are going to present, intensive use is made of two elements: groups and loops.

5.4.1 GROUPS

Groups are named elements which refer to a set of objects (which can be series, series expressions but also other objects), allowing to treat them either as a whole or in sequence.

The statement creating a group is

```
group name-of-the group list-of-elements
```

For instance

```
group g x y z x/y
```

will create a group named `g` containing the three series `x`, `y` and `z` and the ratio of `x` to `y`.

The element must be series of expression, but one can cheat by creating artificial series with the name of the requested element.

One can:

- Group groups
- Add and drop elements from groups:

```
g.add a
```

will add the series `a` to the group `g`.

```
g.drop x
```

will drop the series x from the group g.

Two useful elements can be associated with the group:

`g.@count` is a scalar which contains the number of elements of group g.

`g.@seriesname` is a character vector which contains the names of the series in group g.

Finally, groups can be created through a mask:

<code>group g_fra fra_*</code>	will create a group from all the elements starting with fra_ an underscore.
<code>group g_GDP ???_GDP</code>	will create a group from all the GDPs of OECD countries (using three characters as a label).
<code>group g_3 ???_*</code>	will create a group from all the elements starting with three characters, then

Groups can be used to display a list of series, as spreadsheet or graph, by double-clicking on its name in the workfile window (where they appear as a blue “G” symbol) or calling for it.

The default display is a spreadsheet format, but one can move to graphs using the “View” button then “graph”, or even editing the list of elements by “View” + “group members”.

5.4.2 LOOPS

EViews allow two kinds of loops:

- By elements (a list or a group)

The syntax is:

```
for %parameter list-of-variables or group-name
block of statements including {%parameter} or %parameter
next
```

The block of statements will be repeated in sequence for each element in the list, which will then replace the parameter.

The presence of brackets around the parameter changes its status. With brackets the associated characters are included in the statements, then the brackets are dropped. Without brackets the parameter is considered as a character string variable.

For instance with

```
%1="111"
```

The statement

```
genr xxx={%1}
```

will give to the series xxx the value 111,

while

```
xxx=%1
```

will create a character string with the value "111"

The statement

```
genr xxx=%1
```

will be illegal as it tries to transfer a character string to a series.

We get the message:

```
can not assign string expression to numeric variable in "GENR XXX="111""
```

On the other hand, the statement:

```
%2=%1+"333"
```

Will create a "111333" string, while

```
%2={%1}+"333"
```

will be illegal as it mixes strings and values:

```
Scalar assigned to string in "%2=111+"333""
```

- By integer number.

The syntax is:

```
for !name=first-integer to second-integer by third integer
block of statements including {!parameter}
next
```

The block of statements will be repeated in sequence from first-integer to second-integer, incrementing if necessary by third-integer, the value replacing the parameter.

Integers can be negative. If third-integer is omitted, the increment will be 1.

This type of loop can also be applied to a group

```
for !integer=1 to group-name. @count
%1=group-name. @seriesname(!integer)
block of statements including !integer, %1, {%1}
next
```

- group-name.@count is the number of elements in the group group-name.
- %1 receives the contents of group-name.@seriesname(!i) , the name of the variable in group group-name, with rank !integer.

5.5 COMPARING WORKFILES : THE WFCOMPARE COMMAND

During the modelling process, you often have to compare two sets of information.

In particular, you might want to:

- Make sure that two sets of data are identical. This applies to the results of a program you are running again, maybe after a long delay.
- Control the evolution of historical values for a model data set, showing for instance which equations will have to be estimated again.
- Summarize the results of a residual check, showing for which equations the right hand side (using historical values of the explained variable) is different from the right hand side (the result of the computation). By setting a tolerance level slightly higher than zero (for instance 0.0001) one can restrict the display to the errors deemed significant.
- Or you just might want to know which elements of a set are present in another set, for instance which available series are actually used by one model.

This can be done easily, using the **wfcompare** command.

You can compare elements between workfiles and pages inside the same workfile. EViews will display one line per element, in which will be stated its relation, between: unchanged, modified (numerically), added, deleted, replaced (logically, the last case applies for instance to a linked variable have been modified). A filter can be applied.

For series, a tolerance level can be set, under which the series are not considered modified. The display will tell how many periods show a higher difference.

By default, all elements will be displayed, but one can restrict the case (for instance, to all variables present in both pages with a difference higher than the criterion).

Equations and models are not compared but appear in the list.

The syntax of the **wfcompare** command is:

```
wfcompare(tol=criterion,list=comparison_type) list_of_compared_series list_of_reference_series
```

For more details you should refer to the EViews Help.

For instance if you want to compare all French series (starting with "FRA_") between the pages "base" and "updated", for a tolerance level of 0.00001 one will state:

```
wfcompare(tol=1E-5,list=m) updated\fra_* base\fra_*
```

CHAPTER 6 THE ESTIMATION OF EQUATIONS

We now have

- A full description of the framework of the model, in which all the identities are completely specified, and the intents in terms of behaviors are described as clearly as possible.
- A full database containing all the series in the present model, endogenous and exogenous, with their description.

We have also checked that:

- The specification of identities is consistent with the available data.
- The information obtained on the structure of the model (causalities, interdependencies) is consistent with our economic ideas.

Both the list of variables and equations are available as printable documents.

The next stage is obviously to replace each of the tentative behaviors by actual ones, validated both by economic theory and statistical criteria.

6.1 THE PROCESS OF ESTIMATION

What we are proposing is not a book on econometrics, and anyway we will never be as knowledgeable, by far, as the EViews team, both in terms of theory and ability to teach it.

This means we will not approach the theoretical aspects of the subject, leaving the reader to the use of the books we propose in our bibliography, or even to the EViews Help manuals, which can be actually used as teaching tools, as they are both comprehensive and very progressive in their approach.

But once the modeler is familiar with the concepts, their application to an actual case⁵² is not straightforward at all. This means we think this book can bring a very important contribution: showing how these methods can be used in the process of building our models. The reader will learn how, in very specific cases, first very basic then more operational econometrics can be used (or not used), considering the information he has and the goal he is pursuing.

We shall also show the role econometrics take in the process, not as a single task between data gathering and simulations, but as a recurrent partner in the iterative process of model building.

We shall not only give examples working smoothly from the start, but show also how econometrics can be set aside, and how, in some cases, an initial failure can be transformed into success, with some imagination⁵³.

6.2 SPECIFIC ISSUES

Nevertheless, we feel it will be useful to present two cases, which are not generally treated by manuals, and can lead to wrong decisions, or wrongly evaluating the results of tests.

We shall use a very practical approach.

6.2.1 THE R² OR R-SQUARED

The statistic called "R²" or "R-squared" is the most commonly used to judge the global quality of an estimation. It is defined by the following formula.

$$R^2 = \frac{\sum_{t=1}^T (\hat{x}_t - \bar{x})^2}{\sum_{t=1}^T (x_t - \bar{x})^2}$$

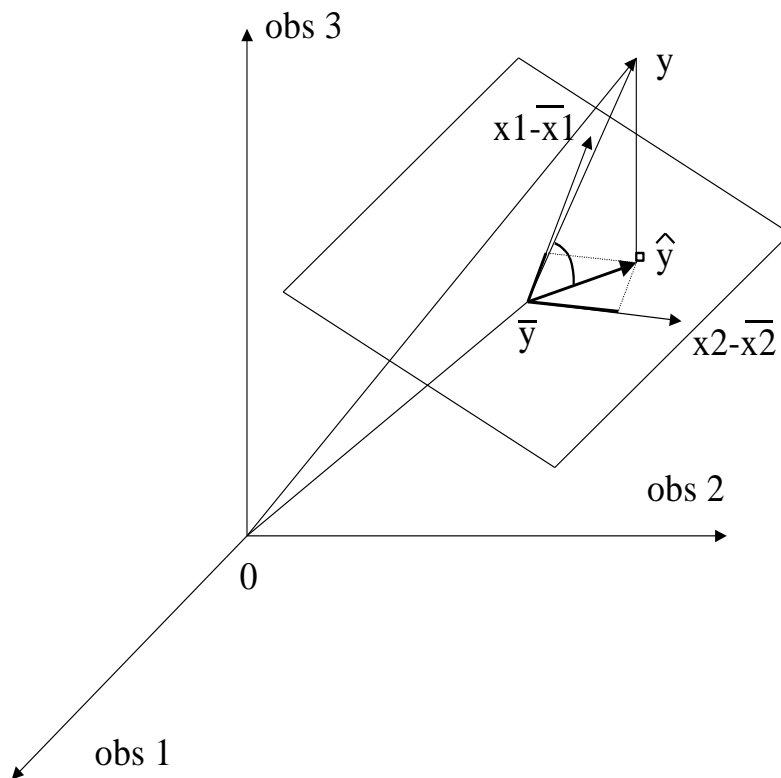
⁵² One in which he is not playing with data, but actually obliged to succeed.

⁵³ Remember David Hendry's four golden rules of econometrics: 1. Think brilliantly, 2. Be infinitely creative, 3. Be outstandingly lucky, 4. Otherwise, stick to being a theorist.

This statistic can therefore be interpreted as the share of the variance of the observed variable x explained by the estimated formula.

A geometrical explanation also can be used: if we consider the space of variables (dimension T = number of observations), the estimation method will consist in minimizing the distance between the explained variable and the space (the plane or hyper plane) generated by the vectors of explanatory series, using combinations of parameter values.

Especially, if the formula is linear relative to estimated parameters and contains a constant term, we can consider the estimation is based on the difference of variables (explained and explanatory) to their means. In this case, minimizing the Euclidian distance will lead (as can be seen on the graph) the vector $(\hat{y}_t - y_t)$ to be orthogonal to the space and therefore to the vector $(\hat{y}_t - \bar{y})$. These two elements represent the non-explained and explained part of $(y_t - \bar{y})$, the variance of which is the sum of their squares. The R^2 can be interpreted as the square of the cosine of the angle between the observed and adjusted series: the closer the R^2 is to 1, the smaller the angle will be and the higher the share of the estimated variable in the explanation of the total variance. The explanation will be perfect if $y - \bar{y}$ belongs to the space, and null if the perpendicular meets the space at the origin.



If the equation presents no constant term, the same reasoning can be applied, but this time the mean is not subtracted. However, the R^2 no longer has the same meaning: instead of variances, direct sum of squares will be used.

We will not go further in the explanation of this test, concentrating instead on its practical properties.

6.2.1.1 Questioning the R-squared

One must be very careful when using the R^2 statistic.

- it increases with trends in variables

The R^2 statistic will be all the higher as the explained variable and at least one of the explanatory variables present a time trend according to the rank of the observation. Thus components of each of these variables on axes of observations will grow in the same or opposite direction (from highly negative to highly positive or the reverse), and give associated vectors very close orientations. On the above graph, the components of variables on the axes will be more or less ordered according to the numbering of the axes themselves. The first observations will be the most negative, then values will grow through zero and reach the most positive ones in the end. The same goes if the ordering follows opposite directions: the estimation will evidence a negative link.

In this case, even in the absence of a true causal relationship between variables, the orientation of the vectors will be similar to a given multiplicative factor, and the R^2 test will seem to validate the formulation. And most time series (like values, quantities or prices), generally present a growing trend, giving this phenomenon a good chance to happen. For example if we apply the previous equation for French imports:

$$(1) \quad \text{Log}(M_t) = a \cdot \text{Log}(TD_t) + b + u_t$$

Replacing TD by any steadily growing (or decreasing) variable⁵⁴ will give a “good” R^2 , better maybe than actual French demand.

Actually it can be shown that testing for each OECD country the estimation of its imports as a function of the demand of any country, the “true” equation does not come always as the best, although it is never far from it.

- It gives misleading diagnoses when comparing estimations explaining different elements.

This happens in particular when we explain the same concept using a different transformation.

Let us consider our equation 14, as

$$(14) \quad \Delta \text{Log}(M_t) = a \cdot \Delta \text{Log}(TD_t) + b + v_t$$

We can see that the time trend has disappeared from both series, and any correlation will come from common deviations around this trend (or rather common changes in the value from one period to another). This is of course a much better proof of a link between the two elements (independently from autocorrelation).

To put the two formulations on equal grounds, they must explain the same element. For this, one can just modify the new equation into:

$$\text{Log}(M_t) = \text{Log}(M_{t-1}) + a \cdot \Delta \text{Log}(TD_t) + b + v_t$$

Compared to the initial formula, this transformation will not change the explanation⁵⁵, as obviously the minimization of the sum of squared residuals represents the same process. The only modified statistic will be the R^2 , which will increase a lot, as an identical element with a high variance (compared to that of $\Delta \text{Log}(M_t)$) has been added on both sides.

⁵⁴ Like Australian demand, or the price of a pack of cigarettes in Uzbekistan.

⁵⁵ Before estimation EViews will move the lagged term to the left.

The choice between the two formulations should not rely on the R^2 but on the autocorrelation of the residual: if u_t is not correlated one should use (1), if it is one should try (2). But in any case the issues will be solved by error correction models and cointegration, which we shall address later..

6.2.2 THE CONSTANT TERM

When observing the validity of individual influences, one element plays a very specific role: the constant term.

This element can have two purposes:

- To manage the fact that the equation does not consider elements as such, but the deviations from their means. In ordinary least squares, even if the final result is a linear formulation of the variables and a constant term, the process actually
 - computes the deviations
 - uses them to estimate a formula with no constant⁵⁶
 - recombines estimated coefficients and means into a constant

This constant is an integral part of the process. It should be included every time at least one of the explanatory elements does not have a zero mean.

- To describe an economic mechanism.

Let us give an example for the first case: if imports have a constant elasticity to demand, we will estimate:

$$\Delta M_t / M_t = a \cdot \Delta TD_t / TD_t$$

Or

$$\text{Log}(M_t) = a \cdot \text{Log}(TD_t) + b$$

but the estimation process will first use the difference to the average to get “a”

$$\text{Log}(M_t) - \text{Log}(\bar{M}) = a \cdot (\text{Log}(TD_t) - \text{Log}(\bar{TD}))$$

or

⁵⁶ As all elements in the formula have zero mean, the sum of the residuals will also

$$\text{Log}(M_t / \bar{M}) = a \cdot \text{Log}(TD_t / \bar{TD})$$

Then the constant

$$b = \text{Log}(\bar{M}) - a \cdot \text{Log}(\bar{TD})$$

We can see in particular the consequences of a change in the units (thousands, millions, billions...). The constant term will absorb it, leaving “a” unchanged. In the absence of “b”, “a” will get a different value, for no economic reason.

Of course, the more “b” is significant, the more its absence will be damaging to the quality of the estimation (and the more “a” will be affected). But this is no reason to judge “b”. It is as if we are weighting an object with a balance: the two platters never have the same weight, and even the damage increases with the difference, it is always useful to correct it. And in our case there is no cost (actually it makes things cheaper, as the cost of the decision process disappears).

It is not frequent for the constant term to have a theoretical meaning. The majority of such cases come from a formula in growth rates or variations, where the constant term will be associated with a trend.

The only justification for the absence of a constant term is when this theoretical constant is null. In this case, observing a significant value becomes a problem. We shall give an example soon.

6.3 APPLICATIONS: OUR MODEL

Let us now apply the above principles to our sample model.

In our model, we have to estimate five equations, for which we have already ideas about their logic:

- The change in inventories, employment and investment should depend on GDP
- Exports and imports should depend on the associated demand (world and domestic) and availability of potential supply.

We shall use each of these equations to illustrate a specific aspect of estimation.

- The change in inventories: general elements, homoscedasticity, presence of a constant term.
- Employment: stationarity, error correction models.
- Investment: the necessity to establish a consistent theoretical equation prior to estimation.
- Exports: autoregressive processes, cointegration, long term stability.
- Imports: going further on cointegration and long term stability.

Each of our formulations will be based on very simple economic ideas, and we shall elect a specification which complies with both econometric tests and economic consistency. They are also chosen in a way which should allow them to merge harmoniously into the model we are building. However, it should be clear that

- Other simple formulations could probably be built on the same sample, with equivalent or maybe better quality.
- Using another sample (another country for instance) the same economic ideas could lead to different formulations (not only different coefficient values).
- Other economic ideas could be applied, with the same complexity.
- To produce a truly operational model, the present framework would have to be developed in a large way. We will present such developments later.

However the model we are building represents in our sense a simplified but consistent summary of the general class of models of this type. Reading descriptive documents for any operational structural model, one will meet many of the ideas we are going to develop.

A note on estimations: one can observe that we are using rather old data, with a bi-yearly periodicity, half way between the two most usual ones: quarterly and yearly. We do have access to the same information on a larger period on a quarterly basis, which would have represented an obvious improvement. However our course proposes also a series of lessons, following the same lines (and using very similar ideas) as the ones we will develop here.

Another reason is that the most recent National Accounts produce chained series which means that the elements "at constant prices" become "at the prices of the previous year". Using them as such makes econometric estimations difficult and modelling almost impossible. And transforming them into "traditional" year based series cannot be done in a rigorous way.

Obviously the lessons (which ask users to solve a set of problems) cannot use the same data as the examples. We have decided that it was more important to keep the shorter periodicity for the tests, which will represent a more important element in the teaching process.

6.3.1 CHANGE IN INVENTORIES

We shall use this simplest estimation to present the basic features of EViews estimation, and also stress the necessity for homoscedasticity.

Our formulation will suppose simply that firms desire a level of stocks proportional to their production (or GDP). For a particular producer, this should be true both for the goods he produces and for the ones he is going to use for production. For instance, a car manufacturer will allow for a given delay between production and sale (for instance three months, which will lead to an inventory level of 1/4th of annual production). And to be sure of the availability of intermediary goods (like steel, tires, electronic components and fuel for machines in this case) he will buy the necessary quantity (proportional to production) sometime in advance.

We shall suppose that firms have achieved, at the previous period, an inventory level IL representing a number of semesters of production:

$$IL_{t-1} = a \cdot Q_{t-1}$$

And they want to keep this level at the present period:

$$IL_t^* = a \cdot Q_t$$

$$IL_t^* = IL_t$$

Then the change in inventory will represent:

$$IC_t = (IL_t - IL_{t-1}) = a \cdot \Delta Q_t$$

This means that contrary to the general case this equation should not include a constant term. Its presence would call for a trend (and a constant) in the equation in levels, with no economic justification. It would also introduce a problem: adding a constant to an explanation in constant Euros would make the equation non-homogenous.

Even then, the equation faces a problem, concerning the residual: between 1963 and 2004, French GDP has been multiplied by 4. We can suppose the level of inventories too (maybe a little less with economies of scale and improved management techniques).

It is difficult to suppose that the unexplained part of the change in inventories is not affected by this evolution. As the variable grows, the error should grow. But to apply the method (OLS), we need the residual to have a constant standard error. Something must be done.

The simplest idea is to suppose that the error grows at the same rate as GDP, which means that if we measure the change in inventories in proportion to GDP, we should get a concept for which the error remains stable. Of course, we shall have to apply the same change to the right hand side, which becomes the relative change in GDP.

To avoid causality problems (for a given semester, demand for IC is partly satisfied by Q) we shall use the previous value of Q.

The equation becomes:

$$IC_t / Q_{t-1} = a \cdot \Delta Q_t / Q_{t-1}$$

6.3.1.1 The basic EViews estimation features

As this is our first example, when shall use it to present the basic estimation features.

Actually the technique will be different according to the stage in the estimation process: whether we are exploring several individual formulations, looking for the best option both in statistical and economic terms, or we have already elected the best one, and want to merge it into our model.

We shall start with the first situation.

The simplest way to estimate an equation under EViews is through the menus, using in succession:

Quick > Estimate equation

A window appears, in which one has to type the formula.

In the case of ordinary least squares, this can be a list of elements separated by blanks, in our case:

IC/Q(-1) D(Q)/Q(-1)

We can also use

IC/Q(-1)=c(1)*D(Q)/Q(-1)

The two methods give exactly the same results (in the first case, the “c” vector will also be filled with the estimated coefficient).

The default method will be Least Squares, appropriate in our case. If the equation was not linear in the coefficients, the second presentation would be automatically called for.

One will note that

- A constant term has to be introduced explicitly (as an additional element called “C”).
- EViews allows to specify a sample, which will be applied only to the particular equation (the current sample is not modified). This is quite useful if some periods have to be excluded from the estimation. This will happen for instance if they are deemed not to follow the estimated behavior (like pre – transition data for Central European countries, China or Vietnam), or observations are also provided over the future (OECD’ Economic Perspectives completes the historical data with the results of its forecasts over the next three years).
- On the contrary, one does not have to care about leaving in the sample periods for which estimation is not possible, due to missing elements or the impossibility to compute a term (for instance the logarithm of a negative value). EViews will eliminate by itself the corresponding periods (and tell you about the reduced sample).

In our case we can use the sample:

smpl 1960s1 2002s1

which means that we consider data from the first semester of 1960 to the last of 2002 (our data is bi-yearly in this particular case).

If the equation is linear in coefficients, EViews recognizes this property, and does not try to iterate on the coefficients, as it knows the values found are the right ones.

Using the “Ok” button gives the following results

We can see that EViews gives the sample used (the relevant periods of our sample). Estimation starts in 1963S2, as the Q series starts only in 1963 and is lagged once.

- We get also the number of periods, and the time and date.
- The other elements are the usual statistics, described earlier. The most important are:
 - The R-squared, the Durbin-Watson test and the Standard Error of regression for global elements.
 - The coefficient, the t-Statistic and the probability for reaching the coefficient value if the true coefficient is null with the estimated standard error.

In our case:

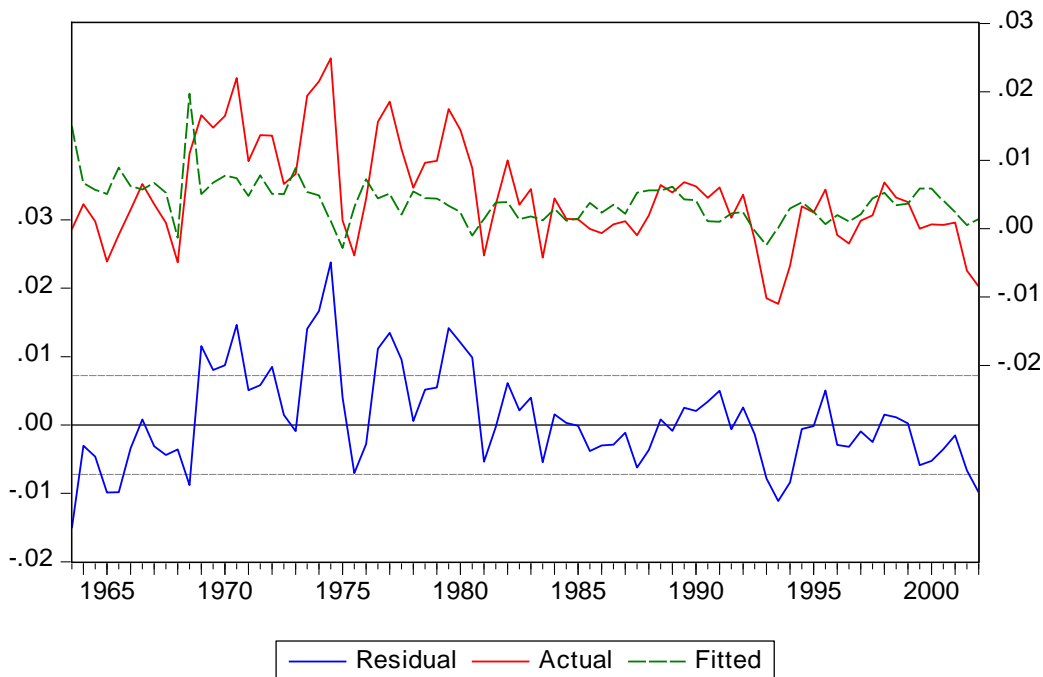
- The R-Squared is very low, even if the extreme variability and the absence of trend of the left-hand element plead in favor of the explanation⁵⁷.

However, as with almost all homogenous estimations, a simple interpretation is available, through the standard error: as the explained variable is measured in points of GDP, this average error represents 0.72 points.

- The coefficient is very significant. The probability of reaching 0.26 for a normal law with mean 0 and standard error .042 is measured as zero. Of course it is not null, but it is lower than 0.00005, and probably much so.
- But the Durbin-Watson test takes an unacceptable value, even if the absence of a constant term (and the subsequent non-zero average of residuals) makes its use questionable.
- The graph of residuals is the second important element for diagnosis. It shows the evolution of actual and estimated series (top of the graph, using the right hand scale) and of the residual (bottom, using the left hand

⁵⁷ If we knew the values for IL, its estimation would get a better R^2 (due to the colinearity of LI and Q). But we would be led to estimate an error correction model on IL, anyway. We have seen the advantage of this formulation, but for the quality to extend to the whole model, all equations must be of this type.

scale, with lines at + and – 1 standard error). This means that inside the band residuals are lower than average, and higher outside it. Of course, it gives only a relative diagnosis.



The graph shows (in our opinion) that the equation provides some explanation, but some periods (1975-1980 in particular) present a large and persistent error, and there seems to be a negative trend on residuals after 1975 (and maybe a positive one before).

In addition to the display of estimation results and graph of residuals, EViews creates several objects:

- A vector of coefficients, contained in the “C” vector. The zero values or the results from the previous regression are replaced⁵⁸.
- A series for the residuals, contained in the “RESID” variable. The “Na” values or the results from the previous regression are replaced⁵⁹.
- A tentative equation, called “Untitled” for the moment, and containing the developed formula, with “c” as the vector of coefficients, with numbers starting from 1. In our case, the formula is obviously

$$IC/Q(-1)=c(1)*D(Q)/Q(-1)$$

Any subsequent estimation will replace this equation by the new “Untitled” version.

⁵⁸ But if the present regression contains fewer coefficients than the previous ones, the additional elements are not put to zero .

⁵⁹ But this time, residuals from previous equations are given either computed values or « NA ».

- EViews provides also several options, accessed from the menu, and which can be useful:
 - **View** gives three representations of the equation: the original statement, and two formulas including coefficients as parameters (the above “c” type) or as values.
 - **“Print”** allows printing the current window: to a printer, to a text file (using characters, which saves space but reduces readability, especially for graphs), or to a graphics RTF file. This last option might call for a monochrome presentation, which is obtained through the « Monochrome » template (the last of the general Graph options).
 - **“Name”** allows creating the equation as a named item in the workfile, with an attached comment. It is important to use it immediately after the estimation, as the temporary equation (named “untitled”) will be replaced by the next estimation.

However, inserting an underscore (“_”) before the name proposed will place the equations in the first positions of the working window.

EViews proposes as a standard name “EQ” followed by a two-digit number, following the lowest one unused at the moment. There are two options:

- * Give a name representative of the equation (like “EQ_X3U” for the third equation estimating X as influenced by the rate of use).
- * Accept the EViews suggestion and rely on the attached comment for the explanation.

Personally we favor the second option:

- * It is simpler and more natural to use.
- * It allows placing all the equation in the same workfile (and window) location.
- * It avoids defining a complex and maybe unclear naming method.
- * The comment zone is much wider and can follow any format, including blanks and special characters.

Actually the item saved is more complex than the actual formula. Double-clicking on it shows that it contains the full representation, including the residual (and actually the standard errors of the coefficients, even if they are not displayed).

- **Forecast** produces a series for the estimated **variable** (or the estimated left-hand expression, generally less interesting), and an associated graph with an error band (and a box with the statistics).

6.3.1.2 An alternate technique: using the command window

Instead of using Quick>Estimate, one can work directly through the command window. One just has to add “ls” before the formula.

```
ls IC/Q(-1)=c(1)*D(Q)/Q(-1)
```

This has several advantages:

- By copying and editing the current equation on the next line of the command box, entering changes is made much easier.
- After a session of estimations, the set can be copied into a program file and reused at will. Management of a set of alternate versions is much easier.
- One can control the size of characters. This is quite interesting when working with a team, or making a presentation, as the normal font is generally quite small.
- The only drawback is sample definition: it has to be entered as a command, not as an item in the “estimate” panel.

6.3.1.3 Other possible specifications

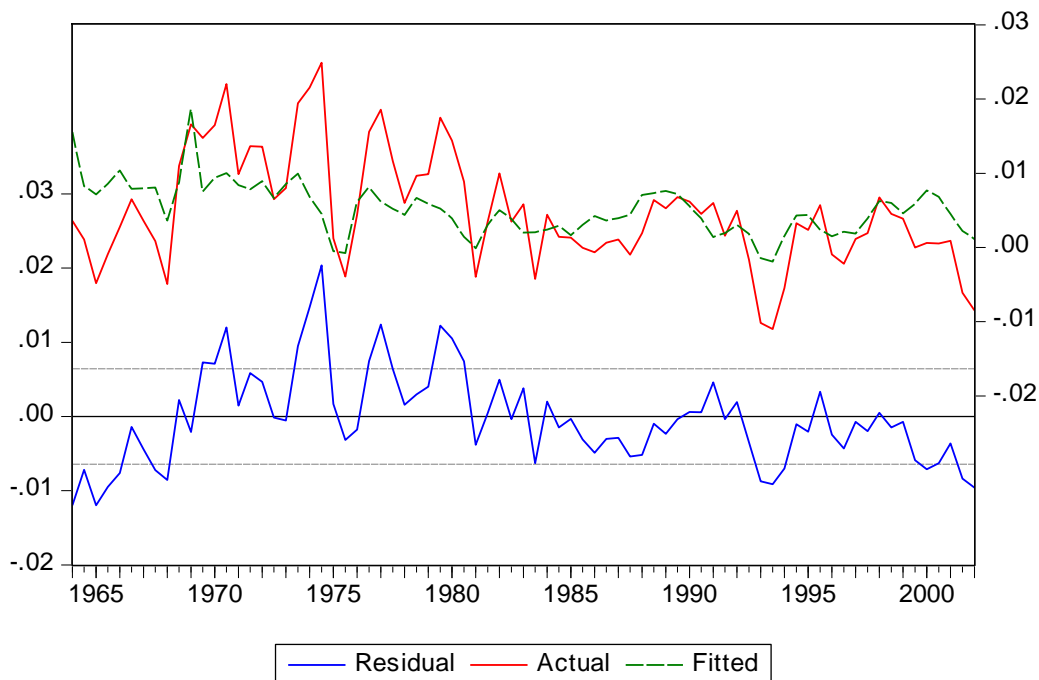
Let us go back to our estimated formula. If we are not satisfied with the previous results, we can try alternate options, without changing the economic background:

Firms could consider the changes in GDP for the last two semesters, with different impacts.

Dependent Variable: IC/Q(-1)
Method: Least Squares
Date: 11/06/12 Time: 19:37
Sample (adjusted): 1964S1 2002S1
Included observations: 77 after adjustments
IC/Q(-1)=C_IC(1)*@PCH(Q)+C_IC(2)*@PCH(Q(-1))

	Coefficient	Std. Error	t-Statistic	Prob.
C IC(1)	0.130473	0.056396	2.313510	0.0234
C IC(2)	0.214330	0.053280	4.022712	0.0001
R-squared	0.280583	Mean dependent var		0.004821
Adjusted R-squared	0.270991	S.D. dependent var		0.007513
S.E. of regression	0.006414	Akaike info criterion		-7.234887
Sum squared resid	0.003086	Schwarz criterion		-7.174009
Log likelihood	280.5432	Hannan-Quinn criter.		-7.210536
Durbin-Watson stat	0.658613			

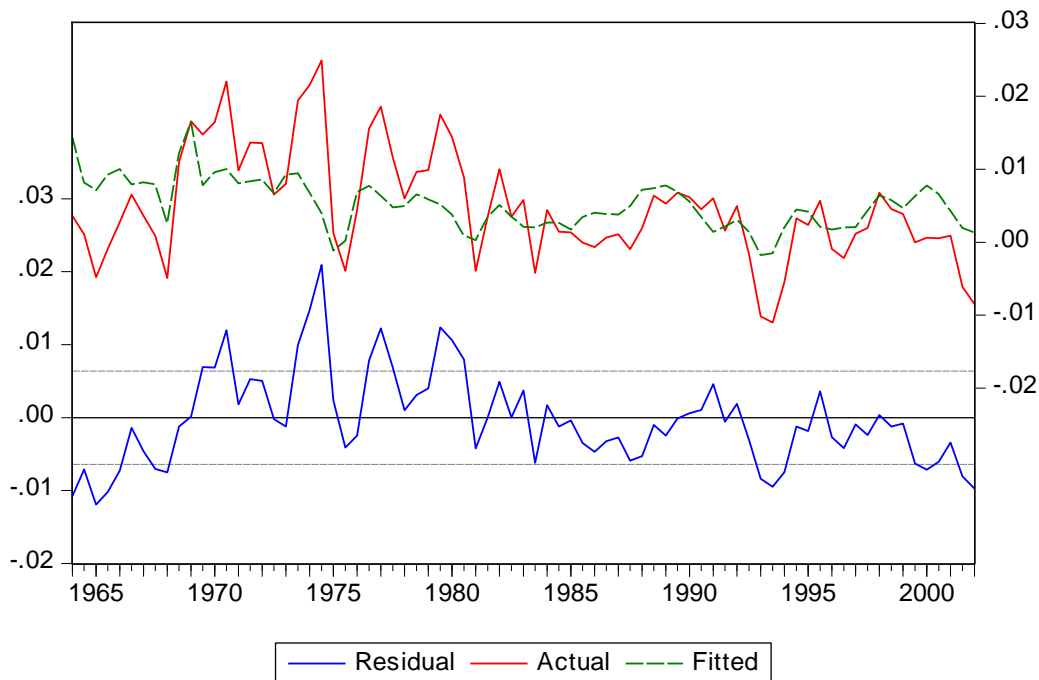
The results are actually better (not the Durbin-Watson test!).



- We can also consider the same coefficient:

Dependent Variable: IC/Q(-1)
 Method: Least Squares
 Date: 11/06/12 Time: 19:37
 Sample (adjusted): 1964S1 2002S1
 Included observations: 77 after adjustments
 IC/Q(-1)=C_IC(1)*@PCH(Q)+C_IC(1)*@PCH(Q(-1))

	Coefficient	Std. Error	t-Statistic	Prob.
C_IC(1)	0.173786	0.020427	8.507613	0.0000
R-squared	0.274067	Mean dependent var		0.004821
Adjusted R-squared	0.274067	S.D. dependent var		0.007513
S.E. of regression	0.006401	Akaike info criterion		-7.251844
Sum squared resid	0.003114	Schwarz criterion		-7.221405
Log likelihood	280.1960	Hannan-Quinn criter.		-7.239669
Durbin-Watson stat	0.628277			



This restriction does not reduce the quality (the R-Squared decreases but the standard error too, a strange result due to the larger number of degrees of freedom, with one less estimated coefficient).

Actual observation of the residuals shows a growing trend before 1975, and a decreasing trend from that date. One can be tempted to take this into account, a prospect which is not completely unlawful: in the latter years the policy of firms has been to reduce the level of inventories (thus their change) and the technical opportunities of implementing it have increased. The pre-1975 increase is harder to explain.

We shall introduce a correcting term, represented by a constant, a trend starting at 1975, and another ending at 1975.

6.3.1.4 Introducing trends in equations

To introduce a trend in an equation (or for that matter any expression depending only on time) two solutions are available:

- Create an (exogenous) variable.
- Introduce directly the associated formula, using a trend variable.

We prefer clearly the second option:

- It reduces the number of elements in the model.
- It requires no extrapolation for simulations over the future.
- It allows a direct interpretation by the modeller or other persons, and does not call for documentation.

We see no relative advantage for the first option.

To define the trend, we have two options:

- Using the @trend or @trendc function of EViews, giving a variable starting from zero and increasing by 1 at each period⁶⁰.
- Creating a specific time trend, taking the value of the year, thus increasing by 1 at each year. Obviously, for non-annual series, one will have to separate sub-periods. The obvious solution is to set the first period to the calendar year, and increase subsequent observations by a fraction, giving 1 over the year.

For the first semester of 1975 the value will be 1975, for the second semester 1975.50.

Unfortunately this technique faces problems with a more than quarterly frequency, a rare occurrence in modelling however.

We shall estimate indifferently:

$$IC/Q(-1) = D(Q)/Q(-1) + D(Q(-1))/Q(-2) + C(1)*(T-1975)*(T \leq 1975) + C(2)*(T > 1975)*(T-1975) + C(3)$$

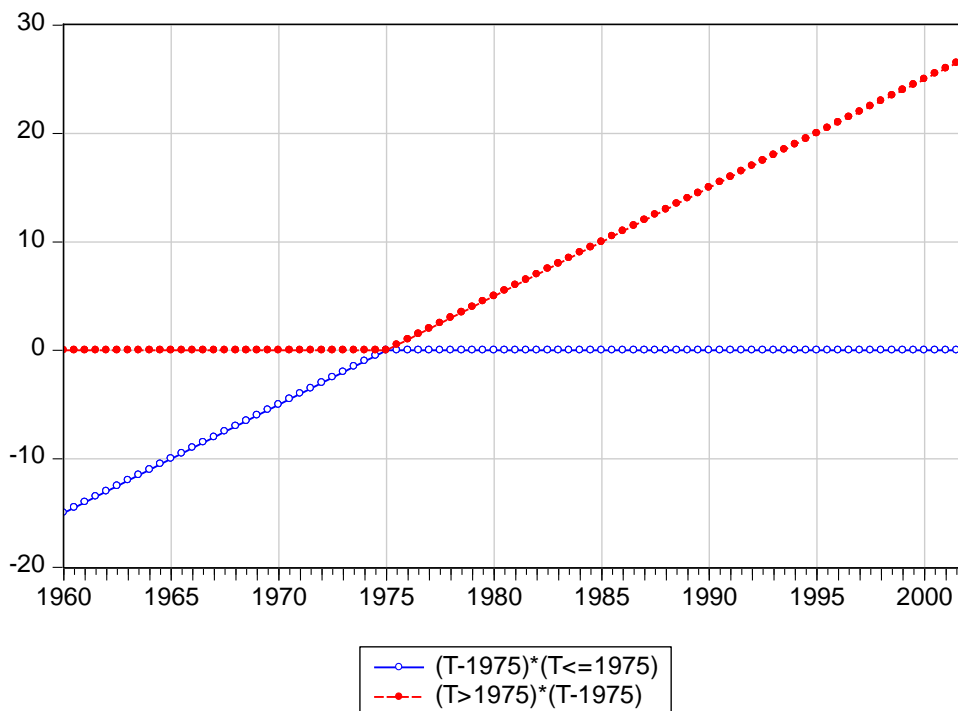
or

$$IC/Q(-1) = C(1)*D(Q)/Q(-1) + C(2)*D(Q(-1))/Q(-2) + C(3)*(T-1975)*(T \leq 1975) + C(4)*(T > 1975)*(T-1975) + C(5)$$

Let us detail the computation of the trends. The first one $(T-1975)*(T \leq 1975)$ multiplies a trend growing to a zero value in the first semester of 1975, by a condition which is true only until 1975. The result is a trend growing (taking less and less negative values) until 1975, when it takes permanently the value 0.

The second one $(T-1975)*(T > 1975)$ creates a trend which is null until 1975, then grows by one each year.

⁶⁰ For any periodicity.

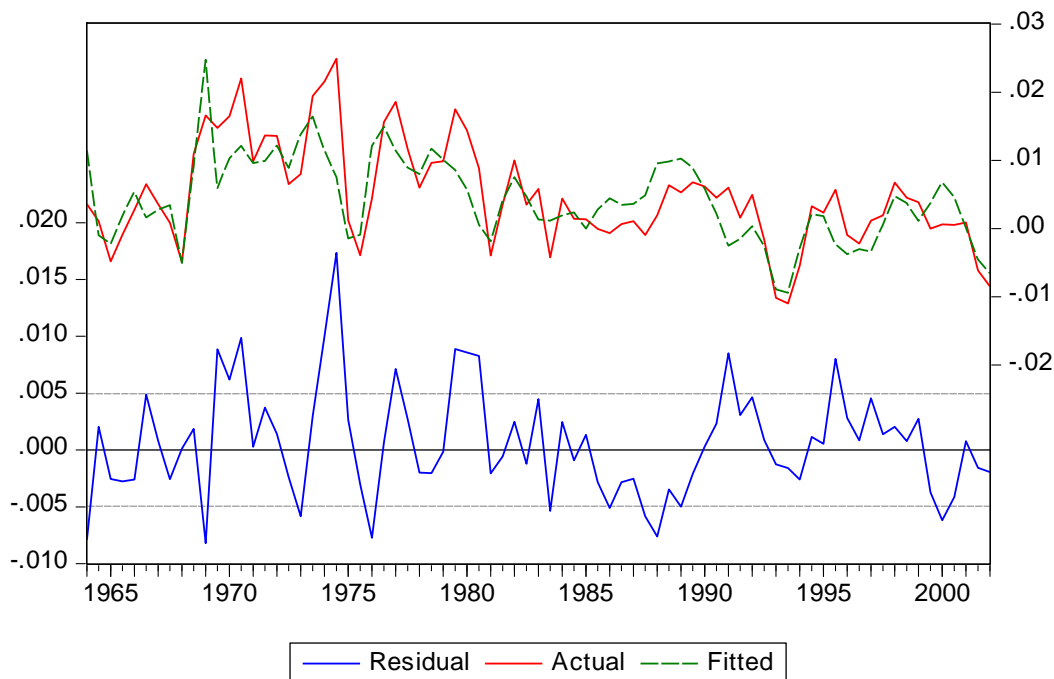


The associated formulation works quite well, with highly significant terms. The standard error is reduced significantly, the R-Squared is rather high for such an erratic dependent variable, and the quality of the explanation is quite high in the end (for the six last years the error is below average, and the last value is relatively very small). However, forecasting the trend will be quite problematic: if we maintain it in the long run, its contribution will become infinitely negative, and the explained ratio too....

And of course, it will probably justified to consider our formula as established ad hoc. In fact we will waive its use.

Dependent Variable: IC/Q(-1)
 Method: Least Squares
 Date: 11/06/12 Time: 19:37
 Sample (adjusted): 1964S1 2002S1
 Included observations: 77 after adjustments
 IC/Q(-1)=C_IC(1)*@PCH(Q)+C_IC(2)*@PCH(Q(-1))+C_IC(3)*(T-1975)
 (T<=1975)+C_IC(4)(T-1975)*(T>=1975)

	Coefficient	Std. Error	t-Statistic	Prob.
C_IC(1)	0.286030	0.048544	5.892212	0.0000
C_IC(2)	0.383594	0.047563	8.064939	0.0000
C_IC(3)	0.001608	0.000253	6.362828	0.0000
C_IC(4)	-0.000327	5.37E-05	-6.083855	0.0000
R-squared	0.582654	Mean dependent var		0.004821
Adjusted R-squared	0.565502	S.D. dependent var		0.007513
S.E. of regression	0.004952	Akaike info criterion		-7.727463
Sum squared resid	0.001790	Schwarz criterion		-7.605707
Log likelihood	301.5073	Hannan-Quinn criter.		-7.678762
Durbin-Watson stat	1.214565			



6.3.1.5 Preparing the equation for the model

Once an equation has been selected for introduction in the model, a different strategy should be used.

If we use the estimated formula, we will face several problems:

- It is not simple to link the equation name with its purpose, which makes the process unclear and forbids to use any automated and systematic process.
- The C vector is used by all equations is only consistent with the last estimated one.
- The residuals cannot be managed simply

Instead, we propose the following organization, deriving all elements from the name of the dependent variable, though a systematic transformation:

- Naming the equation after the estimated variable.

For instance we can call our equation EQ_CI.

- Using the developed specification, with explicit coefficients.
- Naming the coefficient vector after the estimated variable.

For instance we can call it C_CI. Of course this calls for its creation, with a high enough dimension:

```
coef(10) c_ci
```


(we chose 10 as a round number which we know we shall never reach).

- Introducing an additive explicit residual, named after the estimated variable. The reason is the following.
 - It is essential for a model to estimate and simulate the same equation. Of course two versions can be maintained, one being copied into the other after each new estimation. This is:
 - * Tedious.
 - * Difficult to manage.
 - * Error-prone.

It is much better to use a single item. However this faces a problem: one wants access to the residual, in particular for forecasts as we shall see later. And the estimation calls for no residual.

The solution is quite simple: introduce a formal residual, but set it to zero before any estimation.

- Work through a program

This allows:

- Visual control over the specification.
- Easy replication of the estimation (for instance if the data has changed).
- Easy introduction of marginal changes.
- Documentation of the economic context (by introducing comments in the program).

In our case we shall use:

```
coef(10) ec_ci
genr ec_ci=0
equation eq_ci ci/q(-1)=c_ci(1)*@pch(q)+c_ci(2)*@pch(q(-1))+ec_ci
genr ec_ci=resid
```

6.3.2 INVESTMENT: THE NECESSITY TO ESTABLISH A CONSISTENT THEORETICAL EQUATION PRIOR TO ESTIMATION

In this estimation, we shall stress the importance of establishing a sound economic framework before any estimation.

The basic economic idea is quite simple: the purpose of investment is

- To replace discarded capital.
- To allow a higher level production, facing an increase of demand.

Without proceeding further in theory, many formulations can be considered. For instance, investment could have a constant elasticity to GDP, maybe with an error correction term including capital...

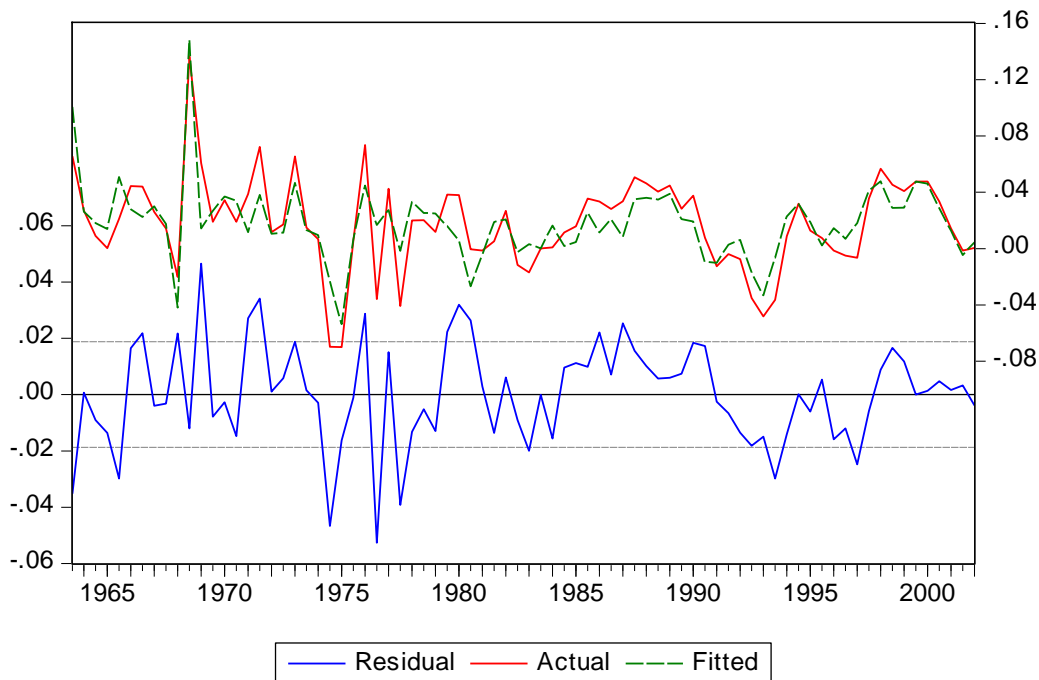
In our sense, trying for the best estimation without considering the economics behind the formula, and especially its formal consequences for model properties, is rather irresponsible. For instance, using the logarithm of investment is quite dangerous. Its value can change in very high proportions, and if we go back to the microeconomic foundation of this behavior, its value could very well be negative, as some firms are led to disinvest from time to time, by selling more capital than they buy.

For instance, the following equation seems to work quite well:

$$\Delta \text{Log}(I_t) = a \cdot \Delta \text{Log}(Q_t) + b \cdot \text{Log}(I_{t-1} / Q_{t-1}) + c \cdot t + d$$

The results are:

Dependent Variable: DLOG(I)				
Method: Least Squares				
Date: 11/07/12 Time: 12:01				
Sample (adjusted): 1963S2 2002S1				
Included observations: 78 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
DLOG(Q)	2.425971	0.195149	12.43140	0.0000
LOG(I(-1)/Q(-1))	-0.076428	0.032747	-2.333903	0.0223
C	-2.458531	0.535731	-4.589116	0.0000
T	0.001145	0.000249	4.605258	0.0000
R-squared	0.684209	Mean dependent var		0.017985
Adjusted R-squared	0.671407	S.D. dependent var		0.032691
S.E. of regression	0.018740	Akaike info criterion		-5.066421
Sum squared resid	0.025987	Schwarz criterion		-4.945564
Log likelihood	201.5904	Hannan-Quinn criter.		-5.018040
F-statistic	53.44416	Durbin-Watson stat		1.658039
Prob(F-statistic)	0.000000			

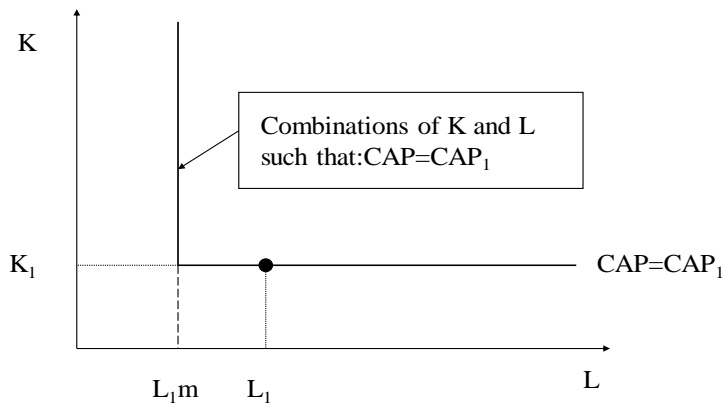


Everything seems to go well: the statistics are quite good (except maybe for the Durbin-Watson test), the signs are right, the graph shows a really strong fit. However, when we merge the equation into a model, its simulation properties will be affected by the base solution: even a very high increase in GDP will have a low impact of the absolute level of investment if it was very low in the previous period.

One can guess that although linking investment (a change in capital) to the change in production seems a natural idea, the jump to the above formulation was a little too fast a move. One should be naturally reticent in taking the logarithm of a growth rate.

We shall try to clarify the economic process through a full logical formalization.

Let us suppose that production follows a “complementary factors” function, which means that to reach a given level of productive capacity, fixed levels of capital and employment are required, and a reduction in one factor cannot be compensated by an increase in the other. This means obviously that the less costly process (optimal) is the one which respects exactly these conditions.



With “pk” productivity of capital, and “pl” productivity of labor, we get :

$$CAP_t = \min(pk_t \cdot K_{t-1}, pl_t \cdot L_t)$$

(the “t-1” means that we shall use the level of capital reached at the end of the previous period).

Actually, for a given level of employment, there is always some short term leverage on production, at least at the macroeconomic level. Temporarily increasing labor productivity by 2% can be easily achieved through extra hours, less vacations, less training courses...

This means capital will be the only limiting factor in the short term.

The capacity equation can be simplified into:

$$CAP_t = pk_t \cdot K_{t-1}$$

Now let us define the rate of use of capacities:

$$UR_t = Q_t / CAP_t$$

$$CAP_t = pk_t \cdot K_{t-1} = Q_t / UR_t$$

Now let us suppose firms actually want to reach a constant target utilization rate UR^* , and expect a production level Q_{t+1}^a . Then by definition:

$$K_t^* = CAP_t^* / pk_t^* = Q_{t+1}^a / UR_t^* / pk_{t+1}$$

$$K_t = CAP_t / pk_t = Q_t / UR_t / pk_t$$

And defining $tx(z)$ as the growth rate of z :

$$tx^*(K_t) \approx tx^*(CAP_t) - tx(pk_t) \approx tx^a(Q_t) - tx^*(UR_t) - tx(pk_t)$$

This means that the target growth rate of capital can be decomposed as the sum of three terms, one with a positive influence:

- The expected growth rate of production

and two negative:

- The target growth rate of the rate of use: if the firms feel their capacities are 1% too high for the present level of production, they can reach the target by decreasing capital by 1% even if production is not expected to change.
- The growth rate of capital productivity: if it increases by 1%, 1% less capital will be needed.

But the element we need is investment. To get it we shall use the definition.

$$K_t = K_{t-1} \cdot (1 - dr_t) + I_t$$

which can be written as

$$tx(K_t) = -dr_t + I_t / K_{t-1}$$

This gives finally:

$$I_t^* / K_{t-1} = tx(K_t) = dr_t + tx^*(K_t) = dr_t + tx^a(Q_t) - tx^*(UR_t) - tx(pk_t)$$

In other words:

- If firms expect a growth rate of 4%, capacities should adapt to that growth
- But if they feel their capacities are under-used by 1%, their desired capacity will only increase by 3%.
- If capital productivity is going to increase by 1%, they will need 1% less capital.
- But once capital growth has been defined, they also have to compensate for depreciation.

If we suppose

- That the depreciation rate is constant, as well as the rate of growth of capital productivity,
- That production growth expectations are based on an average of the previous rates,

And we consider as the rate of use the ratio of actual GDP to a value obtained under normal utilization of factors, which leads to a unitary target.

We get the simplified formula:

$$I_t^* / K_{t-1} = a + \sum_{i=0}^n \alpha_i \cdot tx^a(Q_{t-i}) - tx^*(UR_{t+1})$$

with

$$\sum_{i=0}^n \alpha_i = 1$$

Finally, we can suppose, as we shall do also for employment, that the desired growth of capital is only partially reached in practice, either because firms react cautiously to fluctuations of demand, or because they are constrained by investment programs covering more than one period.

And we shall leave free the coefficients:

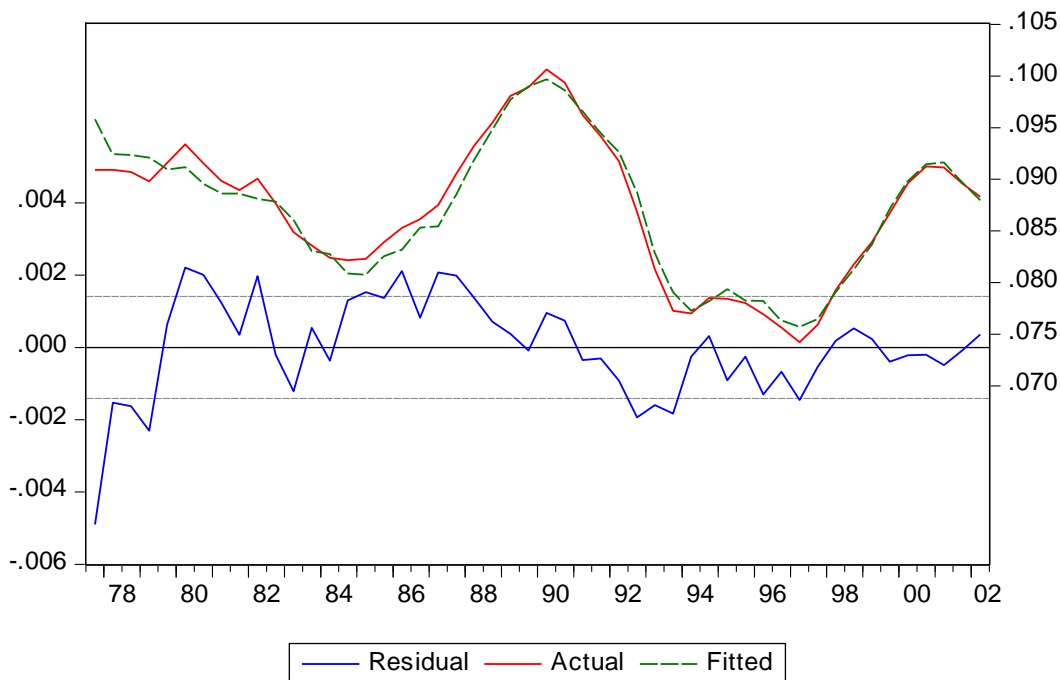
$$I_t^* / K_{t-1} = b \cdot I_{t-1} / K_{t-2} + (1-b) \cdot (a + c \cdot \sum_{i=0}^n \alpha_i \cdot tx^a(Q_{t-i}) - d \cdot tx^*(UR_t))$$

The results are rather satisfactory, with the right sign and acceptable statistics for all explanatory elements. This was not obvious, as their strong correlation (both use Q in the numerator) could have made it difficult for the estimation process to separate their role.

Dependent Variable: $IK(-1)$
Method: Least Squares
Date: 11/07/12 Time: 12:03
Sample (adjusted): 1977S2 2002S1
Included observations: 50 after adjustments
 $IK(-1)=C_I(1)*I(-1)/K(-2)+C_I(2)*UR+C_I(3)*.25*Q/Q(-4)+C_I(4)$

	Coefficient	Std. Error	t-Statistic	Prob.
C_I(1)	0.825643	0.033524	24.62809	0.0000
C_I(2)	0.027926	0.006984	3.998580	0.0002
C_I(3)	0.152870	0.058292	2.622495	0.0118
C_I(4)	-0.052515	0.010337	-5.080438	0.0000

R-squared	0.959393	Mean dependent var	0.087074
Adjusted R-squared	0.956745	S.D. dependent var	0.006785
S.E. of regression	0.001411	Akaike info criterion	-10.21218
Sum squared resid	9.16E-05	Schwarz criterion	-10.05922
Log likelihood	259.3046	Hannan-Quinn criter.	-10.15393
F-statistic	362.2686	Durbin-Watson stat	0.628797
Prob(F-statistic)	0.000000		



The graph of residuals shows that the quality of the explanation grows with time, and is especially good for the last periods. This is rather important for simulations over the future, and one can wonder what we would have done if the sample had been reversed, and the initial residuals had applied to the last periods.

We will deal with this problem of growing errors on recent periods when we address forecasts.

The equation we have built is not only satisfactory by itself, but we can expect it to provide the model with adequate properties. In particular, the long term elasticity of capital to production is now unitary by construction. Starting from a base simulation, a 1% permanent shock on Q will leave the long run value of UR unchanged⁶¹. This gives the same relative variations to production, capacity and (with a constant capital productivity) capital.

The coefficients “a” and “b” determine only the dynamics of the convergence to this target.

Actually we have estimated a kind of error-correction equation, in which the error is the gap between actual and target capacity (the rate of use).

We hope to have made clear that to produce a consistent formulation, in particular in a modelling context, one must start by establishing a sound economic background.

6.3.3 EMPLOYMENT: STATIONARITY, ERROR CORRECTION MODELS, BREAKPOINT TEST.

6.3.3.1 The economic framework

Of course, the employment equation should follow also a complementary factors framework.

In the previous paragraph, we have shown that in this framework the element determining capacity is the sole capital, while firms could ask from workers a temporary increase in productivity, high enough to ensure the needed level of production⁶². Adapting employment to the level required to obtain a “normal” productivity target will be done by steps.

This means estimating employment will allow us to apply the elements on error correction models we have presented earlier, in a very simple framework.

We shall suppose that firms:

- Know the level of production they have to achieve.
- Know also the level of production which should be achieved by each worker under normal circumstances (in other term his normal productivity).

From these two elements they can determine the normal number of workers they need.

But they do not adapt the actual employment level to this target, and this for:

⁶¹ As the left hand side represents the (fixed) long term growth rate of capital.

⁶² This is true in our macroeconomic framework, in which the changes in production are limited, and part of growth is compensated by increases in structural productivity (due for instance to more capital intensive processes). At the firm level, employment can produce bottlenecks. This will be the case if a sudden fashion appears for particular goods requiring specialized craftspeople, even if the tools and machines are available for buying.

- Technical reasons: between the conclusion that more employees are needed and the actual hiring⁶³, firms have to decide on the type of jobs called for, set up their demands, conduct interviews, negotiate wages, establish contracts, get authorizations if they are foreign citizens, maybe ask prospective workers to train... Of course this delay depends heavily on the type of job. And this goes also for laying out workers.
- Behavioral reasons: if facing a hike in production, firms adapt immediately their employment level to a higher target, they might be faced later with over employment if the hike is only temporary. The workers they have trained, maybe at a high cost, have no usefulness at the time they become potentially efficient. And laying them out will call generally for compensations....

6.3.3.2 The formulas: stationarity and error correction

We should realize that we are facing an error correction framework, which we can materialize as:

“Normal” labor productivity does not depend on economic conditions. It might follow a constant trend over the period, such as:

$$\text{Log}(pl_t) = a + b \cdot t$$

Firms use this target to define “normal” employment:

$$LE_t^* = Q_t / pl_t^*$$

They adapt actual employment to this target with some inertia:

$$\Delta \text{Log}(L_t) = \alpha \cdot \Delta \text{Log}(L_t^*) + \beta \cdot \text{Log}(L_{t-1}^* / L_{t-1}) + \gamma + \varepsilon_t$$

We recognize here the error correction framework presented earlier, which requires:

$\text{Log}(L_t^* / L_t)$ to be stationary.

But α does not have to be unitary. However, if we follow the above reasoning, its value should be between 0 and 1, and probably significantly far from each of these bounds.

⁶³ But not the start of actual work: what we measure is the number of workers employed, even if they are still training for instance.

To estimate this system we face an obvious problem: pl^* is not an actual series (LE^* either, but if we know one we know the other).

But if we call “ pl ” the actual level of productivity (Q/LE) we can observe that:

$$\text{Log}(L_t^* / L_t) = \text{Log}((Q_t / pl_t^*) / (Q_t / pl_t)) = -\text{Log}(pl_t^* / pl_t)$$

The stationarity of $\text{Log}(L_t^* / L_t)$ is equivalent to that of $\text{Log}(pl_t^* / pl_t)$

Now it should be obvious that if pl^* and pl have a trend, it must be the same, actually the trend defining completely pl^* . If not, they will diverge over the long run, and we will face infinite under or over employment. So target productivity can be identified using the trend in the actual value, if it exists.

This means we can test the stationarity of the ratio as the stationarity of actual productivity around a trend, a test provided directly by EViews.

We can expect a framework in which actual productivity fluctuates around a regularly growing target, with cycles which we do not expect to be too long, but can last for several periods⁶⁴.

6.3.3.3 The first estimations

First, we compute actual labor productivity

$$\text{genr PROD} = Q / LE$$

and regress it on time:

$$\text{ls log(PROD) c t}$$

to get the structural productivity trend.

⁶⁴ Which will create (acceptable) autocorrelation in the difference to the trend.

Dependent Variable: LOG(Q/LE)

Method: Least Squares

Date: 11/07/12 Time: 12:07

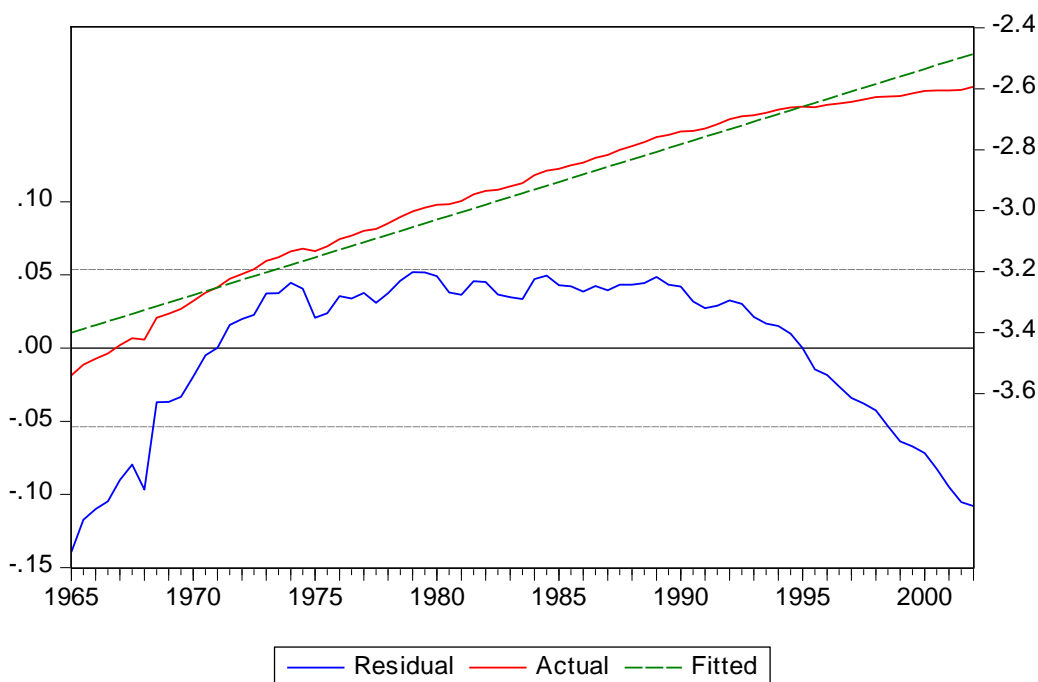
Sample: 1993S1 2002S1

Included observations: 19

LOG(Q/LE)=C(1)+C(2)*T

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-22.34121	0.603556	-37.01598	0.0000
C(2)	0.009865	0.000302	32.64715	0.0000
R-squared	0.984300	Mean dependent var	-2.636851	
Adjusted R-squared	0.983377	S.D. dependent var	0.027976	
S.E. of regression	0.003607	Akaike info criterion	-8.312619	
Sum squared resid	0.000221	Schwarz criterion	-8.213204	
Log likelihood	80.96988	Hannan-Quinn criter.	-8.295794	
F-statistic	1065.836	Durbin-Watson stat	1.123297	
Prob(F-statistic)	0.000000			

Results are quite bad. Of course productivity shows a significant growth, but the standard error is quite high (more than 5 %). More important, the graph of residuals and the auto-correlation test show that we are not meeting the condition we have set: that observed productivity fluctuates around a trend, with potential but not unreasonably long cycles.



The problem apparently lies in the fact that the average growth rate is consistently higher in the first part of the period, and lower later. Seen individually, each sub-period might seem to meet the above condition.

From the graph, we clearly need two breaks. One will observe that the first period follows the first oil shock, and the beginning of a lasting world economic slowdown. The reason for the second break is less clear (some countries like the US and Scandinavia show a break in the opposite direction).

For choosing the most appropriate dates, we can use two methods:

- A visual one: 1973 and 1990 could be chosen, possibly plus or minus 1 year.
- A statistical one: the most appropriate test is the Chow breakpoint test, which we have explained earlier. To make our choice automatic, we shall consider two intervals, and apply the test to all reasonably possible combinations of dates from those intervals. As we could expect, all the tests conclude to a break. But we shall elect the couple associated to the lowest probability (of no break), which means the highest likelihood ratio⁶⁵. Of course, this criterion works only because the sample and the number of breaks remain the same.

The best result corresponds actually to 1973S1 and 1992S1, as shown in this table of log-likelihood ratios.

Log likelihood ratio

dates	1991S2	1992S1	1992S2
1972S2	895	913	908
1973S1	904	928	925
1973S2	895	917	915

The equation for structural productivity is

$$\text{LOG}(\text{PRLE}) = \text{C_PRLE}(1) + \text{C_PRLE}(2) * (\text{T} - 2002) + \text{C_PRLE}(3) * (\text{T} - 1973) * (\text{T} < 1973) + \text{C_PRLE}(4) * (\text{T} - 1992) * (\text{T} < 1992)$$

One will note:

- That we have introduced no residual, contrary to our usual practice.
- That we have introduced reversed trends, which stop after a while instead of starting inside the period.
- Target productivity is not a behavior.

The first element is quite logical: what we are estimating for the model is not actual productivity (this is given in the model by an identity, using actual GDP and employment). We are looking for the **exact** value of target productivity, prone to error only because we have not enough information to produce the true value. If the sample grew, or the periodicity increased, the precision would improve constantly, even if the residual does not decrease. Whereas, in a normal behavioral equation, the residual corresponds to an error on the variable, and cannot be decreased indefinitely, as the identification of the role of explanatory elements becomes less reliable with their number.

⁶⁵ the highest F gives the same conclusion

- Partial trends should apply to past periods.

The reason here is purely technical. Our model will be mainly used on the future, if its period of operational use is longer than its period of production (at least, we hope so). So it is essential to make the forecasting process as easy as possible.

If the partial trends are still active in the future, we shall have to manage them simultaneously. We can expect that we want to control the global trend of labor productivity, if only to make it consistent with our long term evolutions of GDP (which should follow world growth) and employment (which should follow population trends). Obviously, controlling a single global trend is easier than a combination of three trends.

Also, the last trend is the most important for interpretation of model properties, and it is better to make it the easiest to observe.

On the other hand, our technique has no bad points, once it has been understood.

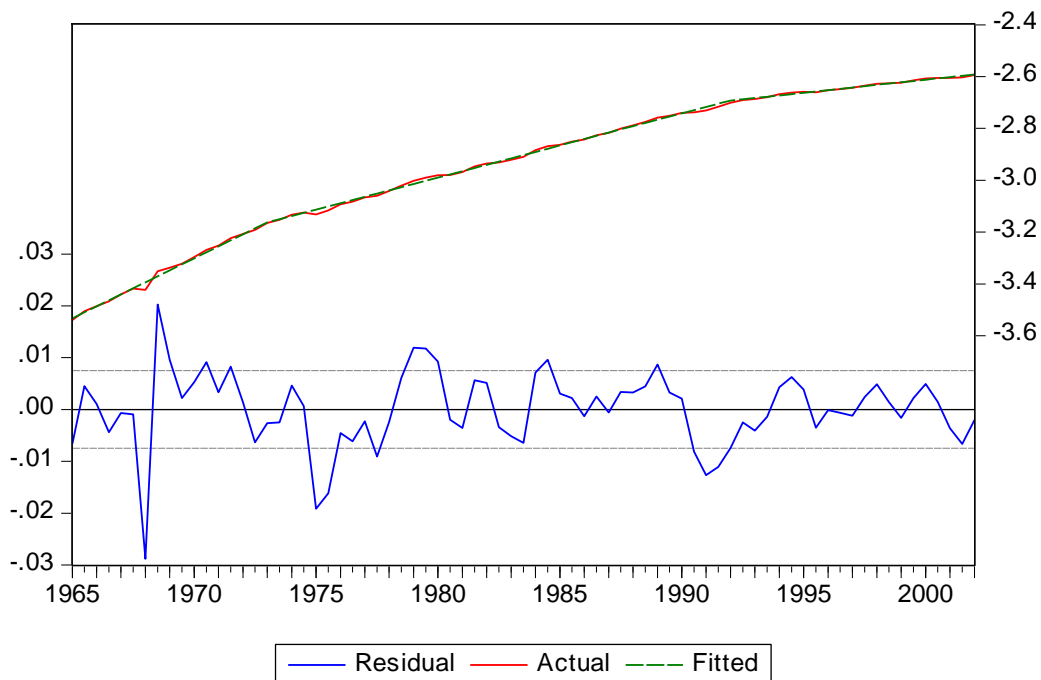
Finally, the reason for starting the trend in 2002 is also associated with handling of its future values. If the global coefficient is changed, this will be the period for a new break, and this is the best period to introduce it.

The results look acceptable, as to the validation of coefficients and the graphs (we are presenting the program version, as the equation will be introduced in the model)⁶⁶.

Dependent Variable: LOG(PRLE)
Method: Least Squares
Date: 11/07/12 Time: 11:59
Sample (adjusted): 1965S1 2002S1
Included observations: 75 after adjustments
LOG(PRLE)=C_PRLE(1)+C_PRLE(2)*(T-2002)+C_PRLE(3)*(T-T1)*(T<T1)
+C_PRLE(4)*(T-T2)*(T<T2)

	Coefficient	Std. Error	t-Statistic	Prob.
C_PRLE(1)	-2.591659	0.002928	-885.1806	0.0000
C_PRLE(2)	0.010124	0.000394	25.66585	0.0000
C_PRLE(3)	0.021567	0.000636	33.91560	0.0000
C_PRLE(4)	0.014596	0.000511	28.58584	0.0000
R-squared	0.999288	Mean dependent var		-2.943006
Adjusted R-squared	0.999258	S.D. dependent var		0.274576
S.E. of regression	0.007478	Akaike info criterion		-6.901879
Sum squared resid	0.003970	Schwarz criterion		-6.778280
Log likelihood	262.8205	Hannan-Quinn criter.		-6.852527
F-statistic	33233.07	Durbin-Watson stat		1.410774
Prob(F-statistic)	0.000000			

⁶⁶ This is not absolutely needed, as a variable depending only on time can be considered exogenous and computed outside the model. But we want to be able to change the assumption in forecasts, and this is the easiest way.



Now we must test the stationarity of the residual. We shall use the Dickey Fuller test (or Phillips – Perron).

First we need to generate from the current RESID a variable containing the residual (the test is going to compute its own RESID, so it is not possible to test on it).

- In program form, the test is conducted by:

```

genr res_prle=resid
uroot(1,p) res_prle
uroot(h,p) res_prle

```

- Using menus, one has to
 - Display the variable
 - Select **View>Unit root test**
 - Choose the method.

Null Hypothesis:
 RES_PRLE has a unit
 root
 Exogenous:
 Constant
 Lag Length: 1
 (Automatic - based on
 SIC, maxlag=11)

	t-	Pr
	Stat ob.	istic *
141	-	
Augmented	5.19	0.

Null Hypothesis: RES_PRLE has a unit root
Exogenous: Constant
Lag Length: 1 (Automatic - based on SIC, maxlag=11)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-5.191811	0.0000
Test critical values:		
1% level	-3.522887	
5% level	-2.901779	
10% level	-2.588280	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(RES_PRLE)
Method: Least Squares
Date: 11/07/12 Time: 12:10
Sample (adjusted): 1966S1 2002S1
Included observations: 73 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RES_PRLE(-1)	-0.735169	0.141602	-5.191811	0.0000
D(RES_PRLE(-1))	0.048991	0.118711	0.412687	0.6811
C	-2.46E-06	0.000833	-0.002958	0.9976

R-squared	0.352694	Mean dependent var	-8.95E-05
Adjusted R-squared	0.334200	S.D. dependent var	0.008722
S.E. of regression	0.007117	Akaike info criterion	-7.012437
Sum squared resid	0.003546	Schwarz criterion	-6.918309
Log likelihood	258.9540	Hannan-Quinn criter.	-6.974926
F-statistic	19.07026	Durbin-Watson stat	1.993440
Prob(F-statistic)	0.000000		

Both tests conclude very strongly the stationarity of the residual.

The values of target productivity and desired employment are given by:

```

genr log(prle_t)= c_prle(1)+c_prle(2)*t+c_prle(3)*(t-1973)*(t<1973)+c_prle(4)*(t-1992)*(t<1992)
genr led=q/prle_t

```

LE will be estimated (using here the developed form) by:

```

equation eq_le.ls dlog(LE)=c_le(1)*dlog(LED)+c_le(2)*log(LED(-1)/LE(-1))+c_le(3)

```

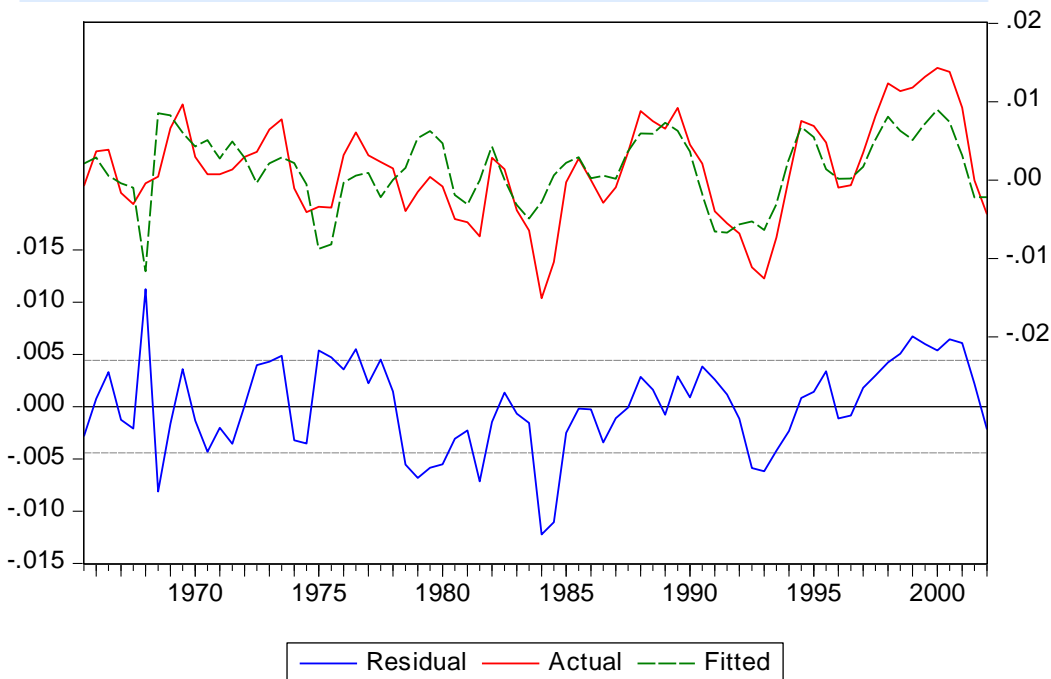
where LED / LE is equal to (Q/LE)/prle, the residual from the previous equation.

The results are rather significant, except for the **last** coefficient.

Dependent Variable: DLOG(LE)
 Method: Least Squares
 Date: 11/07/12 Time: 12:55
 Sample (adjusted): 1965S2 2002S1
 Included observations: 74 after adjustments
 DLOG(LE)=C_LE(1)*DLOG(LED)+C_LE(2)*LOG(LED(-1)/LE(-1))+C_LE(3)

	Coefficient	Std. Error	t-Statistic	Prob.
C_LE(1)	0.422879	0.054347	7.781036	0.0000
C_LE(2)	0.455672	0.074240	6.137818	0.0000
C_LE(3)	0.000732	0.000520	1.407567	0.1636

R-squared	0.514224	Mean dependent var	0.001335
Adjusted R-squared	0.500540	S.D. dependent var	0.006259
S.E. of regression	0.004423	Akaike info criterion	-7.964257
Sum squared resid	0.001389	Schwarz criterion	-7.870849
Log likelihood	297.6775	Hannan-Quinn criter.	-7.926996
F-statistic	37.57891	Durbin-Watson stat	1.015944
Prob(F-statistic)	0.000000		



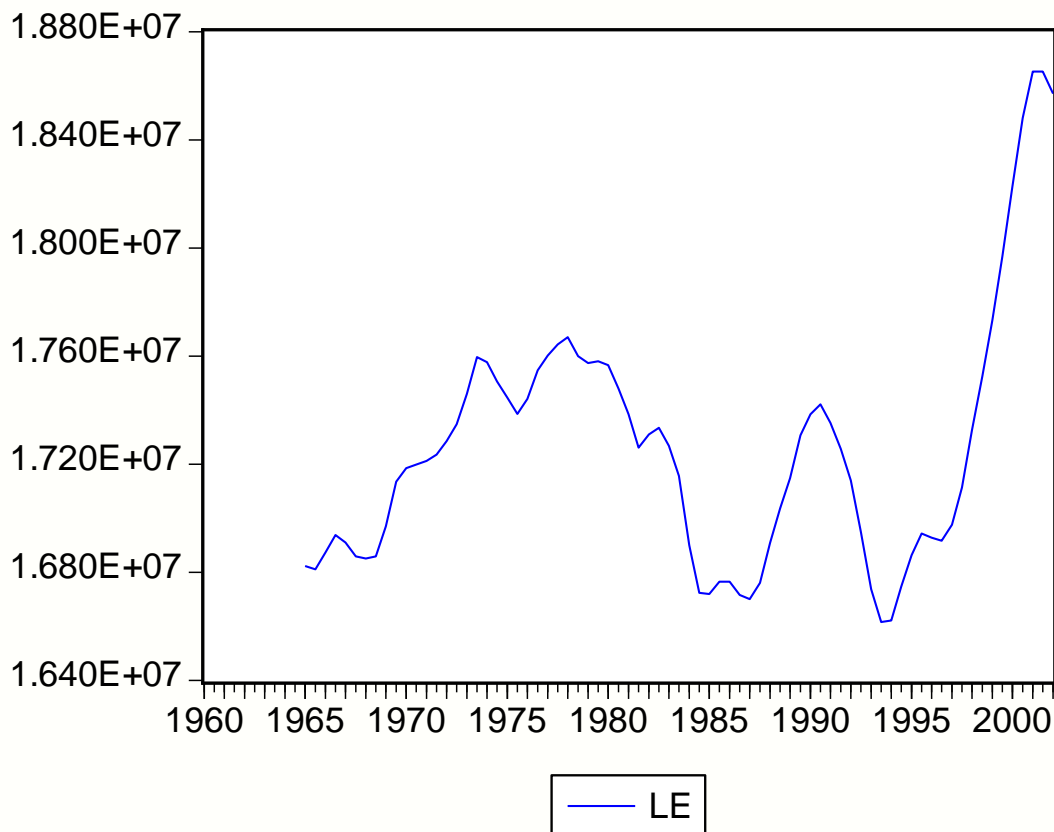
Following the reasoning made earlier, $c_le(3)$ (or rather $c_le(3)/c_le(2)$) should represent the logarithm of the long term gap between the target employment and the level reached. This gap will be significant if both:

- Employment shows a trend (the target is moving), which means that GDP and target productivity show different trends.

- A difference between the growths of GDP and target productivity is not compensated immediately (the value of $c_{le(1)}$ is different from one)

The second condition is clearly met, but not the first.

In the present case, employment shows no obvious trend (positive perhaps?), and **the last coefficient** is not really needed. But there is no reason to drop it (just as there is no reason to drop a non-significant constant term from a linear equation).



As to **the first coefficients**, they look significantly different. The first is more significant, and also higher: this could mean that it is easier (and more desirable) to close the first part of the gap between desired and actual employment.

This will not be true in the US case.

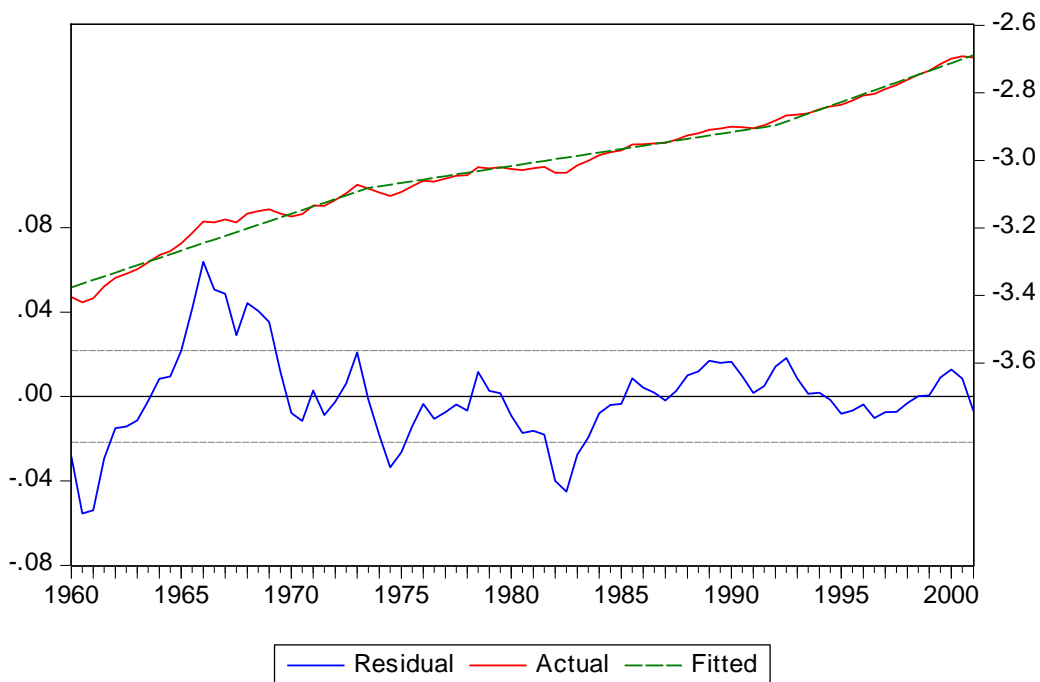
First, the "1992" break exists too, but it is now positive, as shown here:

Dependent Variable: LOG(PRLE)
 Method: Least Squares
 Date: 11/07/12 Time: 12:48
 Sample (adjusted): 1960S1 2001S1
 Included observations: 83 after adjustments

$$\text{LOG}(\text{PRLE}) = \text{C_PRLE}(1) + \text{C_PRLE}(2) * T + \text{C_PRLE}(3) * (T - 1973.5) * (T < 1973.5) + \text{C_PRLE}(4) * (T - 1992) * (T < 1992)$$

	Coefficient	Std. Error	t-Statistic	Prob.
C_PRLE(1)	-48.83508	2.620014	-18.63924	0.0000
C_PRLE(2)	0.023062	0.001313	17.55985	0.0000
C_PRLE(3)	0.011769	0.001133	10.38824	0.0000
C_PRLE(4)	-0.013038	0.001639	-7.954501	0.0000

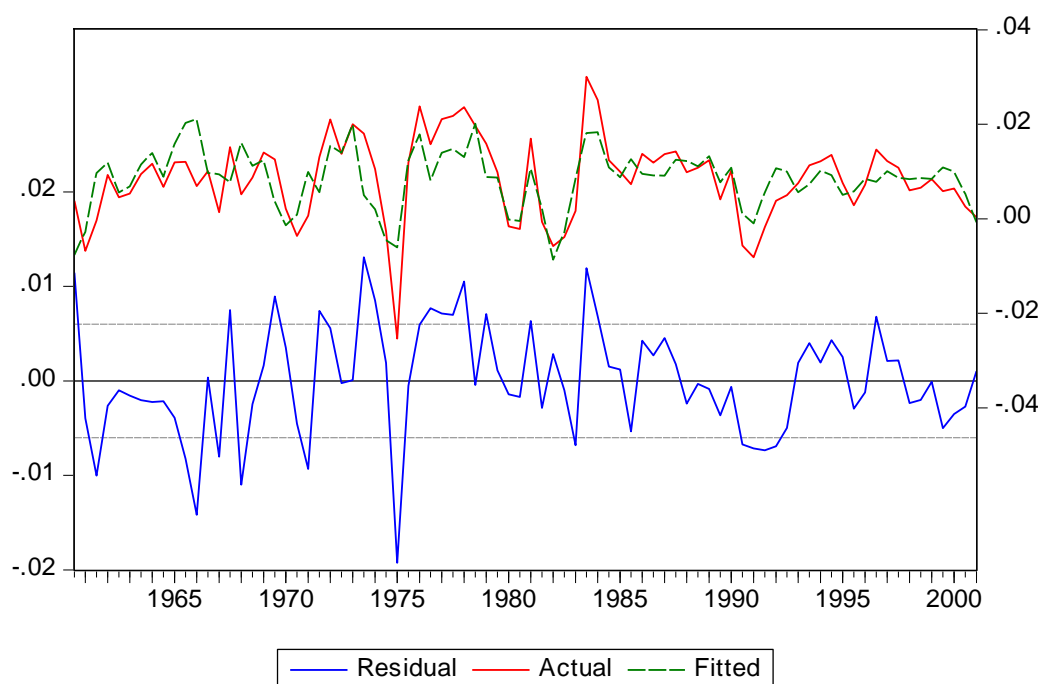
R-squared	0.985989	Mean dependent var	-3.025035
Adjusted R-squared	0.985457	S.D. dependent var	0.180258
S.E. of regression	0.021738	Akaike info criterion	-4.772521
Sum squared resid	0.037331	Schwarz criterion	-4.655951
Log likelihood	202.0596	Hannan-Quinn criter.	-4.725690
F-statistic	1853.184	Durbin-Watson stat	0.252303
Prob(F-statistic)	0.000000		



Second, employment has grown substantially over the sample period, which means that a constant term is called for:

Dependent Variable: DLOG(LE)
 Method: Least Squares
 Date: 11/07/12 Time: 12:50
 Sample (adjusted): 1960S2 2001S1
 Included observations: 82 after adjustments
 $DLOG(LE)=C_LE(1)*DLOG(LED)+C_LE(2)*LOG(LED(-1)/LE(-1))+C_LE(3)$

	Coefficient	Std. Error	t-Statistic	Prob.
C_LE(1)	0.420082	0.045549	9.222675	0.0000
C_LE(2)	0.097051	0.031296	3.101115	0.0027
C_LE(3)	0.004784	0.000774	6.182176	0.0000
R-squared	0.532205	Mean dependent var		0.008455
Adjusted R-squared	0.520362	S.D. dependent var		0.008682
S.E. of regression	0.006013	Akaike info criterion		-7.353944
Sum squared resid	0.002856	Schwarz criterion		-7.265893
Log likelihood	304.5117	Hannan-Quinn criter.		-7.318593
F-statistic	44.93870	Durbin-Watson stat		1.488317
Prob(F-statistic)	0.000000			



6.3.3.4 A virtuous dummy element

Reverting to the French case, one will observe a strange outlier: the year 1968 presents a strong negative residual in the first semester, and a negative one for the last. French people (and people familiar with French post-war history) will certainly recall the May 1968 “student revolution” which lasted roughly from March to June. During that period, the French economic process was heavily disturbed, in particular the transportation system, and GDP decreased (by 0.5% for the semester). If the equation had worked, employment would have decreased too, especially as productivity growth was quite high. On the contrary, it remained almost stable.

The explanation is obvious: firms expected the slump to be purely temporary, and activity to start back after a while (actually they were right, and GDP grew by 7.5% in the next semester, due in part to the higher consumption allowed by “Grenelle⁶⁷” wage negotiations, very favorable to workers). They did not want to lay out (at a high cost) workers whom they would need back later with no guarantee to find the same individuals, familiar with the firms’ techniques. So the employment level was very little affected.

This means that the global behavior does not apply here, and the period has to be either eliminated from the sample, or treated through a specific element, for instance a variable with the value 1 in the first semester and -1 in the second (when employment increased less than the growth in GDP would call for).

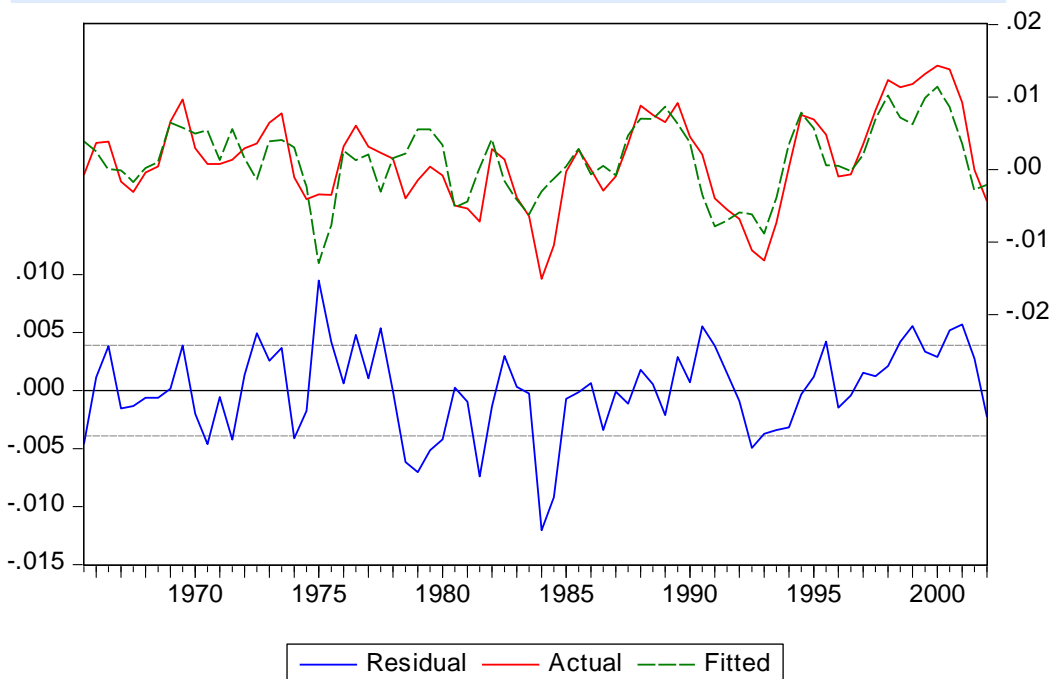
This case is rather interesting: some economists could be tempted to introduce dummies just because the equation does not work, and indeed the results will be improved (including in general the statistics for the explanatory variables). This can probably be called cheating. On the contrary, **not** introducing the present dummy can be considered incorrect: we know that the local behavior did not follow the formulation we have selected, so it has to be modified accordingly.

The global results are slightly improved, and the first coefficient increases significantly (this was to be expected). The introduction of the element was not a negligible issue.

⁶⁷ From the location of the Ministry of Employment where negotiations were conducted.

Dependent Variable: DLOG(LE)
 Method: Least Squares
 Date: 11/07/12 Time: 12:53
 Sample (adjusted): 1965S2 2002S1
 Included observations: 74 after adjustments
 $DLOG(LE)=C_LE(1)*DLOG(LED)+C_LE(2)*LOG(LED(-1)/LE(-1))+C_LE(3)+C_LE(4)*(T=1968.5)-(T=1968))$

	Coefficient	Std. Error	t-Statistic	Prob.
C_LE(1)	0.587984	0.059833	9.827155	0.0000
C_LE(2)	0.411309	0.066185	6.214541	0.0000
C_LE(3)	0.000502	0.000461	1.089292	0.2798
C_LE(4)	-0.016716	0.003625	-4.611097	0.0000
R-squared	0.627400	Mean dependent var		0.001335
Adjusted R-squared	0.611431	S.D. dependent var		0.006259
S.E. of regression	0.003901	Akaike info criterion		-8.202472
Sum squared resid	0.001065	Schwarz criterion		-8.077928
Log likelihood	307.4915	Hannan-Quinn criter.		-8.152790
F-statistic	39.28961	Durbin-Watson stat		1.076553
Prob(F-statistic)	0.000000			



We shall use:

equation eq_prle.ls(p) $\log(prle)=c_prle(1)+c_prle(2)*t+c_prle(3)*(t-1973)*(t<1973)+c_prle(4)*(t-1992)*(t<1992)$

```
equation eq_le.ls(p) dlog(le)=c_le(1)*dlog(led)+c_le(2)*log(led(-1)/le(-1))+c_le(3)+c_le(4)*((t=1968.5)-
(t=1968))+ec_le
```

```
genr ec_le=resid
```

and as we have to introduce two new variables

```
g_vendo.drop led prle_t
g_vendo.add led prle_t
```

Note: the reason for the initial “drop” statement is to avoid the duplication of the elements inside the group, in case the procedure is repeated. If the elements are not present, nothing happens.

6.3.4 EXPORTS: AUTOREGRESSIVE PROCESS, COINTEGRATION, LONG TERM STABILITY.

Estimating exports will be simpler from the theoretical side. We shall use it as an example for introducing first autoregressive processes, then cointegration.

