

# **A seasonal unit root test with STATA**

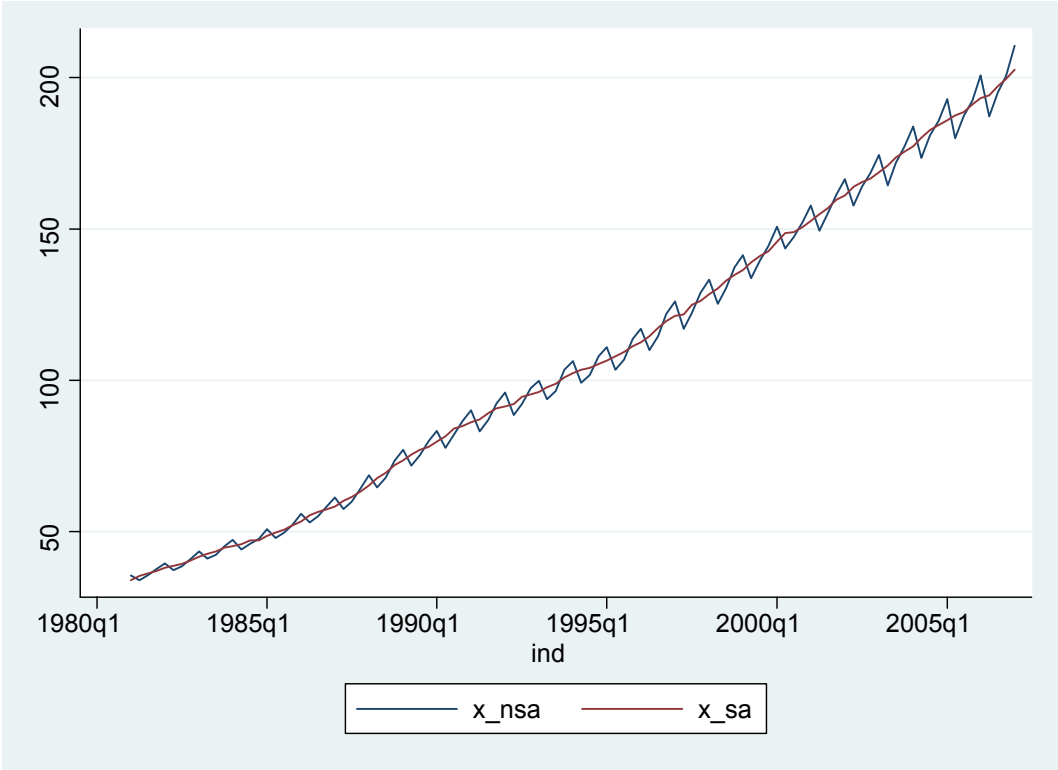
Domenico Depalo  
Tor Vergata University

Milan, October 21<sup>st</sup>, 2008

# 1 Introduction

- Seasonality matters
- For long time econometricians worked with seasonally adjusted data **BUT**
- This practice is sub-optimal for at least two reasons:
  - peak and trough convey information that would be lost during the adjustment
  - in many cases it is in contrast with the economic theory

Figure 1: Seasonality matters



- Solutions can be either
  - use the original data and either control for a set of seasonal dummies
  - redefine the error term to incorporate the seasonal fluctuations
- In fact these are **apparent** solutions, hence
- Seasonality requires a serious systematic approach

---

## 2 Unit roots at various frequencies

- Despite general agreement on the importance of seasonality, less so is for its treatment
- Several way to handle seasonality, each implicitly making **different assumptions**, are
  - a **purely deterministic** seasonal process

$$y = \mathbf{x}\beta + \sum_{i=1}^3 s_i D_i \quad (1)$$

- a **stationary** seasonal process

$$\phi(L)y_t = \epsilon_t \quad (2)$$

$$\phi(L) = (1 - \phi_4 L^4)$$

- an **integrated** seasonal process

$$\begin{aligned} (1 - L^4) y_t &= \epsilon_t \\ &= (1 - L)(1 + L)(1 + L^2) \end{aligned} \quad (3)$$

A general expression for seasonal processes combines cases *i*), *ii*), *iii*) and is compactly represented by

$$d(L)a(L)(y_t - \mu_t) = \epsilon_t,$$

where

- roots of  $a(L) = 0$  lie outside the unit circle,
- roots of  $d(L) = 0$  lie on the unit circle
- $\mu_t = x\beta + \sum_{i=1}^3 s_i D_i$ .

