Theory and Practice of TFP Estimation: the Control Function Approach Using Stata

Vincenzo Mollisi [1]
(with Gabriele Rovigatti [2])

[1] Free University of Bozen-Bolzano Faculty of Economics and Management

[2] University of Chicago Booth School of Business Chicago, IL

April 20, 2017

TFP estimation: the setup

Consider a Cobb-Douglas log production function

$$y_{it} = \alpha + \mathbf{w}_{it}\beta + \mathbf{x}_{it}\gamma + \delta\, l_{it} + \omega_{it} + \varepsilon_{it} \tag{1}$$

where the component ω_{it} is the unobservable productivity or technical efficiency. By assumption, it evolves according to a first-order Markov process

$$\omega_{it} = E(\omega_{it}\,|\Omega_{it-1}) + \xi_{it} = E(\omega_{it}\,|\omega_{it-1}) + \xi_{it} = g(\omega_{it-1}) + \xi_{it} \tag{2}$$

where ξ_{it}, the productivity shock, is uncorrelated with both ω_{it} and the state variable $\mathbf{x}_{it}$.
In order to consistently estimate β and γ - and obtain reliable values of $\hat{\omega_{it}}$ - several methods have been proposed (mainly two steps procedures).

Olley-Pakes

Assumptions

A.1 $i_{it} = f(\mathbf{x}_{it}, \omega_{it})$, the investment policy function, is invertible and monotonically increasing in ω_{it};

A.2 The state variables evolve according to i_{it} and is decided at time $t-1$;

A.3 The free variables $\mathbf{w}_{it}$ are non-dynamic, i.e. their choice does not impact future profits, and are chosen at time t *after* the productivity shock realizes.

Levinsohn-Petrin

Assumptions

B.1 Firms observe their productivity shock and adjust their optimal level of intermediate inputs according to the demand function $m(\omega_{it}, \mathbf{x}_{it})$;

B.2 $m_{it} = f(\mathbf{x}_{it}, \omega_{it})$, the intermediate input function, is invertible and monotonically increasing in ω_{it};

B.3 The state variables evolve according to the investment policy function $i()$ which is decided at time $t-1$;

B.4 The free variables $\mathbf{w}_{it}$ are non-dynamic, i.e. their choice does not impact future profits, and are chosen in t *after* the firm productivity shock realizes.

Ackerberg-Caves-Frazer

Assumptions

C.1 $p_{it} = p_{it}(\mathbf{x}_{it}, w_{it}, \omega_{it})$, the proxy variable policy function, is invertible and monotonically increasing in ω_{it};

C.2 The state variables are decided at time $t - b$;

C.3 The labor input, l_{it}, is chosen at time $t - \zeta$, where $0 < \zeta < 1$. The free variables, $\mathbf{w}_{it}$, are chosen at time t when the firm productivity shock is realized.

Outline of the Algorithm 1

X.1 Under A.1-A.2, B.1-B.2 or C.1 assumptions, a proxy of ω_{it} is obtained through inversion of a policy function of the proxy variable

$$\omega_{it} = f^{-1}(p_{it}, ...) = h(p_{it}, ...) \tag{3}$$

X.2 Plug $h(p_{it}, ...)$ in (1) and estimate non-linearly:

$$y_{it} = \alpha + \mathbf{w}_{it}\beta + \mathbf{x}_{it}\gamma + h(i_{it}, \mathbf{x}_{it}) + \varepsilon_{it} \tag{4}$$

Outline of the Algorithm 2

X.1 After the recover of ω_{it}, exploiting (2), obtain the residuals ξ_{it} (First Stage).