Let us first start with the simplest idea: exports show a constant elasticity to world demand. In other words:

$$\Delta X / X / (\Delta WD / WD) = a$$

or by integration:

$$\text{Log}(X) = a \cdot \text{Log}(WD) + b$$

Estimation should give to a value close to unity, as world demand is measured as the normal demand addressed to France by its clients, and takes into account:

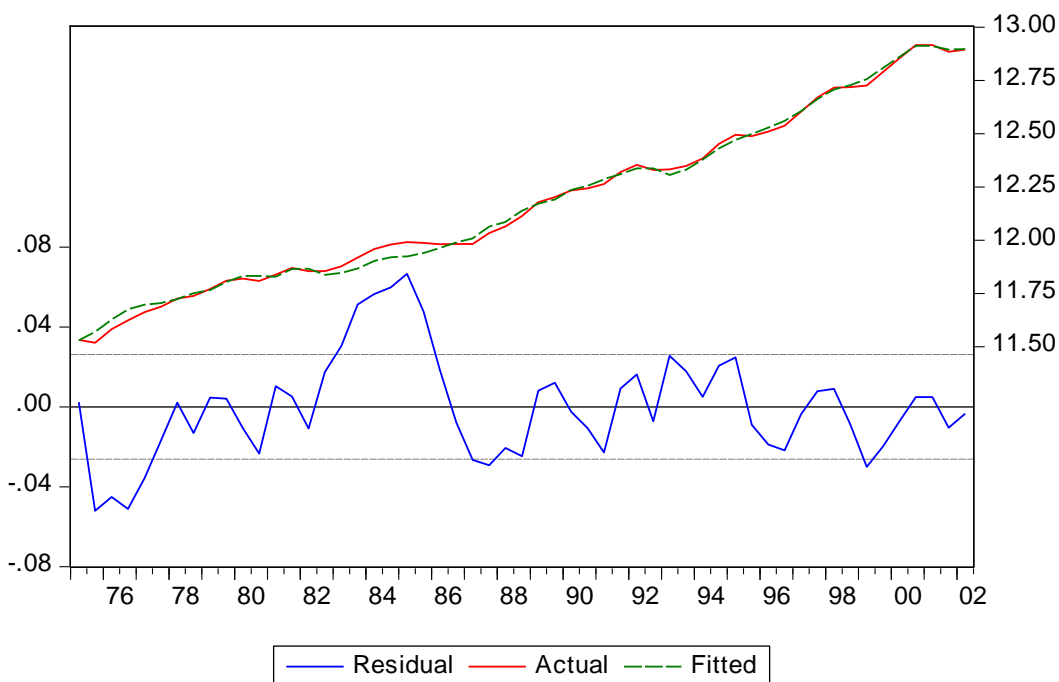
- The expansion of international trade.
- The types of goods France exports,
- The structure of the countries to which France naturally exports.

For instance, both luxury goods and Germany have a higher share in this indicator, compared to the global world market. As a weighting of normal imports, it takes also into account the growing importance of international trade.

Indeed the coefficient we obtain is close to unity.

Dependent Variable: LOG(X)
 Method: Least Squares
 Date: 11/07/12 Time: 11:59
 Sample (adjusted): 1975S1 2002S1
 Included observations: 55 after adjustments
 LOG(X)=C_X(1)*LOG(WD)+C_X(2)

	Coefficient	Std. Error	t-Statistic	Prob.
C_X(1)	1.002987	0.008866	113.1316	0.0000
C_X(2)	-0.105243	0.108654	-0.968605	0.3371
R-squared	0.995876	Mean dependent var	12.18047	
Adjusted R-squared	0.995798	S.D. dependent var	0.403814	
S.E. of regression	0.026176	Akaike info criterion	-4.412289	
Sum squared resid	0.036314	Schwarz criterion	-4.339295	
Log likelihood	123.3380	Hannan-Quinn criter.	-4.384062	
F-statistic	12798.76	Durbin-Watson stat	0.497272	
Prob(F-statistic)	0.000000			



However the low value of the Durbin-Watson test indicates a strongly positive autocorrelation of residuals, and invalidates the formulation.

Let us try to eliminate auto-correlation, supposing that the residual is actually:

$$e_t = \rho \cdot e_{t-1} + u_t$$

where ρ should be significant (positive here), and $u(t)$ has to be independent across time.

The simplest idea is to transform the equation

$$\text{Log}(X_t) = a \cdot \text{Log}(WD_t) + b + e_t$$

which is also true in the previous period.

$$\text{Log}(X_{t-1}) = a \cdot \text{Log}(WD_{t-1}) + b + e_{t-1}$$

We can multiply the second equation by ρ , and subtract it from the first:

$$\text{Log}(X_t) - \rho \cdot \text{Log}(X_{t-1}) = a \cdot (\text{Log}(WD_t) - \rho \cdot \text{Log}(WD_{t-1})) + b \cdot (1 - \rho) + e_t - \rho \cdot e_{t-1}$$

The residual for the new equation is the uncorrelated u :

$$\text{Log}(X_t) - \rho \cdot \text{Log}(X_{t-1}) = a \cdot (\text{Log}(WD_t) - \rho \cdot \text{Log}(WD_{t-1})) + b \cdot (1 - \rho) + u_t$$

6.3.4.1 Introducing an autoregressive process

To estimate the above formula, it is not necessary to establish the full equation (which calls for a full non-OLS specification, as it is not linear in the coefficients).

One can very well use the same presentation as for ordinary least squares, introducing in the estimation window the additional term AR(n), n representing the autocorrelation lag, in our case 1:

$$\text{Is } \text{Log}(X) \text{ log}(WD) \text{ c ar}(1)$$

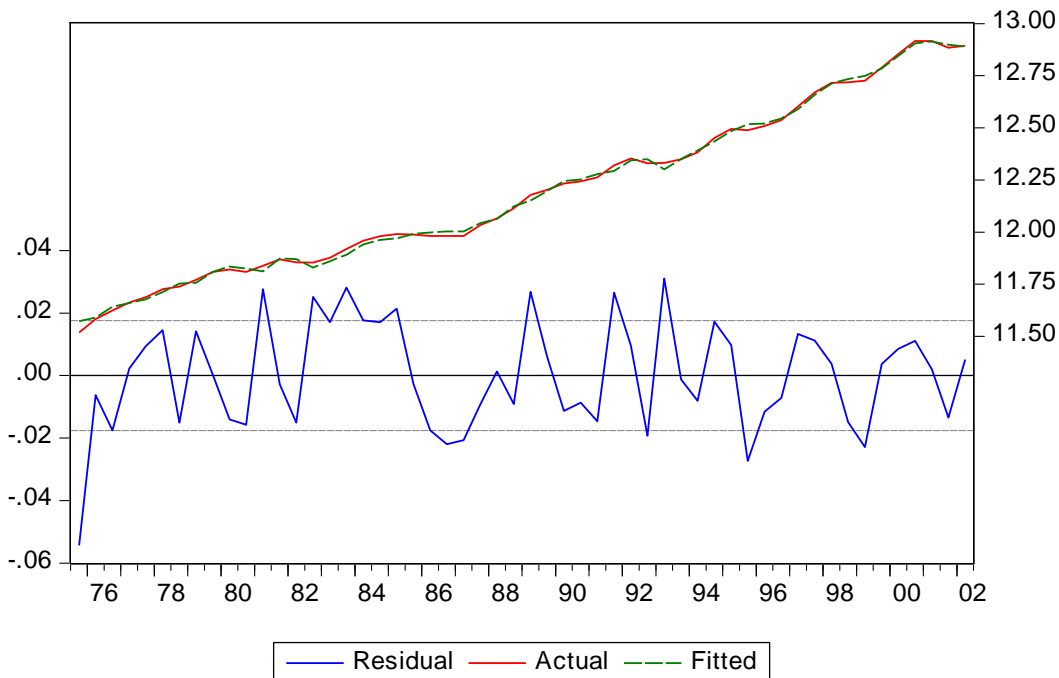
But the application to the developed formula is also quite simple:

$$\text{equation eq_lx. Is } \text{Log}(X) = c_x(1) \cdot \text{log}(WD) + c_x(2) + [\text{ar}(1) = c_x(3)]$$

The results are rather satisfactory: the first coefficient retains the theoretical value, the new coefficient is significant, the global precision is much improved (see also the graph) and the DW test is closer to satisfactory.

Dependent Variable: LOG(X)
 Method: Least Squares
 Date: 11/07/12 Time: 11:59
 Sample (adjusted): 1975S2 2002S1
 Included observations: 54 after adjustments
 Convergence achieved after 9 iterations
 LOG(X)=C_X(1)*LOG(WD)+C_X(2)+[AR(1)=C_X(3)]

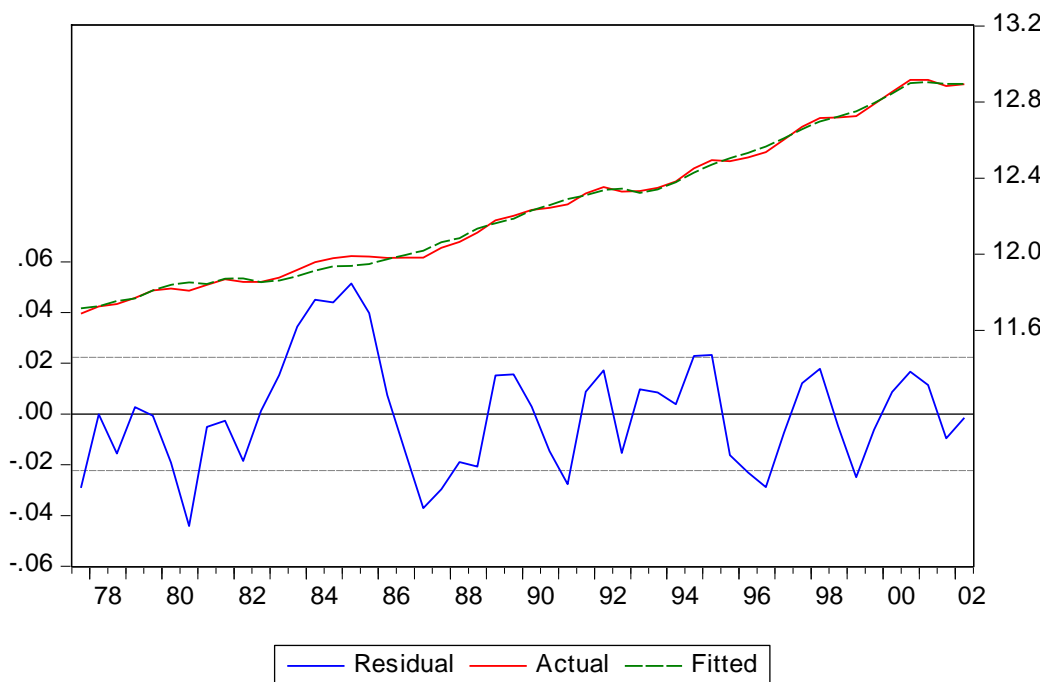
	Coefficient	Std. Error	t-Statistic	Prob.
C_X(1)	0.999049	0.023932	41.74592	0.0000
C_X(2)	-0.057019	0.295426	-0.193006	0.8477
C_X(3)	0.752588	0.092241	8.158958	0.0000
R-squared	0.998113	Mean dependent var		12.19248
Adjusted R-squared	0.998039	S.D. dependent var		0.397575
S.E. of regression	0.017604	Akaike info criterion		-5.187439
Sum squared resid	0.015805	Schwarz criterion		-5.076940
Log likelihood	143.0609	Hannan-Quinn criter.		-5.144824
F-statistic	13491.02	Durbin-Watson stat		1.530307
Prob(F-statistic)	0.000000			
Inverted AR Roots	.75			



However our formulation is a little too simplistic. We want exports to decrease with the rate of use of capacities, representing the fact that if firms are already using selling most of their potential production, they will be to be less dynamic in their search for foreign markets (more on this later).

Let us introduce the rate of use UR in the initial formula (one never knows, it might eliminate autocorrelation). We get:

Dependent Variable: LOG(X)				
Method: Least Squares				
Date: 11/07/12 Time: 12:03				
Sample (adjusted): 1977S2 2002S1				
Included observations: 50 after adjustments				
LOG(X)=C_X(1)*LOG(WD)+C_X(2)+C_X(3)*UR				
	Coefficient	Std. Error	t-Statistic	Prob.
C_X(1)	0.989908	0.008545	115.8478	0.0000
C_X(2)	0.207375	0.120579	1.719825	0.0920
C_X(3)	-0.149706	0.076793	-1.949464	0.0572
R-squared	0.996566	Mean dependent var	12.24014	
Adjusted R-squared	0.996420	S.D. dependent var	0.373424	
S.E. of regression	0.022342	Akaike info criterion	-4.706530	
Sum squared resid	0.023462	Schwarz criterion	-4.591809	
Log likelihood	120.6632	Hannan-Quinn criter.	-4.662843	
F-statistic	6820.462	Durbin-Watson stat	0.752199	
Prob(F-statistic)	0.000000			



Unsurprisingly, the Durbin-Watson test value invalidates the results again (one could always expect a miracle).

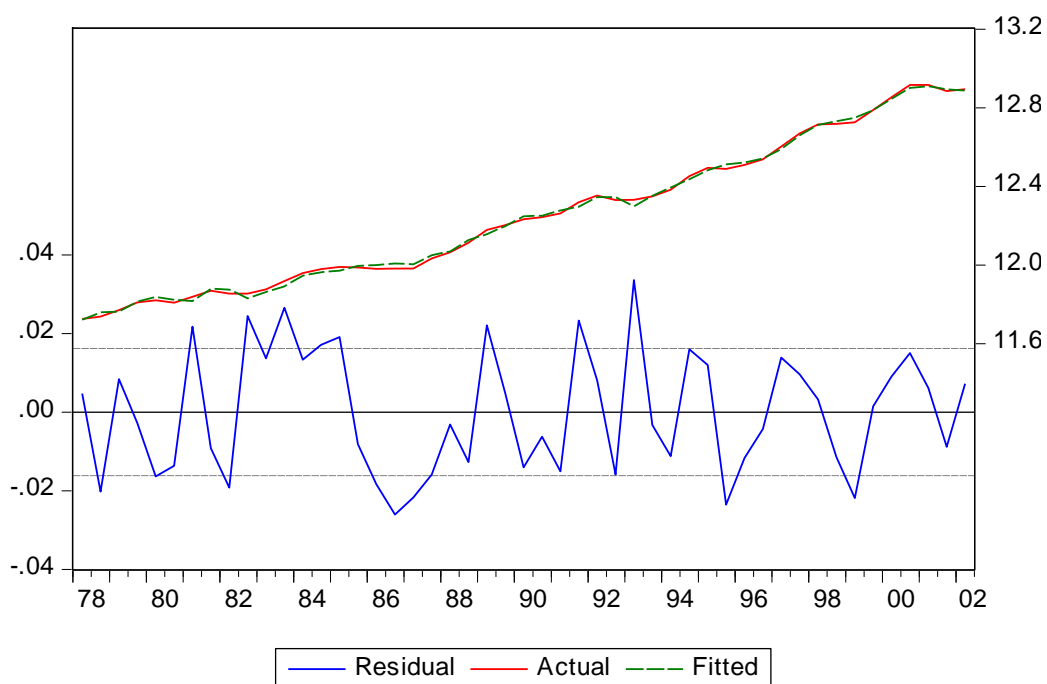
Let us use the same method as before:

Dependent Variable: LOG(X)
Method: Least Squares
Date: 11/07/12 Time: 12:03
Sample (adjusted): 1978S1 2002S1
Included observations: 49 after adjustments
Convergence achieved after 10 iterations
LOG(X)=C_X(1)*LOG(WD)+C_X(2)+C_X(3)*UR+[AR(1)=C_X(4)]

	Coefficient	Std. Error	t-Statistic	Prob.
C_X(1)	0.975444	0.023544	41.43150	0.0000
C_X(2)	0.161832	0.291397	0.555367	0.5814
C_X(3)	0.079844	0.073949	1.079726	0.2860
C_X(4)	0.742440	0.098471	7.539650	0.0000

R-squared	0.998194	Mean dependent var	12.25142
Adjusted R-squared	0.998073	S.D. dependent var	0.368588
S.E. of regression	0.016179	Akaike info criterion	-5.332101
Sum squared resid	0.011779	Schwarz criterion	-5.177666
Log likelihood	134.6365	Hannan-Quinn criter.	-5.273508
F-statistic	8289.252	Durbin-Watson stat	1.692421
Prob(F-statistic)	0.000000		

Inverted AR Roots	.74
-------------------	-----



We have eliminated autocorrelation, but now the coefficient for UR is no longer significant (fortunately as it has the wrong sign).

Let us not despair. If this old fashioned tool did not work, let us try a more modern one: cointegration.

6.3.4.2 Applying cointegration under EViews

As we have stated earlier, to establish cointegration between two elements, one has to prove that in the long run these elements move together, maintaining a bounded “distance” (or rather that a linear combination is bounded), while the value of each of the two elements is unbounded (a necessary condition).

For a group of more than two elements to be cointegrated, no subset of this group must have this property (stationarity of a single element or cointegration of a subset).

If we want to go beyond intuition, the reason for the last condition is that if a cointegrating relation is evidenced between elements, some of which are already cointegrated, one can always recompose the encompassing equation into the true cointegrating equation (considered as a new stationary variable) and other variables.

For instance,

if

$$a \cdot x + b \cdot y + c \cdot z$$

is tested as a cointegrating equation, but:

$$a \cdot x + b' \cdot y$$

is too (we can use the same a as a cointegrating equation is known to a given factor), then

$$a \cdot x + b \cdot y + c \cdot z$$

is equal to:

$$(a \cdot x + b' \cdot y) + (b - b') \cdot y + c \cdot z$$

three new elements, one of which is stationary, which forbids us to test cointegration on the three.

So the two properties must be checked: moving together means both “moving” and “together”.

Using images rather related to stationarity (as they apply to the actual difference of two elements, without weighting coefficients) we can illustrate the concept as

- Astral bodies moving in outer space and linked together by gravity. Their distance is bounded but their position relative to each other is unknown within those bounds, and we do not know if one is leading the other.
- Human beings: if they are always close to each other, they can be decided to be related (love, hate, professional relationship). But only if they move: if they are in jail, a small distance means nothing.

In our example, the first idea could be to test cointegration between X, WD and UR. But to ensure the stability of our long term simulations, we need exports to have a unitary elasticity to WD. If this is not the case, when X reaches a constant growth rate, it will be different from that of WD: either France will become the only exporter in the world (in relative terms) or the role of France in the world market will become infinitely negligible. Both prospects are unacceptable (the first more than the second, admittedly).

This constraint can be enforced very easily by considering only in the long run (cointegrating) equation the ratio of X to WD, which we shall link to the rate of use. We will test cointegration between these two elements.

Let us first test their stationarity. We know how to do it (from the estimation of employment).

In a program, this can be done through:

```

UROOT(1,p) Log(X/WD)
UROOT(1,p) Log(UR)

```

Note: if we use menus, we should first display the group (either by selecting elements in the workfile window or creating a group). The default display mode is “spreadsheet” but the “View” item proposes other modes, among them “cointegration test” (if more than one series is displayed).

Null Hypothesis: LOG(X/WD) has a unit root
 Exogenous: Constant, Linear Trend
 Lag Length: 0 (Automatic - based on SIC, maxlag=10)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-2.709700	0.2370
Test critical values:		
1% level	-4.133838	
5% level	-3.493692	
10% level	-3.175693	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
 Dependent Variable: D(LOG(X/WD))
 Method: Least Squares
 Date: 11/07/12 Time: 13:04
 Sample (adjusted): 1975S2 2002S2
 Included observations: 55 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOG(X(-1)/WD(-1))	-0.249865	0.092211	-2.709700	0.0091
C	-0.019128	0.011568	-1.653591	0.1042
@TREND("1960S1")	3.10E-05	0.000149	0.207341	0.8366

R-squared	0.123932	Mean dependent var	-0.000178
Adjusted R-squared	0.090237	S.D. dependent var	0.018311
S.E. of regression	0.017465	Akaike info criterion	-5.204224
Sum squared resid	0.015862	Schwarz criterion	-5.094733
Log likelihood	146.1162	Hannan-Quinn criter.	-5.161883
F-statistic	3.678052	Durbin-Watson stat	1.532444
Prob(F-statistic)	0.032062		

Null Hypothesis: LOG(UR) has a unit root
 Exogenous: Constant, Linear Trend
 Lag Length: 0 (Automatic - based on SIC, maxlag=10)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-3.181020	0.0999
Test critical values:		
1% level	-4.152511	
5% level	-3.502373	
10% level	-3.180699	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
 Dependent Variable: D(LOG(UR))
 Method: Least Squares
 Date: 11/07/12 Time: 13:05
 Sample (adjusted): 1978S1 2002S2
 Included observations: 50 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOG(UR(-1))	-0.355378	0.111718	-3.181020	0.0026
C	-0.006570	0.020409	-0.321922	0.7489
@TREND("1960S1")	3.81E-05	0.000326	0.116806	0.9075

R-squared	0.177436	Mean dependent var	1.24E-05
Adjusted R-squared	0.142433	S.D. dependent var	0.035782
S.E. of regression	0.033136	Akaike info criterion	-3.918294
Sum squared resid	0.051604	Schwarz criterion	-3.803573
Log likelihood	100.9574	Hannan-Quinn criter.	-3.874608
F-statistic	5.069199	Durbin-Watson stat	1.579826
Prob(F-statistic)	0.010151		

These first tests show that both UR and the ratio of exports to world demand cannot be considered stationary, even around a trend: the T statistic is too low, and the estimated probability for the coefficient to zero is too high^{68, 69}. Let us now see if the two elements are co integrated, using the Johansen test.

For EViews this calls for:

`coint(option,p) list-of-variables or group-name`

⁶⁸ Not extremely high, however.

⁶⁹ We can observe that in a traditional least squares estimation, the same T value would give the opposite diagnosis.

Option represents the type of cointegration tested:

- a No deterministic trend in the data, and no intercept or trend in the cointegrating equation.
- b No deterministic trend in the data, and an intercept but no trend in the cointegrating equation.
- c Linear trend in the data, and an intercept but no trend in the cointegrating equation.
- d Linear trend in the data, and both an intercept and a trend in the cointegrating equation.
- e Quadratic trend in the data, and both an intercept and a trend in the cointegrating equation.
- s Summarize the results of all 5 options (a-e).

In our case, we shall use option d (trend in the cointegrating equation, no trend in the VAR)

`coint(d,p) log(x/wd) log(ur)`

Unrestricted Cointegration Rank Test (Maximum Eigenvalue)				
Hypothesized No. of CE(s)	Eigenvalue	Max-Eigen Statistic	0.05 Critical Value	Prob.**
None *	0.366312	20.98510	19.38704	0.0291
At most 1	0.186003	9.466749	12.51798	0.1535
Max-eigenvalue test indicates 1 cointegrating eqn(s) at the 0.05 level * denotes rejection of the hypothesis at the 0.05 level **MacKinnon-Haug-Michelis (1999) p-values				
Unrestricted Cointegrating Coefficients (normalized by b*S11*b=I):				
LOG(M/TD)	LOG(UR)	@TREND(60S2)		
-7.930144	24.22342	-0.030087		
22.38533	-17.70598	-0.342308		
Unrestricted Adjustment Coefficients (alpha):				
D(LOG(M/TD))	-0.013891	-0.005565		
D(LOG(UR))	-0.020030	0.008593		
1 Cointegrating Equation(s): Log likelihood 203.3606				
Normalized cointegrating coefficients (standard error in parentheses)				
LOG(M/TD)	LOG(UR)	@TREND(60S2)		
1.000000	-3.054601	0.003794		
	(0.48502)	(0.00368)		
Adjustment coefficients (standard error in parentheses)				
D(LOG(M/TD))	0.110155			
	(0.02737)			
D(LOG(UR))	0.158844			
	(0.04042)			

Obviously, cointegration is accepted (a message says so). But we can also:

- Understand the logical process:

EViews tests first if there is no cointegration. If this is accepted (if it shows a high enough probability), the process stops. But here this is refused, as the probability (that there is no cointegration) is too low.

In this case, there is at least one cointegrating equation, and EViews proceeds to testing if there is more than one. This cannot be refused here, as this time the probability is too high.

We have at least one relation, and at most one: we have one.

If the second assumption (at most 1 relation) had not been rejected, there would be at least two and we would have to continue (there cannot be more relations than variables, however).

Evidencing more than one relation is problematic, maybe worse for the model builder than finding no relation (in which case we can always proceed with adding new elements). Even if a cointegrating equation has no implications on causality between elements, we generally intend to include it in a single dynamic formula (a VAR), which does explain a given variable. With two equations, we are stuck with a parasite one, which will be difficult if not impossible to manage in the context of the model (if we stop at econometrics, the problem is smaller).

- Look at the probabilities:

We can observe if the existence (or the rejection) of at least one relation is barely or strongly accepted (and also of only one for that matter).

- Observe the coefficients in the cointegrating equation.

The equation introduces a tradeoff between several concepts (here the share of French exports in world demand and their rate of use). We have always an idea on the sign of the relationship, and also on an interval of economic validity. There is no guarantee that the value will follow these constraints. It can even happen that the right sign obtained by Ordinary Least Squares will become wrong when cointegration is tested on the same relation.

Here it is not too difficult to judge on the soundness of the explanation.

First, the sign is right: exports go down when the rate of use goes up (the sign is positive but both elements are on the same side of the formula).

The size of the coefficient is more difficult to judge. The derivative of the equation relative to Q gives:

$$\Delta(X)/X = -0.62 \Delta(UR)/UR = -0.62 (\Delta(Q)/Q - \Delta(CAP)/CAP).$$

In the short run, CAP does not change:

$$\Delta(X)/X = -0.62 \Delta(UR)/UR = -0.62 \Delta(Q)/Q$$

$$\Delta(X) = -0.62 X/Q \Delta(Q)$$

Let us suppose an increase in Q of 1 billion Euros coming only from local demand FD. In 2004, the share of exports in French GDP was 32%. The exports target will decrease by

$$\Delta X = 0.62 * 0.32 = 205 \text{ millions of Euros.}$$

The substitution effect looks quite reasonable.

Of course:

- In the long run, capacities will build up, UR will get back to its base value (we know that from the investment equation) and the loss will disappear.
- The changes in Q can come also from X, introducing a loop. This means UR might not be the best representative element. Perhaps we should restrict UR to the satisfaction of local demand (but this looks difficult to formulate).

One can also look at the significance of the coefficients, to measure the reliability of the relation. It is satisfactory.

6.3.4.3 Once cointegration has been evidenced

Two things have to be done:

- Storing the cointegrating equation and its parameters.
- Estimating the VAR (the dynamic equation) and creating the associated element.

The first task should be easy, as EViews does display the requested equation, with its values. However:

- This equation is not available as an item.
- The coefficients are not available as a vector (or as scalars).

There is a **trick**, however, which solves this problem. One can estimate a VAR, in other terms a system which includes both a dynamic equation and a cointegrating equation. This is not directly useful for us, as

- This requires to use the same variables in both forms, in other words to extend the unitary elasticity assumption to the dynamic equation, which is neither needed statistically nor realistic from an economic point of view (as we have seen when estimating employment).
- The output does not provide information on the quality of the cointegration, an essential element in our process.

But the good point is that by estimating a VAR using the same elements as the tested cointegration, we get the same cointegrating coefficients, and this time they are stored in a vector! We can then specify the cointegrating equation using these elements.

This has the extremely high advantage of allowing to establish a program which will adapt automatically to any change in the data, an essential element in all operational modelling projects.

For this particular example, after:

```
coint(d,p) log(x/wd) log(ur)
```

we shall use

```
vary _var_x.ec(d,p) 1 1 log(x/wd) log(ur)70
```

to create and estimate the VAR, and store the coefficients in a vector by accessing the first line of matrix `_var_x.b` (not displayed in the workfile window):

```
vector(10) p_x  
p_x(1)=_var_x.b(1,1)  
p_x(2)=_var_x.b(1,2)  
p_x(3)=_var_x.b(1,3)
```

The cointegrating equation will be:

```
0= p_x(1)*log(x/wd)+p_x(2)*log(ur)+p_x(3)*@trend(60S1)
```

Actually the first parameter is not really needed, as it is equal by construction. We think using it makes the equation clearer.

Estimating the dynamic equation calls for the computation of the residual in the cointegrating equation:

```
genr res_x= p_x(1)*log(x/wd)+p_x(2)*log(ur)+p_x(3)*@trend(60S1)
```

Then we estimate the VAR, releasing the constraint on the unitary elasticity of X to WD. In principle, the coefficient should be positive for WD, negative for UR, and we can introduce a lag structure for both elements.

The main issue here is economic validity: we should assess for instance if a given growth in world demand addressed to France (say by 1%) in the absence of capacity problems (this will be the role of UR) will increase exports by less or more than 1%.

In our case unfortunately the change in the rate of use is not significant⁷¹, or any transformation for that matter (it actually takes the wrong sign). We have discarded it.

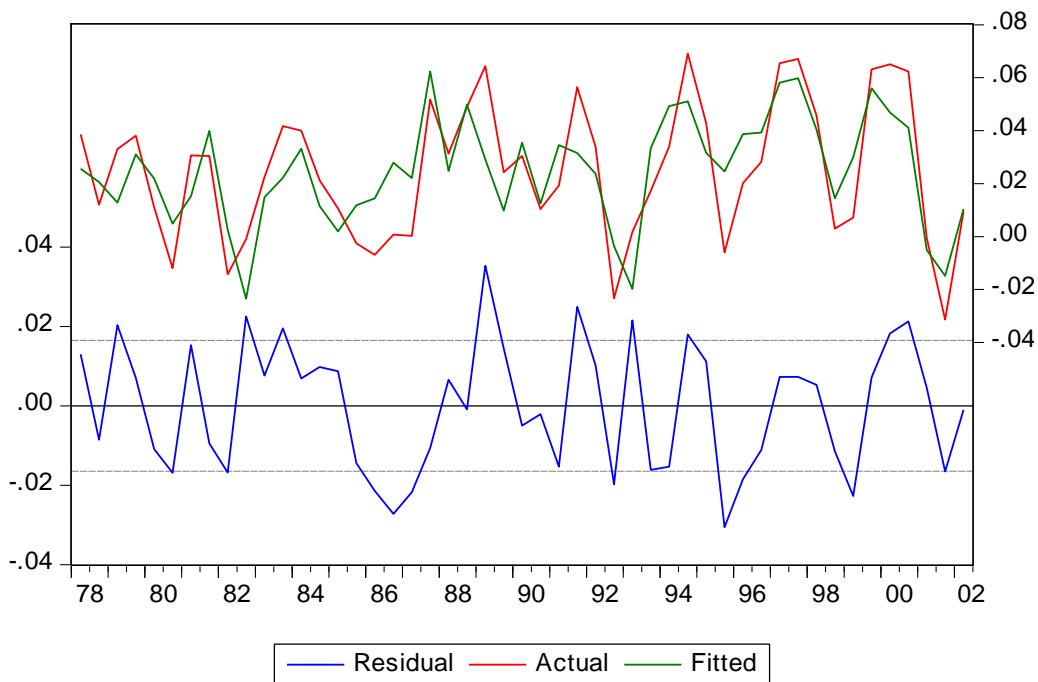
⁷⁰ The figures 1 and 1 indicate the scope of lagged variations of the left-hand side variable in the VAR which will be added to the right hand side. Here it will be 1 to 1 (or 1 with lag 1). The options used are consistent with the ones we have used in `coint` (which have been determined automatically by EViews).

⁷¹ This might be explained by the fact that the impact of the rate of use is on the export contracts, which actual exports follow with some delay. In that case the cointegrating equation might have considered the lagged rate of use.

Dependent Variable: DLOG(X)
Method: Least Squares
Date: 02/21/12 Time: 19:11
Sample (adjusted): 1978S1 2002S1
Included observations: 49 after adjustments
DLOG(X)=C_X(4)*DLOG(WD)+C_X(6)+C_X(7)*RES_X(-1)+EC_X

	Coefficient	Std. Error	t-Statistic	Prob.
C_X(4)	0.940211	0.115121	8.167185	0.0000
C_X(6)	-0.012945	0.006877	-1.882371	0.0661
C_X(7)	-0.195067	0.075051	-2.599137	0.0125

R-squared	0.607911	Mean dependent var	0.024624
Adjusted R-squared	0.590864	S.D. dependent var	0.025734
S.E. of regression	0.016460	Akaike info criterion	-5.316476
Sum squared resid	0.012463	Schwarz criterion	-5.200650
Log likelihood	133.2537	Hannan-Quinn criter.	-5.272532
F-statistic	35.66021	Durbin-Watson stat	1.662233
Prob(F-statistic)	0.000000		



6.3.5 IMPORTS: GOING FURTHER ON COINTEGRATION AND LONG TERM STABILITY.

In a single country model, the rest of the world is exogenous, and imports and exports have to be estimated separately, following of course the same guidelines:

- Imports will depend on demand and capacity utilization, through constant elasticities.

However, the definition of demand is not straightforward. For exports, we just had to consider global imports from each partner country in each product, and compute an average using a double weighting: the role of these partners in French exports, and the structure of products exported by France.

This was possible because we considered the rest of the world as exogenous, and did not try to track the origin of its imports.

Now imports can come from three endogenous elements:

- Local final demand, such as foreign cars.
- The intermediate goods necessary to local firms to satisfy this local demand. For cars it will be electronics, steel, energy to run the machines...
- Identically, intermediate goods necessary to produce exported goods.
- But not finished goods used to satisfy foreign demand: France does not re-export significantly goods without transformation (contrary to Hong Kong for instance).

Basically, two methods can be considered:

- At this stage, our model considers only final demand and exports. Obviously, they do not have the same impact on imports (due to the absence of re-exports). We can generate a global demand by applying to exports a correcting factor. Under rather standard assumptions (same import share in all uses, unitary ratio of intermediate consumption to value added), this factor can be set at 0.5.
- We can also define intermediate consumption, and add it to final demand to get the total demand of the country, a share of which will be imported. This method is obviously more acceptable from the economic point of view. Unfortunately it relies on the computation of intermediate consumption, a variable less accurately measured, and sensitive to categories and the integration of the productive process⁷².

We have chosen this last method nevertheless, favoring economic properties over statistical reliability.

Of course, you can guess we shall try for cointegration, as we did for exports. Let us see however what happens with an autoregressive process, just for the sake of learning its danger.

Our formula will make imports depend on total demand and the rate of use:



⁷² For instance, if good A (say cotton) is used to produce good B (unprinted fabric) which gives good C (printed fabric), both A and B will be counted as intermediary consumption. If the fabric is printed as the same time it is produced, only A will be counted. If we consider value added, the total amount will not change, just the number of elements.

$$\Delta M / M = a \cdot \Delta TD / TD + b \cdot \Delta UR / UR$$

Where

$$TD = FD + IC$$

$$IC = tc \cdot Q$$

And tc is the number of intermediary consumption units required to produce one unit of GDP.

By integration, we get

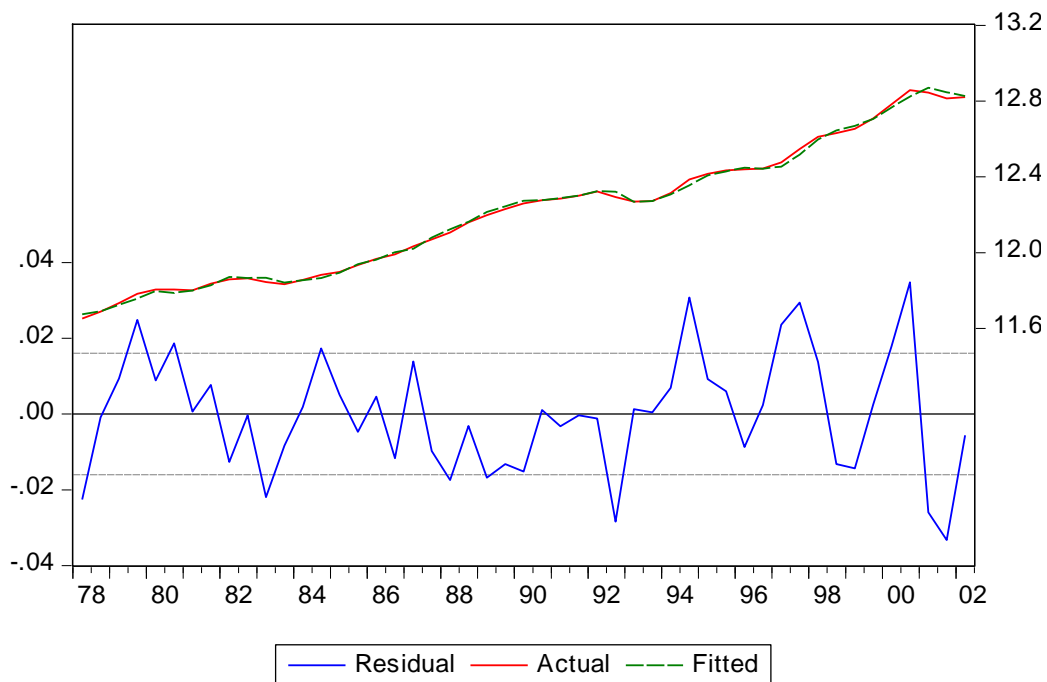
$$\text{Log}(M) = a \cdot \text{Log}(TD) + b \cdot \text{Log}(UR) + c$$

As we could expect, the basic formula faces high autocorrelation (which makes the non-significance of UR an irrelevant issue):

Trying to eliminate autocorrelation does not even work:

Dependent Variable: LOG(M)
 Method: Least Squares
 Date: 11/07/12 Time: 13:20
 Sample (adjusted): 1977S2 2002S1
 Included observations: 50 after adjustments
 LOG(M)=C_M(1)*LOG(TD)+C_M(2)*LOG(UR)+C_M(3)

	Coefficient	Std. Error	t-Statistic	Prob.
C_M(1)	2.395145	0.042125	56.85867	0.0000
C_M(2)	0.162941	0.145531	1.119634	0.2686
C_M(3)	-22.70890	0.614266	-36.96918	0.0000
R-squared	0.986068	Mean dependent var		12.19882
Adjusted R-squared	0.985475	S.D. dependent var		0.355330
S.E. of regression	0.042824	Akaike info criterion		-3.405322
Sum squared resid	0.086192	Schwarz criterion		-3.290601
Log likelihood	88.13305	Hannan-Quinn criter.		-3.361635
F-statistic	1663.282	Durbin-Watson stat		0.165492
Prob(F-statistic)	0.000000			



In addition to the bad statistical results, all the above equations face an economic problem: the coefficient for demand is always quite high. This has to be expected, as in the sample period, imports have permanently grown faster than demand, and the coefficient represents the comparison between growths (especially in the AR formulation).

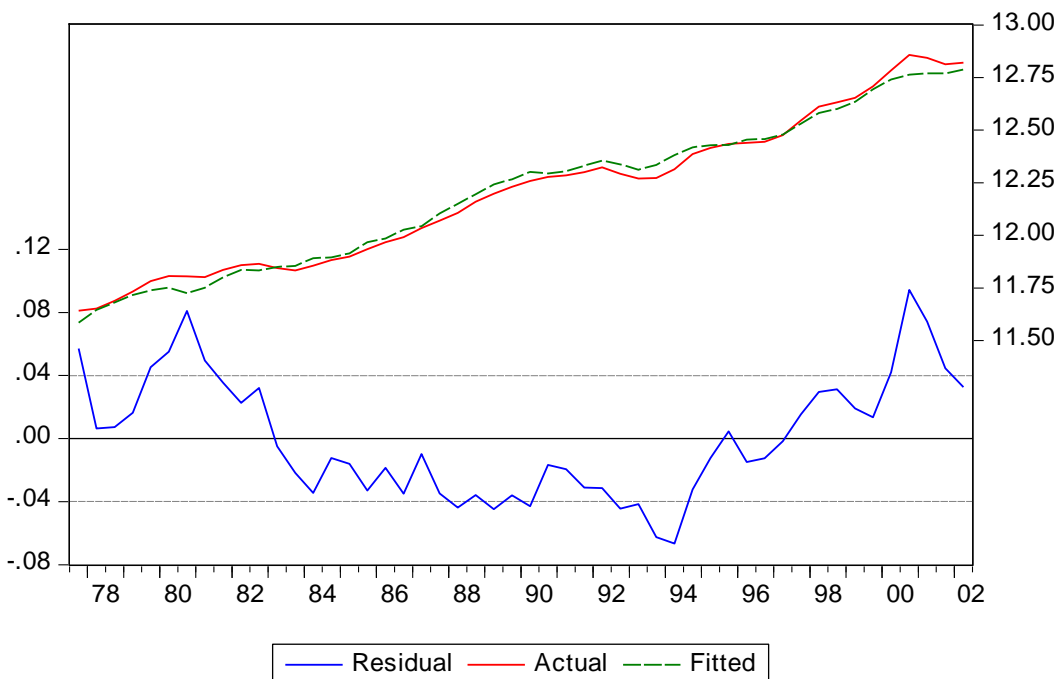
Now the question is: is the growth of demand the only explanation for the growth of imports? In other words, is there not an autonomous force increasing the weight of foreign trade, independently from growth itself? Or: if demand did not grow for a given period, would imports stay stable, or keep part of their momentum?

The best way to answer this question is probably through econometrics, by introducing an additional time trend. This option meets partial success on the economic side, but the obvious autocorrelation forbids its use:

Dependent Variable: LOG(M)
 Method: Least Squares
 Date: 11/07/12 Time: 13:23
 Sample (adjusted): 1977S2 2002S1
 Included observations: 50 after adjustments
 $LOG(M)=C_M(1)*LOG(TD)+C_M(2)*LOG(UR)+C_M(3)+C_M(4)*T$

	Coefficient	Std. Error	t-Statistic	Prob.
C_M(1)	1.472617	0.321303	4.583269	0.0000
C_M(2)	0.380206	0.154757	2.456791	0.0179
C_M(3)	-46.22968	8.150882	-5.671739	0.0000
C_M(4)	0.018580	0.006423	2.892783	0.0058

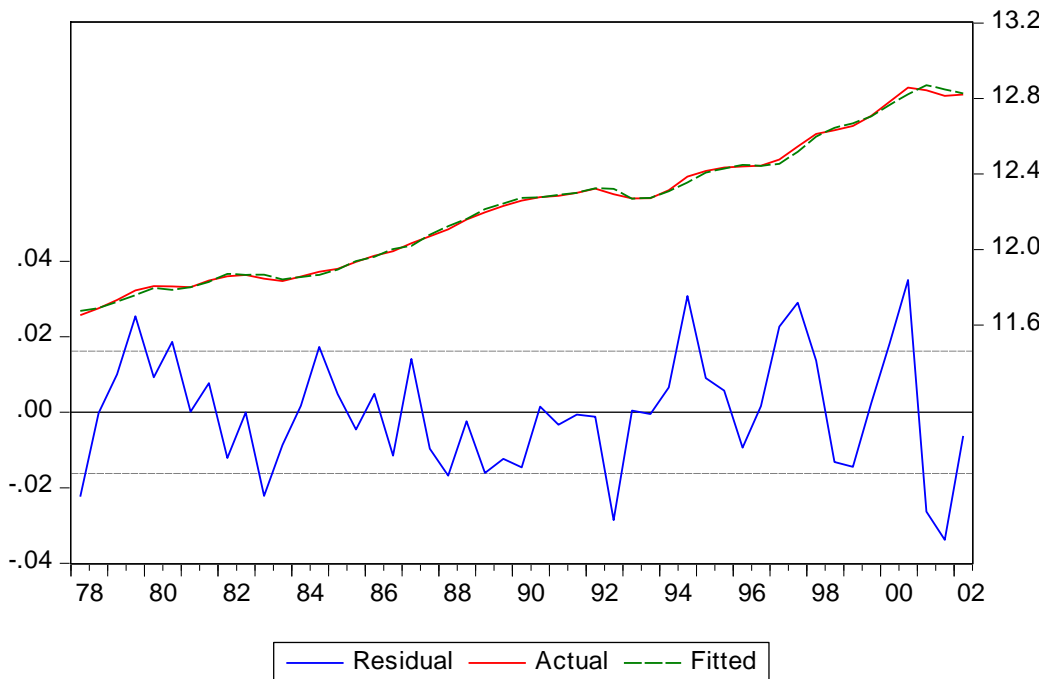
R-squared	0.988212	Mean dependent var	12.19882
Adjusted R-squared	0.987444	S.D. dependent var	0.355330
S.E. of regression	0.039816	Akaike info criterion	-3.532460
Sum squared resid	0.072926	Schwarz criterion	-3.379498
Log likelihood	92.31150	Hannan-Quinn criter.	-3.474211
F-statistic	1285.479	Durbin-Watson stat	0.273643
Prob(F-statistic)	0.000000		



Trying to eliminate autocorrelation, everything breaks down:

Dependent Variable: LOG(M)
 Method: Least Squares
 Date: 11/07/12 Time: 13:25
 Sample (adjusted): 1978S1 2002S1
 Included observations: 49 after adjustments
 Convergence achieved after 19 iterations
 $LOG(M)=C_M(1)*LOG(TD)+C_M(2)*LOG(UR)+C_M(3)+C_M(4)*T$
 $+[AR(1)=C_M(5)]$

	Coefficient	Std. Error	t-Statistic	Prob.
C_M(1)	2.557629	0.350384	7.299512	0.0000
C_M(2)	0.007373	0.076966	0.095791	0.9241
C_M(3)	-28.34794	12.80653	-2.213554	0.0321
C_M(4)	0.001626	0.008525	0.190675	0.8497
C_M(5)	0.903205	0.054688	16.51554	0.0000
R-squared	0.998033	Mean dependent var	12.21019	
Adjusted R-squared	0.997854	S.D. dependent var	0.349705	
S.E. of regression	0.016200	Akaike info criterion	-5.311125	
Sum squared resid	0.011548	Schwarz criterion	-5.118083	
Log likelihood	135.1226	Hannan-Quinn criter.	-5.237885	
F-statistic	5580.647	Durbin-Watson stat	1.352968	
Prob(F-statistic)	0.000000			
Inverted AR Roots	.90			



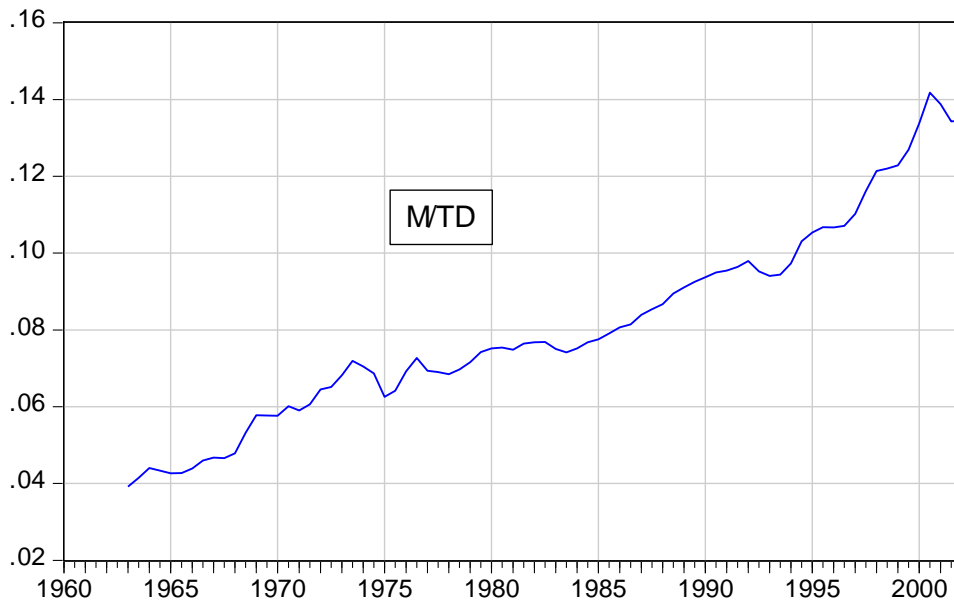
But if we nevertheless use any of the formulas, the model will face a big problem:

Let us take the derivative of the formula without a trend, corrected (partly) for autocorrelation:

$$\Delta(M)/M = 2.62 \Delta(TD)/TD$$

$$\Delta(M) = 2.62 (M/TD) \Delta(TD)$$

Let us consider the evolution of the share of imports in total demand: in 2004 it reaches 15%:



In 2004 this gives

$$\Delta(M) = 2.62 \times 0.15 \Delta(TD) = 0.39 \Delta(TD)$$

$$\Delta(M) = 2.62 \times 0.15 \Delta(TD) = 0.39 (\Delta(FD) + \Delta(Q))$$

$$\Delta(Q) = \Delta(FD) - \Delta(M) = 0.61 \Delta(TD) - 0.39 \Delta(Q)$$

$$\Delta(Q) = 0.61/1.39 \Delta(TD)$$

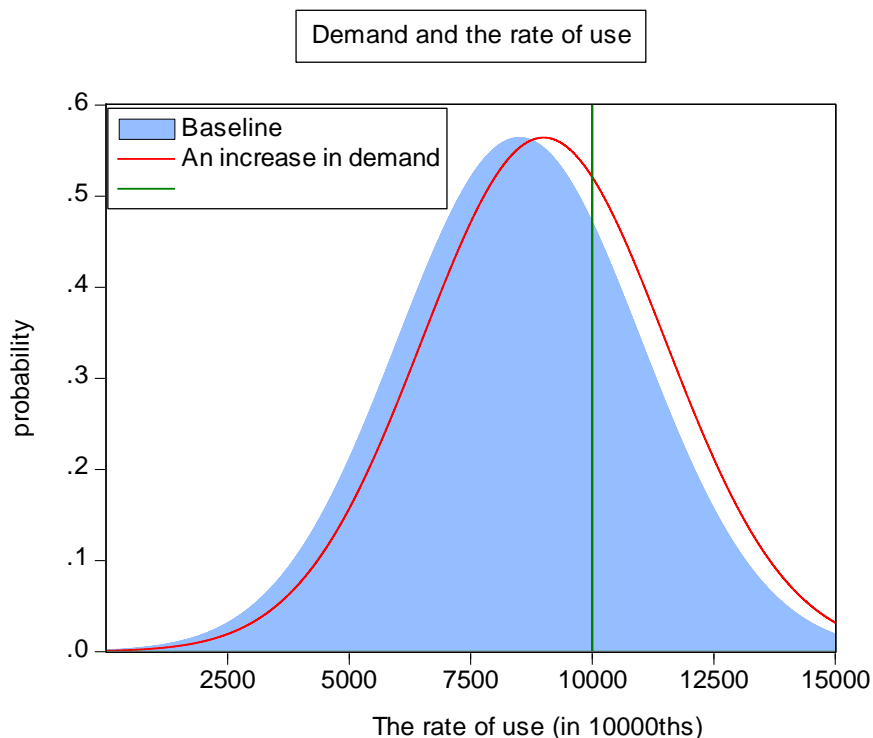
Or more generally using "a"

$$\Delta(Q) = (1 - 2.62 a) / (1 + 2.62 a) \Delta(TD)$$

This means that if "a" keeps growing (say to 0.4) the coefficient will become negative, and it will become possible that an increase in final demand will reduce GDP, with a negative multiplier.

The problem with our formulation is actually very clear: in terms of model properties, it is reasonable to suppose that in the short run, an increase in final demand will increase imports beyond their normal share, by generating local bottlenecks on domestic supply. But the explanatory element should be the rate of use of capacities. At it is fixed in the long run, the share should go back to normal with time.

A rate of use of 85% (a normal value over the whole economy) does not mean that all firms work at 85% capacity, in which case they can easily move to 86% if needed. It means that the rates of use follows a given distribution, some at lower than 85%, some higher, some at 99%, and a finite number at 100% (see graph).



An increase in demand will move the curve to the right, and more firms will face the limit: an increase of 1% will be met halfway by firms starting from a 99.5% rate of use. The additional demand, if clients do not accept local substitutes, will have to be supplied by imports.

However, local firms will react to this situation, and try to gain back lost market shares by increasing their capacities through investment: this is the mechanism we have described earlier. In our small model, the long term rate of use is fixed: the sharing of the additional demand will come back to the base values. These values can increase with time due to the expansion of world trade.

To represent this process, we need a formula which:

- Enforces a unitary elasticity of imports to the total demand variable, with a positive additional effect of the rate of use.
- Allows free elasticities in the short run.

We recognize in the above the same framework as in the exports equation, using cointegration.

Note: as in the exports equation, taking into account the whole model will make the long term elasticity of imports to demand fully unitary, as the rate of use is fixed by the investment equation. A shock on any exogenous element (including the productivity of capital) will bring back UR to its base value, forbidding any trade-off. This will change when we consider the profitability of capital, which will allow the target for UR to vary.

Let us try to estimate such an equation.

We shall start by testing the cointegration between the share of imports in demand: M/TD and the rate of use.

Before, we test the stationarity of M/TD, or rather its logarithm (UR has already been tested):

It is strongly contradicted by the Dickey Fuller test:

Null Hypothesis: LOG(M/TD) has a unit root Exogenous: Constant, Linear Trend Lag Length: 2 (Automatic - based on SIC, maxlag=11)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic				
Test critical values:	1% level		-4.081666	
	5% level		-3.469235	
	10% level		-3.161518	
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation Dependent Variable: D(LOG(M/TD)) Method: Least Squares Date: 11/07/12 Time: 13:29 Sample (adjusted): 1964S2 2002S2 Included observations: 77 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOG(M(-1)/TD(-1))	-0.097183	0.048031	-2.023358	0.0467
D(LOG(M(-1)/TD(-1)))	0.439724	0.104688	4.200318	0.0001
D(LOG(M(-2)/TD(-2)))	-0.337369	0.107119	-3.149483	0.0024
C	-0.294686	0.154078	-1.912578	0.0598
@TREND("1960S1")	0.001282	0.000669	1.915312	0.0594
R-squared	0.277500	Mean dependent var		0.014636
Adjusted R-squared	0.237361	S.D. dependent var		0.029908
S.E. of regression	0.026119	Akaike info criterion		-4.389613
Sum squared resid	0.049117	Schwarz criterion		-4.237418
Log likelihood	174.0001	Hannan-Quinn criter.		-4.328737
F-statistic	6.913494	Durbin-Watson stat		1.974691
Prob(F-statistic)	0.000091			

Now we test cointegration of LOG(UR) and LOG(M /TD)

It fails!

Date: 11/07/12 Time: 13:31
 Sample (adjusted): 1979S1 2002S2
 Included observations: 48 after adjustments
 Trend assumption: Linear deterministic trend (restricted)
 Series: LOG(M/TD) LOG(UR)
 Lags interval (in first differences): 1 to 2

Unrestricted Cointegration Rank Test (Trace)

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.**
None	0.154667	12.10335	25.87211	0.8048
At most 1	0.080687	4.038169	12.51798	0.7366

Trace test indicates no cointegration at the 0.05 level
 * denotes rejection of the hypothesis at the 0.05 level
 **MacKinnon-Haug-Michelis (1999) p-values

Unrestricted Cointegration Rank Test (Maximum Eigenvalue)

Hypothesized No. of CE(s)	Eigenvalue	Max-Eigen Statistic	0.05 Critical Value	Prob.**
None	0.154667	8.065182	19.38704	0.8156
At most 1	0.080687	4.038169	12.51798	0.7366

Max-eigenvalue test indicates no cointegration at the 0.05 level
 * denotes rejection of the hypothesis at the 0.05 level
 **MacKinnon-Haug-Michelis (1999) p-values

Unrestricted Cointegrating Coefficients (normalized by b*S11*b=l):

LOG(M/TD)	LOG(UR)	@TREND(60S2)
-5.221110	36.73565	0.057533
-26.68206	16.01556	0.406982

Unrestricted Adjustment Coefficients (alpha):		
D(LOG(M/TD))	0.000665	0.004179
D(LOG(UR))	-0.010282	0.003897
1 Cointegrating Equation(s): Log likelihood 242.4388		
Normalized cointegrating coefficients (standard error in parentheses)		
LOG(M/TD)	LOG(UR)	@TREND(60S2)
1.000000	-7.035985	-0.011019
	(2.33234)	(0.00511)
Adjustment coefficients (standard error in parentheses)		
D(LOG(M/TD))	-0.003472	
	(0.01192)	
D(LOG(UR))	0.053682	
	(0.02230)	

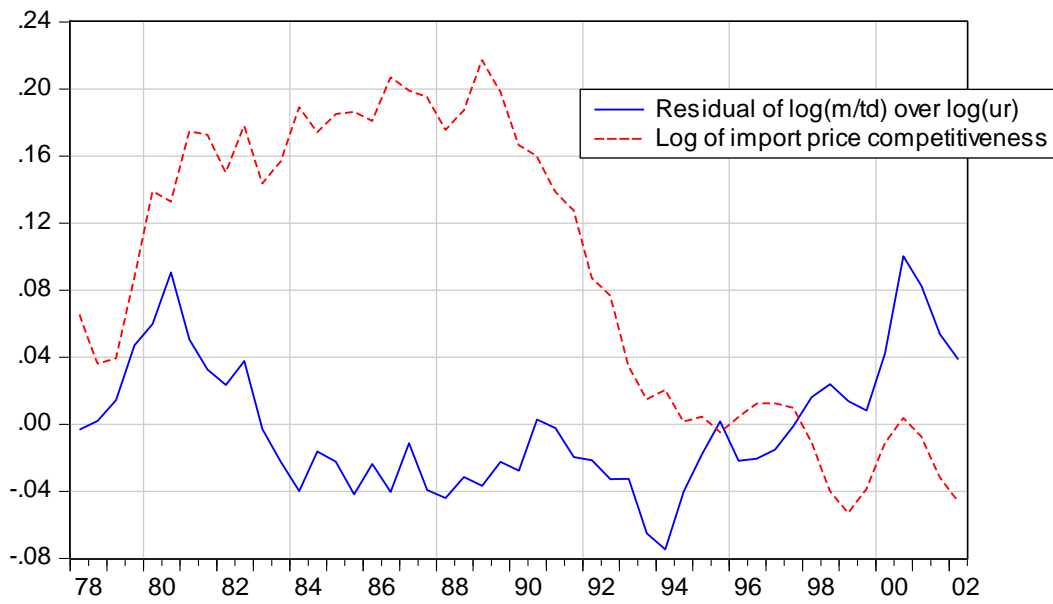
Let us not despair. The absence of cointegration is not so bad⁷³, as it allows us to proceed further. If a set of two variables does not work, why not a set of three?

Now which additional element could we consider? The natural candidate comes both from theory and from the data:

If demand is present but local producers have no capacity problems, how can foreign exporters penetrate a market? Of course, through price competitiveness, in other words by decreasing the import price compared to the local one.

This observation is confirmed by the data. Let us regress the import-demand ratio over the rate of use, consider the residual (the unexplained part) and compare it to the ratio of import to local prices: we observe a clearly negative relation.

⁷³ Identifying two equations would be much worse, especially in a modelling framework.



We observe a clear correlation, especially if we consider variations around a negative trend.

After having tested of course:

- Non-stationarity of Log(COMPM)
- Non- cointegration of Log(COMPM) with both Log(UR) and log(M/TD) individually,

We can test the cointegration of the three elements:

It works!!

Date: 11/07/12 Time: 13:37
 Sample (adjusted): 1979S1 2002S2
 Included observations: 48 after adjustments
 Trend assumption: Linear deterministic trend (restricted)
 Series: LOG(M/TD) LOG(COMPM) LOG(UR)
 Lags interval (in first differences): 1 to 2

Unrestricted Cointegration Rank Test (Trace)

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.**
None *	0.430972	46.69131	42.91525	0.0200
At most 1	0.282157	19.62767	25.87211	0.2454
At most 2	0.074486	3.715476	12.51798	0.7828

Trace test indicates 1 cointegrating eqn(s) at the 0.05 level
 * denotes rejection of the hypothesis at the 0.05 level
 **MacKinnon-Haug-Michelis (1999) p-values

Unrestricted Cointegration Rank Test (Maximum Eigenvalue)

Hypothesized No. of CE(s)	Eigenvalue	Max-Eigen Statistic	0.05 Critical Value	Prob.**
None *	0.430972	27.06363	25.82321	0.0342
At most 1	0.282157	15.91220	19.38704	0.1490
At most 2	0.074486	3.715476	12.51798	0.7828

Max-eigenvalue test indicates 1 cointegrating eqn(s) at the 0.05 level
 * denotes rejection of the hypothesis at the 0.05 level
 **MacKinnon-Haug-Michelis (1999) p-values

Unrestricted Cointegrating Coefficients (normalized by b*S11*b=I):

LOG(M/TD)	LOG(COMPM)	LOG(UR)	@TREND(60S2)
-26.25607	-17.29497	41.40372	0.295194
18.37876	-2.562664	-13.07275	-0.337567
15.08409	9.794651	15.94489	-0.167276

Unrestricted Adjustment Coefficients (alpha):			
D(LOG(M/TD))	0.004500	-0.003607	-0.002787
D(LOG(COM...)	0.005905	0.008244	-0.002935
D(LOG(UR))	-0.010552	-0.001976	-0.006576
1 Cointegrating Equation(s):		Log likelihood	374.4392
Normalized cointegrating coefficients (standard error in parentheses)			
LOG(M/TD)	LOG(COMPM)	LOG(UR)	@TREND(60S2)
1.000000	0.658704 (0.12206)	-1.576920 (0.23954)	-0.011243 (0.00078)
Adjustment coefficients (standard error in parentheses)			
D(LOG(M/TD))	-0.118147 (0.05526)		
D(LOG(COM...)	-0.155046 (0.08330)		
D(LOG(UR))	0.277042 (0.11304)		
2 Cointegrating Equation(s):		Log likelihood	382.3953
Normalized cointegrating coefficients (standard error in parentheses)			
LOG(M/TD)	LOG(COMPM)	LOG(UR)	@TREND(60S2)
1.000000	0.000000	-0.862522 (0.39271)	-0.017123 (0.00092)
0.000000	1.000000	-1.084551 (0.66393)	0.008926 (0.00156)
Adjustment coefficients (standard error in parentheses)			
D(LOG(M/TD))	-0.184434 (0.06492)		
D(LOG(COM...)	-0.003529 (0.09270)		
D(LOG(UR))	0.240731 (0.13762)	0.187552 (0.07508)	

Of course we have to consider the coefficients in the equation. They describe:

- An apparently high sensitivity of imports to the rate of use (but remember the investment equation will stabilize it in the end).

However the true effect is not so high. If the equation was applied to the short term, with the same 0.15 coefficient as before, we would get:

$$\Delta(M) = 1 \times 0.15 \Delta(TD) + 1.57 \times 0.15 \Delta(Q)$$

$$\Delta(M) = 0.15 \Delta(TD) + 0.23 \Delta(Q)$$

And if we do not consider the change in exports:

$$\Delta(FD - Q) = 0.15 \Delta(FD + ct. Q) + 0.23 \Delta(Q)$$

$$\Delta(Q) = 0.85 / 1.38 \Delta(FD) = 0.62 \Delta(FD)$$

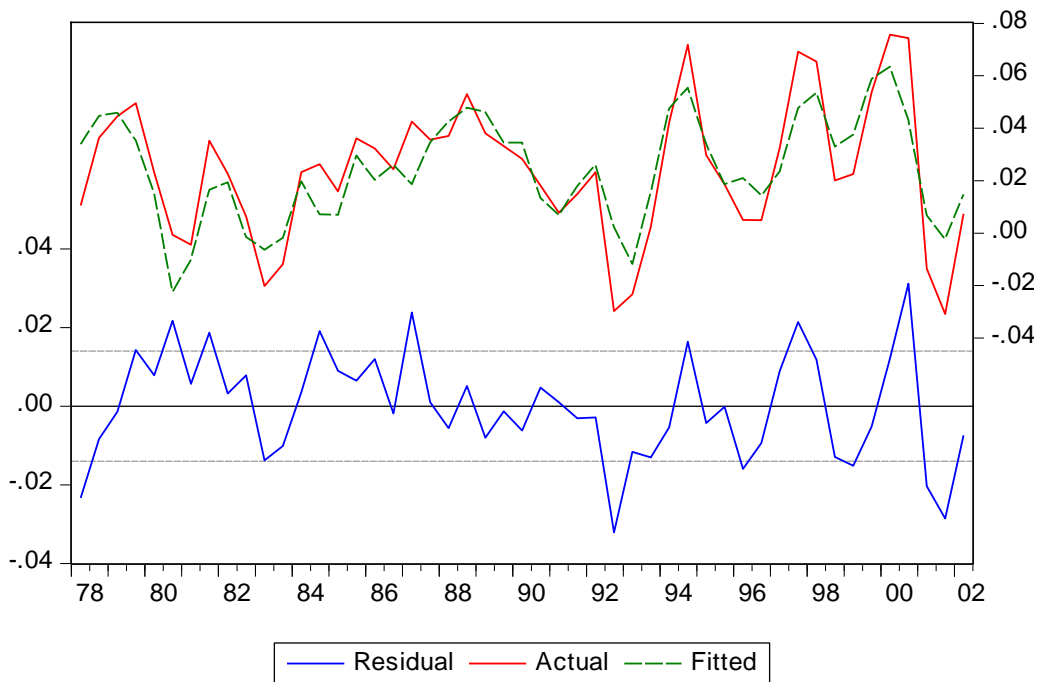
A quite acceptable multiplier for France.

- A much higher sensitivity than in the autoregressive formula.

We can now test the VAR:

Dependent Variable: DLOG(M)
Method: Least Squares
Date: 11/07/12 Time: 14:03
Sample (adjusted): 1978S1 2002S2
Included observations: 50 after adjustments
DLOG(M)=C_M(1)*DLOG(TD)+C_M(2)*DLOG(UR)+C_M(3)*RES_M(-1)
+C_M(4)

	Coefficient	Std. Error	t-Statistic	Prob.
C_M(1)	1.635016	0.411194	3.976267	0.0002
C_M(2)	0.170456	0.083792	2.034268	0.0477
C_M(3)	-0.152136	0.044768	-3.398328	0.0014
C_M(4)	-0.452681	0.132508	-3.416262	0.0013
R-squared	0.684182	Mean dependent var		0.023935
Adjusted R-squared	0.663585	S.D. dependent var		0.025806
S.E. of regression	0.014968	Akaike info criterion		-5.489229
Sum squared resid	0.010305	Schwarz criterion		-5.336267
Log likelihood	141.2307	Hannan-Quinn criter.		-5.430980
F-statistic	33.21779	Durbin-Watson stat		1.408067
Prob(F-statistic)	0.000000			



The results are rather acceptable, even if the UR coefficient is barely significant⁷⁴. But the graph is rather favorable. By the way, it shows that rejecting an R-Squared of 0 is not justified when the dependent element shows a high variability.

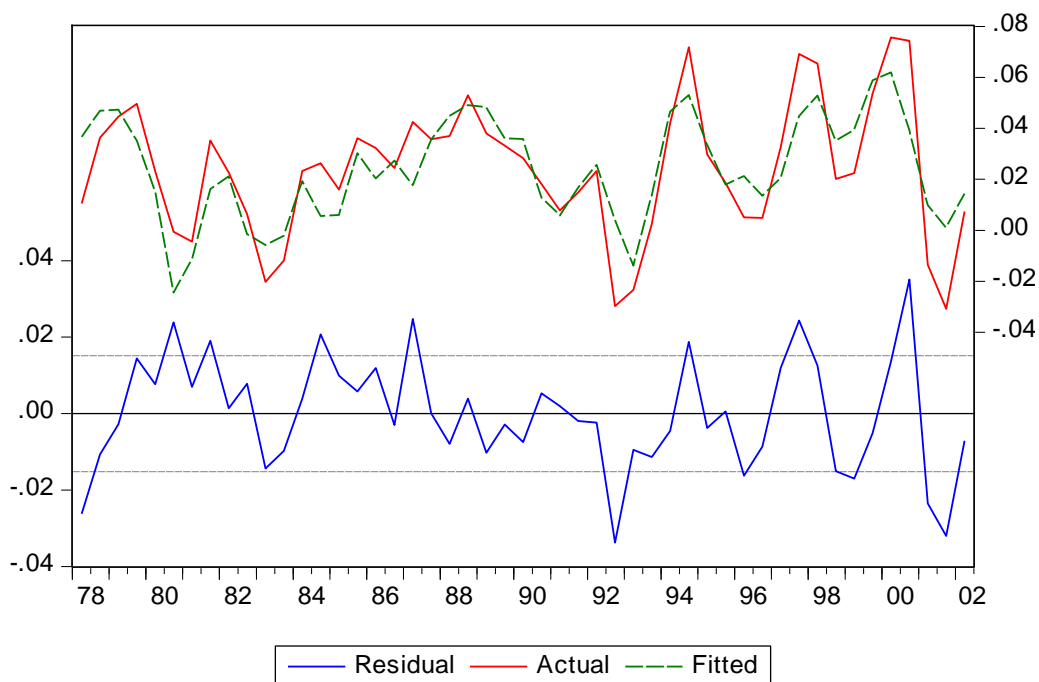
Forcing the demand coefficient to a lower value (say 1.2) increases the coefficient of the rate of use without reducing too much the quality. The reliability of UR increases, due mostly to the higher coefficient (of course UR and TD are correlated as both include Q) and a little to the standard error:

⁷⁴ We see that the t-Statistic is higher than 2, but the p-value higher than 5%!

Dependent Variable: DLOG(M)
 Method: Least Squares
 Date: 11/07/12 Time: 14:04
 Sample (adjusted): 1978S1 2002S2
 Included observations: 50 after adjustments
 $DLOG(M)=1.2*DLOG(TD)+C_M(2)*DLOG(UR)+C_M(3)*RES_M(-1)+C_M(4)$

	Coefficient	Std. Error	t-Statistic	Prob.
C_M(2)	0.230294	0.061900	3.720416	0.0005
C_M(3)	-0.183253	0.033793	-5.422878	0.0000
C_M(4)	-0.542178	0.102118	-5.309318	0.0000

R-squared	0.676498	Mean dependent var	0.023935
Adjusted R-squared	0.662731	S.D. dependent var	0.025806
S.E. of regression	0.014987	Akaike info criterion	-5.505189
Sum squared resid	0.010556	Schwarz criterion	-5.390468
Log likelihood	140.6297	Hannan-Quinn criter.	-5.461503
F-statistic	49.14241	Durbin-Watson stat	1.380208
Prob(F-statistic)	0.000000		



6.3.6 BACK TO THE RESIDUAL CHECK

The method we are using for storing equations has an additional advantage. Now that the residuals have been introduced with their estimated values, all the equations should hold true. The checking process can now be extended to all the endogenous variables.

Theoretically the estimated equations should be consistent with the data, as merging with the model the actual estimated equations ensures the consistency. However:

- If the package does not provide this direct storing, or if the equation had to be normalized by the modeller, editing the formulation could introduce errors.
- The storing of coefficients may have been done badly.
- The text, series or coefficients may have been modified by the user after estimation
- One could have accessed other series or coefficients than the ones used by the estimation (for example one can seek them in another bank including series of similar names).

The reasons for a non-zero residual are less numerous than for identities. They can come only from the fact that equation elements have changed since the last estimation.

Obviously, the main suspect is the data. New data series are due to be inconsistent with the previous estimation, whether it has been updated (moving to a more precise version) or corrected (suppressing an error).

Actually, in EViews, applying a new version of an equation to a model requires, in addition to its estimation, to actually **merge** it again into the model. This will create a new compiled version, without need to explicitly update the model.

Anyway, in our opinion, applying a new estimation should call for a full model re-creation. This is the only way to guarantee a clear and secure update.

For our model, the statements for the residual check will be the following:

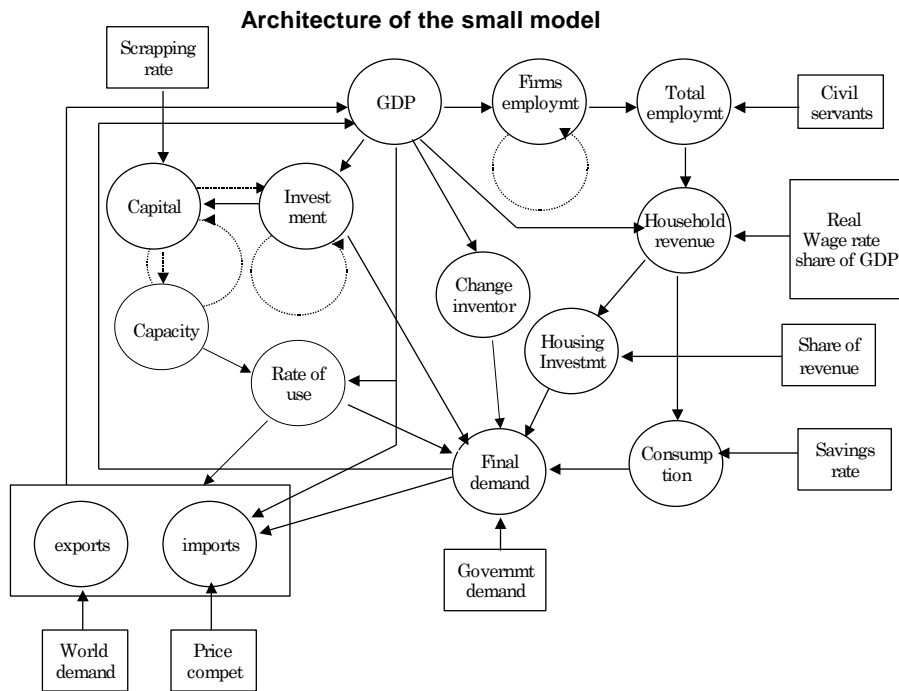
```
' We check the residuals

smpl 1980S2 2002S1
__fra_1.append assign @all _c
solve(d=f) __fra_1

for !i=1 to g_vendo.@count
%2=g_vendo.@seriesname(!i)
genr dc_{%2}={%2}-{%2}_c
genr pc_{%2}=100*dc_{%2}/({%2}+({%2}=0))
next
```

The solution series will have the suffix “_c”, and the residuals the prefix “dc_” for the errors in levels, and “pc_” for the relative errors.

We can now present the framework of the model.



6.3.7 THE PRESENT MODEL

In EViews notations, its specifications are:

$$[1] \text{CAP} = \text{pk} * \text{K}(-1)$$

$$[2] \text{Q} + \text{M} = \text{FD} + \text{X}$$

$$[3] \text{UR} = \text{Q} / \text{CAP}$$

$$[4] \text{CI} = \text{tc} * \text{Q}$$

$$[5] \text{IC}/\text{Q}(-1) = 0.130 * \text{@PCH}(\text{Q}) + 0.214 * \text{@PCH}(\text{Q}(-1)) + \text{EC_IC}$$

$$[6] \text{IP}/\text{K}(-1) = 0.825 * \text{I}(-1)/\text{K}(-2) + 0.0279 * \text{UR} + 0.152 * .25 * \text{Q}/\text{Q}(-4) - 0.0525 + \text{EC_I}$$

$$[7] \log(\text{PRLE_T}) = \text{c_prle}(1) + \text{c_prle}(2) * (\text{t} - 2002) + \text{c_prle}(3) * (\text{t} - \text{t1}) * (\text{t} < \text{t1}) + \text{c_prle}(4) * (\text{t} - \text{t2}) * (\text{t} < \text{t2})$$

$$[8] \text{LED} = \text{Q} / \text{PRLE_T}$$

$$[9] \text{DLOG}(\text{LE}) = 0.587 * \text{DLOG}(\text{LED}) + 0.411 * \text{LOG}(\text{LED}(-1)/\text{LE}(-1)) + 0.000502 - 0.0167 * ((\text{T} = 1968.5) - (\text{T} = 1968)) + \text{EC_LE}$$

$$[10] \text{LT} = \text{LE} + \text{lg}$$

$$[11] \text{RHI} = \text{wr} * \text{LT} + \text{r_rhiq} * \text{Q}$$

$$[12] \text{ IH} = r_{\text{ih}} * \text{RHI}$$

$$[13] \text{ CO} = \text{RHI} * (1 - \text{sr})$$

$$[14] \text{ FD} = \text{CO} + \text{IP} + \text{CI} + \text{IH} + \text{gd}$$

$$[15] \text{ TD} = \text{FD} + \text{tc} * \text{Q}$$

$$[16] \text{ RES_M} = \log(\text{M} / \text{TD}) - 1.322108 * \log(\text{UR}) + 0.419518 * \log(\text{compm}) - 0.012582 * (@\text{trend}(60:1) * (\text{t} \leq 2002) + @\text{elem}(@\text{trend}(60:1), "2002S2") * (\text{t} > 2002))$$

$$[17] \text{ DLOG}(\text{M}) = 1.2 * \text{DLOG}(\text{TD}) + 0.282 * \text{DLOG}(\text{UR}) - 0.212 * \text{RES_M}(-1) - 0.629 + \text{EC_M}$$

$$[18] \text{ RES_X} = \log(\text{X} / \text{wd}) + 0.686 * \log(\text{UR}) - 4.87\text{E-}05 * (@\text{trend}(60:1) * (\text{t} \leq 2002) + @\text{elem}(@\text{trend}(60:1), "2002S2") * (\text{t} > 2002))$$

$$[19] \text{ DLOG}(\text{X}) = 0.940 * \text{DLOG}(\text{wd}) - 0.0129 - 0.195 * \text{RES_X}(-1) + \text{EC_X}$$

$$[20] \text{ K} = \text{K}(-1) * (1 - \text{dr}) + \text{IP}$$

CHAPTER 7: TESTING THE MODEL THROUGH SIMULATIONS OVER THE PAST

We have now achieved the production of a model, for which:

- All the identities are consistent with the data.
- All estimations are in our opinion acceptable from a statistical point of view.
- All the coefficient values and equation specifications seem consistent with economic theory⁷⁵.
- All the necessary connections between model elements are present.
- The estimated equations have been built in such a way that they should provide a long term solution for the full model.

This does not mean our model is acceptable.

- The consistency of the data might hide errors compensating each other.
- Some complex equations might have hidden wrong individual properties.
- Connections between concepts might have been forgotten or wrongly specified.
- The growth rates provided naturally by equations could be the wrong ones.
- Some error correction processes could be diverging instead of converging.
- Putting together individually acceptable formulations might create a system with unacceptable properties.

Let us just give an example: if in the short run increasing government demand by 1000 creates 800 consumption and 600 investment, while exports do not change and imports increase by 300, the individual equations light look acceptable, but the model will diverge immediately through an explosive multiplier effect ($800 + 600 - 300 = 1100$).

Our next goal will be to control:

- That the model can indeed be solved
- That it can be used for forecasts.
- That it can be used for policy analysis.

Actually, our first tests will rather answer the reverse question:

- Is there some indication that the model is unsuitable for forecasts, and for policy analysis?

In our opinion, it is only by simulations over the future (its normal field of operation, actually) that we can really validate the use of a model. But as usual, problems should be diagnosed as soon as possible. And the availability of actual data enhances strongly some of the tests.

Finally, the errors evidenced at this stage might help to build a better forecasting structure.

Let us first address the process of solving the model.

⁷⁵ In some cases we might have been obliged to calibrate the values.

7.1 THE SOLUTION

To solve the model we need to apply a method. Let us present the different algorithms.

7.1.1 GAUSS-SEIDEL

This is the most natural algorithm: one often uses Gauss-Seidel without knowing it, like M. Jourdain (the Bourgeois Gentleman) makes prose.

The method starts from initial values. They can be the historical values on the past, on the future the values computed for the previous period or for an alternate base simulation. The whole set of equations will be applied in a given order, using as input the most recent information (computations replace the initial values). This gives a new set of starting values. The process is repeated, using always the last information available, until the distance between the two last solutions is small enough to be considered negligible. One will then consider that the solution has been reached.

Let us formalize this process.

Considering the model

$$y_t = f_t(y_t, y_{t-1}, x_t, \hat{\alpha})$$

with y_t vector of $y_{i,t}$ and $i \in [1, n]$

As only present values are going to change during computation, we will not consider the other elements, and will drop the time index.

$$y = f(y)$$

We will use it to define the particular endogenous, and an exponent to define the iteration count.

a - We start from y^0 , value at iteration 0.

b - We add 1 to the number of iterations (which we shall note k); this gives to the first iteration the number 1.

c - We compute y_i^k from $i = 1$ to n , taking into account the $i-1$ values we have just produced. This means we compute:

$$y_i^k = f(y_1^k, \dots, y_{i-1}^k, y_i^{k-1}, \dots, y_n^{k-1})$$

(at the first iteration, explanatory elements will take the starting value y^0 if their index is higher than the computed variable)⁷⁶.

d – At the end of the process, we compare y^k and y^{k-1} : if the distance is small enough for every element (we will see later which criteria we use) we stop the process, and take as solution the last value. If not, we check if we have reached the maximum number of iterations, in which case we accept the failure of the algorithm, and stop. Otherwise we resume the process at step b.

Clearly, this algorithm requests an identified model.

7.1.2 RITZ-JORDAN

The Ritz-Jordan method is similar to the one above: it simply abstains from using values computed at the current iteration:

$$y^k = f(y^{k-1})$$

Refusing to take into account the last information, it looks less efficient than Gauss-Seidel. In our opinion, its only interest appears when the model faces convergence problems: it makes their interpretation easier by reducing the interferences between variables.

This method is not provided by EViews.

7.1.3 NEWTON AND ITS VARIANTS

Contrary to the two above, the Newton method applies naturally to non-identified formulations. It represents actually a generalization to an n-dimensional problem of the well-known method using a sequence of linearizations to solve a single equation.

Let us consider the model:

$$f_t(y_t, y_{t-1}, x_t, \hat{\alpha}) = 0$$

that we will simplify as above into:

⁷⁶ This means only variables which are used before they are computed must be given values for initialization. We shall come back to this later.

$$f(y) = 0$$

The linearization of f around a starting solution gives, by calling "fl" the value of f linearized:

$$(\partial f / \partial y)_{y=y^0} \cdot (y - y^0) = fl(y) - f(y^0)$$

Solving the system for $fl(y) = 0$ leads to:

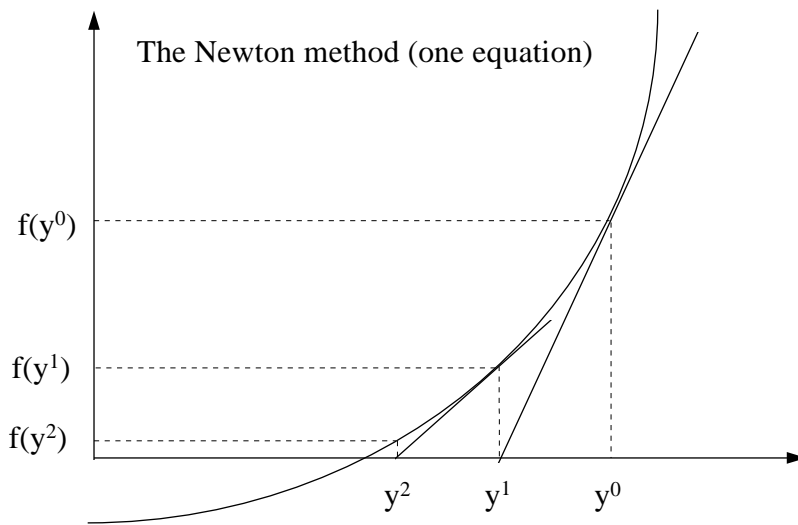
$$y = y^0 - (\partial f / \partial y)_{y=y^0}^{-1} \cdot f(y^0)$$

With an identified system:

$$y - f(y) = 0$$

we would get naturally:

$$y = y^0 - (I - \partial f / \partial y)_{y=y^0}^{-1} \cdot (y^0 - f(y^0))$$



Linearizing the model again, around the new solution y^1 , and solving the new linearized model, we define an iterative process which, as the preceding, will stop when the distance between the two last values gets small enough. Implementing this method is more complex: in addition to inverting a matrix, each iteration involves the computation of a Jacobian. This can be made practically in two ways:

- Analytically, by determining from the start the formal expressions of derivatives. At each iteration, we shall compute them again from the present values of variables. This method supposes either undertaking the derivation "by hand" with a high probability of errors, or having access to an automatic formal deriver, a program analyzing the text of equations to produce the formal expression of their derivatives. To a high initial cost, it opposes a simplification of computations during the iterative process⁷⁷.

EViews allows both methods⁷⁸.

- By finite differences, determining separately each column of the Jacobian by the numerical computation of a limited first order development, applied in turn to each variable. One computes the y vector using the base values, then for starting values differing only by a single variable, and compares the two results to get a column of the Jacobian. One will speak then of a method of secants, or pseudo-Newton.

The derivate formulation becomes then:



⁷⁷ However, changing some model specifications calls for a new global derivation (or a dangerous manual updating).

⁷⁸ Unfortunately, the associated code is not apparently available to the user, which would allow interesting computations.

$$\sum_{j=1}^n \left\{ \left[f_i(y^k + e_j \Delta y_j) - f(y^k) \right] / \Delta y_j \right\} (y_j - y_j^k) = f_i(y) - f_i(y^k)$$

where e_j is a vector of dimension n (number of endogenous), with 1 in position i and 0 otherwise.

In other words, the element of the Jacobian:

$$\partial f_i / \partial y_j \quad \text{will be approximated by} \quad \left| f_i(y^k + e_j \Delta y_j) - f(y^k) \right| / \Delta y_j$$

One will have only to compute the y vector $n+1$ times: one time with no modification and one time for each of the endogenous variables.

The expensive part of this algorithm being clearly the computation of the Jacobian and its inversion, a variant will consist in computing it only each m iterations. The convergence will be slower in terms of number of iterations, but the global cost might decrease.

EViews provides another alternative: Broyden's method, which uses a secant method and does not require to compute the Jacobian at each step. As we shall see later, this method proves often very efficient.

7.1.3.1 The identified case

If the model is in "identified" form:

$$y = f(y)$$

the Newton algorithm will be applied to

$$y - f(y) = 0$$

and the Newton formula becomes:

$$y = y^0 - (I - \partial f / \partial y)_{y=y^0}^{-1} \cdot (y^0 - f(y^0))$$

or

$$y = (I - \frac{\partial f}{\partial y})^{-1}_{y=y^0} \cdot (f(y^0) - \frac{\partial f}{\partial y}_{y=y^0} \cdot y^0)$$

This does not change the technical process.

7.1.4 BROYDEN'S METHOD

Broyden's method (also called secant method) computes the Jacobian only one, in the same way as Newton's, and computes a new value of the variable accordingly.

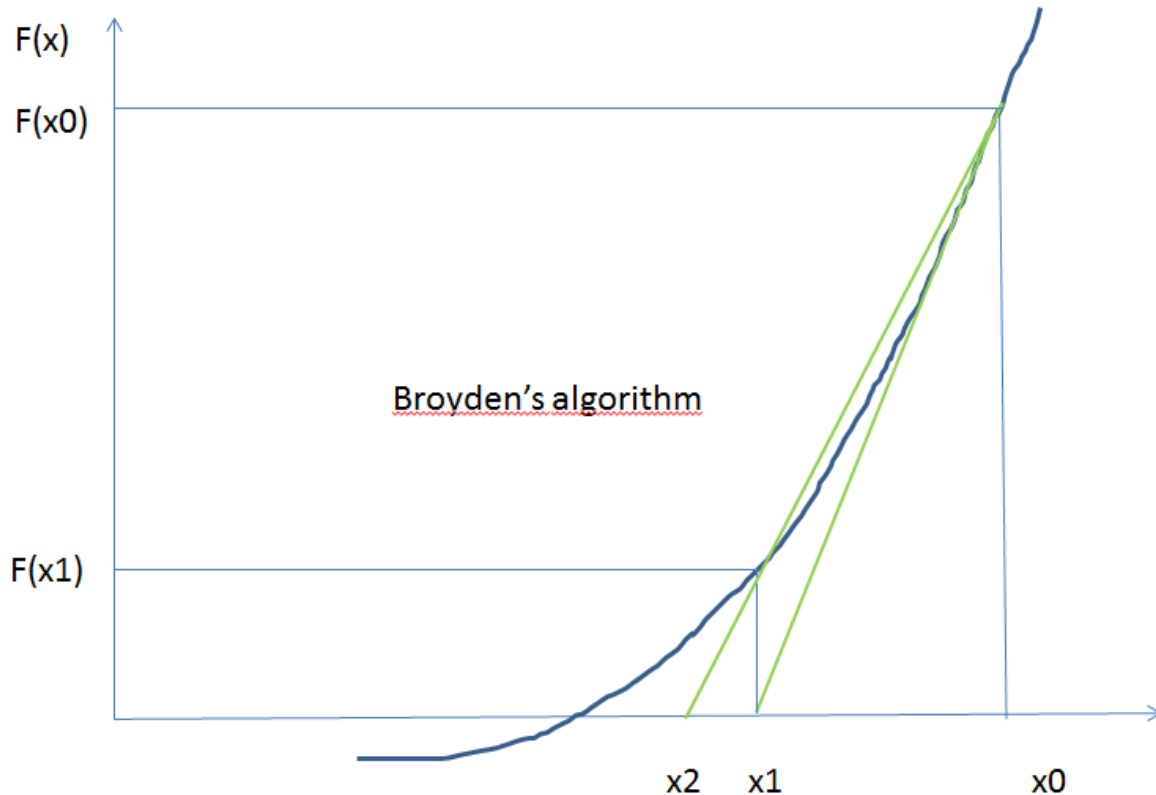
After that, it updates the Jacobian, not by derivation, but by considering the difference between the previous one, and the direction leading from the previous solution to the new one.

The formula for updating the Jacobian is:

$$J_{t+1} = J_t + (F(x_{t+1}) - F(x_t) - J_t \Delta x_t) \cdot \Delta x'_t / (\Delta x'_t \cdot \Delta x_t)$$

where J is the Jacobian, F the function which should reach zero, and x the vector of unknown variables.

Let us clarify all this with a graph based on the single equation case.



We can see that the direction improves with each iteration, less than Newton but more than Gauss-Seidel (for which it does not improve at all).

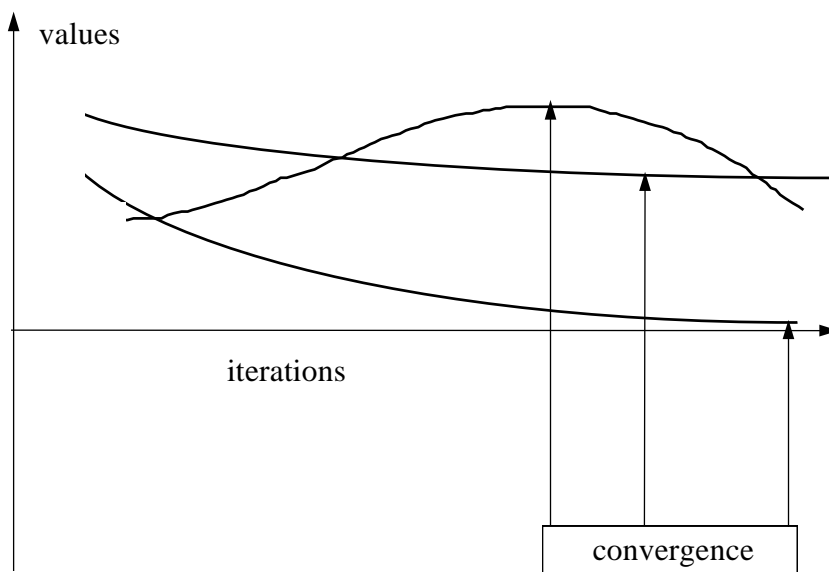
Otherwise the method shares all the characteristics of Newton's, in particular its independence on equation ordering. It takes generally more iterations, but each of them is cheaper (except for the first).

We shall see that on average it looks like the most efficient option on the whole, both in terms of speed and probability of convergence⁷⁹. But the diagnosis is not so clear cut.

7.1.5 ITERATIONS AND TEST OF CONVERGENCE

Methods described above have a common feature: starting from initial values, they apply formulations to get a new set. The process is repeated until the two last sets are sufficiently close to be considered as the solution of the system.

One cannot identify the difference between two iterations with the precision actually reached (or the difference to the solution). This is valid only for alternate processes. For monotonous ones, it actually can be the reverse: the slower the convergence, the smaller the change in the criterion from one iteration to the other, and the higher the chance that the criterion will be reached quite far from the solution. As to cyclical processes, they can reach convergence mistakenly at the top or bottom of a cycle.



So one could criticize this type of method, by stressing that the relative stability of values does not mean that the solution has been found. However, one can observe that if the values do not change, it means that the computation which gave a variable would give the same result with the new values of its explanatory variables: it means also that the equation holds almost true.

⁷⁹ The most important feature in our opinion.

However, it is clear that we do not get the exact solution. This criticism should not be stretched too much: the precision of models is in any event limited, and even supposedly exact algorithms are limited by the precision of computers.

7.1.5.1 The general options

For the algorithm to know at which moment to stop computations, we shall have to establish a test.

In fact, the only criterion used in practice will consider the variation of the whole set of variables in the solution, from an iteration to the other.

It can be measured, for each variable:

in relative values:

$$d_i = \left| \frac{y_i^k - y_i^{k-1}}{y_i^{k-1}} \right|$$

or in levels:

$$d_i = |y_i^k - y_i^{k-1}|$$

As to the condition for accepting convergence, it can be defined:

by variable: $d_i < c_i, \forall i$

on the whole set: $d_i < c_i, \forall i$

or sometimes through a global measure: $f(d) < c$

Generally one will choose a criterion in relative value, each error being compared with a global criterion. This value will have to be small compared to the expected model precision (so that the simulation error will not really contribute to the global error), and to the number of digits used for results interpretation.

The most frequent exception should correspond to variables which, like the trade balance, fluctuate strongly and can even change sign: here the choice of a criterion in level seems a natural solution, which will avoid a non-convergence diagnosis due to negligible fluctuations of a variable if its solution is (by pure chance) very small.

For example, if the convergence threshold is 0.0001 in relative value, convergence will be refused if solutions for the US trade balance alternate by chance between - 1 billion current US Dollars and - 1.0002 billion⁸⁰, while a difference of 200 000 Dollars, at the scale of US foreign trade, is obviously very small. And this difference, which represents less than one

⁸⁰ There is no risk for this in present times.

millionth of US exports and imports, might never be reduced if the computer precision guarantees only 8 significant figures⁸¹.

In practice we shall see that the test could be restricted to a subset of variables in the model, the convergence of which extends mathematically to global convergence.

The value given to the criterion can depend:

- On the algorithm used:
 - In case of Gauss-Seidel, each additional digit bears roughly the same cost.
 - In case of Newton, the number of digits gained increases with the iterations: beyond the minimum level (say 0.01%) a given gain is cheaper and cheaper (this will be developed later).
- On the type of simulation:
 - For a forecast, one will not be too strict, as we all know the precision is quite low anyway. Forecasting growth of 2.05% and 2.07% three years from now delivers the same message, especially the actual growth might materialize as 1% (more on forecast precision later).
 - For a shock analysis, especially if the shock is small, the evaluation of the difference between the two simulations is obviously more affected by the error: decisions increasing in GDP by 0.07% and 0.09% will not be given the same efficiency.
- And perhaps on the stochastic character:

In a stochastic simulation, it is essential that the consequence for the solution of introducing small random residuals is precisely associated with the shock, and not on the simulation process.

As to the number of iterations, it will be used as a limit, after which we suppose that the model has no chance to converge. In practice one never considers stopping an apparently converging process, just because has taken too much time. So the only case is when the process is not progressing, because it is oscillating between two or more solutions, and the deadlock has to be broken. Reverting to the use of damping factors (described later) should solve the problem in the Gauss-Seidel case.

7.1.5.2 The EViews options

Testing convergence under EViews is not very flexible: the only option allowed is the level of the (relative) convergence criterion, and it will apply to all variables.

One can also decide on the maximum number of iterations. For most models, after 1000 iterations, convergence becomes rather improbable. But just to make sure, one can set an initial high number. Observing the actual number required can allow to improve the figure.

⁸¹ Exports and imports will be precise to the 8th digit, but the difference, a million times smaller, to the 2nd only.

7.1.6 STUDY OF THE CONVERGENCE

We are going now to show how the choice of the algorithm affects the convergence process.

Let us begin by stating the problem, and introducing some definitions.

7.1.6.1 The incidence matrix

The incidence matrix of an n-equation model

$$f(y, \dots) = 0 \text{ (n endogenous variables, n equations)}$$

will be defined as the Boolean matrix A (dimension n by n) such that

- $A_{i,j} = 1$ if the variable y_j appears formally, through its present value, in the equation of rank i.
- $A_{i,j} = 0$ otherwise.

We will suppose the model to be normalized, therefore put under the form:

$$y - f(y) = 0$$

where the variable y_i will appear naturally to the left of the equation of rank i: the main diagonal of the matrix will be composed of 1s.

The definition of the incidence matrix, as one can see, depends not only on the model, but also on the ordering of equations, actually the one in which they are going to be computed.