**Thus** stationary components of  $y$  are in  $a(L)$ , while deterministic seasonality is in  $\mu_t$  when there are no seasonal unit roots in  $d(L)$ . Hylleberg, Engle, Granger, Yoo (1993) propose a test to detect seasonal unit roots at different seasonal frequencies, as well as at zero frequency.

Consider the *Lagrangian approximation*:

Any possibly infinite or rational polynomial  $\phi(L)$  which is finite valued at the distinct, nonzero, possibly complex points  $\theta_1, \dots, \theta_p$ , can be expressed in terms of elementary polynomial and a remainder, as follows:

$$\phi(L) = \sum_{k=1}^p \lambda_k \Delta(B) / \delta_k(B) + \Delta(B) \phi^{**}(L)$$

where  $\lambda_k$  are a set of constant,  $\phi^{**}(L)$  is possibly infinite or rational polynomial and  $\Delta(B)$  and  $\delta_k(B)$  are coefficients.

The test is based on

$$\phi(L)y_{4,t} = \pi_1 y_{1,t-1} + \pi_2 y_{2,t-1} + \pi_3 y_{3,t-2} + \pi_4 y_{3,t-1} + \epsilon_t, \quad (4)$$

where

$$\begin{aligned} y_{1,t} &= (1 + L + L^2 + L^3) y_t \\ y_{2,t} &= -(1 - L + L^2 - L^3) y_t \\ y_{3,t} &= (1 - L^2) y_t \\ y_{4,t} &= (1 - L^4) y_t \end{aligned}$$

and  $\pi_i$ s are coefficients for seasonal roots.

Notice, for example that

$$\begin{aligned} y_{1,t} &= (1 + L + L^2 + L^3) y_t \\ &= (1 - L)(1 + L + L^2 + L^3) y_t \end{aligned} \quad (5)$$

$$= (1 - L^4) y_t \quad (6)$$

$$(7)$$



- at root  $1 - L$  the test is on coefficient  $\pi_1 = 0$
- at seasonal root  $1 + L$  the test is on coefficient  $\pi_2 = 0$
- at seasonal roots  $1 + L^2$  the test is joint on coefficients  $\pi_3 = \pi_4 = 0$ .

A natural extension of the seasonal unit root is **seasonal cointegration and seasonal VECM**. One simple method is in Engle and Granger (1993) and is in two step:

1. estimation of the linear combination(s) of levels on data transformed
2. plug the Error Correction in the second step.

### 3 The `root` command

The increasing interest of Stata towards time series analysis is a great opportunity to widen the spectrum of its users. This is the main motivation for this paper. Thanks to the simplicity of data management, the proposed program makes extensive use of Stata's routines for lag operators and command `regress`. The program is:

```
root varname [ if ] [ in ] [, noconstant trend regress lags(#) season(varlist) generate(string)
]
```

**noconstant** suppresses the constant term (intercept) in the model and indicates that the process under the null hypothesis is a random walk without drift. opt `noconstant` cannot be used with the opt `trend` or opt `season(varlist)` options.

**trend** specifies that a trend term be included in the associated regression and that the process under the null hypothesis is a random walk, perhaps with drift. This option may not be used with the **noconstant** option.

**season(varlist)** indicates that the process under the null hypothesis is a random walk augmented for seasonal dummies. It is possible that `varlist` contains only one word (in which case the program builds the dummies) or that `varlist` contains the full set of dummies in which case the program drops the last quarter (due to multicollinearity). This option may not be used with the **noconstant** option.

**regress** specifies that the associated regression table appear in the output. By default, the regression table is not produced.

**lags(#)** specifies the number of lagged difference terms to include in the covariate list.

**generate(string)** generates a set of variables adjusted for seasonal filtering.

## 4 Why a new command?

There exists a routine in Stata (`hegy4`) that performs the HEGY test, however:

- `hegy4` runs a sequential test for the proper number of lags. A Montecarlo experiment reveals that sometimes it is in fact inappropriate. Its performance increases with sample size and with the absolute values of coefficients. However,
  - as either  $\beta_1$  and/or  $\beta_2$  approach, but are different from zero, the practice is questionable
  - Undoubtable advantages are when  $\beta_1$  and/or  $\beta_2$  are indeed zero.