X.2 Form the GMM criterion function by exploiting moment conditios $E[\xi_{it} z_{it}^k]$=0, $\forall k$, where k is the index of the instrument vector $\mathbf{z} = [\mathbf{x}_{it},\, \mathbf{m}_{it-1},\, l_{it-1}]$ (Second stage):

$$[\gamma^*,\, \beta^*,\, \mu^*] = \text{argmax} \left\{ \sum_k \left(\sum_i \sum_t \xi_{it} z_{it}^k \right)^2 \right\} \qquad (5)$$

Wooldridge - System GMM

$$y_{it} = \alpha + \mathbf{w}_{it}\beta + \mathbf{x}_{it}\gamma + h(\mathbf{x}_{it}, \mathbf{m}_{it}) + v_{it} \tag{6}$$

$$y_{it} = \alpha + \mathbf{w}_{it}\beta + \mathbf{x}_{it}\gamma + f[h(\mathbf{x}_{it-1}, \mathbf{m}_{it-1})] + \eta_{it} \tag{7}$$

where $h(\mathbf{x}_{it}, \mathbf{m}_{it}) = \lambda_0 + k(\mathbf{x}_{it}, \mathbf{m}_{it})\lambda_1$. Simple substitutions and a straightforward choice of instruments

$$\mathbf{Z}_{it} = \begin{pmatrix} (1, \mathbf{x}_{it}, \mathbf{w}_{it}, k(\mathbf{x}_{it}, \mathbf{m}_{it})) \\ (1, \mathbf{x}_{it}, \mathbf{w}_{it-1}, k(\mathbf{x}_{it-1}, \mathbf{m}_{it-1})) \end{pmatrix}$$

leads to set the relevant moments like

$$\mathbf{r}_{it}(\theta) = \begin{pmatrix} r_{it1}(\theta) \\ r_{it2}(\theta) \end{pmatrix} = \begin{pmatrix} y_{it} - \zeta - \mathbf{w}_{it}\beta - \mathbf{x}_{it}\gamma - k(\mathbf{x}_{it}, \mathbf{m}_{it})\lambda_1 \\ y_{it} - \theta - \mathbf{w}_{it}\beta - \mathbf{x}_{it}\gamma - k(\mathbf{x}_{it-1}, \mathbf{m}_{it-1})\lambda_1 \end{pmatrix}$$

MrEst

All lags of state and free variables are potentially valid instruments in Wooldridge framework, but each additional lag implies the loss of n observations $\Rightarrow$ This is potentially problematic as most dataset used in the relevant literature are characterized by short panels. We propose to complement Wooldridge estimator with dynamic panel instruments à la Blundell-Bond in order to exploit instrument power without losing information.
More specifically, for each i we define the matrix of dynamic panel instruments like

$$\mathbf{Z} = \begin{bmatrix} \mathbf{z}_2' & \mathbf{z}_3' & \cdots & \mathbf{z}_T' & 0 & 0 & 0 & 0 \\ 0 & 0 & \cdots & 0 & \tilde{\mathbf{z}}_3' & 0 & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & \tilde{\mathbf{z}}_4' & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \tilde{\mathbf{z}}_T' \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$$

prodest

prodest is a Stata - and R - module for production function estimation using the Control Function Approach.

It is able to estimate all above-mentioned models in a unique framework and it is currently the unique module implementing Wooldridge and MrEst estimators.

It is faster than other existing modules thanks to the GMM estimation of the second stage.

Syntax

Prodest

```
prodest depvar [if exp] [in range] , free(varlist)
proxy(varlist) state(varlist) method(name) [valueadded
control(varlist) acf id(varname) t(varname) reps(#) level(#)
poly(#) seed(#) fsresidual(newname) endogenous(varlist) opt
_options]
```

Predict

```
predict newvarname [if exp], [residuals exponential
parameters]
```

Nice Options

Prodest

- `control(varlist)` control variable(s) to be included
- `endogenous(varlist)` endogenous variable(s) to be included
- `attrition` correct for attrition - i.e. firm exit - in the data
- `poly(#)` degree of polynomial approximation for the first stage
- `fsresiduals(newvarname)` store the first stage residuals (OP and LP only) in *newvarname*
- `translog` use a translog production function for estimation
- `optimizer` available optimizers are Nelder Mead (nm), modified Newton-Raphson (nr), Davidon-Fletcher-Powell (dfp), Broyden-Fletcher-Goldfarb-Shanno (bfgs) and Berndt-Hall-Hall-Hausman (bhhh)