The formal presence of a variable in an equation does not necessarily mean a numerical influence: it could be affected by a potentially null coefficient, or intervene only in a branch of an alternative. Often we will not be able to associate to a model a unique incidence matrix, nor a matrix constant with time, except if one considers the total set of potential influences (the matrix will be then the Boolean sum of individual Boolean matrices).

One will also notice that defining the incidence matrix does not require the full formulations, or the knowledge of variable values. We simply need to know the list of variables which appear in each explanation, as well as their instantaneous or lagged character⁸².

Application to our model

⁸² Following our methodology, the incidence matrix can be produced before any estimation.

To apply this technique to our model, we can rely on the block structure provided by EViews, through:

Access to **model (double-click)>View>Block structure**, which gives in our case:

Number of equations: 20
Number of independent blocks: 3
Number of simultaneous blocks: 1
Number of recursive blocks: 2

Block 1: 3 Recursive Equations

cap(1)	x(19)	prle_t(7)		
--------	-------	-----------	--	--

Block 2: 14 Simultaneous Equations

ur(2)	q(3)	ic(4)	ci(5)	i(6)
led(8)	le(9)	lt(10)	rhi(11)	ih(12)
co(13)	fd(14)	td(15)	m(17)	

Block 3: 3 Recursive Equations

res_m(16)	res_x(18)	k(20)		
-----------	-----------	-------	--	--

All these **elements are consistent with the graph of page 196.**

We can use these elements to improve the ordering.

First, we can use the above separation to move the three predetermined variables at the beginning, and the three post determined at the end, which give the following matrix:

	prle_														res_					
	cap	t	x	q	ur	ic	ci	i	led	le	lt	rhi	ih	co	fd	td	m	res_x	m	k
cap	1																			
prle_		1																		
t			1																	
x				1																
q				1											1		1			
ur	1			1	1															
ic				1		1														
ci				1			1													
i				1	1			1												
led		1		1					1											
le		1		1						1										
lt				1							1									
rhi				1								1								
ih				1									1							
co				1										1						
fd				1											1					
td				1																
m				1																
res_x				1																
res_				1																
m				1											1					
k				1																

We can see that the model has been cut into three parts:

- A three equation block, with elements which do not depend on the complement, or on subsequent variables in the same block. The variables in this can then be computed once and for all, in a single iteration, at the beginning of the solving process. Actually they do not depend on any variable in the same block, but this is not required.

This property is called recursiveness, and the block is usually named the **prologue**.

We can see that variables can belong to this block for various reasons:

- Prle_t depends only on time. The only reason for introducing its equation is to allow easy modification in forecasts.
- Cap depends on an exogenous and a predetermined variable.
- X should depend on the rest of the equilibrium (through UR) but this link has not been evidenced statistically, leaving only the instantaneous influence of the exogenous WD.

In practice, however, respecting the convergence threshold will need two iterations, the starting value being different from the solution found, unless the recursivity is known from the start, and the first solution accepted without control⁸³.

- A three equations block, in which elements do not affect the complement, and do not depend on subsequent variables in the same block. These variables can be computed after all the others, once and for all in one pass. Again, they do not depend on any variable in the same block.

In this block we find:

- The residuals for the cointegration equations, which will only be corrected at the next period.
- The end-of-period capital, which obviously cannot affect the equilibrium for the period.

We shall see later another important category: variables with a purely descriptive role, like the government deficit in GDP points.

- The rest of the model is simultaneous, and sometimes called the heart. We can check on the graph that taking any couple of variables in the set, there is at least one sequence of direct causal relationships leading from the first to the second, and vice versa. This means also that exogenizing any element (at a value different from the model solution of course) all the other elements will be affected.

We can now try to better interpret the simultaneity in the heart. The first stage is observing the presence of loop variables.

The incidence matrix allows defining **loop** variables, as variables that enter in an equation of rank lower than the one that defines them, or will be used before they are computed. In matrix notations, we shall have:

for variable j , $\exists A_{i,j} = 1$ such as $i < j$

The variables appearing as an explanatory factor in their own equation of definition also will have to be added to this set. But in practice this case is rather rare.

Let us look at our incidence matrix. Two loop variables are present: FD and M. The reason is that they are used to compute Q, in an equation which appears at the end (of the heart).

Actually X should also be present, but as UR appears only through its lagged value, and WD is exogenous, its exact value can be computed immediately, which means it is located in the prologue. In a way it is now technically exogenous (considering only same period relationships);

Of course, a model can contain a sequence of non-recursive blocks. This will happen for instance for two subsequent non-recursive blocks if elements of the second depend on elements in the first, but not vice-versa. Between the two blocks, a recursive one can appear.

We shall see examples of this situation when we deal with more complex models.

⁸³ Which is of course the case for EViews.

The definition of the set of loop variables presents the following interest: if this set is empty, the model becomes recursive, which means that the sequential calculation of each of the equations (in the chosen order) gives the solution of the system. Values obtained at the end of the first iteration will satisfy the global set of equations, none of these values being questioned by a later modification of an explanatory element. And a second iteration will not modify the result.

This favorable case is rare enough⁸⁴. However, one can often identify simultaneous subsets (or « blocks ») with a recursive structure relative to each other, such that the p first equations of the model are not influenced by the last $n - p$. The process of simulation can be then improved, as it will suffice to solve in sequence two systems of reduced size, allowing to gain time as the cost of solution grows more than proportionally with the number of equations. This property is evident for Newton, where the cost of both Jacobian computation and inversion decrease, less for Gauss-Seidel and Broyden, where the only proof comes from practice.

It is obvious that discovering the above properties and performing the associated reordering are interesting for the model builder, as they allow to improve the organization of the solution process, and therefore reduce computation time. This process will also allow to detect logical errors, for example by evidencing the recursive determination of an element known as belonging to a loop (such as investment in the Keynesian loop). Most software packages, including EViews, take care of this search and the associated reorganization, but further improvement may be sought in the solving process by modifying by hand the order of equation computations.

In the light of previous observations, one can look:

- For the best block-recursive structure.
- Inside each block, for the order which permits the fastest convergence, or ensures its highest probability.

The first goal is indisputable, and in fact the easiest to realize from the algorithmic viewpoint. The separation found is unique, but some orderings of blocks can be equivalent (for example equations using only exogenous or lagged elements can be placed in any order).

The second is much less obvious and in any case more complex. One will seek generally to minimize the number of loop variables. The cost of this technique will depend on the ambition: the search for one set of loop variables from which we cannot eliminate an element (Nepomiaschy and Ravelli) is cheaper than the search for all orderings with the smallest possible number of elements (Gilli and Rossier). The first type of set will be called minimal, the second minimum. In fact, minimizing the number of loop variables might not be a good preparation for the use of the Gauss - Seidel algorithm, as we will see later.

EViews determines automatically the block structure of the model (which is de facto optimal, even if other organizations exist). As to reordering the simultaneous blocks, if it does not apply an optimization algorithm, it determines the loop variables associated with a given ordering (actually associated to the initial one) and places the associated equations at the end of the block. The efficiency of this last action is questionable, as it means that in a given iteration all computations use the previous value of loop variables, delaying somewhat the impact of “new” information.

⁸⁴ And the associated model is probably quite poor.

For instance, in our model, we can reduce the number of loop variables by transferring the equation for Q to the end of the heart:

	prle_															res_				
	cap	t	x	ur	ic	ci	i	led	le	lt	rhi	ih	co	fd	td	m	q	res_x	m	k
cap	1																			
prle_		1																		
t			1																	
x				1																
ur	1			1														1		
ic					1														1	
ci						1														1
i				1			1													1
led		1						1												1
le		1							1											
lt								1		1										
rhi										1	1									1
ih											1	1								
co												1	1							
fd						1	1						1	1	1					
td																1				
m				1												1	1	1		
q			1													1		1	1	
res_x				1																1
res_m					1									1						1
k							1													1

Now Q is the only loop variable (but a strong one as it appears 6 times in that capacity).

The new ordering is:

Prologue:

$$[1] CAP = pk * K(-1)$$

$$[2] \log(PRLE_T) = c_prle(1) + c_prle(2) * (t - 2002) + c_prle(3) * (t - t1) * (t < t1) + c_prle(4) * (t - t2) * (t < t2)$$

$$[3] DLOG(X) = 0.940 * DLOG(wd) - 0.0129 - 0.195 * RES_X(-1) + EC_X$$

Heart:

$$[4] UR = Q / CAP$$

$$[5] CI = tc * Q$$

$$[6] IC/Q(-1) = 0.130 * @PCH(Q) + 0.214 * @PCH(Q(-1)) + EC_IC$$

$$[7] I/K(-1) = 0.825 * I(-1)/K(-2) + 0.0279 * UR + 0.152 * .25 * Q/Q(-4) - 0.0525 + EC_I$$

$$[8] LED = Q / PRLE_T$$

$$[9] DLOG(LE) = 0.587 * DLOG(LED) + 0.411 * LOG(LED(-1)/LE(-1)) + 0.000502 - 0.0167 * ((T = 1968.5) - (T = 1968)) + EC_LE$$

$$[10] LT = LE + lg$$

$$[11] RHI = wr * LT + r_rhiq * Q$$

$$[12] IH = r_ih * RHI$$

$$[13] CO = RHI * (1 - sr)$$

$$[14] FD = CO + IP + CI + IH + gd$$

$$[15] TD = FD + CI$$

$$[16] DLOG(M) = 1.2 * DLOG(FD+CI) + 0.282 * DLOG(UR) - 0.212 * RES_M(-1) - 0.629 + EC_M$$

$$[17] Q + M = FD + X$$

Epilogue:

$$[18] RES_M = \log(M / (FD+CI)) - 1.322108 * \log(UR) + 0.419518 * \log(\text{compm}) - 0.012582 * (@\text{trend}(60:1) * (t \leq 2002) + @\text{elem}(@\text{trend}(60:1), "2002S2") * (t > 2002)))$$

$$[19] RES_X = \log(X / wd) + 0.686 * \log(UR) - 4.87E-05 * (@\text{trend}(60:1) * (t \leq 2002) + @\text{elem}(@\text{trend}(60:1), "2002S2") * (t > 2002)))$$

$$[20] K = K(-1) * (1 - dr) + IP$$

and solving the model will be done in the following way (we shall see later the corresponding statements):

Model: _FRA_1

Date: 02/15/07 Time: 16:52

Sample: 2003S1 2500S2

Solve Options:

Dynamic-Deterministic Simulation

Solver: Gauss-Seidel

Max iterations = 5000, Convergence = 1e-006

Scenario: Scenario 1
Solve begin 16:52:42

2003S1	Block 1 - 3 eqns	Solved (recursive block)
2003S1	Block 2 - 13 eqns	Convergence after 10 iterations
2003S1	Block 3 - 3 eqns	Solved (recursive block)
2003S2	Block 1 - 3 eqns	Solved (recursive block)
2003S2	Block 2 - 13 eqns	Convergence after 10 iterations
2003S2	Block 3 - 3 eqns	Solved (recursive block)
2004S1	Block 1 - 3 eqns	Solved (recursive block)
2004S1	Block 2 - 13 eqns	Convergence after 10 iterations
2004S1	Block 3 - 3 eqns	Solved (recursive block)
2004S2	Block 1 - 3 eqns	Solved (recursive block)
2004S2	Block 2 - 13 eqns	Convergence after 11 iterations
2004S2	Block 3 - 3 eqns	Solved (recursive block)

It is now possible to compare convergence properties of the different methods, beginning with the Gauss-Seidel algorithm.

One can see that, for the Gauss-Seidel algorithm, only loop variables influence (by their starting values) the result of the current iteration. Let us note y_b the vector of loop variables (complement y_c), associated to the current ordering. One could consider replacing in turn each occurrence of a variable already computed by its current expression (possibly already transformed). This will give a model in which only loop variables appear on the right side:

$$y = g(y_b)$$

And for the convergence process, we have only to consider in the y_b elements of the result:

$$y_b = g(y_b)$$

One has just to solve the new (smaller) model, and y_c will be obtained in one pass at the end of the process.

Our model gives:

$$[4] \text{ ur} = f4(q)$$

$$[5] \text{ ic} = f5(q)$$

$$[6] \text{ ci} = f6(q)$$

$$[7] \text{ i} = f7(q, f4(q))$$

[8] led = f8(q)

[9] le = f9(f8(q))

[10] lt=f(f10(f9(f8(q))))

[11] rhi = f11(f(f10(f9(f8(q))),q))

[12] ih = f12(f11(f(f10(f9(f8(q))),q)))

[13] co = f13(f12(f11(f(f10(f9(f8(q))),q))))

[14] fd = f14(f13(f12(f11(f(f10(f9(f8(q))),q))),f7(q,f4(q)), f6(q), f12(f11(f(f10(f9(f8(q))),q))))

[15] td = f(f14(f13(f12(f11(f(f10(f9(f8(q))),q))),f7(q,f4(q)), f6(q), f12(f11(f(f10(f9(f8(q))),q))))),f5(q))

[16] m = f15(f14(f13(f12(f11(f(f10(f9(f8(q))),q))),f7(q,f4(q)), f6(q), f12(f11(f(f10(f9(f8(q))),q))))), f5(q), f4(q))

Even for this simple example, producing the total set of explicit relations would have been difficult if not impossible. But if g is formally unknown, its values can be computed for any y_b (in a typical Gauss- Seidel iteration).

Thus, in our example, the successive application (in the given order) of the formulas in the f system associates to an initial value of Q a numerical value of the associated g (but not its solution of course). No other initial value has to be considered.

Again, this shows that for the convergence, one has only to consider the reduced model:

$$y_b = g(y_b) \text{ or } y_b^{k+1} = g(y_b^k)$$

and that the control of convergence can be limited to these variables.

Let us use this observation to interpret the convergence of the algorithm. We shall linearize the process around the solution:

$$y^k - y^* = g(y^{k-1}) - g(y^*) \approx (\partial g / \partial y)_{y=y^{k-1}} (y^{k-1} - y^*)$$

This approximation, if it can be used in practice (which supposes either that one is near the solution, or that the model is linear enough), shows that the vector of errors is multiplied at each iteration by the Jacobian. If the Jacobian is stable enough in the process, this makes convergence almost linear, and the cost of obtaining an additional solution digit will be independent from its position in the representation of the variable. And in order for the algorithm to converge effectively, the generally necessary and always sufficient condition is that when one raises the Jacobian matrix to a power growing to the infinite, its value goes to zero. An equivalent condition is that the every eigenvalue of the matrix

is strictly lower than 1 in module (one will say that the spectral radius is lower than unity). Thus, in our example, convergence will be assured if:

$$\Delta(g(Q))/Q < 1$$

One sees that minimizing the number of loop variables, if it reduces the size of the Jacobian, has no specific reason to reduce the spectral radius⁸⁵.

Actually it frequently happens that the reordering does not modify in any way the convergence process, as it affects only the point at which it enters in an otherwise unchanged loop. This is the case in our example: by transferring the Q equation to the end, we start with the value for Q in the workfile and not the one which balances the supply-demand equilibrium using the initial values for X, M and FD. But from then on the process is exactly the same. Even more: if the data values meet the equilibrium (they do most of the time), computing Q will not change its value, and the process will be identical even in its figures.

One can even (with no more proof) have the reverse intuition: by concentrating the process on a limited number of variables, one might increase their individual role.

In addition, using an automatic algorithm transfers reorganization control from a logical economist (who has probably chosen a logical choice in terms of causality) to a mathematical blind tool.

Obviously, if the number of loop variables is high, the probability of convergence would be very low if coefficients were chosen randomly. In practice, fortunately, one can associate to the mathematical convergence process a more economic one, the iterative progress to a balanced equilibrium by combining the behaviors of agents (examples can be found in the Keynesian loop, or in the WS-PS wages-prices loop). The probability of success is much improved, and most models converge if they use reasonable assumptions and a reasonable starting point⁸⁶.

Let us illustrate this point by simplifying to the extreme our usual model, taking only into account GDP (Q), private demand (C) and government demand (g). Our model does not consider external trade.

$$(1) C = a Q$$

$$(2) Q = C + g$$

and $0 < a < 1$

With the Gauss-Seidel algorithm, this model will always converge to the solution Q^* , as for the unique loop variable (Q):

⁸⁵ In simpler words : the highest modulus found in the set of eigenvalues.

⁸⁶ Not to converge for improbable assumption values should be considered as a good point.

$$Q^k - Q^* = a \cdot (Q^{k-1} - Q^*)$$

and $|a| < 1$ ⁸⁷

Let us now invert causalities:

$$(1) C = Q - g$$

$$(2) Q = C / a$$

This time the model diverges:

$$Q^k - Q^* = 1/a \cdot (Q^{k-1} - Q^*)$$

and $|1/a| > 1$

but this is normal, as to a convergent economic framework (the Keynesian multiplier) we have substituted a framework where, to simplify:

- Households consume what remains of supply once exogenous demand is satisfied.
- Firms seek to maintain supply in a constant ratio (higher than 1) with consumption.

If such a framework does not converge mathematically, it the same logically: although there exists obviously a solution (the same as higher), an initial error on Q will apply identically to C, then will amplify on Q... And the same instability will be found if we shock public demand (computing the public expenditure multiplier).

One can observe also that with $a > 1$, the first model would diverge (as secondary effects on production would be larger than their source). And the second would converge (as a change in consumption is now reduced on production): the two errors compensate each other.

Formalizing the problem, with the first framework, an initial change in government demand will lead at iteration k, to the cumulated variation:

$$\Delta Q^k = (1 + a + a^2 + \dots + a^k) \cdot \Delta g = (1 - a^{k+1}) / (1 - a) \cdot \Delta g$$

⁸⁷ With $a \geq 1$ the model would diverge, which could be expected from an economic point of view.

which converges to

$$\Delta Q^k = 1/(1-a) \cdot \Delta g$$

if $a < 1$

and with the second

$$\Delta Q^k = (1 + 1/a + (1/a)^2 + \dots + (1/a^k)) \cdot \Delta g = (1 - (1/a)^{k+1}) / (1 - (1/a)) \Delta g$$

which converges only if $a > 1$

The convergence of one model is equivalent to the divergence of the other.

7.1.6.2 Accelerating Gauss-Seidel: damping factors

An efficient technique to accelerate convergence, or to force naturally diverging models to reach a solution, is to introduce one or more «damping factors». This technique is based on the following elementary observation, inspired by the above.

If y is solution of

$$y = f(y)$$

then it will be also the solution of

$$y = D \cdot f(y) + (I - D) \cdot y$$

where y represents the vector of endogenous and D a diagonal matrix.

It goes similarly for g :

if y_b is a solution of

$$y_b = g_b(y_b)$$

then it will be also of:

$$y_b = D \cdot g_b(y_b) + (I - D) \cdot y_b$$

But the new formulation modifies the Gauss - Seidel solving process, by introducing inertia on starting values y_b .

In practice that will mean initializing the iteration k, not by:

$$y_b^{k-1}$$

but by

$$D_b \cdot y_b^{k-1} + (I - D_b) \cdot y_b^{k-2}$$

The change of starting values plays a role only if variables are used before they are computed during the iteration. Therefore damping factors will be applied to loop variables only.

One shows easily that by applying damping factors, we are moving towards a Newton type method, at least in simple cases. Indeed, in a model with a single loop variable, the introduction of damping factors means using:

$$y_b^k = D g_b(y_b^{k-1}) + (I - D) y_b^{k-1}$$

or

$$y_b^k - y_b^{k-1} = D (g_b(y_b^{k-1}) - y_b^{k-1})$$

This is equivalent to the Newton method if one chooses

$$D = (I - \hat{\partial}_{g_b} / \hat{\partial} y_b)_{y_b=y_b^{k-1}}^{-1}$$

or the inverse of the Jacobian of:

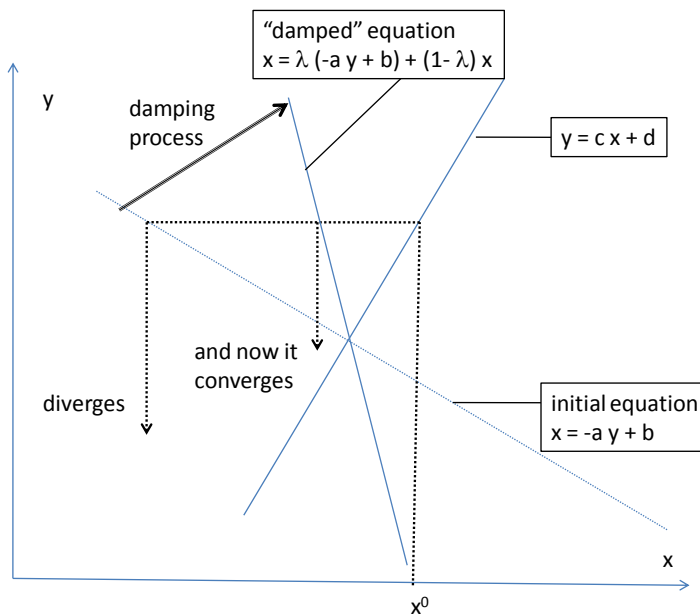
$$y_b - g_b(y_b) = 0$$

This method, more or less efficient following the stability of the Jacobian, will give the solution in a single iteration in the case of a linear model.

Let us complete our presentation by a graphic example.

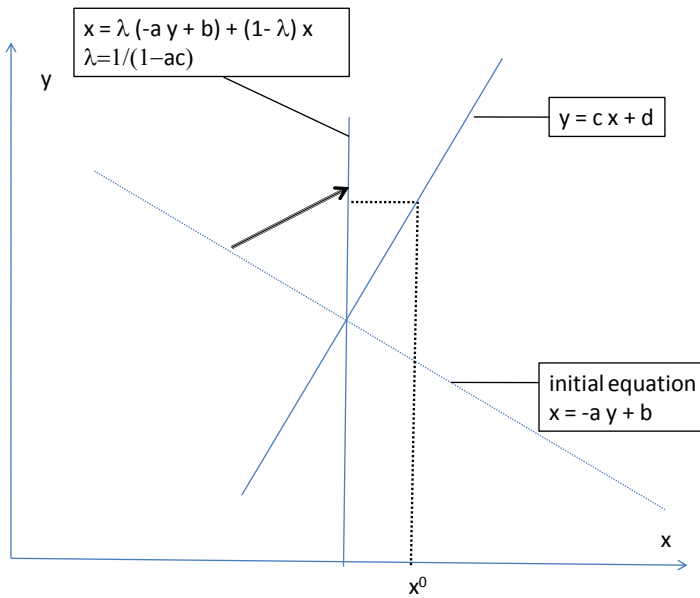
We will use the simple formal model

- (1) $y = c x + d$
- (2) $x = -a y + b$



The basic process clearly diverges, but if we apply a convenient damping factor (we have chosen a value close to 0.5) we can make it converge. What we have done is simply replacing one of the equations ($y = -ax + b$) by another which associates to a given value of y a combination, with constant weights, of the values of x associated to the same y by the two equations. The solution is not changed, but the convergence process is modified.

One could have reached the exact value, by making the new “curve” vertical



Applying damping factors is not however so simple.

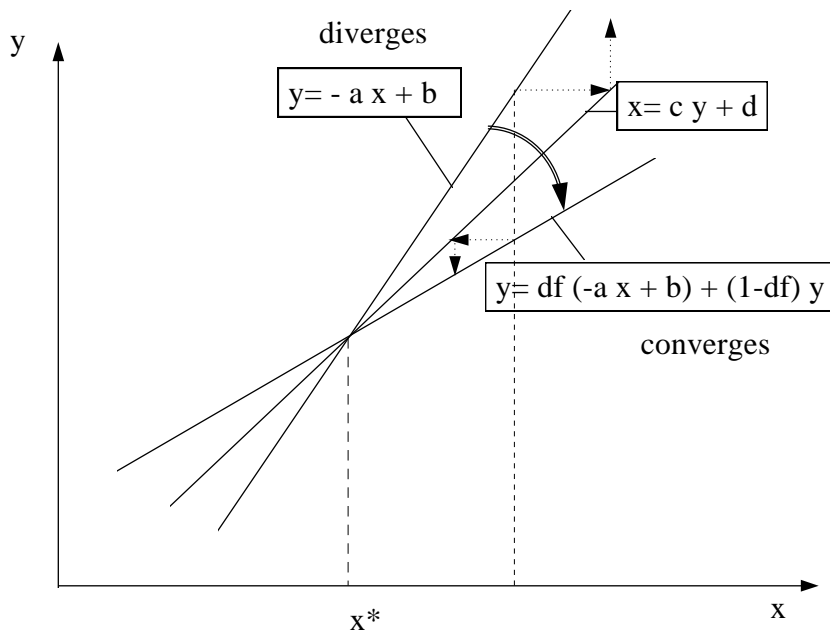
- It can be done only through trial and error, excepts if one solves the problem described above, which is actually more complex than solving the model itself by the Newton method⁸⁸.
- It is only truly efficient for values between 0 and 1 (alternate convergence). If the two curves have slopes with the same orientation, making convergence monotonous, the method becomes more difficult to handle, as an efficient damping factor must now be chosen outside the interval (although it can work, as the following diagram shows).

⁸⁸ One can also concentrate on one loop variable, considered as the most important, and measure its changes for three consecutive iterations. A very rough approximation of the convergence factor can be obtained by:

$$a = (x^k - x^{k-1}) / (x^{k-1} - x^{k-2})$$

and one can use as a damping factor the value:

$$\lambda = 1 / (1 - a)$$



And when the size of the model grows, the searching process becomes rapidly complex.

7.1.6.2.1 More complex models

If the equivalence to Newton cannot be extended to the case of several loop variables, it is only because the method restrains D to a diagonal matrix. One could however consider:

a Looking for the diagonal matrix "closest" to the Jacobian, according to a given measure.

B Decomposing the vector y on the eigenvector basis of the Jacobian, which would allow to diagonalize the process.

$$y^k - y^{k-1} \approx (\partial g / \partial y)_{y=y^{k-1}} (y^{k-1} - y^{k-2}) = V D V^{-1} (y^{k-1} - y^{k-2})$$

$$V^{-1}(y^k - y^{k-1}) \approx D V^{-1} (y^{k-1} - y^{k-2})$$

with D diagonal matrix of the eigenvalues of the Jacobian.

A relative stability of the Jacobian (therefore a good linearity, or an initialization value close to the solution) is then necessary⁸⁹.

⁸⁹ Experience shows that the Jacobian of the linear differentiation is much less stable than the relative one:

- Let us see now how the above considerations influence the convergence of the Newton algorithm.

The usefulness of loop variables, and the minimization of their number, is now indisputable, if one decides to compute the Jacobian by finite differences. Indeed one just has to know starting values for y_b to compute $g(y_b)$ and the Jacobian $\partial g_b / \partial y_b$ (this last by Gauss-Seidel iterations),

- Only the part $\partial g / \partial y_b$ is not identically null.
- And only the part $\partial g_b / \partial y_b$ is going to affect the calculation of the new value of y_b .

We just have to apply the Newton formula to the loop variables alone. Once again, only the values of loop variables of the previous iteration (and their influence on the Jacobian of g) will play a role in the iterative process.

This has obviously the advantage of limiting computations, for the Jacobian itself (calculation of the only non-zero columns) as well as for its inversion.

The usefulness of taking into account loop variables looks clear. The reality is less so:

- The Jacobian matrix associated to the initial formulation f is very sparse (the average number of same period explanatory variables per equation does not grow too fast with the size of the model, and stays generally in the order of 3 or 4). This means the Newton method applied to f can use an inversion algorithm specific to sparse matrices, appreciably more effective. The formulation in g concentrates the process of determination on some variables, but the implicit formulation complexity produces a very dense Jacobian matrix, which can be inverted only by a general algorithm.
- As to the calculation of the matrix itself, it will be obviously more expensive for f than for g , if one uses finite differences. But the Jacobian matrix associated to the f vector can also be determined analytically: then one will just have to compute independently each non zero element of the Jacobian, according to a formula generally simpler than the initial equations⁹⁰. If the average number of instantaneous variables per equation is lower than the number of loop variables, the global cost will decrease. This reasoning is however only valid if one makes abstraction of the (important) cost of the initial formal derivation⁹¹, which supposes also a certain stability of the formulation of the model.

As for the comparison with Gauss-Seidel, one can expect a lower number of iterations, but each of them will be clearly much more expensive.

$$D = (I - \partial g_b / \partial y_b)^{-1}_{y_b = y_b^{k-1}}$$

⁹⁰ For instance when derivation is applied to a sum.

⁹¹ EViews is able to compute the analytical derivatives. However, it does not provide the formulas.

Let us see how the Newton process converges.

For this let us develop to the second order the derivation formula:

$$y^{k+1} - g(y^{k+1}) - (y^k - g(y^k)) = (I - \partial g / \partial y)_{y=y^k} (y^{k+1} - y^k) - {}^T(y^{k+1} - y^k)(\partial^2 g / \partial y^2)_{y=y^k} (y^{k+1} - y^k)$$

plus a third order term.

Using the linearization to determine y^{k+1} eliminates the central terms, leaving:

$$y^{k+1} - g(y^{k+1}) = - {}^T(y^{k+1} - y^k)(\partial^2 g / \partial y^2)_{y=y^k} (y^{k+1} - y^k) + e^3(y^{k+1} - y^k)$$

and as $y^{k+1} - y^k$ is a linear function of $y^k - g(y^k)$ (again by construction), the error $y^{k+1} - g(y^{k+1})$ is a quadratic function of $y^k - g(y^k)$.

Similarly, the distance of y to the solution y^* is a quadratic function of the distance at the previous iteration, as the derivation to the first order has given:

$$(y^{k+1} - g(y^{k+1})) - (y^* - g(y^*)) [= 0] = (I - \partial g / \partial y)_{y=y^k} (y^{k+1} - y^*)$$

The convergence of the algorithm is not therefore uniform: in fact it is going to accelerate with the number of iterations, as the linearized form gets nearer to the derivative at the solution. One will say that it is quadratic in the case of a formal derivation, but only supralinear if one uses finite differences, because one does not compute the exact derivative. The Newton algorithm should be the more advisable as the precision requested from the solution gets higher.

7.1.6.3 Comparison of the relative efficiencies

Let us consider first the ability to find the solution. The convergence of the Gauss-Seidel algorithm appears quite doubtful, if it based on several iterative interacting processes (the number of loop variables) all of which must be converging. In other terms, a matrix must have all its eigenvalues lower than unity in modulus. It is however more frequent than one can expect. As we have seen earlier, the numerical process is generally linked to a set of economic causalities, determining a normally convergent process. Actually, if this process does not converge, the model itself should probably be tested for economic consistency.

The Newton algorithm on the contrary puts no condition on the numerical intensity of relationships, but on its variability (in practice the stability of the Jacobian and in a lesser measure the Hessian). This condition is generally verified in the case of econometric models. In particular, the convergence will be immediate for a linear model showing a unique solution. The Newton algorithm will then be advisable for models presenting convergence difficulties, or for model

builders wanting to free themselves from this concern, which remains always present and sometimes crucial (for instance if they frequently face deadline problems).

On the other hand, considering the speed of convergence, the choice is much less obvious. If the Newton method converges in less iterations, generally between three and five (remember that convergence accelerates with iterations), each iteration is far more expensive. In fact this supplement can be measured, in the case of a derivation using finite differences: each iteration takes as many Gauss-Seidel iterations as the model contains loop variables (plus one), as well as a matrix inversion (growing with the dimension as a third degree polynomial). One can therefore minor the total cost, whatever the method, by a number of Gauss-Seidel iterations.

In practice one observes that the number of iterations necessary for convergence, whatever the method employed, grows slowly enough with the size of the model. On the other hand the number of loop variables is more or less proportional to it.

We shall give now the time taken by each of the methods (in case of success):

- Gauss – Seidel
- Newton with analytical derivations
- Newton with numerical approximation
- Broyden with analytical derivations
- Broyden with numerical approximation.

The test was conducted on a Toshiba Qosmio F60-10J, using an Intel I5 chip and 4 GB memory.

For 10 models of growing size, all built or contributed by the author in the course of cooperation projects and seminars:

- France : a very small French model, resembling the one we are using here
- France cf : a small French model with complementary factors
- France cd : a small French model with Cobb-Douglas. Both these models will be presented later.
- Vietnam : a small Vietnamese model with one product
- Algeria 1 : a small Algerian model with one product (+oil in some cases)
- Algeria 5 : a five product Algerian model
- Algeria QAM : a 19 product quasi-accounting Algerian model
- Vietnam : a 3 product Vietnamese model
- World : a 12 country world model
- China : a 3 product, 4 region model

The first three models are available in the files accompanying the book.

With the following characteristics:

- Prod : number of products
- Coun : number of products or regions
- Per : number of periods
- Equ : number of equations
- Block size : size of the largest non – recursive block
- Feedback : number of feedback (or loop) variables

Countries	prod	coun	per	equ	Block s	Feedba	Gauss-Seid	Newton A	Newton N	Broyden A	Broyden N
France	1	1	q	18	13	2	0,002	failed	failed	0,003	0,004
France cf	1	1	q	87	47	2	0,010	0,018	0,030	0,014	0,014
France cd	1	1	q	88	51	3	0,020	0,025	failed	0,017	0,017
Vietnam	1	1	a	94	55	4	0,030	0,026	0,045	0,019	0,020
Algeria	1	1	a	149	85	7	0,028	0,026	0,036	0,023	0,029
Algeria	5	1	a	630	408	23	failed	0,169	0,237	0,145	0,146
Algeria QAM	19	1	a	1361	922	19	failed	0,312	0,433	0,261	0,234
World	1	12	q	1396	1063	71	0,087	0,084	0,082	0,083	0,083
Vietnam	3	3	a	1858	1260	37	2,446	2,008	2,522	failed	failed
China	3	4	a	2648	1900	45	failed	failed	failed	0,349	0,321

One can observe:

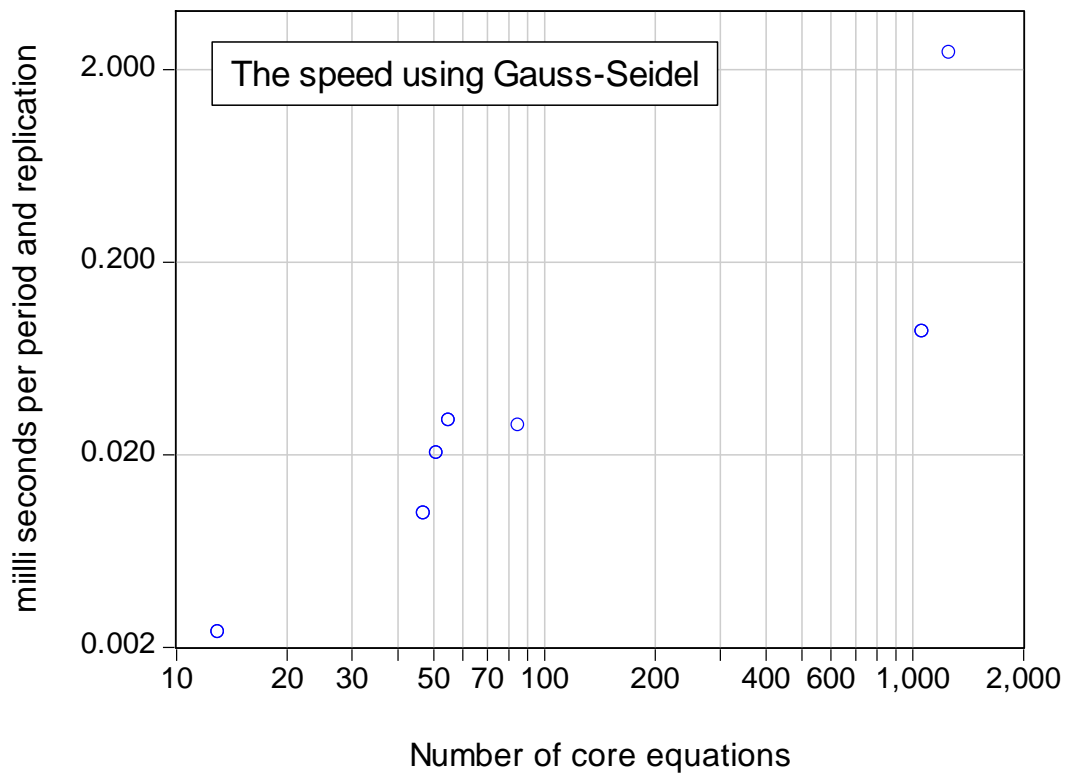
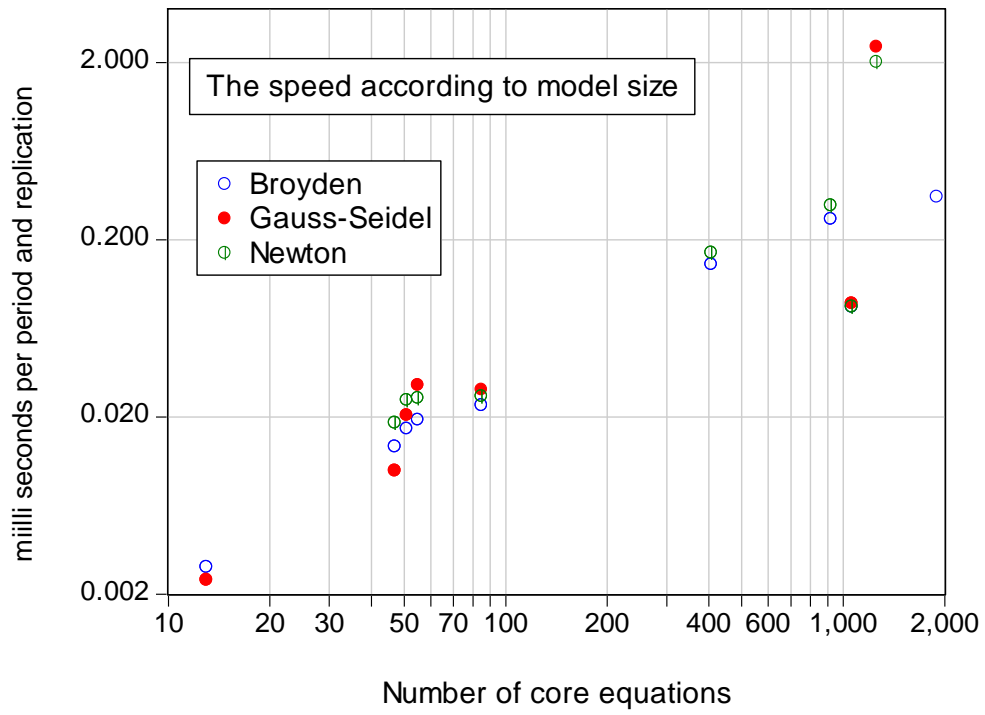
- The bad news is that each method fails at least once, the good that at least one works for any model. If this was true in all cases, convergence would be guaranteed.
- Gauss Seidel works better for small models. This is quite contrary to expectations, as the cost of inverting a matrix of growing size should become more and more expensive as size grows. There might be a cost of entry for Newton.
- Broyden is globally better for large models, for no particular reason. However, it does fail once. But it is the only one working for the Chinese model.
- For the large Vietnamese model, Broyden fails, and Newton is better than Gauss-Seidel.

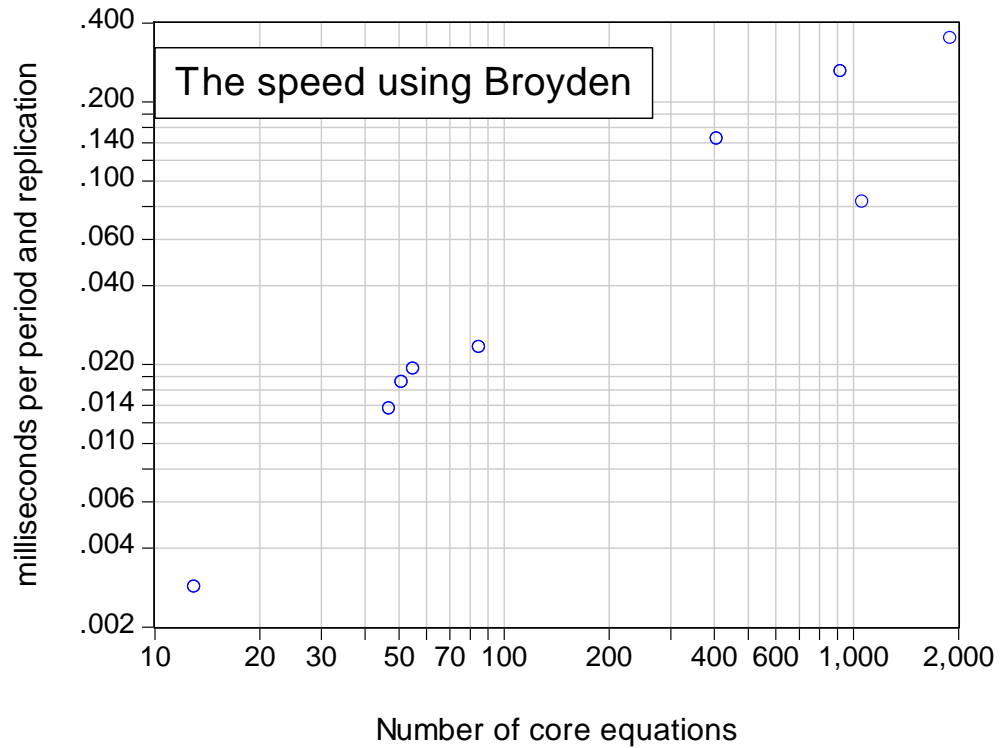
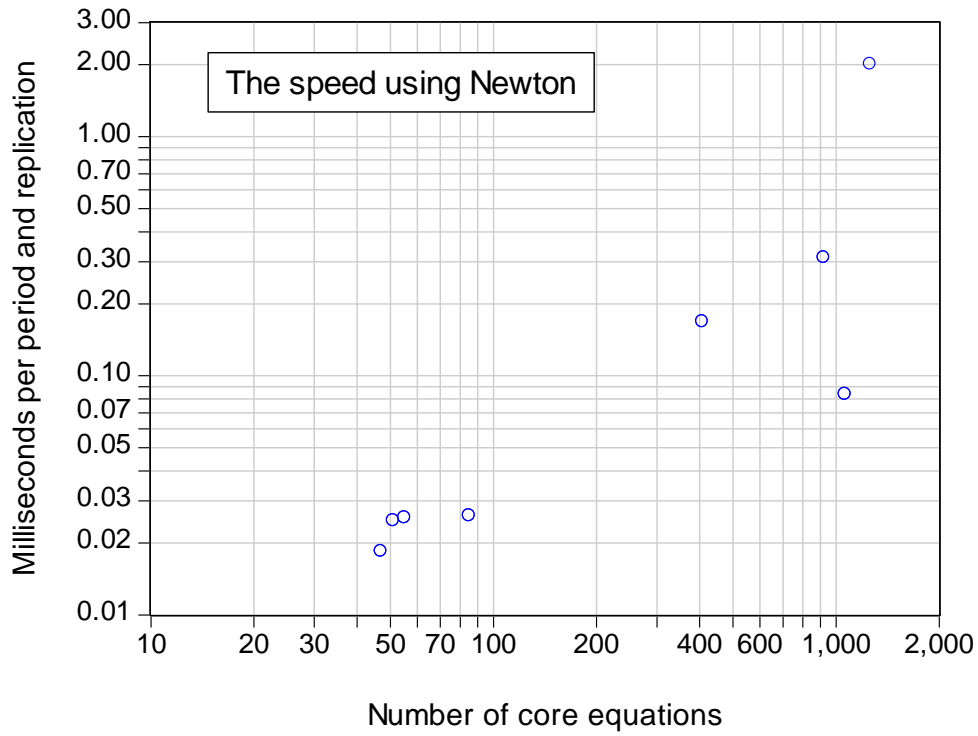
Let us now consider the link between model size and computation cost.

We shall present four graphs, all of them considering the logarithm of both variables:

- A global one for all methods
- Three individual ones dedicated to each method.

They show that the link is globally log-linear, with a slight downward inflexion for the “inversion” methods. As to the elasticity itself, a separate estimation gives a value close to 1, and a global one a value of .99 (with a T-stat of 14.7).



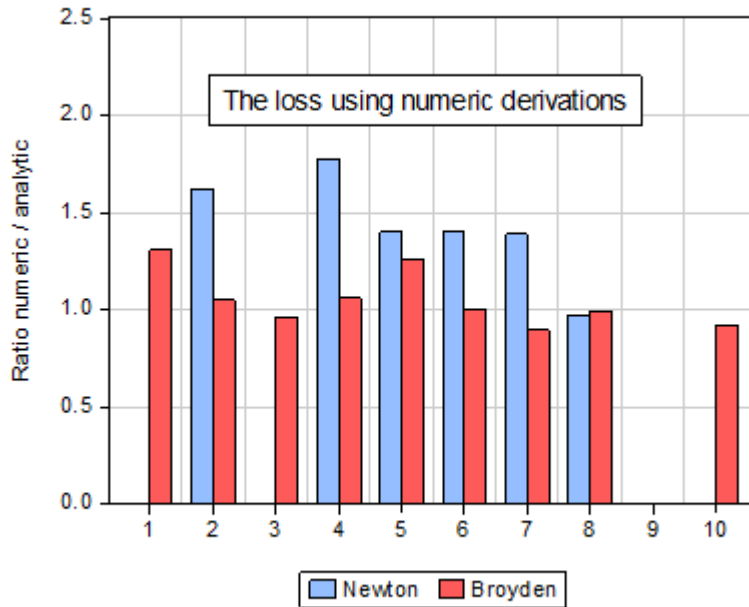


- The two last graphs compare analytical and numerical methods, for both related algorithms.

- The diagnosis is complemented by an histogram presenting the ratio of costs (when both options led to convergence).

They show clearly :

- That establishing analytical derivatives is quite efficient for the Newton method, with an average gain of about 40%.
- That for Broyden (where the derivatives are only computed once) there is no gain, but no loss either.



Countries	Newton analytic	Newton numerical	Newton ratio	Broyden analytic	Broyden numerical	Broyden ratio
France	failed	failed	failed	0,00	0,00	1,31
France cf	0,02	0,03	1,62	0,01	0,01	1,05
France cd	0,02	failed	failed	0,02	0,02	0,96
Vietnam	0,03	0,05	1,78	0,02	0,02	1,06
Algeria	0,03	0,04	1,40	0,02	0,03	1,26
Algeria	0,17	0,24	1,40	0,14	0,15	1,01
Algeria QAM	0,31	0,43	1,39	0,26	0,23	0,89
World	0,08	0,08	0,97	0,08	0,08	0,99
Vietnam	2,01	2,52	1,26	failed	failed	failed
China	failed	failed	failed	0,35	0,32	0,92

7.1.7 SOLVING THE MODEL: BASIC EIEWS TECHNIQUES

The previous program has made model data and equations consistent. The present one will solve it.

7.1.7.1 The syntax

Again, we shall SOLVE the model, but this time the options are more numerous.

Solving the model will generally be done in a program. However, one can use the menus, first by accessing the model item, then using the "solve" item, which gives access to all the features we are going to present.

In a program, these options can be set before any computation, using the SOLVEOPT statement, or at simulation time (SOLVE).

The syntax is the following:

```
Model-name.solveopt(options)
```

Or

```
Model-name.solve(options)
```

Where the options are (taken from the EViews Help file):

- m= (default=5000)** Maximum number of iterations for solution (maximum 100,000).
- c= (default =1e-8)** Convergence criterion. Based upon the maximum change in any of the endogenous variables in the model. You may set a number between 1e-15 and 0.01.
- s=arg** Solution type: "d" (deterministic, default), "m" (stochastic - collect means only), "s" (stochastic - collect means and s.d.), "b" (stochastic - collect means and confidence bounds), "a" (stochastic - collect all; means, s.d. and confidence bounds).
- d=arg** Model solution dynamics: "d" (dynamic solution, default), "s" (static solution), "f" (fitted values - single equation solution). Forecasting calls for a dynamic solution, which is the default option. There are few reasons to choose another (except d=f as above).
- n=arg** NA behavior: "n" (stop on "NA" values), "t" (do not stop when encountering "NA" values). Only applies to deterministic solution; EViews will always stop on "NA" values in stochastic solution.
- a=arg** Alternate scenario solution: "t" (solve both active and alternate scenario and collect deviations for stochastic), "f" (default, solve only the active scenario).
- o=arg** Solution method: "g" (Gauss-Seidel), "n" (Newton), "b" (Broyden).

i=arg For the initialization values: “a” (actual), “p” (previous period solution). The default is “a” if the values are available, “p” otherwise. In forecasts, “p” should be applied, but one can initialize the endogenous which gives the two options (with “a” as the default).

For stochastic simulations:

r=integer (default=1000) Number of stochastic repetitions (used with stochastic "s=" options).

b=number (default=.95) Size of stochastic confidence intervals (used with stochastic "s=" options).

7.1.7.1.1 The suffix

In addition to the above options, EViews allows the user to define the name of solution variables.

Of course, one should not consider using the actual names as such. This will destroy the information on the original values. However, the connection must be as easy as possible. The obvious solution is to use the original names, but modify them by some addition: a given prefix or suffix.

This will call for a specific statement:

```
model_name.assign @all suffix
```

This statement will add to the text of the model *model_name* the statement:

```
assign @all suffix
```

where *suffix* is a string.

For instance,

```
_fra_1.assign @all _b
```

will give to the variables obtained by the next model solution a name composed of the original variable followed by the suffix “_b”. In this case, the solution for Q will be called Q_B.

7.1.7.2 Problem processing techniques

Like all packages of this type, EViews does not always reach a solution. We are going to provide here a typology of techniques one can use for solving convergence problems, focusing on the ones available under EViews, and the way to apply them.

In all cases one must be certain that the residual check shows no error in the first place. This is essential, and very cheap (at least in the observation of errors, and the identification of their location).

7.1.7.2.1 The tools provided by EViews

In addition to changing model specifications, EViews provides a number of specific tools. Some of these can only be specified through menus, but they remain active if one runs a program in the same session.

- Displaying the number of iterations necessary for convergence (only useful of course if the model converges on part of the sample). But if the model starts diverging after a few periods, it is interesting to know if the process is gradual (the number of iterations is growing until breakdown) or sudden (the number is stable until the crash).

This is obtained by

- o double-clicking the model item

- o using:

solve>diagnostics

- o Switching on: Display detailed messages including iteration count by blocks.

A window will open displaying the number of iterations needed for solving each block (several messages per period if the model contains more than one non-recursive block).

This window will remain open after solving, and one can scroll back and forth, to the first message if needed.

```
Model: FRA_1
Date: 11/07/12 Time: 14:41
Sample: 2001S1 2001S2
Solve Options:
    Dynamic-Deterministic Simulation
    Solver: Gauss-Seidel
    Max iterations = 5000, Convergence = 1e-06

Scenario: Scenario 1
Solve begin 14:41:07
    2001S1 Block 1 - 3 eqns          Solved (recursive block)
    2001S1 Block 2 - 14 eqns        Convergence after 14 iterations
    2001S1 Block 3 - 3 eqns          Solved (recursive block)
    2001S2 Block 1 - 3 eqns          Solved (recursive block)
    2001S2 Block 2 - 14 eqns        Convergence after 14 iterations
    2001S2 Block 3 - 3 eqns          Solved (recursive block)
Solve complete 14:41:07
```

- Displaying the values obtained at each iteration for a set of variables.

This is obtained by

- o double-clicking the model item
- o using:

```
solve>diagnostics
```

- o Introducing the names of the variables in the window.

Alternatively, the same can be obtained in a program, using:

```
model_name.trace list_of_variables.
```

This can be useful if you select a given set of important elements, which will then be available at all times after any simulation.

- o After solving, using:

View>trace output

A window will open with the values of the variables at each iteration, for each period (one column of values per variable).

If the variable belongs to a recursive block, only one value will be displayed. In our case, the unique values for CAP and X will precede the iterative process description, then the unique K value.

The first value (iteration 0) is the initial one.

The columns follow the order used in the solving process, not the one in which the list is specified. This is quite useful for following the Gauss-Seidel process.

Iteration	CAP	X	UR	Q	I	FD	M	K
0	1477674.	411669.1						
0			0.967825	1430727.	181579.8	1388969.	381013.8	
1			0.968230	1419625.	179655.5	1344976.	370204.8	
2			0.960716	1386440.	177326.4	1339545.	368264.6	
3			0.940504	1383299.	176118.0	1338384.	366440.3	
4			0.936155	1383611.	175753.1	1337710.	365701.2	
5			0.936344	1383678.	175766.0	1337723.	365717.2	
6			0.936389	1383675.	175768.5	1337726.	365721.7	
7			0.936388	1383673.	175768.2	1337725.	365721.5	
8			0.936386	1383673.	175768.1	1337725.	365721.3	
9			0.936386	1383673.	175768.1	1337725.	365721.3	

0								2085298.
0	1495279.	421079.3						
0			0.968590	1453147.	183168.3	1409358.	386757.6	
1			0.971823	1443679.	178010.0	1350618.	372634.4	
2			0.965492	1399063.	175065.9	1343571.	370409.1	
3			0.938637	1394724.	173433.1	1342002.	367972.8	
4			0.932781	1395107.	172935.2	1341078.	366972.9	
5			0.933008	1395185.	172950.4	1341093.	366991.3	
6			0.933060	1395181.	172953.2	1341095.	366996.5	
7			0.933057	1395178.	172952.9	1341095.	366996.1	
8			0.933055	1395178.	172952.8	1341095.	366995.9	
9			0.933055	1395178.	172952.8	1341095.	366995.9	
0								2105233.

But the interpretation of the evolutions as a column of figures is often difficult, in particular if one tries to evaluate the presence and nature of cycles. It is better to transform these figures into graphs. This can be done in two ways:

- Selecting with the mouse the requested zone (for instance the values for a given period) and copying it into a graphic software such as Excel.

- The same can be done inside EViews but it is a little more complex. One must:
 - o Select with the mouse the requested zone
 - o Create variables associated with the displayed elements (with different names, for instance by inserting a common prefix).
 - o Display the variables as a table filled with “na” values..
 - o Copy the selection into the sheet.
 - o Apply View>Graph.

- Excluding equations from the solving process.

This is very easy. Using **menus**, one has to

- o access the model,
- o use:

Basic options>Edit scenario options

- o In the “exclude series” block, specify to series to exogenize.

Using a **program**, one needs the statement:

```
model-name.EXCLUDE list-of-series
```

The corresponding equation will not be applied, and the variable will retain its initial value.

Note: the list will remain active for all subsequent solutions produced under the same scenario, until it is modified using a different list. Reverting to the full model calls for a statement with a blank list.

One can consider eliminating a single variable, like the change in inventories (if it seems to introduce an explosive process), or a full set, like all the deflators in the model (if the price system seems to explode, or to disturb the computation of the real sector).

However, one must be aware that if an identity variable is excluded, the identity is no longer enforced. Balancing variables should never be excluded. For instance, excluding Q makes supply no longer equal to demand (and the model recursive...).

7.1.7.2.2 Other tools

Some tools can be considered by the user:

7.1.7.2.2.1 USING DAMPING FACTORS

This has been described earlier. Technically, you will have to rewrite the equation. For instance

$$(4) Q + M = FD + X$$

will become:

$$(4) Q = \text{rel_q} * (FD + X - M) + (1 - \text{rel_q}) * Q$$

where damping is enforced by making rel_q different from 1 (lower in general).

It might be useful to set the factor as a series, even if it will be constant over time. To apply a change to a scalar, the model needs to be recompiled (by UPDATE).

Also, this method is a little different from the one described earlier. Application of the factors is not done at the end of the iteration, but at the time the variable is computed. If the variable is used later in the ordering, this affects the convergence process.

7.1.7.2.2 REORDERING THE MODEL

Using Gauss-Seidel, reordering will change the solving process. Not for Newton as the Jacobian is based on derivation of individual equations.

7.1.7.2.3 CHANGING PARAMETER VALUES

Changing the values of parameters will obviously modify the solving process. In particular, a coefficient which looks too high can be reduced or even eliminated. This could be the case for instance if the intensity for the short term accelerator effect in the investment equation.

This technique can be used to eliminate a specific influence of a variable, leaving unchanged its other roles.

This calls for a good knowledge of the properties of the model.

7.1.7.2.4 CHANGING EQUATION SPECIFICATIONS

A mechanism which seems to create problems can be replaced by another. For instance, the purchasing power parity assumption for the exchange rate can be smoothed or replaced by a fixed rate.

7.1.7.2.3 Applying these tools

Let us now propose a methodology.

1 - The model converges for the whole period: OK

2 – The model does not converge at the first period.

2.1 The maximum number of iterations has been reached (the number after $m=$ in the SOLVE or SOLVEOPT statement, default=5000).

This is not too bad: making the model converge should not be too difficult, most of the time.

First, check which variables have not converged, by selecting a few of them (the loop variables in the case of Gauss-Seidel) and displaying their values.

Technically, this done by:

Accessing the model (double-clicking on the model item in the workfile).

Selecting “Diagnostics”

Putting the names of the variables in the “Trace variables” window.

Solving the model

Selecting View > Trace output.

2.1.1 No variable has converged (or just a few)

In case of a few, they should belong to the prologue.

If **Gauss-Seidel**:

- Check that we are not facing oscillations between two very close values (purely numerical problem due to the precision of computer computations) in which case the chosen threshold is too small.

Otherwise:

- Look for another ordering, checking first the economic logic of the present one.
- Introduce one or several damping factors (computed by trial and error) on the main loops.
- If Newton (rare): check that all starting values are reasonable; search for other starting values.

2.1.2 A single variable (or just a few) does not meet the criterion.

Normally this comes from:

- Pure bad luck: a variable which can change signs, which means that it is computed as a difference (like the change in inventories or the trade balance) takes values much lower than its components, and loses precision digits in the process. The same is true also for growth rates, when the associated element is stable over time.
- A small non-recursive block which does not converge: we shall see that in operational models the state budget belongs to the epilogue and is generally recursive, but it includes often the following loop:

Deficit -> interests paid -> deficit

which might not converge.

Note: in the last case, if the block is small enough, one can transform it into a recursive block by editing the equations.

The situation is frustrating, as we feel we have a naturally converging model, but with EViews we cannot:

- Adapt the criterion to the variable, we could check if it is appropriate, in level and in mode of computation. On balances or growth rates it is preferable to control differences in levels, actually following the normal presentation of the variable: a change in inflation will be presented in points, an increase in government deficit in billions, while for a gain on GDP one will think in percentage points.

- o Adapt the criterion to the period: release the value for the non-converging period and move to the normal one afterwards.

So we are left with the methods of 2.1.1

2.2 The model crashes due to missing data

Obviously, the data should be provided. In the following example, estimation has not been conducted on the full sample, the residual `res_m` has not been extended. The determination of `M` lacks its lagged value.

```

Model: _FRA_1
Date: 11/07/12 Time: 16:15
Sample: 2003S1 2050S2
Solve Options:
  Dynamic-Deterministic Simulation
  Solver: Gauss-Seidel
  Max iterations = 5000, Convergence = 1e-006

Scenario: Scenario 1
Solve begin 16:15:25
  2003S1 NA generated for M
  2003S1 NA generated for Q
  2003S1 NA generated for I
  2003S1 NA generated for LED
  2003S1 NA generated for LE
  2003S1 NA generated for LT
  2003S1 NA generated for RHI
  2003S1 NA generated for IH
  2003S1 NA generated for CO
  2003S1 NA generated for IC
  2003S1 NA generated for FD
  2003S1 NA generated for UR
Solve terminated - Unable to compute due to missing data in "UR = Q / CAP"

```

2.3 The model crashes, as some variables reach an unacceptable value (for example the argument of a logarithm takes a negative value).

EViews will provide the user with a message stating the place the error occurs, like:

```

Log of non positive number in DLOG(LE) = C_LE(1) * DLOG(LED) + C_LE(2) * LOG(LED( - 1) / LE( - 1))
+ C_LE(3) + ec_le
" in "SOLVE(N=T M=10001 C=1E-6 O=G D=D I=A) _FRA_1"

```

First, this does not mean that the equation is at fault: if the employment equation crashes through a negative argument of a logarithm, it means that target employment LED is negative, which can only come from the fact that Q (GDP) is itself negative. As it is computed through:

$$[4] Q + M = FD + X$$

it means that FD or X is too small or negative, or M too high.

Moving back through the causalities is not always simple, however, and sometimes we get no information at all, like in (Newton):

near singular matrix in "SOLVE(N=T M=10001 C=1E-6 O=N D=D I=P) _FRA_1"

In practice, one should first look at the number of iterations performed before the crash.

2.3.1 If it is small (2, 3 or even one) for Gauss-Seidel, or 1 for Newton, there is probably a technical problem, as an apparently acceptable model should not perform so badly.

Sources for technical problems can be:

- Your model is indeed very wrong. For Gauss-Seidel, the eigenvalues are much higher than 1⁹². For Newton, the Jacobian leads in a very wrong direction (maybe because the starting point is unrealistic).
- Your model cannot respond to abnormal values of the endogenous. For instance, you are taking the logarithm of a variable which you expect to be positive, but due to bad model properties it is computed as negative. Investment is a good candidate here, if GDP goes down too much.

$$I/K(-1)=0.825*I(-1)/K(-2)+0.0279*UR+0.152*.25*Q/Q(-4)-0.0525+EC_I$$

- Your model is not consistent with the data. Of course you have performed a residual check, but maybe something has changed since. One should perform it again.
- The other cases apply mostly to simulations outside the sample period:
- The exogenous assumptions have been given abnormal values.

For instance with WD=0 we get:

⁹² For instance, an error correcting mechanism is strong, and its sign is wrong.

Solve terminated - Log of non positive number in "DLOG(X) = @COEF(1) * DLOG(WD) + @COEF(2) + @COEF(3) * RES_X(-1) + EC_X"

- The initial values are abnormal, and the model crashes before it reaches an acceptable region.

We can display the associated variable values. Beyond the erroneous variable itself, one will be interested in each of its determinants, as the logical source of the problem will be often located far upstream of the diagnosed equation. For example a negative number of the unemployment level (used in a logarithm) might come from overestimating demand, therefore production and the employment level itself (which becomes higher than the work force).

To solve the problem, it will be necessary in this case to retrace the causal chaining

For instance if M has been initialized at 10^{15} , we get:

```
Solve begin 18:07:52
  2003S1      Block 1 - 3 eqns      Solved (recursive block)
  2003S1      Block 2 - 13 eqns     No convergence after 3 iterations
Solve terminated - Log of non positive number in "DLOG(LE) = @COEF(1) * DLOG(LED) + @COEF(2) *
LOG(LED(-1) / LE(-1)) + @COEF(3) ..."
```

And we can follow the evolution of variables in the "Trace output" display:

Date	Block	Iteration	X	UR	Q	LED	LE	FD	M
2003S1	0	0	411669.1						
2003S1	1	0		0.964223	-2.50E+14	-8.11E+14	18608119	1.23E+12	1.00E+15
2003S1	1	1		-1.69E+08	-9.99E+14	-1.30E+16	18608119	4.42E+12	1.00E+15
2003S1	1	2		-6.76E+08	-9.96E+14	-1.29E+16	18608119		

For Gauss-Seidel, an efficient method (not provided by EViews) consists in changing the algorithm to Ritz-Jordan, which takes into account only values of the previous iteration, eliminating interferences and allowing to improve individualizing divergent processes. At the beginning of a process, this method individualizes immediately the faulty equation (less efficiently than a residual check, however).

2.3.2 If the number of iterations before the crash is higher, two situations have to be considered:

A The model has no solution.

B The model has a solution which it is unable to reach.

With two options for the latter, in the case of Gauss-Seidel:

B1 You have ordered the equations in the same way as the equilibrating process it describes.

B2 Your ordering is not consistent with this process.

Unfortunately, there is no way to separate the cases. The first advice is simple and quite cheap to apply: try the alternate algorithm!

There is a good chance that:

- Newton does not converge due to colinearities in the Jacobian matrix, or to initial values too far from the solution, making the linearization inefficient. This does not affect Gauss-Seidel too much.
- The Jacobian matrix associated with Gauss-Seidel (a different one) contains eigenvalues with modulus higher than 1, which is not a problem for Newton.

This could actually solve your problem. However, in case B1 you should be concerned with the stability of the equilibrium you model describes.

If this does not work, the solution is partially different for the two algorithms:

- For both algorithms, you can drop one or several equations, and try solving the model again. The choice depends on the variables which seem to lead the divergence: either the ones which actually take the worst values, or the elements which influence them. By forcing a variable to take a given value, one can often measure its contribution to the problem (especially if the variable goes through unacceptable values) by observing if the problem disappears. This method has however the same limit as the others: the exogeneity of a variable destroys all influences forwarded by it, and making the error disappear does not mean we have exogenized the faulty equation.
- You can also change the values of parameters, or drop an individual influence by setting its coefficient to zero. This is different from dropping the variable altogether: in our model, you can drop the role of the rate of use in exports but leave it in imports. By applying variations to some strategic coefficients, chosen either because they play an important role in the model (in the case of a general absence of convergence), or because they appear linked to the diagnosed error (in case of a local error), one might establish a connection between these coefficients and the intensity of the problem. However this method, once again, does not locate necessarily the origin of the problem, which can come from far upstream. And it remains to determine how to use the information, if one excludes setting arbitrarily the coefficient to the value ensuring convergence
- You can try alternate (probably simpler⁹³) formulations. For instance, you could replace a complex CES function using the relative cost of factors in a sophisticated way, by a very simple complementary factors function. Of course, you can start by exogenizing the relative cost of factors.

⁹³ If the model you want to produce is quite complex, it might be a good idea to start with a simpler version, and introduce layers of complexity one by one.

If the argument of a logarithm takes a negative value, you can introduce a test avoiding this situation (for example replacing $\text{Log}(x)$ by $\text{Log}(\max(x, a))$, with $a > 0$). The system might actually converge to an acceptable value ($x > 0$)⁹⁴, but reaching a solution with $x < 0$ will give you information anyway⁹⁵.

- For Newton, try other starting points. You should try to guess as well as possible the domain in which you solution will reside, and initialize the variables within it. This is easy on the past.
- Check the derivatives of your equations. You can do it by sight, or by computations (a little harder, but EViews provides the elementary tools). If you start from quasi-solution values but the gradient is wrong, you can go in the wrong direction, perhaps by a wide margin.
- For Gauss-Seidel, try damping factors. Technically, you will have to rewrite the equation.

If you have tried to do it, but things are actually worse, there is a good chance that your model is not naturally converging for economic reasons, but it might also have no solution.

But in any event the change must not reduce the theoretical validity of the model: one does not try to enforce a solution thanks to an ad hoc formulation, but rather to correct an inconsistency.

Finally, you can change the simulation period (all cases). It is interesting to know if the problem is general, or appears only for some periods. Examining the particular aspects of the faulty periods (slow growth, strong inflation, important over employment...) can give useful information on the origins of the error. Using it can however be difficult. One specific case is the role of external trade: in forecasts, too fast an expansion compared to local production and demand can drive the multiplier to abnormal values, which will disappear if we simulate the model on the first periods available.

3 The model converges for a while, then crashes.

This situation has good and bad points:

- The model has no pure technical flaws, except if you have introduced an error in the future assumptions, or used a formulation which becomes wrong as it changes over time (a dummy with a wrong coefficient, a condition which can only happen in the future...).
- But these non-technical flaws are harder to solve. Your model might describe a sequence of equilibrating processes, which converge initially, but change through dynamics which makes them diverging after a while. For instance, the expanding effect in the Keynesian multiplier might get higher than one, making the process explosive (the same with the price-wage loop). Or the error-correction process is actually error-increasing.

⁹⁴ A rare situation in which the studying the error actually solves the problem.

⁹⁵ When things go wrong, the potential for solving the problem is of course quite different if you have a (bad) solution or nothings at all. This means it is generally essential to manage to reach a solution, even in a non-acceptable framework.

Let us take the example of a car: if it does not start, you might have forgotten to put gas, or the battery might be flat. This can be solved easily. If it starts but does not run fast enough, or veers to the left systematically, the problem might be quite worse (and more difficult to locate).

- There is also less chance that the model is really wrong. And you can always start again under new assumptions. The motor of the car might explode, at a high replacement cost. For a model, you can recreate a new version for free, and repeat the process indefinitely.

But the best point is:

We get more information

We can:

- Observe the way the model is diverging, and which variables (or sets) appear to lead the way. For instance, variables at constant prices might follow an acceptable path, and prices explode. Or unemployment might explode while employment is acceptable, driving wages and the partition of value added between workers and firms....

We can see if the divergence is monotonous or cyclical, regular or erratic, sudden or progressive.

- We can shock assumptions and look at the consequences. They can be acceptable in the beginning, which means that the problem lies in the dynamics. They can grow wrong in a regular way, or explode suddenly.

For instance, a growing call for labor in the productive process can make unemployment become negative in time, while used as a logarithm in the wage equation⁹⁶. Or as we have seen earlier, an accidentally very low level of investment can make the associated equation explode suddenly. Or exports growing much slower than imports can make GDP become negative.

- All this can be combined with the exclusion of equations, to show the propagation of a disease introduced by a specific mechanism.

We can guess already that interpreting this information will be easier with a longer simulation, starting from on a regular base trajectory. In other terms, a forecast, or rather a simulation over the future.

- Finally, the computation of eigenvalues of $(I - \partial g / \partial y)^{-1}$ when one uses the Gauss - Seidel can help to understand the convergence process, and evidence the mathematical characteristics of a possible divergence (intensity and number of divergent processes, presence of cycles). Processing this information can be done by a gradual elimination of loop influences, analogous to the one recommended by Deleau and Malgrange (1978) to study dynamic model properties. In the best of cases, an association will be obtained between each

⁹⁶ Replacing the logarithm by a level does not change the situation, as negative unemployment is of course unacceptable.

divergent process and a single loop variable (or two for a couple of conjugate eigenvalues, producing a cyclical process).

The interpretation and solution of a convergence problem are not simple, essentially because of interferences due to the interactive nature of the system. It will be often necessary to use several methods, and to repeat each of them several times, which will not dispense with a thorough reflection on the logical structure of the model. It is therefore essential for the model builder to know his model well, which will be all the more difficult as its size grows.

7.2 A FIRST VALIDATION

Although the convergence of a model is not without connection to its theoretical validity, it is not enough to ensure it. For this, a certain number of more or less exact methods must be applied.

7.2.1 EX POST SIMULATIONS

First we want to know if the model is able to perform realistic forecasts. For this, two techniques are available: ex-post simulations (or ex-post forecasts) over the past, and stochastic simulations, generally over the future. For now, we shall limit ourselves to the first case.

Our present goal is to know if the model will produce reliable forecasts. For this we have a problem: we do not know the future (if we did we would not need the model anyway), so we have to rely on the past, which provides limited information. We are not in the situation of a scientist who can produce relations as exact as he wants through the required number of experiments.

So we shall perform simulations on the past, but put the model in a situation as close as possible as the one it will meet in the future. And there is one element it certainly will not have available: the residuals in the estimated equations, in other terms the unexplained part of the behaviors.

What he knows is their most probable values, the average of their distribution: zero.

So we shall simulate the model on the past, with zero residuals. This test is called "ex post simulations".

They are undertaken on the total set of periods used for the estimation of coefficients: it may seem that the closer simulation results will be to the observed values, the better the model should be.

One will then compare model results to reality, through tables of residuals, graphs showing the two curves, or statistics.

In practice one shall perform:

- Static simulations, which use historical values for lagged variables. We will get in fact a sequence of independent simulations with a one-period horizon.
- Dynamic simulations, which will use the simulation results of previous periods (except of course for the initial ones).

In general this last criterion will be favored, as it is more representative of the future use of the model. One can also assume that it is more difficult to meet, but nothing truly proves it.

In addition to the visual observation of errors (made easier by a graphic presentation) information can be synthesized using the following criteria:

- average absolute error
- average absolute percentage of error.
- standard error (square root of the average squared error)
- standard error divided by the average of observations means.
- bias

One will generally start with one of the first four, which we shall now present with

\hat{x}_t as the simulated value, and x_t as the observed value.

The root mean square error:

$$\text{RMSE} = \sqrt{1/T \sum_{t=1}^T (x_t - \hat{x}_t)^2}$$

The mean absolute error:

$$\text{MAE} = 1/T \sum_{t=1}^T |x_t - \hat{x}_t|$$

The mean absolute percentage error:

$$\text{MAPE} = 100/T \sum_{t=1}^T |(x_t - \hat{x}_t)/x_t|$$

The root mean square error, normalized:

$$\text{RMSEr} = \sqrt{1/T \sum_{t=1}^T (x_t - \hat{x}_t)^2} / \bar{x}$$

The choice between a criterion in level or in relative value will come as usual from the nature of the variable. In general the relative criterion will be more relevant (and more telling): an error of 1% on household consumption (CO) is more easily interpreted than an error of 3 billion 1980 US Dollars. The absolute criterion will be used for elements presenting a strong variability, especially if their sign is not determined: the error on the trade balance will be measured naturally in billions of current units (Euros or US Dollars). But to better understand the figure, it will be more appropriate to divide it by a some normalizing element, like GDP. Levels can be used also if the notion of relativity is already present in

the variable itself (growth rates, ratios): the absolute error on the variable «growth rate of wages» will give the relative precision on the salary itself. The error on the profits ratio will also be better measured in points⁹⁷.

As for the choice between the two types of moments: first and second order, it will be often forced, packages providing in general only one of the two statistics. One should note that:

- The first moment is always lower than the second, when the formula is of the same type (Cauchy-Schwartz inequality).
- This difference is all the more important as individual errors are of unequal size: a simulation including one or two particularly badly simulated observations will be particularly hurt by the second criterion.

Finally, another question appears, in the case of dynamic simulations of quantities or especially indexes: should one consider the error between simulated and observed levels of the variable, or the additional error introduced by the simulation of the period? In others terms, should one consider the error on the level or on the growth rate? In the case of a price index, both can be meaningful: in economic forecasts, or when judging the economic efficiency of a policy, one generally uses the inflation level, but the cumulated evolution on the price levels is the determinant of price competitiveness and of purchasing power.

And the result can be completely different, as the following example will show.

The simulation of the price index p , from $t = 1$ to T (let us suppose T even), uses an econometric formulation in growth rates, where the formula:

$$p_t = p_{t-1} \cdot (1 + f_t) + e_t$$

can be affected by two sets of errors:

$$(1) \quad e_1 = a, e_T = -a, e_t = 0 \quad \text{for other values of } t.$$

$$(2) \quad e_t = a \quad \text{when } t \text{ is even, } e_t = -a \quad \text{when } t \text{ is odd}$$

(the sum of errors is zero in both cases).

We shall not discuss the way these results have been obtained: the first estimation looks skewed, at least for small samples, while the second should be corrected for autocorrelation. We have to admit the problems we shall meet come at least partially from faulty estimations.

⁹⁷ We get back to the same notion as usual: to choose the best criterion, one just has to consider the way the variable (and its changes) is normally presented.

The criterion "absolute mean relative error" will give the results:

	error on the level	error on the rate
case (1)	$a (T - 1) / T$	$2 a / T$
case (2)	$a / 2$	a

The order is strongly inverted according to the criterion used.

In more general terms, the issue is the following: is a simulation which presents a lasting gap to observed reality, but a relatively stable trend parallel to it, better than a simulation which oscillates around true values, with the same error band as the first⁹⁸?

More generally, the difference between the two criteria comes from the dynamic character of the model. At a given period, the error comes from the accumulation of present and past ones, the impact of the latter reduced in principle with time.

The growth in the variance of the global error will be reduced by the fact that the sum of original residuals is zero, so positive and negative ones will alternate. However, this condition is only met on the whole sample. It is quite possible, even with no apparent autocorrelation, that the current sum has a significantly positive or negative value.

So it is not surprising with highly dynamic equations that the smallest simulation error is quite often found at the last period, and the largest error close to the middle.

But in some occasions, an accumulation of errors can lead the variable away from its observed value. Let us take the following example, simplified from the DMS model.

The wage rate depends on the consumption price (in growth rates) and the unemployment rate (unemployment divided by the total active population).

$$tx(W_t) = a \cdot tx(PC_t) - b \cdot UN_t / (L_t + UN_t) + c$$

Unemployment decreases with job creation, with a lower intensity:

$$\Delta UN_t = -d \cdot \Delta L_t$$

Both equations are estimated, and subject to errors. Let us suppose the second error is not auto correlated. On the series UN the error will be the sum of non-correlated errors: it can very well keep the same sign over the period, for

⁹⁸ We prefer the first case.

instance if the few first errors are mostly positive. In the end the error will cancel out, but it will have introduced on W itself a cumulative positive error which can make its level diverge markedly from its historical value⁹⁹.

7.2.1.1 Reassessing the criterion

Actually the criterion is not so relevant, and its use is highly questionable, for an evident reason: the numerical reality one seeks to approach was known at estimation. It has been used to determine values of coefficients, but also to choose among possible formulations. Even a scrupulous model builder will be led sometimes to favor equations that give results close to reality, to the detriment of theoretical quality.

An extreme case is the introduction of dummy variables, taking the value 1 in one or several periods, and zero otherwise: they will allow the equation to go perfectly through a point quite badly explained. One does not see why this introduction, which improves the criterion, can improve the model and its properties.

It is clear in this case that the additional explanation does not add anything to the model, but rather masks a problem (all the more as these variables are attached to the most badly explained periods). They can be accepted only if the period presents a specific feature, impossible to translate into an explanatory element or behavior (as in the May 1968 case we have seen earlier). But even then one runs the risk, by eliminating the main fluctuations of the explained series, of making the interpretation of its behavior more difficult. And the problem of forecasting these variables will remain.

More generally, there is no strict connection between the proximity of formula results to the observed reality and its intrinsic quality (one can observe that very accurate estimations can provide coefficients values entirely unacceptable by any theory, for example presenting wrong significant signs). A model produced by selecting systematically the formulations giving the smallest distance to observed values, to the possible detriment of economic meaning, might give good ex-post simulations¹⁰⁰. By pushing the argument to the extreme, one can end up with a model describing exactly reality, provided enough variables are added¹⁰¹.

But a model determined according to these principles will probably fail the more difficult test of the analytic shocks, which we shall present later.

It is nevertheless also clear that a model failing to describe past evolutions will have to be rejected or at least corrected (but by theoretical rather than technical contributions), as the quality of its forecasts will be even lower. Passing the present test is therefore necessary, but not sufficient, to guarantee model quality.

Let us consider the formation of simulation errors, originating in the residuals (we shall consider later the consequences of errors on coefficients and the actual form of the equations). They have three sources:

- The original estimation errors: it is clear that the smaller the estimation errors, the better the expected adequacy of simulation results to observed reality.

⁹⁹ Of course, this would not happen with error correction formulations.

¹⁰⁰ Although it is quite possible that in an incoherent model the instantaneous and dynamic links between variables exert a strong expanding effect on originally small estimation errors..

¹⁰¹ But the small explanation given by the last variables introduced should give them very low acceptability statistics.

- The multiplying effects of the instantaneous links between variables. Taking into account as explanatory elements variables affected by estimation errors (directly or indirectly) will increase the number of error sources and in most cases the global error (in the absence of correlation, the variance of a sum is the sum of variances).

This is not guaranteed however, as some errors can compensate by pure chance, and some model mechanisms can play a moderating role. For instance, a positive error on demand will see its effect on production dampened by the increase in imports it will generate. This will happen in our model:

$$[14] \text{FD} = \text{CO} + \text{IP} + \text{IC} + \text{IH} + \text{gd}$$

$$[15] \text{TD} = \text{FD} + \text{CI}$$

$$[16] \text{DLOG}(\text{M}) = 1.2 \cdot \text{DLOG}(\text{TD}) + 0.282 \cdot \text{DLOG}(\text{UR}) - 0.212 \cdot \text{RES_M}(-1) - 0.629 + \text{EC_M}$$

$$[17] \text{Q} + \text{M} = \text{FD} + \text{X}$$

A positive error on IP and FD will increase M, which will limit the impact on Q.

- The dynamic error (in dynamic simulations), coming from lagged variables which are now the result of a simulation. The error should grow with each period simulated, as again the number of error sources increases. However if we impose to all our estimated equations an error correcting structure, the problem is partially solved: if a variable moves away from a target value, the behavior of model agent(s) will drive it back to this target (it is not surprising that an “error correcting” model corrects the errors better than others). With these models, medium-long term simulations can actually give better results than the short-medium term ones, and ex post simulations can be more accurate at the end than in the middle¹⁰².

We shall improve our understanding of the implications of error correction when we address stochastic simulations.

7.2.1.2 Partial tests

To better interpret the simulation errors, some of their sources can be eliminated

- By exogenizing variables.

The variables will take their exact values, which should reduce the global error. The measure of the decrease gives some idea as to the contribution of the variable.

However this technique will mix two effects, in the case of a behavioral variable: suppressing the estimation error and the model errors transiting through this variable. This will make it difficult to interpret. In the case of identities, the change in the error will show clearly the contribution of the variable.

¹⁰² On the past, this is favored by the fact that the sum of the historical errors is zero.

One can consider:

- Eliminating a single variable (or a small set) and observing the decrease of the residual compared to the full model.
- Eliminating a large set of variables (for instance all the prices) and observe the expansion of the residual through model mechanisms.

With EViews this is done quite simply through the EXCLUDE statement.

- By re-introducing residuals (starting from a fully null set) :

This is done by setting the estimation residuals to their observed values, starting in two directions:

- Keeping every residual null, except for one equation. The (probable) reduction of the error will show the gains to be expected from improving the equation.
- Setting every residual to its historical value, except for one equation. The remaining error will show how the model dynamics expands the original estimation error.

Of course, combinations of variables also can be used: for instance observing the global error coming from external trade, or the error remaining if all prices showed no residual error....

With EViews, we can simply put the requested residuals to zero (to avoid losing their estimated value, they can be stored in an alternate variable). However, there is a much better technique, which uses the SCENARIO option. We shall present it shortly.

- by using static simulations, and comparing them to dynamic ones

This allows to separate the impact of the error on the first period, and from its dynamic influences.

7.2.2 EX-POST FORECASTS

To make the above criterion lose its artificial character, it looks more appropriate to simulate the model over a period which has not been used for estimation. This can be done by excluding from the estimation sample the last known periods, and using them to check the quality. If we eliminate the first periods, checking the quality on these periods is not very relevant for the purpose of the model: forecasting.

But this introduces several problems:

- It reduces the size of the sample, and therefore limits the significance of estimated coefficients. To reproduce the conditions of the test, it is necessary to remove a high number of periods (5 at least in the case of a medium-term annual model).
- Evaluating the result faces the same subjectivity problems as the previous test.
- But the main problem is that if we select models on the basis of the test, we use actually the same technique as on the full sample, and the same “ad-hoc” critique applies.

Everything depends on the way the test is conducted.

- First, it is always applied to a model (or a set of models) which has been estimated initially on the full sample. One never restricts the sample during the whole estimation process, then tests the models on the eliminated periods. This would be quite inconvenient, as these periods are the most representative of the domain on which the model will be used: the future.

So all the equations have been validated on the full sample. When we restrict the period, we get other equations which might or might not be satisfying from an economic and statistical point of view. If the equations have remained stable, the residuals are probably of the same size, and it is quite improbable that the multiplying effects of the model properties will change very much, so the ex post error should not be much larger.

But if the results change, we will question the specifications themselves.

This means that the only message from the new test is the stability of the set of equations, which could have been evidenced using a simple Chow breakpoint test, a more objective criterion. What we still get is a more global (and perhaps utilitarian) measure.

Of course, if the modelling process had been applied to the reduced sample, the process would be much more valid, provided one uses the result as an absolute measure of model quality, and not as a way to select the best model in a set. In this case, the process is not much different from estimating on the full sample. The choice relies on a test (obtaining good results on out of sample simulations) which basically requests equation stability, which means that the equations obtained are not very far from the ones from the full sample.

In other words, the success of the ex post forecast requires that the estimations work on the full sample, and that they keep stable if we eliminate the last periods. Of course, stability is a requirement for model reliability. But it is not a proof: an equation inverting a true and strong link between two variables (making demand depend on imports for instance) should remain stable over any sample.

7.2.3 SOLVING THE MODEL: SCENARIOS

To apply the above techniques, an EViews feature is quite useful: the scenarios.

When one wants to change the way the model is solved, it can mean:

- Solving the same model under
 - Different assumptions on exogenous variables
 - Different horizons
 - Different options for formulations allowing alternate versions (for instance the exchange rate equation can propose a constant nominal rate, a constant real rate or uncovered interest rate parity).
- Solving different coexisting versions of the same model.
 - One can be official, having passed all the requested tests and familiar to the operational forecasters, and the other can be more experimental, in the process of being tested.

The **scenario** option provides an efficient way of dealing with this issue.

- It works first through the creation of a new scenario.

In a program, the statement is:

```
model-name.scenario(n) scenario-name
```

This will create a blank scenario even if it exists already. If a previously defined scenario should be reused, one should state:

```
model-name.scenario scenario-name
```

Then one should:

- Assign a suffix to the scenario/model solution:

```
model-name.assign @all suffix
```

Using EViews 8 this can actually be obtained in one step:

```
model-name.scenario(a=suffix) scenario-name
```

- Create assumptions associated with the scenario, using the name of the variables followed by the suffix.
- State that the solution should look for the alternate versions of these variables by using the override statement:

```
model-name.override list-of-overriden-exogenous
```

For instance to solve our model **_fra_1** dynamically, with zero residuals for exports and imports (series called `m_ec` and `x_ec`) and the current suffix “_s”.

```
' We solve over the past  
' with residuals set to zero  
' using the suffix « _s »
```



```

genr m_ec_s=0
genr x_ec_s=0

_fra_1.scenario "scenario 1"
_fra_1.append assign @all _s
_fra_1.solveopt(n=t m=1000,c=1e-6,o=n,d=d)
_fra_1.override ec_m ec_x
solve _fra_1

```

To override variables, two operations are needed:

- Stating that the variables have to be overridden.
- Creating the alternate assumptions using the scenario suffix.

If one step is missing, the consequences are different.

If a variable is omitted from the list, the override operation will not be applied, without any adverse consequence (but no message). This allows to define a pool of potentially overridden variables in which the choice will be made later.

If a variable in the list has not been created, with the adequate suffix, EViews will refuse to solve the model, just as if the original variable did not exist.

If the program switches to another scenario, the options associated to the previous scenario will be put aside but remain associated with it, until they are modified under this very scenario. A new override statement cancels any previous overrides.

To eliminate overrides, one can just specify a blank list.

EViews 8 provides an alternate option: you can drop overrides individually through the REVERT command:

```
model_name.revert list-of-unoverriden-series
```

But the most important addition is the possibility to specify directly the changes in the assumptions, using ADJUST.

The syntax is:

```
model_name.adjust(init=initial_series) series adjustment
```

For instance the statements :

```
_mod_1.scenario(a=_2) "scenario 2"  
_mod_1.adjust(init="scenario 1") gdp =+10000
```

will create a series called gdp_2 with a value 10000 higher than the value from "scenario 1" (maybe called gdp_1). The series will be added to the "override" list, and to the "exclude" list as well if the variable is endogenous.

Be careful to introduce a space before the "=" sign, or gdp will take the value 10000!

Introducing a change involving elements, such as series and parameters, is possible but more difficult. One can prefer using actual expressions, as we have done in our examples.

7.2.4 OVERRIDING THROUGH MENUS

You have certainly understood that in our opinion menus are not the best way to manage alternate scenarios. However one might want to use this option to make a quick study, for instance to test a simple assumption.

This is done by accessing the model, then its "Variables" presentation, under a specific scenario (excluding "Baseline").

Right-clicking on a variable name will open a box allowing you to change its override status, using "Edit Override".

If the variable is not currently overridden, it will be added to the list (endogenous variables will be endogenized first, by adding them to the EXCLUDE list).

Then a window will open, with four columns.

- The overriding values (this is where you shall enter your changes).
- The base values.
- The difference in level (to the actual value or the baseline scenario)
- The same difference in percentage.

Just as in programs, clicking on the REVERT item will get the series back to the original values, and the endogenous back to their original status.

For an endogenous EXCLUDEd variable (through any method), REINCLUDE will also change its status back, but will create an add factor for the variable, with null values.

7.2.5 OUR EXAMPLE

We shall now simulate the model with residuals set to zero, on the 1978S2 - 2002S1 period, the longest one for which all estimations could be conducted.

- ' Ex-post simulation
- ' We solve over the past
- ' with residuals set to zero

```

smpl 1978S2 2002S1
for !i=1 to g_vbeha.@count
%2=g_vbeha.@seriesname(!i)
genr ec_{%2}_s=0
next

_fra_1.scenario "scenario 1"
_fra_1.append assign @all _s
_fra_1.override ec_i ec_ci ec_le ec_m ec_x
solve(d=d) _fra_1

' We compute the residuals and their absolute averages
' - the relative errors (psa_*)
' - the absolute errors on the growth rates (psd_*)

' We put them in a (19 x 2) matrix

matrix(20,2) v_psm_a
for !i=1 to g_vendo.@count
%st1=g_vendo.@seriesname(!i)
%st2=g_vendo.@seriesname(!i)+"_s"
series psa_{%st1}=100*({%st2}-{%st1})/({%st1}+({%st1}=0))
series tc_{%st2}=100*({%st2}-{%st2}(-1))/({%st2}(-1)+({%st2}(-1)=0))
series tc_{%st1}=100*({%st1}-{%st1}(-1))/({%st1}(-1)+({%st1}(-1)=0))
series psd_{%st1}=tc_{%st2}-tc_{%st1}
v_psm_a(!i,1)=(@mean(@abs(psa_{%st1})))
v_psm_a(!i,2)=(@mean(@abs(psd_{%st1})))
next

```

These statements might look a little complex. They just compute the mean of the absolute value of both criteria.

We get the following results

	Simulation errors	
	Percentage error on the level	Error on the growth rate
CAP	1.69	0.54
CO	1.43	0.38
FD	1.36	0.49
I	2.85	0.47
IH	1.30	0.34
K	2.70	0.43
LE	1.68	0.45
LED	1.36	0.40
LT	1.29	0.35
M	2.81	1.40
PRLE_T	-	-

The quality model ¹⁰³ .	Q	1.36	0.40
estimation	RHI	1.30	0.34
our	TD	1.36	0.45
an error	UR	1.00	0.46
appearing at	X	1.77	1.36

looks rather acceptable, considering the size of the residuals, which have a zero sum over the period. And estimations, except for the change in inventories, follow a correction framework. This means that any residual a given period is corrected with time, sometimes very fast. The variance of the global error will grow, but converge in the long run to a finite value, as we shall observe later.

7.2.6 ANALYTIC SHOCKS

If ex-post simulations give a valued criterion (more or less valid) of the quality of the model in forecasts, analytic shocks allow judging its properties (so the reliability of policy studies) by studying in terms of quantified economic logic the results of model simulations. Thus the model will show if its quantitative properties are consistent with the economic theory that has presided over its construction, and not too atypical as compared to the other models of the same class (the second condition is not of course necessary, but to ignore it begs for further study and strong arguments).

For this, a set of simple changes will be undertaken independently, on the main assumptions, trying to encompass each of the main areas described by the model. Every type of domain must be treated: agents, production function, supply and demand effects. This means we will study the sensitivity of the economy to a variation of exogenous demand (by State investments for example), the behavior of households (by a variation of income tax), of firms (by a variation of their social contributions), the foreign trade (by the impact of a devaluation or a change in the tariffs rates), the role of capital productivity, of an exogenous price increase. The quantitative and logical analysis of the evolutions of the main economic elements will give indications on model quality, by comparison with economic theory and properties of other models.

If the information gained from this criterion is probably more instructive, it also loses its objective character: a model which makes households decrease their consumption if employment improves will be clearly wrong, but an original behavior compared to the set of other models (for example if exports show a particularly strong sensitivity to price competitiveness, or depend on the age of capital) will not be conclusive, in the absence of a general consensus on the theoretical value. This might even, in some cases, affect the sign of the connection: facing a rise of inflation, households can increase their savings (to maintain their purchasing power) or decrease them (to buy immediately what will soon become more expensive).

In practice, to validate an atypical behavior, one will have to identify the source of the originality. Sometimes it will have been expected before any simulation, as it comes from an original feature introduced voluntarily into the model.

For instance, most models suppose that an increase in local demand will be partly satisfied by limiting exports. But our model can say also that the investment efforts it requires, in addition to increasing productive capacity through capital, allows local producers to supply goods better adapted to potential demand, making them more attractive to foreign

¹⁰³ But one could also suggest that the smaller the model, the larger the domain of the assumptions and the smaller the sources of uncertainty.

importers (this is called non-price competitiveness). Then a shock on demand might after some time bring an increase of exports, compensating the basic substitution effect¹⁰⁴.

In general, to make the interpretation easier, the change in assumptions will not fluctuate with time. One will distinguish however three types of shocks, according to the nature of the variation:

- One-time shocks, where the modification disappears after one period. For example the State will increase its investment for one budgetary year. In this case one will be often interested, not only in the evolution of effects (which should disappear with time), but also in the cumulating of these effects over time (such as the total gain on production).
- Sustained shocks, where the modification is maintained for a number of periods, which can be the full sample used in the simulation. For example a reduction of the income tax rate will be maintained on the period of simulation.
- Cumulative shocks, where the shock is amplified at each period by an identical amount. Their interpretation will be often difficult.

It is clear that if one wants to correctly interpret these shocks, they must remain logically acceptable. The consequence for the type of shock is not too clear.

In summary, one must consider two criteria:

- The acceptability of the shock.
- The readability of results.

This allows discarding one-time shocks in general, especially for policy variables. They fail most of the time on both elements. Except for an additional investment or subsidy, policy decisions are generally taken to last for a while, especially if they apply to tax rates or social security benefits. Structural changes (gains on factor productivity) are also permanent. As to external shocks (like an increase in world demand) they could very well be one-time.

But the most decisive argument is more down to earth: most of the interpretation will be done through graphs, and one-time shocks leads back all the variables to the baseline scenario. After a few periods, the return to zero of variations will make it impossible to separate the lines from each other, creating in general a high confusion close to the time axis, while in the case of a sustained shock the stabilization of consequences will create a set of more parallel (or slowly evolving) lines. As to cumulative shocks, they will produce expanding graphs and make the interpretation of the initial changes (associated with a small scale) more difficult.

So we are left with sustained shocks, which are indeed the easiest to read, provided the shock is reasonable. This applies first to the intensity: the less reasonable the shock, the more difficult its interpretation. This means the shock must be

¹⁰⁴ We observe here an important feature: the originality can lie not on the sign in the short run, but in its reversal in the long run. More generally, original long term features (favored by the error correction format) can appear only with time.

small enough (for instance affordable in terms of budgetary cost) and large enough (why should the government bother with such small decisions?).

But it applies also to the cumulated cost of the shock, in the case of policy decisions. If the period is long enough, and the shock important enough, one can consider that its cost will become unbearable after a number of periods. And the economic agents know about this. So not only should we stop the shock in time, but the model should adapt the behavior of agents to this future event. This is not possible with our present model, in which our agents look only backwards in time.

This is the most basic principle in the Lucas critique, which we shall approach later, when we address rational expectations.

Of course, this long term problem will be more accurate when we consider forecasts: we could stress that the sample period is small enough to allow sustained policy changes.

7.2.6.1 The choice of modified assumptions

Concerning modified variables, one will look for a set that asks contributions from all the mechanisms of the model: as the purpose is to validate the whole model, no abnormal behavior should remain. An initial investment in the full examination of these elements will avoid later problems, the identification of which will prove to be far more expensive (provided it can be done).

Especially, one will seek to put into play the whole set of agents, and the whole set of domains: production, labor market, prices and wages, external trade, financial mechanisms. Supply and demand oriented influences must all be studied. However, if there was only one test to realize, the most representative of a usual model would consist in increasing the exogenous part of demand, thus determining the Keynesian multiplier. We shall come back to this in the following paragraph.

7.2.6.2 The choice of the results considered.

All variations of endogenous elements may or may not be examined, according to the size of the model. In many cases, the first analyses will be taken on aggregated results, resorting only to detailed values if a problem comes to light.