- It is possible to study seasonal cointegration starting from transformed variables. The interested reader is referred to that paper for further details, but for what matters here we can build stationary combinations from transformed variables in levels and study a VAR augmented for these components in a seasonal error correction model (further explained in the next section). We view this as a key distinction to push efforts towards a new command, i.e. `sroot`.
- Consider estimating the cointegrating vector

$$x_{1,t} = Ax_{2,t} + u_t$$

- The presence of seasonal unit roots causes that  $A$  may not be consistently estimated. It is because if  $x_{1,t}$  and  $x_{2,t}$  are cointegrated at both zero and seasonal frequency, with cointegrating vector  $\alpha_1$  and  $\alpha_s$ , with  $\alpha_1 \neq \alpha_s$ , it is unclear which value of  $A$  will be estimated by classical approach.
- A very minor difference is the option `regress`.

## 5 Example

In this section we use the command `sroot` to test for the presence of the unit root at seasonal frequency for the series of consumption in UK. The data are from the National Institute of Statistics for the year 1955-2006, on quarterly basis. We first test for the presence of unit root for NSA data:

```
. sroot x_nsa,lag(4) trend season(quarter) regress gen(pi1 pi2 pi3)
```

HEGY test for SEASONAL unit roots				Number of obs	=	200
Test	1% Critical	5% Critical	10% Critical			
Statistic	Value	Value	Value			
Z(t) - Fr 0	0.330	-4.050	-3.490	-3.180		
Z(t) - Fr 1/2	2.739	-3.520	-2.910	-2.600		
Z(t) - L. Ann.	1.014	-4.040	-3.410	-3.100		
Z(t) - Annual	-0.023	-2.650	-1.920	-1.480		
Joint Annual	0.514	8.960	6.570	5.560		
All SEAS. fr.	2.842	0.000	5.100	5.890		
All freq.	2.166	0.000	5.610	6.380		

x_nsa	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
x_nsa						
Freq.0	.0001847	.0005594	0.33	0.742	-.0009189	.0012882
Freq.1/2	.0461743	.0168582	2.74	0.007	.0129175	.0794311
L. Annual	.0154886	.0152743	1.01	0.312	-.0146435	.0456206
Annual	-.0003434	.0152415	-0.02	0.982	-.0304109	.0297241
LD.	.7711865	.0777132	9.92	0.000	.6178793	.9244938
L2D.	.0894235	.0952005	0.94	0.349	-.0983814	.2772285
L3D.	.1412389	.0953463	1.48	0.140	-.0468536	.3293314
L4D.	-.2178033	.0753951	-2.89	0.004	-.3665376	-.069069
_trend	.0117578	.0031245	3.76	0.000	.005594	.0179216
Q1	-.1432412	.1811381	-0.79	0.430	-.5005779	.2140954

---

Q2	-.0581299	.1837385	-0.32	0.752	-.4205965	.3043368
Q3	.1162713	.1815115	0.64	0.523	-.2418021	.4743447
_cons	-.3112752	.1896996	-1.64	0.103	-.6855015	.0629511

