Initial Conditions and Stationarity Tests

Fabio Busetti Bank of Italy, Research Department

July 2, 2009

Abstract

It is showed that the magnitude of the initial condition of a highly persistent process affects the properties of stationarity tests applied to the first differences, with a "relatively large" initial condition implying non-negligible oversizing. The finding is important for time series analysis of convergence; one implication is that the properties of both levels and first differences of the series should be analysed. An example with price indices and inflation rates is provided.

KEYWORDS: Convergence, Near unit root, Stability.

Keywords: C12, C22

1. Introduction

Several papers have addressed the effect of initial conditions on the properties of unit root tests. It has been found that when the data are highly persistent (in the sense of near unit root) the power properties of various unit root tests are deeply affected by the magnitude of the initial condition; see Müller and Elliott (2003) for a unified treatment.

An alternative approach to unit root inference is the so-called KPSS stationarity test (Kwiatkowski *et al.*, 1992): the null hypothesis is weak dependence and the test is consistent against the unit root alternative.[1] Müller (2005) shows that,

[1] The KPSS test generalizes to the case of weak dependence a statistic previously proposed for i.i.d. series. Gardner (1969) obtains the statistic for detecting a change in the mean under a uniform prior on the time of the break. Nyblom and Makelainen (1983) shows that the statistic provides a locally optimal test against a Gaussian random walk plus noise model (i.e. a unit root process), while Nyblom (1989) considers a more general alternative hypothesis of parameter variation in the form of a martingale. Furthermore, Lee and Schmidt (1996) show that the KPSS test is consistent also against a stationary long memory process.

under a near unit root data generating process, the probability that KPSS rejects the null hypothesis tends to one asymptotically; this occurs independently of the size of the initial condition.

Despite taking first differences largely offsets the near unit root, this paper shows that the magnitude of the initial condition of a highly persistent (level) process affects the properties of the KPSS stationarity test applied to first differences. In particular, a "relatively large" initial condition implies non negligible oversizing and it does provide evidence against stationarity of first differences. This finding appears important for time series analysis of convergence, e.g. when stationarity and unit root tests are applied to inflation differentials across geographical areas (as in Busetti, Fabiani and Harvey, 2006). An empirical illustration is provided.

2. The main result

Consider the near unit root data generating process

$$y_t = d_t + u_t \quad (2.1)$$

$$u_t = \rho_T u_{t-1} + \varepsilon_t, \quad (2.2)$$

$$\rho_T = \exp(c/T), \quad (2.3)$$

where d_t is a deterministic component and ε_t is a zero mean sequence satisfying a Functional Central Limit Theorem with (strictly positive) long run variance ω_ε^2. As $\rho_T = 1 + c/T + o(T^{-1})$, it follows that y_t is a highly persistent process for $c < 0$ and a unit root process for $c = 0$. As in Müller and Elliott (2003), it is assumed that

$$u_0 = \omega_\varepsilon M T^{\frac{1}{2}} \quad (2.4)$$

for some constant M, which represents the magnitude of the initial condition of the series. Making u_0 of order $T^{\frac{1}{2}}$ ensures that it does not vanish asymptotically, allowing to study the properties of the tests under both "small" and "large" initial conditions.

If $d_t = \beta_0$, a constant level, Δy_t has zero mean and thus the appropriate test statistic of stationarity[2] for first differences is

$$\zeta_0 = \frac{\sum_{t=1}^{T}\left(\sum_{j=1}^{t}\Delta y_j\right)^2}{T^2\widehat{\omega}_{\Delta y}^2},$$

where $\widehat{\omega}_{\Delta y}^2 = \sum_{j=-m}^{m} w_j\hat{\gamma}_j$, is an estimate of the long run variance of first differences, with $\widehat{\gamma}_j$ being the sample autocovariance of first differences at lag j and m a bandwidth parameter such that, as $T \to \infty$, $m \to \infty$ and $m/T \to 0$; a common choice for the autocovariance weights is the Bartlett kernel $w_j = 1 - |j|/(m+1)$.

If there is a linear trend in the levels, i.e. $d_t = \beta_0 + \beta_1 t$, then first differences should be demeaned and the appropriate statistic for stationarity is

$$\zeta_1 = \frac{\sum_{t=1}^{T}\left(\sum_{j=1}^{t}(\Delta y_j - \overline{\Delta y})\right)^2}{T^2\widehat{\omega}_{\Delta y}^2}.$$

Proposition 2.1. *Let the data be generated by (2.1)-(2.4). (a) If $d_t = \beta_0$ then*

$$\zeta_0 \xrightarrow{d} \int_0^1 K_0\left(r; c, M\right)^2 dr, \tag{2.5}$$

where $K_0\left(r; c, M\right) = W(r) + c\int_0^r e^{(r-s)c}W(s)ds + M\left(\exp(cr) - 1\right)$, and $W(r)$ is a standard Wiener process. (b) If $d_t = \beta_0 + \beta_1 t$ then

$$\zeta_1 \xrightarrow{d} \int_0^1 K_1\left(r; c, M\right)^2 dr, \tag{2.6}$$

where $K_1\left(r; c, M\right) = K_0\left(r; c, M\right) - rK_0\left(1; c, M\right)$.