- In the case of a macro-economic model, the main elements will be:
 - The supply and demand equilibrium: gross domestic product, demand and its decomposition, exports and imports (the ratio of the variation of GDP to the increase of exogenous demand defines the Keynesian multiplier).
 - Prices and wages, interest rates.
 - Employment and unemployment.
 - Foreign and budget balances.
 - Some ratios: savings ratio, profits, utilization of production capacities, financial balances and debts as a share of GDP.

7.2.7 OUR EXAMPLE

In our example, the number of available assumptions is quite low. The most reasonable choices are:

- Government demand GD
- World demand WD

And maybe also:

- Capital productivity PK
- The savings rate SR
- The depreciation rate DR¹⁰⁵
- The share of non-wage household revenue in GDP R_RHIQ
- The share of household housing investment in revenue R_HI
- The residuals in each of the five estimated equations.

The main elements are demand oriented ones. We shall select the most widely accepted: government demand GD.

We shall use the following statements, taken from a program (we have left the comments):

```
' We produce a shock

' First, an unshocked simulation
' The suffix is _b
' We override no element (we make sure of this through a blank list).

smpl 1980S2 2002S1
_fra_1.scenario "scenario 1"
_fra_1.append assign @all _b
_fra_1.override
_fra_1.solveopt(n=t m=1000,c=1e-6,o=n,d=d)
solve _fra_1

' Now, the shocked simulation
```

¹⁰⁵ Shocking DR introduces an interesting problem, as it is de facto forbidden. Indeed, the development of our investment equation is based on the fact that DR is constant, which allows to integrate it to the constant term, so changing it violates model specifications. It is the only shock which will actually modify the long term rate of use, and this for wrong reasons. To allow it to change, one can simply estimate the change in capital instead of investment (the assumption on DR stability disappears) and rearrange the present capital equation so that it now determines investment.

- ' The suffix is `_v`
- ' We compute the shock on `gd` (government demand) by 1 point of GDP
- ' We apply it to the shocked assumption
- ' We override `gd`

```
smpl 1980S2 2002S1
```

```
genr dv_gd=.01*q_b*(t>=1981)
genr gd_v=gd+dv_gd
```

```
_fra_1.scenario "scenario 1"
_fra_1.append assign @all _v
_fra_1.override gd
solve _fra_1
```

- ' We compute the differences
- ' in absolute and relative terms
- ' They will be called `dv_*` and `pv_*`

```
for !i=1 to g_vendo.@count
%2=g_vendo.@seriesname(!i)
series dv_{%2}={%2}_v-{%2}_b
series pv_{%2}=100*dv_{%2}/({%2}_b+({%2}_b=0))
next
```

A few additional comments.

- We have started the shock a little after the start of the simulation. This allows checking that in absence of shock the two simulations give the same result, which should mean that the difference can be interpreted as coming only from the shock. Additional differences can come from:
 - o A different starting date
 - o A different model.
 - o Different assumptions.

This is not an absolute proof, however:

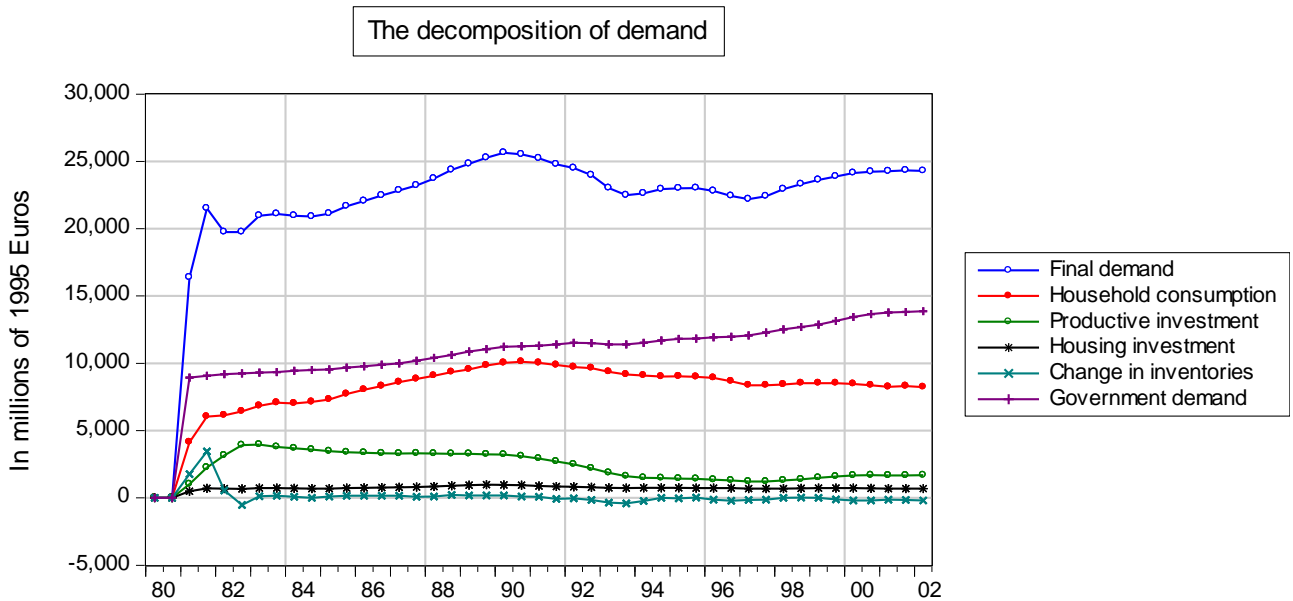
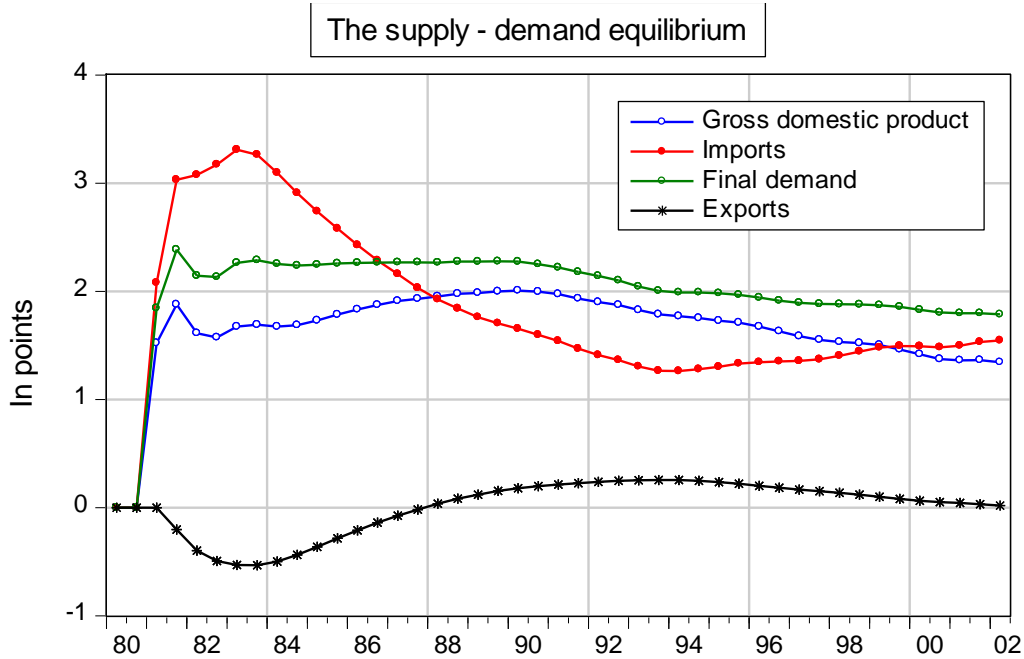
- o An additional change in the series might start after 1981.
- o The specification of the model might contain changes which apply only after 1981 (for instance a new tax) or only if a condition is met, which will occur later (like a high level of inflation).

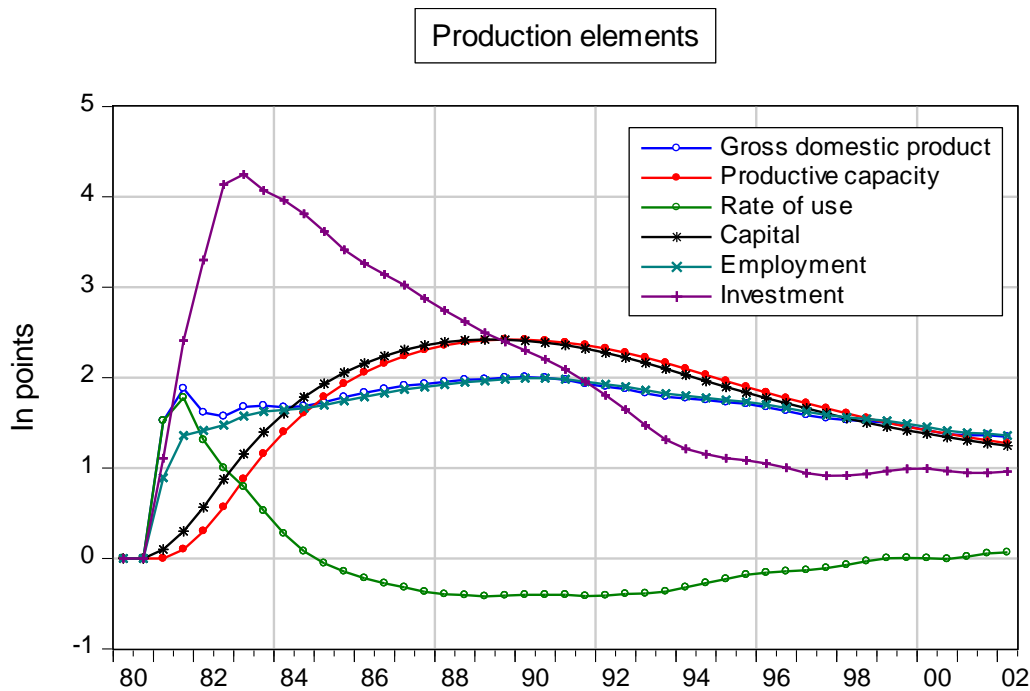
To guarantee the absence of such errors, it is advisable to simulate again the baseline scenario just before the shock, even it has been produced earlier. The efficiency of present algorithms and the speed of present computers make the cost negligible in most cases.

- We produced the alternate assumption in two steps, computing first the shock itself, with the same format ("`DV_`" and the name of the variable) as the absolute endogenous deviations. This will allow us to compare the ex-ante and ex post changes in GDP and create a graph presenting all changes.

7.2.7.1 The results

We shall base our commentary on three graphs showing: the evolution of the supply-demand equilibrium in relative terms, the decomposition of final demand in absolute terms (including the initial shock), and the evolution of elements linked to production.





The properties shown by these graphs are quite usual for this type of model (considering the reduced size of our version).

- The ex-ante demand variation is amplified by the increase in production factors (investment and employment), the latter affecting household revenue and consumption. These two elements keep increasing for a while, with some inertia, traditional for an error correcting process.
- Later, both changes become proportional to GDP. But while households obtain as revenue a high share of the increase, the motivation for investment decreases as capital adapts to GDP. In the long run, the only incentive left is the replacement of a higher level of discarded capital.
- Capital and capacity adapt slowly, but this inertia makes them overshoot the target. A cyclical process appears on the rate of use, with a rather fast decreasing intensity.
- External trade plays a simple role.
 - The change in exports follows the rate of use, with a negative evolution limited to the first periods.
 - Imports take a large initial share in the additional demand, when capacities have not adapted. They compensate the ex-post increase in demand, reducing the multiplier value to unity. In following periods, capacities adapt, and the multiplier increases regularly. The overshooting of capital actually favors local firms in the medium term.

On the whole, although this model remains overly simplistic, it looks rather sound if we consider its size.

7.2.7.2 The drawbacks of shocks over the past

Although producing shocks over the past for a model which will be used on the future does not face the same criticisms as the ex-post tests, they still have to meet several problems.

- They are conducted on a limited period. If one wants to observe the long term properties of the model, the available period will generally be too small. The problem will increase if data production has started recently, or if the country has undergone a recent transformation making the use of older data unacceptable. For Central and Eastern Europe countries, acceptable series start in 1995 at best.
- They are conducted on the wrong period. If one wants to interpret the dynamic properties of a model over a sufficiently long term, it is necessary to start the shocks quite far in the past. At that time, the economy of the country might have presented different characteristics. Even if the economic behavior was the same, some structural parameters were quite different, like the importance of external trade, or the relative roles of labor and capital in the production process
- They are based on a more or less unstable baseline. No model is completely linear (as it will always include both sums and logarithms). This means that the response to shocks will depend on the baseline, which can present irregularities. For example, a given increase in government consumption will generate more GDP if the activity is low, and the country can use its idle capacities to respond to demand.

This means that it will be difficult to decide if irregular variations in the results of the test come from model properties or from a non-smooth baseline.

Obviously, all these problems disappear when the shock applies to forecasts:

- The period is unlimited, allowing to evidence completely convergence processes and in particular cycles.
- As tests are conducted on the same period as future operational applications, the evaluation will be more relevant
- The present context and future issues are more familiar to the modeller
- As the baseline is smooth (especially in the medium and long term when it matters) any irregularity will be allocated to model properties.

This is why we suggest to limit to a minimum the studies on the past, and proceed to simulations on the future. This is true even for control of convergence: it is on the future and not on the past that the model will have to converge. But a limited series of immediate tests can evidence problems right away, which can help to gain time in the global process by staying on the right path.

CHAPTER 8: TESTING THE MODEL OVER THE FUTURE

The tests performed on the sample period were not so satisfying, for several reasons:

- For simulations, even if it was the only way to check results against actual data, the fact that this data had actually been used to produce the estimations could not avoid the tests from being flawed, whatever the precautions we had taken.
- For shock analysis, the sample period was generally too short to evidence long term properties and to measure cycles, and the irregularities (to say the least) in the base trajectory have been transmitted to the results (due to the non-linearity of the model), making the diagnosis on model stability unclear.
- Moreover, the results applied to the historical period, which is not the true field for future operational uses of the model.
- To obtain enough information, shocks have to be conducted on at least ten years, making the starting period quite far from the present.

This leads to the natural idea: test the model on the future.

- We shall have initial information on the reliability of spontaneous forecasts, and of their distance to what we expect of the near future.
- The results will be more representative of future use.
- The actual results can be interpreted as the actual consequences of policy decisions.
- The first periods of the shock will be representative of the present efficiency of present policies.
- The tests can be conducted on a period of any length, allowing to observe convergence and cycles.
- With regularly growing assumptions we can test that the simulation is regular, and that it converges to both a steady state growth and a long term stable solution.
- Applying to these regular solution constant shocks we can check that we get smooth evolutions, and we can interpret them easier.
- We have enough observations to treat the Lucas critique.

There are only two drawbacks:

- We cannot check the simulation results against true values. We shall try to prove this is not so important, and can be replaced by other tests.
- We do have to produce a simulation over the future, an unknown domain in which convergence might be more difficult to achieve.

But this convergence will have to be obtained anyway. So our first test will be:

- **Can the model converge at any horizon?**

In our sense, this is a prerequisite for structural econometric models. This is where this type of model has an advantage over CGE and VARs: to produce dynamic simulations which show the way the economy converges to a long term

equilibrium, and allow a full interpretation of the process. If this is missing (if the model converges only over the medium term) this advantage is lost.

So our first issue will be:

8.1 MAKING THE MODEL CONVERGE IN THE LONG RUN

First, let us stress that this is not the same as producing a long term forecast. What we are considering is the very long run (the first converging version of the model can take several centuries to stabilize). We are not going to suppose that the model will remain valid for such a period. What we want to achieve is:

A technical reproduction on a long period of the mechanisms associated with the equations we have evidenced, as if they remained valid forever¹⁰⁶. The period is much longer than the longest on which the model will be used.

For instance, many models present a cyclical behavior, linking investment, capital, the output gap and profitability. The length of this cycle can exceed 30 years. To check that its period is constant, and that it is converging regularly, we need at least three cycles, which means that our simulations have to last until 2100.

So at the beginning of the tests, one should consider a very long period.

This could be compared to testing a car (actually a prototype of a car) on an artificial track, for as long as it can run, exceeding by far its normal life (say several millions of kilometers), at a constant speed. And also changing one of the car elements, or the driving conditions, and observing the consequences.

What we want to get is information which will help us to build a better model, which will be used later in quite different circumstances (shorter but less regular environment).

We have to keep this in mind when we undertake our simulation.

8.1.1 THE ASSUMPTIONS

To produce a long term simulation, we need assumptions on the exogenous elements.

The choice of these assumptions can have several goals:

- Producing short term evolutions which coincide with what we know already of the near future (or of the recent past which we had to simulate in the absence of complete information).
- Producing reasonable medium term solutions, consistent with the goals of the Government, or of what we expect from the local economy (a high but decreasing Budget deficit for instance).
- Allowing the model to converge regularly to a long term equilibrium, which we shall examine to understand model properties.

¹⁰⁶ But we know this is not the case

For the time being, we shall concentrate on the last goal. Of course, we shall control that the two first targets are at least partially met, as an unreasonable simulation will react to shocks in an unreliable way, due to the model non linearities, making it more difficult to interpret.

This means we shall not differentiate the assumptions which share the same dimension, but apply to them common growth rates. In practice three rates have to be defined (in addition to a null one):

- o A rate for populations (including employment).
- o A rate for variables at constant prices.
- o A rate for deflators.

Of course, the second rate could be replaced by labor productivity. Combining it with employment we would get volumes.

Actually, we shall see that most models include essentially undimensioned assumptions. This is to be looked for as it suppresses the need to make normally independent assumptions grow at the same rate (in our model, Government demand GD and World demand WD).

8.1.1.1 A particular case: the estimation residuals

In general (this is the case for all the equations in our model) the estimated expression has been defined without dimension, if only to eliminate the risk of heteroscedasticity. And anyway, the most probable value of a residual is supposed to be zero. But one can suppose that the last value obtained is a better indicator of the future, in particular if the previous values held the same sign. For the time being, the choice is not so important, as we are not overly concerned with short term solutions. But it will become crucial later, when we approach actual forecasts.

8.1.2 ADAPTING THE FORMULATIONS

To make the model converge in the long run, applying any formula to variables which grow at the theoretical rate should produce a result which grows at its own theoretical rate (this is easily controlled as we shall see later).

Identities will provide this property automatically, if they are specified correctly and if the assumptions are well defined.

For estimated equations, the error correction format allows to introduce the necessary constraints in the long term equation. One does not have to be concerned with the VAR, which includes only growth rates, stable in the long run by definition. But one has to manage (generally suppress) any additional trend.

Let us explain this on our exports equation. We have:

$$(3) \text{DLog}(x) = 0.940 * \text{DLog}(wd) - 0.0129 - 0.195 * \text{res}_x(-1) + \text{ec}_x$$

$$(18) \text{res}_x = \text{Log}(x/wd) + 0.686 * \text{Log}(ur) - 4.87 \text{ e-}05 * @\text{trend}(60:1)$$

Let us suppose the rest of the model (in particular the investment equation) has allowed UR to stabilize in the long run. Then any growth in res_w will come from:

- X growing at a different rate from WD.
- The trend.

Let us now suppress the trend, for instance by replacing its expression by¹⁰⁷:

```
@trend(60:1)*(t<=2002)+@elem(@trend(60:1),"2002S1")*(t>2002)
```

Starting in the second semester of 2002 (represented by t>2002) the trend will be replaced by its 2002S1 value.

We have (with $\Delta\log(\text{WD})$ and ec_x constant)

$$\Delta\text{Log}(X_t) = a - b \cdot res_x_{t-1}$$

$$res_x_t = \text{Log}(X_t / \text{WD}_t) + c$$

giving

$$\text{Log}(X_t) = \text{Log}(X_{t-1}) + a + c - b \cdot (\text{Log}(X_{t-1}) - \text{Log}(\text{WD}_{t-1}))$$

and

$$\text{Log}(X_t) = (1 - b) \cdot \text{Log}(X_{t-1}) + a + c - b \cdot \text{Log}(\text{WD}_{t-1})$$

The growth rate of X will converge to that of WD.

In very special cases, the trend must not be suppressed but replaced by a given value, as it applies to non-constant ratios.

For instance in the employment equation, the structural labor productivity follows an estimated trend, estimated by:

¹⁰⁷ Let us recall the syntax for the @elem function:

@elem(series,"date") gives a scalar with the value of the series for that particular date.

$$\text{Log}(Q_t / LE_t) = a + b \cdot t$$

In the long run, the growth rates of quantities and employment are fixed, which defines the growth rate of labor productivity. The estimated coefficient has to be replaced by this value.

8.1.3 IMPROVING THE CHANCE (AND SPEED) OF CONVERGENCE

We have presented earlier some techniques. They can be applied indifferently to simulations over the past or the future. Few new elements will be specific to forecasts.

Actually the only significant one is associated with the starting values of the endogenous. EViews allows initializing the solving process either by “historical” values or those obtained at the previous period. This is done through the option: i=a (actuals) or i=p(previous) in the SOLVE or SOLVEOPT statement. In forecasts, only the last one should be available.

But actually one can define values on the future:

- By computing the theoretical values, applying to the endogenous variables their theoretical growth rate.
- By giving to the actual names the value of a previous solution, which should use comparable assumptions to the current problem.

These values are considered as “actual” by EViews, even if they correspond to future periods¹⁰⁸. They can be used as starting points, and it is possible (in particular in the second case) that they will improve the speed and probability of convergence. The issue is whether the growth rate of variables is higher than the difference for the same period between the two solutions. The two cases are quite possible, depending on the difference between assumptions in the second case.

In any case, the general principle applies: if one option does not work and another is available, just try it!

8.1.4 SOLVING PARTIAL MODELS

As presented earlier, an efficient way of diagnosing the origin of convergence problems is to exclude endogenous variable(s) from the solution. This calls for generating values for these variables. Generating starting values for all variables allows to use this feature at no further cost.

8.1.5 CHECKING THE EXISTENCE OF A LONG TERM SOLUTION

We shall now present a very efficient way of checking that model specifications allow a long term solution.

For this solution to exist, the following must be true:

All the exogenous assumptions sharing the same dimension must grow in the long run at a common rate.

¹⁰⁸ EViews leaves to the user the management of the present date.

Of course, their evolution can differ from this rate for the first periods.

As stated earlier, in practice three basic rates will appear:

- variables at constant prices (let us call it q)
- deflators (let us call it p)
- populations (let us call it n)

The other rates will be combinations of the three (in addition to stability). The main ones are labor productivity and purchasing power of wages ($q-n$).

The trends must be stopped, or set to theoretical values (like the trend in labor productivity).

In estimated formulas, the long term equations (in principle cointegrating relations) should use elements with zero dimension, and the dynamic equations too. This is not completely necessary, but it makes things much simpler, and is required anyway if we want the equations to be homoscedastic.

For dynamic equations, using logarithms (in general their variations) will provide this property naturally.

Under the above conditions, every formula must give a solution growing at the rate associated with the element it defines. For instance, the equation for imports at constant prices must give a result growing at rate q .

It is enough that a single equation does not meet these conditions, to forbid the whole model from having a long term solution, let alone reach it.

In addition, the estimated formulas must lead to the solution and not away from it. In particular the sign of the coefficients associated with error correcting terms must be the right one, and the value not high enough to create a diverging process.

If these conditions are not met, the model will probably explode in the long run, and it is quite possible that this process affects already the medium term solutions. To be sure this is not the case, the only option is to check the presence of a long term solution.

Of course, one can start by simulating the whole model over a long enough period. Unfortunately, even if high care has been taken, at least one error will generally remain, propagating to the whole system without generally giving clear clues as to the origin of the problem.

However, a very simple technique can be applied: producing a residual check over the future!

One just has to generate (as proposed earlier) a full set of exogenous and endogenous elements following the above constraints. Then the model is solved using the "fit" option (separate computation of each equation) and the growth rate of the result is compared to the growth rate of the variable the equation defines. The two must be identical. If not, the elements in the equation are not consistent with its formulation.

The origins of problems are somewhat different from a normal residual check. The most frequent are:

- A trend has been forgotten or miscalculated in the equation, for instance the opening of world trade in the imports equation¹⁰⁹.
 - The dimension of one or more element on the right hand side is not consistent with its role. This can apply to identities or estimated formulas.

For instance the exogenous variable associated with social benefits is computed in purchasing power per head, while the social benefits equation just multiplies this element by the deflator. Or a linear estimation adds up elements with various dimensions.

- The dimension of the endogenous variable is inconsistent with the formula. For instance a CES production function uses the wrong coefficients, used as exponents.
- We have decided to apply different formulas for the past or the future, for instance, blocking the trend after the estimation period, the second type could contain errors.

For instance, a trend term such as:

$$C(1)*((t-2007) (t<2007)+0.02*t*(t>=2007))$$

instead of:

$$C(1)*((t-2007) (t<2007)+0.02*(t-2007)*(t>=2007))$$

will make the element jump from 0 to $0.02*2007$ when forecasts start in 2007.

The obvious advantage of this technique is to identify most of the erroneous equations in the set. But as in the more usual application, a zero residual does not prove the absence of error, and a non-zero one can come from several errors. This means that in general reaching an (apparently) cleaned version will take several iterations.

The programs provided at the end of the book will systematically use this technique.

8.2 CONVERGENCE PROBLEMS IN THE SHORT RUN

¹⁰⁹ This trend can be continued in the initial forecasting periods, but it must decrease and disappear after a while. In this case, one must check the residuals only after this condition is met.

Success in the above test does not guarantee reaching a consistent solution, however: now the problem lies in the initial periods, and particularly in the very first one.¹¹⁰And the above method cannot yet be applied in all cases, as the lagged elements on the right hand side do not yet grow at the theoretical rate.

The most immediate danger is an error in the assumptions: the above program, which sets the growth rates to all the exogenous variables, should avoid this problem. We have checked that their growth rates are correct, and consistent with the formulas in which they are used. The only error could come from values given explicitly (for instance because we have advance information on some elements, or because we are trying to specify endogenous target by manipulating exogenous instruments).

Finally, it is possible that the management of residuals in the first forecast periods brings high variations, and in particular strong cycles, which

- make one of the elements pass through unacceptable values (such as negative levels of goods, for which we consider the logarithm)
- or produce a model with diverging properties. This can be true for Gauss-Seidel, if the linearized model is very different from the right one, it can contain eigenvalues with a higher than one modulus. And the Newton method can make the solution move in the wrong direction.

As to the remedies, they are not different from the ones developed in the previous chapter.

8.3 CONVERGENCE PROBLEMS IN THE MEDIUM RUN

Now if the model has started converging, and contains a converging error-correcting process, it should give solutions over the whole period.

Of course this is not true if the second condition is not met: in that case we know nothing about the long run properties, and divergence can occur at any time (actually the greatest danger appears when convergence is still obtained at the end of the forecasting period, but numerical properties are already wrong, not yet bad enough to be noticed).

Otherwise, the only danger which remains is if the cycles are too strong, which can only occur if the model is wrong. But now we have help, just as if we tested a motor which runs for a while before breaking down: it is much easier to understand the reason than if it does not start at all.

The main solutions have been developed earlier:

- Applying shocks to the main assumptions and observing the consequences. Maybe the diverging process will start before the breakdown, and its reasons can be made explicit.
- Excluding some variables, or changing (increasing or decreasing) local explanations. Now we can observe how these changes affect the horizon of the crash, and the properties: less cyclical or more consistent with economic theory. This is much more informative than just trying to get a solution.

¹¹⁰ Just as takeoff is the most dangerous part of a plane flight.

8.4 TESTING THE RESULTS

Let us now suppose that our model is converging in the long run. It means that we have efficiently managed to establish a long term solution.

8.4.1 OF SIMULATIONS

What we should check at this stage is how the model behaves in the short and medium runs. Of course, we have introduced assumptions too crude to provide a realistic forecast. In particular, setting from the first the growth of world demand to the common value q will stop the expansion of exports (imports should follow). A more realistic option should be to reduce it gradually. However:

- The impact of this gradual decrease will change as the model evolves, so it will have to be reconsidered with time.
- This will go against the main present goal in simulating the model over the future, which is obtaining trajectories which are stable, thus easier to interpret, in a future as near as possible.

In practice the quality of our simulations has just to be reasonable enough not to question the teachings of sensitivities to shocks. They will represent the most important diagnosis.

8.4.2 OF SHOCKS

For the shocks, we shall use the same « Scenario » feature as in the previous chapter. We have already described the advantages of waiting for this stage to make the tests. The situation has not changed much.

As we have stated earlier, it is on the future that shocks are better performed:

- They can last longer.
- They start from a smoother baseline.
- The period is the same as for future uses.

In particular, one can check the exact long run values and ratios. Long term convergence itself is guaranteed by the previous production of a long term scenario, provided the shock is not too large.

Concerning the period itself, it is better to start with a very long one (like two or three centuries). This might look unreasonably high, but it allows:

- To check that indeed the long run is reached numerically.
- To interpret the convergence (or divergence) of potential cycles, which have to be observed in a high enough number. Cycles with a 50 years period are not uncommon (especially if the model slightly overdoes it) so observing 4 instances requires 200 years...

Of course once the model is stabilized, the cycles will probably converge faster, and the test period can be reduced, allowing to concentrate of the short-medium term properties (which will be much more visible in shortened graphs).

8.5 EIEWS FEATURES

We will now see how we can apply the above principles to EViews.

8.5.1 PRODUCING SIMULATIONS

We shall now describe in turn the phases of the simulation process.

8.5.1.1 Preparing the workfile

Let us see how we can manage this on our simple model. We shall concentrate on the specific features (the full program is made available later).

The first task is producing a « forecast » file. The period should be quite large

```
'
' we set the workfile to mode, annual from 1970 to 2200
'
close _fra_1.wf1
close _fra_1.wf1
wfoopen _fra_1.wf1
wfsave _fra_1.wf1

' We expand the workfile until the year 2200 (reduced later)

pagestruct(end=2200) *

' We set the two generic growth rates
' to values associated to "round" Dlogs
' This will help to check convergence to the theoretical value
' for variables with complex rates (like labor productivity)111

scalar txq=exp(0.015)-1
scalar txn=exp(0.001)-1

' We extrapolate the time trend

smpl 2002S2 2500S2
genr T=t(-1)+0.5
```

¹¹¹ This option can be questioned. Using actual growth rates with round value is simpler. However the growth rate of composite variables such as quantities at current prices, productivities, purchasing powers.... will contain a second order term which makes them lose this property.

Considering changes in logarithms eliminates this term, but calls for a (tedious) explanation, as the notion is not straightforward. In actual forecasts, where the results have to be presented to a public, it is better to use growth rates.

8.5.1.2 Producing the assumptions

Now we will produce the assumptions.

Fortunately, most of the assumptions have no dimension, and are independent from the growth rates.

The only exceptions are:

- Government and world demand, gd and wd , quantities which grow at txq .
- Government employment lg which grows at txn
- The real wage rate which grows as labor productivity, at $txq-txn$ (in $\Delta \log$ s).

```
' exogenous

for %2 gd wd
  genr {%2}={%2}*(-1)*(1+txq)
next

for %2 rfdx r_rhiq r_ih dr sr rvat pk compm tc ec_i ec_le ec_m ec_x ec_ci
  genr {%2}={%2}*(-1)
next

for %2 lg
  genr {%2}={%2}*(-1)*(1+txn)
next

for %2 wr
  genr {%2}={%2}*(-1)*(1+txq)/(1+txn)
next
```

We can also initialize the endogenous (as justified earlier).

```
for %2 ic i m x co fd k ih rhi q cap ci
  genr {%2}={%2}*(-1)*(1+txq)
next

for %2 ur res_m res_x
  genr {%2}={%2}*(-1)
next

for %2 le lt led
  genr {%2}={%2}*(-1)*(1+txn)
next
```

```
for %2 prle_t
genr {%2}={%2}*(-1)*(1+txq)/(1+txn)
next
```

Changing coefficients

For our simulation to provide a steady state solution, we only need to change one coefficient: the trend in labor productivity which must follow $txq-txn$ (in $\Delta \log s$) at each period. As it applies to t which grows by 0.5, we must multiply the coefficient by $2^{112.v}$. And as we have seen, our manipulation of trends makes the change in coefficient apply only from 2004, without calling into question the previous values of trend productivity.

```
c_prle(2)=2*log((1+txq)/(1+txn))
```

8.5.2 PRODUCING A BASE SOLUTION

To produce a base solution, we use the same technique as usual. We compute the growth rates of the endogenous to check that they converge to the long term values (zero, txq , txn or a combination of both).

Two remarks:

- We shall also compute the theoretical growth rates, and the difference between actual and theoretical, which should decrease to zero in the long run.

Actually we shall compute the $\Delta \log s$, for which the variables associated with combinations of growth rates (like labor productivity) will nevertheless give round values, making a visual check easier.

As stated earlier, this technique represents a trick, and one can rely on actual growth rates. It is in our sense less efficient, but it is much easier to explain (look the length of my message, and I am not sure it actually went through).

- The blank “exclude” statement insures against any previous exclusion associated with the present scenario. It is free, and one cannot be sure that it is not needed.

```
wfsave _fra_1
smpl 2003S1 2200S2
_fra_1.scenario "Scenario 1"
_fra_1.append assign @all _p
_fra_1.override
_fra_1.solveopt(n=t m=1000,c=1e-6,o=g,d=d)
solve _fra_1
'
```

¹¹² 4 for a quarterly model

```
' We compute the growth rate of all variables
' and compare it with the theoretical value
```

```
for !i=1 to g_vendo.@count
%st2=g_vendo.@seriesname(!i)
genr tc_{%st2}=100*log(((%st2)_p+{%st2}_p=0)/{%st2}_p(-1)+{%st2}_p(-1)=0)))
genr tc0_{%st2}=100*log(((%st2)+{%st2}=0)/{%st2}(-1)+{%st2}(-1)=0)))
genr dtc_{%st2}=tc_{%st2}-tc0_{%st2}
next
```

8.5.3 PRODUCING SHOCKS

To produce shocks we use the same technique as before. We apply to *gd* an increase of 1 GDP point (using *q_p*, the GDP level computed by the baseline simulation). We solve the model using the suffix “_v”. Then we compute the absolute and relative differences (destroying first any variable of this type, to avoid confusions).

Finally we export the results to Excel (this is becoming less and less necessary).

A trick: when a simulation breaks down in a program which contains several SOLVE statements for a given model, the EViews message will not let you identify which of them produced the error. A simple way to make this identification is to change the maximum number of iterations in the statement (1000 into 1001, 1002,....), as this statement will be displayed in the message.

```
genr gd_v=gd+.01*q_p*(t>=2004)

_fra_1.solveopt(n=t m=1000,c=1e-6,o=g,d=d)
_fra_1.scenario "Scenario 1"
_fra_1.append assign @all _v
_fra_1.override gd
smpl 2003S1 2200S2
solve(m=1001) _fra_1
delete dv_* pv_*
for !i=1 to g_vendo.@count
%st1=g_vendo.@seriesname(!i)
series dv_{%st1}={%st1}_v-{%st1}_p
series pv_{%st1}=100*dv_{%st1}/{%st1}_p
next

group shocks pv_* dv_*
write(t=xls) varia_pfra.xls chocs
```

8.5.4 CHANGING MODEL SPECIFICATIONS

To test model properties in the long run, it can be interesting

- To test alternate estimated formulas, the properties of which have to be ascertained.

- To allow some of its equations to follow formulations different from estimated ones, either through a change of coefficient values, or even a change in the actual formula.

This can be done of course by producing a new model creation program. But this will have several drawbacks.

- It will take more time and involve more statements.
- It will not single out an “official” version.
- One will have to maintain several model-building programs, making file management less clear.
- The forecasting program will have to be run anyway.

In our sense it is better:

- To maintain an official version of the model, representing the best tested version according to the present information.
- To introduce the changes in the “forecasting” program.

This can be done in two ways.

- If a new official version has been estimated for some equation, the full set of associated statements should be specified (residual set to zero, defining the equation and estimating it, storing the estimated residual, merging the equation).
- If one is just testing an alternate version with at least one fixed coefficient, with or without re-estimating the rest, we suggest the following procedure, which we shall present using investment as an example:
 - Copy into the “forecasting” program the set of statements, as above.
 - State a vector of parameters (which will not be estimated).

```
vector(10) p_i
```

- Estimate the equation as usual, with a different name (here: the normal name followed by “1”).

```
coef(10) c_i
smpl 1977S1 2002S1
genr ec_i=0
equation eq_i1.ls i/k(-1)=c_i(1)*i(-1)/k(-2)+c_i(2)*ur+c_i(3)*.25*q/q(-4)+c_i(4)+ec_i
genr ec_i=resid
```

This step can be bypassed if you are sure that the present equation is correct, and you know the estimation results. However, we advise it as it displays the situation and helps you control the process.

- Transfer the coefficients into the vector elements, making any desired change in the values. Here we want to set the tension coefficient to 0.5.

```
p_i(1)=c_i(1)
p_i(2)=0.5
p_i(3)=c_i(3)
```

- Replace all “c_” coefficients by “p_” parameters, except for the constant term (this will allow EViews to estimate the equation, a process which requires at least one coefficient).¹¹³
- Estimate the equation using the normal name.
- Store the new residual value.

```
equation eq_i.ls i/k(-1)=p_i(1)*i(-1)/k(-2)+p_i(2)*ur+p_i(3)*.25*q/q(-4)+c_i(4)+ec_i
genr pq_ec=resid
_fra_1.merge eq_i
```

We could modify more than one coefficient, of course, or none if we want to be ready for a future modification.

To apply this technique, we can consider two general strategies.

- Start with no changes, and copy in succession each equation which requires it.
- Copy at the beginning all equation-building elements into the forecasting program, with no value changes (all the “c_” elements are transferred into equivalent “p_”s). The specifications have changed, but not the numerical properties.
- The requested paragraphs will be changed in turn later, as needed. This calls for an initial investment, but makes the following changes easier and less error prone.

As you can guess, we favor the second option. But one should consider the share of equations he intends to change¹¹⁴.

Note: when you are testing new values, you should keep the old one as a comment, which will allow you to backtrack if the situation deteriorates, and also to recall the level of your change.

8.5.5 UPDATING THE MODEL SPECIFICATIONS

The above tests, or other considerations like a desire to introduce a new theoretical feature, can lead the modeler to update the model. This can represent a very simple change, like replacing the unemployment rate by its logarithm in the wage equation, or a very complex one, like changing the whole production function.

¹¹³ We have no simple solution if the equation has no constant term (as for the change in inventories in our example).

¹¹⁴ Which is difficult to forecast.

This process can lead to the introduction of new variables, or completely new behavioral equations.

Once the new set has been established, two options are available:

- If the model specifications are produced by a single program, one can simply edit it. Of course, one should keep a version of the old program, in case mistakes are made or the new model proves unacceptable, for any reason.
- One can keep the previous model creation program (or programs) and create an additional updating program applying the changes.

As you can expect, we favor the first option, which is both safer (recreating the model is completely safe forward) and clearer (the specifications of the model can be observed by sight, especially if it is not too large).

EViews 8 greatly improves the management of model equations, especially identities. Until EViews 7:

- Dropping an identity was not possible.
- Adding a new version actually duplicated the definition, so you had more equations than endogenous.
- Dropping an estimated equation could only be done through EXCLUDE, difficult to manage, as to keep it excluded one had to specify it again in all following EXCLUDE statements.
- To replace an estimated equation you had to use the same name, otherwise you faced the same duplication problem as above.

Now you can (if you choose that option¹¹⁵):

- Drop any kind of equation, using :

```
model_name.drop variable_name (for identities)
```

```
model_name.droplink equation_name (for estimated equations)
```

For instance you can specify:

```
_fra_1.drop GDP ' GDP is made exogenous
```

```
_fra_1.droplink eq_i ' eq_i, the equation for investment is dropped, and I becomes exogenous
```

- Replace the formula using:

```
model_name.replace variable_name (for identities)
```

```
model_name.replacelink old_equation_name new_equation_name (for estimated equations)
```

¹¹⁵ Remember that in our opinion, the most efficient method for editing a model is to destroy the previous version and create a new one from scratch.

```
_fra_1.replace GDP = .....  
_fra_1.replacelink eq_i eq_i_new
```

In addition you can also replace all occurrences of a model variable using a new name, through:

```
model_name.replacevar old_variable_name new_variable_name
```

For instance:

```
_fra_1.replacevar GDP Q ' replaces GDP by Q in the whole model (but not in estimated relations)
```

8.6 RATIONAL EXPECTATIONS

They are called by EViews « model consistent expectations » and their rationale is a little different from the usual one.

8.6.1 THE FRAMEWORK

In a rational expectations framework, we suppose that some of the agents (at least) are aware of the future assumptions, and are able to evaluate their influence on some (at least) of the model variables. This does not need the knowledge of the actual model equations, just the mathematical « application » from the assumptions to the endogenous.

As they will use this knowledge in their present decisions, some of the present variables will depend on future values, either of assumptions or endogenous elements depending on these assumptions.

To take into account rational expectations one does not need to believe in them. Interpreting the differences in economic behavior (and their consequences for the equilibrium) between forward and backward looking agents is quite interesting if only from a theoretical point of view. The following example will shed some light on this point.

8.6.2 CONSEQUENCES FOR MODEL SIMULATIONS

In this context, we can no longer compute the solution for each period separately, moving from the past to the future. The solution for a given (future period) will depend on values for which the solution has not been yet obtained.

This introduces two problems:

- Finding a way to take into account future values, belonging to the forecasting period.
- Finding a way to end the simulation, the last simulations taking into account values outside the forecasted period. This calls for the definition of « terminal conditions ».

8.6.2.1 Inside the forecasting period

The problem is purely technical. Two strategies are considered at present.

- The Fair – Taylor algorithm: it solves the model for each period in succession (using the Gauss-Seidel, Newton or Broyden algorithm each time), then goes back to the first one and iterates (in a Gauss – Seidel way) until convergence of the whole system. This is the technique used by EViews.
- The Laffargue (1990) algorithm: basically, it adds a time dimension to the model equations (duplicating them as many times as there are periods) and solves the associated model as a whole, using the Newton algorithm with possible improvements (using the fact that the matrix is band diagonal, the number of non-zero matrixes on a given line depending on the number of backward and forward lags).

Obviously, applying the last algorithm was only made possible by the increase in the computing power of computers, and more importantly of the memory capacity, as a very large set of values has to be managed at the same time.

8.6.2.2 Outside the forecasting period

EViews provides the usual options for terminal conditions, which are controlled by

```
t=arg
```

in the SOLVEOPT statement

arg can take the values:

```
"u" (user supplied – actuals, default)
"l" (constant level)
"d" (constant difference)
"g" (constant growth rate)
```

8.6.3 TECHNICAL ELEMENTS

EViews can detect the presence of future endogenous variables, and will switch automatically to forward looking algorithm (actually to Fair Taylor). One can use the same statements as usual, or change the terminal conditions from default values, if deemed useful.

8.6.4 OUR EXAMPLE

In our example, we shall limit the case to the investment equation.

In the present model, we use:

$$I/K(-1)=C_I(1)*I(-1)/K(-2)+C_I(2)*UR+C_I(3)*.125*Q/Q(-8)+C_I(4)$$

Dependent Variable: I/K(-1)
 Method: Least Squares
 Date: 11/07/12 Time: 12:03
 Sample (adjusted): 1977S2 2002S1
 Included observations: 50 after adjustments
 I/K(-1)=C_I(1)*I(-1)/K(-2)+C_I(2)*UR+C_I(3)*.25*Q/Q(-4)+C_I(4)+EC_I

	Coefficient	Std. Error	t-Statistic	Prob.
C_I(1)	0.825643	0.033524	24.62809	0.0000
C_I(2)	0.027926	0.006984	3.998580	0.0002
C_I(3)	0.152870	0.058292	2.622495	0.0118
C_I(4)	-0.052515	0.010337	-5.080438	0.0000
R-squared	0.959393	Mean dependent var		0.087074
Adjusted R-squared	0.956745	S.D. dependent var		0.006785
S.E. of regression	0.001411	Akaike info criterion		-10.21218
Sum squared resid	9.16E-05	Schwarz criterion		-10.05922
Log likelihood	259.3046	Hannan-Quinn criter.		-10.15393
F-statistic	362.2686	Durbin-Watson stat		0.628797
Prob(F-statistic)	0.000000			

Now we shall consider four options:

The basic case (full backward looking)

$$I/K(-1)=C_I(1)*I(-1)/K(-2)+C_I(2)*UR+C_I(3)*.25*(Q-Q(-4))/Q(-4)+C_I(4)$$

A mixed backward – forward looking case

$$I/K(-1)=C_I(1)*I(-1)/K(-2)+C_I(2)*UR+C_I(3)*.25*(Q(2)-Q(-2))/Q(-2)+C_I(4)$$

A full forward looking case

$$I/K(-1)=C_I(1)*I(-1)/K(-2)+C_I(2)*UR+C_I(3)*.25*(Q(4)-Q)/Q+C_I(4)$$

Another full forward looking case, extending the knowledge to a longer horizon

$$I/K(-1)=C_I(1)*I(-1)/K(-2)+C_I(2)*UR+C_I(3)*.05*(Q(20)-Q)/Q+C_I(4)$$

The global influence of Q is the same in each case, only the lags differ.

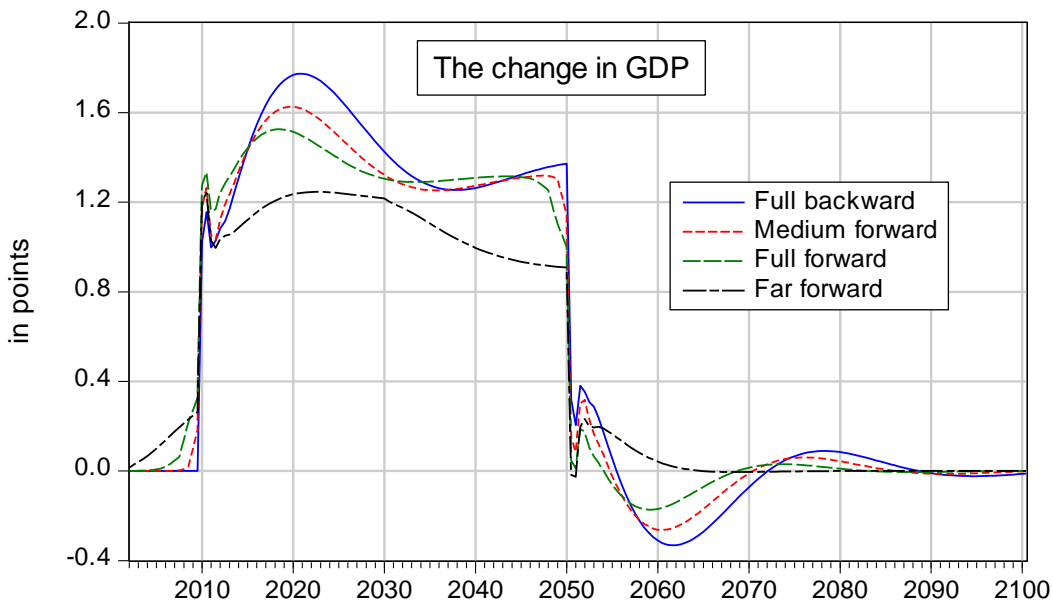
8.6.5 THE TEST

We shall consider that the State increases its demand by 1% of GDP (as in the previous shocks) but only from 2010 to 2050. The length of this period might seem too high, but it will be necessary for the analysis¹¹⁶.

Of course, one could also apply the technique to other elements: employment and change in inventories in particular. One should be able to do it easily.

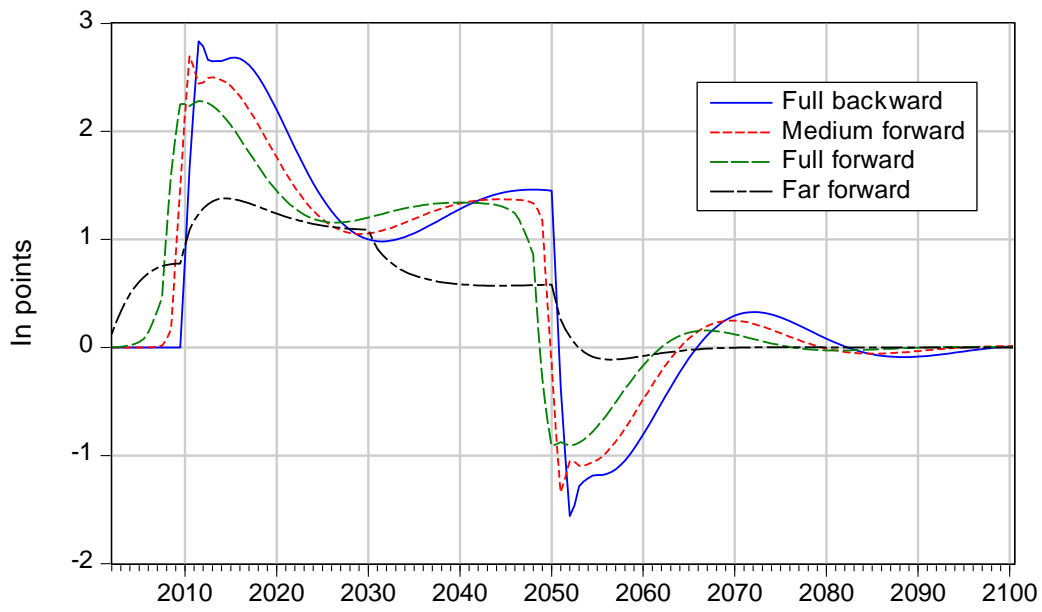
8.6.6 THE RESULTS

We shall concentrate on a few graphs. All changes are measured in percentage from the baseline.

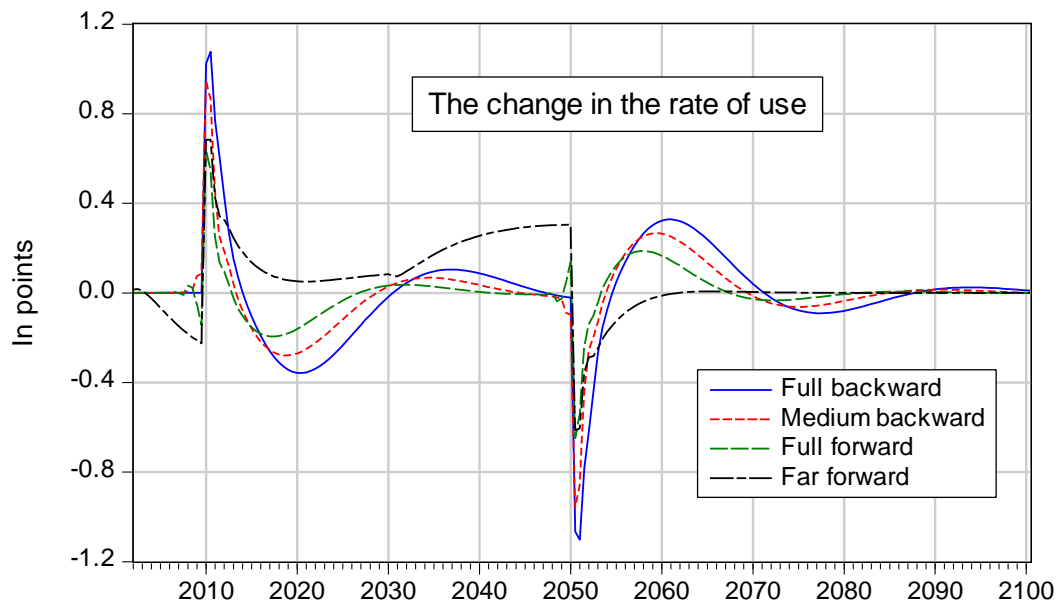


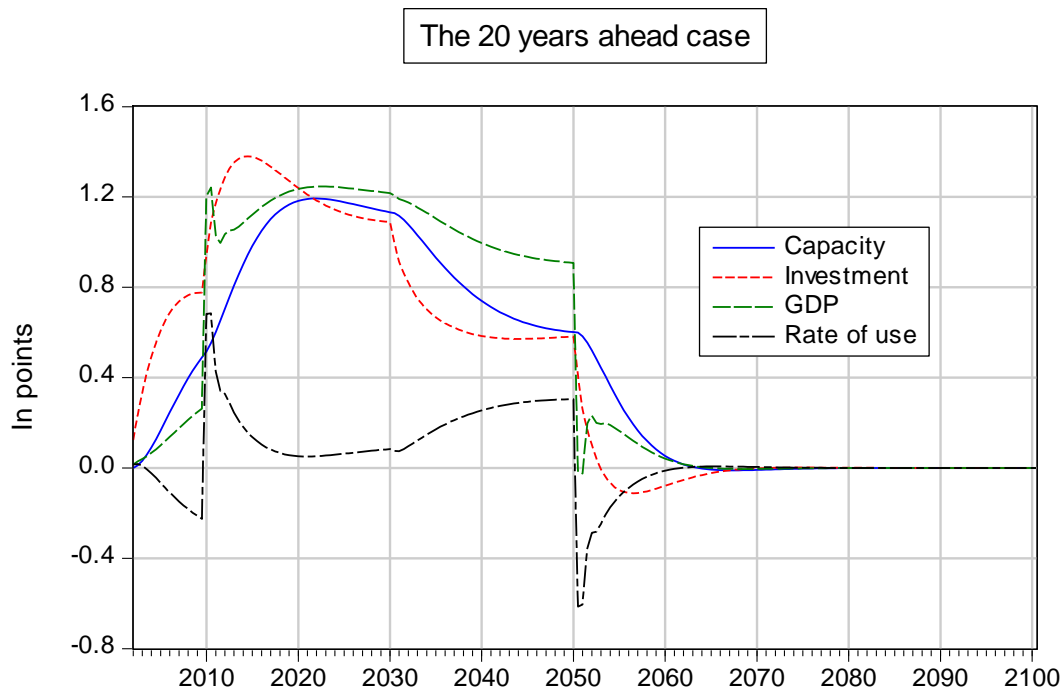
¹¹⁶ In fact, this decision will lead to an increase in the debt primary of the State by 40 percentage points of GDP (excluding the additional interest). The level of debt is certainly unbearable, reducing the interest of the study, particularly concerning the reaction of other agents.

The change in investment



The change in the rate of use





We can see very clearly generally logical elements:

- The more the investment decision is based on future GDP, the more investment adapts in advance.
- This is also true if the horizon increases. In this case, the weight of the periods associated with a change in decisions goes up.
- This produces a smoother adaptation to a change in Government decisions, in both events. The intensity of the cycles decreases, and they converge faster.
- With a 40 year horizon (case 4), the output gap does not change too much. Its higher values appear before the changes, when the investment decisions (and the resulting capital changes) are made according to the future, not the present, conditions. The negative overshooting of capacities observed in the last period is not present, and the convergence of the rate of use is monotonous.
- The changes appear sooner before the shock than the forward horizon itself. The early response of forward looking firms is taken into account even earlier by other firms, introducing a cascading effect (rather limited, however).

8.7 STOCHASTIC SIMULATIONS

For scientific purposes but also the analysis of a given model, the uncertainty of the model in forecasting may be directly measured. This error can come from several sources:

8.7.1 PURELY STATISTICAL ERRORS

This means

- The presence of a random residual term in behavioral equations
- The subsequent uncertainty on coefficients, even with an exact specification.

When a model is produced, some of the equations are estimated using an econometric technique, from simple ordinary least squares to the more sophisticated cointegration. This process provides:

- An estimate of the standard error of the estimation.
- Estimates on the standard errors of the estimated coefficients.

The precision of these two elements depends on sample size, in a different way.

- A larger sample should produce better estimates of coefficients, and reduce their error.
- But for the estimation residual, once the sample is large enough for estimation to be applied, enlarging it provides essentially a better knowledge of its distribution (in practice its standard error). This is due to the fact that even with a large sample, introducing new explanations in addition to five or six variables leads very quickly to colinearity problems.

8.7.2 FORMULATION AND ASSUMPTION ERRORS

Actually, these two errors are only half of the types facing model forecasts. One must also consider

- The fact that the estimated formula itself could be wrong. Even if a formula is accepted by all tests, one can always find another formula (sometimes with very different properties) which will show the same statistical quality, or better. And even if the formula applied to the past, the agents could very well modify their behavior later.
- The error on the assumptions used for the forecast, for which the model should not be accounted responsible.

8.7.3 CONSIDERING THE ERRORS

In practice only the last first types can be measured without too much difficulty, using statistics (supposing the last two are absent): the estimation of coefficients gives under some assumptions an estimation of the law of estimation residuals, as well as of the law of coefficients themselves. The moments of projected variables then can be computed, either:

- By drawing at random a sample of residuals, undertaking the associated projection, and observing the statistical characteristics of results (a « Monte Carlo » technique).
- By determining, analytically or numerically, the linear transformation from residuals to the model solution, and by applying this transformation directly to the law of residuals, to obtain the law of forecasted variables.
- More simply and under clearly less restrictive assumptions, by drawing randomly a sample from the sequence of observed residuals (with or without putting back the elements selected into the pool). This technique is "bootstrapping".

The same type of method applies to the uncertainty due to coefficients:

- By repeating enough times the following sequence of operations: drawing a vector of residuals on the period of estimation (using the estimated law or the "bootstrap" technique), simulating the model on this period to obtain a new sample of coherent variables (in the model terms), re-estimating coefficients based on this sample, and forecasting the re-estimated model. One will then measure the moments of the sample of projections.
- By using the law of coefficients to generate a sample of models, which when forecasted will provide a sample of variable values.
- By inferring analytically the law of the projected variable from the law of coefficients, and using the matrix transformation (Jacobian) to link the variables to the coefficients.

Three main error types should be considered:

- The bias: in the case of a model with non - linear properties (relative to the endogenous, both instantaneous and lagged), the mean of the solution will not be identical to the deterministic solution (obtained with a null residual): this introduces a bias.

The difference comes of course from the non-linearities of the equations, and any economic model presents non-linearities. Obvious cases are the presence of variables at current prices, product of a variable at constant prices by a deflator, or variables computed by applying a growth rate to their past value.

If one is only interested in this accuracy, a single experiment (with a large number of replications) will tell of the stochastic improvement. If it is small, one can stick with deterministic (and cheaper) simulations.

One must be aware that coefficient uncertainty can introduce non linearities even if the model is linear relative to coefficients (actually it is not linear to the set: variables + coefficients put together).

Let us take the simplest model (Q =GDP, C = private demand, g =public demand):

$$Q_t = C_t + g_t$$
$$C_t = a \cdot Q_t$$

It can be written as:

$$Q_t = g_t / (1 - a)$$

Which is not linear relative to a.

- the standard error: this time the interest lies in the dispersion of solutions around their average: this criterion will allow for assessing the reliability of results (starting possibly from the mean of stochastic simulations) by evaluating a confidence interval, or range of possible values.
- The distribution: what we want is a graph of the probability distribution of the random solutions. Obviously producing a consistent graph calls for more simulations than the previous cases. One interesting message is its symmetric character. Actually EViews takes into account dissymmetry, even without a graph, by producing a confidence interval eliminating the highest and lowest 2.5% of the values.

The analysis of the uncertainty is not necessarily limited to values of variables. It also can be measured on multipliers (or derivatives of variables relative to the exogenous), therefore on the efficiency of economic policies, or on eigenvalues of the structural form, thus on the probability of model convergence (see the following paragraph).

Finally, one could consider the covariance between equations. Cointegration should eliminate the problem, as it individualizes a group of variables or concepts which evolve independently from the rest of the model (provided each set contains a single equation). However, this is not completely true, as a variable can belong to several cointegrating equations, establishing a link between them.

If one considers also the error on coefficients, the process is a little more complex. Again, the sets of cointegrating equations are independent, but they always use more than one coefficient (one has to consider the dynamic equation and the long term relationship). In this case, it is necessary to take into account the correlation between coefficients, and the drawing will call for a multivariate normal law, with a non-diagonal covariance matrix.

The two last forms of error, on the functional form and the assumptions, are more difficult to assess. However:

- If we have recent information on some variables, the main aggregates (supply-demand balance, inflation) and especially on the model assumptions (external environment, public policy), we can simulate the model over these periods and observe if the evolution of the main aggregates is correct.

A rough evaluation of the functional error can be obtained by reducing the estimation sample, and simulating the model over the remaining periods (as in an ex-post forecast). Then the size of the residuals can be compared with the theoretical standard error obtained from the previous method (which supposed the use of the true equations). If the error is much higher than the standard error (for instance systematically more than twice its value) one can surmise that the estimated functions are unable to forecast the future, and the functional form must be wrong (or the agent has changed his behavior, which is in fact the same). This technique is similar to the Chow forecast test.

This technique does not face the critique presented above, as we refuse to get back to the choice of model.

- As to the error on assumptions, it is outside the responsibility of the model (only of the forecast producers). But an estimate of it can be obtained by applying an ARMA process to the external assumptions (not the instruments), and observing the reliability of the associated simulation. The error will then supposed to be the part of the assumption which cannot be forecasted.

8.7.4 THE INTEREST OF THE TECHNIQUE

It has a cost, and an advantage. The cost is the increased length of computations.

To get an accurate measure of the distribution, a minimum of 500 replications is considered necessary, 1000 or even 10000 being a better figure. This is clearly the main reason for the relatively recent use of this technique: 30 years ago, solving 1000 times even a small model could take several hours on a mainframe computer. Now the price can be afforded in most cases: on an average microcomputer, the simulation over 20 periods of the 705 equations model MacSim-CAN takes less than 2 seconds. Multiplied by 1000, we get about 30 minutes, a reasonable figure which might look large however to contemporaneous modellers, especially if simulations are part of an iterative process looking for the best forecast.

8.7.5 BACK TO OUR EXAMPLE

We shall solve the `_fra_1` model over 100 periods, starting as usual from 2002S1.

The only necessary change concerns the **`solveopt`** statement, where « `d=d` » is followed by « `d=a` »:

In addition, the **`solveopt`** option « `r=integer` » controls the number of replications (default: `r=1000`), and « `b=number` » controls the confidence interval (default: `b=.95`)

```
_fra_1.solveopt(n=t m=1000,c=1e-6,o=g,d=d,s=a)
_fra_1.solveopt(n=t m=1000,c=1e-6,o=g,d=d,s=a,r=500,b=.98)
```

Actually, the « `a` » option is not the only one associated with stochastic simulations. It is the one which gives the most information, including the confidence intervals, while « `m` » gives only the mean, « `s` » only the standard deviation and « `b` » the means and the standard deviation.

In our case, the solution will create for each variable in the model:

- A series for the mean, adding « `m` » to the name of the simulated variable. For GDP and a current suffix of F, the name of this variable will be « `GDP_FM` ».
- A series for the standard error, using « `S` », such as `GDP_FS`
- Series for the higher and lower bounds, using « `H` » and « `L` », such as `GDP_FH` and `GDP_FL`

8.7.6 THE RESULTS

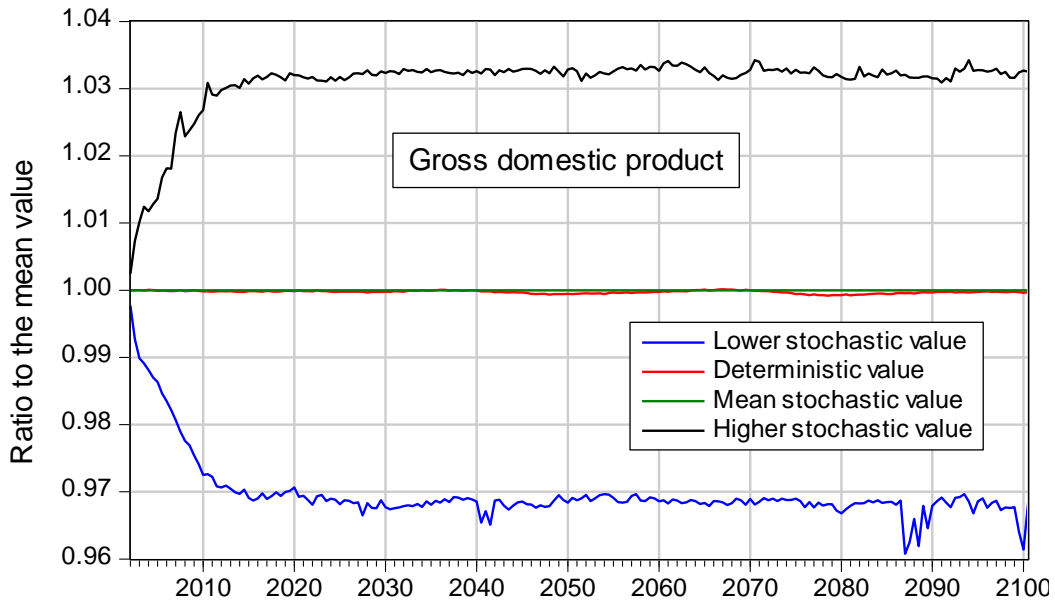
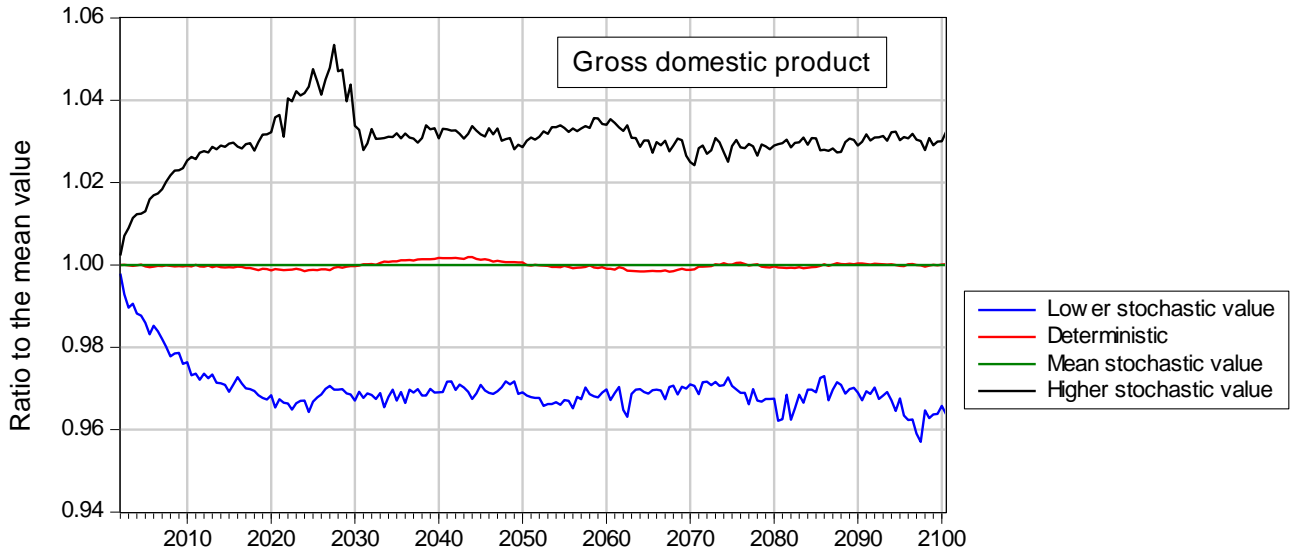
We shall perform a stochastic simulation on our small model, under the same conditions as the deterministic forecast, which we will actually conduct again, to measure the error it introduces.

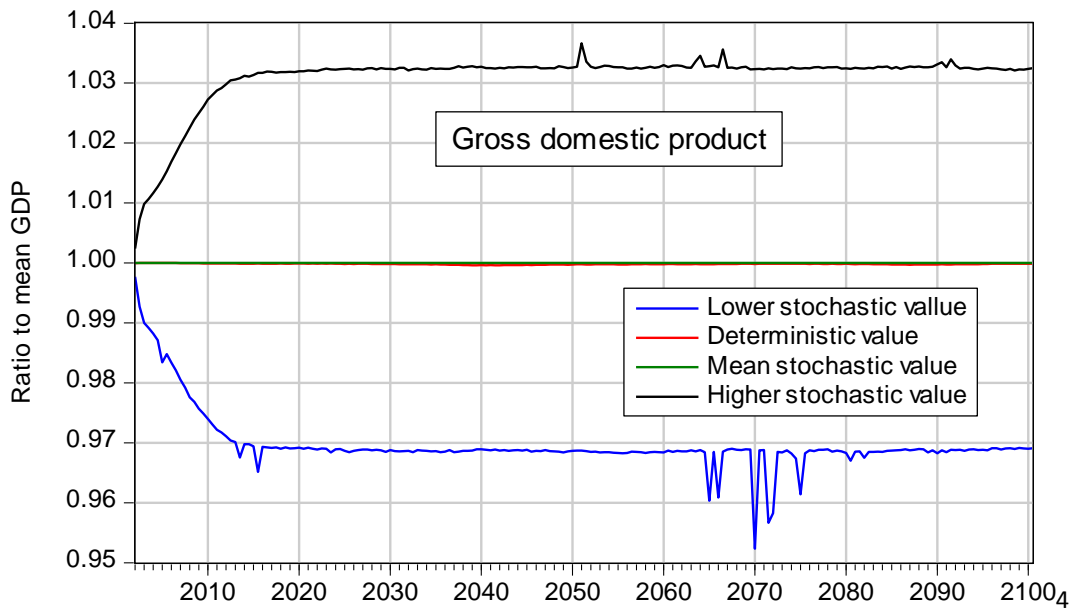
The program is very similar to the previous one, except for the line:

```
_fra_1.solveopt(n=t m=1000,c=1e-6,o=g,d=d,s=a)
```

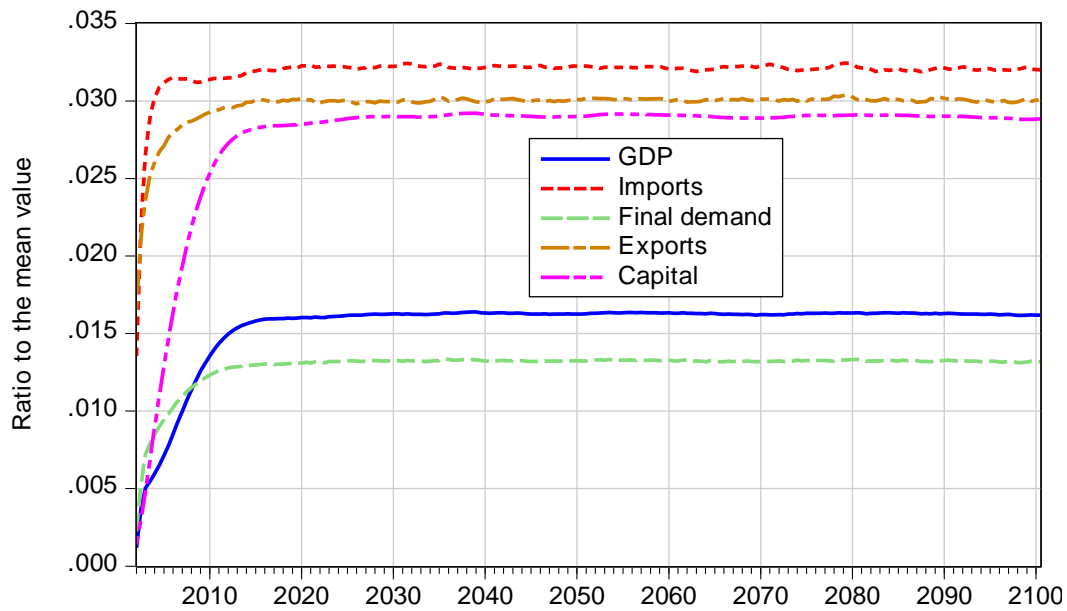
The following graphs show:

- For GDP, the evolution of its higher bound, lower bound and deterministic value, as a ratio to the average value.
- The ratio of the standard error to the average value, for a set of variables. Two graphs are actually provided, one with 100 replications, the other with 1000.





The standard errors



In the above graph, one will find from top to bottom: Imports, exports, capital, GDP, final demand.

In addition, the following table gives the average relative difference from the mean, compared with the error obtained from the ex-post simulation on the past.

Variable	Stochastic error	Ex-post error
CAP	2.85	1.69
CO	1.31	1.43
FD	1.38	1.36
I	4.36	2.85
K	2.85	2.70
LE	1.68	1.68
LED	1.59	1.36
LT	1.28	1.29
M	3.23	2.81
Q	1.62	1.36
RHI	1.31	1.30
UR	1.35	1.00
X	2.99	1.77

One can observe:

- That the error is relatively small.
- That the error increases at first for all variables, as the sources increase.
- That it converges in the medium term, faster for the trade elements (which use a cointegrating framework). This is consistent with stationarity properties.
- That the evolution is much smoother for capital (its values are highly correlated).
- That trade elements show the highest variability (and are correlated). This correlation does not come from Q.
- Comparing the ex post error, we can observe that the stochastic error is generally larger, in particular for the capacity variables. This can be due to the fact that the sum of the residuals is no longer zero, which helped the solution to get back to the baseline in the end.

8.8 GOING FURTHER: STUDYING MODEL PROPERTIES

In addition to the use of direct simulations, the mathematical functions provided by EViews allow to apply more complex techniques.

8.8.1 EIGENVALUE ANALYSIS

This technique can help to understand the dynamics of model convergence or divergence, and interpret them in economic terms.

Its application calls for the model to be linearized around a reference solution (now Δ stands for the difference between two simulations):

$$\Delta y_t = A_t \cdot \Delta y_{t-1} + B_t \cdot \Delta x_t$$

If A is supposed to be constant with time, a shock Δx undertaken at time t (only) on the vector of the exogenous x will lead at period t+k to a change in y:

$$\Delta y_{t+k} = A^k \cdot B \cdot \Delta x_t$$

and long-term consequences will depend therefore on the evolution of powers of A^k , when k tends to infinite. Three cases can occur:

A^k decreases to 0.

A^k increases to infinite (or at least one of its terms does).

A^k stabilizes at a finite, non-null limit.

Each of these three cases associates to the measure of the spectral radius of A (larger modulus of its eigenvalues): this radius will have to be respectively lower, higher or equal to unity. This will lead us to reason in eigenvalues, by stating:

$$A = V^{-1} \cdot \Lambda \cdot V$$

(Λ diagonal, containing eigenvalues of A, and V base of associated eigenvectors).

One obtains then:

$$V \cdot \Delta y_{t+k} = V \cdot A^k \cdot \Delta y_t = \Lambda^k \cdot V \cdot \Delta y_t = \Lambda^k \cdot V \cdot B_t \cdot \Delta x_t$$

Therefore, for a given shock on x, its dynamics can be studied by decomposing first period effects $B_t \cdot \Delta x_t$ on the eigenvector base of A, each element of this decomposition evolving then with time proportionally to the corresponding eigenvalue. Evolutions can be:

- Divergent, convergent or stationary, according to the modulus of the eigenvalue (higher, lower or equal to one).
- Monotonous or cyclic, according to the real or complex nature of the eigenvalue.

Actually the stability of A over time is much more assured if one considers relative variations:

$$\Delta y_t / y_t = A_t \cdot \Delta y_{t-1} / y_{t-1} + B_t \cdot \Delta x_t / x_t$$

This property has two explanations

- Most estimated equations are formulated in logarithms.
- By dividing the variation by the value, the result should be stable with time.

Indeed, this modification is equivalent to a replacement of each variable by its logarithm in model equations.

This does not change fundamentally the above considerations, except that convergence and non-convergence will be associated now to relative variations: an eigenvalue larger than one will mean that it is the relative difference to the variable to the base value which will tend to infinite.

As most variables grow with time, this criterion for convergence will be less restrictive than the first. But the following question must be considered: must one associate to divergence a growing absolute variation or a relative one? For instance, let us suppose households benefit from a tax exemption, at two different periods in time. To compare the two, do they consider levels at constant prices, or the ratios to their present income? The truth is certainly between the two.

8.8.1.1 First example: a very simplified model

As an example, let use an extreme simplification of our usual model.

$$(1) I_t = a \cdot (Q_t - Q_{t-1}) + b$$

Firms invest to adapt their productive capacities to the evolution of demand.

$$(2) CO_t = c \cdot Q_t + d \cdot Q_{t-1} + e$$

Households consume a share of the present and previous production levels (introducing a lag makes the example more interesting).

$$(3) Q_t = I_t + CO_t + g_t$$

Production adapts immediately to demand, composed of investment, consumption and State demand.

Reasoning in variations, one gets:

$$\begin{pmatrix} \Delta I \\ \Delta CO \\ \Delta Q \end{pmatrix}_t = \begin{pmatrix} 0 & 0 & a \\ 0 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} \Delta I \\ \Delta CO \\ \Delta Q \end{pmatrix}_t + \begin{pmatrix} 0 & 0 & -a \\ 0 & 0 & c \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \Delta I \\ \Delta CO \\ \Delta Q \end{pmatrix}_{t-1} + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \Delta g_t$$

or

$$\begin{pmatrix} \Delta I \\ \Delta CO \\ \Delta Q \end{pmatrix}_t = 1/(1-a) \begin{pmatrix} 0 & 0 & (c-1) \\ 0 & 0 & c(c-a) \\ 0 & 0 & (c-a) \end{pmatrix} \begin{pmatrix} \Delta I \\ \Delta CO \\ \Delta Q \end{pmatrix}_{t-1} + 1/(1-a) \begin{pmatrix} a \\ 0 \\ 1 \end{pmatrix} \Delta g_t$$

This process contains a single lagged variable, and a single eigenvalue:

$$\lambda = (d-a)/(1-a-c)$$

and the multiplier at period k of a unitary change in State demand at period 0 will be:

$$1/(1-a-c) \cdot \lambda^k$$

At this level of simplification a formal interpretation is still possible. The eigenvalue adds two lagged dynamic effects:

- The sensitivity of consumption to past households income: d.
- The sensitivity of investments to the previous level of production, representative of capacity levels acquired earlier: -a.

These two effects are amplified by the multiplier mechanism, due to the present influence of investment and consumption on present production: $1 / (1 - a - c)$.

One can therefore logically suppose:

- That a, c and d are positive.
- That $0 < c + d < 1$: households consume a part of their additional revenue, itself lower than additional production.

Our purpose is not to theorize on these formulations. However, we can identify some special cases:

- $d + c = 1$ (unitary impact of Q on consumption); then the eigenvalue is 1, and while the dynamic process does not converge, nor does it diverge: the increase in consumption maintains the level of production in the next period, which stabilizes investment, which has no further impact on growth Q. A demand shock at period 1 will maintain its effects indefinitely.

In other words, the multiplying effect: $1/(1-a-c)$ is compensated in the next period by the reducing effect: $d-a$ or $(1-a-c) - (1-d-c)$ lower than $(1-a-c)$ if $d+c < 1$.

Or: with $c+d < 1$, the multiplying effect will spread to the two periods with an intensity lower than the initial shock. The process will converge.

- $c = 0$ (consumption independent of Q); then the eigenvalue is $-a/(1-a)$, and the dynamics converge if $a < 1/2$. Otherwise, getting back to the initial level of investment represents more than half the previous increase of production, and the multiplier more than doubles this change, which leads to an alternate diverging process.
- $a = 0$; then the eigenvalue is c , leading to a general convergence if $|c| \in]0, 1[$

More generally, when $0 < a < 1$ and $0 < c + d < 1$ the process will converge if:

$$c + d > 2a - 1$$

Which is a less restrictive condition than $a > 1/2$ (if $c=0$), because the positive effect of the consumption can then compensate the alternate divergence presented higher.

We can see that even for a very simple example the formal interpretation can present some complexity. For more realistic cases, one will have to resort to numerical techniques, interpreting results in terms of economic mechanisms by associating them to variables of the model.

For this one can (following a methodology introduced by Deleau and Malgrange (1978)) eliminate for each variable in turn its lagged influence (by deleting the associated column and line of A , then by re-computing the eigenvalues). In the most favorable case, a single eigenvalue will disappear, creating a natural association to the eliminated variable. In other cases research will be more difficult, and might need the simultaneous elimination of groups of variables, which one will associate to a group of eigenvalues of the same dimension.

Our example is too simple to present this method, as it contains only a single lagged influence (and therefore a single eigenvalue). But the analysis we have just made gives a basic illustration of the method. To extend the dynamic characteristics of the above model, one could have added an equation for capital and productive capacity.

8.8.1.2 Second example: the role of the Taylor rule

We shall now explain, on a very crude example, the role of the real element of the Taylor rule in our type of model.

Let us remind that the Taylor rule represents the behavior of a central bank which uses the interest rate to target inflation (and also reduce its volatility). See TAYLOR JB. (1993).

It will increase the rate if:

- Inflation is higher than the target (2% for the European Central Bank).
- Tensions appear on productive capacities (the “output gap”) signaling potential inflation.

This action should reduce demand (investment and consumption), activity and therefore inflation.

The original formula is:

$$IR_t = tx(p_t) + 0.5 \cdot (tx(p_t) - tx^*(p_t)) + 0.5 \cdot gap$$

With

- IR the interest rate
- P the price index (deflator of consumption?)
- $tx^*(p)$ the inflation target
- gap the output gap indicator, normally the relative difference between the actual and potential values of output.

Let us consider the following model, measured in real terms, a simplification of our example without external trade

$$(1) \dot{i}_t / k_t = 0.65 \dot{i}_{t-1} / k_{t-1} + 0.10[(q_t - q_{t-1}) / q_{t-1} + (ur_t - ur^*) / ur^*] + d + a(ur_t - ur^*)$$

$$(2) c_t = r_t q_t (1 - b(ur_t - ur^*))$$

$$(3) ur_t = q_t / (k_t p k_t)$$

$$(4) k_t = k_{t-1} (1 - dd_t) + i_{t-1}$$

$$(5) q_t = c_t + \dot{i}_t + g_t$$

with the endogenous variables:

c household consumption
i productive investment
k capital
q GDP
ur the rate of use of productive capacity

the exogenous variables:

dd the depreciation rate
g government consumption
pk the productivity of capital
r the ratio of consumption to GDP
ur* the target rate of use

For the Taylor rule, we have only considered the role of the output gap, which we have represented (this can be questioned) by the rate of use of capacities.

So we shall use the parameters:

- a the sensitivity of investment to the rate of use
- b the sensitivity of consumption to the rate of use

If we compute the dynamics of this model by differentiating its equations numerically,

$$\Delta y_t / y_t = A \Delta y_t / y_t + B \Delta y_{t-1} / y_{t-1} + C \Delta x_t / x_t$$

where A, B and C are relative Jacobians supposed constant with time, we get

$$\Delta y_t / y_t = (I - A)^{-1} (B \Delta y_{t-1} / y_{t-1} + C \Delta x_t / x_t)$$

To describe the dynamic process, we must now compute the eigenvalues of $(I - A)^{-1} B$, a 6x6 matrix.

But of course, independently from the values of a and b, we should get only two non-zero eigenvalues, as there are only two equations with lagged elements (and three such elements).

Starting with a=b=0, we get two complex eigenvalues, with

Modulus: .977 period: 18.8

Now if we put the value of a to 0.03, consistent with our Taylor rule (coefficients of -.06 in the investment equation, and .5 in the Taylor rule itself) we get:

Modulus: .872 period: 20.6

The period not much affected, but the convergence is much faster. After 10 periods an initial shock is reduced by 75% (instead of 21%).

Considering b (the role of the interest rate in consumption) we get only a small increase in the period, and in the speed of convergence.

In conclusion: the dynamics are almost fully explained by the real elements of capital formation, including the real interest rate.

The elements allowing to product the above study can be very simply implemented using EViews.

8.8.1.3 Steady state growth paths

Let us go further in determining the conditions for the existence of a path of model solutions, such that the whole set of variables grows at a constant rate (a « steady state growth path »). The application of these conditions to formulations of the model will lead, by editing some equations, to the production of a long-term model, with potentially very different causalities compared to the original model. One will study also the stability of this path, by introducing an initial gap through a one-time shock, and observing if the long term dynamics leads back to the path or away from it (again, reasoning in absolute or relative terms).

These techniques can be clarified through our simple model. Let us recall its formulations (in EViews format):

```

CAP = pk * K(-1)

Q + M = FD + X

UR = Q / CAP

IC = tc * Q

CI / Q(-1) = - 0.157 * @pch(Q) - 0.0464 * @pch(Q(-1)) + ec_ci

I / K(-1) = 0.825 * I(-1) / K(-2) + 0.0279 * UR + 0.152 * .25 * Q / Q(-4) - 0.0525 + ec_i

Log(PRLE_T) = 2.591 + 0.0279 * (t - 2002) + 0.0215 * (t - t1) * (t<t1) + 0.0145 * (t - t2) * (t<t2)

LED = Q / PRLE_T

DLog(LE) = 0.422 * DLog(LED) + 0.455 * Log(LED(-1) / LE(-1)) + 0.000731 + ec_le

LT = LE + lg

RHI = wr * LT + r_rhiq * Q

IH = r_ih * RHI

CO = RHI * (1 - sr)

FD = CO + I + gd + CI + IH

TD = FD + IC

RES_M = Log(M / TD) + 1.322 * Log(UR) + 0.420 * Log(compm) + 0.0126 * (@TREND(60:1) * (t<=2002)
+ @ELEM(@TREND(60:1) , "2002S2") * (t>2002))

DLog(M) = 1.200 * DLog(TD) + 0.282 * DLog(UR) - 0.212 * RES_M(- 1) - 0.629 + ec_m

```

$$\text{RES_X} = \text{Log}(X / \text{wd}) + 0.686 * \text{Log}(\text{UR}) - 4.87 \text{ E-05} * (@\text{TREND}(60:1) * (\text{T} \leq 2002) + @\text{ELEM}(@\text{TREND}(60:1), "2002S2") * (\text{T} > 2002))$$

$$\text{DLog}(X) = 0.940 * \text{DLog}(\text{wd}) - 0.0129 - 0.195 * \text{RES_X}(-1) + \text{ec_x}$$

$$K = K(-1) * (1 - \text{dr}) + I$$

If we call

- txq the growth rate of quantities
- txn the growth rate of populations.

We get:

$$1. \text{CAP} = \text{pk} * K / (1 + \text{txq})$$

$$2. Q + M = \text{FD} + X$$

$$3. \text{UR} = Q / \text{CAP}$$

$$4. \text{IC} = \text{tc} * Q$$

$$5. \text{CI} / Q / (1 + \text{txq}) = -0.157 * \text{txq} - 0.0464 * \text{txq} + \text{ec_ci}$$

$$6. I / K / (1 + \text{txq}) = 0.825 * I / K / (1 + \text{txq}) + 0.0279 * \text{UR} + 0.152 * .25 * ((1 + \text{txq})^4 - 1) - 0.0525 + \text{EC_I}$$

$$7. \text{Log}(\text{PRLE_T}) = 2.591 + 0.0279 * (t - 2002) + 0.0215 * (t - t1) * (t < t1) + 0.0145 * (t - t2) * (t < t2)$$

$$8. \text{LED} = Q / \text{PRLE_T}$$

$$9. \text{Log}(1 + \text{txn}) = 0.422 * \text{Log}(1 + \text{txn}) + 0.455 * \text{Log}(\text{LED} / \text{LE}) + 0.000731 + \text{ec_le}$$

$$10. \text{LT} = \text{LE} + \text{lg}$$

$$11. \text{RHI} = \text{wr} * \text{LT} + \text{r_rhiq} * Q$$

$$12. \text{IH} = \text{r_ih} * \text{H} * \text{RHI}$$

$$13. \text{CO} = \text{RHI} * (1 - \text{sr})$$

$$14. \text{FD} = \text{CO} + I + \text{gd} + \text{CI} + \text{IH}$$

$$15. \text{TD} = \text{FD} + \text{CI}$$

$$16. \text{RES_M} = \text{Log}(M / \text{TD}) + 1.504 * \text{Log}(\text{UR}) + 0.483 * \text{Log}(\text{COMPM}) + 0.0124 * (@\text{TREND}(60:1) * (\text{T} \leq 2002) + @\text{ELEM}(@\text{TREND}(60:1), "2002S2") * (\text{T} > 2002))$$

$$17. \text{Log}(1+\text{txq}) = 1.200 * \text{Log}(1+\text{txq}) - 0.282 * \text{RES_M} - 0.629 + \text{EC_M}$$

$$18. \text{RES_X} = \text{Log}(X / \text{wd}) + 0.686 * \text{Log}(\text{UR}) - 4.87 \text{ E-}05 * (@\text{TREND}(60:1) * (\text{T} \leq 2002) + @\text{ELEM}(@\text{TREND}(60:1), "2002S2") * (\text{T} > 2002))$$

$$19. \text{Log}(1+\text{txq}) = 0.940 * \text{Log}(1+\text{txq}) - 0.0129 - 0.195 * \text{RES_X} + \text{ec_x}$$

$$20. K = K / (1+\text{txq}) * (1 - \text{dr}) + I$$

First we can observe that in each equation, provided that all the explanatory elements grow at the theoretical rate, the explained element will also. In particular, all additive terms should grow, taking into account the variables which enter them, at this rate.

This means that there exists a consistent solution in which all elements grow at the theoretical rate. As the model has only one solution, this is the solution it should reach, provided the error correction mechanisms lead in the right direction (from the coefficients, we can suppose they do).

We can also build the long term model, which follow a logic somewhat different from the dynamic one, as disequilibria have been stabilized.

- From (20) we get I/K as the rate which allows to adapt capital to both growth and depreciation:

$$I/K = 1 - 1 / (1+\text{txq}) * (1 - \text{dr})$$

- From (6) we get UR depending on the speed of adaptation. Even if adaptation was immediate (dynamic homogeneity) UR would still depend on txq, as a higher expected growth calls for an effort in creating additional capital, for use at the next period:

$$I / K / (1+\text{txq}) = 0.825 * I / K / (1+\text{txq}) + 0.0279 * \text{UR} + 0.152 * .25 * ((1+\text{txq})^4 - 1) - 0.0525 + \text{EC_I}$$

- From (18) we get RES_X.
- From (16) we get RES_M.
- From (9) we get LED/LE.

If adaptation was immediate the three would not depend on txq or txn. This would also be the case for UR if the coefficients complied exactly with the error correction framework.

Here the gap will increase with the growth rate, and decrease with the speed of adaptation.

- From RES_X, UR and (17) we get X/wd and X :
- From RES_M, UR and (15) we get $M/(FD+IC) = M/(FD+tc*Q)$
- From UR, (1) and (3) we get Q/K ; from (6) we get I/Q
- From (5) we get CI/Q
- From (8) and (9) we get LE/Q
- From (10) and (11) we get RHI as a term proportional to Q , plus $wr*lg$
- From (12) and (13) we get CO and IH as proportional to RHI, thus to Q and $wr*lg$
- From (14) we get FD as proportional to Q plus a term proportional to $wr*lg$, plus gd
- This gives M/Q
- This gives Q from (2)
- And all the other elements.

Finally, all the terms in the supply – demand equilibrium are defined as proportional to Q , except for:

- Household consumption and investment which are partially proportional to government employment.
- Government demand which is exogenous¹¹⁷.

The sum of these elements makes demand partially exogenous.

- Exports which are proportional to world demand.

We can see that the logic of the model causalities is quite different from the dynamic model. In particular,

- The rate of use depends on the growth rate of the economy and the depreciation rate.
- The gap between targets and actual values too (with UR as an eventual intermediary).
- Knowing this gap we can define all the real elements as a share of GDP, except for exports and an exogenous element of demand. Let us call it dx .

The supply – demand equilibrium

$$Q + M = FD + X$$

becomes:

¹¹⁷ Of course the assumptions will give them values proportional to Q .

$$Q + a \cdot (b \cdot Q + dx) = b \cdot Q + dx + c \cdot wd$$

$$Q = ((1 - a) \cdot dx + wd) / (1 - (1 - a) \cdot b)$$

With a and b constant elements depending on exogenous assumptions and the growth rate txq (and also txn).

This means that if

- gd was defined as a constant share of demand
- lg was defined as a constant share to LE (or of $LT = LE + lg$)

(which seem logical and even necessary assumptions),

the exogenous world demand level would define all elements in the local economy (as world inflation will drive the local one with a more complex model).

8.8.2 THE CASE OF ERROR CORRECTION MODELS: A SIMPLE EXAMPLE

Let us now show how an error correction formulation makes the interpretation of model dynamics both easier and more meaningful.

Let us introduce the following model

$$(1) CO_t = a \cdot Q_t$$

$$(2) Q_t = CO_t + I_t + g_t$$

$$(3) K_t = K_{t-1} \cdot (1 - dr_t) + I_t$$

$$(4) \Delta K_t = \alpha \cdot b \cdot \Delta Q_t + \beta \cdot (b \cdot Q_{t-1} - K_{t-1})$$

We have defined an error correction model, where the target consists in making K proportional to Q (this equates to a constant rate of use of capacities if one supposes constant the productivity of capital).

Let us consider the dynamics of this model, along the lines of the eigenvalue analysis presented earlier.

By eliminating CO and I , we can reduce the system into:

$$(1) Q_t = a \cdot Q_t + K_t - K_{t-1} \cdot (1 - dr_t) + g_t$$

$$(2) K_t = \alpha \cdot b \cdot Q_t + (\beta - \alpha) \cdot b \cdot Q_{t-1} + (1 - \beta) \cdot K_{t-1}$$

$$\begin{pmatrix} \Delta Q_t \\ \Delta K_t \end{pmatrix} = \begin{pmatrix} a & 1 \\ \alpha b & 0 \end{pmatrix} \begin{pmatrix} \Delta Q_t \\ \Delta K_t \end{pmatrix} + \begin{pmatrix} 0 & -(1-d) \\ b(\beta - \alpha) & 1 - \beta \end{pmatrix} \begin{pmatrix} \Delta Q_{t-1} \\ \Delta K_{t-1} \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} g$$

We can apply to this system the same technique as before, with normally two eigenvalues. The process might be a little too complex for this simple formulation. We shall simplify it by supposing that the speed of adaptation is the same for a new gap, due to a change in the target, and a pre-existing gap:

$$\alpha = \beta$$

The system becomes

$$\begin{pmatrix} \Delta Q_t \\ \Delta K_t \end{pmatrix} = \begin{pmatrix} 1-a & -1 \\ 1-\alpha b & 1 \end{pmatrix}^{-1} \begin{pmatrix} 0 & -(1-d) \\ 0 & 1-\alpha \end{pmatrix} \begin{pmatrix} \Delta Q_{t-1} \\ \Delta K_{t-1} \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} g_t$$

with one eigenvalue:

$$\lambda = 1 - \alpha + \alpha(d - \alpha) / (1 - a - b\alpha)$$

Let us interpret this element.

If $d = \alpha$, the eigenvalue is $(1 - \alpha)$. This means an initial gap between the variable and its target will close with a reason $1 - \alpha$. The reason for this is clear. If we consider the equation system, the capital will depreciate at the speed d , which is exactly the rate at which the firms « want » it to decrease if a gap has appeared in previous periods, due for instance to a shock on g . The ex-ante variation will be the same as the desired one, and no additional mechanism will appear.

On the contrary, if d is higher than α for instance, the natural decrease in capital will be higher than the desired one. Firms will have to invest the difference, leading to an increase in capital made larger by the presence of the multiplier.

But for the process to diverge, the difference between d and α must be large, or the multiplier effect very important (which would lead to problems with any model).

It might look strange that an ex-ante decrease in capital (through a higher d) could lead to an ex-post increase; This is because the behavior of firms as to capital is set by (2), taking into account the variation of production but not the additional depreciation rate, which is compensated automatically through higher investment (thus production, and investment again).

Let us now consider the long-term model.

It is now clear that all variables: CO, Q, I, g, CAP grow at the same rate on the steady state growth path. This rate (let us call it q) is set by « g ».

The long-term model is

$$(1) CO = a \cdot Q$$

$$(2) Q = CO + I + g$$

$$(3) K = K \cdot (1 - dr) / (1 + q) + I$$

$$(4) K \cdot (q + \beta) / (1 + q) = (\alpha \cdot q + \beta) \cdot b \cdot Q / (1 + q)$$

or

$$(1) CO = a \cdot Q$$

$$(2) Q = CO + I + g$$

$$(3) I / K = q + dr$$

$$(4) K / Q = b \cdot (\alpha \cdot q + \beta) / (q + \beta)$$

We observe that none of the causal explanations of the original model has changed. The ratio of production to capital is fixed, actually based on the target in the original formulation. And investment maintains this constraint, by leading to an implicit evolution of productive capacity equal to the growth rate (taking into account depreciation).