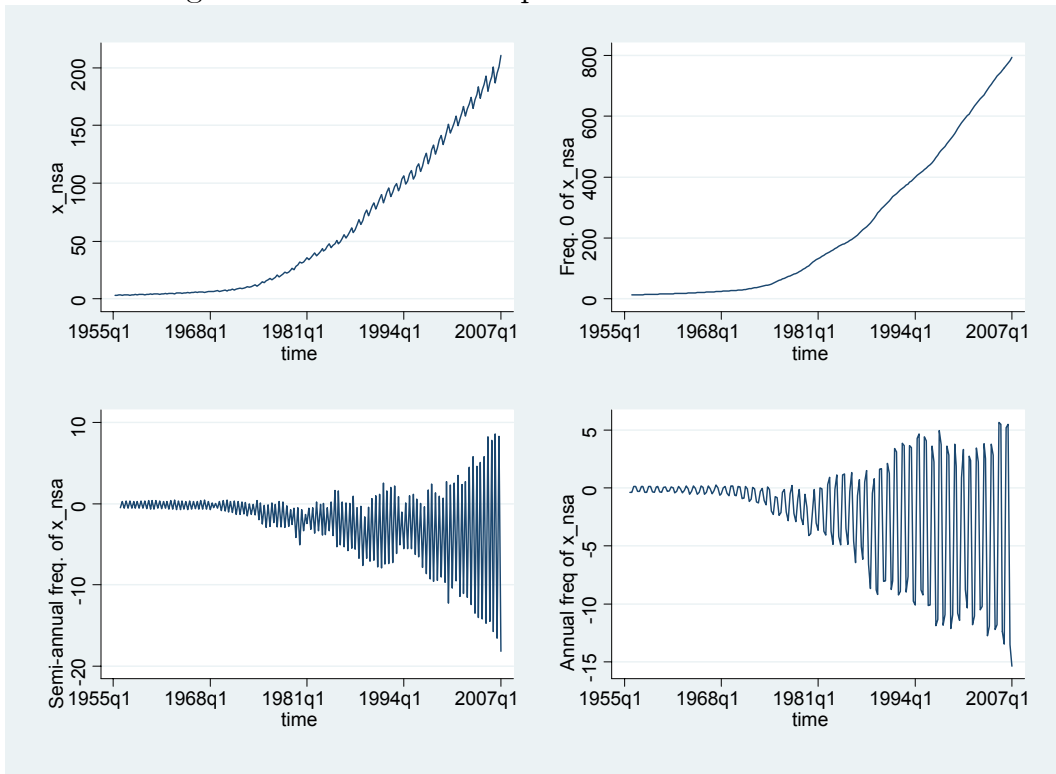
---



- According to the t statistics from Frequency 0, we do not reject that  $\pi_1$  is different from zero at conventional confidence level  $\Rightarrow$  we cannot reject that the time series has a unit root at frequency zero.
- According to Section 2, test statistic and critical values for this frequency could have been obtained from those already tabulated from the Dickey-Fuller test.
- For frequency  $\pi/2$  we cannot reject the presence of a (seasonal) unit root. The t-statistic is 2.739 against a critical value of -2.910 at 5% confidence level.
- Again, the t-statistics could be obtained from a regression involving the transformation of the data as in  $y_2$ .
- For remaining frequencies the results should be read properly. From Section 2 we know that results depend on the **joint test** on coefficients. With this model specification, we cannot reject the unit root at annual frequency based on line **Joint Annual**.
- From line **All SEAS. fr.** we do not reject the joint significance of seasonal unit roots
- From line **All freq.** the joint significance of the full set of unit roots, at seasonal and nonseasonal frequencies.

- The actual existence of seasonal unit roots in the series of consumption sheds more light on the potentially dramatic impact that a sub-optimal econometric technique can have on a test of an economic theory. In this sense, the results from `sroot` are important *per se*. However we can go a step further.
- The option `generate` is crucial to consider one possible extension of the unit root at seasonal frequencies, namely **cointegration at seasonal frequencies**.
- We plot the transformed series in Figure 2 as they are generated by `sroot`.

Figure 2: Series of consumption and its transformations



As a practical example, consider consumption ( $\mathbf{x\_nsa}$ ) and income ( $\mathbf{y\_nsa}$ ). To estimate seasonal cointegration among these variables at frequencies 0 and 1/2, as is often the case of interest, a researcher needs the following lines of commands:

```

qui sroot x_nsa,lag(4) gen(pi1x pi2x pi3x)
qui sroot y_nsa,lag(4) gen(pi1y pi2y pi3y)
qui reg pi1x pi1y
qui predict res1 // coint. vec. at freq. 0
qui reg pi2x pi2y
qui predict res2 // coint. vec. at freq. 1/2

```

where  $\mathbf{res1}$  and  $\mathbf{res2}$  are stationary cointegrating vectors at specified frequencies, to form a seasonal VECM of

$$\Delta_4 \mathbf{Y}_t = \mathbf{A}_1 \mathbf{res1}_{t-1} + \mathbf{A}_2 \mathbf{res2}_{t-1} + \sum_{i=1}^{p-4} \Gamma(\mathbf{i}) \Delta_4 \mathbf{Y}_{t-i}$$

and  $\Delta_4 \mathbf{Y}_t = (\Delta_4 \mathbf{x\_nsa}_t, \Delta_4 \mathbf{y\_nsa}_t)'$ ,  $\mathbf{A}_i$  vectors of adjustment parameters and  $\Gamma(\mathbf{i})$  a  $2 \times 2$  matrix of parameters.

## 6 Conclusion

- The new command `sroot`, which implements a test to detect unit roots at frequencies other than zero, in quarterly data.
- Although the usual practice is to work with seasonally adjusted data, we view this as a weak solution  $\Rightarrow$  one should always consider using seasonally unadjusted data
- Cointegration at seasonal frequencies can be studied through the option `generate` in `sroot`, even though more efficient methods are available in the literature.
- **!!!THANK YOU for being patient!!!**