OP - comparison

Table: Olley-Pakes (1996) confront: Chile value added

	OLS	FE	Levpet	Prodest	Opreg	Prodest_exit
main						
β_k	0.116***	0.0828***	0.402***	0.402***	0.408***	0.398***
	(0.00127)	(0.00126)	(0.00939)	(0.00879)	(0.00798)	(0.0104)
β_{skil}	0.668***	0.458***	0.313***	0.313***	0.313***	0.313***
	(0.00317)	(0.00341)	(0.00734)	(0.00571)	(0.00610)	(0.00571)
β_{unskil}	0.436***	0.339***	0.224***	0.224***	0.224***	0.224***
	(0.00266)	(0.00283)	(0.00643)	(0.00607)	(0.00472)	(0.00607)
time	0.0630	0.582	56.28	55.67	154.6	199.7
N	91598	91598	60253	60253	60253	60253

Column (1) reports results of a linear regression of log output - value added - on free and state variables, in column (2) we add individual fixed effects; column (3) uses the user-written command *levpet* (`levpet va, free(skilled unskilled) capital(k) proxy(inv) reps(50) valueadded`) with *investment* as proxy variable; in column (4) and (6) we perform the same exercise with *prodest* (`prodest va, free(skilled unskilled) state(k) proxy(inv) met(op) valueadded reps(50) [attrition]`), with and without the attrition; lastly, column (5) reports parameter estimates computed by the *opreg* command (`opreg va, exit(exit) free(skilled unskilled) proxy(inv) state(k)`)

LP - comparison

Table: LP (2003) confront: Chile value added

	OLS	FE	Levpet	Prodest	Prodest_exit	Wooldridge
main						
β_k	0.116***	0.0828***	0.146***	0.146***	0.147***	0.149***
	(0.00127)	(0.00126)	(0.00413)	(0.00408)	(0.00439)	(0.00307)
β_{skil}	0.668***	0.458***	0.293***	0.293***	0.293***	0.288***
	(0.00317)	(0.00341)	(0.00604)	(0.00551)	(0.00551)	(0.00400)
β_{unskil}	0.436***	0.339***	0.179***	0.179***	0.179***	0.194***
	(0.00266)	(0.00283)	(0.00490)	(0.00513)	(0.00513)	(0.00337)
time	0.0630	0.540	110.4	73.06	437.5	219.7
N	91598	91598	91598	91598	91598	69376

Column (1) reports results of a linear regression of log output - value added - on free and state variables, in column (2) we add individual fixed effects; column (3) reports results using the user-written command *levpet* (`levpet va, free(skilled unskilled) capital(k) proxy(water ele) reps(50) valueadded`) with *investment* as proxy variable; in column (4) and (5) we perform the same exercise with *prodest* (`prodest va, free(skilled unskilled) state(k) proxy(water ele) met(lp) valueadded reps(50) [attrition]`), with and without the attrition; at last, column (6) reports the estimation with *prodest* using the Wooldridge method with a second order polynomial (`prodest va, free(skilled unskilled) state(k) proxy(water ele) poly(2) met(wrdg) valueadded`).