But one also can see that the target is reached only in special conditions. This is a standard characteristic for an error correction model.

Equation (4) shows that the cases are:

$$\alpha = 1 \text{ (dynamic homogeneity, gaps are closed immediately)}$$

$$q = 0 \text{ (the target does not move)}$$

This is true for any starting values, including those which meet $K / Q = b$: if q or α are not zero, a gap will appear, which will never close.

We shall stop here, as our purpose was only to illustrate the process, and show that error correction models make the dynamics more interesting, as well as easier to study.

CHAPTER 9: USING MODELS

In chapter 2 we began to present modelling applications. We are going now to be more precise, concentrating on practical aspects.

9.1 OPERATIONAL DIAGNOSES

The first application we quoted, and the most natural apparently, is to request from the model a realistic diagnosis of economic problems. This diagnosis can lean on scenarios or shocks.

The originator of this type of study can be:

- A governmental organization, such as the local Ministry of Finance, which wants to get information on the future evolution of the economy it manages, or the efficiency of the policy decisions it contemplates.
- An international institution, such as the IMF or the OECD, or a Non-Governmental institute, which wants to assess the future of the world economy.
- An academic research unit or individual, which wants to study the impact of policy decisions or structural changes, according to a given theoretical model.
- A private firm dealing in economic services, which provides forecasts and studies to subscribers, either as a publication or a separate study.

9.1.1 SCENARIOS AND THEIR DIFFERENT TYPES

Two types of scenario may be distinguished.

- **Trend forecasts**, which will use the most probable assumptions (in the case of a unique scenario), or a set of possible assumptions, covering the scope of foreseeable evolutions.

Assumptions will often be developed by experts from outside the area of model building. Let us take as an example: the forecasts of MESANGE, a French macroeconomic model considering policy decisions as exogenous, managed by the French Ministry of Finance, in two of its Directorates: at INSEE, the National Institute for Statistics and Economic Studies at the General Direction of Treasury and Economic Policy (DGTPE).

Variables describing foreign environment are valued through simulations of a set of tools:

- The MZE (Modèle Zone Euro or EuroZone model) model of the CEPII (a French economic institution specialized in international studies), managed in the same institutions.
- The NIGEM world model, managed and leased by the British NIESR (National Institute for Economics and Social Research).
- Results from economic forecasts of OECD, the UN and the IMF.
- Numerical evaluations made by experts in international economics and trade, without using any kind of model.

- The observation of international developments, like the Greek crisis, and the assessment of potential evolutions.

The government decisions are specified following discussions with the DGTPPE and especially the people which produce the numerical evaluation of next year's budget.

As for assumptions on the evolution of the population and its structure, they will use the results of studies produced by INSEE or Ministry of Labor demographers.

This type of scenario will give indications of the most probable future evolution, or on the range of foreseeable evolutions, to the best knowledge of the available experts.

Production of the final forecast will require a large number of model simulations, accompanied by meetings and discussions. It lasts usually several weeks (let us say 2 to 4). The improved speed of computation helps, but not too much, as it is mostly used to improve the quality of the results.

- **Normative forecasts**, where results are asked to meet a certain number of conditions. These constraints are set before the development of assumptions.

This type of scenario describes a situation in which efforts are made, relative to the foreseeable evolution of the economy, to reach some targets (such as meeting the criteria for joining the European Monetary Union, usually called the Maastricht criteria). It will essentially give information on the extent of expected difficulties, and efforts required, associated to a possible path.

In general the client of this type of study will also be the State, or a public institution. In cooperation with the model managers, it will set the constraints, and the policy elements used to meet them¹¹⁸. But non-governmental institutions and researchers can also be interested in this type of issue, perhaps more than the production of a forecast as it involves more economic reasoning.

If precise goals are set (such as a budget deficit of 3 GDP points), one can work either by a sequence of approximations on the normal model, or by modifying model specifications by an explicit reversal of causalities. This last technique introduces two problems:

- A formal one: for the model to preserve a unique solution, the number of endogenous elements which are now given a set value must be equal to the number of instruments that will allow to reach them. For example, if one wants to balance both the commercial balance and the State budget, it will be necessary to free two decision variables and two only, to avoid either impossibility or an indetermination¹¹⁹.

¹¹⁸ However in some cases forecasters will also include in the potential set of instruments some assumptions which are not manageable by the State, such as an exogenous change on the behavior of agents, or structural parameters such as the productivity of factors. This is not unacceptable, provided the option is clearly displayed in the associated publication. The initial question was then: how should the behavior of agents, or structural parameters, have to change in order to reach a specific objective?

¹¹⁹ Except if the number of instruments is larger than targets, but limits are set on their variations, as well as a priority order. Then each instrument will be used in turn, switching to a new one if the constraint has been reached. In

- A technical one: often it will be too complex to produce an identified model which takes into account this reversal explicitly, and if the solving algorithm requests such a formulation, it will be necessary to revert to the other technique, determining the solution by a sequence of approximations. Let us suppose for instance that we want to balance foreign trade by increasing subsidies on investment. To transform the model, one will have to identify a chain of causalities leading from subsidies to foreign trade (which can be quite long) and revert each one in turn, through the associated equation. To this obviously difficult if not unfeasible technique, one can prefer a numerical, iterative one: computing outside the model an approximation of the necessary subsidy amount, using it to solve the standard model, then correcting the change from the observation of the remaining error. This sequence of approximations should decrease to an acceptable value.

At the end of the book, we propose an alternate method, which allows solving the model for given values of a number of endogenous variables, provided the same number of exogenous variables are released. This method does not work all the time, but if it does the solution is quasi-immediate, and can be applied under EViews without accessing the model code. The program itself is provided, and can be adapted in a few minutes to any existing model managed under EViews.

But in any case, results cannot be taken for more than what they are, the economic equilibrium associated to assumptions deviating from normal evolution, whose probability of occurrence cannot be judged. To earn the right to use this method, one has to display clearly and probably justify the options he has chosen.

Under those restrictions, one is not limited in the elements he can use. For instance one can modify the estimation residual in the investment equation to measure by how much private firms should increase it (all things being equal) for the trade balance to get back to equilibrium in the long run. The less controlled the instruments, the less the simulation can be considered as a policy result. But this does not make the results (and the observation of causal sequences) less interesting from a theoretical point of view.

9.1.2 MANAGING ACTUAL FORECASTS: MANAGING THE RESIDUALS

On the estimation period, the sensible option for residuals is of course to put them as the difference between the actual values and the one given by the estimated formula. Now we have to set values to them, starting with the first forecast period.

Two main (and extreme) options are available.

- Putting them to zero (in principle, the value with the highest probability).
- Keeping the last value in the estimation period (the most recent information we have).

In other words: should we rely on the recent past, or summarize all the past information we have, giving equal weight to each piece, independently from its age?

Each choice can be dangerous:

The first will refuse to take into account recent elements, like a feature which appeared only in the last periods, and could not be evidenced due to its shortness (otherwise it should appear in the actual formula). Also, it will produce

practice we will be left with as many unknowns as constraints. One can also create a synthetic instrument by applying predefined weights to a larger set.

shocks on the initial growth rates, which will be difficult to explain. For instance if the last residual on household consumption was 2%, a structural decrease of 1% (taking only into account the formula and its variables, and excluding the residual) will be transformed into a growth of 1%.

The second will violate the zero mean assumption, and suppose that a possibly high shock will be maintained for the whole period of the forecast.

In our opinion, no option clearly dominates under all circumstances. One should first consider the profile of the last residuals.

First, their size: obviously if the last residual is close to zero, the problem is minor.

This means of course that one will favor estimations with a low last residual, even if it is not really politically correct.

If it is large, one should look at the immediately preceding values. If they are too, there is clearly a problem with the equation, and something should be done to reduce it. This is also true if they are smaller but growing: a trend (or a variable with a trend) might be missing. With lower values, the above applies too, of course.

Maybe the best option is to decide that the last residual is indeed going to last for a while, but with an intensity decreasing probably to zero.

But in any case one should first try to understand the reason for the value. This might make things clear as to the way to manage it.

9.1.3 MANAGING ACTUAL FORECASTS: THE TARGETING OF SIMULATIONS

In the normative case, one has to manipulate the exogenous in order to reach a certain solution. But this happens also in trend forecasts. One cannot expect the model to give spontaneously an image of the future consistent with what we expect, especially in the short run where we already have some (partial) information about what is happening.

A targeting procedure has to be applied, which may be automated, by the application of specific techniques.

9.1.3.1 on the past

We might want the base simulation to reproduce exactly the known past. This operation is not futile: it is advisable to start the analytic shocks from the historical value, eliminating second-order errors.

The technique is simple: we apply to behavioral equations the estimation residual. Each individual equation will provide historical values, and the global absence of error makes historical values the solution of the system.

9.1.3.2 in forecasts

In forecasts this method will seek to give some variables values belonging to a certain interval, through the choice of assumptions and of estimation residuals.

In the short term the goal will be to adapt results to already known values (or known intervals) of some elements of the model. Thus at the end of the year, we know already global growth and some of its components, global inflation, unemployment and the trade balance, and the forecast should provide their values, if only for show. Reproducing wrongly the already known past casts strong doubts on the forecasting quality of the model.

In the medium term one will seek to keep the general evolution within limits judged reasonable, in particular meeting some general conditions. Thus one will not accept too large a gap between the country described by the model and the rest of the world, as to inflation and growth.

But the fragmentary character of the information will prohibit the use of the method given in the previous paragraph. The model builder will work in general by iterations, sometimes relaxing some constraints if they prove difficult to meet. Or he can try the specific algorithm we are proposing.

9.1.4 DIFFERENT TYPES OF SHOCKS

We are going now to categorize the different types of shocks.

9.1.4.1 Analytic and complex shocks

We have already presented analytical shocks, used to interpret and validate model properties, by comparison with economic theory and properties of the other models. This set of tests also allows the interpretation of model responses, making the later interpretation of more complex shocks easier. Actually, operational modelling teams always produce a “shocks catalogue”¹²⁰ giving the responses of the model to changes in the main assumptions. Given the numerical quasi-linearity of models, one will get a realistic enough approximation of a complex set of changes by combining linearly the results of these individual shocks, using a spreadsheet for instance or more simply an EViews program (combining in one page the contents of several “single shock” pages with the same structure).

Going even further, a few months spent working with a model will give its user the ability to anticipate (at least roughly) its reactions to given changes, and this without any simulation. He will already have a basic “shocks catalogue” available in his mind.

But if the intensity of the shock changes with time (for instance to take into account a gradual loosening of quotas in a trade agreement) we face a larger problem. As we have already stated, most macroeconomic models are roughly linear at a given period, but their Jacobian is not stable over time, mostly because of the growing role of external trade. Of course, the slower the change, the larger the problem. But in the above case, the expected change in the structure of exports and clients should call for a specific forecast anyway.

One will also use this type of shock for operational studies

- If requested diagnoses are associated to simple decisions.
- If one seeks to decompose the effects of more complex policies.

But to get the exact consequences of complex shocks, one will have to apply them simultaneously to the base simulation, with a true operational purpose.

9.1.4.2 External and policy shocks

All these shocks fall also in three categories:

- External environment shocks, which seek to measure the consequences of non-controllable events. For single country model, the question asked is then: *what are the consequences for my national economy of a change in external conditions?*

Two subcategories can be considered.

- Shocks on foreign assumptions, associated for single country models with the situation in other countries, or elements defined at the world level. A typical shock would represent a depression in Asia, or an increase in the price of oil, followed possibly by a decrease in world demand and an increase in world inflation. Technically, these elements should not be exogenous, as the modelled country or set of countries have some impact on the economy of the rest of the world. But the cost of modelling this feedback is generally much too high (it could involve the production of a world model) compared to the limited improvement of model properties. Of course, the approximation (and the associated error) increases with the size of the modelled country. If it is reasonable to consider the world outside Denmark as exogenous to this country's events (we have taken voluntarily a country outside the Euro Zone), this is not so true for Germany (a large country which shares a currency with a significant part of the world) and even less for the whole Euro Zone itself.
- Shocks on local uncontrolled variables. Examples include technical progress, population, or climate. Again, one can argue that each of these elements is not completely exogenous: economic growth accelerates technical progress, the birth rate changes with wealth and health of households, growth brings pollution and climate changes. But the associated mechanisms are difficult to establish, formally complex and limited in their short and medium term impact. Therefore the cost of taking them into consideration is generally much higher than the benefits, except for very specific fields.

The only exception concerns long term studies. But although it is true that we have solved our model over a very long period, the goal was only to observe its properties from a technical point of view, and not to produce an actual forecast. A true long term model would require a rather different framework, with new formulations relying more on a formal economic reasoning than on actual estimations, as the sample is generally too small to evidence statistically long term evolutions.

- Economic policy shocks, where interest lies in the consequences of choices on institutional instruments, decision variables if one considers that the State (or its impersonator) is the client of the study. The question is now: what will happen if I take the following decisions? And when these shocks are given an objective (for example: to reduce unemployment by a given margin without eroding too much the other criteria), they will call for an iterative targeting process analogous to that of normative scenarios.

These elements are not exogenous to the country modeled, as the decisions taken by the Government are largely based on the local economic situation. But the producer of the study wants to be allowed to manage the decisions, and not to see the model do it in his place. Moreover, the State has few goals (unemployment, inflation, budget and foreign deficits, growth) and a lot of potential instruments. Evidencing statistically the link between decisions on the instruments and the present state of the goals is very difficult if not impossible. The few studies on that subject do not give in general reliable answers.

- The two approaches may also be combined however, if we look for the decisions which allow to face (or take advantage of) a given evolution of external conditions. One could for instance determine how the State can use his own tools to fight losses in trade balance coming from a given depression in world demand. Or an oil producing country can decide on the use of the additional revenue coming from an increase in the price of a barrel. This type of exercise is parent again to normative scenarios.

It will be necessary, in all these cases, to care for the presence of formal connections between exogenous variables. For example, one cannot suppose an increase of the price of oil ¹²¹without taking into account its inflationary effects on the production price in non-oil-producing countries, or the depression it will cause in world economy. Similarly it seems difficult to increase the number of civil servants without considering an increase in State consumption (heating, lighting, and various equipment) and investments (offices, hospitals, schools) according to the type of positions created.

A very important issue is the impact of public spending (investment, consumption or hiring of civil servants, separately or together) on local supply. Of course, they modify demand (new civil servants increase household revenue and reduce unemployment). But they should also have a positive effect on potential production, for many reasons. To give just two examples:

- Creating or improving roads, or the lighting of streets, increases the productivity of the transportation sector, and of the whole economy (agricultural goods can be delivered fresh over longer distances, attrition from accidents will diminish...). Some completely new productions can become feasible (access to water can allow to produce new kinds of vegetables).
- Buying educational books, or hiring more teachers, will improve students' skills and their productivity when they join the workforce.

Unfortunately, the influence of these elements is very difficult to quantify. It is left to the discretion of the modeler, who more often than not keeps it null. A more reasonable option would certainly be to suppose that the effect on global productive capacity would be the same as actual productive investment, but with a longer delay in its impact. This would mean that the main difference in behavior between the public and private sectors would be the lower concern for quick returns, and the more widespread effect.

- Finally, one might want to modify structural parameters, such as the productivity of factors, or the depreciation rate. Falling in the same category will be the residuals on some behavioral equations: the ratio of household consumption to revenue can be modified ex-ante, as well as the sharing of value added between firms and workers. According to the formulation, the error-correcting mechanisms and multipliers can lead to increased or decreased ex-post changes¹²².

Of course, these shocks fall in the first category concerning their feasibility, as external uncontrolled elements.

9.1.5 THE FORECAST: TECHNICAL ASPECTS

¹²¹ Otherwise the modeled country will be unduly subject to losses in competitiveness coming from the increase of its own inflation.

¹²² The price-wage loop will increase the deflationary effect of a wage hike, but the impact of additional demand on imports will be reduced by capacity adaptation.

It would appear that once a coherent model has been developed, with its global precision and economic properties validated by the previous tests, the only efforts remaining for producing a correct forecast would focus on the definition of assumptions. It is unfortunately not so, and a good deal of know-how will be necessary for the model builder to produce acceptable forecasts, bearing especially on the following points:

- How to judge the likely character of the forecast.
- How to integrate evolutions of the endogenous elements, known but not yet precisely quantified (concerning especially the first forecasted periods). This goes from partially known values (as the recent monthly evolution of unemployment for a quarterly simulation) to vaguer elements, such as the anticipations provided by non-modelling experts. At the time of writing (February 2012) forecasting the European economy calls for a lot of out-of-model
- Once anomalies concerning the near future are corrected, how these corrections should be taken into account for next periods. For example, if one decides that estimates of exports have to be decreased in the first period compared to the value given by the equation, should this gap be maintained in the medium term?
- How to interpret results, beyond the simple observation of figures.

In addition, some specific difficulties can appear when generating assumptions:

- In general, they need the intervention of external expertise, as the macro-economic model builder cannot be fully competent in the whole set of areas concerned. In some cases, it even will be necessary to call for another model to obtain coherent assumptions (for example a world model will provide assumptions on international environment).
- External information, even quantified, is often presented neither in the units nor in the categories used by the model (like trade agreements or the results of wage negotiations).
- The transformation of some qualitative information (climate, strikes, tensions, expectations.....) into numerical values is even more complex.

9.1.6 CHANGING MODEL SPECIFICATIONS

9.1.6.1 The reasons

Normally, the modeller should make its forecasts on the same model he has estimated. To proceed differently, there should be some acceptable reason.

One can think of four cases:

1. It is known that a behavior has changed, in a given way: for instance the Central Bank has decided to change its method for inflation targeting, or it has just become independent and can decide on the interest rate. Or a new government has just decided to stabilize the social security deficit, putting the burden on the contributions of firms and workers, in equal shares.

2. A behavior was present from the first but we had not enough data, or the variations of the explanatory element were too small to support it. For instance, the country faced no supply problem, but now capacities are growing slower than demand and disequilibrium has to be taken into account¹²³.
3. A behavior appeared late in the sample period, and there were not enough periods to support it. For instance, a new law introduced part-time work, with a strong impact on the employment process.
4. It is known that the economy of the future will follow different behaviors compared to the past. This can apply to a country undergoing transition.

In the first case, the behavior is supposed to be known. The associated equation should be introduced, for instance a Taylor rule.

The second case is more difficult to solve. It applies for instance to a period for which inflation was flat, while in the future we can expect it to fluctuate much more. Or local inflation was similar to that of partner countries (stabilizing competitiveness) while differences are starting to appear.

In this case there is no other way than taking values either from economic theory, measurement on other countries and the observation of resulting properties.

The last two cases can be generally associated with:

- Developing countries
- Transition countries (some countries can be both like China or Vietnam, and maybe Romania and Bulgaria).

Identifying transition countries

In transition countries, the behavior of all agents has changed as market economy mechanisms have appeared. But although the technical transition could be quite fast (like German reunification) the transition process has been generally much slower.

One can separate two cases:

- The Central and Eastern Europe countries, with the addition of Russia and maybe the former USSR republics.
- Asia: mostly China and Vietnam, possibly Laos and Cambodia.

The difference of course is that the latter are still under Communist regimes, and that they can be considered as developing countries, which is also probably the case of the less advanced elements of the first set, like Bulgaria, Romania and Ukraine¹²⁴.

Actually in our opinion, if we consider the present situation, the type of regime is not the most important, as for the most part China and Vietnam behave economically like market economies (perhaps more than present Russia). The only issue is the date at which the transition can be considered as advanced enough for the economy to be modelled in this

¹²³ It was present from the first but had no opportunity to play a role.

¹²⁴ Although the appreciation is quite dependent on criteria (look at China for instance).

way. Officially, the date was 1978 for China, and 1986 for Vietnam. But the process was quite gradual, and one should consider delaying the application of a structural model by at least 10 years.

For the Eastern bloc, the transition started at the beginning of the nineties, and can be considered to have proceeded much faster, in particular because the economy of these countries had been following western standards (in framework and level) 50 years before. This means modelling could start in 1994-1995, with additional elements:

- The statistical system took some time in changing methods, in particular from the Social Accounting Matrix system to western-type National Accounts (like the UN defined SCN95).
- Now almost all of these countries provide quarterly accounts. This multiplies the number of observations by four, which is crucial in this case (although the problem is decreasing with time). Unfortunately, the gain in information is much lower.
- Some specific events took place during the period, like the partition of Czechoslovakia in 1993, or joining the European Union (even the Eurozone).

Concerning the data

Now several situations can appear:

- We now have enough data to estimate equations. This is for instance the case of Poland, where data is available from 1995 to 2011 (68 observations), using the same methodology, and with behaviors which can be considered (with some doubt nevertheless) as typical of market economy.

But we have also to consider the last two cases:

- In case 4, the equation has to be entirely calibrated, with values taken either from economic theory, measurement on other countries and the observation of resulting properties.
- In case 3 also, except that some additional information can be obtained from estimations. But it should not be used alone, even if all statistics look good (unless we get extremely good statistics on a sample not too small).
- In any case, if the influence of an element (such as profitability on capital or unemployment on the wage rate) is considered to be growing in the present period, the associated coefficient should also show a growth in forecasts, probably reduced with time to converge to a constant. This process (initial level of the increase, speed of convergence) is the responsibility of the modeler, helped by the observation of model properties.

9.1.6.2 The technique

We can use exactly the same technique as for testing model properties over the future, except that the role of changes in parameter values and formulations will increase.

9.1.7 OPTIMAL CONTROL

This technique consists in computing the values of instruments which gets model variables as close as possible to a given objective function, under constraint of the model. Compared to the reversal of causalities, the model user will set here a unique target, normally unreachable, formalized in most cases by a quadratic function of one or several variables.

Then he will apply a maximization algorithm under constraints (equations of the model and bounds on the evolution of instruments). This will have to be done outside the modelling package (while spreadsheets provide a simplified version of this technique).

This technique has been quite popular in the fifties and sixties in western countries like France (managed by the Planning Agency) and remained so in the socialist countries for a longer time¹²⁵. Now they have been more or less abandoned, and the remaining applications limited to economic research.

9.2 TEACHING WITH MODELS

The associated models will allow educational or scientific applications. Diagnoses can be observed for themselves, as the quantified illustration of some theories. But their small size makes them efficient for the application of specific validation techniques, useful especially if they have been designed to reproduce the properties of a larger operational model.

The simplicity of these models gives them the natural role of a learning instrument. According to the public and the goal, different tools may be used.

- Extremely simplified representations, based on formal examples and fixed parameters, clarify theoretical frameworks (Solow model, IS-LM model, Mundell-Fleming Fair model...). By formalizing them on a numerical example and solving them under various assumptions, one can better appreciate their properties (the simplicity of these models often allows a graphical display of their equations as dynamic curves). Courses in macroeconomics are now often accompanied by computer programs running these representations.
- More applied models, reproducing as faithfully as possible the properties of larger models used for operational diagnoses. Rather than illustrating an economic theory, the goal here is to put together a limited number of elementary mechanisms, evidencing the main interactions governing the economic equilibrium. This will generally be done through shocks, concentrating on the impact of the different economic policy instruments and the orders of magnitude of their effects.

In this last case, even if the precision of these models remains limited, producing acceptable diagnoses requires to apply them to real cases. Hence they generally will be based on real data, and often estimated econometrically. And to make the interpretation easier, one will generally process current issues: for example measuring consequences of a decrease in work duration, for the French economy of 1998.

9.3 PRESENTATION OF RESULTS

9.3.1 GENERAL ISSUES

The role played in the efficiency of the simulation process by the quality of results presentation is often underestimated. To researchers, it represents often a minor task, to which they are not very well prepared. Nevertheless, this element is often a necessary condition to capture the attention of the audience and insure the success of a study. While a

¹²⁵ Not surprisingly, Russia has produced the best theoreticians in this field (like Pontryagin) and in the connected field of matrix theory (like Gantmacher

mediocre but well-presented work will sometimes have some success, a work of quality but badly presented will generally fail in getting any audience, by discouraging its potential listeners or readers, including the referees¹²⁶.

A good presentation has to be:

- Error free (this is more difficult than it seems). A single visible error will cast doubt on the whole set of results.
- Explicit (the meaning of figures presented has to be evident, in particular their classification).
- Not too long.
- Pleasing to the eye (not too dense), but not too flashy (one must not think the author is trying to hide problems, or has diverted time from his research).
- Clearly synthesizing the main teachings, in the right order.
- Allowing fast access to a particular element of information.
- Adapted both to the nature of the work and to its public (for different audiences, different presentations can be produced for the same work¹²⁷).

The two main elements of any presentation are tables and graphs, and one must keep in mind that these two types of presentation, while complementary, must be conceived simultaneously.

9.3.2 TABLES

There are two main kinds of graphs, depending to the public:

- Basic tables, designed for the model builder and his working partners, use simple formats to summarize the whole set of model information. They allow a quick interpretation of the main characteristics of simulations and shocks. For a macro economic model, the table will generally contain the evolution of the main components of the supply-demand equilibrium (at constant prices) as well as the evolution of prices, employment, external and budgetary balances, and some ratios.

The definition of elements presented will be limited, sometimes even to a name representative of the concept. It will be simpler to implement if adequate procedures have been created by the user inside the model-building software.

For more detailed models, one will not have to consult the whole set of results, except to find reasons for the abnormal evolution of some aggregated elements.

- More elaborate tables, designed for presentation to a more or less initiated public (the general public, the clients of the study, partners that have collaborated directly to it, other researchers, non-specialists of the

¹²⁶ But the truth of this argument depends also heavily on the celebrity of the author.

¹²⁷ This is made easier by presentation packages.

subject). Even in the case of a working paper, a certain level of quality is needed. This is made easier by the most sophisticated model-building packages, and their production can be used for some types of publications not requiring too high a level. The quality of EViews tables has increased dramatically with the last versions, allowing to choose fonts, symbols and borders. In our opinion, it is no longer necessary to transfer information to presentation packages, except in very specific cases.

In practice, one observes that, even for if the model is completely assimilated, the production of relatively elaborate tables makes work easier, and allows to locate, not only the properties, but also the problems. It is a good investment to produce them as soon as possible, starting with the development phase. Once their framework has been defined, these tables can be reused indefinitely without additional cost.

9.3.3 GRAPHS

The advantage of graphs is to clarify series evolutions by a more telling presentation. The disadvantage is the absence of precise figures. Graphs therefore should be used:

- Along with tables, to complement the detailed results by more synthetic and easily accessible information.
- Alone, if the message to transmit represents a simple synthesis.

The main graph type uses:

- Points, connected or not by segments, or more rarely by curves (differentiated by a color, a type of line, and often a symbol).
- Histograms, each observation being associated to a set of vertical bars (juxtaposed or piled) each measuring the size of a variable.
- Pies¹²⁸, associating to a single observation a circle whose slices correspond to the share of each variable in the total.

This list is not restrictive, as the imagination can go far in search of visual synthesis. For example, bars in histograms are often replaced by appropriate drawings (persons, factories, products...).

Concerning graphs (and presentations as a whole) the modeling packages have made considerable progress in recent years. Some offer sophisticated graphic features available through a large set of user-friendly functions, and EViews is not the last in that regard. One will be able for example, immediately after simulation, to display the evolution of historical and simulated variables or of selected expressions. This display mode will be used in the model development process, allowing a better interpretation of intermediate results, and revealing problems.

This means, even for publication, one can consider using EViews graphs. Of course they lack some features like 3D histograms or items (like plants) of variable size, but they take advantage of their specific purpose: displaying a limited

¹²⁸ In French, “Camembert”, a region Taped une équation ici. al cheese with a round shape.

number of time series, in color or monochrome. The language used might be a little awkward (but explicit) but once a graph has been designed, the statements can be used to initialize the next one.

Graphs can be transferred to

9.3.3.1 Graph of model solutions

For graphs, model solutions represent a specific case.

The most interesting feature is the MAKEGRAPH command, specially adapted to model solutions. The syntax is:

```
model_name.makegraph(options) graph_name model_series_names
```

If no option is specified only the current scenario solution is displayed.

The main options are:

```
a : include actuals  
c : compare active to baseline  
d : include deviations from baseline (as an additional graph in the same frame)  
n : do not include active scenario.
```

The results can be presented raw or transformed, and for stochastic simulations a confidence interval can be displayed.

Once the graph is created, the usual modifiers can be applied (legend, type of line, colors....).

One has to note that the graph must not preexist. If it does it must be deleted first (using the “noerr” modifier to avoid a possible error message).

```
delete(norerr) name_of the graph
```

CHAPTER 10: APPLYING THE ABOVE PRINCIPLES TO OPERATIONAL CASES

In the preceding chapters, we have addressed all the technical aspects of models construction and applications, illustrating them by a simple example. Even if the techniques proposed have been quite general, we have not addressed the more economic features of operational models. This is what we are going to do now.

Having mastered the techniques for producing a very small model, the professional will face two kinds of difficulties:

- The structure of the new model has to be more consistent with economic theory (the one he has in mind). The detail of the specifications must meet his requirements, in terms of assumptions, mechanisms and results.
- The methodology for producing and managing the model call for slightly different techniques, adapted to a problem of higher complexity.

In the following pages, we shall address both these questions. As we are not giving a course in macroeconomics¹²⁹, we shall only describe the elements which in our opinion are requested for the production of an operational model, the choices one has to make, and give some directions as to the most promising options. We leave to the reader the choice of actual decisions and the introduction of additional features he considers adapted to his problem.

So please do not criticize our explanations for being too basic¹³⁰. If this is your feeling:

- You are a good enough economist to proceed by yourself
- Or you can use any of the more advanced textbooks in the list we provide.

Again, let us stress that you are not reading a book on sophisticated econometrics or economics. Our goal is to show how the two fields can be merged into a working model, concentrating on the technical aspects of the task.

In any case, the elements we will describe should be interesting, if interpreted as a description of the main options used by mainstream operational structural models (maybe a little on the old fashioned side).

For this reason, we do not cite a source in which the mechanisms we describe are presented. Rather, we give a list of books on macroeconomics, or full model presentations, in which most of them will be found in higher (and probably better) detail.

We will be more directive on the EViews aspects, of course.

10.1 THE ACCOUNTING FRAMEWORK

Before we start dealing with macroeconomic theory, we must describe the framework and the concepts we shall use. We have already started, of course, but in operational cases we must be much more precise and consistent.

As for econometrics, we do not have the ambition of giving a course in national accounting. At the end of the book one will find a list of associated publications, from simple courses to official manuals stating the official concepts used at present by local or international organizations.

In our opinion, the best reference in that field is:

LEQUILLER F, BLADES D, (2006), Understanding National Accounts, OECD Publishing, downloadable at: <http://www.oecd.org/dataoecd/37/12/38451313.pdf>

To define the framework of a macroeconomic model, one must define:

- The agents involved
- The operations they share

¹²⁹ Neither did we give a course in econometrics.

¹³⁰ Some people will certainly decide they are wrong. This is another problem, but most of the features we present have been used in several models (this is not a proof of their validity of course).

- The products associated with these operations.

According to the problem the model is designed to solve, these elements will be more or less detailed.

We shall start with a single good, single country model (or a series of unrelated ones), then extend the classification. The specifics of the building process will also be dealt with separately.

10.1.1 THE AGENTS: A FIRST DEFINITION

The operations described in a model will be managed by agents, characterized by their role. We shall separate:

- The firms, which buy and produce goods and market services, distribute revenue.
- The households, who receive revenue and consume part of it.
- The financial institutions, which manage financial transactions: lending, borrowing, insuring.
- The administrations, which gather taxes from the above agents, and address social needs through redistribution and production of associated elements.
- The rest of the world, representing the foreign agents trading with the first four.

10.1.2 THE OPERATIONS

They are divided into:

- Operations on goods and services: a good (material, such as a bottle of wine or a CD player) or a service (immaterial, such as transportation from one place to another or a music lesson) is traded between two agents, generally in exchange for money. But it can be bartered (a kingdom for a horse) or given for free, most of the time by some administration (free school or free medicine). In the last case the transaction is considered as “non-market”.
- Transfers from one agent to the other, which can come in exchange for something, separately from the above transaction. For instance, wages are paid to households by firms, but they do not buy the actual good, just the contribution to its production. Alternatively, pensions are financed by contributions, but not necessarily paid by the beneficiaries, and in any case with some delay. And income taxes are not linked with any service (although they will be used in part for the good of the taxpayer, at least in principle).
- Things are not always so simple: an independent taxi driver will sell a service (and use the money as revenue). If he belongs to a company he will earn wages (a transfer). The company will sell the service and transfer part of the revenue to him.

10.1.3 THE INTEGRATED ECONOMIC ACCOUNTS

Once these elements have been defined, they can be presented in a table, allowing one line per operation, and one column per agent. Actually, an additional column will be introduced for the first “goods and services” part: for each operation, on the left for total expenditures, on the right for total revenue.

This is particularly useful for transfers: in general, wages will be an expenditure for firms and a revenue for workers. But all agents give wages, including households: this technique allows to separate the two types: paid and received.

- In addition, the table will describe the steps in the economic process: the operations will be distributed in sequential sub-tables, each presenting a balance as the last line.

We shall find in succession:

- The “production” account, describing all operations linked to production, but also imports and exports. It gives value added. This part will call for an additional column, as lines are not balanced (exports represents an expenditure for the Rest of the world, but the revenue is not distributed among agents).
- The “primary distribution of income” account, introducing the immaterial expenses: wages, direct taxes, subsidies (-). It gives the operating surplus.
- The “secondary distribution of income” account, further distributing revenue: dividends, revenue of property, and also social transfers from the administration. The balance is disposable income.
- The “use of disposable income” account, describing essentially household consumption. The balance is savings.
- The “capital” account, describing the use of savings to acquire goods. It includes various forms of investment and the change in inventories. Its balance is the financial capacity.
- The “financial” account, which explains the financial capacity. It describes money, stocks, and credits. Its balance is zero

10.1.4 SECTORS, BRANCHES AND PRODUCTS,

Now we can separate the “firms” or rather “non-financial firms” into categories, according to its type of production. Three notions appear here:

- The sector: it categorizes a firm according to its main activity (car making for the Peugeot car manufacturer).
- The branch: it separates the firm into each of its activities (spare parts production and distribution for the Peugeot car manufacturer).
- The product: it represents the good or service actually produced by the firm. It is normally equivalent to the branch, but not in all countries: the difference can come from “fatal products” coming automatically from a process designed to create another good (such as hydrogen coming from oil refining).

As we shall see later, this decomposition calls for the definition of intermediate consumption, a two-dimensional variable associated to the consumption of one good to produce another. This can be true also for investment (in one product by one branch) or inventories¹³¹.

¹³¹ Even though in this case some elements are identically null, such as investment in the energy product or inventories in financial services.

10.1.5 AGENTS SUBDIVISIONS

National accounts can further subdivide the agents. For instance one can separate:

- Households into individual (a family) and collective (a convent).
- Financial institutions into the Central Bank, other banks, and insurance companies.
- Firms into State owned, individual firms and companies, local or foreign owned.
- The administration into central government, local agencies and specific agencies, such as the social security agency.
- The rest of the world into countries and zones (even for a single country model). It will allow to make the model assumptions more explicit.

Of course, other more statistical classifications can be used, like the revenue level of the household or the occupation of its head.

10.1.6 A MULTI COUNTRY MODEL

We shall develop later the specifics of a multi-country model. Let us just say for now:

- Of course, the system must be duplicated for each of the countries described. Specific elements can be introduced in some cases (like individualizing oil for producing countries).
- Consistency must be enforced between the exports of each country and the imports of others. This can be done either by identifying individual trade flows, or by creating a specific export demand variable for each country, based on the global demand of each of its clients.
- Finally, the categories can be countries, unions (like the European Union) or related zones (like Sub-Saharan Africa).

10.2 A SINGLE COUNTRY, SINGLE PRODUCT MODEL

We shall start with the economic developments needed by our model to reach operational status.

10.2.1 THE ECONOMIC ASPECTS

The very small model we have built as an example is of course too limited to be used for operational studies. It presents nevertheless:

- A link between production and revenue.
- External trade, depending on available productive capacity.
- Production defined as local demand plus net exports.
- A simple production function.

- A description of the output gap, its consequences and the way it is closed.

What do we need in addition, at the least?

- A price system, and its links with the real sector (in both directions).
- A better description of the behavior of firms and households.
- A financial sector.
- A full description of the State budget and its instruments.

This will also call for a redefinition of the scope of assumptions.

10.2.1.1 The productive process

This part of the model (one speaks often of “blocks”) will not define production, but rather potential production (or productive capacity), as a function of available factors.

Why not actual production itself? There are two ways to consider production:

- Actual local production, contributing with foreign exporters to the satisfaction of demand (both local and foreign) demand, in a share depending on relative prices and available capacities).
- Potential production, given by the production function, taking into account the level of factors (capital and labor), themselves chosen by firms according to their relative costs, expected demand, and profits conditions.

We want our model to follow the most logical causal sequence, which is:

- Defining target capacity depending on profit conditions and expected demand.
- Choosing the optimal level of factors allowing this capacity.
- The actual levels will adapt, giving potential production.
- Global demand will follow, and will be shared between local and foreign producers to give actual production.
- Imperfect knowledge of future demand, technical difficulties, and concerns in a fast adaptation of factors will contribute to the creation of a gap between potential and actual value.

The comparison between actual and potential production will play an important role in some behaviors.

This is the sequence that the model will describe, actual production being obtained late in the process, once demand is known (as in the small model).

This capacity for production will be measured:

- For employment, in man - years or man - quarters according to model periodicity
- For capital, at constant prices, in the currency of the country.

The function can also include:

- Energy consumption
- Intermediate goods (like raw materials).

Actually, capacities are generally defined in terms of value added, a more reliable notion as we have explained earlier. This means the two last elements are not taken into account, or rather their level will come automatically from value added itself.

The first issue concerns the logical link between capacity and factors. We have already seen:

- Complementary factors. For a given capacity, there is a single optimal process using a fixed combination of labor and capital. Starting from an optimal combination, adding to the process a quantity of one factor does not increase capacity, or allow using less of the other factor. This capacity is obviously optimal regardless of the relative costs. Actually labor productivity has generally some flexibility, and capital is the truly constraining factor, as temporary and limited increases in labor productivity can be achieved (for instance by increasing the number of hours worked).

This is the simplest option, in its formulation, estimation and understanding of properties. Operational models use generally more sophisticated frameworks:

- Cobb-Douglas. The elasticity of substitution is unitary. This means that if the ratio of the cost of labor to capital changes by 1%, the optimal ratio of capital to labor will change by 1% too, for a given capacity requirement.
- CES (Constant elasticity of substitution). Now the elasticity can take any fixed value (with the right sign).

Of course, the CES option covers both others (with fixed elasticities of 0 and 1 respectively).

The framework calls also for:

- A definition of the relative cost.

The relative cost of labor and capital is not just measured by the ratio of the wage rate to the investment deflator. One has to take also into account:

- Social contributions of firms: they contribute to the cost of labor.
- The interest rate: while capital is bought immediately¹³², labor can be bought (rented) when the time comes (slavery has been abolished for some time now). So a firm which has money can save it, and one which has not does not have to borrow.
- The depreciation rate: capital wears out, while when a worker “wears out” through old age or sickness, he will leave and can be replaced by a new one at no cost except training (pensions have already been saved as a share of wages).
- The future evolution of wages: if wages are currently growing faster than inflation, firms can expect labor to become less competitive. The gain from having output transferred to fast developing countries becomes lower as they close the gap with developed ones. This applies in particular to present China.

¹³² Actually some forms of capital (like buildings, computers or patents) can be rented or leased.

- The possible changes in technology.

The issue here is to decide if the technology decided at investment time (which defines the roles of labor and capital) can change later.

Basically, the options are:

- A single available technology (Clay-Clay).
- A technology chosen at installation time, with no later change (Putty-Clay). This means basically that the “complementary factors” option applies to factors once they are installed.
- A technology with a permanent possibility of change (Putty - Putty). The same substitution option applies to factors at any period.

10.2.1.1.1A specific problem: the statistical determination of productive capacity

To determine capacity, we have several options, depending on the available information

- In some countries (such as France), a survey asks firms by how much they could increase their production using the present factors (hiring more people if necessary). This gives the firm’s capacity. Using the same weights as for computing actual production, one gets a comparable measure of capacity, and the rate of use as a ratio of the global values.

Then we shall use the capacity series to estimate its equation. For this, we can specify the actual behavior of firms, and optimize their profits under a capacity constraint using the formula we want to estimate. This applies when the factors are substitutable (otherwise the optimum solution is set from the start, and does not depend on relative costs). Taking the derivative of the function according to both labor and capital will give a set of equations with common coefficients, which one can estimate as a system. This method takes into account fully and explicitly the role of the relative costs.

- If we know only the level of factors (capital is sometimes missing in the country’s statistics), we can specify the production function, and estimate its parameters over the actual values of production. We can suppose that the estimated formula gives normal production, and the residual is the output gap. Again, the ratio of actual to “normal” production gives the rate of use, but this time to a constant factor (the average rate of use).

We can also (a better solution in our opinion) apply the first method, using actual production instead of capacity. Again, the estimated capacity (reconstructed by applying the production function to the estimated factors, considered as optimal) will give a normal level of production, and the difference to actual production the output gap.

- If we do not have this information, we can always smooth production, and use the result as a “normal production” level (at a normal rate of use of capacities). For this, applying to actual data a Hodrick-Prescott filter is the most usual technique. If we suppose the “normal” rate of use of capacities constant over time, we get capacity at an unknown multiplicative factor.

This technique does not require a choice of production function, or the availability of a series for capital (which is often absent or unreliable). Neither does it provides it, which will be a problem for model specification.

10.2.1.2 The change in inventories

We see no specific reason to modify the framework used by the small model. More sophisticated formulations could use:

- A full error-correction framework, provided we knew the level of inventories.
- An influence of demand: if it goes up suddenly, some of it can be met by using inventories. This element will be difficult to introduce, as it calls for a negative influence, while value added has a positive one, and both elements are positively correlated. This means the over estimation of one coefficient can be compensated by over estimating the second too.
- An influence of prices: the more expensive the inventories, the shorter the time they will be stored.

10.2.1.3 Unemployment

This is a new concept, which would have fitted easily into the small model (but it would make it a little larger).¹³³

We shall consider that the variations of employment do not transfer fully to unemployment. Job creation will attract to the labor market previously inactive persons, who shall take some of the jobs offered: the work force (employed + unemployed) will increase.

For instance, creating a firm in a low industrialized zone will allow housewives to combine employment with domestic work.¹³⁴ Or employees of a closing down factory will not necessarily remain in the labor market if their qualification is not required elsewhere.

But the level of unemployment should also influence its dynamics. If it is high, the incentive to join the work force will be lower. Favorable employment prospects will lead young people living with their parents to start their working life. On the contrary, a depressed labor market will persuade aged workers to retire earlier (and they will be incited to). And some of the unemployed will stop looking for a job, and leave the work force.

Also, the higher the unemployment level, the higher the quality of the best unemployed. Observing the situation, the average unemployed people will lower their probability of getting a job, leading them to leave the work force.

On the contrary, at a low level of unemployment, the unemployed will feel that they stand a good chance over their competitors, most of them being either inefficient or not really looking for employment.

This obviously corresponds to an error correction framework, leading to a target rate of unemployment (and also of participation of potential workers to the labor force, as we shall see).

10.2.1.4 The price system

¹³³ We have tested it, and it works..

¹³⁴ Which is not considered as employment (maybe because it is not paid, and does not affect GDP, even if paid housework does).

The role of prices in a model is of course essential. But it is not so simple to introduce, even for a minimal model like the one we presented above. In this case, several deflators have to be introduced simultaneously, associated with the elements in the supply - demand equilibrium:

- GDP
- Final demand
- Exports
- Imports.

And in addition:

- Wages (possibly including social security contributions)
- Possibly, deflators for each element in the decomposition of demand: consumption, investment, government demand....
- The price of foreign currency (the exchange rate)
- The prices of lending and borrowing (the interest rates)

Moreover, trade prices have to be defined including and excluding taxes. This distinction applies to external trade (for defining competitiveness and trade balance) and local demand (for defining final and intermediate consumptions).

Not all these elements have to be estimated. Behaviors should be associated with:

- GDP (firms decide on the price at which they sell, once they take into account the cost of input).
- Exports (local exporters do the same).
- Imports (now we consider foreign exporters)¹³⁵.
- Wages (the result of a negotiation between workers and firm managers).

Final demand price should be used to balance supply and demand at current prices. The model is giving a balanced set of four elements at constant prices, and three of the deflators have already been decided. The demand price should balance:

$$P_{fd} \cdot FD + P_x \cdot X = P_q \cdot Q + P_m \cdot M$$

Or

$$P_{fd} = (P_q \cdot Q + P_m \cdot M - P_x \cdot X) / FD$$

Let us now address the links between prices

¹³⁵ Remember we are building a single country model. The description of trade will be different with several connected countries.

- In the system, the deflators will depend on each other. For the time being, we will only give indications. A more detailed reasoning will come with actual estimations.
- The GDP deflator depends on the wage rate, or rather the wage cost.

If wage costs go up, firms will have to increase prices to keep their margins.

They do not have to do it immediately, and they are not obliged if they want to keep their competitiveness on the local and foreign markets (for exporting firms).

And actually it might be better to use the global cost, including amortization of capital.

- The wage rate depends on the consumption price, but maybe also on the value added price.

If prices go up, workers ask for a raise in wages to sustain their purchasing power. But again, firms are less liable to accept raises if they were not able to increase their own price.

- Trade prices depend on the cost supported by the exporter, and on the price set by its competitors. This means they have to maintain their margins and their competitiveness at the same time.

This behavior is obviously based on production prices, the price at which they sell. This means the cost of intermediate consumptions has to be taken into account. For instance, a country having access to cheap oil will be able to export at lower prices, even at the same cost in value added (and the same margins). But this introduces a problem, as until now the single product feature allowed us to discard intermediate consumption, a variable difficult to manage as its value depends on the classification.

The behavior also has to apply to the same currency. If the export price uses the currency of the exporter, the price of its competitors measured in foreign currency has to be corrected by the exchange rate.

- The price of demand depends on the price at which local producers and foreign exporters sell on the local market.

This uses the identity above.

Another important issue concerns the separation between the prices at which local firms sell on the local and foreign markets (the export price).

Two behaviors can be considered:

- The firms define both selling prices separately. Local firms start by defining a price for selling on the local market, using the above behavior. Then the export price will average this price and that of competitors.
- The firms define first a global selling price, allowing to reach a global margin rate, then they chose a combination of the two prices which meets this target. This means that a decrease in the export price (possibly designed to stay competitive facing foreign deflation) will have to be compensated by an increase in the local selling price.

The choice will have a strong impact on the price system. Actually the second option will increase the intensity of the price-wage loop: if local costs go up, firms refuse to apply completely these costs to exports (as they do not want to lose their competitiveness), and maintaining global margins calls for a larger increase in local selling prices (which does not happen if targets are defined separately).

- This equilibrium is subject to external influences, either endogenous or exogenous.
 - Endogenous
 - * If labor productivity goes up, firms need fewer workers and can pay them more. They can also lower their prices.
 - * If output is too low compared to capacities, firms can first lower prices to sell more (later they can adapt their capacities).
 - * If unemployment goes down, workers can increase their demands without the risk of firm managers to look elsewhere.
 - Exogenous.

The concerned variables are the indirect tax rates.

One will generally consider:

- * VAT.
- * The other indirect tax rates, such as the tax on tobacco, gas, alcohol.
- * Tariffs.

And also

- * The rate of social security contributions by firms.

If indirect¹³⁶ tax rates (such as VAT, tax on gas, cigarettes, social contributions paid by firms...) go up then firms should adapt their price if they want to keep their margins.

Two elements:

- First, it is quite important to separate these taxes in a model, for the usual reason: their base is different, and their impact on the economy also.
- * VAT applies only to value added. But the most important feature is that it does not apply to exports (exporters can deduct it before they sell abroad), and they apply to imports. VAT on foreign cars is the same as on local

¹³⁶ These taxes are called indirect because they are not paid directly by the ultimate payer, contrarily to income tax, corporate tax....

ones, and applies to the total value. And when the car firm looks for electronic equipment, increasing VAT on this good will not change its decision on its origin as it can deduct VAT anyway¹³⁷.

- * On the contrary the other indirect taxes apply only to local productions, even though the impact of this difference is not so high, as imported goods are often taxed at the moment they are sold. For instance, the tax on tobacco applies also to imported cigarettes, and the tax on alcohol to imported whisky.
- * Concerning local tariffs, they are not deductible in the general case. This means that a change in their rate will affect directly the competitiveness of imported goods, unless the importer decides to compensate the effect by adapting its margins.
- * As to tariffs applied to local products by foreign countries, they affect directly the competitiveness of exports. This means they have to be considered, even though their statistical value is not directly obtained from the national accounts.
- Second, formalizing the role of taxes relies obviously on the rates, the variables decided by the state (or foreign states for tariffs on local exports). These rates will affect deflators, and allow computing the amount of the tax once the base is known.

It should be clear that the right way to formalize these taxes is to set the rate as an exogenous decision variable, and not to estimate the amount as some modellers might be tempted to do. This allows handling the decision easily, both in forecasts and shock analysis. And on the past, the technique is quite simple: the tax amount and the base are known, and this allows computing the rate, used as an exogenous ratio. The associated identity (tax = rate x base) will hold true. We shall see later how to handle these rates on the future.

Obviously, the rate obtained will be different from the legal one (generally lower!).

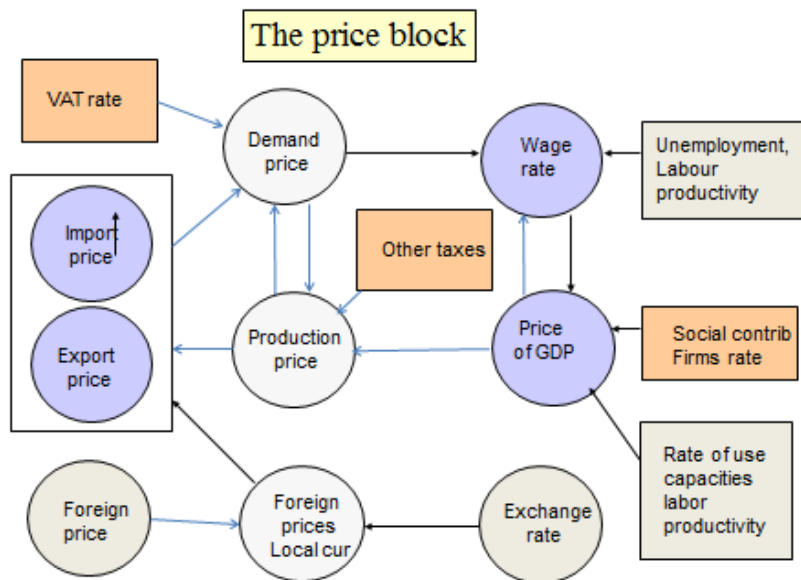
This technique is consistent with the general approach: make the government decisions exogenous, but identify first what represents the true decision.

- Prices can also influence real elements
 - The selling price of local producers determines the quantities they will sell. This is also true of exporters, through the comparison between their export price and the price on the market on which they sell
 - The relative costs of labor and capital influence the choice of the factors in the productive process.
 - More generally, ratios of prices affect the ratios of elements (or the shares in a total). For a given global consumption level, reducing the price of one good will increase its share.
 - A higher inflation reduces the purchasing power of previous savings, calling for a larger reconstitution effort.

¹³⁷ It might not apply the whole of the increase, however, if it fears a drop in its sales.

- And of course prices enter the definition of variables at current prices, when they are separated into volume and deflator (elements in the trade balance, wages...). For the elements in a sum, a different evolution of deflators will change the shares at current prices.

All this is described by the following graph.



10.2.1.5 The account of firms

We have already dealt with the supply side, defining the adaptation of production factors: employment and capital, to target capacity, as well as the decision on prices, based generally on the short term maximization of profits.

This means that most of the remaining equations will be definitions, describing the firms account without calling for any theoretical elements.

There are two exceptions:

- The tax on profits, which should be again computed by applying a rate to a base. This is more complex than usual, however, as:
 - Computing profits in a model is quite complex, and not all models are able to do it. Sometimes it is necessary to use a proxy, making the apparent rate more difficult to interpret.
 - The timetable for the tax calls for a dynamic equation, as the tax is not generally paid in the same period as the associated profits (but there can be a provision to pay immediately). So a formula describing the mechanism must be established.
 - The tax on negative profits is not negative, but null, introducing a bias on the apparent rate.

- Dividends paid by firms, which can be estimated or constructed through an identity (using a rate in the same manner as taxes). Again, one must decide on the dynamics, as dividends follow the profits. Also, the beneficiary of dividends has to be identified (sharing must be done between the five usual agents).

Of course, the complexity of formulations (and even the identification of elements such as dividends) depends on the role of the model, whether it is used by researchers trying to answer global theoretical issues, or by policy advisers addressing in detail the evolution of the next State budget.

10.2.1.6 The behavior of households

On the contrary, we have not yet addressed the decisions of households.

Basically:

- Households obtain revenue from several sources, the main ones being:
 - Wages
 - The revenue of individual workers
 - Social benefits of various kinds
 - Interests from loans
 - Dividends
 - Renting lodgings to other households (a service)¹³⁸
- They use this revenue:
 - To pay the income tax
 - To consume various goods and services
 - To save, and in particular in housing, but also in deposits, bonds, stocks and goods (such as art).

To be considered operational, even a single product model must use some detail, as the economic processes through which these revenues are obtained, and the consequences of spending decisions, are quite different from each other.

Another principle of modelling: favor the detail which allows separating behaviors.

Basically

- For revenue:
 - Wages paid by firms should be the product of an average wage rate (coming from the price block) by the number of workers (from the production block).
 - The number of civil servants will generally be exogenous, but not the wage rate, which can be different from firms'.

¹³⁸ Actually, it would be strange to consider that if a household buys the apartment it was renting, the service disappears and GDP decreases. For that reason, housing owners are considered by National Accounts as paying to themselves a fictitious rent.

- Wages paid by households (mainly for housekeeping) can be identified or not, according to the type of model.
- Social benefits are generally separated in five types: sickness, family subsidies, unemployment benefits, invalidity from working accidents, pensions. It is clear that:
 - * Each of these elements depends on inflation, but at different degrees.
 - * Most of them depend on population, and often a given type of population. For instance, the number of children, the number of people having reached retirement age, or of unemployed.
 - * All of them depend on economic activity, again in a variable way. For instance, unemployment benefits decrease with GDP, working accidents increase, and pensions should increase (in principle) with the revenue from the contributions which finance them.
 - * They also depend on a decision made by the State (the purchasing power is maintained).

This means an operational model should try to separate these items, to take into account their differences in behavior.

In this way the model will show naturally:

- * The change in benefits with the number of beneficiaries.
- * The change in benefits with the decision.
- The interests will be described globally, in a subsequent paragraph. Let us only stress that for households the interest rates (lending and borrowing) can deviate from market values through state intervention. In France, a limited amount of savings benefits from a higher guaranteed rate, and borrowing to buy housing can be done at a lower rate (0% in some cases).
- Dividends will be treated later with the firms' account.
- As to the revenue from housing (rents), its role in a model is limited, as it mostly represents a transfer from households to other households. For owners of property, it is even a transfer within the same household. There are reasons to consider it, however: it can be subject to taxation, and it enters GDP.

One should not consider marginal elements, such as lottery winnings, inheritance, donations, fines...

- Finally, one can formalize the transfers from abroad (or to abroad). For developing countries remittances can represent a sizable share of household revenue (more than one third of GDP for Tajikistan). For a single country model they should be exogenous, perhaps even in current terms (a notable exception to the general principle).
- For expenditures:
 - The income tax should be computed as a rate applied to revenue before tax, obtaining the historical values of the apparent rate by dividing the amount by the base. The model will then get the tax by applying the exogenous rate to the base. The base poses the usual dynamic problem: the tax can be paid after the revenue is obtained (with a provision mechanism).

Also, applying an average rate to all households can be acceptable for forecasts (which allow this rate to change with time), but less for the shocks addressed to a category of households at one extremity of the spectrum: in a traditional macroeconomic model, a decrease in the tax on large or an increase in benefits for the poor, of the same ex ante size,

will have the same ex post consequences¹³⁹. To eliminate this error, an ad hoc correction has to be made on the savings rate itself.

This problem appears in most models, coming from the fact is that the tools to solve it are not available. National Accounts separate firms using the goods they produce, but not households according to any categorization, including the level of revenue. Some surveys address the problem, and their teachings could be used to create specific data. This means some solution might be found, but without doubt at a high cost. Actually, the same problem arises if one wants to separate firms not according to sectors, but to size, considering that small firms act differently from large ones.

- Once the disposable income is known, all that remains is to separate it into consumption and savings, considered as whole in most models (for multi-product models the situation will be more complex).

The most common technique is to compute consumption first, as a ratio to revenue, then savings as a residual. We shall develop this with estimations.

Consumption is generally determined at constant prices (which means in purchasing power). The usual determinants are:

- The level of revenue (measured also in purchasing power). The higher the revenue, the higher the consumption level, but the lower the share of consumption (the poor do not save, and remember that buying a house is considered as savings).
- The recent evolution of revenue. Households take some time in adapting their behavior to an increase (or decrease) in revenue. And a sudden hike (especially if it is destined to be permanent, like a promotion) can lead them to invest in housing, which can actually decrease consumption for a while.
- Inflation (the “real holdings” effect). Present savings contain a large share of monetary elements (deposits, bonds with fixed rates...). Current inflation reduces their purchasing power, which has to be complemented by additional savings. The effort is proportional to the inflation level.
- The unemployment rate. For employed workers, an increase in the chance of losing their jobs (measured more by the change in the rate than its value¹⁴⁰) leads them to save a larger share of their present revenue, if they want to optimize their utility across time.
- The (short term) interest rate: in general, people prefer satisfying a given need now than later. But this has a cost, the interest they have to pay. The lower the rate, the more they will indulge in immediate consumption.

This is particularly true for durable goods: if a household wants to watch flat screen TV (and thinks that after its purchase, in its whole life it will have enough resources to afford a set) the only reason for not buying one right now (and increasing its satisfaction permanently) is the actualized cost, which is lowered with a decrease in interest rates. What it has to consider is not the cost of the good, but the cost of making its acquisition earlier.

¹³⁹ Of course, the impact on consumption will be higher if the increase concerns the poor.

¹⁴⁰ Although the actual rate plays also a role: a higher value implies a higher turnover, and a high risk of participating in the turnover.

If the good is perfectly durable, and can be sold back at its original value at constant prices, things are just as if it was renting the good. If the interest rate is divided by 2, the “price” of the good is divided by 2.

For non-durable goods, the situation is different. The household has already optimized its consumption over time. If the interest rate changes, it might be tempted to consume earlier, but if the marginal utility of the good is decreasing fast, the pattern of consumption will not be much affected. A person dreaming of visiting the pyramids, and saving for that purpose, might make the trip earlier but will not do it again.

What matters is the real rate:

- They allow comparing goods at constant prices.
- If households assume their revenue will grow with inflation, they will optimize in real terms.

Once consumption is determined, savings are computed as a residual, and generally as a global element. This option can be discussed, as different kinds of savings can be assumed to follow different behaviors.

In particular, housing investment is negatively affected by interest rates (a specific rate, but one can assume it follows the global rate) while financial savings are positively so. Buying a house calls for obtaining a given good and asking another agent to provide the collateral, in return for interests. Buying a bond means lending collateral to another agent to use it as a spending tool (maybe to buy a durable good), in return for interests but this time in the other direction.

10.2.1.7 External trade

In a single country model, the rest of the world is exogenous.

This means that we consider only influences from the world to the country, and not the other way around.

Of course this is not really exact, even for the smallest of countries (or in that regard for a region, a town or an individual): by increasing your consumption and so local production, you create a fraction of a job, a small amount of household revenue, and again more consumption.

What we consider is that the influence is too small to have a sizable effect, and that the cost of producing (and running) a model describing it is too high compared to the gain in the accuracy of results. This is essentially true for lower or medium sized countries like Latvia or Bolivia, much less so for larger countries like France, and quite untrue for the USA or the European Union considered as a whole. For instance when we used the MacSim world model for a shock analysis, the French Keynesian multiplier for 2000 was 1.3 if we ran the full model, but only 1.1 if we ran the French model by itself. The iterative feedbacks of German imports from France, coming from the increase of German exports, will have the largest share in the difference. Considering the evolution of world trade, the present difference should be even wider.

This means that the exchanges of the country have to be considered from the point of the country:

- Exports are the share of production of goods and services which is sold by the country to the rest of the world.
- Imports are the share of local demand for goods and services which is not produced in the country, but bought from the rest of the world.

Both elements will be computed using the currency of the country. However, using constant prices will mean using the exchange rate of the base year, so the currency issue is not relevant, introducing only a scaling by a constant factor¹⁴¹.

However, the trade elements having the same nature, their logical determinants will be the same. The main difference will come only from the relative size of the two markets (buyer and seller) in the trading operation: the single country's importance (or GDP) will always be much lower than that of the rest of the world, although this is less obvious again if we model the USA or the European Union as a whole.

These elements will be:

- Demand: for a country to sell a given good to a partner country, demand for this good must be present, part of this demand must be addressed to the world market, and the quality of local products must appeal to the importing country. For instance, French exports of wine will depend on the world demand for wine, and the natural preference of importing countries for foreign wine (starting with their status as wine producers).and French wine in particular

Defining demand introduces two problems.

- For imports, we have already seen that including intermediate consumption in the supply-demand equilibrium (thus considering production on one side and total local demand on the other) is quite a problem for models, as the level of intermediate consumption depends on the number of steps in the production process. The single product feature has until now eliminated the need for considering intermediate consumption. But imports do contain intermediate goods, whether they represent energy (oil, gas, even electricity) or primary goods (from untreated wood to electronic components). And these intermediate goods are necessary to exports.

A simple solution is to consider that the ratio of intermediary consumption to value added. Looking at the figures, we can indeed observe that the "technical coefficients", the number of units needed to produce a unit of value added or GDP, is rather constant.

So we have just to consider a composite demand, as the sum of final demand itself, and intermediate consumption as a function of GDP (or rather value added, as intermediate consumption excludes VAT).

In countries in which the trade balance is more or less in equilibrium, we might consider using a combination of final local demand and exports.

- Price competitiveness: to decide whether to buy a good from a local or foreign producer, a country will compare the local price with the foreign exporters' price. And to choose among potential sellers, the importing country will consider their relative price at a given quality (remember that the deflators consider goods at the same quality level, an increase in quality improving the value at constant prices).

We generally observe that the relative price is less of an issue when the buyer contemplates buying local or foreign goods, than when he has to choose between foreign sellers. This follows economic logic: local goods are supposedly designed for the local market, and some goods are not or hardly substitutable (local bus tickets or newspapers).

¹⁴¹ This is only true if we consider a single rest of the world, or we measure it in a single currency. More on this later.

This means in our case that the sensitivity of exports to price competitiveness should be higher than for imports. Exports depend on world demand to the world market, and once a country has decided to import, the price will play a more important role than in the import decision itself.

Finally, of course measuring competitiveness must use deflators defined in the same currency. It can be any currency, as applying the same exchange rate to both elements of the ratio will not change its value. In the case of exports, this means that measuring their deflator in local currency calls for a foreign price measured in the same units. As the exchange rate is identified, this foreign price will be endogenous, as the product of two assumptions: the foreign price in foreign currency, and the correcting exchange rate (a deflator). It is perhaps more logical (and equivalent in practice) to consider both prices in foreign currency, the local one being corrected by the symmetric exchange rate.

- The available capacities

The third element is the potential to supply additional demand, which means the presence of available productive capacities. The relevant variable is naturally the rate of use of capacities, independent from the size of the economy.

The choice of this option is not so straightforward, however. One could argue that as long as the rate is lower than one, additional demand can be satisfied. We have already shown that this is not true: demand concerns a wide range of products, and one cannot generally be substituted for another, in particular in the short term. And some products may see their capacity completely saturated.

Let us explain the process again, this time in an import oriented way.

Actually, the average rate of use is based on a distribution of rate values, from zero (hopefully a few cases) to one (probably a sizable number). When global demand increases, it addresses a range of firms, in variable intensity. Some of these demands will be addressed to firms already unable to provide more, and some others will make them reach that level. The proportion of firms working at full capacity will grow. Of course, in some cases, another available product can represent a substitute. But the most simple option is to import the same product, as the missing product should be available somewhere in the world (maybe at a higher price, but this should be treated by the price competitiveness).

The “missing” demand increases with the share of firms which cannot increase their production, having reached capacity.

Of course, this phenomenon applies essentially in the short term, as firms will react by investing, which will increase capacity and close the output gap with time. But this process can be slow, even if full adaptation should be obtained in the very long run.

But if we follow the above reasoning, we observe:

- That the larger the country, the lower the probability that a given absolute, but also **relative** increase in demand will face local supply problems. This increase in demand will be more diversified, and the available capacities will be more diversified too¹⁴².

¹⁴² This would not happen if the additional demand was in a specific good.

- That in our case, the rest of the world should not face any supply problem, which means that for both our country's imports and exports, only the local rate of use should be taken into account.
- And a last condition can appear for the exporting country. If the world requires a given good, the characteristics of that good produced in the country must also be adapted to the particular demand, which can be quite specific. For instance, facing an increase in the foreign demand for cars, a country might be able to supply them at a competitive price, but the type of cars they propose might be unsuitable. This might concern the size, the sophistication, the powering energy, the security features... Of course one cannot consider going in such a detail in a model, even if the data was available (which it is not).

Unfortunately, finding an element describing this feature is less straightforward than above, especially for a single product model. The simplest idea is to use the age of capital, assuming that a younger productive process will be better adapted to present demand¹⁴³. For instance, a recently built car factory might follow market trends in producing smaller cars, or more energy efficient ones. The age of capital can be derived simply from the chronology of investment and depreciation, if we consider that this depreciation applies equally to all generations of present capital, or that capital disappears brutally a given number of years after its implementation. Another assumption leads to more complex, but manageable, formulas.

10.2.1.8 The Budget

In operational models, describing fully and consistently the Government budget is an absolute requirement.

This is true even if the model is not going to be used by Government advisers, but rather by experts in applied economics. The general goal of these researchers is to assess the consequences for the economy of Government decisions, external events, or structural changes, considering the most probable impact or the range of possibilities, possibly under different model formulations (like different options on the interest rate). The approach might be more or less applied (the advisers might try to produce an image of the next budget, to be presented to the Parliament, and the scientists will try to see how the adoption by the Central Bank of a Taylor rule will stabilize the economy), but the tool required is quite similar.

As we have stated above, the best way of defining the associated equations is to build identities, computing an endogenous revenue or expenditure as the product of an endogenous base by an exogenous rate. The equations will hold true over the past, and the modeller will be responsible for (and allowed to) establishing future assumptions on the rate. Of course, he does not have to keep this rate constant, and can rely on an estimated trend as a starting base. But the final decision will be his.

This technique answers to the following objection: if we consider VAT, even with constant legal rates, the apparent rate will change (grow) with the affluence of households, able to increase the share of highly taxed products in their consumption. One solution is to establish a trend, used as a base value, and to deviate from this trend as a policy decision.

If these principles are followed, it will be possible to produce a table showing the evolution of all budget elements, in current terms and in GDP points, both features being obviously required for presentations.

¹⁴³ Especially foreign demand as its role is increasing with time, so its influence on the nature of investment will be higher in later periods.

Another important principle of modelling: if you cannot choose between the possible presentations for a given concept (value at constant prices, at current prices, growth rate, and ratio to another variable) just look at how this concept is presented in economic publications (focusing on the ones designed for the general public). Or wait until you will have to use the figures in your own presentations, then measure your reaction and that of the public.

10.2.1.9 Financial and monetary elements

In any model, this represents the most variable and controversial part. The first models had little or no financial equations. Even at this stage, the financial block can be limited to the definition of a few rates, and their impact on the real sector (these rates can even be exogenous, generally in real terms). On the contrary, this block can be so developed that the purpose of defining a real sector can be considered as a way to complete the links between financial elements, for instance describing the creation and origin of additional lending if a decrease in interest rates draws investment upward.

In our opinion, even a real side oriented model should include:

- A base interest rate set by the Central Bank of the country.
- A short and a long term rates in the currency of the country.
- An average rate on current net borrowings.
- A rate on the present debt, being computed from the chronology of past rates, perhaps as an autoregressive function.
- One or several foreign rates, applied to borrowings in foreign currency both in the country and in the rest of the world.
- The net interests paid by all (five) agents, considering two currencies for the interests paid to the Rest of the World.

An example of this framework will be presented soon.

From this basic option, developments can consider:

- Identifying the debt of agents (or their financial holdings).
- Separating it into currencies (local, US Dollars, maybe Euros for non EMU countries).
- Separating it into short term and long term.
- In addition, one or more forms of money supply can be formalized.

Most of these equations should be established as identities, based on available data or assumptions. Exceptions can concern:

- The Central Bank rate, following perhaps a Taylor rule, but not necessarily. Actually, the same model should allow several options (using a separating parameter).
- The short term and long term rates could include a risk premium, depending for instance on the current budget deficit or its most recent values.

- The spread between long and short term could depend on growth expectations (more true if they are partly or totally rational) and the health of the local economy.

We shall stop here, as financial issues are less a purpose of this publication.

10.2.2 THE EViews PROGRAMS

We shall now present EViews programs producing the data and the framework for a model of the type we have just described. We shall use the French case again, but we will switch to quarterly data, making our example more operational.

Note: all the programs presented here are available either on CD or on the model site.

Just as in the previous example, we shall start by stating completely the identities, but limiting the definition of the behavioral equations to a declaration of the type of function we intend to estimate (this will be done later).

As we have stated before, producing the data and stating the equations can be done in any order, but creating the groups and predicting the residual check must wait for both tasks to be completed.

However, we shall separate the program which creates the data, and the one which creates the equations (and produces the residual check).

The reasons for this shall be made clear later, but we can already state the main issue: during the model building process, both tasks will evolve, but separately most of the time. Changing the model specifications will be much more frequent than changing the data set, and we want to avoid running a task without reason.

It seemed to us that our presentation would be clearer if we started with model production. It might be also the more natural sequence for the model builder: first to look generally at the data available (without considering individual series in detail¹⁴⁴) then writing down the equations in sequence, observing more precisely if each series is available (directly or through a transformation), then creating the data building program.

10.2.2.1 Producing the model

Let us now see how the economic elements we have described can be combined into an EViews program, along the lines of the simpler case.

In this program, we shall expand the principles of the method we have outlined earlier, to create the framework of a small operational model. As in the example, we shall stop before any estimation. What we should get is a set of equations in which:

- Identities are fully defined.
- Behavioral equations are defined as identities presenting, using specific notations, the explained variable and its explanatory elements.

The following text will essentially contain the EViews statements, and comments present in the file itself (a quote mark appears at the start of the line). Additional comments will appear without quote mark, and use a different font. They are generally associated to features requested longer explanations, which would have taken too much place in the program.

¹⁴⁴ For instance one can start from the list of variables presented at the end of the book.

' We start with the usual definition of the directory

```
cd "c:\program files\evIEWS5\book\fra_0"
```

' We state that any test results will be exported to a Rich Text Format file named _mod_1.rtf

' output(r) _mod_1 would create a basic text file.

```
output(r) mod_1
```

' We close the input and output file

' in case a version is already open

```
close data_1  
close mod_1  
open data_1  
wfsave mod_1
```

' We define the sample period as the maximum available

' restrictions will appear with actual estimations

```
smpl 1960Q1 2005Q4
```

' We now need to create a new (blank) model

'

' Trick

' We give the model a name starting with « _ »

' « _ » is the first element in any alphabetical ordering

' As the workfile elements are displayed in alphabetical order

' this means it will appear first in the display

' (of course among elements using the same trick).

' This is quite useful for frequently accessed elements

' In case the model already exists, we delete it (the "noerr" option allows failure if the model does not exist).

```
delete(noerr) _mod_1
```

' So this means we start afresh

If we do not do this, the previous equations will be retained, even if the new equations redefine the same variables (even using identical formulations). This means variables would be defined twice, which EViews will obviously refuse,

producing an erroneous model. This happens because identities are « appended » as text to the model specification. Until EViews 8, there was no straight possibility to delete an identity from a model.

We will not present them, as we consider much more efficient to state the full model in the same program. We have already presented the reasons for this method: clarity of the model text, better error identification, easier transfers, and easier management of programs.

This is not true for estimated equations, as we shall see later.

```
model _mod_1
```

```
' We define the f scalar, for the production of the behavioral « identities »
```

```
scalar f
```

We start with the production block

```
'-----  
' The production block  
'-----
```

```
' The rate of use of capacities is the ratio of actual GDP to potential GDP
```

```
_mod_1.append UR=Q/CAP
```

```
' GDP balances the supply – demand equilibrium
```

```
_mod_1.append GDP+M=FD+X
```

```
' To get value added we subtract VAT at constant prices, the product of the base year VAT rate by final demand excluding VAT, at constant prices
```

Identifying value added is necessary to compute firm's margins and deflators excluding VAT.

In a model at constant prices, the VAT rate would be more or less constant, and VAT proportional to GDP. But here the use of VAT at current prices, and a deflator excluding VAT, calls for excluding VAT from a denominator measured at constant prices. We shall use value added Q in the production function (even though GDP would provide a similar explanation).

```
_mod_1.append Q=GDPM-r_vat0*FD/(1+r_vat0)
```

‘ Investment depends on GDP, the rate of use, profitability, previous capital and the long term real interest rate. If we consider capital – labor substitution, it will depend on the relative cost of factors.

We see that we can already introduce lags (for the previous level of capital) and elements of formulations (for the real interest rate).

```
_mod_1.append I=f*(Q+UR+RPROF+K(-1)+(IRL-100*@pchy(PC))+RELC)
```

We have chosen to measure capital at the end of the period.
It is the sum of the non-discarded previous capital
and the investment implemented during the period.

Actually what we consider here is the investment purchased during the period. There is no guarantee that it can be used readily for production.

```
_mod_1.append K=K(-1)*(1-rdep)+I
```

‘ The change in inventories will depend on GDP

```
_mod_1.append IC=f*(Q)
```

‘ Employment by firms too, and also possibly on the relative cost of factors.

```
_mod_1.append LF=f*(Q+RELC)
```

‘ Productivity of labor is the ratio of value added to firms’ employment

```
_mod_1.append PL=Q/LF
```

‘ Productive capacity depends on firms’ employment and the initial level of capital)

```
_mod_1.append CAP=f*(LE+K(-1))
```

- ' Total employment introduces civil servants
- ' actually we should also identify households' employees
- ' such as maids and janitors

_mod_1.append $LT=LF+lg$

- ' The actual work force depends on employment and the potential work force, in practice the population in age of working.

_mod_1.append $POPAC=f*(LT+pop65)$

- ' The unemployed are the jobless inside the work force

_mod_1.append $UN=POPAC-LT$

- ' The unemployment rate

_mod_1.append $UNR=UN/POPAC$

' The price block

- ' The unitary wage cost is the wage
- ' (including contributions by firms)
- ' necessary to produce one unit of value added

_mod_1.append $UWC=WR*(1+r_scf)/PL$

- ' The value added deflator will depend on UWC
- ' and the rate of use

_mod_1.append $PQ=f*(UWC+UR)$

- ' The production price aggregates the value added deflator
- ' and the price of intermediary consumption
- ' weighted by its influence :

- ' the demand price excluding VAT
- ' It will be needed for defining the trade prices.

```
_mod_1.append PP=(PQ+tc*PFDXT)/(1+tc)
```

- ' The deflator of final demand is the ratio of demand
- ' at current and constant prices.
- ' It balances the supply – demand equilibrium at current prices
- ' All other elements are or will be computed elsewhere

```
_mod_1.append PFD=(GDPMVAL+MVAL-XVAL)/(GDPM+M-X)
```

- ' We compute also the deflator excluding VAT
- ' by inverting the relation with PFD:

```
_mod_1.append PFDXT=PFD*(1+r_vat0)/(1+r_vat)
```

- 'The deflators for demand elements depend on the global one.
- 'The relation can be estimated, or we can apply an exogenous ratio

```
_mod_1.append PC=f*(PFD)
_mod_1.append PI=f*(PFD)
_mod_1.append PIG=f*(PFD)
```

- 'The wage rate depends on deflators:
- 'PC from the point of view of workers
- 'PQ from the point of views of firm managers
- 'And also on labor productivity and unemployment

```
_mod_1.append WR=f*(PC+LP+UNR+PQ)
```

- 'The trade deflators in local currency depend on the local and foreign production prices
- 'which have to be converted in local currency through the exchange rate

```
_mod_1.append PX=f*(PP+ppx+ER)
_mod_1.append PM=f*(PP+ppx+ER)
```

- ' The exchange rate can be exogenous

- ‘ or depend on the inflation differential (PPP assumption)

```
_mod_1.append ER=f*(PP+ppx+ERX)
```

- ‘ The short term interest rate can be exogenous
- ‘ in nominal or real terms
- ‘ or depend on inflation and the output gap (Taylor formula)

```
_mod_1.append IRS=f*(IRSX+(IRSR+100*@pchy(PC))+ (150*&pchy(PC)+50*(UR-urd)/urd))
```

- ‘ The long term interest rate depends on a lag structure of short term rates
- ‘ possibly with a spread

```
_mod_1.append IRL=f*(IRS+spread)
```

- ‘ The rate on new borrowings is an average of short and long term rates

```
_mod_1.append IR=f*(IRS+IRL)
```

- ‘ The rate on previous borrowings depends on the previous rate
- ‘ and the current rate
- ‘ according to the reimbursement speed

```
_mod_1.append IRM=f*(IRM(-1)+IR)
```

The households block

- ‘ The firms' wages

```
_mod_1.append WF=WR*LF
```

- ‘ The civil servants wages
- ‘ could be different on average from firms'

_mod_1.append $WG=WR*lg$

‘ Total wages

_mod_1.append $W=WF+WG$

‘ Social benefits are exogenous in purchasing power per head
‘ They could be separated into risks

_mod_1.append $SOCB=socbr*PC*popt$

‘ Additional revenue can be linked to value added

_mod_1.append $REVQ=r_{revq}*QVAL$

‘ or be exogenous in purchasing power

_mod_1.append $RE VX=r_{revx}*PFD$

‘ More complex formulations can be used
‘ Social security contributions use an exogenous rate

_mod_1.append $SCW=r_{scw}*W$

‘ Household income is defined as a sum of its elements

_mod_1.append $HI=W-SCW+REVQ+RE VX+SOCB$

‘ Income tax applies to the revenue of the previous year
‘ again, this can be made more complex
‘ or use the present value of revenue
‘ according to the country's rules

_mod_1.append $ICT=r_{ict}*HI(-1)$

‘ Now we compute disposable income

```
_mod_1.append HDI=HI-ICT
```

‘ also in purchasing power

```
_mod_1.append HRDI=HDI/PC
```

‘ Household consumption will depend on:

- ‘ * real disposable income
- ‘ * inflation
- ‘ * unemployment
- ‘ * the short term interest rate

```
_mod_1.append HCO=f*(HRDI+PC+UNR+(IRS-100*@pchy(PC)))
```

‘ * a more explicit relation could be used

```
_mod_1.append HCO=f*(HRDI+HCO(-1)/HRDI(-1)+PC+d(UNR)+(IRS-100*@pch(PC)))
```

or even:

```
_mod_1.append Dlog(HCO) =f*(Dlog(HRDI)+Log(HCO(-1)/HRDI(-1)+Dlog(PC)+d(UNR) +(IRS-100*@pch(PC)))
```

- ‘ This will not change the diagnosis on model structure
- ‘ But will state the type of relationship one wants to estimate
- ‘ As a personal reminder or a presentation to partners

```
-----  
‘ The firms block  
-----
```

‘ Value added at current terms

```
_mod_1.append QVAL=PQ*Q
```

‘ The value added tax

_mod_1.append $VAT=r_vat*PFDXT*FD/(1+r_vat0)$

‘ Gross Domestic Product

_mod_1.append $GDPMVAL=QVAL+VAT$

‘ And its deflator

_mod_1.append $PGDPM=GDPMVAL/GDPM$

‘ Subsidies are proportional to value added

_mod_1.append $SUBS=r_subs*QVAL$

‘ Margins are computed as:
‘ value added plus subsidies minus « other indirect taxes »
‘ minus wages including firms' social security contributions

_mod_1.append $MARG=QVAL*(1+r_subs-r_oit)-WR*LF*(1+r_scf)$

‘ The margins rate

_mod_1.append $RMARG=MARG/QVAL$

‘ The tax on profits is based on past profits
‘ A more complex rule can be applied

_mod_1.append $IFP=(PROF(-1)+IFP(-1))*r_ifp$

‘ Profits exclude:
‘ * household revenue from production
‘ * the tax on profits
‘ * net interests paid

_mod_1.append $PROF=MARG-REVQ-IFP-NIF$

- ' The profits rate applies to capital
- ' at the cost of renewal

_mod_1.append RPROF=PROF/(PI*K(-1))

- ' We can also compute the ratio of margins to capital
- ' a more stable and reliable concept

_mod_1.append RPROB=MARG/(PI*K(-1))

- ' The balance of forms excludes spending on
 - ' * productive investment
 - ' * the change in inventories
- ' as margins includes yet unsold value added

_mod_1.append FCAPF=PROF-PI*I-PFD*IC

- ' Net interests paid depend on
 - ' * their past value
 - ' * the rate on past debts
 - ' * the rate on new debts
 - ' * the balance

_mod_1.append NIF=f*(NIF(-1)+IRM+IR+FCAPF)I

' The external trade block

- ' The import price including tariffs

_mod_1.append PMT=PM*(1+r_tar)/(1+r_tar0)

- ' Import price competitiveness compares
 - ' * the local production price
 - ' * the import price including tariffs

_mod_1.append $COMPM = PMT/PP$

- Final demand at constant prices is the sum of its components including a residual proportional to GDP (a reasonable assumption)
- Housing investment could be identified

_mod_1.append $FD = COH + I + IC + CG + IG + fdxr * Q$

- Imports depend on
 - * Final demand and intermediate demand (proportional to value added)
 - * The rate of use of capacities
 - * Price competitiveness

_mod_1.append $M = f * (FD + Q + UR + COMPM)$

- Export price competitiveness compares
 - * the foreign production price
 - * the local export price including tariffs
- The exchange rate corrects the currency difference

_mod_1.append $COMPX = PX * (1 + r_tarx) / (1 + r_tarx0) / (PPX * ER)$

- Exports depend on
 - * World demand (both final and intermediate)
 - * The rate of use of capacities
 - * Price competitiveness

_mod_1.append $X = f * (WD + UR + COMPX)$

- Trade flows are computed in current terms

_mod_1.append $MVAL = PM * M$
_mod_1.append $XVAL = PX * X$

- ' The export import ratios are computed:
- ' * at current prices
- ' * at constant prices
- ' * for the deflators

```
_mod_1.append RCVAL=XVAL/MVAL
_mod_1.append RCVOL=X/M
_mod_1.append TTRAD=PX/PM
```

- ' The trade balance

```
_mod_1.append TRB=XVAL-MVAL
```

- ' Net interests paid to the Rest of the world
- ' are separated depend on currency.

- ' They depend on:
- ' * their past value
- ' * the rate on past debts
- ' * the rate on new debts
- ' the international rate is used for interests
- ' in foreign currency
- ' * the balance
- ' * the exchange rate for the debt in foreign currency

```
_mod_1.append NIXL=f*(NIXL(-1)+IRM+IR+TRB)
_mod_1.append NIXX=f*(NIXL(-1)+IRMX+IRX+TRB+ER)
_mod_1.append NIX=NIXL+NIXX
```

- ' The financing capacity

```
_mod_1.append FCAPX=TRB-NIX
```

' The State budget block

- ' Most of its elements have been computed already
- ' as transfers between the State and another agent

‘ Social contributions paid by firms

_mod_1.append $SCF=r_scf*Wf$

‘ Other indirect taxes

_mod_1.append $OIT=r_oit*QVAL$

‘ Tariffs

_mod_1.append $TAR=r_tar*MVAL$

‘ Social contributions paid by the State

_mod_1.append $SCG=R_SCG*WG$

‘ Total revenue

_mod_1.append $REVG=SCF+SCG+SCW+OIT+IFP+ICT+VAT+TAR+r_revg*QVAL$

‘ Government investment at current prices

_mod_1.append $IGV=IG*PIG$

‘ Government consumption at current prices

_mod_1.append $CGV=CG*PFD$

‘ Government demand at current prices

_mod_1.append $FDGV=CGV+IGV$

‘ Net interests paid by the State depend on

‘ * their past value

‘ * the rate on past debts

- ' * the rate on new debts
- ' * the balance

```
_mod_1.append NIG=NIG(-1)*IRM/IRM(-1)-IR/100*FCAPG
```

- ' Total expenditures

```
_mod_1.append EXPG=FDGV+WG+SUBS+SOCB+NIG+SCG+r_exp*QVAL
```

- ' Government balance

```
_mod_1.append FCAPG=REVG-EXPG
```

- ' Government balance in GDP points

```
_mod_1.append FCAPGP=100*FCAPG/GDPMVAL
```

- ' Total GDP
- ' The sum of market and non-market GDP
- ' equated to the total wage cost of the State

```
_mod_1.append GDPVAL = GDPMVAL+WG+SCG
```

```
'-----  
' End of model specifications  
'-----
```

10.2.2.2 Producing the data: an OECD example

Now we will present a program creating the data for the above model. We have used a very simple case, in which the model builder has access to the OECD "Economic Perspectives", a file containing about 5000 quarterly series describing the world economy. Each OECD country is described individually using the same concepts (with a few exceptions), and other important countries (like China) or zones (like Latin America) using less detail. However, the definitions vary slightly across countries, and although the same set of series is always technically present, some of them do not contain any value, or very few. For instance, the notion of savings is not always the same, and capital stock is not always available.

As we shall see, this base contains all the data we need for France, with one or two minor exceptions.

Obviously, this program is quite case dependent, and a user starting from another base (such as IMF) will use a rather different one. But:

- The user might actually start from this OECD base, a quite popular one.
- Some of the tasks (creating the workfile, producing a trend...) will appear in all cases.
- And actually the program will not be so different. Almost of the variables accessed in the OECD base will be available in any base of this type, and one will just have just to replace the "OECD" names, changing some of the concepts if needed.

The only additional problems for the prospective user of the program will come from:

- Unavailable data which has to be estimated, or guessed.
- Additional data required by the introduction of new elements in the model.

```
'=====
'  An example of data transfer
'=====

'  This program will start from the original French data
'  provided by OECD Economic perspectives and named fra_*,
'  the prefix used by OECD to identify French statistics

'  We decide on the directory
'  This is not generally necessary
'  except if one works on several projects
'  or maintains several directories for the same project
'  It guarantees trial versions do not destroy official ones

cd "c:\program files\EViews5\Book\fra_cf"

'  It will create the data for our model
'  with the prefix f_*

'-----
```

' This technique can be used with any source
' where the original series use the same prefix
' the results will be created with any different prefix

' In the best case, if the set available is the same (or larger) than the OECD set
' one has just to replace the OECD names in the following statements
'-----

' we close the original file fra_1
' containing the sole fra_* data
' and some global OECD series named OECD_*

' we close also the file which will receive the French data
' in case it is already open
' having two versions of the same file open in memory is quite dangerous...

```
close fra_1  
close data_1
```

' We open the original file (presently closed)
' and save it under the name data_1 for the French data

```
open fra_1  
save data_1
```

' Now the file should contain only original data
' called fra_*

' we delete any existing f_* series
' just in case, this should not happen...

```
delete f_*
```

' We have to make an assumption on the sharing of indirect taxes
' into VAT and other indirect taxes
' as OECD provides only a global variable

' p_oit = assumption on the share of oit in indirect taxes
,

```
scalar f_p_oit=0.2
```

' we create a time trend
' with the value of the year for the first quarter

' to which we add 0.25 for each following quarter of the year

' 1994 : 1994Q1
' 1994.25: 1994Q2
' 1994.50: 1994Q3
' 1994.75: 1994Q4
' 1995.00: 1995Q1

' This will be quite useful to:

' create yearly time trends
' replace actual dummy variables by expressions using logical conditions
' much easier to manage

```
smpl 1962Q1 1962Q1  
genr f_t=1962  
smpl 1962Q2 2004Q4  
genr f_t=f_t(-1)+0.25
```

' Now we start with the supply - demand equilibrium

```
smpl 1962Q1 2004Q4  
genr f_gdpval=fra_gdp
```

' Gross domestic product at current prices

```
genr f_gdp=fra_GDPV
```

' Gross domestic product at constant prices

```
-----  
' Supply  
-----
```

' The model separates market GDP

```
genr f_gdpmval=fra_gdp-fra_cgw
```

' = *gross domestic product at current prices – Government wage consumption at current prices*

```
genr f_pcog=fra_cg/fra_cgv
```

' = Government total consumption at current prices / = Government total consumption at constant prices

```
genr f_gdpm=fra_GDPV-fra_CGw/f_pcog  
genr f_pgdpm=f_gdpmval/f_gdpm
```

' Trade at current and constant prices, deflators

```
genr f_m=fra_MGSV  
genr f_x=fra_XGSV  
genr f_pm=fra_PMGS  
genr f_px=fra_PXGS  
genr f_xval=f_px*f_x  
genr f_mval=f_pm*f_m
```

' The deflator of final demand (including VAT and OIT)

```
genr f_pfd=(f_GDPMVAL+f_MVAL-f_XVAL)/(f_GDPM+f_M-f_X)  
genr f_fd=fra_TDDV-fra_CGw/f_pcog
```

' Identifying the indirect taxes This separation is important as VAT applies to final demand

```
genr f_oit= f_p_oit* fra_TIND  
genr f_vat = (1-f_p_oit)* fra_TIND  
genr f_r_vat =f_vat/(f_fd*f_pfd-f_vat)  
scalar f_r_vat0=@elem(f_r_vat,"1995")  
genr f_r_oit=f_oit/(f_gdpmval-f_vat-f_oit)  
scalar f_r_oit0=@elem(f_r_oit,"1995")  
genr f_pfdxt=f_pfd*(1+f_r_vat0)/(1+f_r_vat)
```

' Value added excluding VAT (but including OIT)

```
genr f_qval=f_gdpmval-f_vat  
genr f_q = f_gdpm-f_r_vat0*f_fd/(1+f_r_vat0)  
genr f_pq=f_qval/f_q
```

' Capital, rate of use, capacity

```
genr f_k=fra_KBV
genr f_ur=fra_GDPV/fra_GDPVTR
genr f_cap=f_q/f_ur
genr f_pk=f_cap/f_k(-1)
genr f_urd=1
```

'-----
' Demand
'-----

' Intermediate consumption
' Not identified in the OECD data base
' We specify the ratio as 1 (a common value)

```
genr f_tc=1
genr f_ic=f_tc*f_q
```

' Production price

```
genr f_pp=(f_qval+f_ic*f_pfdxt)/(f_q+f_ic)
```

' Elements of demand and their deflators
' Household consumption

```
genr f_coh=fra_CPV
genr f_pcoh=fra_cp/fra_cpv
```

' Investment, depreciation rate

```
genr f_i=fra_IBV
genr f_pi=fra_ib/fra_ib
genr f_rdep = (f_k(-1) + f_i - f_k) / f_k(-1)
```

' Housing investment

```
genr f_hih=fra_IHV
```

' Government investment (excluding State firms)

```
genr f_ig=fra_IGV
genr f_igv=fra_IG
genr f_pig=f_igv/f_ig
```

' Government consumption (excluding State firms)

```
genr f_cogv=fra_cg-fra_cgw
genr f_cog=(fra_cg-fra_cgw)/f_pcog
```

' Government demand

```
genr f_fdgv=f_cogv+f_igv
genr f_fdg=f_cog+f_ig
```

' Change in inventories

```
genr f_ci=fra_iskv
```

' Individual demand deflator ratios

```
genr f_r_pi=f_pi/f_pfd
genr f_r_pcoh=f_pcoh/f_pfd
genr f_r_pig=f_pig/f_pfd
genr f_r_pcog=f_pcog/f_pfd
```

' Wage rate

```
genr f_wr=fra_WAGE/(fra_ET-fra_ES)
genr f_r_scf = fra_WSSS/fra_WAGE -1
```

' Checking that the decomposition of final demand is correct

```
genr f_fdxr=(f_fd-f_coh-f_i-f_hih-f_ic-f_fdg)/f_q
```

'-----

' employment, unemployment, population

' Employment

```
gener f_lt=fra_ET
gener f_lg=fra_EG
gener f_lf=fra_ETB
```

' Labour productivity

```
gener f_pl=f_q / f_lf
```

' Wages and wage cost

```
gener f_w=f_wr*f_lt
gener f_wf=f_lf*f_wr
gener f_uwc=f_wr*(1+f_r_scf)/f_pl
```

' Unemployment

```
gener f_un=fra_UN
gener f_unr=fra_UN/(fra_ET+fra_UN)
```

' Population

```
gener f_pop=fra_POPT
gener f_pop65=fra_POPT
gener f_popac=f_lt+f_un
```

' Financial elements

' Exchange rate

```
gener f_er=1/(fra_EXCHEB/@elem(fra_EXCHEB,"1995"))
gener f_erx=f_er
```


' Nominal interest rates

```
gener f_irs=fra_IRS
gener f_irl=fra_IRL
gener f_ir = fra_irwyp
gener f_irm=f_ir
gener f_irsx=fra_irsaf
gener f_irlx=fra_irlaf
gener f_irmx=fra_irfor
gener f_irx=fra_irfor
```

' Real interest rates

' Here we assume the formulas used for the short term
' and the average interest rate on past debts
' their logic will be explained later

```
gener f_irr=f_ir-100*@pchy(f_pcoh)
gener f_irl_ec=0
gener f_irsr=f_irs-100*@pchy(f_pcoh)
gener f_irst=f_irs-(150*@pchy(f_PCoh)+50*(f_UR-f_urd)/f_urd)
scalar p_f_irm=0.8
gener f_irm_er=f_IRM-(p_f_irm*f_IRM(-1)+(1-p_f_irm)*f_IR)
```

' relative cost (for substitution between labor and capital)

```
gener f_relc=f_wr*(1+f_r_scf)/f_pi/(f_ir/100-@pchy(f_pcoh)-4*log(1-f_rdep))
gener f_spread=3
```

' Households

' Global elements

```
gener f_hi =fra_YRH-fra_TRPH
gener f_ict=fra_tyh
gener f_r_ict = fra_TYH / f_hi(-1)
gener f_hdi=f_hi-fra_tyh
gener f_hrdis=f_hdi/f_pcoh
```

```
gener f_sr=1-f_coh/(f_hrdi)
```

' Individual elements

```
gener f_socb=fra_trrh  
gener f_scw= fra_TRPH - f_r_scf*f_w  
gener f_r_scw=f_scw/f_w  
gener f_rpro=f_hi-(f_w-f_scw+f_socb)  
gener f_revx=(1-0.5)*f_rpro  
gener f_revq=0.5*f_rpro  
gener f_r_revx=f_revx/f_pfd  
gener f_r_revq=f_revq/f_qval  
gener f_wg=f_wr*f_lg  
gener f_hdir=f_hdi/f_pcoh
```

```
'-----  
' external trade  
'-----
```

' Tariffs

```
gener f_tar=0  
gener f_r_tar=f_tar/f_mval  
gener f_r_tarx=0  
scalar f_r_tarx0=@elem(f_r_tarx,"1995")  
scalar f_r_tar0=@elem(f_r_tar,"1995")
```

' Foreign elements

```
gener f_wd = fra_XMVMKT  
gener f_ppx=OECD_PGDP
```

' Competitiveness

```
gener f_PMT=f_PM*(1+f_r_tar)/(1+f_r_tar0)  
gener f_compm=f_pmt/f_pp  
gener f_COMPX=f_PX*(1+f_r_tarx)/(1+f_r_tarx0)/(f_PPX*f_ER)
```

' Ratios and balances

```

gener f_rcvol=f_x/f_m
gener f_rcval=f_xval/f_mval
gener f_ttrad=f_px/f_pm
gener f_trb=f_xval-f_mval
gener f_FCAPX=fra_FBGSV

```

' Interests paid

- ' In the OECD data set, the interests paid to the RoW are not identified
- ' We apply the formulas used by the model
- ' They will be explained later

```

smpl 1990Q1 2004Q4
scalar p_nix=0.5
gener f_nixl=-f_ir*f_trb/40*p_nix
gener f_nixx=-f_irx*f_trb/40*(1-p_nix)
smpl 1990Q2 2004Q4
gener f_nixl=(f_nixl(-1)*f_irm/f_irm(-1)-f_ir*f_trb/400*p_nix)/(1-f_ir/400)
gener f_nixx=(f_nixx(-1)*f_irmx/f_irmx(-1)*f_er/f_er(-1)-f_irx*f_trb/400*(1-p_nix))/(1-f_irx/400)
gener f_nix=f_nixl+f_nixx
gener f_fcpx=f_trb-f_nix
smpl 1962Q1 2004Q4

```

' Government budget

' Revenue

```

gener f_scf=f_r_scf*f_wf
gener f_r_scg= f_r_scf
gener f_scg=f_wg*f_r_scf
gener f_ifp=fra_tyb
gener f_r_oit=f_oit/(f_qval)
gener f_revf=f_ict+f_oit+f_vat+f_scf+f_scg+f_tar+f_scw+f_ifp
gener f_r_revf=0
gener f_recg=0
gener f_r_recg=0

```

' Expenditures

```

gener f_socbr = fra_TRRH /f_pcoh/f_pop

```

```
gener f_subs=fra_TSUB
gener f_r_subs = fra_TSUB / (f_qual)
```

' Interest rates and balance

```
gener f_nig=fra_gnintp
gener f_fcpg=-FRA_NLG
gener f_nig_er=(f_NIG-(f_NIG(-1)*f_IRM/f_IRM(-1)-f_IR/400*f_FCAPG))/f_qual
gener f_expg=f_revq-f_fcpg
gener f_r_expg=(f_expg-(f_fdgv+f_wg+f_scg+f_nig+f_socb+f_subs))/f_qual
```

' Balance in GDP points

```
gener f_fcpgp=100*f_fcpg/f_gdpval
```

' Firms account

' Margins

```
gener f_marg=f_qual*(1+f_r_subs-f_r_oit)-f_wf*(1+f_r_scf)
gener f_fmarg =f_marg / f_qual
```

' Interests paid, profits and balance

```
gener f_prof=fra_PROF
smpl 1978Q1 2004Q4
gener f_fcpcf=-fra_NLB

smpl 1978Q1 1978Q1
gener f_nif=f_ir*f_fcpcf/11
smpl 1978Q2 2004Q4
gener f_nif=f_nif(-1)*f_irm/f_irm(-1)-f_ir/400*f_fcpcf
smpl 1962Q1 2004Q4
gener f_rprob = f_marg/(f_pfd*f_k(-1))
gener f_prof1=f_marg-f_revq-f_ifp-f_nif
gener f_prof_er=(f_prof-f_prof1)/f_qual
gener f_r_IFP=f_ifp/(f_PROF(-1)+f_IFP(-1))
gener f_rprof=f_prof/(f_pfd*f_k(-1))
gener f_fcpcf_er=(f_fcpcf-(f_prof-f_pi*f_i-f_pfd*f_ci))/f_qual
```

```
-----  
' Damping factors  
-----
```

```
gener f_relax_q=1  
gener f_relax_pfd=1  
  
save data_1
```

10.2.2.3 Creating the model groups and checking the data – equations consistency

These two tasks can be started only when the model has been defined, and the data created.