ACF - comparison

Table: ACF (2015) comparison: Chilean dataset

	GO			VA - II			VA		
	LP	ACFest	Prodest	LP	ACFest	Prodest	LP	ACFest	Prodest
β_{skil}	0.268***	1.991***	0.427***	0.322***	-0.147***	0.701***	0.309***	-0.212***	0.702***
	(0.006)	(0.380)	(0.005)	(0.006)	(0.040)	(0.008)	(0.006)	(0.042)	(0.007)
β_{unskil}	0.160***	-0.528***	0.279***	0.210***	-0.089***	0.467***	0.192***	-0.161***	0.467***
	(0.006)	(0.185)	(0.006)	(0.005)	(0.032)	(0.002)	(0.005)	(0.037)	(0.002)
β_k	0.073***	0.069***	0.039***	0.139***	0.252***	0.060***	0.143***	0.269***	0.057***
	(0.003)	(0.011)	(0.003)	(0.004)	(0.008)	(0.004)	(0.004)	(0.008)	(0.004)
time	140	792	415	85	234	330	93	294	297
N	93,191	71,369	93,191	91,598	70,238	91,598	91,598	70,238	91,598

In colums (1)-(3) the dependent variable is *log(gross output)* - GO - in (4)-(9) is *log(value added)* - VA. (1), (4) and (7) report the benchmark Levinsohn-Petrin estimates; (2), (5) and (8) report results obtained on Chilean data using the user-written command *acfest* with 50 bootstrap repetitions (`acfest [go/va], free(skilled unskilled) proxy(ele) state(k) nbs(50) robust [va] [second]`), whereas columns (3), (6) and (9) refer to the same models estimated with *prodest* (`prodest [go/va], free(skilled unskilled) proxy(ele) state(k) acf reps(50) [va] [poly(2)]`). Value added models have been estimated with a second degree - columns (4)-(6) - and third-degree polynomials - columns (7)-(9).

Wooldridge and MrEst

Table: Bias and Mean Squared Error - DGP 2

Panel (a): Levinsohn-Petrin

	$\hat{\beta}_{sk}$	$\hat{\beta}_{unsk}$	$\hat{\beta}_{k}$	MSE
Levinsohn-Petrin	0.303 (0.121)	0.228 (0.086)	0.039 (0.045)	0.000 (0.000)

Panel (b): Wrdg and MrEst: Bias + MSE

	$Bias_{sk}$	$Bias_{unsk}$	$Bias_{k}$	MSE
Wooldridge	-0.080 (0.097)	-0.062 (0.079)	-0.001 (0.016)	0.009 (0.013)
MrEst - 2 lags	-0.024 (0.042)	-0.018 (0.034)	0.001 (0.018)	0.001 (0.002)
MrEst - 3 lags	-0.025 (0.040)	-0.016 (0.034)	0.002 (0.018)	0.001 (0.001)

in panel (a) we report the average $\hat{\beta}$ value of Levinsohn and Petrin estimation on 60 subsets (i.e. industry sectors, according to the CIIU2 variable) of Chilean firm-level data. These are the benchmark values: we define $Bias_j = \hat{\beta}_j - \beta_j^{lp}, \forall j \in [sk, unsk, k]$ and $MSE = \mathrm{E}(Bias_j^2)$. Panel (b) reports the average bias and the MSE, with their standard deviations, of Wooldridge and MrEst models (various lags).

Table: MrEst - MSE with simulated data (DGP2)

Panel (a): $n \rightarrow \infty$, fixed T

	(1)	(2)	(3)	(4)	(5)	(6)
MrEst - 2 lags	2.439	2.196	2.354	1.847	1.797	1.729
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
MrEst - 3 lags	2.370	2.178	2.344	1.836	1.787	1.719
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
MrEst - 4 lags	2.304	2.151	2.329	1.824	1.777	1.709
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
N	1500	3000	5000	6500	8000	10000

Panel (b): increasing T, fixed n

	(1)	(2)	(3)	(4)	(5)	(6)
MrEst - 2 lags	2.795	3.100	2.730	2.431	2.529	2.354
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
MrEst - 3 lags	2.799	3.101	2.729	2.428	2.518	2.344
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
MrEst - 4 lags	2.797	3.097	2.719	2.420	2.508	2.329
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
N	2500	3000	3500	4000	4500	5000

MSE of MrEst with 2,3 and 4 lags on simulated data - DGP3, no measurement error - with increasing number of firms in the sample and T = 10 fixed.