Creating the groups needs not only the model, but also the data, as EViews groups can only be built from (existing) series.

This program can be run after the two tasks, in the same session. This is why we did not introduce the usual statements specifying the directory and opening the workfile.

The program is model independent. It:

- Creates a list of endogenous and exogenous variables.
- Solves the model equations separately from each other twice, for different values of the scalar « f ».
- Separates the endogenous into identity and behavioral using the fact that the equations for the latter contain the scalar « f »
- Computes the level and relative differences between historical values and the results of one of the simulations above.

```
=====  
' creating the groups  
=====
```

```
' We create groups for the endogenous and exogenous
```

```
' The statement « makegroup »
```

```
' * applies to the model « _mod_1 »
```

```
' * creates a group of endogenous (@endog) or exogenous (@exog)
```

```
' * with a name _g_vendo or _g_vexog
```

```
' The modifiers specify the elements in the group
```

```
' « n » means we do not want the suffixed elements
```

```
' (suffixed by the current suffix, actually « _c »)
```

```
' « a » means we want the actual names (no suffix)
```

```
' This overrides strange EViews default options
```

```
_mod_1.makegroup(a,n) _g_vendo @endog
_mod_1.makegroup(a,n) _g_vexog @exog
```

- ' Now we want to separate identity and behavioral variables
- ' We shall use a trick
- ' The future behavioral equations use the scalar "f"
- ' We solve the model twice for different values of "f"
- ' using different suffixes for the solution

- ' We need to "update" the model
- ' for the parameter change to be taken into account

```
smpl 1995Q1 2004Q4
f=1
_mod_1.append assign @all _d
solve(d=f) _mod_1
f=2
_mod_1.update
_mod_1.append assign @all _c
solve(d=f) _mod_1
```

- ' We create two empty groups

```
group _g_vbeha
group _g_viden
```

- ' Then we consider in turn each endogenous variable
- ' `_g_vendo.@count` is the number of elements in `_g_vendo`
- ' `_g_vendo.@seriesname(!i)` is the name of the *i*th element in the group
- ' We consider the variable built from this name and the two suffixes "`_c`" and "`_d`"
- ' If the values are different then the variable is behavioral
- ' We test the condition on a given period
- ' (but not the base year as there is a chance the right hand side gives zero
- ' in which case the method does not work)
- ' If it is true we add the element to the list of behavioral elements
- ' otherwise to the list of identity elements
- ' « `endif` » close the condition
- ' « `next` » closes the loop

- ' the brackets « `{` » and « `}` » delimitate the parameter
- ' they are dropped after the replacement

```
for !i=1 to _g_vendo.@count
%1=_g_vendo.@seriesname(!i)
scalar {%1}_eq=(@elem({%1}_c,"2000Q1")<>@elem({%1}_d,"2000Q1"))
```

```

if {%1}_eq=1 then
_g_vbeha.add {%1}
else
_g_viden.add {%1}
endif
next

```

```

' =====
'  computing the residuals
' =====

' Now we check the residuals

' We define the suffix as « _c »

```

```

_mod_1.append assign @all _c

```

```

' We reduce the sample to a period
' for which each equation can be computed
' taking into account the presence of lags

```

```

smpl 1995Q2 2004Q4

```

```

' But only for identities
' However, we were able to compute the « behavioral »
' over the given period
' which means that we will be able to estimate

' The dc_{%1} are the absolute differences
' The pc_{%1} the relative ones
' To avoid dividing by zero, we use a trick
' We add to the variable a boolean one
' Testing if the variable is zero
' If it is true we get 0 divided by 1 = 0
' If not the computation is not affected
'

' This means we have just modified the outcome of divisions.
' so that dividing zero by zero now gives zero

```

```

for !i=1 to _g_viden.@count
%1=_g_viden.@seriesname(!i)
gener dc_{%1}={%1}-{%1}_c
gener pc_{%1}=100*dc_{%1}/({%1}+({%1}=0))
next

```

10.2.2.4 Estimating the equations: examples

We shall now apply the above principles to the estimation of equations. Basically, this means replacing in the previous program the identities declaring the modeller's intentions by actual equations.

We shall only present the new program sequences, associated with this process. The full model building program is available as an annex, and includes a summary of the comments we are going to make. It is also available as a .prg file on the model site.

Of course, the results we shall obtain are specific to the French case. They should only be considered as a (working) example, and will probably prove much less directly useful than the previous programs.

One will observe that we are not always following a pure and clean methodology. Our opinion is that we behave better than most. And anyway rather than a perfect element, what we are proposing here an example of the general methodology used by model builders, including deviations from the "politically correct" path.

For instance, we shall use cointegration only once, and build most error correction models in one step. You can guess (rightly) that we could not find a formulation fitting the required conditions, either because one of the elements was stationary, cointegration was not present, or the coefficients in the cointegrating equation were not significant or satisfactory.

We shall deal with the equations in sequence, considering two cases for the production function: complementary factors and Cobb-Douglas. The choice will affect other elements, as estimating a different capacity will affect the rate of use, an explanatory variable.

10.2.2.4.1 The complementary factors case

Let us first deal with the complementary factors case. The framework is very similar to the one we have seen earlier. Investment and employment are estimated separately.

10.2.2.4.2 Estimating investment

As in the small model, capacity will be defined by capital.

But now we shall estimate capital as the decision variable, and deduct investment from the previously capital equation, as the level which allows to reach the estimated capital value accounting for depreciation:

$$I_t = K_t - K_{t-1} \cdot (1 - dr_t)$$

The capital equation will be:

$$\Delta \text{Log}(K_t) = a \cdot \Delta \text{Log}(K_{t-1}) + b \cdot (0.5 \cdot \text{Log}(Q_t / Q_{t-4}) - \text{Log}(UR_t)) + c \cdot 1/3 \cdot \sum_{i=0}^2 RPROB_{t-i} + d \cdot (t - 2005) + e$$

or in EViews terms:

```
equation _eq_k.ls(p) dlog(k)=c_k(1)*dlog(k(-1))+0.50*c_k(2)*(q-q(-4))/q(-4)
-c_k(2)*(1-ur)/ur+c_k(4)*@movav(rprob,3)+c_k(5)*(t-2004)*(t<=2004))+c_k(6)+k_ec
```

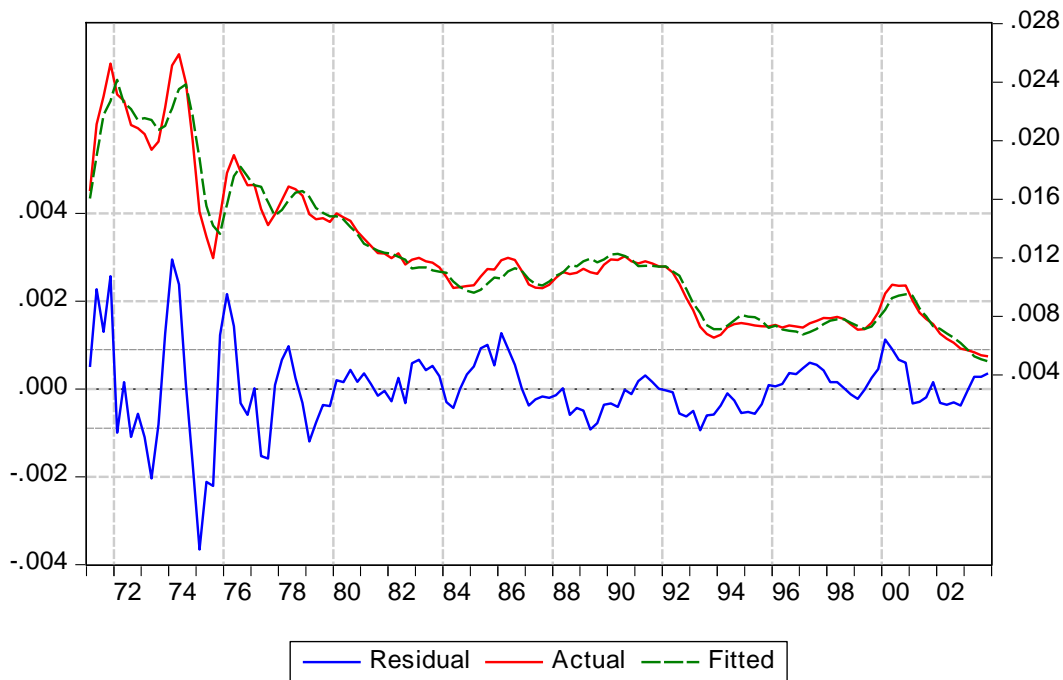
In our equation, the change in capital will come from:

- The change of value added over the last year, representing the firms' expectations for future growth.
- The target change in the rate of use.
- The expected profitability of capital, represented by the average of the profits rate over the last three quarters.

We also had to introduce a time trend, which we shall stop at the end of the sample period.

Dependent Variable: DLOG(K)
Method: Least Squares
Date: 11/07/12 Time: 18:51
Sample (adjusted): 1971Q1 2003Q4
Included observations: 132 after adjustments
DLOG(K)=C_K(1)*DLOG(K(-1))+0.50*C_K(2)*(Q-Q(-4))/Q(-4)-C_K(2)*(URD-UR)/UR+C_K(4)*@MOVAV(RPROB,3)+C_K(5)*(T-2005)*(T<=2005))+C_K(6)

	Coefficient	Std. Error	t-Statistic	Prob.
C_K(1)	0.602302	0.046208	13.03471	0.0000
C_K(2)	0.019733	0.004273	4.618419	0.0000
C_K(4)	0.011318	0.002696	4.198355	0.0001
C_K(5)	-0.000117	1.85E-05	-6.338253	0.0000
C_K(6)	0.000221	0.000460	0.480779	0.6315
R-squared	0.966687	Mean dependent var		0.012048
Adjusted R-squared	0.965638	S.D. dependent var		0.004818
S.E. of regression	0.000893	Akaike info criterion		-11.16664
Sum squared resid	0.000101	Schwarz criterion		-11.05745
Log likelihood	741.9985	Hannan-Quinn criter.		-11.12227
F-statistic	921.3380	Durbin-Watson stat		0.788728
Prob(F-statistic)	0.000000			



The results are satisfactory even though we had to equalize the coefficients for the accelerator and the rate of use, and introduce a trend which will be dropped in the future.

10.2.2.4.2.1 ESTIMATING EMPLOYMENT (COMPLEMENTARY FACTORS CASE)

Again, we shall use the same framework as in the small model. We start by defining the labor productivity trend:

```

smpl 1962Q1 2003Q4
coef(10) c_plt
equation _eq_plt.ls(p) LOG(Q/LF) = c_plt(1) + (t-2004)*(c_plt(2)*(t<=2004)
+4*log((1+txq)/(1+txn))*(t>2004))+c_plt(3)*((T<=1975)*(T-1975)) + c_plt(4)*((T<=1992)*(T-1992))
_eq_plt.resids(p)
close _eq_plt
smpl 1962Q1 2004Q4
genr log(plt) = c_plt(1) + (t-2004)*(c_plt(2)*(t<=2004)+4*log((1+txq)/(1+txn))*(t>2004))
+c_plt(3)*((T<=1975)*(T-1975)) + c_plt(4)*((T<=1992)*(T-1992))

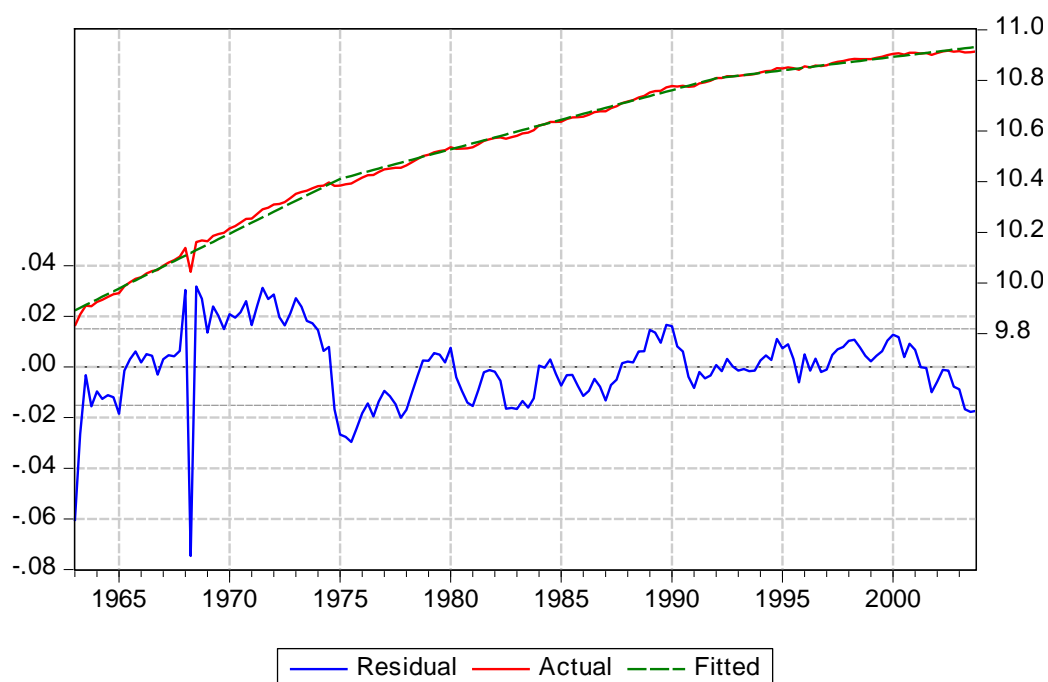
genr lfd=q/plt

```

Dependent Variable: LOG(Q/LF)
 Method: Least Squares
 Date: 11/07/12 Time: 18:53
 Sample (adjusted): 1963Q1 2003Q4
 Included observations: 164 after adjustments

$$\text{LOG(Q/LF)} = C_PLT(1) + (T-2005)*(C_PLT(2)*(T \leq 2005) + 4*\text{LOG}((1+TXQ)/(1+TXN))*(T > 2005)) + C_PLT(3)*((T \leq 1975)*(T-1975)) + C_PLT(4)*((T \leq 1992)*(T-1992))$$

	Coefficient	Std. Error	t-Statistic	Prob.
C_PLT(1)	10.94511	0.004502	2431.041	0.0000
C_PLT(2)	0.010535	0.000469	22.47874	0.0000
C_PLT(3)	0.019946	0.000646	30.89304	0.0000
C_PLT(4)	0.012821	0.000658	19.49688	0.0000
R-squared	0.997580	Mean dependent var	10.54892	
Adjusted R-squared	0.997534	S.D. dependent var	0.303824	
S.E. of regression	0.015086	Akaike info criterion	-5.525963	
Sum squared resid	0.036416	Schwarz criterion	-5.450357	
Log likelihood	457.1290	Hannan-Quinn criter.	-5.495270	
F-statistic	21983.12	Durbin-Watson stat	0.812272	
Prob(F-statistic)	0.000000			



The two breaks take place in 1972 and 1992 (first quarters in each case).

As in the previous case, we observe that actual labor productivity is stationary around the trend.

This gives us target employment, from which we build the error-correction model:

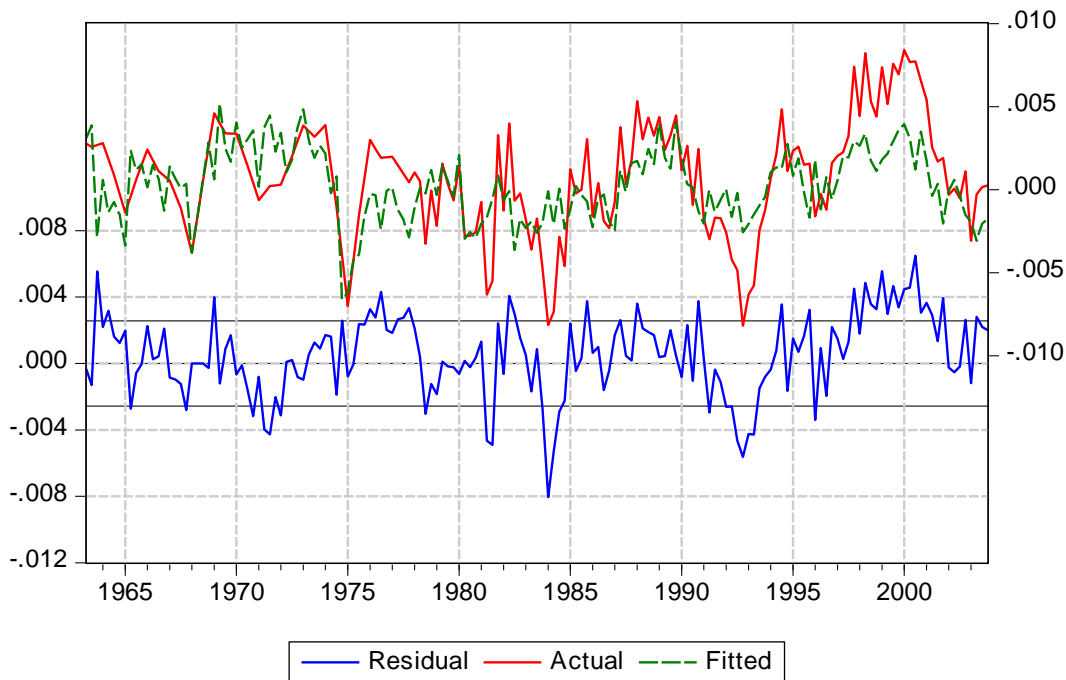
```

coef(10) c_lf
smpl 1962Q1 2004Q4
genr lf_ec=0
smpl 1962Q1 2003Q4
equation _eq_lf.ls(p) dlog(lf)=c_lf(1)*dlog(lfd)+c_lf(2)*log(lfd(-1)/lf(-
1))+c_lf(3)*(t=1968.25)+c_lf(4)*(t=1968.50)+c_lf(5)*(t=1968)+lf_ec
_eq_lf.resids(p)
close _eq_lf
genr lf_ec=resid

```

Dependent Variable: DLOG(LF)
 Method: Least Squares
 Date: 11/07/12 Time: 18:54
 Sample (adjusted): 1963Q2 2003Q4
 Included observations: 163 after adjustments
 DLOG(LF)=C_LF(1)*DLOG(LFD)+C_LF(2)*LOG(LFD(-1)/LF(-1))+C_LF(3)
 (T=1968.25)+C_LF(4)(T=1968.50)+C_LF(5)*(T=1968)

	Coefficient	Std. Error	t-Statistic	Prob.
C_LF(1)	0.260651	0.028089	9.279331	0.0000
C_LF(2)	0.107863	0.015245	7.075301	0.0000
C_LF(3)	0.022898	0.003906	5.861669	0.0000
C_LF(4)	-0.019260	0.003948	-4.877970	0.0000
C_LF(5)	-0.009769	0.002640	-3.700988	0.0003
R-squared	0.395867	Mean dependent var		0.000753
Adjusted R-squared	0.380573	S.D. dependent var		0.003267
S.E. of regression	0.002571	Akaike info criterion		-9.058593
Sum squared resid	0.001045	Schwarz criterion		-8.963693
Log likelihood	743.2754	Hannan-Quinn criter.		-9.020065
Durbin-Watson stat	0.907945			



This time we identify three dummy variables for 1968, a very particular year in France especially for the labor market.

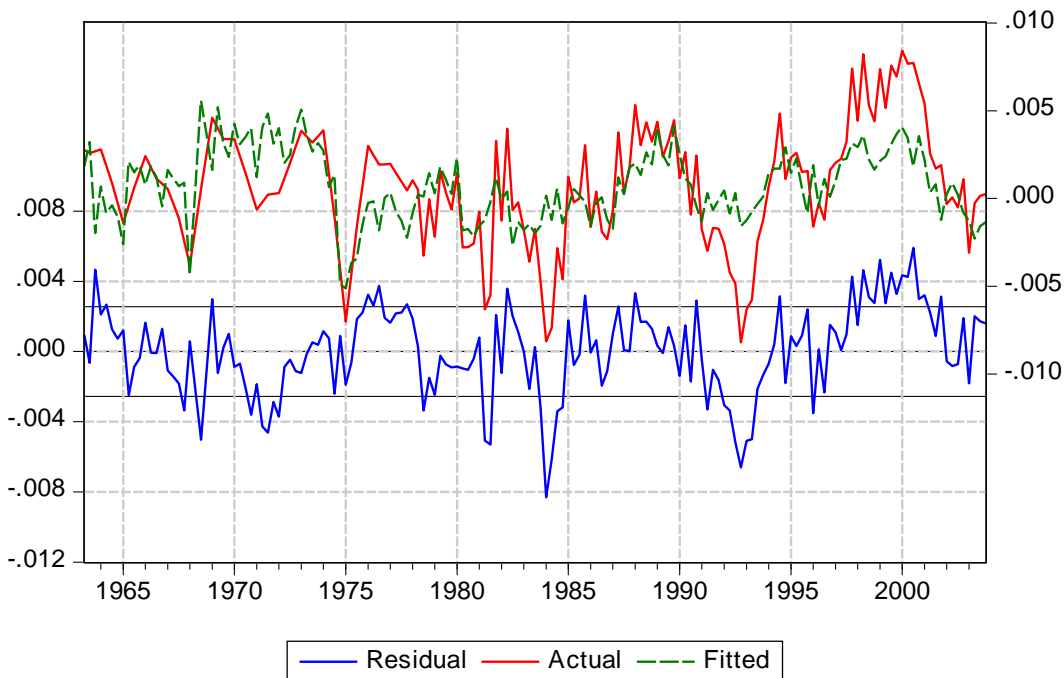
During the whole second quarter, a so-called “student revolution” actually blocked the economy, in particular transportation. GDP went down sharply, but employment did not really follow, as managers guessed (rightly) that this situation was transitory. This calls for a dummy for that quarter (with a positive coefficient). The negative coefficients are also explained by the particular situation. In the first quarter, GDP was actually growing without job creation. In the third, the increase in GDP could be met by people not laid out earlier.

Actually, this transitory process could be expressed by a single dummy variable with values -0.5, 1 and -0.5 in the three relevant quarters. The zero sum makes the impact transitory.

Dependent Variable: DLOG(LF)
Method: Least Squares
Date: 11/07/12 Time: 18:58
Sample (adjusted): 1963Q2 2004Q4
Included observations: 167 after adjustments
DLOG(LF)=C_LF(1)*DLOG(LFD)+C_LF(2)*LOG(LFD(-1)/LF(-1))+C_LF(3)
(-0.5(T=1968)+1*(T=1968.25)-0.5*(T=1968.50))+C_LF(4)

	Coefficient	Std. Error	t-Statistic	Prob.
C_LF(1)	0.220392	0.024174	9.116813	0.0000
C_LF(2)	0.111036	0.014601	7.604636	0.0000
C_LF(3)	0.020235	0.003322	6.091305	0.0000
C_LF(4)	0.000546	0.000198	2.761393	0.0064

R-squared	0.394013	Mean dependent var	0.000756
Adjusted R-squared	0.382860	S.D. dependent var	0.003228
S.E. of regression	0.002536	Akaike info criterion	-9.093091
Sum squared resid	0.001048	Schwarz criterion	-9.018408
Log likelihood	763.2731	Hannan-Quinn criter.	-9.062779
F-statistic	35.32752	Durbin-Watson stat	0.822083
Prob(F-statistic)	0.000000		



10.2.2.4.3 The Cobb-Douglas case

In the framework of a policy model, the complementary factors option is quite questionable and limited. But it is also quite simple to implement, and leads to easier interpretation of properties. Moving to a Cobb-Douglas specification might be simple at first sight. For instance we could consider modifying slightly the complementary factors framework,

defining employment and capital separately based on a capacity target, and simply establishing the capacity using the Cobb-Douglas function.

But then we lose the most interesting property of this framework: taking into account the (endogenous) sensitivity of labor and capital factors to the ratio of their relative costs.

This means that a shock producing a 1% increase in target capacity would lead in the long run to the same relative increase in both factors, even if the same shock has modified the ratio of labor and capital costs.

We think this is too simplistic, and our formulas will take into account this effect. This will show for instance that a decrease in the social security contributions of firms will especially favor employment, and that the effect of a demand shock on unemployment and wages will favor capital.

The Cobb-Douglas assumption supposes a unitary elasticity of the share of factors to the relative cost.

However, using this assumption calls for a much more complicated framework. Let us consider its elements in turn.

10.2.2.4.3.1 THE FRAMEWORK

We shall now describe the framework in detail (more than we did earlier).

10.2.2.4.3.1.1 MARGINS MAXIMIZATION

In this framework, firms will try to maximize their margins

$$(1) \quad pq_t \cdot Q_t - w \cos t_t \cdot LE_t - k \cos t_t \cdot K_t$$

under constraint of the production function:

$$(2) \quad \text{Log}(CAP_t) = \alpha \cdot \text{Log}(LE_t) + (1 - \alpha) \cdot \text{Log}(K_t) + \beta \cdot t + \gamma$$

which leads us to maximizing

$$(3) \quad pq_t \cdot Q_t - w \cos t_t \cdot LE_t - k \cos t_t \cdot K_t - \lambda \cdot \left(CAP_t - \exp(\beta \cdot t + \gamma) \cdot LE_t^\alpha K_t^{1-\alpha} \right)$$

relative to both LE_t and K_{t-1} .

Derivation of (3) gives:

$$-w \cos t_t - \lambda \cdot \alpha \cdot \exp(\beta \cdot t + \gamma) \cdot LE_t^{\alpha-1} K_t^{1-\alpha} = 0$$

Or

$$(4) -w \cos t_t - \lambda \cdot \alpha \cdot \exp(\beta \cdot t + \gamma) \cdot (K_t / LE_t)^{1-\alpha} = 0$$

and equivalently

$$(5) -k \cos t_t - \lambda \cdot (1-\alpha) \cdot \exp(\beta \cdot t + \gamma) \cdot (K_t / LE_t)^\alpha = 0$$

Dividing (4) by (5) on both sides gives:

$$(6) w \cos t_t / k \cos t_t = \alpha / (1-\alpha) \cdot (K_t / LE_t)$$

which shows indeed the unitary elasticity of the ratio of factors to the ratio of costs.

From (2) and (6) we get:

$$(7) \begin{aligned} \text{Log}(LE_t / CAP_t) &= (1-\alpha) \cdot \text{Log}(LE_t / K_t) - \beta \cdot t - \gamma \\ &= (1-\alpha) \cdot \text{Log}(\alpha / (1-\alpha) k \cos t_t / w \cos t) - \beta \cdot t - \gamma \end{aligned}$$

$$(8) \begin{aligned} \text{Log}(K_t / CAP_t) &= -\alpha \cdot \text{Log}(LE_t / K_t) - \beta \cdot t - \gamma \\ &= -\alpha \cdot \text{Log}(\alpha / (1-\alpha) k \cos t_t / w \cos t) - \beta \cdot t - \gamma \end{aligned}$$

But to apply this framework to a full model, we have to take into account several elements.

10.2.2.4.3.1.2 TARGETS AND ACTUAL VALUES

The above presentation applies to targets: knowing the relative costs, firms will estimate a target level of capacity, then the target levels of factors which allow reaching this capacity.

This means that we have to define this capacity, then compute the target level for both factors, then define the process leading from target values to actual values.

10.2.2.4.3.1.3 CAPACITY

For defining capacity, we shall suppose that firms have a target rate of use, constant over the period. This means that target capacity is proportional to production (or rather value added). In the estimation of equations (7) and (8), we will replace the ratio of factors to capacity by the ratio of factors to production, without loss of generality (the target rate of use will be absorbed by the constant term).

10.2.2.4.3.1.4 THE TIME FACTOR

In the above framework, we have used only instantaneous elements. But we have to consider the nature of our variables.

- Capital and capacity are measured at a given point in time
- Employment, in our definition, is an average level across one period (one quarter).

We shall suppose that target capital, employment (and implicit target capacity) are given by the system, using as target the actual level of production, but that actual capacity for the period is given by actual employment and the initial level of capital. As capital is measured as end-of-period, we shall use the lagged value.

10.2.2.4.3.1.5 THE INERTIA OF FACTORS

We shall also assess that optimal decisions are not implemented immediately. As explained before, the reasons are both technical (the length of the investing process) and psychological (risk aversion). So firms go only part of the way to the target, starting from the previous decision level. We shall try to estimate the inertia factor, allowing different values for labor and capital.

The first factor should be less inert, as the penalty for errors is lower (we have to consider annual wages compared to full cost of capital), and managing their consequences is also easier (laying down workers is easier than selling back unneeded equipment).

10.2.2.4.3.1.6 THE RELATIVE COST

Of course, it should compare the price of capital (actually investment¹⁴⁵) to the price of labor (the wage rate). Actually, things are a little more complex.

- The wage rate should include social contributions.

¹⁴⁵ Remember that the evolution of capital quality changes its value at constant prices, not its deflator, measured at a given quality (or efficiency). This means that the present deflator applies also to past values of capital, as the efficiency of a quantity is independent from the time at which it has been purchased. Applied to a given capital level, it will give the value of investment necessary to replace this capital, with the same efficiency.

- Once purchased, capital can be used as long as it is not destroyed or obsolete¹⁴⁶, whereas employment is bought for a single period¹⁴⁷.
- The price of capital should take into account the fact that it has to be purchased at once, whereas the alternate factor, labor, is paid for at the moment it is used, or even later. This delay should call for the introduction of the interest rate.
- The price of labor is measured in time spent. This means it is expected to increase in real terms with labor productivity.
- Capital depreciates over time. For workers, the efficiency generally increases then decreases, depending on the product. But firms can always replace older workers by new ones, at minimal cost (retirement financing is generally included in the wage cost).

We shall actually compare the yearly wage cost with the price of investment, equivalent to the price of capital at the cost of renewal. One has to consider that the increase in the efficiency of capital is included in the variable at constant prices, not in the deflator (this is called the « quality effect »). To spread the cost of capital over its period of use, we shall divide its deflator by an estimated factor, which should associate more or less with the number of periods of its productive life.

$$relc = (wf \cdot (1 + r_{scf}) / pi) \cdot f$$

$$f = 1 / ((ir - pc) - 4 \cdot \text{Log}(1 - rdep))$$

Some explanations are needed for the last term. The increase in the cost of capital for a given period is the sum of the interest rate and the annual depreciation:

$$z = (ir - pc) + 4 \cdot \text{Log}(1 - rdep)$$

This means the ratio between the cost of labor and capital will give to the first an advantage of:

$$1 / (1 + z^k)$$

Or over the full period of

$$1 / (1 + z) + 1 / (1 + z)^2 + 1 / (1 + z)^3 + \dots = 1 / z$$

¹⁴⁶ Maybe with a decreasing efficiency.

¹⁴⁷ This was different in the Roman Empire, when you could buy slaves. Then labor and capital had closer properties.

10.2.2.4.3.2 THE ESTIMATION

Let us see how we can apply this technique in our model, under EViews. We have seen that we have to estimate two equations, with common coefficients. This will be done through a system, a feature which we will have now the opportunity to present.

First let us compute the relative cost, as explained above:

```
smp1 1962Q1 2004Q4
gener relc=wr*(1+r_scf)/pi/(ir/100-@pchy(pc)-4*log(1-rdep))
gener k_ec=0
gener lf_ec=0
```

Now we have to estimate a system of two equations with common coefficients.

- As usual we create two vectors: coefficients and parameters.
- We destroy any preexisting system of the same name.
- We initialize the values with a guess.
- (this is important as convergence of this system is a little tricky under EViews)

```
coef(10) c_cd
vector(10) p_cd
delete cd
smp1 1962Q1 2003Q4
system cd
c_cd(4)=0.1
c_cd(1)=100
c_cd(3)=0.02 ' the yearly growth of total factor productivity: guessed at 2%
c_cd(2)=0.65 ' the share of employment in the process, guessed at 0.65148
p_cd(8)=0.0
p_cd(9)=0.0
cd.append log(k/q(-1))=-c_cd(3)*(t-2005)-c_cd(7)+c_cd(2)*log(relc)
cd.append log(lf/q(-1))=-c_cd(3)*(t-2005)-c_cd(1)+(c_cd(2)-1)*log(relc)
```

Two methods are available; Full Information Maximum Likelihood (FIML), and Seemingly Unrelated Regression (SUR, Zellner). We test both but we chose the latter.

¹⁴⁸ A figure also close to the share of the wage cost in the total.

System: CD
 Estimation Method: Seemingly Unrelated Regression
 Date: 11/07/12 Time: 19:18
 Sample: 1968Q1 2005Q4
 Included observations: 152
 Total system (balanced) observations 304
 Linear estimation after one-step weighting matrix

	Coefficient	Std. Error	t-Statistic	Prob.
C_CD(3)	0.005050	0.000587	8.605400	0.0000
C_CD(7)	6.454626	0.166179	38.84134	0.0000
C_CD(2)	0.637878	0.014030	45.46507	0.0000
C_CD(4)	0.015137	0.002106	7.187983	0.0000
C_CD(1)	6.908534	0.166183	41.57180	0.0000
Determinant residual covariance		7.51E-05		

Equation: $\text{LOG}(K/Q(-1)) = -C_CD(3)*(T-2005) - C_CD(7) + C_CD(2) * \text{LOG}(\text{RELC}) - C_CD(4)*(T \leq 1982)*(T-1982)$

Observations: 152

R-squared	0.154620	Mean dependent var	0.480164
Adjusted R-squared	0.137484	S.D. dependent var	0.277316
S.E. of regression	0.257548	Sum squared resid	9.816973
Durbin-Watson stat	0.010576		

Equation: $\text{LOG}(LF/Q(-1)) = -C_CD(3)*(T-2005) - C_CD(1) + (C_CD(2)-1) * \text{LOG}(\text{RELC}) - C_CD(4)*(T \leq 1982)*(T-1982)$

Observations: 152

R-squared	-0.210476	Mean dependent var	-10.63960
Adjusted R-squared	-0.235013	S.D. dependent var	0.232456
S.E. of regression	0.258331	Sum squared resid	9.876747
Durbin-Watson stat	0.005369		

The estimation shows a 0.64 coefficient for labor (a very usual value) and a significant positive trend in global factor productivity.

We generate the estimated ("desired") series as target values.

```

smpl 1970Q4 2004Q4
genr log(kd/q(-1))=-c_cd(3)*(t-2005)-c_cd(7)+c_cd(2)*log(relc)
genr log(lfd/q(-1))=-c_cd(3)*(t-2005)-c_cd(1)+(c_cd(2)-1)*log(relc)

```

We estimate the actual results as weighted averages of the estimated result and the previous actual value, with an additional trend.

Actually we have fixed the inertia factors, at levels which give to the model reasonable properties.¹⁴⁹

```
smpl 1970Q4 2003Q4
coef(10) c_k
vector(10) p_k
genr k_ec=0
coef(10) c_lf
vector(10) p_lf
genr lf_ec=0
p_lf(1)=0.20
p_k(1)=1.00
equation _eq_k.ls(p) log(k/q(-1))=p_k(1)*log(kd/q)+(1-p_k(1))*log(k(-1)/q(-1))+c_k(2)+c_k(3)*(t-
2005)*(t<=2005)+k_ec
genr k_ec=resid
equation _eq_lf.ls(p) log(lf/q(-1))=p_lf(1)*log(lfd/q)+(1-p_lf(1))*log(lf(-1)/q(-1))+c_lf(2)+c_lf(3)*(t-
2005)*(t<=2005)+lf_ec
genr lf_ec=resid
```

Finally we generate capacity applying the estimated formula to actual factor values, and we compute the rate of use of capacities.

```
genr log(cap)=c_cd(7)*(1-c_cd(2))+c_cd(1)*c_cd(2)+c_cd(3)*(t-
2005)*(t<=2005)+c_cd(2)*4*log((1+txq)/(1+txn))*(t-2005)*(t>2005)+c_cd(2)*log(lf)+(1-c_cd(2))*log(k(-1))
genr ur=q/cap
```

We have to observe that this rate of use is different from the one we got with the complementary factors function. This means that the equations in which it enters will have to be estimated again: value added deflator, exports and imports.

10.2.2.4.4 The change in inventories

We shall use the same explanation as the simple model, except for a (negative) role of previous decisions, and a trend ending in 1985.

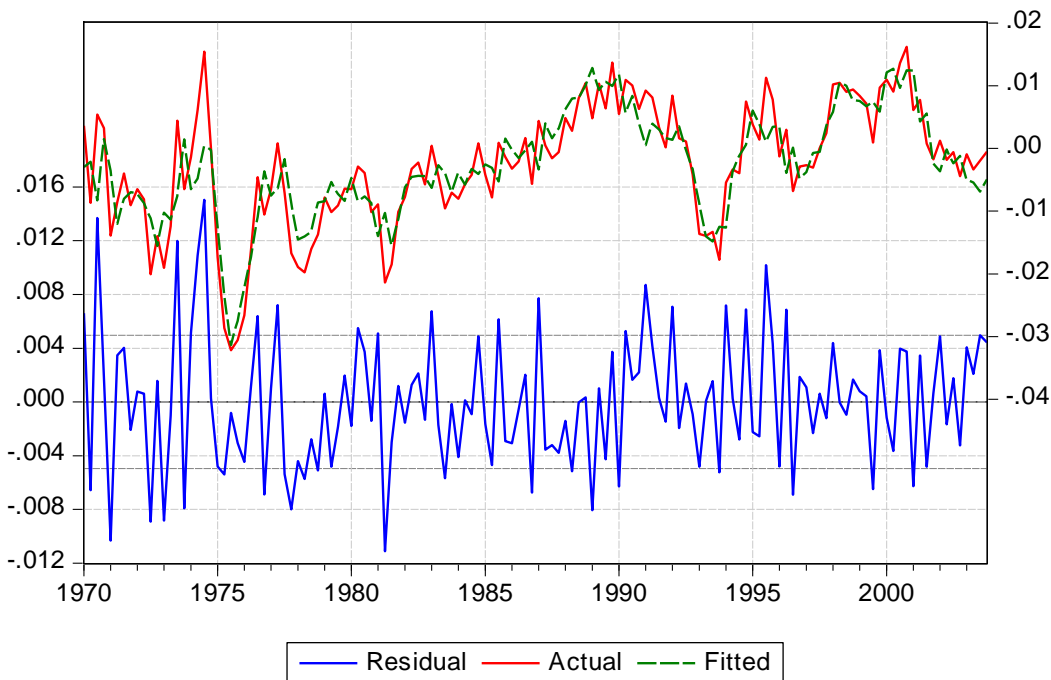
```
coef(10) c_ci
smpl 1962Q1 2004Q4
genr ci_ec=0
smpl 1963Q1 2003Q4
equation _eq_ci.ls(p) ci/q(-1)=c_ci(1)*@pchy(q)+c_ci(2)+c_ci(3)*ci(-1)/q(-2)+c_ci(4)*(t-
1985)*(t<=1985)+ci_ec
_eq_ci.resids(p)
```

¹⁴⁹ OK, we do a little cheating here, please forgive us.

close_eq_ci
 genr ci_ec=resid

Dependent Variable: CI/Q(-1)
 Method: Least Squares
 Date: 11/07/12 Time: 19:24
 Sample: 1970Q1 2003Q4
 Included observations: 136
 $CI/Q(-1) = C_CI(1) * @PCHY(Q) + C_CI(2) + C_CI(3) * CI(-1)/Q(-2) + C_CI(4) * (T - 1985) * (T \leq 1985)$

	Coefficient	Std. Error	t-Statistic	Prob.
C_CI(1)	0.195389	0.026646	7.332822	0.0000
C_CI(2)	-0.003401	0.000717	-4.744389	0.0000
C_CI(3)	0.584125	0.053030	11.01501	0.0000
C_CI(4)	0.000754	0.000119	6.332001	0.0000
R-squared	0.742572	Mean dependent var		-0.002669
Adjusted R-squared	0.736722	S.D. dependent var		0.009661
S.E. of regression	0.004957	Akaike info criterion		-7.747087
Sum squared resid	0.003243	Schwarz criterion		-7.661420
Log likelihood	530.8019	Hannan-Quinn criter.		-7.712274
F-statistic	126.9217	Durbin-Watson stat		2.260290
Prob(F-statistic)	0.000000			



10.2.2.4.5 Unemployment

This is a new equation compared to the small model. Actually it seemed clearer to us to model the work force (employment + unemployment). As it depends on employment, the quality of estimation will be exactly the same (only the R-squared will change). The coefficient of the work force POPAC will be higher by 1 compared to a formulation using unemployment UN.

Otherwise, the equation follows fully the framework defined above.

The short term sensitivities are reasonable, and the long term ones too, with a slightly lower value¹⁵⁰. We can observe a very slow return convergence to the long term relation (with a non-reliable intensity), and a stronger long term sensitivity of the work force to employment (thus a lower impact on unemployment). It is logical to expect that a permanent improvement of the labor market will attract more and more job seekers.

```
coef(10) c_popac

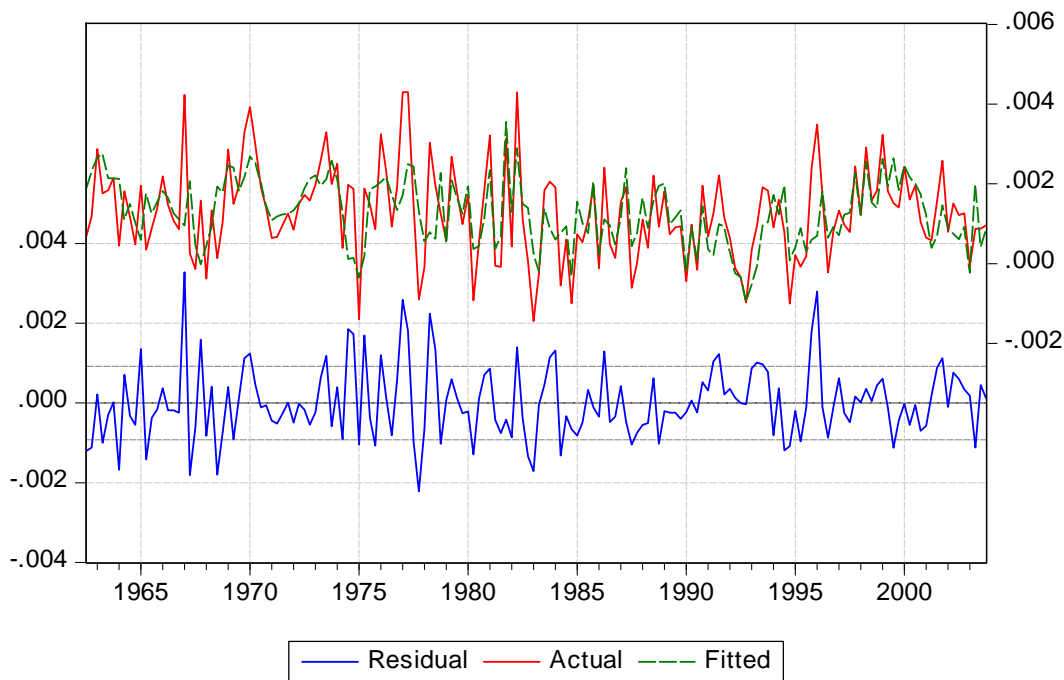
smp1 1962Q1 2002Q4
genr popac_ec=0
equation _eq_popac.ls(pd(popac)/pop65(-1)=c_popac(1)*d(lt)/pop65(-1)
+c_popac(2)*d(pop65)/pop65(-1)-c_popac(3)*(popac(-1)/pop65(-1)-c_popac(4)*lt(-1)/pop65(-1)-
c_popac(5))+c_popac(6)*(t-2003)+popac_ec
_eq_popac.resids(p)
genr popac_ec=resid
close _eq_popac
```

¹⁵⁰ For the potential work force, the T-stat is lower than 2, but the probability is higher than 5%.

Dependent Variable: D(POPAC)/POP65(-1)
 Method: Least Squares
 Date: 11/07/12 Time: 19:40
 Sample (adjusted): 1962Q3 2003Q4
 Included observations: 166 after adjustments
 Convergence achieved after 121 iterations

$$D(POPAC)/POP65(-1) = C_POPAC(1) * D(LT)/POP65(-1) + C_POPAC(2) * D(POP65)/POP65(-1) - C_POPAC(3) * (POPAC(-1)/POP65(-1)) - C_POPAC(4) * LT(-1)/POP65(-1) - C_POPAC(5) + [AR(1) = C_POPAC(6)]$$

	Coefficient	Std. Error	t-Statistic	Prob.
C_POPAC(1)	0.698524	0.073991	9.440704	0.0000
C_POPAC(2)	0.394457	0.203308	1.940193	0.0541
C_POPAC(3)	0.621335	0.074271	8.365737	0.0000
C_POPAC(4)	0.432522	0.073996	5.845199	0.0000
C_POPAC(5)	1.731148	83.13383	0.020824	0.9834
C_POPAC(6)	0.999840	0.010059	99.39601	0.0000
R-squared	0.362344	Mean dependent var		0.001278
Adjusted R-squared	0.342417	S.D. dependent var		0.001138
S.E. of regression	0.000923	Akaike info criterion		-11.10322
Sum squared resid	0.000136	Schwarz criterion		-10.99074
Log likelihood	927.5673	Hannan-Quinn criter.		-11.05756
Durbin-Watson stat	1.812878			
Inverted AR Roots	1.00			



10.2.2.4.6 The value added deflator

The formula will change according to the type of production function: the measure of the rate of use will be different, and in the Cobb Douglas case the cost will include amortization of capital.

In both cases the equation follows an error correction format: a long term relationship between the rate of use and the margins rate (or rather the share of the cost in value added), and a dynamic equation freeing the elasticity of the deflator to the cost.

Let us first present the complementary factors case.

We have constrained to the same value the short and long term sensitivities to the rate of use.

```

smpl 1962Q1 2004Q4
coef(10) c_pq
genr pq_ec=0
smpl 1963Q1 2003Q4
equation _eq_pq.ls(p) dlog(pq)=c_pq(1)*dlog(uwc)+c_pq(2)*dlog(ur)+c_pq(3)*log(pq(-1)/uwc(-1))-
c_pq(2)*c_pq(3)*log(ur(-1))+c_pq(5)+pq_ec
_eq_pq.resids(p)
close _eq_pq

```

Dependent Variable: DLOG(PQ)

Method: Least Squares

Date: 11/07/12 Time: 19:49

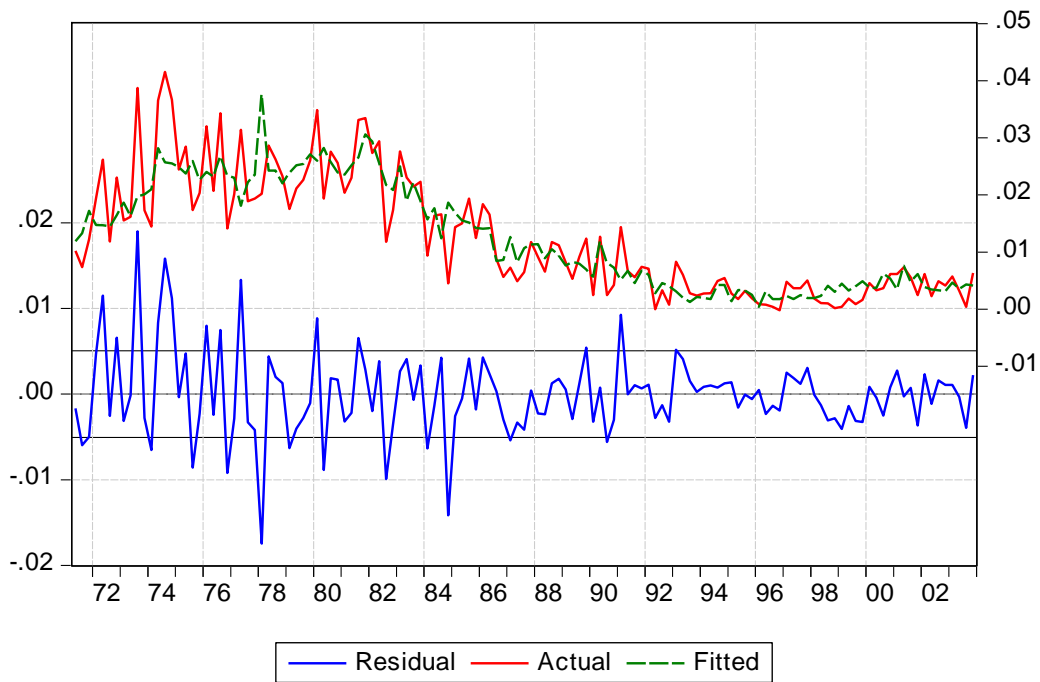
Sample (adjusted): 1971Q2 2003Q4

Included observations: 131 after adjustments

Convergence achieved after 1 iteration

DLOG(PQ)=C_PQ(1)*DLOG(UWC)+C_PQ(2)*DLOG(UR)+C_PQ(3)
 *LOG(PQ(-1)/UWC(-1))-C_PQ(2)*C_PQ(3)*LOG(UR(-1))+C_PQ(5)

	Coefficient	Std. Error	t-Statistic	Prob.
C_PQ(1)	0.415570	0.045536	9.126161	0.0000
C_PQ(2)	0.354049	0.086013	4.116222	0.0001
C_PQ(3)	-0.065965	0.007525	-8.765824	0.0000
C_PQ(5)	0.037465	0.003775	9.923875	0.0000
R-squared	0.773324	Mean dependent var		0.012717
Adjusted R-squared	0.767969	S.D. dependent var		0.010493
S.E. of regression	0.005054	Akaike info criterion		-7.707104
Sum squared resid	0.003244	Schwarz criterion		-7.619311
Log likelihood	508.8153	Hannan-Quinn criter.		-7.671430
F-statistic	144.4234	Durbin-Watson stat		2.015563
Prob(F-statistic)	0.000000			



In the Cobb-Douglas case, we have to consider the total cost: wage and capital.

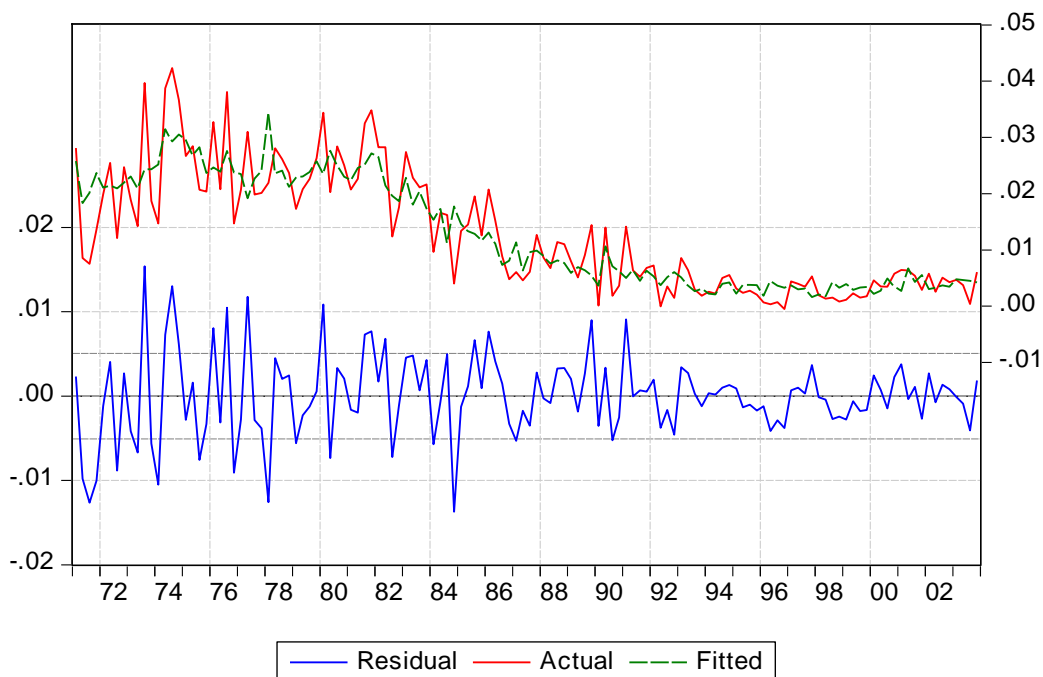
$$\text{genr cost} = (\text{wr} * \text{lf} * (1 + \text{r_scf}) + 0.05 * \text{pi} * \text{k}(-1)) / \text{q}$$

We distribute the cost of capital over 20 periods (5 years). Reducing this value would not change much the properties.

The coefficients are rather different from the previous case: stronger dynamic effect, slower correction.

Dependent Variable: DLOG(PQ)
 Method: Least Squares
 Date: 11/07/12 Time: 19:18
 Sample (adjusted): 1971Q1 2003Q4
 Included observations: 132 after adjustments
 Convergence achieved after 5 iterations
 $DLOG(PQ) = C_PQ(1) * DLOG(COST) + C_PQ(2) * DLOG(UR) + C_PQ(3) * LOG(PQ(-1)/COST(-1)) - C_PQ(2) * C_PQ(3) * LOG(UR(-1)) + C_PQ(5) + PQ_EC$

	Coefficient	Std. Error	t-Statistic	Prob.
C_PQ(1)	0.441903	0.077407	5.708801	0.0000
C_PQ(2)	0.295159	0.120499	2.449460	0.0157
C_PQ(3)	-0.052352	0.008027	-6.522219	0.0000
C_PQ(5)	0.022348	0.002811	7.949870	0.0000
R-squared	0.783973	Mean dependent var		0.012987
Adjusted R-squared	0.778910	S.D. dependent var		0.010789
S.E. of regression	0.005073	Akaike info criterion		-7.699873
Sum squared resid	0.003294	Schwarz criterion		-7.612515
Log likelihood	512.1916	Hannan-Quinn criter.		-7.664375
F-statistic	154.8399	Durbin-Watson stat		2.056374
Prob(F-statistic)	0.000000			



10.2.2.4.7 The wage rate

The wage rate estimation is the only occasion which allows us to use cointegration, and thus to conform to the accepted principles of econometrics.

The estimation process is exactly the same as for the small model.

We suppose that the wage cost is indexed in the long run: for 50% on the value added deflator (the firms want to stabilize the share of wages in value added) and for 50% on the consumption deflator (workers want to their purchasing power to follow the gains in productivity). The weighting could be adapted through a parameter.

So first we test that the stationarity of `_luwc`:

```

genr _luwc=log(uwc)-0.5*log(pq)-(1-0.5)*log(pcoh)

uroot(p) _luwc
    
```

Null Hypothesis: `_LUWC` has a unit root
 Exogenous: Constant
 Lag Length: 0 (Automatic - based on SIC, maxlag=12)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-0.209301	0.9332
Test critical values:		
1% level	-3.482453	
5% level	-2.884291	
10% level	-2.578981	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
 Dependent Variable: `D(_LUWC)`
 Method: Least Squares
 Date: 11/07/12 Time: 19:59
 Sample (adjusted): 1972Q2 2003Q4
 Included observations: 127 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
<code>_LUWC(-1)</code>	-0.001769	0.008454	-0.209301	0.8346
<code>C</code>	-0.001650	0.003746	-0.440503	0.6603

R-squared	0.000350	Mean dependent var	-0.000881
Adjusted R-squared	-0.007647	S.D. dependent var	0.008313
S.E. of regression	0.008345	Akaike info criterion	-6.718639
Sum squared resid	0.008705	Schwarz criterion	-6.673848
Log likelihood	428.6336	Hannan-Quinn criter.	-6.700441
F-statistic	0.043807	Durbin-Watson stat	1.855823
Prob(F-statistic)	0.834554		

Then that of the unemployment rate itself.

uroot(p) unr

Null Hypothesis: UNR has a unit root

Exogenous: Constant

Lag Length: 1 (Automatic - based on SIC, maxlag=13)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-1.390809	0.5858
Test critical values:		
1% level	-3.468980	
5% level	-2.878413	
10% level	-2.575844	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation

Dependent Variable: D(UNR)

Method: Least Squares

Date: 11/07/12 Time: 20:01

Sample (adjusted): 1962Q3 2004Q4

Included observations: 170 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
UNR(-1)	-0.004118	0.002961	-1.390809	0.1661
D(UNR(-1))	0.674893	0.056522	11.94040	0.0000
C	0.000450	0.000238	1.888933	0.0606
R-squared	0.469104	Mean dependent var		0.000495
Adjusted R-squared	0.462746	S.D. dependent var		0.001945
S.E. of regression	0.001426	Akaike info criterion		-10.25100
Sum squared resid	0.000339	Schwarz criterion		-10.19566
Log likelihood	874.3347	Hannan-Quinn criter.		-10.22854
F-statistic	73.78123	Durbin-Watson stat		1.881471
Prob(F-statistic)	0.000000			

Both fail.

Then we test the cointegration of the two elements, which works.

coint(b,p) _luwc unr

Date: 11/07/12 Time: 20:03
 Sample (adjusted): 1973Q2 2003Q4
 Included observations: 123 after adjustments
 Trend assumption: No deterministic trend (restricted constant)
 Series: _LUWC UNR
 Lags interval (in first differences): 1 to 4

Unrestricted Cointegration Rank Test (Trace)

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.**
None *	0.153423	27.57871	20.26184	0.0041
At most 1	0.056032	7.092588	9.164546	0.1216

Trace test indicates 1 cointegrating eqn(s) at the 0.05 level

* denotes rejection of the hypothesis at the 0.05 level

**MacKinnon-Haug-Michelis (1999) p-values

Unrestricted Cointegration Rank Test (Maximum Eigenvalue)

Hypothesized No. of CE(s)	Eigenvalue	Max-Eigen Statistic	0.05 Critical Value	Prob.**
None *	0.153423	20.48612	15.89210	0.0088
At most 1	0.056032	7.092588	9.164546	0.1216

Max-eigenvalue test indicates 1 cointegrating eqn(s) at the 0.05 level

* denotes rejection of the hypothesis at the 0.05 level

**MacKinnon-Haug-Michelis (1999) p-values

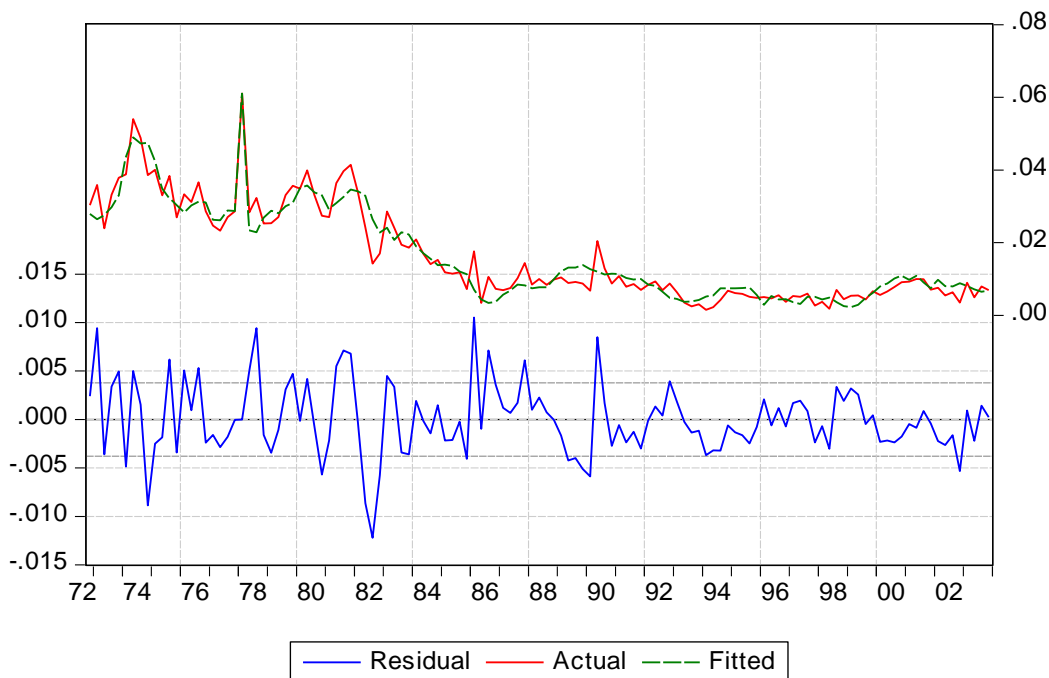
Unrestricted Cointegrating Coefficients (normalized by $b^*S11*b=l$):		
_LUWC	UNR	C
-13.09680	-61.22113	-0.735028
9.578167	-2.900417	5.183698
Unrestricted Adjustment Coefficients (alpha):		
D(_LUWC)	0.003119	-0.000294
D(UNR)	-1.40E-05	0.000290
1 Cointegrating Equation(s):		Log likelihood 1080.812
Normalized cointegrating coefficients (standard error in parentheses)		
_LUWC	UNR	C
1.000000	4.674509	0.056123
	(0.72130)	(0.06394)
Adjustment coefficients (standard error in parentheses)		
D(_LUWC)	-0.040853	
	(0.00912)	
D(UNR)	0.000183	
	(0.00150)	

The coefficient for UNR in the cointegrating equation is significant with the right sign. It remains to be seen if this equation gives good properties to the full model.

For the dynamic equation, we had to set the indexation as globally unitary (dynamic homogeneity) with a lag structure. All the elements are quite significant, but the error correction is quite slow (a result we found in other circumstances for the same equation).

Dependent Variable: DLOG(WR)
Method: Least Squares
Date: 11/07/12 Time: 20:13
Sample (adjusted): 1972Q4 2003Q4
Included observations: 125 after adjustments
DLOG(WR)=C_WR(1)*@MOVAV(DLOG(PCOH),4)+C_WR(2)*D(UNR)
+C_WR(3)*RES_WR(-1)+C_WR(7)*(T=1978)+C_WR(8)*T

	Coefficient	Std. Error	t-Statistic	Prob.
C_WR(1)	1.148306	0.036142	31.77180	0.0000
C_WR(2)	-0.694329	0.180518	-3.846317	0.0002
C_WR(3)	-0.038498	0.004146	-9.285473	0.0000
C_WR(7)	0.033340	0.003844	8.673487	0.0000
C_WR(8)	1.29E-06	2.97E-07	4.330384	0.0000
R-squared	0.920773	Mean dependent var		0.016535
Adjusted R-squared	0.918132	S.D. dependent var		0.013225
S.E. of regression	0.003784	Akaike info criterion		-8.276957
Sum squared resid	0.001718	Schwarz criterion		-8.163824
Log likelihood	522.3098	Hannan-Quinn criter.		-8.230997
Durbin-Watson stat	1.594350			



10.2.2.4.8 The trade deflators

For this equation we will not use an error correction format, rather a homogenous indexation. This means we shall not separate the short and long term sensitivities.

As usual for this type of equation:

- Exporters show a higher attention to their costs than to the price of their competitors. This is particular true for the exporters to France, as the role of French prices is not completely significant.

This will dampen the dynamics of the price-wage loop. The impact of trade on the price of demand can be explained in the following way.

- Imports are a share of global demand. They are bought at the import price. The higher its sensitivity to foreign costs, the higher the difference to the local production price, and the higher the reducing impact of imports on the global demand price.
- If local producers decided on their selling price on the local and foreign markets independently (a possible behavior that we did not consider), the sensitivity of the demand price to local costs would clearly be less than one.
- But in our framework, the production price is decided globally, and the lower sensitivity of the export price has to be balanced by a higher sensitivity of the price at which they sell on the local market. The higher the impact of local costs on the export price, the lower the necessary compensation.

One can see that in the transition from production to demand price, the higher the role of the production cost in the price set by the exporter, the higher the first (negative) effect and the lower the second (positive) one.¹⁵¹

In the extreme, if all exporters take only into account their costs, the import price will not be affected, and as the export price will change just as the global production price, no compensation will be needed. The damping effect will be maximal.

- The additional trend is negative and quite significant. It probably represents a structural but permanent shift in traded goods to the ones which present the lowest price.

To reach a steady state in the long run, these trends will have to be suppressed after a while. Here we did it immediately, but true forecasts should call for a gradual decrease.

```
coef(10) c_px
smpl 1962Q1 2004Q4
genr px_ec=0
smpl 1962Q1 2003Q4
equation _eq_px.ls(p) log(px)=c_px(4)*log(pp)+(1-c_px(4))*log(ppx*er)+c_px(6)+c_px(5)*(t-
2005)*(t<=2005)+[ar(1)=c_px(7)]+px_ec
_eq_px.resids(p)
close _eq_px
```

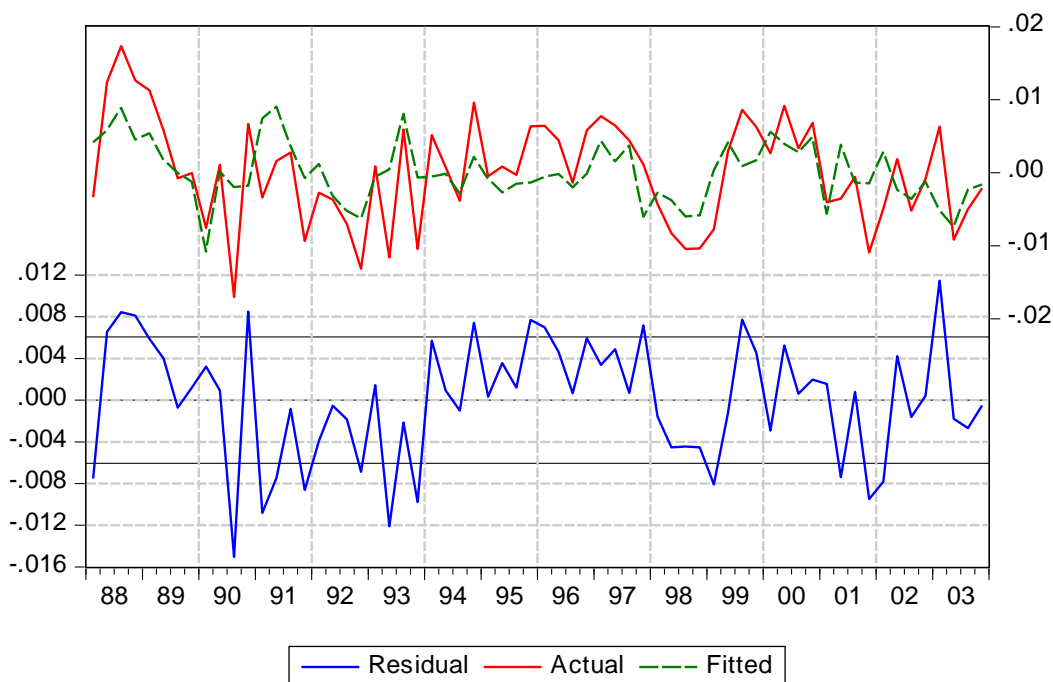
¹⁵¹ This could be formalized easily, but we hope the message is already clear.

Dependent Variable: LOG(PX)
 Method: Least Squares
 Date: 11/07/12 Time: 20:17
 Sample (adjusted): 1970Q2 2003Q4
 Included observations: 135 after adjustments
 Convergence achieved after 1 iteration
 $LOG(PX)=C_PX(4)*LOG(PP)+(1-C_PX(4))*LOG(PPX*ER)+C_PX(6)$
 $+C_PX(5)*(T-2005)*(T<=2005)+[AR(1)=C_PX(7)]$

	Coefficient	Std. Error	t-Statistic	Prob.
C_PX(4)	0.614657	0.046654	13.17491	0.0000
C_PX(6)	-0.172383	0.035436	-4.864641	0.0000
C_PX(5)	-0.022531	0.003025	-7.447879	0.0000
C_PX(7)	0.963293	0.018031	53.42319	0.0000

R-squared	0.999545	Mean dependent var	-0.287933
Adjusted R-squared	0.999535	S.D. dependent var	0.445430
S.E. of regression	0.009606	Akaike info criterion	-6.423668
Sum squared resid	0.012088	Schwarz criterion	-6.337586
Log likelihood	437.5976	Hannan-Quinn criter.	-6.388687
F-statistic	95996.91	Durbin-Watson stat	1.682038
Prob(F-statistic)	0.000000		

Inverted AR Roots	.96
-------------------	-----



coef(10) c_pm
 smpl 1962Q1 2004Q4
 genr pm_ec=0

```

smpl 1962Q1 2003Q4
equation _eq_pm.ls(p) log(pm)=c_pm(4)*log(pp)+(1-c_pm(4))*log(ppx*er)+c_pm(6)+c_pm(5)*(t-
2005)*(t<=2005)+[ar(1)=c_pm(7)]+pm_ec
_eq_pm.resids(p)
close _eq_pm

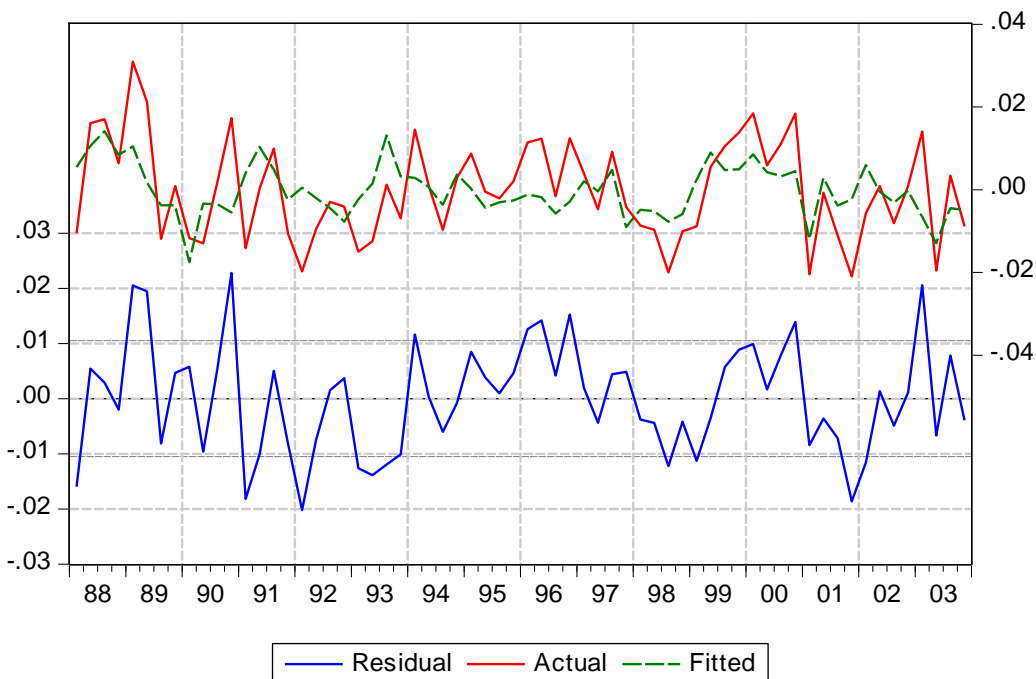
```

Dependent Variable: LOG(PM)
Method: Least Squares
Date: 11/07/12 Time: 20:23
Sample (adjusted): 1970Q2 2003Q4
Included observations: 135 after adjustments
Convergence achieved after 1 iteration
LOG(PM)=C_PM(4)*LOG(PP)+(1-C_PM(4))*LOG(PPX*ER)+C_PM(6)
+C_PM(5)*(T-2005)*(T<=2005)+[AR(1)=C_PM(7)]

	Coefficient	Std. Error	t-Statistic	Prob.
C_PM(4)	0.118328	0.107132	1.104509	0.2714
C_PM(6)	-0.270806	0.078177	-3.464005	0.0007
C_PM(5)	-0.034081	0.006262	-5.442496	0.0000
C_PM(7)	0.960815	0.018142	52.95997	0.0000

R-squared	0.997835	Mean dependent var	-0.264537
Adjusted R-squared	0.997785	S.D. dependent var	0.469326
S.E. of regression	0.022088	Akaike info criterion	-4.758370
Sum squared resid	0.063913	Schwarz criterion	-4.672288
Log likelihood	325.1900	Hannan-Quinn criter.	-4.723389
F-statistic	20122.07	Durbin-Watson stat	1.417189
Prob(F-statistic)	0.000000		

Inverted AR Roots	.96
-------------------	-----



10.2.2.4.9 Household consumption

Our equation follows as usual an error correction specification (estimated in one step!) following almost completely the framework presented earlier.

The change in consumption depends on:

- The change in real income (over the last year).
- The change in unemployment.
- Inflation (the « real holdings » effect).
- An error correction term.
- A negative time trend, representing the increase in the wealth of households and their saving potential (in particular their accession to housing ownership).

The only influence we could not evidence is that of the real short term interest rate.

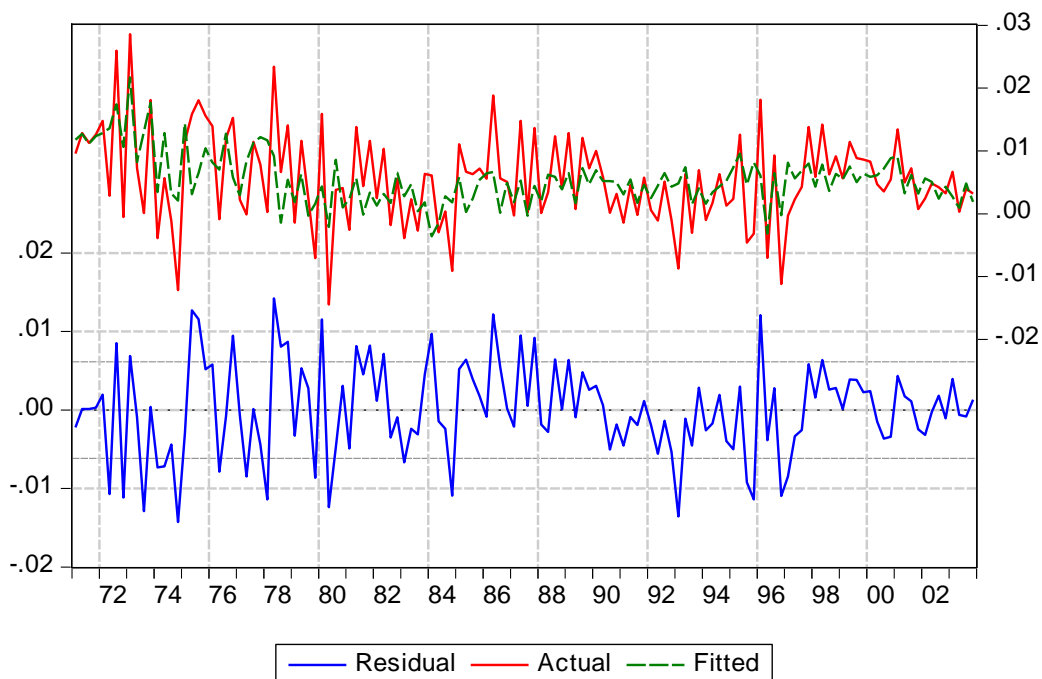
```
coef(10) c_coh
smpl 1962Q1 2004Q4
genr coh_ec=0
smpl 1962Q1 2003Q4
equation _eq_coh.ls(p) dlog(coh)=c_coh(1)*.25*log(hrldi/hrldi(-
4))+c_coh(2)*dlog(unr)+c_coh(3)*log(pcoh/pcoh(-4))+c_coh(5)*dlog(coh(-1))+c_coh(6)+c_coh(7)*log(coh(-
1)/hrldi(-1))+c_coh(8)*(t-2005)*(t<=2005)+coh_ec
_eq_coh.resids(p)
```

Dependent Variable: DLOG(COH)
 Method: Least Squares
 Date: 11/07/12 Time: 20:45
 Sample (adjusted): 1971Q1 2003Q4
 Included observations: 132 after adjustments

$$DLOG(COH) = C_COH(1) \cdot .25 \cdot LOG(HRDI/HRDI(-4)) + C_COH(2) \cdot DLOG(UNR) + C_COH(3) \cdot LOG(PCOH/PCOH(-4)) + C_COH(5) \cdot DLOG(COH(-1)) + C_COH(6) + C_COH(7) \cdot LOG(COH(-1)/HRDI(-1)) + C_COH(8) \cdot (T-2005) \cdot (T \leq 2005)$$

	Coefficient	Std. Error	t-Statistic	Prob.
C_COH(1)	0.445908	0.158975	2.804900	0.0058
C_COH(2)	-0.043832	0.019937	-2.198509	0.0298
C_COH(3)	-0.064211	0.027487	-2.336033	0.0211
C_COH(5)	-0.330953	0.079638	-4.155725	0.0001
C_COH(6)	-0.003384	0.002005	-1.688118	0.0939
C_COH(7)	-0.066734	0.023394	-2.852604	0.0051
C_COH(8)	-0.000269	0.000109	-2.461291	0.0152

R-squared	0.328381	Mean dependent var	0.005510
Adjusted R-squared	0.296144	S.D. dependent var	0.007319
S.E. of regression	0.006140	Akaike info criterion	-7.296357
Sum squared resid	0.004713	Schwarz criterion	-7.143482
Log likelihood	488.5596	Hannan-Quinn criter.	-7.234236
F-statistic	10.18626	Durbin-Watson stat	1.932245
Prob(F-statistic)	0.000000		



10.2.2.4.10 Exports

For exports we shall use again an error correction framework, estimated in one pass.

- The substitution effect appears through the average of the rate of use over the last two quarters (the dynamic and long term effects are not otherwise separated). Price competitiveness uses the same technique.
- A significant trend had to be introduced and will be stopped in the future. It is possible that the measure of world demand is biased.

```

gener x_ec=0
smpl 1962Q1 2003Q4
equation _eq_x.ls(p) dlog(x)=c_x(1)*dlog(wd)+c_x(2)*log(x(-1)/wd(-1))+c_x(3)*0.5*(log(ur)+log(ur(-1)))+c_x(4)*0.5*(log(compx(-1))+log(compx))+c_x(5)+c_x(6)*(t-2005)*(t<=2005)+x_ec
_eq_x.resids(p)
close _eq_x
gener x_ec=resid
    
```

Dependent Variable: DLOG(X)

Method: Least Squares

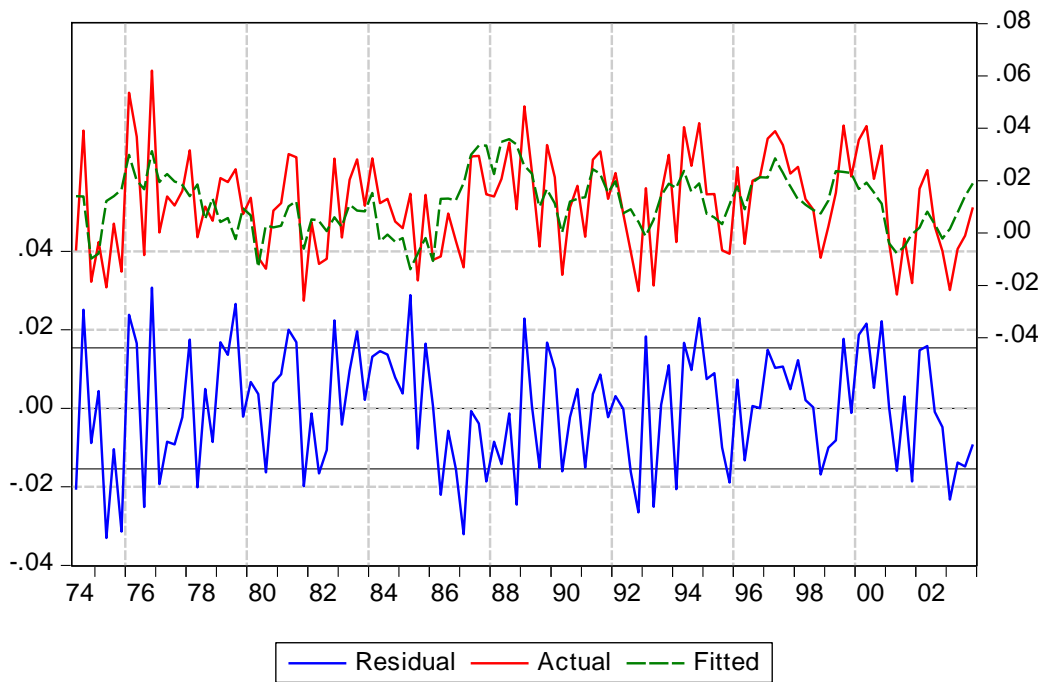
Date: 11/07/12 Time: 20:46

Sample (adjusted): 1974Q2 2003Q4

Included observations: 119 after adjustments

$$DLOG(X) = C_X(1) * DLOG(WD) + C_X(2) * LOG(X(-1)/WD(-1)) + C_X(3) * 0.5 * (LOG(UR) + LOG(UR(-1))) + C_X(4) * 0.5 * (LOG(COMPX(-1)) + LOG(COMPX)) + C_X(5) + C_X(6) * (T-2005) * (T \leq 2005)$$

	Coefficient	Std. Error	t-Statistic	Prob.
C_X(1)	0.385320	0.083551	4.611796	0.0000
C_X(2)	-0.230013	0.050940	-4.515338	0.0000
C_X(3)	-0.232284	0.087825	-2.644856	0.0093
C_X(4)	-0.183650	0.052253	-3.514653	0.0006
C_X(5)	-0.034008	0.016122	-2.109421	0.0371
C_X(6)	-0.007193	0.001669	-4.309415	0.0000
R-squared	0.340963	Mean dependent var		0.011474
Adjusted R-squared	0.311802	S.D. dependent var		0.018593
S.E. of regression	0.015424	Akaike info criterion		-5.456639
Sum squared resid	0.026884	Schwarz criterion		-5.316515
Log likelihood	330.6700	Hannan-Quinn criter.		-5.399739
F-statistic	11.69246	Durbin-Watson stat		2.024301
Prob(F-statistic)	0.000000			

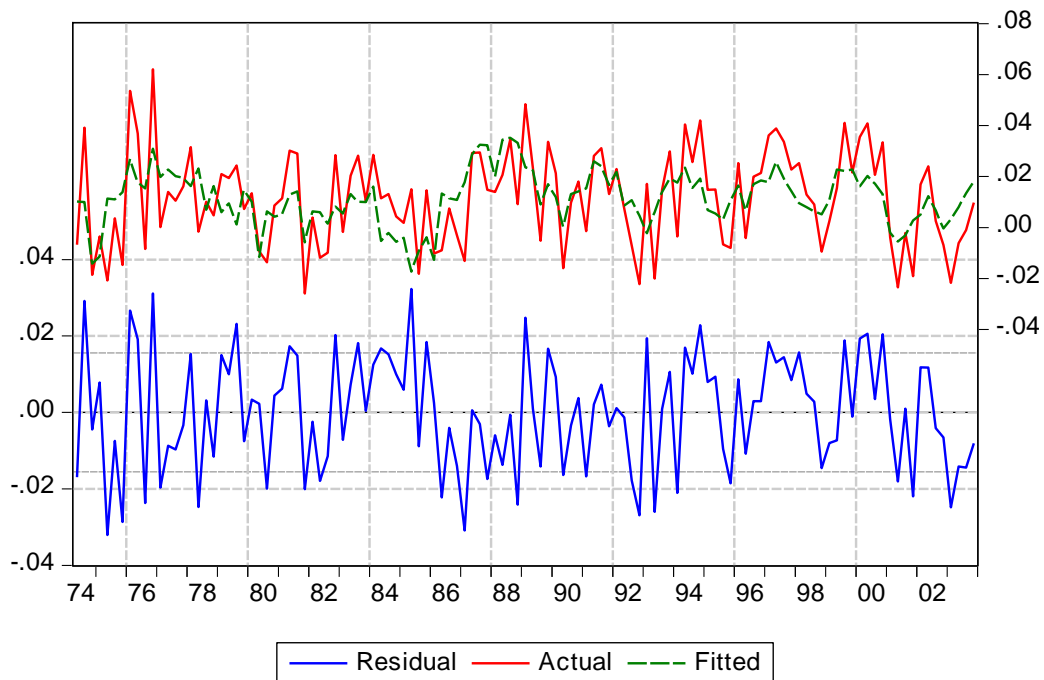


The Cobb-Douglas case gives similar results with a different UR:

Dependent Variable: DLOG(X)
 Method: Least Squares
 Date: 11/07/12 Time: 19:18
 Sample (adjusted): 1974Q2 2003Q4
 Included observations: 119 after adjustments

$$DLOG(X) = C_X(1) * DLOG(WD) + C_X(2) * LOG(X(-1)) / WD(-1) + C_X(3) * 0.5 * (LOG(UR) + LOG(UR(-1))) + C_X(4) * 0.5 * (LOG(COMPX(-1)) + LOG(COMPX)) + C_X(5) + C_X(6) * (T - 2005) * (T \leq 2005) + X_EC$$

	Coefficient	Std. Error	t-Statistic	Prob.
C_X(1)	0.416462	0.084406	4.934023	0.0000
C_X(2)	-0.188997	0.048520	-3.895210	0.0002
C_X(3)	-0.221336	0.136816	-1.617768	0.1085
C_X(4)	-0.199217	0.065871	-3.024365	0.0031
C_X(5)	-0.046719	0.023643	-1.976001	0.0506
C_X(6)	-0.007443	0.002100	-3.544910	0.0006
R-squared	0.316007	Mean dependent var		0.011474
Adjusted R-squared	0.285742	S.D. dependent var		0.018593
S.E. of regression	0.015714	Akaike info criterion		-5.419471
Sum squared resid	0.027902	Schwarz criterion		-5.279347
Log likelihood	328.4585	Hannan-Quinn criter.		-5.362571
F-statistic	10.44128	Durbin-Watson stat		2.071681
Prob(F-statistic)	0.000000			



10.2.2.4.11 Imports

For imports we use the same framework, but:

- We had to set the dynamic demand coefficient to unity.
- The rate of use is not lagged.
- Price competitiveness is measured over 6 periods.

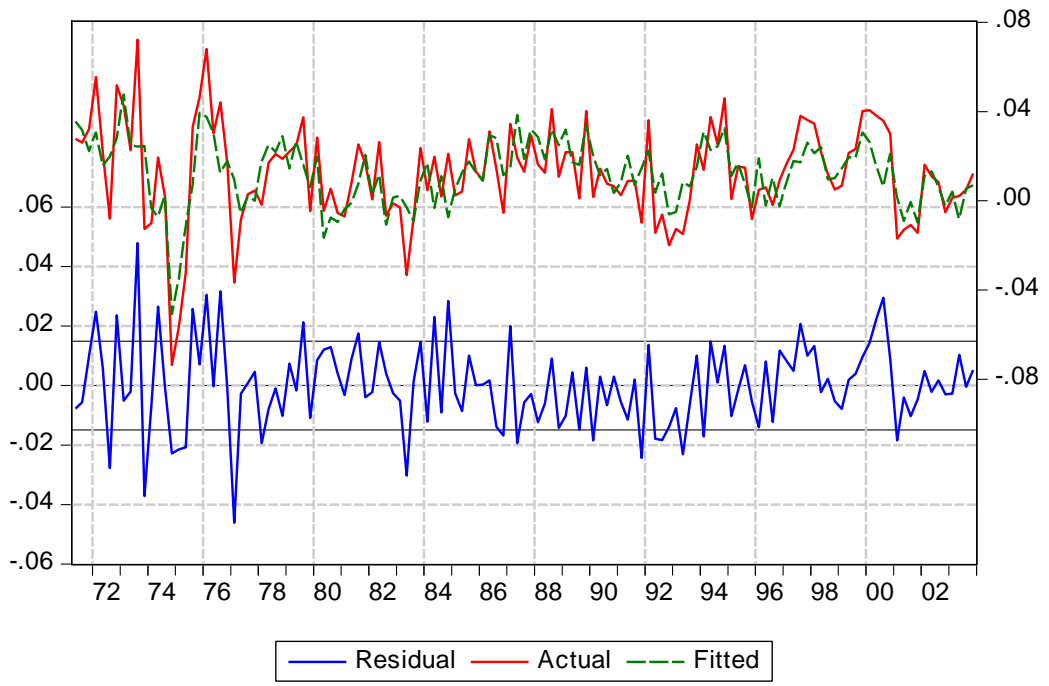
```
coef(10) c_m
smpl 1962Q1 2004Q4
genr m_ec=0
smpl 1962Q1 2003Q4
equation _eq_m.ls(p)
dlog(m)=dlog(fd+tc*q)+c_m(2)*log(ur)+c_m(3)*log(@movav(compm,6))+c_m(4)+c_m(5)*(t-
2005)*(t<=2005)+[ar(1)=c_m(6)]+c_m(7)*log(m(-1)/(fd(-1)+tc(-1)*q(-1)))+m_ec
_eq_m.resids(p)
close _eq_m
genr resid_m=resid
```


Dependent Variable: DLOG(M)
Method: Least Squares
Date: 11/07/12 Time: 20:49
Sample (adjusted): 1971Q2 2003Q4
Included observations: 131 after adjustments
Convergence achieved after 1 iteration
DLOG(M)=DLOG(FD+TC*Q)+C_M(2)*LOG(UR)+C_M(3)
LOG(@MOVAV(COMP6,6))+C_M(4)+C_M(5)(T-2005)*(T<=2005)
+[AR(1)=C_M(6)]+C_M(7)*LOG(M(-1)/(FD(-1)+TC(-1)*Q(-1)))

	Coefficient	Std. Error	t-Statistic	Prob.
C_M(2)	0.854879	0.259322	3.296598	0.0013
C_M(3)	-0.219006	0.101418	-2.159443	0.0327
C_M(4)	-1.364308	0.175134	-7.790062	0.0000
C_M(5)	0.017528	0.002816	6.225037	0.0000
C_M(7)	-0.782949	0.097880	-7.999036	0.0000
C_M(6)	0.889459	0.052156	17.05367	0.0000

R-squared	0.545765	Mean dependent var	0.012263
Adjusted R-squared	0.527596	S.D. dependent var	0.021698
S.E. of regression	0.014914	Akaike info criterion	-5.528350
Sum squared resid	0.027802	Schwarz criterion	-5.396661
Log likelihood	368.1069	Hannan-Quinn criter.	-5.474839
Durbin-Watson stat	2.113970		

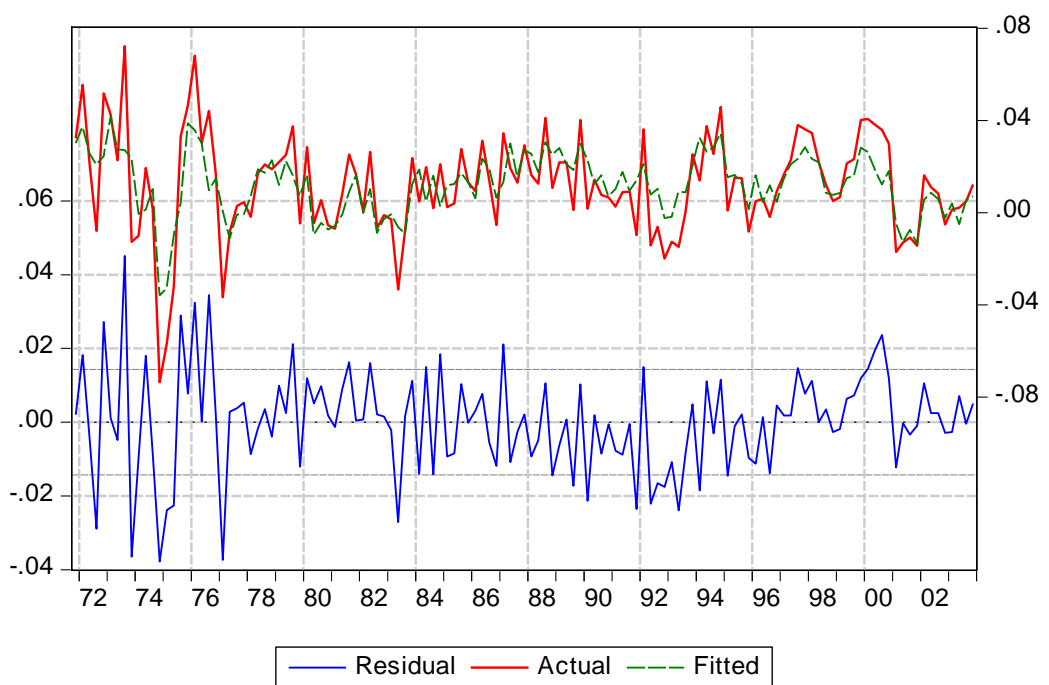
Inverted AR Roots	.89
-------------------	-----



The Cobb-Douglas case gives different, and less significant results:

Dependent Variable: DLOG(M)
 Method: Least Squares
 Date: 11/07/12 Time: 19:18
 Sample (adjusted): 1971Q4 2003Q4
 Included observations: 129 after adjustments
 Convergence achieved after 11 iterations
 $DLOG(M) = DLOG(FD + 0.5 * X) + C_M(2) * LOG(UR) + C_M(3) * LOG(@MOVAV(COMPM,6)) + C_M(4) + C_M(5) * (T - 2005) * (T \leq 2005) + [AR(1) = C_M(6)] + C_M(7) * LOG(M(-1) / (FD(-1) + 0.5 * X(-1))) + M_EC$

	Coefficient	Std. Error	t-Statistic	Prob.
C_M(2)	0.109425	0.154100	0.710092	0.4790
C_M(3)	-0.056667	0.031988	-1.771487	0.0790
C_M(4)	-0.250145	0.093434	-2.677242	0.0084
C_M(5)	0.004009	0.001571	2.551585	0.0119
C_M(7)	-0.201193	0.071715	-2.805436	0.0058
C_M(6)	0.390800	0.121756	3.209690	0.0017
R-squared	0.575925	Mean dependent var		0.012036
Adjusted R-squared	0.558686	S.D. dependent var		0.021789
S.E. of regression	0.014475	Akaike info criterion		-5.587442
Sum squared resid	0.025770	Schwarz criterion		-5.454428
Log likelihood	366.3900	Hannan-Quinn criter.		-5.533396
Durbin-Watson stat	2.089100			
Inverted AR Roots	.39			



10.2.2.5 Checking again model – data consistency

Now that equations have been estimated, we should check again data-equations consistency, this time on all equations.

The identities should still hold true, and adding the estimation residual at the end of the behavioral equations should make them hold true, too.

So this process could look unnecessary. This will generally be true if all estimations have been repeated immediately before the check, which means that they have been grouped, along with the definition of identities, in the same program. This is actually the best technique, as it allows producing the full model through a single program, and not through a sequence (which needs some organization and can lead to errors).

But if the estimates somewhat dated, and if one is not sure that the data have not been modified since, errors might have appeared.

And in any case, this check is almost free (the control that all the residuals are negligible can be done with a program, with a single answer) so there is no reason not to do it.

Of course, our example gives acceptable results¹⁵².

```
smpl 1995Q2 2004Q4
```

```
' Now we compute the residuals
' for all the endogenous
'
' The process is the same
'
' The dc_{%1} are the absolute differences
' The pc_{%1} the relative ones
' To avoid dividing by zero, we use trick 1
' We add to the variable a boolean one
' Testing if the variable is zero
' If it is true we get 0 divided by 1 = 0
' If not the computation is not affected
'
' This means we have just modified the outcome of divisions.
' Dividing zero by zero now gives zero
```

```
for !i=1 to _g_vendo.@count
%1=_g_viden.@seriesname(!i)
genr dc_{%1}={%1}-{%1}_c
genr pc_{%1}=100*dc_{%1}/({%1}+({%1}=0))
next
```

10.2.2.6 Solving the model on the future

¹⁵² But achieving this result took some time...

We are coming now to the crucial part of model testing, observing how it performs on the field where it will be used: the future.

Globally, the technique is the same as explained earlier. The differences are minimal.

We start with the usual statements: changing the directory, creating a special expanded file called "proj_1".

```
cd "c:\Program Files\reviews5\book\fra_cf"

' we set the workfile to proj_1, quarterly from 1962 to 2100
'
close proj_1.wf1
close mod_1.wf1
open mod_1.wf1
save proj_1
pagestruct(end=2100Q4) *
```

Now we are using deflators. For the long term assumptions, we need to include their growth rate:

```
' we define long-term growth rates

scalar txq=exp(0.005)-1
scalar txn=exp(0.002)-1
scalar txp=exp(0.006)-1
```

Obviously, the set of exogenous variables, and of assumptions to define, is larger.

This program simulates the model over the future

```
' we expand series for projection: EXOGENOUS variables
```

```
' constant expansion
```

```
smp1 2005Q1 2100Q4
for %1 COH_EC ERX FDXR K_EC CI_EC IR_ER IRL_ER IRMX IRSR IRST IRSX IRX LF_EC M_EC NIF_ER
NIG_ER PK PM_EC POPAC_EC PQ_EC PX_EC R_EXPG R_IGT R_IFP R_OIT R_PCOH R_PCOG R_PI
R_PIG R_REVG R_REVQ R_SCF R_SCG R_SCW R_SUBS R_TAR R_TARX R_VAT RDEP TC URD
WR_EC X_EC IRM_ER relax_q relax_pfd fcapf_er prof_er
genr {%1}={%1}(-1)
next
```

' expansion with GDP growth rate

```
for %1 COG IG WD R_REVX HIH
gener {%1}={%1}(-1)*(1+txq)
next
```

' expansion with population growth rate

```
for %1 LG POP65 POP
gener {%1}={%1}(-1)*(1+txn)
next
```

' expansion with deflators growth rate

```
for %1 PPX
gener {%1}={%1}(-1)*(1+txp)
next
```

' expansion with special growth rate

```
for %1 SOCBR
gener {%1}={%1}(-1)*(1+txq)/(1+txn)
next
```

```
gener t=t(-1)+0.25
```

It will be also useful (but not necessary) to create values for the endogenous:

```
smpl 2005Q1 2100Q4
```

' we expand series for projection: we initialize ENDOGENOUS variables

' constant expansion

```
for %1 COMPM COMPX ER FCAPGP IR IRL IRM IRS RCVAL RCVOL RES_WR RMARG RPROB RPROF
TTRAD UNR UR
gener {%1}={%1}(-1)
next
```

' expansion with long-term GDP growth rate

```
for %1 CAP COH FD HRDI I CI K M Q X GDPM IC
genr {%1}={%1}(-1)*(1+txq)
next
```

' expansion with population growth rate

```
for %1 LFD LF LT POPAC UN popt
genr {%1}={%1}(-1)*(1+txn)
next
```

' expansion with deflators growth rate

```
for %1 PCOH PFD PFDXT PI PIG PM PMT PP PQ PX UWC
genr {%1}={%1}(-1)*(1+txp)
next
next
```

' expansion with variables at current prices growth rate

```
for %1 CGV EXPG FCAPF FCAPG FCAPX FDGV HDI HI ICT IFP IGV MARG MVAL NIF NIG NIX NIXL NIXX
OIT PROF GDPMVAL QVAL REVG REVQ REVX SCF SCG SCW SOCB SUBS TAR TRB VAT W WF WG
XVAL
genr {%1}={%1}(-1)*(1+txp)*(1+txq)
next
```

' expansion with labour productivity growth rate

```
for %1 PL PLT
genr {%1}={%1}(-1)*(1+txq)/(1+txn)
next
```

' expansion with labor productivity at current prices growth rate

```
for %1 WR
genr {%1}={%1}(-1)*(1+txp)*(1+txq)/(1+txn)
```

But the essential issue is the guarantee that a stationary path will be obtained.

10.2.2.6.1 General elements

The same elements as above apply to this case.

But the higher complexity of the problem makes the methods proposed earlier more efficient if not necessary:

- Making shocks on the main assumptions and observing the consequences.
- Excluding some variables, or changing (increasing, decreasing or suppressing) local explanations.

10.2.2.6.2 Ensuring convergence in the long run: further elements

Let us return to the constraints the model must follow, and see how the above principles apply to a more complex case.

- First, care must be taken to avoid any remaining trend.

As we have said earlier, all elasticities should be unitary in the estimated equations, except of course for elements without dimension (ratios). The easiest method is obviously to use an error correction framework: the constraints can be set in the cointegrating equations (estimating them in one pass does not change this property, just the validity of the method....). They can be released in the VAR (the dynamic equation) at no cost to convergence, in principle.

One will observe that almost all our estimated equations, even if cointegration was applied only once, contain a long term expression linking ratios without dimension. The only exceptions are:

- The trade prices where no difference is made between short and long term, but homogeneity is obtained by constraints on coefficients.
- For investment:
 - In the complementary factors case, most elements represent derivatives (with no dimension). The only exceptions are the rate of use and the profits rate, both ratios which should stabilize in the long run. However, the rate of use is no longer fixed as in the simpler case.
 - In the Cobb-Douglas case, the capital – output ratio is linked to a ratio of prices (labor to capital).

The only real problem lies with employment, which has a specific dimension.

- We have already treated the complementary factors case. The trend in labor productivity is estimated. In the long run, both trends in employment (populations) and value added (a variable at constant prices) are fixed, and so is the trend in labor productivity, with a value which must replace the estimated one¹⁵³.
- For the Cobb-Douglas case, three elements are taken into account to define productive capacity:
 - * The trend in global factor productivity, at present estimated: a
 - * The contribution of labor: $\alpha \cdot txn$
 - * The contribution of capital: $(1 - \alpha) \cdot txq$

Basically, the total growth rate of capacity should be the same as all variables at constant prices, which gives:

$$txq = a + \alpha \cdot txn + (1 - \alpha) \cdot txq$$

$$a = \alpha \cdot (txq - txn)$$

This formula is easy to interpret: capital units will have a constant productivity (remember that the value at constant prices includes the increase in quality). Labor units do not, so to achieve their share of the increase in capacity their productivity must increase as the difference between growth of quantities and growth of labor.

In this way the total growth of capacity will meet the constraint.

To present the issue in a clearer way, it might be better to consider, not employment, but employment “efficient units” in which each worker is value at his potential contribution to production, just like capital. The value of these units at constant prices will grow as production, and there will be no need for an additional productivity trend.

Once these conditions are set, the model should have a long term solution, to which the formulations (in particular error-correction behaviors) should make the model converge.

10.2.2.6.3 Making the model converge in the short run

The problems are the same as in the simple case, but

- The danger is higher as more elements and mechanisms are concerned. The probability of spurious cycles grows.
- Deflators are generally more volatile than quantities (being less “real”).
- The risk of making an error on assumptions is higher, as the process of definition is more complex: trying to produce an accurate forecast on a partially known future will lead us to define some of them explicitly.

¹⁵³ Or one of trends must be set to agree with the estimated value.

- The trends which have to be blocked are more numerous, and they might appear in more complex equations.
- Finally, it is possible that the management of residuals in the first forecast periods brings high variations, and in particular strong cycles.

10.2.2.6.4 Making the model converge in the medium run

Again, the main danger lies in the cycles, with a higher probability.

But the fact that the model has converged earlier will provide us with important information, and the possibility to conduct additional tests, as we have already seen.

10.2.2.6.5 Specific forecast elements

Now that we have made the model converge over the whole period, we can test its properties through responses to shocks. Then we can move to actual forecasts.

10.2.2.7 Producing a forecast

The technique is the same as usual. Here we use the Gauss-Seidel algorithm.

We check that the $\Delta \log s$ converge to the theoretical values. To avoid taking the logarithm of negative values (this happens when the variable changes signs from one period to the other¹⁵⁴) we eliminate the associated computations.

' we solve the model in the future (no shock)

```
smpi 2005Q1 2100Q4
```

' The letter for the base simulation is b

```
_mod_1.append assign @all _b
```

' We use the Gauss-Seidel method

```
_mod_1.solveopt(n=t m=1001,c=1e-6,o=g,d=d)
_mod_1.scenario "scenario 1"
```

' We make sure that the full model is solved

¹⁵⁴ With growth rates, we would get a value, but it would have no meaning.

```
_mod_1.exclude
smpl 2005Q1 2100Q4
solve(n=t m=1001 c=1e-6 o=g d=d) _mod_1
```

- ' We compute the change in logarithms
- ' Only when the sign does not change

```
for !i=1 to _g_vendo.@count
%st1=_g_vendo.@seriesname(!i)
smpl 2005Q1 2100Q4
series _z=({%st1}_1+({%st1}_1=0))/({%st1}_1(-1)+({%st1}_1(-1)=0))
series {%st1}_tc=na
smpl 2005Q1 2100Q4 if _z>0
series {%st1}_tc=log(_z)

smpl 2005Q1 2100Q4
series _z=({%st1}+({%st1}=0))/({%st1}(-1)+({%st1}(-1)=0))
series {%st1}_tc0=na
smpl 2005Q1 2100Q4 if _z>0
series {%st1}_tc0=log(_z)
series {%st1}_DTC=100*({%st1}_tc-{{%st1}_tc0})
next
```

10.2.2.8 Producing shocks on assumptions

Compared to the previous example, the process will be a little more complex. In particular, we shall consider seven shocks.

Comments are imbedded in the program.

- ' Now we shall produce a set of shocks (in the present case 7)

```
smpl 2000 2500
```

- ' The group called shocks_v will contain the list of 7 shocked variables

```
group shocks_v ig erx r_vat r_tar r_tarx wd m_ec
```

- ' The group called shocks_l will contain 7 letters associated to each shock
- ' but the letters must be known as series

- ' as in EViews groups can only contain series
- ' We create the artificial series
- ' but only for the non-existing ones of course

```
for %z g r t f y w n j
if @isobject(%z) =0 then
gener {%z}=na
endif
next
```

- ' We create the group

```
group shocks_l g r t f y w n j
```

- ' Now we compute the additional change for each assumption,
- ' using the name of the variable
- ' The shock will only start in 2006
- ' leaving one unshocked period
- ' to check that the difference comes only from the shock
- ' The name of the assumption will combine the shocked variable with the letter
- ' 1 - Shock g: +1 GDP point on Government demand

```
gener ig_g=ig+.01*gdpm_b*(t>=2006)
```

- ' 2 - Shock r: 1% devaluation of the Euro

```
gener erx_r=erx*(1+.01*(t>=2006))
```

- ' 3 - Shock t: -1 point on the VAT rate

```
gener r_vat_t=r_vat-.01*(t>=2006)
```

- ' 4 - Shock f: -1 point on the local tariffs rate

```
gener r_tar_f=r_tar-.01*(t>=2006)
```

' 5 - Shock y: -1 point on the foreign tariffs rate

```
genr r_tarx_y=r_tarx-.01*(t>=2006)
```

' 6 - Shock w: +1% on World demand addressed to France

```
genr wd_w=wd*(1+.01*(t>=2006))
```

' 7 - Shock n: +1% on ex-ante imports

```
genr m_ec_n=m_ec+.01*(t>=2006)
```

' The loop on the shocks

```
for !j=1 to shocks_v.@count
```

```
  smpl 2005Q1 2100Q4
```

' We get the name of the shocked variable and the associated letter

```
  %2=shocks_v.@seriesname(!j)
```

```
  %3=shocks_l.@seriesname(!j)
```

' We set the solution suffix using the letter

' We control no endogenous is excluded from the simulation

' We override the variable associated to the current shock

' We solve the model

```
  _mod_1.scenario "scenario 1"
```

```
  _mod_1.append assign @all _{%3}
```

```
  _mod_1.exclude
```

```
  _mod_1.override {%2}
```

```
  _mod_1.solve(n=t m=1002 c=1e-6 o=g d=d)
```

' We compute the difference to the base simulation, in absolute and relative terms

```
for !i=1 to _g_vendo.@count
```

```

%st1=_g_vendo.@seriesname(!i)
series d{%3}_{%st1}={%st1}_{%3}-{%st1}_b
series p{%3}_{%st1}=100* d{%3}_{%st1}/({%st1}_b+ ({%st1}_b=0))
series dv_{%st1}=d{%3}_{%st1}
series pv_{%st1}=p{%3}_{%st1}
next

```

- ' We create groups for variations for a specific list of important variables
- ' the number is restricted to make a table legible

```

group g_v{%3} P{%3}_PFD P{%3}_PM P{%3}_PX P{%3}_PQ P{%3}_WR P{%3}_FD P{%3}_gdpm
P{%3}_X P{%3}_M P{%3}_I P{%3}_COH P{%3}_UR P{%3}_CAP P{%3}_LF P{%3}_K P{%3}_UNR
D{%3}_FCAPGP P{%3}_RCVAL P{%3}_RCVOL P{%3}_TTRAD

```

- ' or very important ones
- ' displayed in a graph
- ' the number is further restricted to make the graph legible

```

group g_w{%3} P{%3}_FD P{%3}_gdpm P{%3}_X P{%3}_M P{%3}_PQ

```

- ' We create a group for variations for all variables

```

group g_v2 pV_* dV_*

```

- ' We store all shocks in Excels files (2025 and 2100)
- ' using the letter associated to the shock

```

smpl 2005Q1 2025Q4
write(t=xls) v2_{%3}.xls g_v2
smpl 2005Q1 2100Q4
write(t=xls) w2_{%3}.xls g_v2

next

```

10.2.2.9 Applying the programs

We shall now present the results obtained from the previous programs.

10.2.2.9.1 Producing the model framework

Basically, the program creates an item called `mod_1` in the workfile. Of course, this model cannot be solved at this time. However, accessing this item gives alternately

- the list of variables,
- the list of equations,
- the original source code,
- the block structure.

The list of equations and the code repeat the model creating statements, and are of little interest. The list of variables, as stated earlier, helps locate problems such as logically endogenous variables with no equation (they appear as exogenous) or typing errors in variable names (they create also an exogenous item).

The last element is the most interesting. Let us show the result for our model. First the complementary factors version.

```
Number of equations: 85
Number of independent blocks: 5
Number of simultaneous blocks: 2
Number of recursive blocks: 3

Block 1: 2 Recursive Equations

    ict(40)          ifp(51)

Block 2: 55 Simultaneous Equations

    gdpm(1)          q(2)          ur(3)          i(4)          ci(6)
    lf(7)            pl(8)          lt(9)          cap(10)
    popac(11)        un(12)         unr(13)        uwc(14)
    pq(15)           pp(16)         pfd(17)        pfdxt(18)
    pcoh(19)        pi(21)         wr(23)         px(24)
    pm(25)           er(26)         irs(27)        irl(28)
    ir(29)           irm(30)        relc(31)       wf(32)
    wg(33)           w(34)          socb(35)       revq(36)
    revx(37)         scw(38)        hi(39)         hdi(41)
    hr(42)           coh(43)        qval(44)       vat(45)
    gdpmval(46)     marg(49)       nif(52)        prof(53)
    rprof(54)       fcapf(56)     pmt(57)        compm(58)
    fd(59)           m(60)          compx(61)      x(62)
    mval(63)        xval(64)

Block 3: 23 Recursive Equations

    k(5)             pcog(20)       pig(22)        pgdpm(47)
    subs(48)         rmarg(50)      rprob(55)     rcval(65)
    rcvol(66)        ttrad(67)     trb(68)       nixl(69)
```

nixx(70)	nix(71)	fcapx(72)	scf(73)
oit(74)	tar(75)	scg(76)	revg(77)
igv(78)	cogv(79)	fdgv(80)	

Block 4: 3 Simultaneous Equations

nig(81)	expg(82)	fcapg(83)
---------	----------	-----------

Block 5: 2 Recursive Equations

fcapgp(84)	gdpval(85)
------------	------------

What the above shows is that the 85 equations in our model can be decomposed into a sequence of five « blocks ». Three of them are defined as recursive, two as simultaneous.

Let us define these notions, which are associated to the model solving process.

- Recursive means that the set can be arranged in such a way that no variable appears in any equation before the one in which it is computed. Obviously this means that computations will give to each variable its exact value, and after each variable has been computed once no further computation is needed.
- Simultaneous means that each variable in the set depends on a variable computed later, either directly or through a sequence of (possibly recursive) influences. Therefore this variable will not take its exact value immediately, except if the variable(s) responsible for the non-recursiveness was given its exact value in the beginning.

Starting from any order, EViews is able to build an initial recursive set, as variables which depend only on exogenous, lagged, or variables which have been included earlier in the set. We will call this set a « prologue ».

It is also able to build a final recursive set, as variables which influence no variable in the rest of the set, or only variables which have already been included in the set. We will call this set an « epilogue ».

Once this is done, the rest of the model could be considered simultaneous. However, this set can eventually be separated again in two consecutive ones, if no variable in the second set influences any variable in the first. The first set can have an epilogue, the second one a prologue, which can be joined into an intermediate recursive block. And the process can be applied again to the two simultaneous blocks, until partition is no longer possible.

In our case, we observe:

- A small initial recursive block with ICT (Income tax) and IFP (Tax on firms' profits). They obviously depend on an exogenous tax rate and revenue from the previous period.
- A large (55 equations) block containing two interconnected loops: the Keynesian supply – demand – supply equilibrating process, and the wage - price loop. Taking any couple of variables in the set, one can link the first to the second, and the second to the first.

- A large (22 equations) recursive block, containing mostly descriptive elements (the terms of trade), items in the State budget (amount of tariffs) and variables influencing only the future (the capital stock)
- A small non-recursive 3 equations block, linking Government balance (revenue – expenditures) interests paid (depending on the balance) and expenditures (including interests). These elements depend on the rest of the model, but have no influence, except of course for the balance in GDP points, which appears in the last (simultaneous) block, along with global GDP at current prices which is purely descriptive.
- This defines a final two-equation block.

One could question the absence of a 3 equations block for firms, similar to the one obtained for Government. It is actually integrated into the 55 equations block.

Then the Cobb-Douglas version

Number of equations: 88

Number of independent blocks: 7

Number of simultaneous blocks: 3

Number of recursive blocks: 4

Block 1: 1 Recursive Equations

ict(43)

Block 2: 51 Simultaneous Equations

gdpm(1)	q(2)	ur(3)	kd(4)
lfd(5)	lf(6)	ic(8)	lt(9)
id(10)	i(11)	cap(13)	popac(14)
un(15)	unr(16)	pq(18)	pp(19)
pdf(20)	pdfxt(21)	pc(22)	pi(23)
wr(26)	cost(27)	px(28)	pm(29)
er(30)	irs(31)	irl(32)	ir(33)
relc(35)	revq(36)	revx(37)	socb(38)
wg(39)	w(40)	scw(41)	hi(42)
hdi(44)	hrdi(45)	coh(46)	qval(47)
gdpmval(48)	wf(50)	pmt(60)	compm(61)
fd(62)	m(63)	compx(64)	x(65)
mval(66)	xval(67)	vat(76)	

Block 3: 11 Recursive Equations

pl(7)	k(12)	uwc(17)	pig(24)
res_wr(25)	irm(34)	pgdpm(49)	subs(51)
marg(52)	rmarg(53)	ifp(54)	

Block 4: 3 Simultaneous Equations

nif(55) prof(56) fcapf(59)

Block 5: 18 Recursive Equations

rprof(57) rprob(58) rcval(68) rcvol(69)
ttrad(70) trb(71) nixl(72) nixx(73)
nix(74) fcapx(75) scf(77) oit(78)
tar(79) scg(80) revg(81) igv(82)
cgv(83) fdgv(84)

Block 6: 3 Simultaneous Equations

nig(85) expg(86) fcapg(87)

Block 7: 1 Recursive Equations

fcapgp(88)

The number of block is now 7. This comes from the fact that profits no longer influence investment as a variable. For firms, the loop: interests – profits – balance is now disconnected from the main loop, just as for Government.

10.2.2.9.2 Producing the data

We hope the comments included in this program make it self-explanatory. Basically, the OECD data is used to create the model series in sequence, and the results are saved. Of course, elements used in any computation must have been created in a previous statement (if the sequence defined a model it would have to be recursive).

To adapt this program to another model using the same concepts, we advise to replace all references to the original “FRA_” by the specific name in the original file. This should produce a nearly correct version, with some exceptions:

- Unavailable variables which will have to be created (a common example is capital)
- Variables known using a different definition. This will change the logic of the statements. For instance OECD provides wages including and not including contributions, but not the contributions themselves, which have to be computed as a difference. If they are available, a direct transfer is possible.

10.2.2.9.3 Creating the model groups and checking data – equations

This program is also self-explanatory. The elements produced are:

- Groups for the endogenous (separated automatically into identity and behavioral) and exogenous.

As explained in the program, the fact that the behavioral equations, and those only, contain the element « f » is used to separate them from the identities.

One has just to run two « residual check » simulations with different values of “f”, and identify as behavioral the variables for which the results are different.

- A set of absolute and relative differences between the historical series and the result given by the associated equation.

Let us make a comment on checking that the residual is null.

Actually, the value of the residual is seldom exactly null. Due to the limited precision of EViews, we generally get a value like 10^{-6} (percent, so zero to an 8 digits precision). Once the model reaches a certain size, it becomes difficult to check by sight that the value is sufficiently small (the error can concern a limited range of periods, even a single one).

There are at least two ways to treat the problem, one immediate but limited, the other harder to implement but more efficient and exact.

- The first is to produce an EViews graph. If all residuals are negligible, we should get a set of Brownian motions, with a very low higher and lower maximum values. So we know immediately if there are no errors. But it is difficult to identify the culprits (and an EViews graph is limited in series display).
- The second is to export the relative errors to an Excel sheet, sort the file for a given year, and concentrate on the top and bottom values. The process is repeated until only negligible errors remain.

10.2.2.9.4 Estimating the equations

We have already presented its elements in detail.

10.2.2.9.5 Solving the model on the future

The solving process gives the same results as usual. If convergence is achieved (our case), no message is produced. If not, we have described earlier the techniques one can use to solve the problem.

10.2.2.9.6 Producing a forecast

In our case, the goal is not to produce an actual forecast, but a simulation giving results acceptable enough to be used as a base for the shocks which will follow.

This property can be controlled by displaying the growth rates of the main variables in the short and medium runs.

In our case, the model converges normally under Gauss-Seidel, to values not very different from history for the first periods (we have only presented the convergence of the rate of use UR and the value added deflator PQ).

Model: _MOD_1

Date: 06/12/11 Time: 16:00

Sample: 2005Q1 2020Q4

Solve Options:

Dynamic-Deterministic Simulation

Solver: Gauss-Seidel

Max iterations = 5000, Convergence = 1e-008

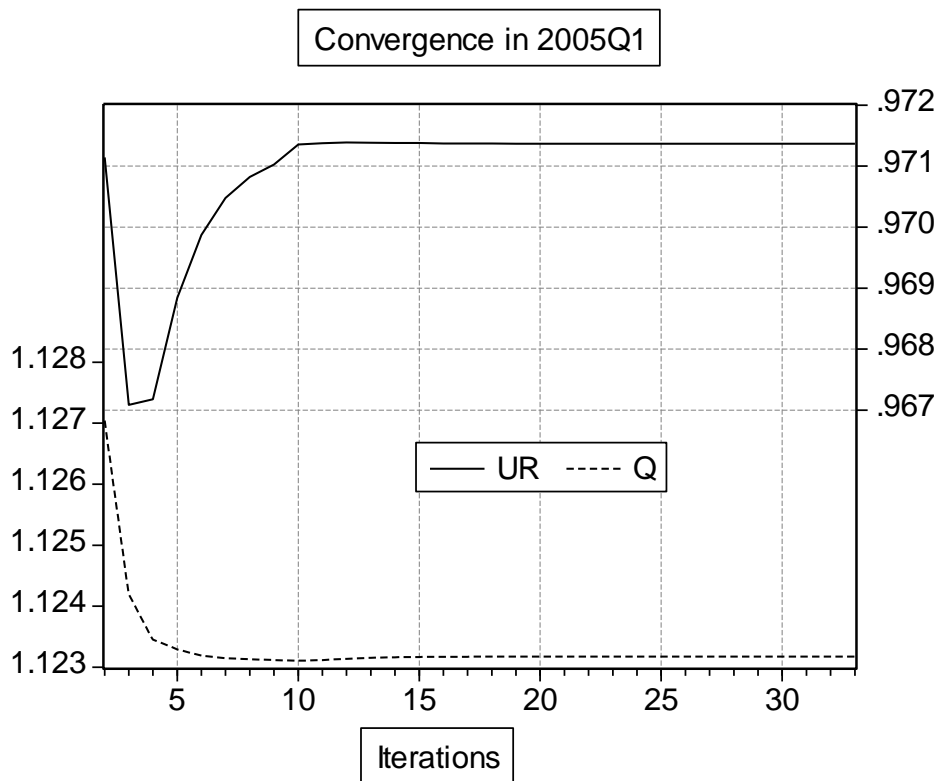
Scenario: Scenario 1

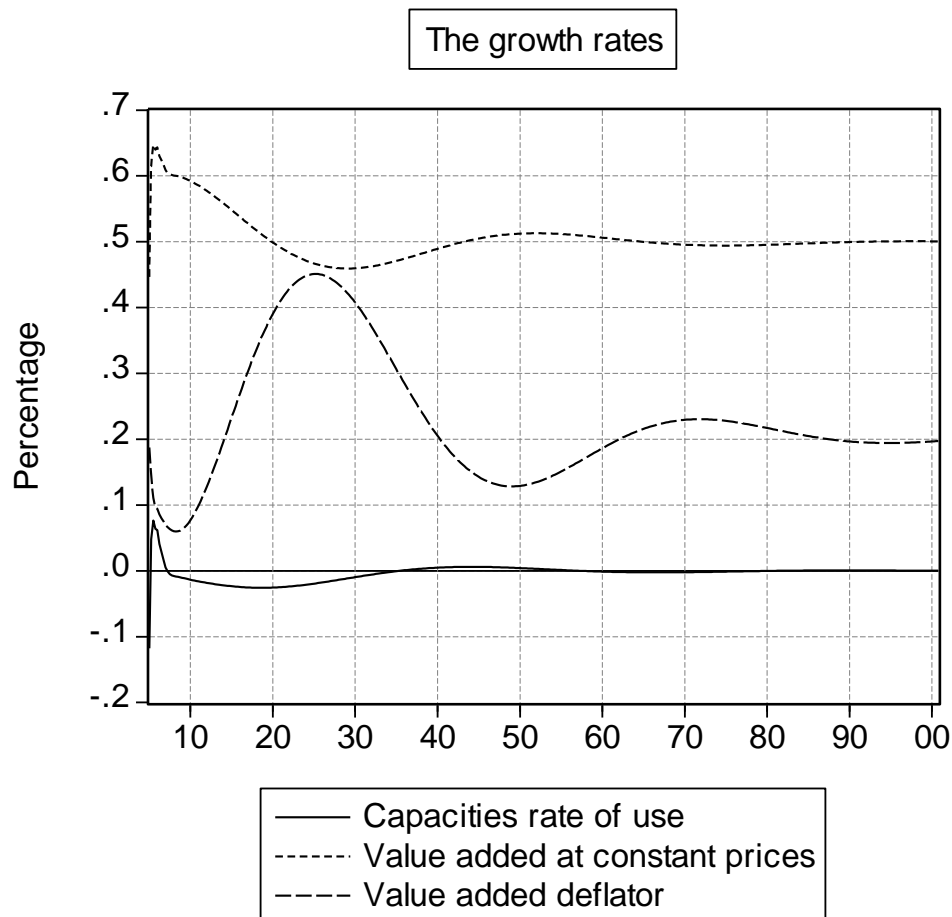
Solve begin 16:00:24

2005Q1	Block 1 - 1 eqns	Solved (recursive block)
2005Q1	Block 2 - 51 eqns	Convergence after 41 iterations
2005Q1	Block 3 - 11 eqns	Solved (recursive block)
2005Q1	Block 4 - 3 eqns	Convergence after 7 iterations
2005Q1	Block 5 - 18 eqns	Solved (recursive block)
2005Q1	Block 6 - 3 eqns	Convergence after 7 iterations
2005Q1	Block 7 - 1 eqns	Solved (recursive block)
2005Q2	Block 1 - 1 eqns	Solved (recursive block)
2005Q2	Block 2 - 51 eqns	Convergence after 33 iterations
2005Q2	Block 3 - 11 eqns	Solved (recursive block)
2005Q2	Block 4 - 3 eqns	Convergence after 7 iterations
2005Q2	Block 5 - 18 eqns	Solved (recursive block)
2005Q2	Block 6 - 3 eqns	Convergence after 8 iterations
2005Q2	Block 7 - 1 eqns	Solved (recursive block)
2005Q3	Block 1 - 1 eqns	Solved (recursive block)
2005Q3	Block 2 - 51 eqns	Convergence after 34 iterations
2005Q3	Block 3 - 11 eqns	Solved (recursive block)
2005Q3	Block 4 - 3 eqns	Convergence after 7 iterations
2005Q3	Block 5 - 18 eqns	Solved (recursive block)
2005Q3	Block 6 - 3 eqns	Convergence after 7 iterations
2005Q3	Block 7 - 1 eqns	Solved (recursive block)

2020Q1	Block 1 - 1 eqns	Solved (recursive block)
2020Q1	Block 2 - 51 eqns	Convergence after 43 iterations
2020Q1	Block 3 - 11 eqns	Solved (recursive block)
2020Q1	Block 4 - 3 eqns	Convergence after 9 iterations
2020Q1	Block 5 - 18 eqns	Solved (recursive block)
2020Q1	Block 6 - 3 eqns	Convergence after 8 iterations
2020Q1	Block 7 - 1 eqns	Solved (recursive block)
2020Q2	Block 1 - 1 eqns	Solved (recursive block)
2020Q2	Block 2 - 51 eqns	Convergence after 44 iterations
2020Q2	Block 3 - 11 eqns	Solved (recursive block)
2020Q2	Block 4 - 3 eqns	Convergence after 9 iterations
2020Q2	Block 5 - 18 eqns	Solved (recursive block)
2020Q2	Block 6 - 3 eqns	Convergence after 8 iterations
2020Q2	Block 7 - 1 eqns	Solved (recursive block)
2020Q3	Block 1 - 1 eqns	Solved (recursive block)
2020Q3	Block 2 - 51 eqns	Convergence after 44 iterations
2020Q3	Block 3 - 11 eqns	Solved (recursive block)

2020Q3	Block 4 - 3 eqns	Convergence after 10 iterations
2020Q3	Block 5 - 18 eqns	Solved (recursive block)
2020Q3	Block 6 - 3 eqns	Convergence after 8 iterations
2020Q3	Block 7 - 1 eqns	Solved (recursive block)
2020Q4	Block 1 - 1 eqns	Solved (recursive block)
2020Q4	Block 2 - 51 eqns	Convergence after 44 iterations
2020Q4	Block 3 - 11 eqns	Solved (recursive block)
2020Q4	Block 4 - 3 eqns	Convergence after 9 iterations
2020Q4	Block 5 - 18 eqns	Solved (recursive block)
2020Q4	Block 6 - 3 eqns	Convergence after 8 iterations
2020Q4	Block 7 - 1 eqns	Solved (recursive block)





If we consider now the evolution with time, we can see that all growth rates stabilize after 100 years. But the quantities and the output gap do it rather quickly, as well, while the deflator keeps oscillating for a long time.

10.2.2.9.7 Producing shocks on assumptions: the Complementary factors case

Let us now see how our model answers to shocks on its assumptions. We shall be brief, and only present and comment the results for four of them: our goal is mostly to show that the present framework and the associated estimations can provide a model with consistent properties. With a little bit of luck you should be able to do the same, or probably better. The model(s) we are presenting are far from perfect, and we did not try to make them so¹⁵⁵.

As presented earlier, we are considering two production functions: complementary factors and Cobb-Douglas. Using the two associated models, we will see that their properties are sizably different, due mostly (but not only) to this feature.

We shall start with the first case.

¹⁵⁵ Actually, we can identify a few coefficients which, with different values, would improve slightly model properties. Of course, we will not apply these changes.

All shocks will start in the first quarter of 2006, and will be sustained for the whole period. To make interpretation easier, we shall limit the periods to the first fifteen years (for the short and medium term dynamics) and the last three (for the long term impacts).

10.2.2.9.7.1 AN INCREASE IN GOVERNMENT DEMAND

In this shock, we increase Government demand by 1% of the baseline GDP. This is the first shock one has in mind, as it defines the Keynesian multiplier, showing the way the country's economic mechanisms modify the consequences of an external shock on demand. Technically the multiplier will be the ratio between the ex post and ex ante changes in GDP, the latter being the change in one of the exogenous demand elements.

Observing the role of Government demand (investment IG or consumption CG) we can see that it affects only:

- Final demand, and GDP through the supply – demand equilibrium.
- Government expenses and deficit.

As stated earlier, we do not consider the increase in global productivity coming from Government investment.

Let us concentrate on the supply – demand equilibrium:

$$GDP + M = FD + X$$

With:

$$FD = COH + IC + cg + ig + fdxr * Q$$

In terms of GDP, the ex-ante impact of the shock is of course 1% (this is the reason for the size we have chosen). The evolution to the ex post value comes from the endogenous elements:

- Trade elements: exports and imports.
- Demand elements: consumption, investment, changes in inventories.

Obviously we shall get, for ex post GDP:

- A positive effect from demand, as the need for additional capacities will increase investment, and the new jobs will produce wages, household revenue and consumption.
- But a negative one from imports, as a share of this additional demand will have to be imported.

Two additional effects have to be considered:

- In the short and medium runs, demand will meet a local capacity constraint for some products, which will have to be imported. And as local producers will use a larger part of their capacity to satisfy local demand, they will be less active in looking for export markets.

This effect will disappear gradually as local producers adapt their capacities (through investment).

- Inflation will appear, due to disequilibria:
 - On capacities, as the higher level of production (compared to capacities through the rate of use) will allow optimizing firms to increase their prices. This effect will disappear in the long run.
 - On wages, as the lower level of unemployment will improve the negotiating power of workers. As long as unemployment is reduced, this effect will remain.

Inflation will reduce the competitiveness of local producers, both on the foreign and the local markets.

The following graph illustrates our comments. Ex-post demand grows by 1.5% (compared to 1% ex ante, the French net trade balance being roughly at equilibrium). The multiplier is affected in the short and medium terms by capacity bottlenecks, in the long run by losses in competitiveness. It is actually in the medium run that the loss is the lowest (inflation is still limited, capacities are largely adapted).

The two other graphs show the adaptation of factors to the new production level (with two different time spans). We can see that:

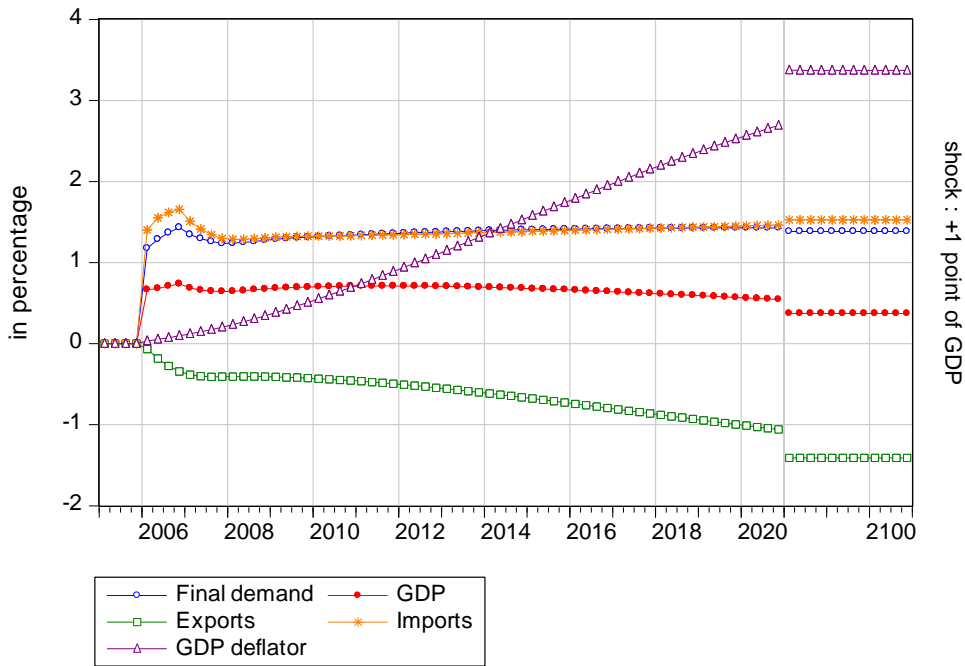
- Employment adapts much faster (following the estimated coefficients).
- In the long run, the rate of use does not revert to the baseline value. This is not too difficult to explain. As the export price is more sensitive to local costs than the import price, a change in the local value added deflator will see its impact dampened on the local demand price, defined by an identity.

$$PFD = (GDPMVAL + MVAL - XVAL) / (GDPM + M - X)$$

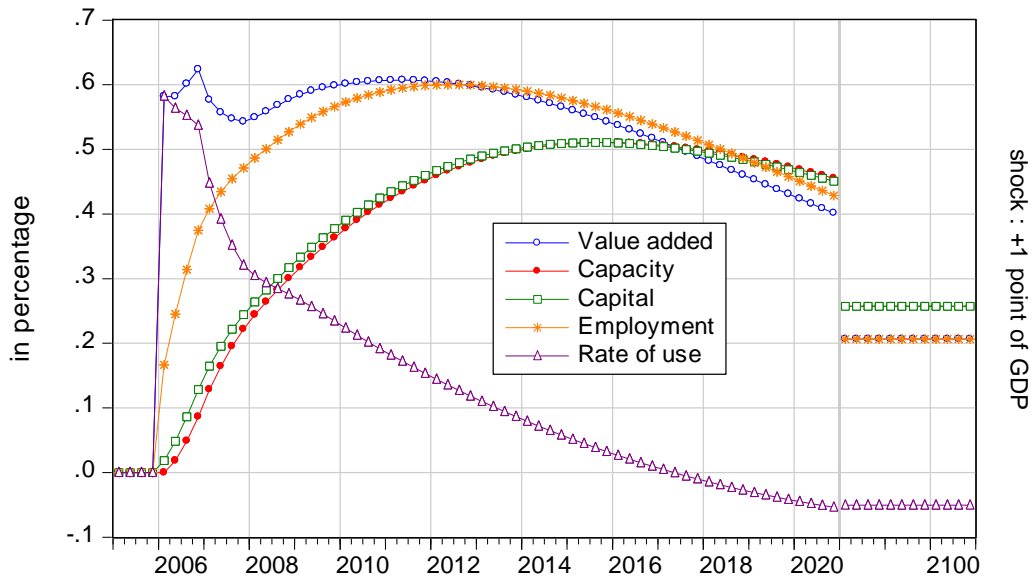
- The relatively lower cost of investment, compared to the revenue which finances it, will increase profitability and allow firms to afford a lower rate of use, following the mechanism presented in the estimation of the value added deflator.
- The lower long run activity levels are of course due to inflation, which takes a long time in reaching its long term value (we can see that the error correcting coefficients are quite low).

By running the associated programs, the reader will be able to observe other elements, such as the ex post reduction of the budgetary cost in the short run (due to the various additional tax revenues) followed by an expansion (as the Government has to pay the interest on the accumulated debt). But he is also expected to produce his own model, and to observe how his changes (limited or extensive) modify model properties.

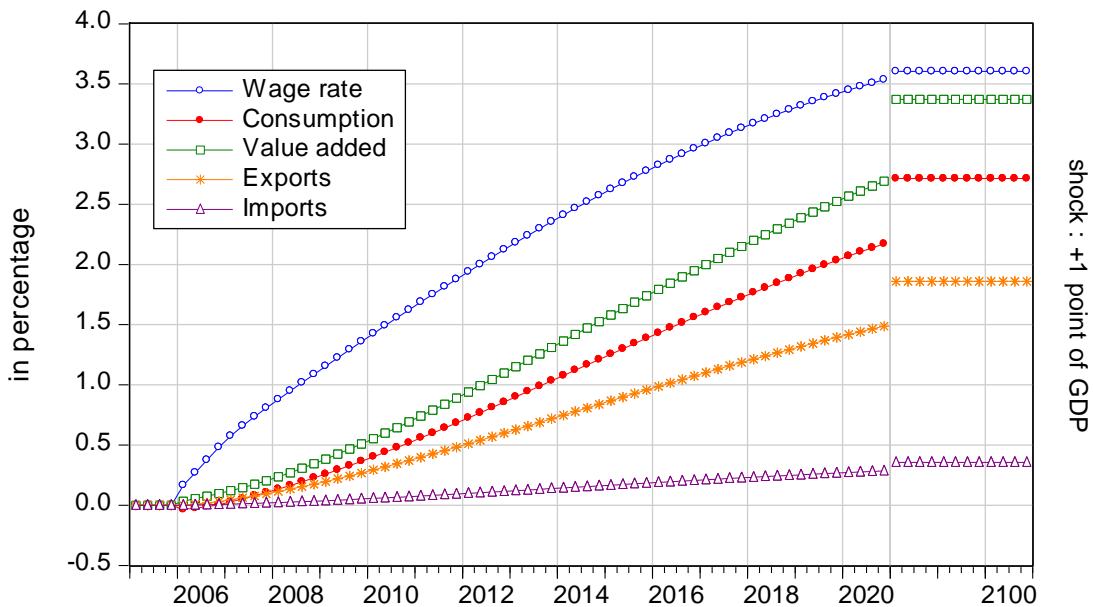
Government demand : The supply - demand equilibrium



Government demand : The production elements



Government demand : The prices



10.2.2.9.7.2 A DEVALUATION (OF THE FRENCH FRANC!)

Of course this shock is not realistic, as France shares with other countries its present currency.

We will show the results nonetheless, to illustrate the general consequences of a devaluation, using our proposed model framework. Let us suppose France had not joined the Euro zone (or left it!).

We will depreciate the French currency (the Franc?) by 1%.

Our first graph presents quantities, the second prices.

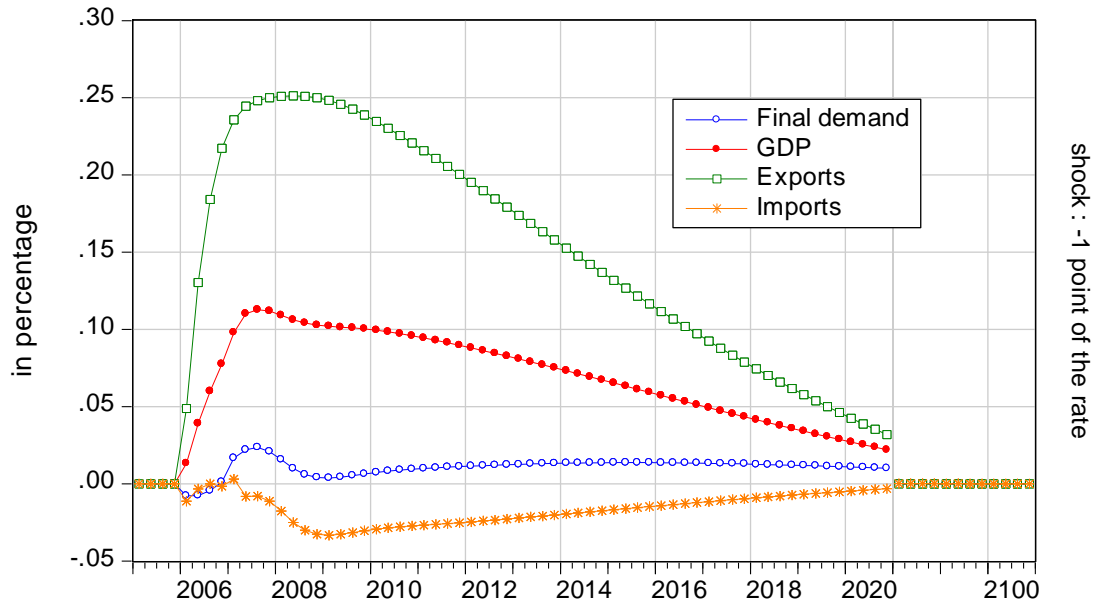
Concerning the latter, we see that in the long run they all increase by 1%. But the increase is slow, and presents some overshooting in the medium run (due to the inertia coming from low error correcting coefficients). Logically, the imports (essentially) and exports deflators are the fastest to take the decision into account, and also present the lowest overshooting.

Concerning quantities, exports profit the most from the shock, improving GDP. But imports actually increase a little, the gains in competitiveness being more than balanced by the additional demand (not only from local investment and consumption, but also from exports which call for importing intermediary goods).

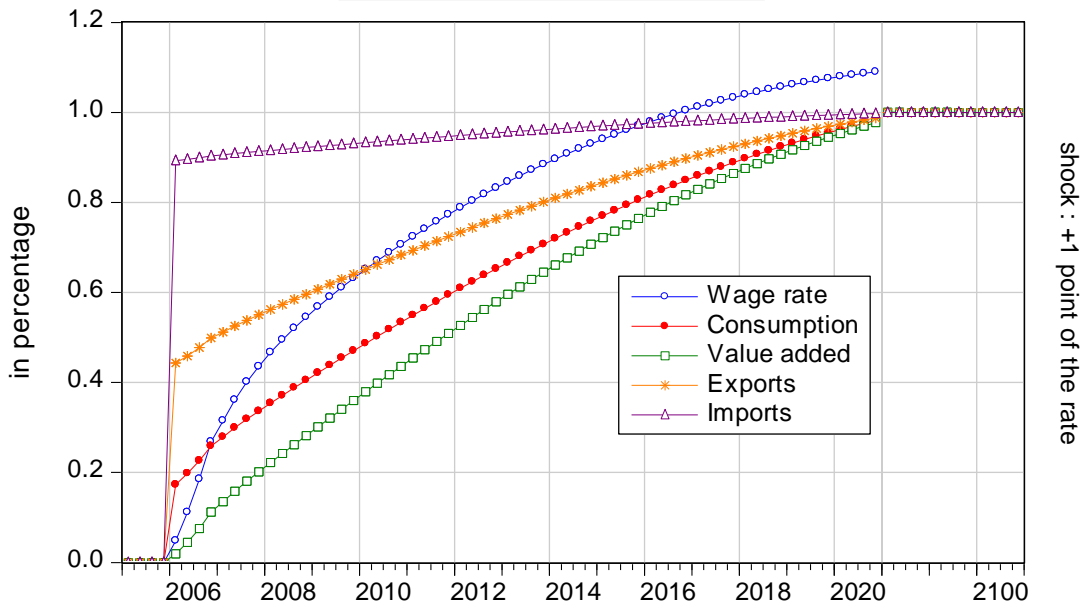
In the long run, the full adaptation of inflation makes all effects disappear, after a temporary reversal due to the overshooting of local deflators.

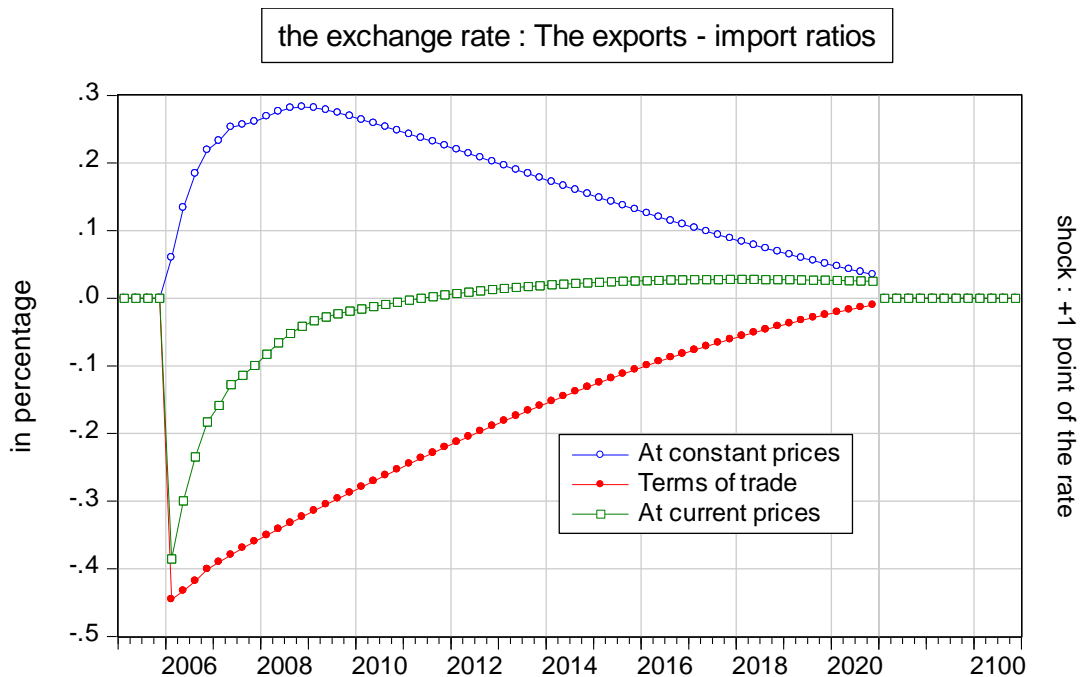
The last graph presents the evolution of trade. We see that the gains in competitiveness (associated with a loss on the terms of trade) are reduced gradually from the first, but that the gains in real trade take some time in reaching the highest level (due to the initial bottlenecks and increase in demand). A very limited gain at current prices appears only in the medium run.

shock on the exchange rate : The supply - demand equilibrium



the exchange rate : The prices





10.2.2.9.7.3 A DECREASE IN THE FOREIGN TARIFFS RATE

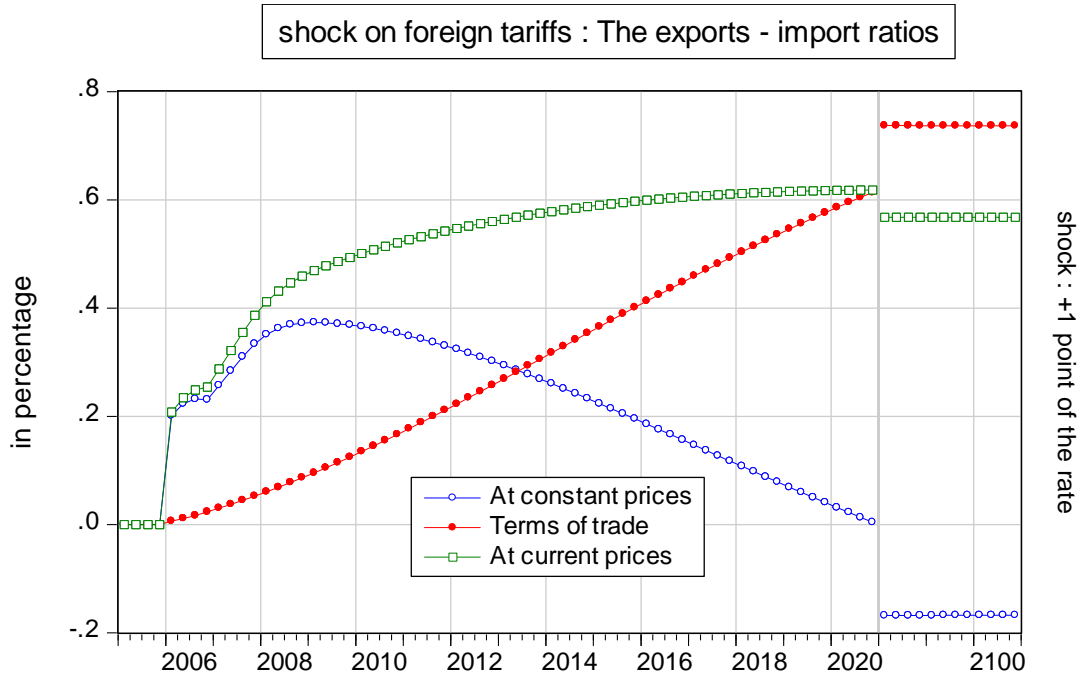
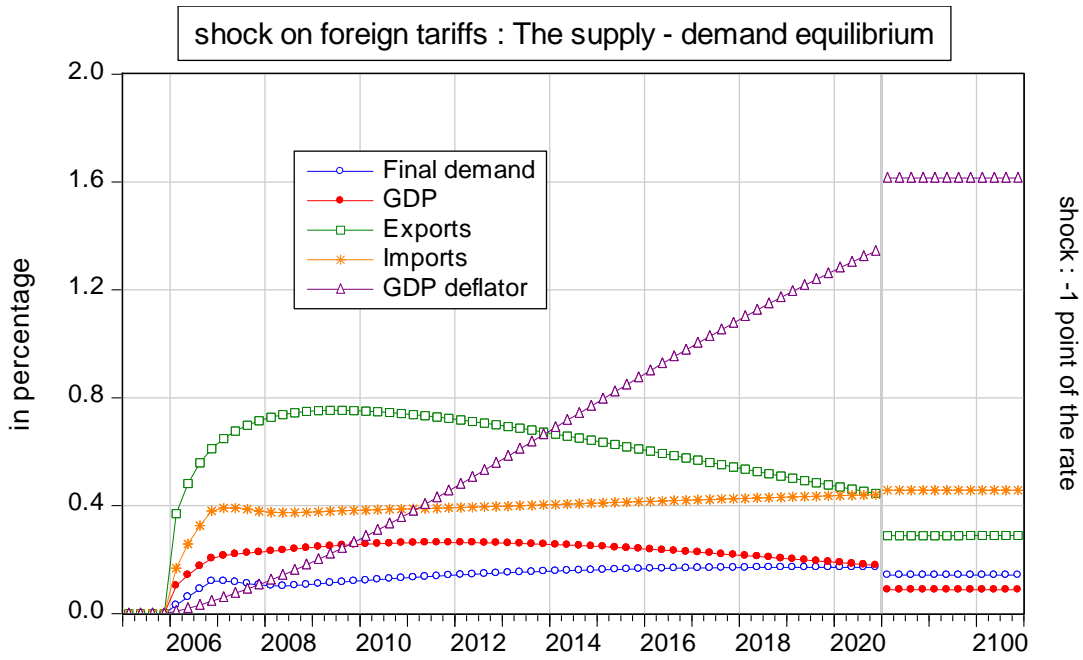
As you have probably guessed, this shock is essentially demand-oriented:

The improved competitiveness will increase the demand addressed to France, with basically the same effects on GDP as Government demand, from the supply – demand equilibrium.

Of course lower import prices from France will reduce inflation in the rest of the world, which should affect French import prices. World GDP should also change (in a dubious way, positive through disinflation, negative through the higher share of France in world trade). But both these effects can be considered negligible, especially if we consider the cost of the alternative: building a reliable world model.

It is interesting to observe that the increase in local activity inverts the gain in trade at constant prices in the long run, but not as current ones as this loss is due to local inflation, which “improves” also the terms of trade.

On the whole, the current gain stabilizes after a while, when capacities have adapted to the new demand level.



10.2.2.9.7.4 A DECREASE IN THE LOCAL TARIFFS RATE

This shock brings traditionally the most complex mechanisms, and its consequences are quite volatile from one model to the other. Two main channels have to be considered:

- The improved imports competitiveness increases their share in local demand, reducing local output, with the traditional consequences of a demand shock, only negative.
- The lower import prices brings global disinflation, especially for demand but also for value added through lower wages (indexed on the consumer price). This disinflation helps local firms to compete with foreign producers on the foreign and local markets, reducing the initial gap on the first.
- The cost of capital (part of it imported) decreases more than the value added price. This increases profitability and creates investment and productive capacity per se, independently from demand. Firms working at full capacity will see the constraint released¹⁵⁶, and local producers as a whole will gain market share on the local and foreign markets. At the same time, to improve demand for these new capacities, they will decrease their prices.

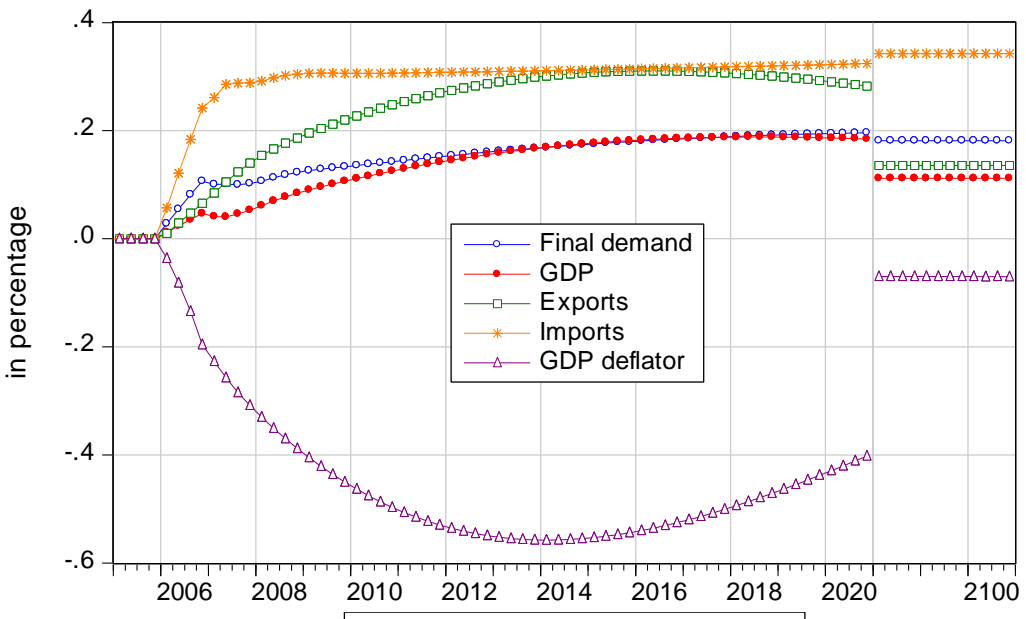
We can see:

- That imports show the highest increase, as the only ex ante positively influenced element. But the improvement of local competitiveness reduces that effect, and allows exports to increase, without compensating the initial gap. But GDP actually increases. This would be strange if net exports was the only determinant of GDP, but we observe also a significant increase in local final demand, coming from the cheaper consumption and investment.
- That the decrease in inflation follows a logical order (the same as in the legend). We observe in sequence:
 - The import price excluding tariffs, linked essentially to the world price.
 - The export price, less sensitive to the world price.
 - The demand price, a weighted average of imports and local prices (with a small impact from exports).
 - The production price, a weighted average of value added and demand (intermediate consumption).
 - The value added price, the most subject to deflation, as in addition to the lower consumption price wages are sensitive to the higher unemployment.
 - The wage rate, for the same reason.
 - The import price including tariffs, of course.

Finally, the trade balance is constantly worsened, at a rather constant level, first through the increase in real imports, then by the loss on the terms of trade. If the trade balance in real terms is allowed to recover, it is at a loss on relative prices.

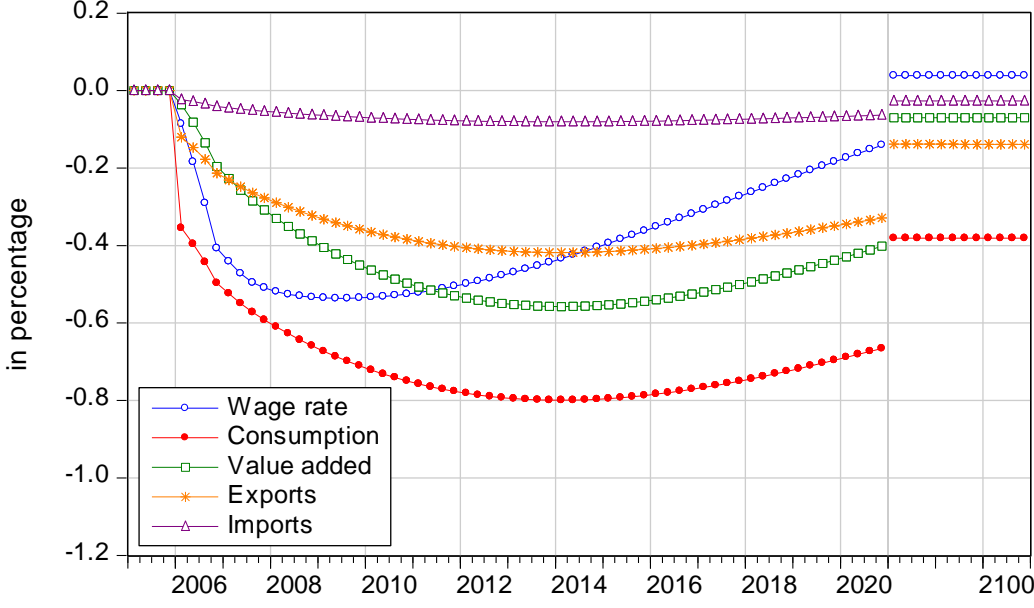
¹⁵⁶ And possibly reappear at a higher production level.

shock on local tariffs : The supply - demand equilibrium

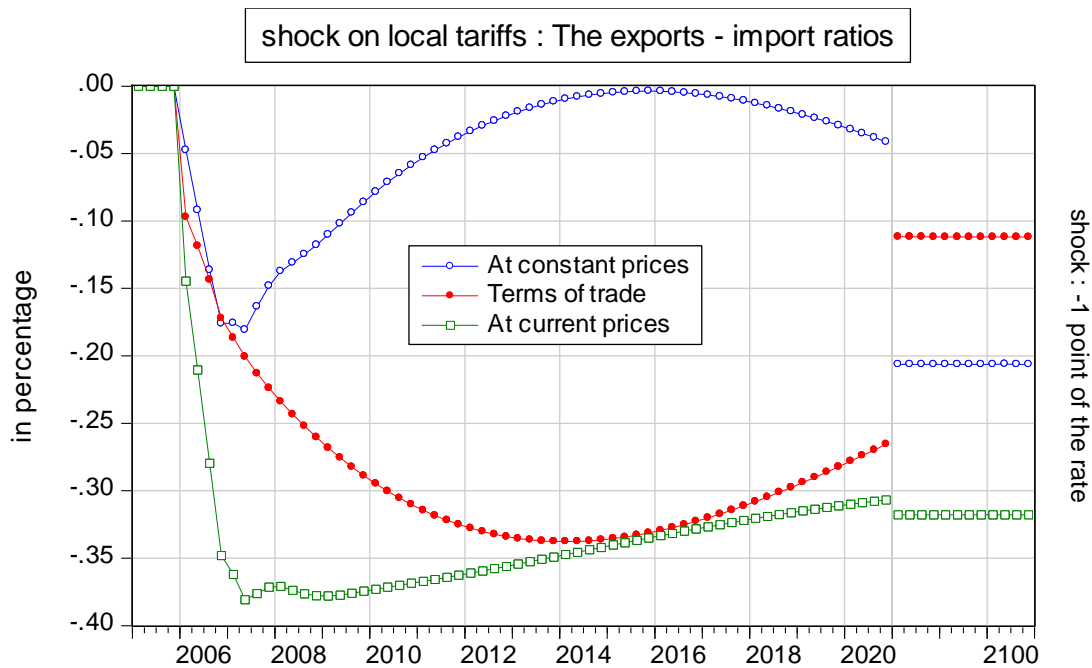


shock : -1 point of the rate

shock on local tariffs : The prices



shock : -1 point of the rate



10.2.2.9.8 The Cobb-Douglas case

Let us see now how the use of a Cobb – Douglas production function modifies the diagnosis. First, let us introduce a shock on Government demand.

To make the results comparable to the Complementary Factors case, we shall calibrate the import equation, giving to price competitiveness the same elasticity.

Re-estimating the other coefficients, we observe the values do not change too much. We have decided to keep the original ones.

This is done through:

```
' estimating imports

vector(10) p_m
smpl 1962Q1 2004Q4
genr m_ec=0
smpl 1962Q1 2003Q4
equation _eq_m1.ls(p)
dlog(m)=1*dlog(fd+q)+c_m(2)*log(ur)+c_m(3)*log(@movav(compm,6))+c_m(4)+c_m(5)*(t-
2005)*(t<=2005)+[ar(1)=c_m(6)]+c_m(6)*log(m(-1)/(fd(-1)+q(-1)))+m_ec
p_m(1)=1
p_m(2)=c_m(2)
p_m(3)=-0.22
p_m(5)=c_m(5)
p_m(6)=c_m(6)
```

```
equation _eq_m.ls(p)
dlog(m)=p_m(1)*dlog(fd+tc*q)+p_m(2)*log(ur)+p_m(3)*log(@movav(compm,6))+c_m(4)+p_m(5)*(t-
2005)*(t<=2005)+[ar(1)=c_m(6)]+p_m(6)*log(m(-1)/(fd(-1)+q(-1)))+m_ec
_mod_1.merge _eq_m
```

Reestimating would have been done through (remember that the elasticity to demand is calibrated in all cases):

```
vector(10) p_m
smpl 1962Q1 2004Q4
genr m_ec=0
smpl 1962Q1 2003Q4
p_m(1)=1
p_m(3)=-0.22

equation _eq_m.ls(p)
dlog(m)=p_m(1)*dlog(fd+tc*q)+c_m(2)*log(ur)+p_m(3)*log(@movav(compm,6))+c_m(4)+c_m(5)*(t-
2005)*(t<=2005)+[ar(1)=c_m(6)]+c_m(6)*log(m(-1)/(fd(-1)+q(-1)))+m_ec
_mod_1.merge _eq_m
```

10.2.2.9.8.1 AN INCREASE IN GOVERNMENT DEMAND

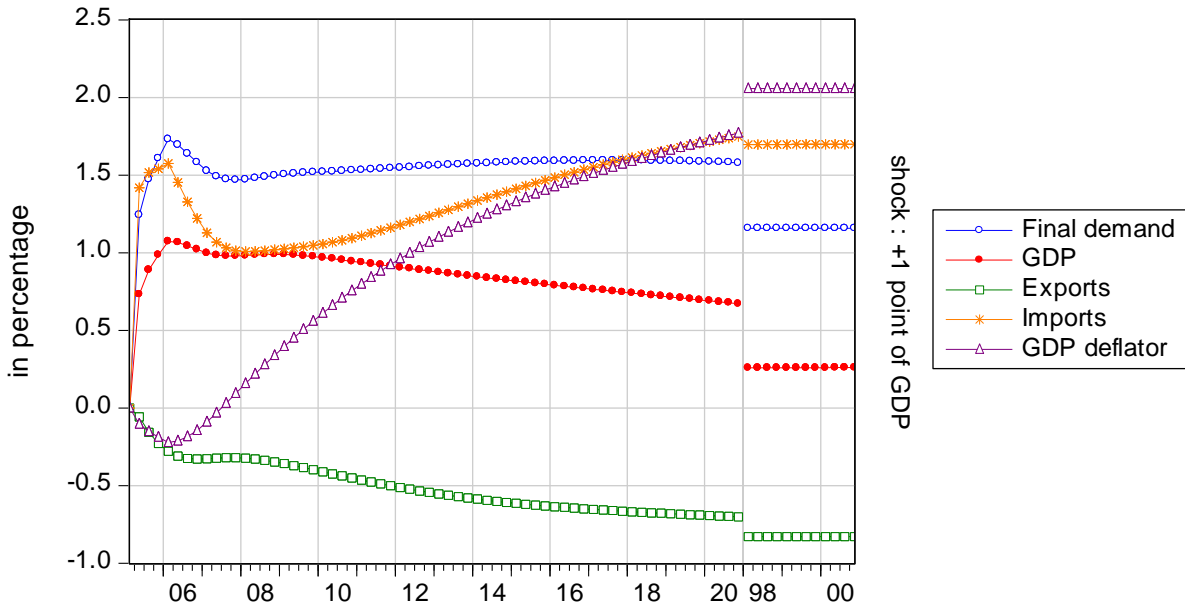
Again, we increase Government demand by 1% of the baseline GDP. Compared to the previous case:

The inflationary effect is lower, with similar dynamics. This comes from the cost variable used by firms to target their margins: to wages is now added the cost of capital. Both are indexed on the demand price, which grows less than the value added one. But on wages this impact on is more than offset by the gains from lower unemployment. This means that compared to the value added deflator, the wage cost goes up, the capital cost goes down.

The multiplier is higher (a logical consequence). The loss on exports is smaller, imports increase less than final demand (but remember they depend also on exports).

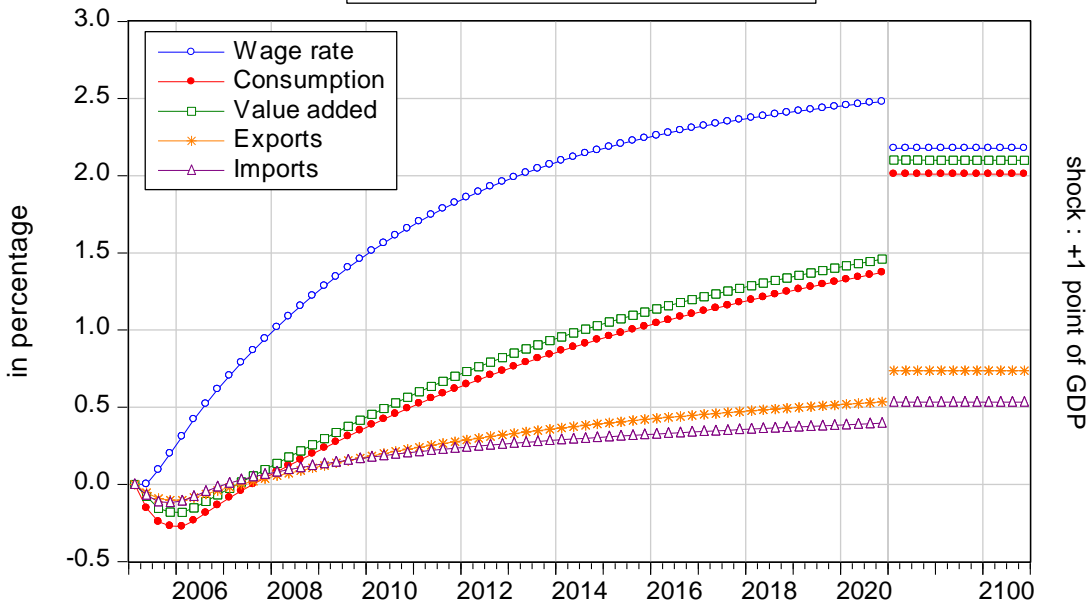
The short term presents a minor cycle.

Government demand : The supply - demand equilibrium



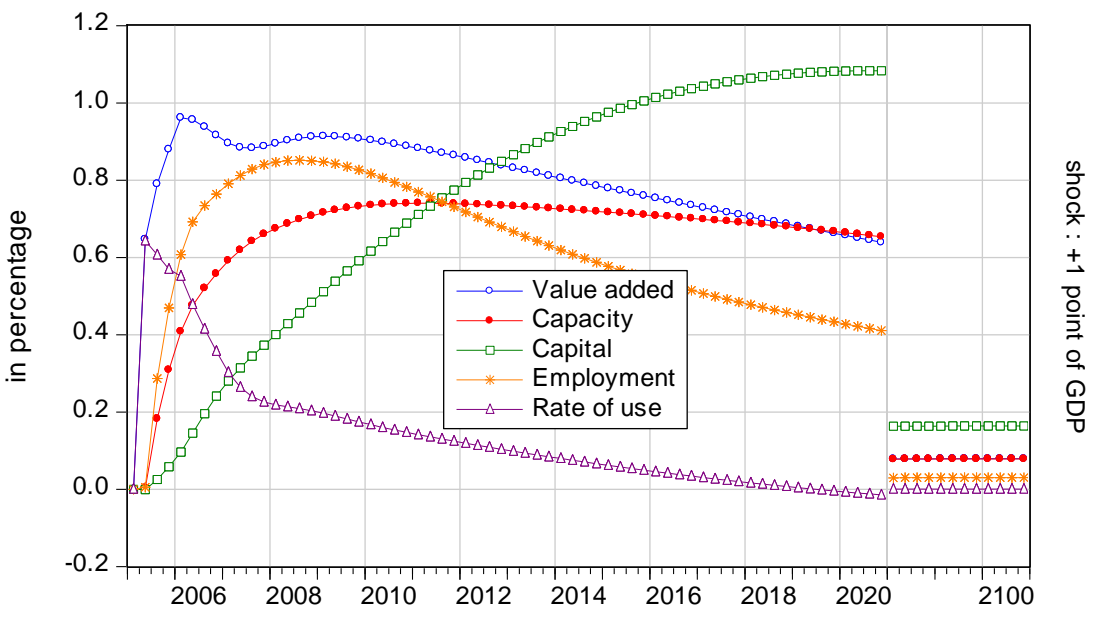
The following graph lists price elements.

Government demand : The prices

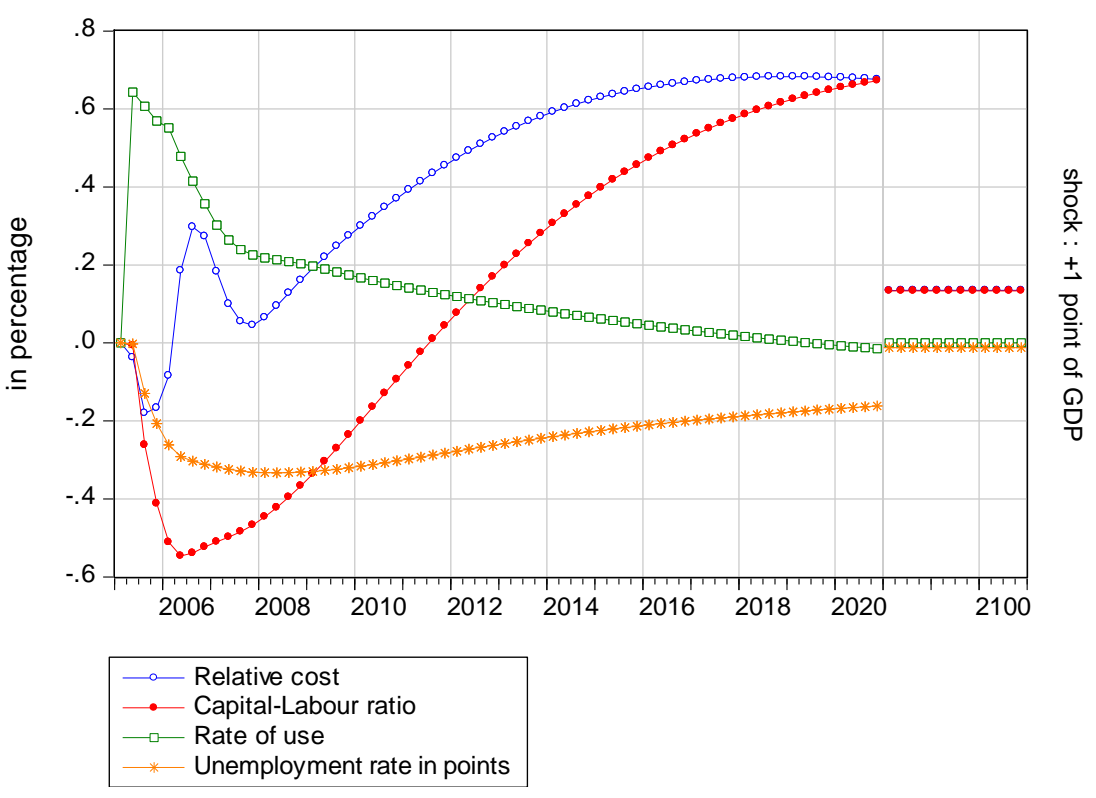


The next graphs present the productive process, and the relative evolutions of factors and their cost.

Government demand : The production elements



Government demand : The ratios



We see that

- The cost of labor increases compared to capital (from the lower unemployment rate).
- This makes capital increase more than labor.

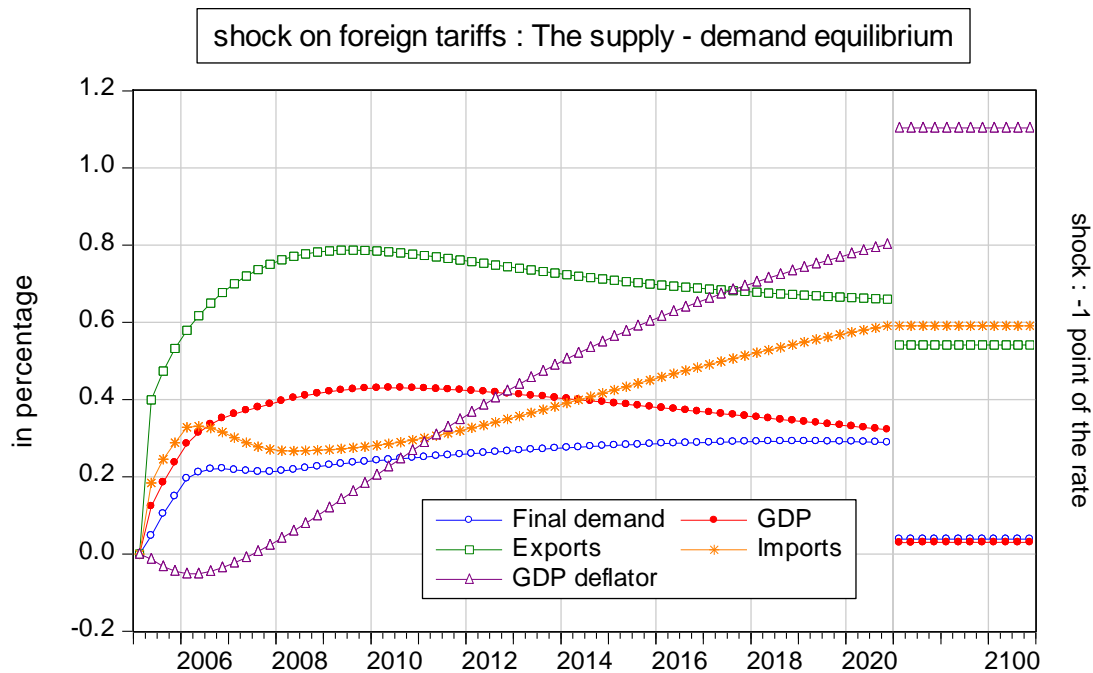
- But the effect takes some time in taking place, principally due to the inertia on investment.

10.2.2.9.8.2 A DEVALUATION (OF THE FRENCH FRANC!)

The devaluation case does not show any interesting difference.

10.2.2.9.8.3 A DECREASE IN THE FOREIGN TARIFFS RATE

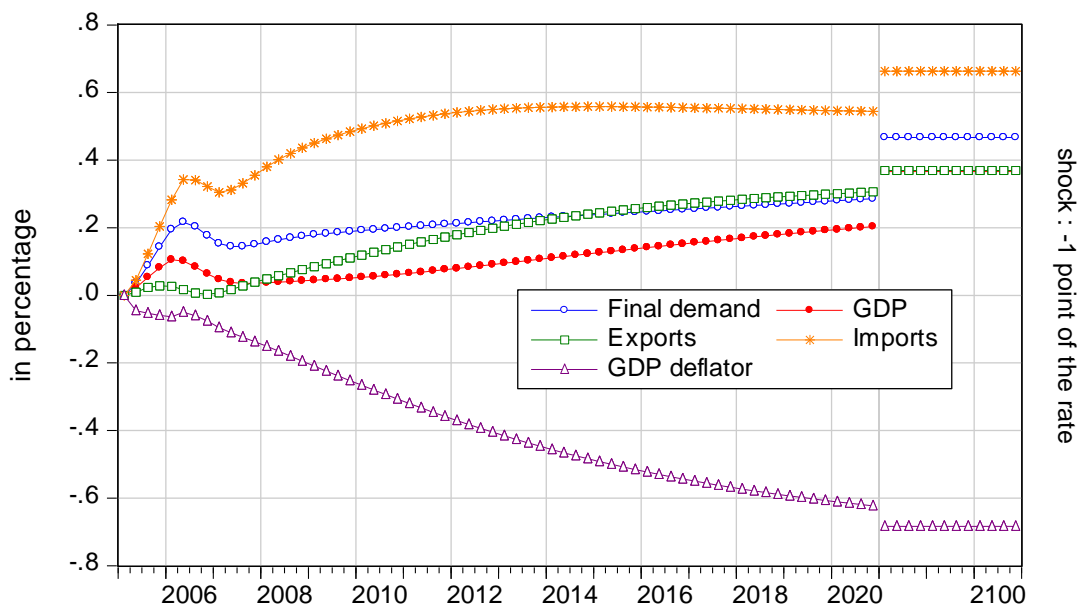
For this other demand shock, the changes come only from the lower inflation.



10.2.2.9.8.4 A DECREASE IN THE LOCAL TARIFFS RATE

The main impact of the change in formulation is quite logical: as the new definition of the cost includes –largely imported) equipment goods, it goes down more, leading to more disinflation and higher increase in trade. Of course, this increase leads to more imports, and GDP does not really improve.

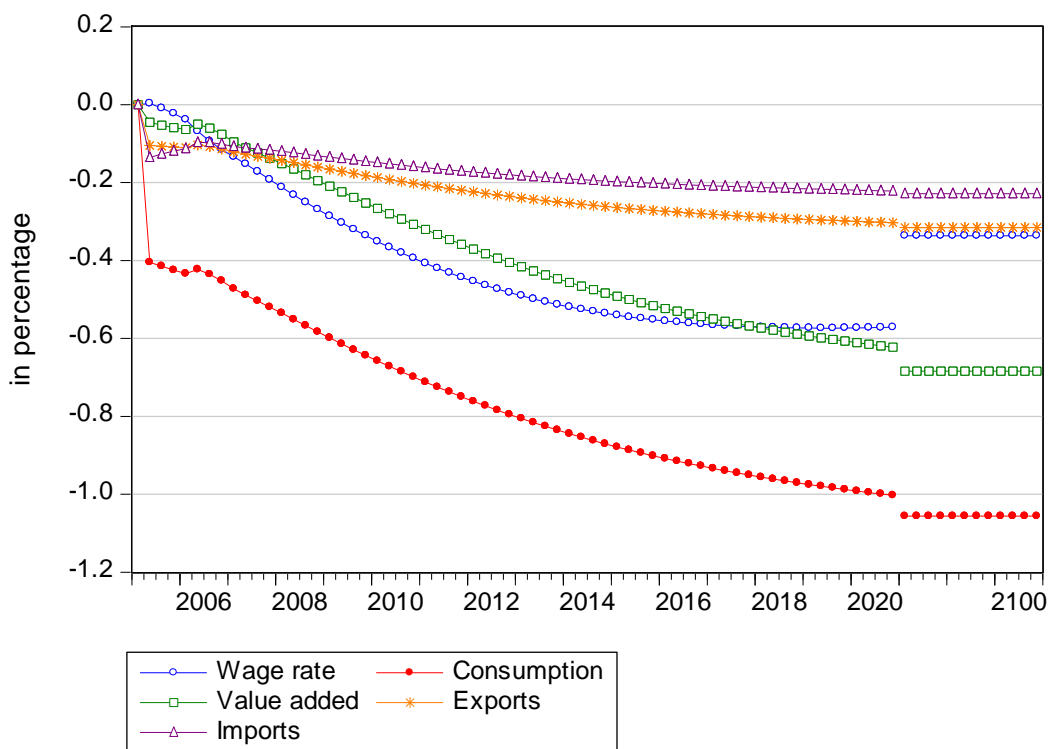
shock on local tariffs : The supply - demand equilibrium



shock : -1 point of the rate

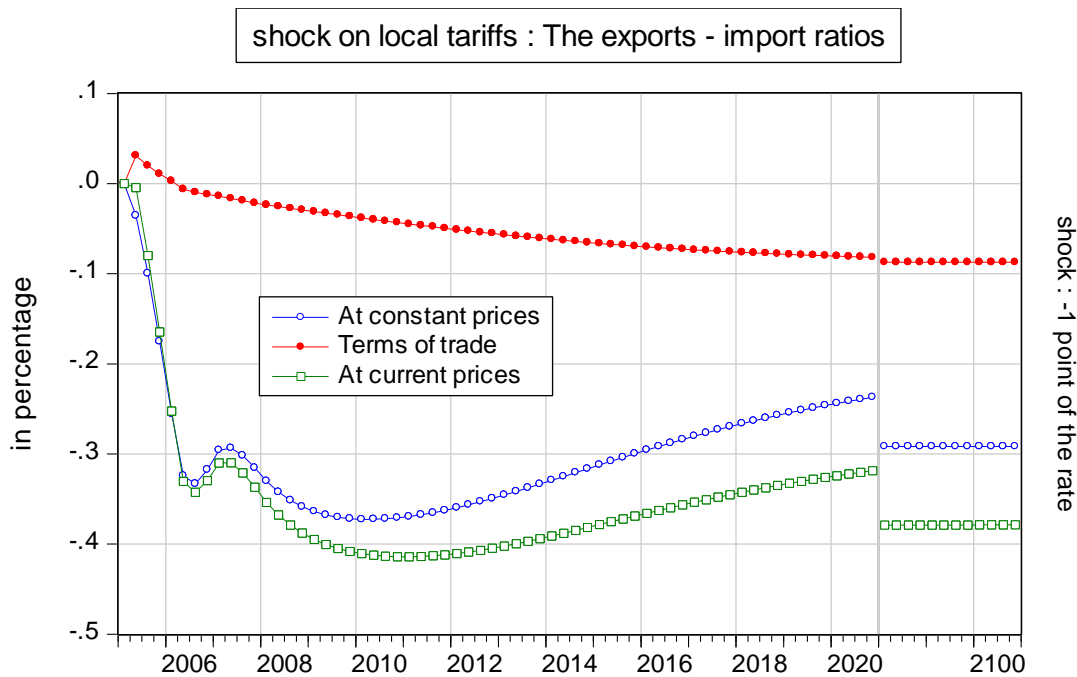
Concerning prices, the evolution is much more regular, and convergence to the long term decrease is monotonous.

shock on local tariffs : The prices



shock : -1 point of the rate

As to the trade balances, we can see again that at current prices a higher improvement of the balance at real prices being compensated (and also justified by) by a higher loss on the terms of trade. But as for prices the evolution is much more regular than in the complementary factors case.



10.3 A SINGLE COUNTRY, MULTI PRODUCT MODEL

Let us now consider separating the products of our model into several categories. We shall address in turn:

- The reasons for product decomposition
- The specific features introduced by product decomposition, and the way it can be treated.

10.3.1 THE MAIN REASONS FOR PRODUCT DECOMPOSITION

There are three main reasons for product decomposition:

- The decomposition allows to take into account differences in the values of structural elements.

This appears when structural parameters take different values from one product to the other, which implies that they will react differently to external influences (exogenous or endogenous), or exogenous assumptions concerning the product.

- The decomposition allows to evidence differences in the sensitivity to some explanatory elements.
- For some economic element or field, the links between variables follow different mechanisms, calling for different specifications which cannot be summarized by mathematical aggregation.

10.3.1.1 Structural differences

They can be identified easily within the framework of our model. The main elements are:

- Factor productivity: the quantity of labor and / or capital necessary to produce one unit of good can be different. This is particularly true for agriculture, in which labor productivity is particularly low¹⁵⁷. And also for services, where the need for capital is generally limited.
- The wage rate: the average wage obtained by workers can be quite different (linked often to labor productivity, which can offset the effect when we consider the unit cost).
- The different shares in each demand item, allowing to present different sensitivities to a given increase in consumption or in investment.
- The amount of intermediate consumption of each good needed to produce one unit of a given good.
- The sharing of production destination between the local and foreign markets (exports).
- The sharing of local demand between the local and foreign producers (imports).
- The separation between firms, households, the state and foreign agents in the local production process, with different consequences on revenue and its use.
- The taxation of goods (VAT, other indirect taxes, tariffs on local imports and exports).

All these elements will be taken into account at no cost, by separating the goods in the model.

10.3.1.2 Differences in sensitivity

This can apply to all parameters in estimated equations. Of course, one should be interested on the most important ones, and economic theory can point out for which elements significant differences can be expected. For instance, the need for inventories might be more important in the manufacturing sector, or the role of unemployment limited in defining the agricultural wage, or the sensitivity of external trade to price competitiveness can be low for services.

10.3.1.3 Differences in logical behavior

This time the very formulation of a behavior, or its causal logic, will be different.

This applies essentially to two domains.

10.3.1.3.1 The production process

- In the agricultural sector, one can assume that producers maximize their output. For this, they can use plantations (fields and trees), and animals, or a combination of both (like raising trout in ponds). They need also a minimum amount of labor, tools and intermediate consumption goods (like fodder). Increasing these levels can improve output, as well as better infrastructures (irrigation, storage and transportation facilities)¹⁵⁸.

¹⁵⁷ Especially in family managed units.

¹⁵⁸ Remember that production includes only the goods actually bought (consider the supply-demand equilibrium) or entering inventories, which do not apply to agricultural goods (processed goods are considered industry). This means

But once these are defined, production follows, corrected heavily by unforecasted climatic conditions. This is also mostly true of fishing. The impact of an increase in demand is limited at least in the short run (even if perhaps more cattle will be slaughtered, or more fish will be caught).

- In the manufacturing sector, we can assume the availability of several processes, among which the producer will choose according to the relative costs of capital and labor. Assuming a constant and unitary elasticity leads to a Cobb-Douglas formulation. We can also choose a simple complementary factors specification, as well as a more general CES function. Actually, one of the advantages of product identification is to associate a sophisticated function to the goods which actually deserve it. In the single good case, the elements of the alternative: using a simple function or associating a Cobb-Douglas to the whole economic spectrum are both harmful to model quality.

But the main advantage of this identification is to define a productive capacity, and the associated rate of use. This element will play an important role (as we have seen already) in three elements: the role of producers in the satisfaction of local and foreign demands, the investment behavior, and the short-term choice by firms of the mark-up applied to the production cost.

- Finally, in the services sector, the choice of factors can depend on the relative cost (computers will replace people), but the role of capacity is less clear, as in many cases producers have a considerable leverage on output, independently of installed capital and even employment: a sudden and high increase in the demand for touristic trips can generally be met by travel agencies, if customers are willing to change their plans. The quality of the service will probably decrease, but not its cost, the element by which the output is measured. Dissatisfaction of customers does not reduce the value of a bought good, except if some reimbursement is done.

10.3.1.3.2 External trade

The trade in agricultural goods, manufactured goods and services follows obviously different rules.

- For agriculture, the export price is generally set to the world price for a given quality, and a given share of the local production is proposed on the foreign market, according to the type and quality of goods, local policy, pre-established contracts.
- For manufactured goods, additionally to demand, price competitiveness is essential, as well as available capacity which can create temporary bottlenecks on the supply of specific elements.
- For services, the determinants are completely different and much less clear, but the amounts traded are less important (except for tourism which is quite difficult to model).

spoiled products are not considered, and improving transportation will increase agricultural production, all things being equal.

10.3.1.3.3 Consequences for product decomposition

The above remarks should have made clear that the minimal decomposition should introduce three categories: primary, secondary and tertiary products.

Going further, one can consider

- Using more products (more on this later).
- The most immediate idea concerns energy, especially for oil or gas producing countries. It is clear that the level of production can change widely according to the will of the producer, that the variable production costs are relatively small, and that the whole production should have no problem being sold, at a price highly variable but defined by the world market.
- This is also true in general of mining, which should therefore be counted with the secondary goods, in its own category, or together with energy.
- Construction can also be identified, as it follows the other products, and household revenue. The production process is relatively straightforward for a given type, and it is neither imported nor exported, although its inputs can be (like wood and steel).
- Introducing the type of ownership. Firms can be divided in state and private, and the latter into individual firms and companies. Foreign owned firms can be identified (using FDI, repatriating profits).

10.3.2 INTRODUCING INTERMEDIATE CONSUMPTION

A very important development corresponds to intermediate consumption.

Any model, for any product, has to follow the equilibrium:

$$\text{Production} + \text{imports} = \text{Total demand} + \text{exports}$$

However, at the global level, subtracting total intermediate consumption from each side gives:

$$\text{GDP} + \text{imports} = \text{Final demand} + \text{exports}$$

This almost frees the model from defining intermediate consumption, a concept which is not easy to handle, as it depends from the product decomposition itself, more precisely on the number of stages in the production process.

Its only remaining role is the fact that in trade equations, price competitiveness is based on production prices: preference for goods from a country is obviously based on the price at which they sell.

This means that the trade prices (exports and imports) are production prices, and that competitiveness compares the global cost of the exporter and the price at which its competitors sell (both including intermediate consumption). This

means for example that a country which has access to cheap oil or gas (like the US, Russia or Venezuela) has a comparative advantage over Germany or China, both against exporters to its market and against other competitors on foreign markets¹⁵⁹.

The above equation

$$\text{Production} + \text{Imports} = \text{Total demand} + \text{Exports}$$

Can be written as:

$$\text{Value added} + \text{intermediate consumption by the product} + \text{imports} = \text{Total demand} + \text{intermediate consumption of the product} + \text{exports}$$

This calls for the identification of a matrix of intermediary consumptions, a square one with the dimension of the decomposition.

Typically, the definition of each of these elements at constant prices will suppose that producing each unit of a given good requires a given amount of each good (including itself).

At current prices, one could apply the deflator of the product used (excluding VAT, as it does not apply here) at the global demand level, or at the global intermediate consumption level. The availability of the matrix at current prices improves the situation by allowing to define an individual deflator for each cell.

10.3.3 SPECIFIC SECTORAL ISSUES

Now that we know the reasons for product decomposition, let us see how this affects model structure.

The most immediate change is of course the multiplication of categories. According to the item, the extension can apply to a different criterion, or even two criteria at the same time.

A precise list of the items belonging to each category will be provided as an annex.

First we can consider the **products**:

Product 1: agricultural products (such as non-processed food)

Product 2: manufactured goods (such as household appliances)

Product 3: services (such as transportation).

¹⁵⁹ However, as we are using deflators, this means the trade prices for these countries will be less sensitive to the cost of energy, which represents a lower share of the total cost. They will gain competitiveness if the price goes up, but lose if it goes down (as they do not profit from the shock).

This decomposition applies naturally to traded elements, such as consumption or exports.

But the same decomposition will apply to **productive units**, according to their output. Farms, industrial plants and travel agencies will be classified in each of our three categories, in order.

And here a distinction must be made, as a given unit can produce different goods. In that case:

- The whole unit will be classified in the **branch** of its main activity.
- It will be fractioned into products, each share being allocated to the **sector** it represents.

In our model, we shall not separate branches and sectors, and we shall use alternately the two terms. But branches and products can play different roles, and appear sometimes for the same variable, introducing a double indexation.

In addition, it might look interesting to proceed further in the decomposition, in two cases:

- For agriculture (and fishing) the model would benefit from a separation into artisanal and industrial units:
 - Artisanal units use less capital (machinery) and much more labor, with a lower productivity.
 - Artisanal employment will not be very sensitive to output (family units will often include a generally inactive work force, which can be called for if needed).
 - They also use less intermediate goods (such as fertilizer and fodder).
 - Their revenue is entirely appropriated by households.
 - The export a lower share of their production.

This is particularly interesting in developing countries, in which the role of agriculture is still important.

- The manufacturing sector includes energy and mining. Oil, gas and ore are traded through a process different from other manufacturing products:
 - The international price is fixed, and competitiveness plays no role (even for imports).
 - Capacity also plays no role in the short run: quantities exported are decided by the exporter, and imports are directly linked to demand¹⁶⁰.

This means we shall separate in our example:

- Agriculture and fishing production into artisanal and industrial
- Manufacturing exports and imports into energy and non-energy products. We will not include mining, even if it plays a high role in the economy of some countries (Chile, Morocco, Australia...).

This is an example of the (frequent) case where decompositions can lead to further decompositions:

- it is only for agriculture that artisanal/industrial decomposition is really efficient,

¹⁶⁰ But if local production goes down, imports can increase.

- And excluding energy from manufactured trade comes from the fact that in that regard, it behaves more like the primary product. This calls for a local correction of the general decomposition, if we want the model to present reliable trade equations.

10.3.4 THE DATA

For detailed data, very often the information gets scarcer. The following problems can occur:

- Detail on goods and services at current prices is known only from the production side, not the demand side.
- For the detailed series which are available, deflators are known only at a more aggregate level, making the computation of variables at constant prices approximate.
- Investment (and capital) is not available by product, or by investing branch, or (most often) in two-dimensional detail. The same problem can appear for the change in inventories.

Fortunately investment is made essentially in manufacturing goods and construction (plus livestock and plants for agriculture). If the model uses a global secondary product, the dimension of the products is not required¹⁶¹. In any case, many cells in the matrix will contain null values.

Intermediate consumption by product and use is available more frequently, but not always (this means the input-output matrix is not entirely known, perhaps only for some periods).

- Employment is not detailed by branch, or by categories of firm's ownership.
- The part of the account of firms giving transfers (wages, subsidies...) is not detailed by branch.
- The period for which data is known is shorter, or more difficult to collect. For instance, tables might be produced yearly for the current year.

10.3.5 CONSEQUENCES FOR ESTIMATIONS

If the sample gets too short, we can still try to estimate. But we should not rely on results, however favorable. We should just use them as another indication on the associated behavior, along with economic theory, observations on the way the local economy works, and estimations at the global level.

But very often the above problems will not allow to estimate individual equations at the detailed level.

One can:

- Calibrate the equations using theoretical coefficients.
- Use the results estimated at the aggregate level.

¹⁶¹ Although investment contains a share of services (producing plans for buildings, patents).

- Take the coefficients from other models, describing a similar country (Thailand for Vietnam for instance, or Austria for the Czech Republic). This option presents a surprising good point: the country used will generally be more advanced (as it has a better and more developed data system) which means that its estimations are based on a period which should represent the future of our modeled country, the actual period on which economic studies will be performed.

And actually, this difficulty has another silver lining: partially free from econometric restrictions, we can apply the formulations which conform the most to economic theory, in specification and value of coefficients. In particular, we can apply error-correction frameworks in each formula, leading to a model which will:

- Provide a long term solution on any future period, without any change to specifications (with a few restrictions on assumptions).
- Separate completely the long term specification from the dynamics leading to it.
- Allow to interpret completely and easily the coefficients (using values which conform to theory).

10.3.6 THE PRODUCTION FUNCTION

As we have stated above, we can use different solutions for each branch.

10.3.6.1 Primary product

This is a field is quite difficult to manage. Capacity can depend on:

- Capital in product 1: Land made ready for use, plantations, and cattle.
- Capital in product 2: Machinery, more or less sophisticated.
- Public capital: availability of water and electricity, road network.
- Services capital: transportation, storage.
- Intermediate inputs: fodder, fertilizer.
- Climatic conditions (exogenous of course).

Formulations can be more or less complex, from a simple production function (complementary or Cobb-Douglas) to a logical system including conditions and non-continuous functions. Output could also be decomposed into more categories (agriculture/fishing/forestry) or into products (rice and coffee for Vietnam, fruits for Central America...). In the last case, physical quantities could be used (tons, liters, numbers).

But basically, the short run definition of quantities produced should not depend so much on demand, but rather on potential production including climatic factors (in the long run of course, profitability and market size will affect the creation of capacities).

This has an important impact on model properties. As the supply – demand equilibrium still has to apply, we need a new balancing element, which can be imports, exports, demand or one of its components. The simplest solution is to use imports, which means that:

- Local demand has to be satisfied¹⁶²
- Exports are determined by world demand and competitiveness (with a limited impact as agricultural export prices are mostly determined at the world level).
- The share of global demand which cannot be satisfied locally has to be imported¹⁶³.

Other frameworks can be considered, such as:

- Local demand has to be satisfied
- Imports are controlled by local agents or the State.
- What is left from local production is exported.

Or:

- Exports are controlled by the state or local agents, according to market conditions.
- Local agents are allowed to buy the non-exported share of local production.
- Imports are also controlled.
- This gives total local demand.

It is quite possible to consider several of these frameworks simultaneously, but only if two or three categories are defined, each with its four elements which verify the particular equilibrium.

We can expect the consequences for model properties of the actual choice to be quite important.

10.3.6.2 Industrial product

One of the main purposes of the decomposition is to identify a sophisticated production function for the manufacturing sector, leaving out the other branches for which this choice is much less natural, and would have a negative impact on the quality of a global estimation.

A moderately complex option is the Cobb-Douglas function, which we presented earlier.

10.3.7 UNEMPLOYMENT

Unemployment follows the same logic as the simple model. However one could define different sensitivities to employment according to the branch. For instance in manufacturing, the higher share of qualified workers could lead job creation to call more for immediately productive unemployed.

10.3.8 CHANGE IN INVENTORIES

The new element is that we should develop the equation in two directions: the good in the inventory, and the branch using it. This means that in principle we will have at least 9 variables. However, services are not stored as inventories, and many agricultural products cannot be either.

¹⁶² But remember it is endogenous, so it follows global activity.

¹⁶³ But if the price of imported goods is too high, consumption will decrease and will also move to other goods.

To set the coefficients, we can consider theory, expert advice or the ratio of total inventory change to the change in value added.

10.3.9 HOUSEHOLD CONSUMPTION

The decomposition brings a new feature: consumption has to be divided into products.

We first compute global consumption according to the single product option.

Then we separate consumption into products, using an error correction framework, with a target sharing depending on the relative prices of consumption goods. After the decomposition has been applied, we need to control that the sum is consistent with the total (this can be done using a system or applying a correcting factor).

The situation is simpler if we consider two products, as we only need a single ratio, once global consumption is known.

10.3.10 EXTERNAL TRADE

This is the second field in which product decomposition allows to introduce differences in the formulas across products.

This originality will come from:

- Different weights: the share of exports in production, and imports in demand, is different from one product to the other.
- Different sensitivities: estimating the role of price competitiveness can give different answers; if the equation has to be calibrated, economic theory could lead to use different values.
- Different formulations: the role of tensions on capacities can be limited to the manufacturing product.

10.3.11 WAGES

For wages, we should use the same (theoretical) framework as in the simple model.

However, the indexation process is a little more complex. We still have a choice (in particular in the long run) between an indexation on the consumption price and the value added deflator. But if the first element is measured at the global level, the second corresponds to the branch. This means we can observe the consequences of dissymmetric shocks (such as an increase in subsidies to one sector), or shocks having different sectoral impacts (like an increase in the margins of exporting firms, mostly industrial).

Also, the sensitivity to unemployment can be quite different, leading to dissymmetric inflationary properties.

Finally, identifying the artisanal and industrial branches of agriculture will allow considering different productivities and wage levels.

10.3.12 PRICES

10.3.12.1 The value added deflator

With different production frameworks, the cost variable used in the long term target will change. And the role of the rate of use should be more intense in manufacturing.

10.3.12.2 The production price

This is a domain in which the complexity increases: the intermediate consumption of a branch is the sum of individual two-dimensional elements, each valued at its own price (which has to be defined).

10.3.12.3 The trade prices

Once production prices are known, we can move to the trade prices. To define the price of competitors, we need to identify an average foreign production price for each of our products, which will depend on the structure of traded goods, and of customers and suppliers.

10.3.12.4 The demand prices

The above deflators allow us to compute value added and trade at current prices. As we know already the values at constant prices, the demand deflators can be defined by identities.

However, one must check that, at current prices, the data still verifies the equilibrium between demand and supply. A small error can be accepted, and either

- Forgotten: this is dangerous as the model will never get a perfect historical fit, once estimated equations have been fed their residuals.
- Treated as a (multiplicative) residual. We have seen that it is a much better option.

10.3.12.5 The government budget

As in the simple model, we have to define a detailed government budget. The elements associated with goods (demand and taxes) will have to be separated. Others will not (transfers to households in particular).

10.3.12.6 The EViews program

We can now build a decomposed model, based on the same framework as our single product one, and using the same concepts.

We have decided the problem was too complex to be presented here. We shall only give a table presenting the decomposition level applied to these concepts. It can add no, one or two dimensions.

10.3.12.7 A list of elements

We present here a list of elements, coming actually from a 3-product operational model for Vietnam. It applies the artisanal – industrial decomposition of agriculture we presented earlier.

In the “categories” column,

T stands for total,

1, 2 and 3 for the associated products (or branches, or sectors),

1a and 1i for the decomposition of agriculture into artisanal and industrial.

2e and 2m for the decomposition of product 2 into energy and the rest (mostly associated to manufacturing).

name	type	definition	Units	Type	categories
CAP	Endog	Productive capacity	Prix constants	Produit	1,2,3
CI	Endog	Change in inventories	Prix constants	Produit	T, 1,1a,1i,2,3
COG	Exog	Consumption of Administrations	Prix constants	Produit	1,2,3
COGV	Endog	Consumption of Administrations	Prix courants	Produit	T
COH	Endog	Household consumption	Prix constants	Produit	T,1,2,3
COMPM	Endog	Imports price competitiveness (import / local production price)		Produit	1,2i,3
COMPX	Endog	Exports price competitiveness (export/ foreign production price)		Produit	1,2e,2i,3
COST	Endog	Cost of wages and capital	Prix constants	Branche	2
CPI	Endog	Consumption price Deflator	Déflateur année de base = 1	Produit	T
CRV	Endog	Export - import ratio at current prices	Ratio	Produit	T
CRVOL	Endog	Export - import ratio at constant prices	Ratio	Produit	T
ER	Endog	Exchange rate (value of the local currency)	Déflateur année de base = 1		T
EXPG	Endog	Expenditures of Administrations	Prix courants		T
FBAL	Endog	Financing Capacity of Firms	Prix courants	Produit	1i,2,3
FD	Endog	Final local demand	Prix constants	Produit	T,1,2,3
FDGV	Endog	Final demand of Administrations at current prices	Prix courants	Total	T
FDI	Endog	Foreign direct investment at constant prices	Prix courants	Produit	1,2,3
FDV	Endog	Final local demand	Prix courants	Produit	T,1,2,3
FOBAL	Endog	National commercial balance	Prix courants	Produit	T
GBAL	Endog	Financing Capacity of Administrations	Prix courants	Produit	T

GDPM	Endog	Market Gross Domestic Product	Prix constants	Produit	T,1,2,3
GDPMV	Endog	Market Gross Domestic Product at current prices	Prix courants	Produit	T,1,2,3
GDPV	Endog	Total Gross Domestic Product at current prices	Prix courants	Produit	T
HDI	Endog	Households disposable income	Prix courants	Produit	T
HI	Endog	Household revenue	Prix courants	Produit	T
HRDI	Endog	Household disposable income in purchasing power	Prix constants	Produit	T
IC	Endog	Intermediate consumption at constant prices	Prix constants	Produit x Branche	(1,2,3)*(1,2,3)
ICV	Endog	Intermediate consumption at current prices	Prix courants	Produit x Branche	(1,2,3)*(1,2,3)
IG	Exog	Government investment at constant prices	Prix constants	Produit	1,2,3
IGV	Endog	Government investment at current prices	Prix courants	Total	T,1,2,3
IHH	Endog	Housing households investment	Prix constants	Total	T,2,3
INCT	Endog	Income tax	Prix courants	Total	T
IP	Endog	Productive investment	Prix constants	Total	(1,2,3)*(T,1a,1i,1,2,3) ,T*T
IPD	Endog	Target productive investment	Prix constants	Produit Branche	by 2*2
IR	Endog	Current interest rate on new loans	Points		T
IRL	Endog	Long term interest rate on new loans	Points		T
IRM	Endog	Average interest rate on current debts	Points		T
IRS	Endog	Short term interest rate on new loans	Points		T
K	Endog	Productive capital	Prix constants	Produit x Branche	(1,2,3)*(T,1a,1i,1,2,3) ,T*T
KD	Endog	Target productive capital	Prix constants	Produit x Branche	2*2

KV	Endog	Productive capital at current prices	Prix courants	Branche	1a,1i,1,2,3
L	Endog	Total employment	Thsds persons	Branche	T,1a,1i,1,2,3
L_G	Exog	Employment	Thsds persons		
L_M	Endog	Employment	Thsds persons		
LD	Endog	Target Firms employment	Milliers de personnes	Branche	1a,1i,2,3
LF	Endog	Firms employment	Milliers de personnes	Branche	1a,1i,2,3
LH	Exog	Household productive employment	Milliers de personnes	Branche	1a,1i,2,3
LP	Endog	Labour productivity	Milliers de personnes	Branche	T,1,2,3
LPTA	Exog	Labour productivity trend	Milliers de personnes	Branche	1a,1i,3
LS	Endog	Wage earners	Milliers de personnes	Total	1,1a,1i,2,3
LS_F	Endog	Wage earners firms	Milliers de personnes	Total	T,1a,1i,2,3
LS_H	Endog	Wage earners households	Milliers de personnes	Total	T,1a,1i,2,3
M	Endog	Imports	Prix constants	Total	T,1,2e,2i,3
MARG	Endog	Margins of firms	Prix courants	Branche	1,1a,1i,2,3
MV	Endog	Imports at current prices	Prix courants	Total	T,1,2e,2i,3
NIF	Endog	Net interests paid by Firms	Prix courants	Branche	1i,2,3
NIG	Endog	Net interests paid by Government	Prix courants		
NIX	Endog	Net interests paid to the Rest of the World	Prix courants		
OIT	Endog	Other indirect taxes	Prix courants	Produit	T,1,1a,1i,1,2,3
PCOG	Endog	Government consumption price Deflator	Déflateur année de base = 1	Produit	1,2,3
PENS	Endog	Pensions	Prix courants		

PENSR	Exog	Pensions rate	Prix constants		
PEX	Endog	Export price Deflator	Déflateur année de base = 1	Produit	T,1,2,2e,2i,3
PEXT	Endog	Export price Deflator including tariffs	Déflateur année de base = 1, USD	Produit	1,2e,2i,3
PFD	Endog	Local demand price Deflator	Déflateur année de base = 1	Produit	T,1,2,3
PFDG	Endog	Government demand price Deflator	Déflateur année de base = 1	Produit	T
PFDXT	Endog	Local demand price Deflator excluding VAT	Déflateur année de base = 1	Produit	1,2,3
PGDPM	Endog	Market GDP price Deflator	Déflateur année de base = 1	Produit	T
PIC	Endog	Price of intermediate consumption	Déflateur année de base = 1	Produit x Branche	(1,2,3)*(1,2,3)
PIG	Endog	Government investment price Deflator	Déflateur année de base = 1	Produit	2,3
PIHH	Endog	Household housing investment price Deflator	Déflateur année de base = 1	Produit	T
PIM	Endog	Import price Deflator	Déflateur année de base = 1	Produit	T,1,2e,2i,3
PIMT	Endog	Import price Deflator including tariffs	Déflateur année de base = 1	Produit	1,2e,2i,3
PIMU	Exog	Energy import price in US Dollars	Déflateur année de base = 1	Produit	2e
PIP	Endog	Productive investment price Deflator	Déflateur année de base = 1	Produit	T,1,2,3
POP	Exog	Population	Milliers de personnes		
POPAG	Exog	Population in age of working	Milliers de personnes		
PP	Endog	Production price Deflator	Déflateur année de base = 1	Produit	1,2,3
PQ	Endog	Value added price Deflator	Déflateur année de base = 1	Produit	1,2,3
PROF	Endog	Profits of Firms	Prix courants	Branche	1i,2,3
PT	Endog	Tax on profits	Prix courants	Branche	T,1i,2,3
PWM	Exog	World price imports weighting	Déflateur année de base = 1, USD	Produit	2i,3

PWX	Endog	World price exports weighting	Déflateur année de base = 1	Produit	T,1,2e,2i,3
Q	Endog	Value added	Prix constants	Produit	T,1,1a,1i,2,3
QA	Endog	Target value added	Prix constants	Produit	2
QV	Endog	Value added at current prices	Prix courants	Produit	T,1,1a,1i,2,3
R_COGV	Exog	Ratio of non-wage Govt consumption to GDP			
R_CPI	Exog	Ratio of CPI to Final demand deflator			
R_EXPG	Exog	Ratio of non-identified Administrations demand to GDP			
R_IHH	Exog	Ratio of Housing investment to household revenue		Branche	2,3
R_INCT	Exog	Income tax rate			
R_IP	Exog	Ratio of productive investment to GDP		Produit x Branche	3*(1,1a,1i,2,3)
R_LFA	Exog	Share of Firms employment in total		Branche	1a,1i,2,3
R_OIT	Endog	Rate of Other Indirect taxes (than VAT)		Branche	1,1a,1i,2,3
R_PCOG	Exog	Ratio of Govt consumption to final demand deflator		Branche	1,2,3
R_PFDG	Exog	Ratio of Govt demand to final demand deflator			
R_PIC	Exog	Ratio of Interm. consumption to Demand (excluding taxes) deflator		Produit x Branche	(1,2,3)*(1,2,3)
R_PIG	Exog	Ratio of Govt investment to Final Demand deflator		Produit	2,3
R_PIHH	Exog	Ratio of Housing investment to final demand deflator			
R_PIP	Exog	Ratio of Productive investment to final demand deflator		Produit	1,2,3
R_PT1	Exog	Rate of the tax on profits		Branche	1i,2,3
R_REVG	Exog	Ratio of other Administrations revenue to GDP			
R_REVQ	Exog	Ratio to production of other Household revenue from production		Branche	1i,2,3

R_REVX	Exog	Non wage non production Household revenue, in purchasing power			
R_SCF	Endog	Rate of social contributions by Firms		Branche	1,1a,1i,2,3
R_SCF_G	Exog	Rate of social contributions by Government			
R_SCF_H	Exog	Rate of social contributions by Households			
R_SCW_G	Exog	Rate of social contributions by wage earners		Government	
R_SCWA	Exog	Rate of social contributions by wage earners		Branche	1a,1i,2,3
R_SUBS	Exog	Rate of subsidies (to value added at current prices)		Branche	1,1a,1i,2,3
R_TARIFF	Exog	Rate of Local tariffs		Branche	1,2,2e,2i,3
R_TARX	Exog	Rate of Foreign tariffs		Branche	1,2e,2i,3
R_UN	Endog	Unemployment rate			
R_VAT	Exog	VAT rate		Produit	1,2,3
R_WG	Exog	Ratio of Government to market wage rate			
R1_A	Exog	Share of artisanal in Agriculture			
R1_I	Exog	Share of industry in Agriculture			
RDEP	Exog	Depreciation rate of capital		Produit x Branche	(1,2,3)*(1a,1i,2,3)
RELC	Endog	Labour - Capital relative cost		Branche	2
REVG	Endog	Revenue of Administrations			Eq364
REVQ	Endog	Revenue of Households from production, other than wages.		Branche	T,1i,2,3
REVX	Endog	Non wage non production Household revenue	Prix constants		
RFDI	Exog	Share of FDI in investment		Produit	1,2,3
RMARG	Endog	Margins ratio		Branche	1,1a,1i,2,3

RPROB	Endog	Margins ratio including capital amortization		Branche	1i,2,3
RPRO	Endog	Profits ratio		Branche	1i,2,3
RS_F	Exog	Share of wage earners in Firms		Branche	1a,1i,2,3
RS_H	Exog	Share of wage earners in Households		Branche	1a,1i,2,3
SCF	Endog	Social contributions by Firms	Prix courants	Branche	T,1,1a,1i,2,3
SCF_G	Endog	Social contributions by Government	Prix courants		
SCF_H	Endog	Social contributions by Households	Prix courants		
SCW	Endog	Social contributions by Wage earners	Prix courants		
SOCB	Endog	Social benefits	Prix courants		
SOCBR	Exog	Social benefits in purchasing power	Prix constants		
SUBS	Endog	Subsidies to firms	Prix courants	Branche	T,1,1a,1i,2,3
TARIFF	Endog	Local tariffs	Prix courants	Produit	T,1,2,2e,2i,3
TC	Exog	Technical coefficient		Produit x Branche	(1,2,3)*(1,2,3)
TRBAL	Endog	Trade balance	Prix courants	Produit	T
TTRAD	Endog	Terms of trade (export - import price ratio)	Milliers de personnes		
UN	Endog	Unemployment	Milliers de personnes		
UN_R	Exog	Unemployment	Milliers de personnes		
UN_U	Endog	Unemployment			
UR	Endog	Rate of use of capacities		Branche	1,2,3
VAT	Endog	Value Added Tax	Prix courants	Total	T,1,2,3
W	Endog	Market wage rate	Milliers aux prx courants	Branche	1,1a,1i,2,3

W_G	Endog	Government wage rate	Milliers aux prx courants		
WAGE	Endog	Total wages	Prix courants	Branche	T,1,1a,1i,2,3
WAGE_F	Endog	Wages paid by Firms	Prix courants	Total	T,2,3
WAGE_G	Endog	Wages paid by Administrations	Prix courants		T
WAGE_H	Endog	Wages paid by Households	Prix courants	Total	T,2,3
WCOST	Endog	Unitary wage cost	Milliers aux prx courants	Branche	1,2,3
WD	Exog	World demand at constant prices	Indice à prix constants	Produit	1,2e,2i,3
WF	Endog	Wage rate paid by Firms	Milliers aux prix courants	Branche	2,3
WH	Endog	Wage rate paid by Households	Milliers aux prx courants	Branche	2,3
X	Endog	Exports	Prix constants	Total	T,1,2,2e,2i,3
XV	Endog	Exports at current prices	Prix courants	Total	T,1,2,2e,2i,3

10.4 A MULTI COUNTRY, SINGLE PRODUCT MODEL

We shall now address the issue of building, maintaining and using a model describing the economy of more than one country. We shall also consider the similar issue of a single country but multi-region model.

10.4.1 FIRST ISSUE: THE MODELS

To produce a multi – country model, we first need single country ones. Two issues have to be considered:

- The degree of similitude of the country models
- The description of the Rest of the World

10.4.1.1 The single country models

We shall suppose the country models follow the lines used above. What matters now is the differences allowed in the specifications, from one country to the other. Several cases can be considered:

A - The models are allowed to be completely different. This even applies the concepts used, such as the product decomposition. The only condition is that, one identified the elements linking the models (basically the trade variables at constant and current prices) they can be converted in such a way that they can be transferred as input to the other models. But for instance a single product and multi product models can coexist. The transfer process will apply aggregation or disaggregation as required. In this case the names of the variables will probably be different too.

B - The models use the same concepts but the economic framework is different. For instance a model can use a Cobb-Douglas production function, the other a CES.

C - The models use the same equations but the coefficients are different. Some exceptions can appear, for instance an additional lagged variable, but the logic remains the same. In general these differences should be due to estimation, which means that a common data set must be available.

D - The models use the same equations and the same coefficients. This means that either the formulas have been estimated as a panel (by polling all the countries data together) or calibrated.

The choice of option depends on the case:

- Option A can be considered if the models already exist, which simplifies the process. The case is even stronger if the models have to keep on living as independent versions.

This is the case for instance for the LINK project, which links into a multi-country version, under the management of UN and the University of Toronto, models maintained as independent forecasting tools by country institutes.

Another example is the grafting of existing a national model in an international environment, also existing. For example, the MODUX model managed by the Institute statistque Luxembourg (STATEC) has been integrated into

the environment MacSim (see below) in order to enrich its projections and especially its impact analyzes. Such a possibility is provided by the managers of MacSim (which basically means the author of this book).

The drawback is of course that the models can follow different economic properties, and the differences in mechanisms, leading to different sensitivities to shocks, are not necessarily justified by reliable country specifics (if they are, then option A is clearly optimal). As stated earlier, econometrics based on the same data can justify formulations with quite different properties, even by pure chance (two economists with the same economic philosophy can reach different formulations through the meanders of successive estimations).

- Option B is option A with no initial models. Then the models will be easier to manage, but the above problems remain.
- Option C is even simpler to manage and its properties easier to interpret. The only drawback is the constraints on model formulations, which forbids introducing country-specific formulations. But of course, some minimal differences can be accepted, such as a longer lagged structure or the long run indexation of wages on different deflators.
- Option D limits even more country specifics. What only remains is the effect of the size of the country, the structure of its partners (imports and exports) and structural parameters like labor and capital productivity, the weight of the public sector, the tax structure, the sharing of production revenue between the firms and the workers.

The only justification is just this: that the only difference in properties comes from undisputable elements, the consequences of which they describe faithfully.

From the above elements we can derive the following suggestion: if the models do not preexist, option C is clearly the best, with some elements from option B if they are really justified, in particular by the observation of country characteristics.

For instance for an oil producing country the trade in energy elements could be individualized. For the interest rate, identifying the policy of the central bank should lead to the introduction of the associated rule. And maybe if there is sufficient proof that factor substitution follows a zero or unitary elasticity, the associated production function (complementary factors or Cobb-Douglas) should be used.

10.4.2 SECOND ISSUE: INTER-COUNTRY TRADE

Once the single country models have been defined, they have to be merged into a single entity.

The main issue is to make consistent the imports and exports of each country, at constant and current prices.

The obvious option is to associate behaviors with the actual decision processes:

- Imports are decided by the importing countries, and exports adapt to them.
- Export prices are decided by the exporters, and import prices are computed by identities, taking into account the relative exchange rate.

Once imports of a country are decided, they must be shared between the countries which export to it. There are two options:

- Identifying the individual imports by their source, and converting them into individual exports from the supplying country. This is obviously the most logical solution. It has the essential advantage of making exports and imports consistent at the world level.
- However, some models use a more devious technique: they compute global potential exports of a country by weighting by constant shares the imports of its partners.

Of course total exports and imports will have different values, and some correction must be made. The main advantage of this method is to allow estimation of exports based on actual global values. Also, it avoids identifying a large number of trade flows (growing as the square of the number of countries).

We favor strongly the first option.

- It identifies more variables, but the associated information is important. The interpretation of the consequences of shocks (on domestic or trade related variables) is much clearer and more informative.
- It does not call for corrections, giving automatically consistent results.
- It does not require more information: the weights used by the other method can be used to compute the trade flows.
- But the most important in our sense is that it locates the decision process where it belongs: exports are not decided by exporters, but by importing countries which, once they have decided to import, choose among the potential suppliers, essentially by considering relative prices.

Of course, it does not allow global estimation. The individual trade flow equations will have to be either calibrated (perhaps using values from global estimations) or estimated using panel techniques.

From now on, we shall concentrate on this technique.

10.4.3 A CONSISTENT METHOD: MACSIM

We shall present now a consistent method for defining in sequence the above elements. It is used by the MacSim system (2001, 2012).

We shall start, not with imports but with the price system, as prices depend only on local elements, while imports depend on price competitiveness.

10.4.3.1 The price system

We start by establishing a coherent system for trade prices. For this we shall start from the export price formulations (estimated or set). For clarity we shall use calibrated equations, but the method applies to estimated ones also.

We shall suppose that the traditional equation applies to each client country. But this time the exports destination will be identified.

We shall consider i as the exporting country, j as the destination.

To get individual export prices, we correct by the exchange rate:

$$\text{Log}(pex_{i,j}) = a \cdot \text{Log}(pp_i) + (1-a) \cdot (\text{Log}(pp_j ch_i / ch_j) + e \cdot t + f)$$

Where i is the exporter and j the client, and ch_i is an index, representing the evolution (from the base year) of the value of a common currency (actually the US dollar) compared to the currency of country i . If the currency of i depreciates, the index grows.

$$pim_{j,i} = pex_{i,j} ch_j / ch_i$$

and we get the global import price through an average:

$$pim_j = \frac{\sum_i m_{j,i} pim_{j,i}}{\sum_i m_{j,i}}$$

(we shall see later how we compute the trade flows at constant prices).

10.4.3.2 The foreign rates of use

In the single country models, we have seen the fundamental role of the rate of use of capacities for the local country. But we supposed the foreign capacities to be infinite.

Now we can consider, to define the imports of a country, the rate of use of providing countries. If the rate of use increases in Germany, it will lose market shares in France, relative to other exporters but also to French manufacturers.

To get the average, we shall use the same method as above (without currency corrections).

$$urx_j = \frac{\sum_i m_{j,i} ur_i}{\sum_i m_{j,i}}$$

10.4.3.3 The global imports

We can now determine the global imports of country i , by modifying slightly the equation from our single country model, to take into account the capacity of exporters:

$$\text{Log}(m_i / td_i) = b \cdot [\text{Log}(ur_i) - c \cdot \text{Log}(urx_i)] + d \cdot \text{Log}(pm_i / pp_i) + e$$

To simplify the explanation, we have presented only the long term part. The dynamic equation should contain the same change.

This means that, as in the single country models, a general decrease in the available capacity of exporters will reduce their exports, through a substitution effect. The coefficient c , lower than one, takes into account the larger size of this set of countries.

10.4.3.4 The trade flows

Finally, we have to separate imports into individual exports. Once again, we shall take into account relative competitiveness, and fluctuations in available capacities, relative to the above average. Actually, rather than estimating an equation (we do not have the associated data) we shall use:

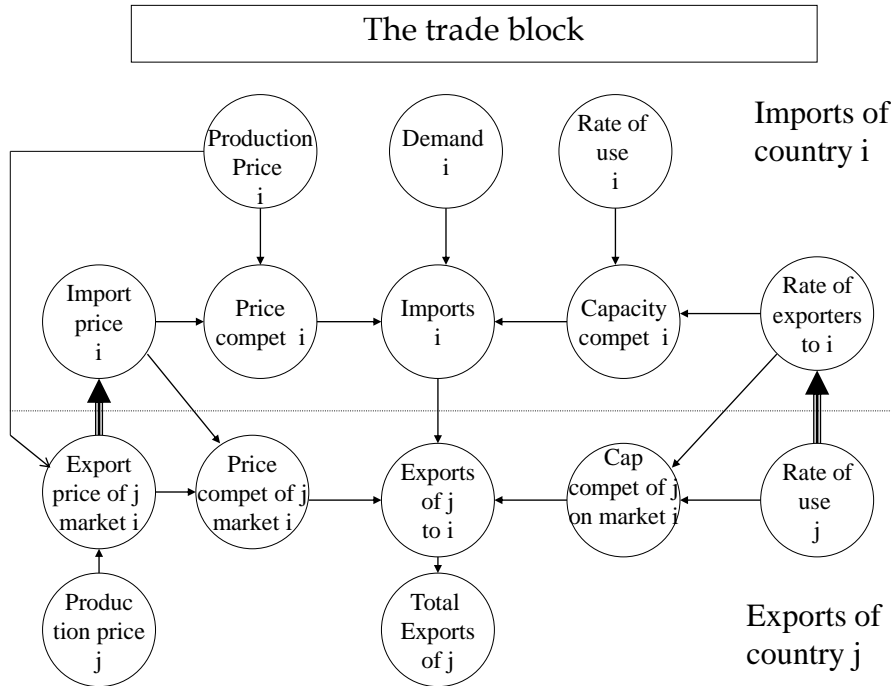
$$m_{i,j} = m_i \cdot b_{i,j} \left[1 - \alpha_i (pex_j ch_i / ch_j - pim_i) - \beta_i (ut_j - utx_i) \right]$$

which means that (as for the single country models) exporters to one country will increase a « natural » share with competitiveness and available capacity, this time relative to their competitors. More precisely, the A and B matrixes (one derived from the other) represent the sharing of bilateral trade associated with identical rates of use and competitiveness for all countries (not necessarily the most natural assumption, nor the present situation). Scalars α and β are fixed, and will be set at a higher value than for the import equation (between 0.5 and 1.0). This should represent the higher versatility of the choice of its providers by a country, once the decision has been made not to buy local products.

One will observe that this technique guarantees the identity of the sum of individual exports with its global value, without any correction.

Of course, the coefficients can be different from one market to another, but not within one market.

The system can be summarized with this graph:



In addition to the above, we can now introduce accounting equations:

Exports from country i to j:

$$x_{i,j} = m_{j,i} \cdot \text{usd}0_j / \text{usd}0_i$$

Where $\text{usd}0_i$ represents the base year value of the currency of country i, in US Dollars.

Total exports are computed as a sum

$$x_i = \sum x_{i,j}$$

The average export price of country i:

$$pex_i = \sum pex_{i,j} \cdot x_{i,j} / x_i$$

10.4.3.5 Third issue: common behaviors

But we have also to consider behaviors common to a set of countries, which take into account global variables. The most obvious case is the presence of a global formal agreement, such as the use of a common currency by the European Monetary Union.

This option is very easy to implement: one just has to create a new “country” using aggregated concepts, considering only the relevant variables.

10.4.3.6 The Rest of the World

If the scope of the countries described is large enough, we cannot expect that shocks applied to a large subset or the whole model will have no feedback through the economy of the remaining countries. This is true in our MacSim system, which models most of Europe (the Euro zone + the United Kingdom) and also the US and Japan. Moreover, with the technique we are proposing (computing the shares of each country in each country’s imports) endogenizing the Rest of the World allows its exports to conform to the global framework, avoiding any unbalance.

But of course, we are not going to estimate a ROW model: we do not have the data, and the heterogeneity is too high.

What we propose is:

- To define the trade elements as usual: imports are shared according to the above framework, and exports participate to the above competition.
- In addition, to create a set of multipliers, defining the feedback of the Rest of the World exports on its GDP and imports, and linking its imports to its final demand.

It should be quantified to give a response similar to the single country models, with the exception of infinite (or quasi-infinite) capacity and high sluggishness of prices.

This can be done quite simply.

10.4.3.7 The sequence of production

To produce a multi-country model, one should proceed through the following sequence:

10.4.3.7.1 Deciding on the model framework

We shall suppose that we build our model from scratch, which in this case should lead to a set of country models using the same economic framework but estimated separately, with limited variations in the estimated formulas.

In this case one should consider:

- The data available.
- The type of behaviors the model should describe.
- The expected degree of detail.

This should allow deciding:

- The variables used by the model (they have to be useful and available).
- The full form of the identities.
- The type of behaviors which should be estimated, including the explanatory elements.

10.4.3.7.2 Producing a set of country models

Then one should build a set of country models, using the technique described earlier. Of course, a global data set has to be produced, containing series for each single model, with the same name except for an identifier. For instance French exports can be called FR_X.

As usual, concerning behavioral equations, one should start with declaration of intent before moving to estimations.

Several sequences can be considered. One can build models one by one, or apply simultaneously each phase (data production, framework, estimation, testing) to the whole set. Probably the best solution is to start with one country, go through the whole process, and apply it to the set of remaining countries, this time phase by phase simultaneously.

This allows measuring the difficulties, to solve the general problem on a single case, and define at a limited cost the full procedure, which will then be applied to the global case. In particular, if problems were to be met at a later stage, this would probably call for reconsidering the initial ones. Having to modify the full set of programs can then become quite costly.

The choice is not so simple, however, as:

- Running estimations for the same equations on a set of countries can be organized quite easily using EViews loops.
- But the estimation processes must succeed together, as the equations are not allowed to differ too much. An interesting direction found for a given country, but not associated to its specific features, can be applied only if it works also for other countries.

This means more iterations will be needed than for a single country model. The only issue is to limit their number.

10.4.3.7.3 Assembling the models

Once satisfying single country versions have been produced, assembling them is essentially technical.

10.4.3.7.3.1 COMPUTING THE TRADE FLOWS

If formalizing the trade between countries uses explicit individual flows (our preferred option) one needs a matrix of transfers. It can represent either the share of exporters in each country's imports, or the share of clients in each country's exports. Both matrixes are needed, but one can be built from the other in an exact way. The matrix elements should be considered as series. If it is only available for some periods, it should be extrapolated, but in any case, the modeler should be allowed to make it change with time in the future. This is particularly essential for transition economies: the share of Russia in Polish exports has decreased in the last decades, but it is not stabilized and could increase again in the future.

From these matrixes the individual trade flows can be computed. If the matrix is known only at current (or constant) prices, a global assumption has to be made, the simplest being that the same export (or import) price applies to one country whatever its destination (or origin).

Of course adding up the result in the other dimension does not give the actual value. For instance if we start from imports and share them between exporters, the sum of individual exports of one country will show a difference with the national accounts. The issue is the size of this difference: it should not be higher than the residual of the equation if it had been estimated, otherwise a correction could be made (but this would bring back an inconsistency at the world level).

Of course, the forecast results will be consistent, as the model system is by construction. And the consequences of shocks too, as the difference between two consistent systems. We are using the error observed on the past as an indication of the global precision of the model on the future, not of its internal consistency.

10.4.3.7.3.2 FORMALIZING THE TRADE FLOWS AT CONSTANT PRICES

We can do it using the framework defined earlier. We have only to consider the following technical issues:

- The export equations of single models will disappear.
- The trade flows have two dimensions, and call for an additional loop.
- If the trade flows are not available as series, the coefficients will have to be calibrated. They must be common for a given market, and we cannot expect the choice between import origins to use the same price elasticity as the choice between imports and local products.

One advantage of our proposal is that trade flows will add up to imports without any correction, as the weighted differences to the weighted mean cancel out naturally.

10.4.3.7.3.3 FORMALIZING THE TRADE DEFLATORS

If imports are decided by importers, the transaction price is decided by the exporter.

This time we can estimate the global export price of each country (depending on local costs and the price applied by competitors). Then we can apply the results to each export market, using as competitors' price the one specific of this market.

Apart from this, we have to suppose that for each exporter the same formula applies to all its clients.

10.4.3.7.3.4 INTRODUCING IDENTITIES

Finally we need to compute identities, like imports and exports at current prices.

10.4.3.7.3.5 PRODUCING COMMON ELEMENTS

The other issue is the presence of an agreement between countries, defining a variable common to each of them, based on one or several common elements.

For instance, under purchasing power parity, the exchange rate of the Euro will apply to all Euro zone countries, based on Euro zone inflation.

Equations will have to be built:

- Summarizing the explanatory indicators.
- Computing the common element, generally giving a choice between several options (real interest rate or Taylor rule, fixed real or nominal exchange rate).

The formula must allow changing the list of countries concerned. This is very easy:

- For computing the indicators, the element of each country will be multiplied by an indicator (1=in, 0=out) to indicate if it participates in the weighted sum.
- For the use of the common result, an option must be present in each country's equation, using the same indicator (if 1, use the common value, if 0, use a country specific formula).

Even in the absence of agreement, it can be useful to compute common indicators (for instance the GDP of the global European Union, Euro zone or not).

10.4.3.7.3.6 THE REST OF THE WORLD

The Rest of the World "model" will need the usual trade flows equations, and in addition:

- An equation linking GDP and local demand to exports.
- A global imports equation linked to RoW demand and price competitiveness (its prices can be affected by the other countries inflation). Introducing an influence of the rate of use should depend on the size of the ROW.

10.4.3.7.3.7 SOLVING THE MODEL

The procedure for solving the model is the same. The number of equations is just larger, and the testing procedure more complex.

Another problem is the convergence: with less exogenous variables the solution is less anchored on assumptions, and the probability of divergence might increase. But one can also consider that an individual problem can become diluted in the global equilibrating process. Nothing is clear on that issue.

The most important issue is perhaps the price system: with less anchoring, even if each country's inflation converges in the long run to a common value, this can happen at highly heterogeneous levels. If a country's price indexes are higher than the others, the inflation created normally by a demand shock can be offset by the increase in the share of its lower priced imports. The overall effect can be deflationary, introducing properties which should be considered abnormal, especially as they are due to the distance to the base year.

This can be solved by managing the system of residuals in such a way that the long term prices remain rather close to each other. The targeting technique we will present soon can be used here¹⁶⁴.

10.4.3.7.3.8 USING THE MODEL

Of course if this model is more expensive to produce, it allows also a wider range of studies, and makes them more reliable compared to the single country case. As we have said earlier, the demand multiplier is 20% higher if we run a French model in a multi-country context, compared to a stand-alone use of the same model.

In addition to a better study of changes in the local assumptions of a single country, for this country and abroad, this type of model allows in particular to study consequences of:

- Global events such as a slowing down of the world growth, or a change in the price of oil, taking into account its consequences for all countries. We no longer need to make assumptions consistent for the rest of the world.
- Global trade agreements.
- Global policies (for instance a reduction of VAT in the European Union).

These elements can be assessed as a whole, and by comparing their consequences for each country.

All this can be done taking into account the role of common rules, such as the establishment of the Euro Zone, which can be changed at will.

10.4.4 THE EViews PROGRAM

The sequence of EViews programs is too complicated and too problem-specific to be presented here. However, a model is working, and is used in operational studies. Its elements are available, and can be provided through a special agreement.

10.5 A REGIONAL MODEL

Building a regional model presents the same features as in the multi-country case, with some differences:

- There is no currency issue, nor any difference in the definition of monetary rules. But the level of interest rates can vary.
- Variables such as various deflators are not independent from one region to the other.
- Variables such as unemployment have to be computed at the country level, or at least regional variables will influence strongly other regions, especially if we consider migrations.
- “Importing” from other regions is much easier. This means in our sense that instead of a sequence: definition of imports then separation into providers, all forms of demand from one region should be

¹⁶⁴ We use it quite often in practice.

separated immediately into providers, the region itself being treated like the others. But of course this region must be given a higher potential share, at the same level of competitiveness. Some products will have to come from the region, in particular services.

- But this information has less chance of being available.
- Migrations are much easier, and should be taken into account, through both their determinants and their effects. Transfer of household revenue is also easier and more natural.
- One should also consider the difficulty in transporting goods from one region to the other.

All these elements have been taken into account in one of our projects, presented as an example in the annex. We shall not however provide the associated programs, due to the cost of documenting them.

10.6 A MULTI COUNTRY, MULTI PRODUCT MODEL

Basically, the issues combine the features (and the difficulties) of both dimensions: countries and products.

- In addition to combining the features described earlier, the only problem lies in our opinion in the much higher complexity of the problem. Even if the EViews programs can be compacted through the use of loops, the actual number of variables is much higher, which means that the control of model properties will be more complex. In practice it will have to be conducted first at a global level, then as a detailed one.
- Also, the assumptions for forecasts and response to shocks will be more difficult to produce, especially if one wants to profit from the additional detail (which should be expected).
- Also, the requirements in terms of data are of course more important.

10.6.1 THE EIEWS PROGRAM

Although such a program has actually been created, its size is too important to present if here.

The technology is not too complex, as it just combines the elements from its two dimensions.

ANNEX 1: A SIMPLE METHOD ALLOWING TO SWITCH THE ENDOGENOUS/EXOGENOUS STATUS OF MODEL VARIABLES

We shall present here a method allowing making a model reach given values for a set of endogenous variables, by freeing the values of an equivalent set of exogenous elements.

This method can treat single or multi period simulations, and rational expectations can be considered at a small price in complexity (but a higher one in computation time, especially in the last case)

At first, we shall consider solving a backward looking model for a single period.

11.1 THE PROBLEM

Let us consider the model

$$y = f(y, x)$$

with:

y vector of endogenous with dimension m

x vector of endogenous with dimension n

We have suppressed the time index, as past (and future) values do not change in the solving process.

Let us suppose we want to get:

$$y^s = y^{s*}$$

For a subset of y of dimension p

through the endogenization of a subset of x with the same dimension, called x^s.

11.2 THE LINEAR CASE

If f was a linear function:

$$y = A \cdot y + B \cdot x + b$$

we would have

$$y = (I - A)^{-1}(B \cdot x + b)$$

Let us call C the sub-matrix $(I - A)^{-1} \cdot B$ (dimension $(p \times n)$ limited to the lines of y^s et C^s the sub-matrix $(p \times p)$ of C associated to the elements of x^s .

We have:

$$y^{s*} = C^s \cdot (x^{s*} - x^{s0}) + C \cdot x^0 + c$$

The values allowing reaching y^{s*} are obtained immediately through:

$$x^{s*} = x^{s0} + (C^s)^{-1}(y^{s*} - C \cdot x^0 - c)$$

$$x^{s*} = x^{s0} + (C^s)^{-1}(y^{s*} - y^{s0})$$

The C matrixes can be obtained easily by a matrix transformation, once A, B and b are known.

11.3 THE NON-LINEAR CASE

Let us now consider the non-linear case. We still have:

$$y = f(y, x)$$

and we still want to get:

$$y^s = y^{s*}$$

We shall start by supposing that the model is quasi-linear, and apply the above method. The C matrix becomes:

$$\partial g(y) / \partial g(x)$$

Where g represents the function $y = g(x)$ where the y vector has been eliminated from the right term.

It is unfortunately clear that identifying the g function is difficult if not impossible in practice. On the other hand, the values of the Jacobian are easy to obtain, or its approximation through finite differences.

One just has to solve the model for starting values of x , apply in sequence small shocks to each of the exogenous, and compute the associated solution. We get:

$$\partial g(y_i) / \partial g(x_j) \approx (g_i(x + \Delta x_j) - g_i(x)) / \Delta x_j$$

where Δx_j is a vector of dimension n with null values except for x_j , changed by Δx_j .

To get the equivalent of matrix C , we just have to select in the Jacobian the columns and lines associated with x^s and y^s .

The solution uses the same formula as the linear case:

$$x^{s*} = x^{s0} + (C^s)^{-1}(y^{s*} - y^{s0})$$

One shall notice that solving the model and computing the changes are only required for the exogenous concerned, and the inversion applies to a much reduced ($p \times p$) matrix.

The only difference with the linear case is that the solution is not exact, due to the linear approximation. But one just has to extend the process, starting from the new values obtained for x (actually limited to x^s). The process remains exactly the same, and can be repeated as many times as necessary. And as the linear approximation becomes more and more accurate (in term of properties) when we get closer to the solution (especially if the model is locally convex), the share of error eliminated each time becomes higher and higher, and convergence should be reached after a few iterations.

We identify here a Newton type method, which is characterized by a supralinear convergence property. An interesting feature is that we know the actual precision of the solution, contrary to other algorithms which consider only the speed of convergence.

11.4 A FEW QUESTIONINGS

Now, several questions appear:

The method we suggest looks quite simple for a package which does not provide an automatic process (such as TROLL), and represents an efficient alternative to more complex (and less efficient) tools. One can question why it is not proposed right away by packages, and applied more often by modellers (unless we are not up to date as to the present fashions).

Several answers can be considered:

- It does not work in practice, or only in very simple (non-operational) cases. We shall prove it works, at least for simple but operational models. There is no reason (but it remains to be proved) that it does not work also for the majority of reasonable cases.
- It does not work always. But then we can at least try, and maybe improve it in “difficult” cases. We can expect two main cases for failure:
 - The model is highly nonlinear. This is one of the problems with the Newton method, which can lead the solution to absurd values if the Jacobian computed is very far from its “exact” value. It is essential to start from values rather close to the solution, which should be the case in operational cases, if the forecasting quality of the model is high enough to provide spontaneously reasonable, if not acceptable, solutions.
 - The Jacobian is highly collinear, and cannot be inverted, or suggests once inverted very high changes in the exogenous (at least two of them). This case is less problematic, as another method (« by hand » included) will face the same problems, and here at least the problems will be evidenced immediately, and maybe interpreted through the observation of the matrix. The time gained can be used to look in other directions.
- It does not apply to the problem. Indeed the method is limited to reaching of a given set of n endogenous values by freeing n exogenous elements. A professional modeller might want to spread the charge of the process on more exogenous elements. But if the weight of the contributions is known, the method can still be applied: several exogenous will be affected, but the algorithm looks only for one value. For instance if the rates of social contributions by employers and workers are used simultaneously with the same variation, this variation becomes the unique endogenous element.

More problematic is the case of limits on changes, set by the user¹⁶⁵: a given exogenous is used until it reaches a limit, in which case the computation switches to a new element. This case can also be treated automatically, but the program become more complex, and the solution is not guaranteed within the acceptable range (but neither “by hand” anyway).

- Finally if the method supposes backward looking processes, it can also be applied to forward looking models. Two techniques can be considered, associated to the current methods for the simple case:
 - Fair Taylor: we iterate on the period by period solution, initializing future values by thus of the last iteration (at the beginning, the historical, theoretical or previously computed values).
 - Lafarge: we solve the global models where equations are repeated along the time dimension.

¹⁶⁵ Mostly for sociological reasons : it not reasonable to expect an increase in the income tax by 20% to accepted by households.

But now we have to consider two nested loops, and the process can become unacceptably slow. Specific improvements could be found.

11.5 GENERAL CONCLUSION

Of two things one: there are good reasons not to use this method, and it would be interesting to know them. Or there are not, and beyond the questioning of reasons for the present situation, a set of fully fledged tests should be conducted in my sense.

11.6 A TEST ON A VERY SMALL MODEL.

We shall present here the conditions and results of a test conducted on a very small model, called pip, with 17 equations (5 estimated) used by Jean Louis Brillet for his course in modeling techniques.

We have supposed a change in GDP (called Q) and imports at constant prices (M) by respectively +5% and -5% compared for the historical values for the same period. The change in values goes much beyond usual cases, and their direction is rather opposed: an increase in GDP generates imports, even if all things being equal an increase in imports reduces GDP.

We shall use as targeting variables the real wage rate WR and world demand WD. These elements are rather independent, first from each other (this should increase their efficiency and accuracy, as a team with different skills), and also from the targeted elements.

We will give:

- The EViews program, clearly short in statements (but not in comments). It can apply to any model, the only information requested (in addition to the access to the mode itself) being the period concerned, the lists of targeted variables (with their values) those of targeting elements.
- The results from iterations. In this list, the “_cur” suffix is associated with current values (of endogenous and exogenous) and the “_star” suffix to the endogenous targets.

The « delta » values represent the difference between target and present solution, for each targeted endogenous. We observe that we need in practice 4 or rather 3 iterations (the first one consisting in simulating the model with the initial values). The relative higher gain is generally obtained at the second.

As an indication, running the program takes 0,003 seconds on a portable computer of average power.

We shall the replicate this test on a larger model.

11.6.1 THE PROGRAM

```
=====
'
' This program computes for a set of exogenous variables
' the values
' which allow a set of endogenous to reach given solutions
'
' The two sets must have the same dimension
'
' The pic_b model is used as an example
' but the program is quite general
'
' The present program applies to a single period
' but it can be generalized quite easily
'
=====

' We set the directory

cd "c:\program files\evIEWS5\pic\"

' Results will be dumped as text to a file called jacob.txt
' overriding any previous elements

output(t,o) jacob
'
' We create a specific workfile called pic_star
' from the original pic_b

' We make sure no file of this name is open at present
'
'

close pic_b.wf1
close pic_star

wfoopen pic_b
wfsave pic_star

' We create a new model
' Basically, the purpose is to present the model used as an example

delete(noerr) _pic_b

model _pic_b
_pic_b.append cap=pk*k(-1)
_pic_b.append ur=q/cap
_pic_b.append q+m=fd+x
_pic_b.merge eq_i
_pic_b.append log(prle_t)=c_prle(1)+c_prle(2)*(t-2002)+c_prle(3)*(t-t1)*(t<t1)+c_prle(4)*(t-t2)*(t<t2)
_pic_b.append led=q/prle_t
```

```

_pic_b.merge eq_le
_pic_b.append lt=le+lg
_pic_b.append rhi=wr*lt+xhr*q
_pic_b.append co=rhi*(1-sr)
_pic_b.merge eq_ic
_pic_b.append fd=co+i+gd+ic+rfdx*q
_pic_b.append          res_m=log(m/(fd+0.5*x))          -1.322108*log(ur)+0.419518*log(compm)-
0.012582*( @trend(60:1)*(t<=2002)+@elem(@trend(60:1),"2002S2")*(t>2002))
_pic_b.merge eq_m
_pic_b.append          res_x=log(x/wd)+0.686672*log(ur)-4.87E-
05*( @trend(60:1)*(t<=2002)+@elem(@trend(60:1),"2002S2")*(t>2002))
_pic_b.merge eq_x

_pic_b.append k=k(-1)*(1-dr)+i

'   Now we start the actual program

'   We shall use the period 2001S1

smpl 2001S1 2001S1

'   We define the lists of elements which change status
'   exogenous to endogenous
'   endogenous to exogenous
'   The numbers must be the same

'   If not, the program stops with a message

group g_vexog1 wd wr
group g_vendo1 q m

!t1=g_vexog1.@count
!t2=g_vendo1.@count
if !t1<>!t2 then
statusline !t1 exogenous !t2 endogenous: stop
stop
endif

'   nstar is the common number

scalar nstar=g_vexog1.@count

'   We delete Scenario 1 (it exists)

_pic_b.scenario(d) "Scenario 1"

'   We build Scenario 0 as the base
'   It will contain the current solution
'   The associated suffix will be "_0"
'   We declare all the overridden exogenous
'   (with the currently best solution)

```

```

' (note: we start with a blank list and fill it one by one)

_pic_b.scenario(n) "Scenario 0"

_pic_b.append assign @all _0

_pic_b.override
for !i=1 to g_vexog1.@count
%1=g_vexog1.@seriesname(!i)
_pic_b.override(m) {%1}
next

' Now we create the alternate scenarios
' for computation of the Jacobian
' They are numbered 1 to nstar (here 2)
' The suffix is "_n"
' The exogenous are the same as Scenario 0

' Note: all exogenous have to be overridden
' but only one will change from the _0 value
' for a given scenario

for !i=1 to nstar
_pic_b.scenario(n,a={!i},i="Scenario 0") "Scenario {!i}"
_pic_b.append assign @all _{!i}
next

' Now we give the values for the targets
' We used arbitrary numbers

smpl 2000S1 2001S1
genr q_star=q*(1+0.05*(t=2001))
genr m_star=m*(1-0.05*(t=2001))

' Now we define the Jacobian

matrix(2,2) jacob

' Now we initialize the exogenous instruments
' used to reach the target

for !i=1 to g_vexog1.@count
%1=g_vexog1.@seriesname(!i)
genr {%1}_cur={%1}
genr {%1}_0={%1}
for !j=1 to g_vexog1.@count
genr {%1}_{!j}={%1}
next
next

```

```

'   And the endogenous target variables

for !i=1 to g_vendo1.@count
%1=g_vendo1.@seriesname(!i)
genr {%1}_0={%1}
genr {%1}_cur={%1}_0
next

'   We set the maximum number of iterations
'   and the convergence criterion

scalar nitmax=20
scalar sconv=1e-6

'   We initialize as control variables
'   the number of iterations
'   And the variable controlling the convergence

!niter=1
!iconv=0

'   We double the elements so they can be displayed

scalar niter=0
scalar iconv=0

'   We display several elements for control

print g_vexog1 g_vendo1
print *_cur *_star

'-----
'   Now we start the loop
'-----

smpl 2001S1 2001S1

'   It will run while convergence is not achieved
'   and the maximum number of loops is not reached

while !niter<=nitmax and !iconv=0
'   We test for convergence
'   We suppose it is achieved (!iconv=1)
'   Then we look at the relative difference
'   between the present solution and the target
'   for each targetted endogenous

'   If the relative difference is higher than the criterion
'   for any target
'   convergence is not yet reached (!iconv=0)

```

```

'     else we stop

!iconv=1
scalar iconv=1
for !j=1 to g_vendo1.@count
%1=g_vendo1.@seriesname(!j)
genr delta_{!j}={({%1}_star-{%1}_cur)/{%1}_cur
scalar delta=@elem(delta_{!j}, "2001S1")

if @abs(delta)>sconv then
!iconv=0
scalar iconv=0
endif
next

if !iconv=1 then
stop
endif

'     We increment the iteration counter

!niter=!niter+1
scalar niter=niter+1

'     We display the differences

print delta_*

'     We compute the base solution
'     defined as Scenario 0

_pic_b.scenario "Scenario 0"
solve _pic_b

'     The solution values are set as current

for !i=1 to g_vendo1.@count
%1=g_vendo1.@seriesname(!i)
genr {%1}_cur={%1}_0
next

'     We display the current solution (exogenous + endogenous) and the targets

print niter
print *_cur *_star

'     Now we compute the Jacobian
'     We start a loop on the shocked instruments

for !i=1 to g_vexog1.@count

```

```

' For each exogenous, we compute the whole set of overriding instruments
' shocking only the current one by 0.1 %
' while the others keep their base value
' This is necessary to implement the current changes

for !j=1 to g_vexog1.@count
%2=g_vexog1.@seriesname(!j)
genr {%2}_!i={%2}_cur*(1+.001*(!i=!j))
next

' We solve the model under Scenario !i

_pic_b.scenario "Scenario !i"
_pic_b.solve

' Now we compute the relative change in the target endogenous
' This will give a column of the Jacobian matrix

for !j=1 to g_vendo1.@count
%2=g_vendo1.@seriesname(!j)
_z=(((%2)_!i-{%2}_cur)/{%2}_cur)/0.001
jacob(!j,!i)=@elem(_z,"2001S1")
next

next

' This ends the computation of the Jacobian

' We compute its inverse

matrix jacob_inv=@inverse(jacob)

' We apply to each exogenous a change equal to
' the inverse of the jacobian
' times the remaining error.

' This gives a new solution for the exogenous
' We start the process again until convergence (hopefully)

for !i=1 to g_vexog1.@count
%1=g_vexog1.@seriesname(!i)
genr {%1}_old={%1}_cur
for !j=1 to g_vendo1.@count
%2=g_vendo1.@seriesname(!j)
genr {%1}_cur={%1}_cur+{%1}_old*jacob_inv(!i,!j)*({%2}_star-{%2}_cur)/{%2}_cur
next
genr {%1}_!i={%1}_cur
genr {%1}_0={%1}_cur
next
wend

```

11.6.2 THE RESULTS

Iteration 0: Initial values

```
=====
obs      2000S1  2000S2  2001S1
=====
WD      411615.4  432806.5  439347.6
WR      0.028897  0.029101  0.029511
Q       1344432.  1365342.  1385977.
M       356344.0  383838.0  389639.0
=====
```

```
=====
obs      2000S1  2000S2  2001S1
=====
M_CUR   356344.0  383838.0  389639.0
Q_CUR   1344432.  1365342.  1385977.
WD_CUR  411615.4  432806.5  439347.6
WR_CUR  0.028897  0.029101  0.029511
M_STAR  356344.0  383838.0  370157.0
Q_STAR  1344432.  1365342.  1455275.
=====
```

```
=====
obs      2001S1
=====
DELTA_1  0.050000
DELTA_2 -0.050000
=====
```

Iteration 1: solution with the initial exogenous

```
=====
obs      2001S1
=====
M_CUR   409107.1
Q_CUR   1383326.
WD_CUR  439347.6
WR_CUR  0.029511
M_STAR  370157.0
Q_STAR  1455275.

DELTA_1  0.052012  The model solution gives larger errors.
DELTA_2 -0.095208  We have not started targeting yet.
```


=====
Iteration 2

=====
obs 2001S1
=====
M_CUR 365029.1
Q_CUR 1441045.
WD_CUR 887589.2
WR_CUR 0.012164
M_STAR 370157.0
Q_STAR 1455275.

DELTA_1 0.009875
DELTA_2 0.014048

=====

Iteration 3

=====
obs 2001S1
=====
M_CUR 370184.2
Q_CUR 1455269.
WD_CUR 893720.4
WR_CUR 0.012746
M_STAR 370157.0
Q_STAR 1455275.

DELTA_1 4.35E-06
DELTA_2 -7.34E-05

=====

Iteration 4: Convergence!

=====
obs 2001S1
=====
M_CUR 370157.0
Q_CUR 1455275.
WD_CUR 893915.9
WR_CUR 0.012738
M_STAR 370157.0
Q_STAR 1455275.

DELTA_1 2.25E-07

=====

11.7 A TEST ON A LARGER MODEL

We reproduced the test on cases closer to operational.

Note: applying the program to a new model is immediate, provided it is available in an EViews workfile along with the required data. One has only to update the two lists of targets and instruments, set the values for the targets, and call for the workfile. In all it should take less than ten minutes.

First case: the quarterly model of the French economy, which we have just used to illustrate modelling under EViews. It includes 88 equations, 11 of which are behavioral. Its specifications are quite traditional, and all behavioral estimations are validated by econometrics. It contains a Keynesian and a price-wage loops.

We have also selected a more operational case. We ask the model to reach;

- An unemployment rate of 6%
- An inflation rate of 2%.
- A null trade balance (exports = imports at current prices).
- A budget deficit of 1% of GDP.

For this we shall use the following instruments:

- The rates of social contributions by both firms and workers.
- Government investment at constant prices.
- The number of civil servants.

The algorithm converges again, this time in 5 iterations. Computation time is 0.231 seconds.

Second case: a multi country yearly model for the Andean community (5 countries including Venezuela and a simplified rest of the world). It includes 770 equations, 11 x 5 of which are behavioral. Its specifications are again quite traditional, and all behavioral estimations are calibrated this time. It formalizes completely the trade flows.

This time we set as targets the Government deficit of each country (at 3% of GDP) and as instruments in sequence:

- Government investment
- The VAT rate
- The rate of social security contributions by workers
- The rate of social security contributions by firms.
- The exchange rate
- An increase in the long term interest rate.

THz system converged in all cases, taking between 3.7 and 5.2 seconds (5 iterations most of the time).

and finally:

- targets: the Government deficit of each country (at 3% of GDP) and balanced trade (net exports=0)
- instruments: Government investment and subsidies to firms.

converges in 7 .2 seconds.

A LIST OF USEFUL SERIES FOR PRODUCING A SINGLE COUNTRY, SINGLE PRODUCT MODEL

We shall give now a list of series normally useful for producing a simple but operational model. Of course this is only a proposal, designed as a starting point to which the user is welcome to apply as many changes as he wishes.

Just as we have done for model specifications, we think it is easier to start from a proposal, even rather far from what the modeller has in mind, than from nothing. Convergence to an acceptable version can start immediately, by a series of acceptations, rejections and changes based on an actual solution.

12.1 THE PERIODICITY

It should be **quarterly**, annual if quarterly data is not available. This is not so important for population, or for stock variables: capital, housing, credits and debts.

12.2 GENERAL ELEMENTS

If several series are available for the same concept (as unemployment) the first priority is to get one series, the second to get all available series with their definition. Of course if several sets of series are available for the same domain (like exports and imports in both local currency and dollars) the priority is to make the whole set as consistent **as possible**. This means for instance that being part of the major set of data used can be considered more important than perfect adaptation to the concept required.

We are not considering **product decomposition**. But oil could be separated in terms of production, exports, and relation to the State budget.

The agent decomposition is:

Households
Firms (private and public)
Financial institutions (non-state)
State (including local administrations and Social Security)
Rest of the World.

12.3 THE SERIES

The most important series are underlined.

Some of the variables can be computed from others. They can be provided or not.

We need

12.3.1 THE SUPPLY - DEMAND VARIABLES AT CONSTANT PRICES (WHOLE ECONOMY)

Total local demand (including intermediary consumption)

Intermediary consumption by use and product
Total local demand
Household consumption at constant prices
Housing investment by households at constant prices
Housing investment by the State at constant prices
Productive investment by firms at constant prices
Government investment at constant prices
Government consumption at constant prices
Changes in inventories at constant prices.
Exports at constant prices (constant terms)
Imports at constant prices (constant terms)
GDP at constant prices

12.3.2 THE SAME ELEMENTS AT CURRENT PRICES

12.3.3 THE SAME ELEMENTS DECOMPOSED INTO PRODUCTS

All elements have an additional dimension (3) except for

intermediate consumption (3 x 3)
productive investment (3 x 3)

12.4 PRODUCTIVE CAPACITY

Productive capital at constant prices
If not, capital depreciation at constant prices
Productive capital (material) at constant prices
Productive capital (building) at constant prices
Rate of use of capacities (ratio)

12.5 EMPLOYMENT

Employment of firms (global)
Employment of households
Employment of financial institutions
Employment of the State (including local administrations)
Self employed
Family workers
Unemployment
Unfilled job vacancies
Weekly work duration
Population
Population in age of working (or 15-65 years)

12.6 PRICE DEFLATORS

The deflators associated to the supply - demand equilibrium (but we can compute them ourselves if we have elements at both current and constant prices).

The consumer price index

The production price index

The yearly wages (global)

The yearly wages (firms)

The yearly wages (State)

The yearly wages (employed by households)

The yearly wages (by product)

12.7 HOUSEHOLDS ACCOUNT

All in current terms

Revenue before income tax

Revenue after income tax

Revenue from wages

Individual workers revenue

Pensions

Social security benefits (global)

Social security benefits (illness).

Social security benefits (family).

Revenue from housing

Interests received.

Income tax

Social security contributions

Interests paid.

Housing capital

12.8 FIRMS ACCOUNT (ALL TYPES OF FIRMS)

All elements in current terms

Value added

Intermediary consumption

Tax on production

Wages

Social security contributions paid by firms

Revenue of individual workers

Dividends paid

Dividends received

Interests paid

Interests received

Tax on profits

Productive investment

Change in inventories

Balance

12.9 REST OF THE WORLD

Trade balance

Financing capacity

Balance of services

Exchange rate to the dollar

Exchange rate to the partners of the country.

Balance of payments

Capital movements

Wage transfers of the country's expatriate workers.

Demand in the Rest of the World (weighted by the share of countries in local exports).

12.10 GOVERNMENT ACCOUNT (ELEMENTS NOT DESCRIBED EARLIER)

All elements in current terms

VAT

Other indirect taxes

Tariffs

12.11 FINANCIAL DOMAIN

Short term interest rate

Long term interest rate

Interest rate on consumption

Interest rate on housing investment

All elements in current terms

Stock of debt for firms (global)

Stock of debt for firms (private)

Stock of debt for firms (public)

Stock of debt for the State

Stock of debt for households

Capital movements

Foreign direct investment (in current terms)

Foreign direct investment (in constant terms)

A LIST OF USEFUL SERIES FOR A THREE SECTOR MODEL

13.1 THE PERIODICITY

Again, it should be **quarterly**. But this is not so important for population, or for stock variables: capital, housing, credits and debts.

13.2 GENERAL ELEMENTS

If several series are available for the same concept (as unemployment) the first priority is to get one series, the second to get all available series with their definition. Of course if several sets of series are available for the same domain (like exports and imports in both local currency and dollars) the priority is to get one consistent **set**.

The agent decomposition is:

- Households
- Firms (private and public)
- Financial institutions (non-state)
- State (including local administrations and Social Security)
- Rest of the World

In addition to non-market services (separated in the first version) the **product decomposition** is:

- Agriculture, fishing and forestry
- Manufacturing
- Market services

This decomposition applies to series noted *

In addition, **separations** should be considered for:

- Agricultural production into firms and family
- This applies to production, labor, investment, wages, and revenue

- Manufacturing trade, into energy and non-energy
- This applies to local and world demand, imports and exports at constant and current prices.

13.3 THE SERIES

The most important series are noted by *.

Some of the variables can be computed from others. They can be provided or not.

We need

13.3.1 THE SUPPLY - DEMAND VARIABLES AT CONSTANT PRICES (WHOLE ECONOMY)

- * Total local demand (including intermediary consumption)
- * Intermediary consumption by use and product
- * Total final local demand
- * Household consumption at constant prices
- * Housing investment by households at constant prices
- * Housing investment by the State at constant prices
- * Productive investment by firms at constant prices
- * Government investment at constant prices
- * Government consumption at constant prices
- * Changes in inventories at constant prices.
- * Exports at constant prices (constant terms)
- * Imports at constant prices (constant terms)
- * GDP at constant prices

13.3.2 * THE SAME ELEMENTS AT CURRENT PRICES

All elements have dimension 3 except for

Intermediate consumption (3 x 3)
Productive investment (3 x 3)
Change in inventories (3 x 3)

13.4 PRODUCTIVE CAPACITY

- * Productive capital at constant prices
- If not, capital depreciation at constant prices
- * Productive capital (material) at constant prices
 - * Productive capital (building) at constant prices.
 - * Rate of use of capacities (ratio).

13.5 EMPLOYMENT

- * Employment of firms (global).
- Employment of households
Employment of financial institutions
Employment of the State (including local administrations)
- * Wage earners in firms (global).
- Employment of households
Employment of financial institutions
Employment of the State (including local administrations)
- * Self employed

- * Family workers
- Unemployment
- Unfilled job vacancies
- Weekly work duration
- Population
- Population in age of working (or 15-65 years)

13.6 PRICE DEFLATORS

The deflators associated to the supply - demand equilibrium (but we can compute them ourselves if we have elements at both current and constant prices).

- The consumer price index
- The production price index
- * The yearly wages (global)
- * The yearly wages (firms)
- The yearly wages (State)
- The yearly wages (employed by households)
- The yearly wages (by product)

13.7 HOUSEHOLDS ACCOUNT

All in current terms

- Revenue before income tax
- Revenue after income tax

- * Revenue from wages
- * Individual workers revenue
- Pensions
- Social security benefits (global)
- Social security benefits (illness).
- Social security benefits (family).
- Revenue from housing
- Interests received.

- Income tax
- Social security contributions
- Interests paid.

Housing capital

13.8 FIRMS ACCOUNT (ALL TYPES OF FIRMS)

All elements in current terms

- * Value added

- * Intermediary consumption
- * Tax on production
- * Wages
- * Social security contributions paid by firms

- * Revenue of individual workers
- * Dividends paid
- * Dividends received
- * Interests paid
- * Interests received
- * Tax on profits

- * Productive investment
- * Change in inventories
- * Balance.

13.9 REST OF THE WORLD

Trade balance

Financing capacity

Balance of services

*Exchange rate to the dollar

Exchange rate to the partners of the country

Balance of payments

Capital movements

*FDI

*Wage transfers of the country's expatriate workers.

*Demand in the Rest of the World (weighted by the share of countries in local exports).

13.10 GOVERNMENT ACCOUNT (NOT DESCRIBED EARLIER)

All elements in current terms.

* VAT

* Other indirect taxes

* Tariffs

13.11 FINANCIAL DOMAIN

Short term interest rate

Long term interest rate

Interest rate on consumption

Interest rate on housing investment

All elements in current terms

Stock of debt for firms (global)

- * Stock of debt for firms (private)
- * Stock of debt for firms (public)
- Stock of debt for the State
- Stock of debt for households
- * Capital movements
- * Foreign direct investment (in current terms)
- * Foreign direct investment (in constant terms)

EViews 8 ADDITIONS FOR MODELLING

We are now going to summarize the new options and functions proposed by version 8 of EViews, and which apply to the modelling task. Each of them has already been addressed in the relevant part of the book, but we thought also interesting to give users a synthetic presentation¹⁶⁶.

Some of these functions are not directly related to modelling, but improve in some way the modelling process.

14.1 SERIES MANAGEMENT

- EViews 8 allows to introduce one's own labels, for instance the country for a multinational model, the agent for an accounting one, or the fact that a series belongs to a particular model.

For instance you can use:

```
HI.label(agent) Households  
MARG.label(agent) Firms
```

- If the workfile window screen is in "Display+" mode, you can sort the elements according to their characteristics. In addition to the name, the type and the time of last modification (or creation) you have access to the description.

Moreover, if you right click on one of the column headings, and choose "Edit Columns" you can display additional columns for any of the label types, including the ones you have created.

This can prove quite useful, as it allows you to filter and sort on any criterion, provided you have introduced it as a label.

This criterion can be for instance:

- The agent concerned
- The country
- The association with a given model
- The formula in the model
- The formula used to create the series (if any)¹⁶⁷
- The type within this model (exogenous, endogenous, identity, behavior...)
- The sub-type: for exogenous it can be policy, foreign, structural. For endogenous it can be behavior or identity.

¹⁶⁶ Especially as some users might have skipped them along with features they feel they already know well enough.

¹⁶⁷ You can also use the "source"

Once the display is produced, it can be transferred to a table, which can be edited (lines, fonts...) and used for presentations.

For instance, one can produce a table for a model, with columns for type, agent, units, source, identity / behavior.... This table can be sorted using any of the criteria.

These new functions allow table production to be integrated in the modelling process, a very powerful information tool for both model development and documentation.

For instance you could use:

```
F_HDI.label(d) Disposable income
U_MARG.label(d) Margins
F_HDI.label(model) France small
U_MARG.label(model) USA small
F_HDI.label(agent) Households
U_MARG.label(agent) Firms
```

and produce sorted tables according to any of the three criteria.

- You can compare elements between workfiles and pages inside the same workfile. EViews will display one line per element, in which will be stated its relation, between: unchanged, modified (numerically), added, deleted, replaced (logically, the last case applies for instance to a linked variable have been modified). A filter can be applied.

For series, a tolerance level can be set, under which the series are not considered modified. The display will tell how many periods show a higher difference.

By default, all elements will be displayed, but one can restrict the case (for instance, to all variables present in both pages with a difference higher than the criterion).

Equations and models are not compared but appear in the list.

The associated command is **wfcompare**.

```
wfcompare(tol=criterion,list=comparison_type) list_of_compared_series list_of_reference_series
```

For more details you should refer to the EViews Help.

For instance if you want to compare all French series (starting with "FRA_") between the pages "base" and "updated", for a tolerance level of 0.00001 one will state :

```
wfcompare(tol=1E-5,list=m) updated\fra_* base\fra_*
```

This command can be particularly useful

- To control the evolution of historical values for a model data set, showing which equations will have to be estimated again.
- To summarize the results of a residual check, showing for which equations the right hand side (using historical values of the explained variable) is different from the right hand side (the result of the computation). By setting a tolerance level slightly higher than zero (for instance 0.0001) one can restrict the display to the errors deemed significant.

14.2 PROGRAMS

EViews 8 improves the way programs can be run.

You can run part of a program, by selecting it with the mouse (in the usual Windows way), clicking on the right button, and choosing “Run Selected”.

This is generally more efficient than the previous method of copying the selected part into a blank program, and running it. However the new method does not allow editing, useful when one wants to run a selected AND modified set.

Symmetrically one can exclude temporarily from execution part of a program, by “commenting it out”. To do this, one should select the relevant part, click on the right button, and choose “Comment Selection”. To reactivate the statements, one should select them again and use “Uncomment Selection”.

This can be a little dangerous, especially if you (like myself) have the reflex of saving the program before each execution. To avoid destroying the original, one can save first the modified program under another name¹⁶⁸.

Finally, one can ask a column of numbers to be displayed left of the program lines. This is particularly efficient if you use the “Go To Line” statement¹⁶⁹.

14.3 MODEL MANAGEMENT

EViews 8 greatly improves the management of model equations, especially identities. Until EViews 7:

- Dropping an identity was not possible.
- Adding a new version actually duplicated the definition, so you had more equations than endogenous.
- Dropping an estimated equation could only be done through EXCLUDE, difficult to manage, as to keep it excluded one had to specify it again in all following EXCLUDE statements.
- To replace an estimated equation you had to use the same name, otherwise you faced the same duplication problem as above.

¹⁶⁸ Only once of course.

¹⁶⁹ However, you have to be careful to update the numbers when the program changes.

Now you can (if you choose that option¹⁷⁰):

- Drop any kind of equation, using :

```
model_name.drop variable_name (for identities)
```

```
model_name.droplink equation_name (for estimated equations)
```

For instance you can specify:

```
_fra_1.drop GDP ' GDP is made exogenous
```

```
_fra_1.droplink eq_i ' eq_i, the equation for investment is dropped, and I becomes exogenous
```

- Replace the formula using :

```
model_name.replace variable_name (for identities)
```

```
model_name.replacelink old_equation_name new_equation_name (for estimated equations)
```

```
_fra_1.replace GDP = .....
```

```
_fra_1.replacelink eq_i eq_i_new
```

- In addition you can also replace all occurrences of a model variable using a new name, through:

```
model_name.replacevar old_variable_name new_variable_name
```

For instance:

```
_fra_1.replacevar GDP Q ' replaces GDP by Q in the whole model (but not in estimated relations)
```

14.4 MODEL SOLVING MANAGEMENT

¹⁷⁰ Remember that in our opinion, the most efficient method for editing a model is to destroy the previous version and create a new one from scratch.

In the chapter about model solving, we proposed to override a list of series by creating a new list from scratch, and in particular to create a blank list to run a solution with no overrides.

EViews 8 provides an alternate option: you can drop overrides individually through the REVERT command:

```
model_name.revert list-of-unoverriden-series
```

But the most important addition is the possibility to specify directly the changes in the assumptions, using ADJUST.

The syntax is:

```
model_name.adjust(init=initial_series) series adjustment
```

For instance the statements :

```
_mod_1.scenario(a=_2) "scenario 2"  
_mod_1.adjust(init="scenario 1") gdp =+10000
```

will create a series called gdp_2 with a value 10000 higher than the value from "scenario 1" (maybe called gdp_1). The series will be added to the "override" list, and to the "exclude" list as well if the variable is endogenous.

Be careful to introduce a space before the "=" sign, or gdp will take the value 10000!

Introducing a change involving elements, such as series and parameters, is possible but more difficult. One can prefer using actual expressions, as we have done in our examples.

14.5 DISPLAYS

The graphic functions of EViews have been improved. You can now:

- Decide on the scope of the graph using a slide line, available at the bottom.
- Add custom arrows.
- Export the graph to a PDF file.

But the most interesting addition is the MAKEGRAPH command, specially adapted to model solutions. The syntax is:

```
model_name.makegraph(options) graph_name model_series_names
```

If no option is specified only the current scenario solution is displayed.

The main options are:

a : include actuals
c : compare active to baseline
d : include deviations from baseline (as an additional graph in the same frame)
n : do not include active scenario.

The results can be presented raw or transformed, and for stochastic simulations a confidence interval can be displayed.

Once the graph is created, the usual modifiers can be applied (legend, type of line, colors....).

One has to note that the graph must not preexist. If it does it must be deleted first (using the “noerr” modifier to avoid a possible error message).

`delete(norerr) name_of_the_graph`

BIBLIOGRAPHY

We have limited this work to elementary econometrics notions, and we have used only macroeconomics to illustrate some methods. We give you now a list of more specialized, or less basic, reference works.

Econometrics and statistics

BALTAGI, B.H., (2007) *Econometric Analysis of Panel Data*, Wiley, 4th Ed.

BANERJEE A, GALBRAITH JW, DOLADO J, HENDRY DF, (1993) *Co-Integration, Error Correction, and the Econometric Analysis of Non-Stationary Data (Advanced Texts in Econometrics)*, Oxford University Press

CHAREMZA W, DEADMAN D,(1997) *New Directions in Econometric Practice: General to Specific Modelling, Cointegration, and Vector Auto regression*, Edward Edgar Pub

DIEBOLD, F.X., (2000), *Elements of Forecasting*, South Western, 2ème Ed.

ENDERS, W., (2004), *Applied Econometric Time Series*, Wiley Series in Probability and Statistics, Wiley, 2nd Ed.

GOMEZ V, Maravall A(1996). *Programs TRAMO (Time series Regression with ARIMA noise, Missing observations, and Outliers) and SEATS (Signal Extraction in ARIMA Time Series). Instructions for the User.* Working Paper 9628, Banco de España.

GOURIEROUX C, MONFORT A, VUONG Q (translator) (1995) *Statistics and Econometric Models (Themes in modern econometrics)s*, Cambridge University Press.

- GOURIEROUX C, MONFORT A, GALLO G (translator) (1997) Time Series and Dynamic Models (Themes in Modern Econometrics) Cambridge University Press.
- GREENE W.H., (2011), Econometric analysis, Pearson , 7th Ed.
- CLEMENTS P, HENDRY D, (1998), Forecasting Economic Time Series, Cambridge University Press.
- HAMILTON JD, (1994), Times Series Analysis, Princeton University Press.
- HARVEY A C,(1990) The Econometric Analysis of Time Series; MIT Press.
- HENDRY DF, (1995) Dynamic Econometrics, Oxford University Press.
- JUSELIUS, K, (2006), The Cointegrated VAR Model: Methodology and Applications, Oxford University Press.
- LADIRAY D, QUENNEVILLE B, (2001), Seasonal adjustment with the X-11 Method, Springer.
- ROBERT E. LUCAS, Jr. (1976), Econometric policy evaluation: a critique.' Carnegie-Rochester Conference Series on Public Policy' 1, pp. 19–46.
- LÜTKEPOHL H., KRÄTZIG, (2004), Applied Time Series Econometrics, Cambridge.
- MADDALA GS, (1994), Econometric Methods and Applications (Economists of the Twentieth Century), Edward Elgar Pub.
- RAO B, (2007),Cointegration: For the Applied Economist, St Martins Press, 2nd Edition.
- X-13-ARIMA-SEATS Reference Manual, (2013),Version 0.3,Time Series Research Staff, Statistical Research Division, U.S. Census Bureau, <http://www.census.gov/ts/x13as/docX13AS.pdf>

Macroeconomics

- AGENOR, P.-R., MONTIEL, P.J., (2008), Development Macroeconomics, Princeton University Press, 3rd Ed.
- AGHION P, HOWITT P, (2008), The Economics of Growth, MIT Press.
- BARRO R.J. et X. SALA-I-MARTIN, (2003), Economic Growth, Mc Graw-Hill, New York (2nd edition).
- BEHRMAN, J., & SRINIVASAN, T.N., (Eds.) 1995, Handbook of Development Economics, Vol III., Elsevier.
- BLANCHARD O J (2009) Macroeconomics, 5th Ed., Prentice Hall .
- BLANCHARD O J, FISHER S (1989) Lectures on Macroeconomics, MIT Press.
- BURDA M, WYPLOZ C (2009), Macroeconomics, An European text, 5th ED , Oxford University Press.
- CHENERY, H., et SRINIVASAN, 1989, Handbook of Development Economics, North Holland, Amsterdam.
- GALI, J (2008), Monetary Policy, Inflation, and the Business Cycle. Princeton University Press
- GROSSMAN, G., M., 1995, Handbook of International Economics, Vol.III, Elsevier, North Holland.

LEQUILLER F, BLADES D, (2006), Understanding National Accounts, OECD Publishing, downloadable at : <http://www.oecd.org/dataoecd/37/12/38451313.pdf>

MUET P A, Théories et modèles de la macroéconomie, (1990) Economica.

OBSFELD, M., & ROGOFF, K., (1996), Foundations of International Macroeconomics, MIT Press.

RAY, D., (1998), Development Economics, Princeton University Press,

ROMER D (2011) Advanced Macroeconomics, McGraw-Hill, 3th edition .

TAYLOR, J.B., and WOODFORD, M., (1999), Handbook of Macroeconomics, Elsevier, North Holland.

TAYLOR, J.B., (1993) "Discretion versus Policy Rules in Practice," Carnegie-Rochester Conference Series on Public Policy, 39, pp.195-21

WOODFORD M (2003), Interest and Prices: Foundations of a Theory of Monetary Policy. Princeton University Press

Models and modelling

AMMAN H, KENDRICK D.A., RUST J., (eds.), (1996), Handbook of computational economics : H vol. 1. North-Holland, Amsterdam,

AMMAN H, RUSTEM B, WHINSTON A (1996) Computational Approaches to Economic Problems (Advances in Computational Economics, Vol. 6, Kluwer Academic Publishing.

ARTUS P, DELEAU M, MALGRANGE P (1986) , Modélisation Macroéconomique, Economica.

BODKIN R., KLEIN L., MARWAH K., (1991), A History of Macroeconometric Model Building. Edward Elgar.

BRILLET JL (1994) Modélisation Econométrique: Principes et Techniques, Economica.

CANOVA, F., 2007, Methods for Applied Macroeconomic Research, Princeton University Press.

CARNOT, N. , TISSOT, B., 2002, La Prévision Economique, Economica.

DEJONG, D. N. with DAVE C. (2007), Structural Macroeconometrics. Princeton University Press

.FAIR RC (1994), Testing macroeconometric Models, Harvard University Press.

FAVERO, C.A., 2001, APPLIED MACROECONOMETRICS, OXFORD.

FISHER P (1992), Rational Expectations in Macroeconomic Models (Advanced Studies in Theoretical & Applied Econometrics, Vol. 26), Kluwer.

FOUQUET D et alii (1978) DMS, Modèle Dynamique Multi Sectoriel, Collections de l'INSEE no 64-65.

GILLI M (1996), Computational Economic Systems: Models, Methods & Econometrics (Advances in Computational Economics, Vol 5), Kluwer Academic Publishing.

- GILLI M, ROSSIER E (1980), Understanding and Solving Complex Systems, Automatica, Vol. 17, pages 647-652.
- DERVIS, K., DE MELO, J. et ROBINSON, S., 1982, General Equilibrium Models for Development Policy, World Bank, Washington D.C..
- GANTMACHER, F. (1959), Theory of matrices, AMS Chelsea publishing
- LAXTON D (1998) Multimod Mark III: The Core Dynamic and Steady-State Models, IMF
 Pindyck R, Rubinfeld D (1997), Econometric Models and Economic Forecasts; Cambridge University Press.
- INTRILIGATOR M, BODKIN G, HSIAO C. (1995), Econometric Models, Techniques and Applications , Prentice Hall Press.
- KLEIN L (1991) Comparative Performance of U.S. Econometric Models, Oxford University press
- LAFFARGUE, JP (1990) Résolution d'un modèle économique avec anticipations rationnelles, Annales d'Economie et de Statistiques no 17.
- NEPOMIASCHY P, RAVELLI A (1978), Adapted Methods for Solving and Optimizing Quasi--Triangular Econometric Models , Annals of Economic and Social Measurement.
- PONTRAGIN L.S, BOLTJANSKII V.G, AMKRELIDZE R. V, MISCHENKO E. F (1962) The Mathematical Theory of Optimal Processes (Russian), English translation: Interscience.
- SADOULET, E., DE JANVRY (1995), Quantitative Development Analysis, John Hopkins University Press.
- SBODORNE A., TAMBALOTTI A., RAO K., WALSH K., (2010), 'Policy analysis using DSGE models: an introduction'. Federal Reserve Bank of New York Economic Policy Review 16 (2).
- SHOVEN J, WHALLEY J (1992) Applying General Equilibrium (Cambridge Surveys of Economic Literature) Cambridge University Press
- TAYLOR, L. (Ed.) 1990, Socially Relevant Policy Analysis, Structuralist Computable General Equilibrium Models for the Developing World, MIT Press., Cambridge, Mass.
- WHITNEY, J.D., 1994, A Course in macroeconomic modelling and forecasting, Harvester Wheatsheaf.

INDEX

- access (to data), 66, 68, 69, 75, 308
- aggregation, 54, 69, 70, 72
- algorithm, 36, 44, 49, 71, 186, 187, 190, 193, 195, 201, 203, 205, 206, 212, 213, 214, 233, 298, 300
- analytic (shocks), 239, 246, 300
- autocorrelation, 237
- average (solution), 276, 277
- average absolute error, 235
- behavioral equations, 17, 21, 22, 47, 101, 238, 275, 300
- bias, 235, 276
- block-recursive, 200
- blocks (model), 200
- bootstrap, 276
- Broyden (algorithm), 190, 191
- complex shocks, 48, 300, 301
- constants with dimension, 43
- continuous time models, 29
- convergence (study of), 190, 193, 194, 195, 200, 203, 204, 205, 207, 211, 214, 233, 283, 285
- convergence criterion, 187, 193, 194, 227, 236
- convergence of the solution, 46, 192, 194, 198, 203, 210, 212, 214, 215, 234, 283, 285
- convergence problems, 46, 192, 198, 200, 205, 231, 234
- convergent, 206, 214, 283
- convexity of models, 37
- cumulated shock, 247
- cyclic dynamics, 283
- damping factors, 207, 208, 209, 210, 211, 227
- data correction, 40, 238
- data definition, 73, 309
- data format, 66, 67, 68, 75
- data gathering, 50, 65
- Deleau and Malgrange, 234, 286
- deterministic (solution), 276
- differentiability of models, 33
- dimension, 43
- direct access to data, 29, 68
- disaggregation, 69, 70, 71
- discretization, 24
- divergent dynamics, 207, 230, 233, 283, 286, 289
- documentation of data, 73
- documentation of models, 61
- dummy variables, 25, 238
- Durbin - Watson statistic, 115
- dynamic models, 26
- dynamic simulations, 56, 235, 236, 239, 240, 285
- economic policy, 17, 48, 236, 296, 297, 301, 302, 307
- economic policy (shock), 48, 301, 302, 307
- efficiency, 277
- eigenvalues of the Jacobian, 205, 212, 214, 233, 277, 282, 283, 284, 285, 286, 293, 294
- eigenvectors of the Jacobian, 212, 283
- elimination, 29, 234
- endogenous (variables)**, 37, 195

error (estimation), 22, 227, 235, 238, 239, 240, 241, 276, 300
 error (simulation), 22, 100, 101, 102, 183, 194, 222, 227, 231, 235, 236, 237, 238, 239, 240, 241, 274, 276, 277, 278, 298, 300
 error correction models, 293, 295
 estimated coefficients, 21, 102, 183, 205, 231, 234, 238, 275, 276
 estimation, 21, 22, 29, 47, 60, 61, 102, 111, 234, 237, 239, 241
 estimation methods, 29, 50, 71, 239
 exogeneity, 231
 exogenous (variables), 33, 41, 47, 297
 explained (variable), 22, 112
 explanatory (variable), 19, 25, 199, 200, 238, 239
 ex-post (simulations), 241
 ex-post (simulations), 234, 240
 external methods, 52
 external shocks, 301, 308
 field (models), 53, 54, 61
 finite differences, 190, 212, 213, 214
 forecasts, 303
 formal transformations, 72
 Gauss-Seidel (algorithm), 200, 201, 203, 204, 208, 212, 213, 215, 230
 Gilli-Rossier method, 200
 global compatibility (specifications), 37
 graphs, 73, 234, 308, 309, 310
 histograms, 309, 310
 homogeneity, 43
 horizon, 55, 56, 235
 horizontal (aggregation), 54
 identifiability, 46
 identified (formulations), 190
 identities, 19, 47, 240
 incidence matrix, 195, 196, 199
 information, 73, 308, 309
 instruments, 17, 278, 297, 298, 302, 306, 307
 internal (methods), 52
 interpretation, 54, 310
 iterations (simulations), 187, 189, 190, 192, 193, 195, 198, 200, 203, 204, 205, 207, 208, 209, 212, 213, 214, 215, 226, 230
 Jacobian matrix, 189, 190, 200, 205, 209, 212, 213, 214, 276
 lags, 240, 284, 286
 lengthening (of data), 72
 linearity (equations), 43
 linearity (estimation), 113
 linearity (models), 30, 32, 34, 51, 205, 209, 212, 214
 linearization (Newton), 188, 213
 linearized (models), 282
 loop (variables), 199, 200, 201, 203, 205, 206, 208, 209, 212, 213, 214, 215, 234
 maintained (shock), 247, 304
 means, 41, 187
 methods of analysis, 274, 282, 286, 293
 minimal (set of loop variables), 200
 minimum (set of loop variables), 200
 model analysis, 30, 52, 54, 56, 246, 274, 277, 282, 286, 293
 model continuity, 33, 34
 model dynamics (analysis of), 28, 283, 285, 293, 295
 Monte Carlo, 275
 Nepomiaschy-Ravelli, 200

Newton (algorithm), 188, 190, 191, 200, 209, 212, 213, 214
 classification, 61, 71, 308
 non - linear (model), 276
 non - linearity (equations), 31
 normalization (equations), 44, 46
 normalized (equations), 183, 196, 235
normative (scenarios), 297, 302
 one-time shocks, 247
 optimal control, 306
 organization of the modelling process, 60, 61, 74, 75
 parameters, 18, 20, 33, 113, 297, 307
 periodicity, 24, 25, 29, 56, 69, 71, 73
 physical units, 65
 pies, 309
 preliminary processing of data, 69
 presentation of results, 235, 308, 309
 processing (of data), 69
 processing of problems, 49, 74, 221
 R square (test), 112
 random term, 21, 22, 23, 275
 rational expectations, 28
 recursivity, 199
 residual (estimation), 22, 102, 276, 300
 residual check, 99, 101, 102, 222
 resolution, 30, 186, 200, 298
 Ritz-Jordan (algorithm), 187, 230
scenarios, 48, 55, 296, 297, 302
 scientific applications of models, 52, 274, 307
 seasonality, 25
 shocks, 43, 48, 51, 56, 239, 246, 247, 282, 283, 285, 288, 294, 296, 300, 301, 302, 307, 308
 simulation error due to coefficients, 241, 275, 276
 simulation error due to residuals, 101, 275, 300
 simulations, 55, 234, 235, 236, 240, 277, 297, 299, 308
 size of models, 28, 40, 50, 51, 53, 61, 200, 213, 215, 234, 248, 307
 smoothing (of data), 71
 software aspects, 67, 68, 73, 74, 200, 309
 specification (of the model), 47, 102, 111
 specification constraints, 47
 stability (dynamics), 213, 283, 288
 standard error (estimation), 23, 278
 standard error (simulation), 235, 277
 static (models), 26
 static (simulations), 235
 steady state paths, 283, 288, 294
 suppression (of data), 73
 tables, 73, 308, 309
 temporal (aspect), 24, 66
 term (short, medium, long), 56
 term (short, medium, long), 17, 43, 55
 term (short, medium, long), 240
 term (short, medium, long), 300
 term (short, medium, long), 300
 term (short, medium, long), 304
 tests (estimation), 25, 113, 238
 theoretical models, 29, 51, 57, 307
 time dimension, 24
 trend (estimates), 114, 237
 trend forecasts, 296

types of error, 22, 101, 238, 239, 274, 276, 277, 278

types of model, 48

uncertainty, 274, 275, 276, 277

uniqueness of the solution, 36

units (models), 236

update (of data), 73, 74

use (models), 48, 55, 213

validation of the model, 49, 234, 307

variables, 15, 187, 236, 239

vertical (aggregation), 54