The proposition is proved in the appendix. Note that for $c = 0$ (irrespectively of M) the two stochastic processes $K_0\left(.; c, M\right)$, $K_1\left(.; c, M\right)$ reduce to, respectively, a Brownian motion and a Brownian bridge, thus providing the same limiting representation as in Kwiatkowski et al. (1992).[3]

[2]The statistic ζ_0 corresponds to the KPPS test of Kwiatkoski et al. (2001) but without demeaning (or detrending) of the series.

[3]The stochastic process $K_0\left(r; c, M\right)$ contains the extra term $M\left(\exp(cr) - 1\right)$, not present in the usual asymptotic framework. This induces size distortion if the series in level is a highly persistent process, in the sense of a near unit root ($c < 0$ with a nonzero initial condition).

For a given significance level α, as $T \to \infty$ the probability of rejecting the null hypothesis of stationarity (the asymptotic size) is given by

$$\pi_{j,\alpha}(c, M) = \Pr\left\{ \int_0^1 K_j(r; c, M)^2 \, dr > q_{j,1-\alpha} \right\}, \qquad j = 1, 2,$$

where $q_{j,1-\alpha}$ is the $(1-\alpha)$ quantiles of $\int_0^1 K_j(.; 0, M)^2 \, dr$; e.g. for tests run at 5% significance level $q_{0,0.95} = 1.656$, $q_{1,0.95} = 0.461$.The size distortion depends on the two parameters c and M, with $\pi_{j,\alpha}(0, M) = \alpha$ for every M.

For different values of the persistence parameter c, the amount of distortion $\pi_{j,\alpha}(c, M)$ is graphed in figure 1,2 for $j = 0, 1$ respectively and $\alpha = 5\%$. For the case $d_t = \beta_0$, it is seen that the stationarity test is oversized for values of the initial condition $M > 1.2$ (provided of course $c < 0$); the degree of oversizing then increases quite dramatically with M if $c \leq -2$. The case of linear trend in the level (figure 2) displays less distortion, but still non-negligible for $c \leq -2$. Additional unreported simulations show that the limiting behaviors graphed in figures 1,2 provide reasonably good approximations in finite samples. The choice of bandwidth parameter appears to make, in general, little difference if a standard rule as in Kwiatowski et al. (1992, p.169) is used; however if persistence is relatively low choosing higher values of m helps reducing distortion.

3. Empirical example and discussion

Figure 3 shows the difference of the logarithm of Consumer Price Index between two Italian regional capitals, Milan and Ancona, for the period 1970M1-2002M12. The base year is 2002, so by construction the final year average of the series is equal to zero. The picture shows high persistence and a large initial condition. The p-value of the standard ADF test (with a constant term included in the regression) is equal to .06, while the test τ^* of Busetti, Fabiani and Harvey (2006, p.866) provides stronger evidence against unit root with a p-value of 0.01; there is therefore evidence of converging behavior in the (log) relative prices between the cities: this is convergence in the levels.

But what about first differences, i.e. inflation rates? The application of stationarity tests to the first differences of the (log) relative prices (the so-called inflation differential) provides evidence of instability: ξ_0 rejects the null of stationarity at 1% significance for values of the truncation parameter $m \geq 6$, while ξ_1 rejects at 5%. This is entirely due to the distortions induced by a large initial condition. In fact, the first value of the series is equal to 0.111 which corresponds to a value of M between 2.1 and 2.3 (using the estimates of long run variance); figure 1 shows that a lot of over-sizing can be expected in this range of values of the initial condition.

Note however that unit root tests applied to the inflation differential (correctly) provide a sound rejection of their null hypothesis: the apparent conflict in the outcomes of unit root and stationarity tests is therefore fully explained. Intuitively, during the convergence process in relative prices, inflation rates have been, on average, consistently higher in the smaller city of Ancona (at least until the late 80's, as can be seen from the figure), which is the reason underlying rejection of stationarity tests. The main message is that, to avoid misinterpretation of time series analyses of convergence, one should combine evidence of both unit root and stationarity tests in levels and first differences, taking into account of the possible effects of initial conditions on both tests.

References

[1] Busetti, F., Fabiani, S. and A.C. Harvey (2006), Convergence of prices and rates of inflation, *Oxford Bulletin of Economics and Statistics* 68, 863-877

[2] Kwiatkowski, D., Phillips, P.C.B., Schmidt, P. and Y. Shin (1992), Testing the null hypothesis of stationarity against the alternative of a unit root: how sure are we that economic time series have a unit root?, *Journal of Econometrics* 44, 159-178.

[3] Gardner, L.A. (1969), On detecting changes in the mean of normal variates, *Annals of Mathematical Statistics* 40, 116-126.

[4] Lee, D. and P. Schmidt (1996), On the power of the KPSS test against fractionally integrated alternatives, *Journal of Econometrics* 73, 285-302.

[5] Müller, U.K. (2005), Size and power of tests of stationarity in highly autocorrelated time series, *Journal of Econometrics*, Vol. 128, pp. 195–213.

[6] Müller, U.K. and G. Elliott (2003), Tests for unit roots and the initial condition, *Econometrica* 71, 1269-1286.

[7] Nyblom, J. and T. Mäkeläinen (1983), Comparison of tests for the presence of random walk coefficients in a simple linear model, *Journal of the American Statistical Association* 78, 856-864.

[8] Nyblom, J. (1989), Testing for the constancy of parameters over time, *Journal of the American Statistical Association* 84, 223-230.

[9] Phillips, P.C.B. (1987), Towards a unified asymptotic theory for autoregression, *Biometrika* 74, 535-547.

APPENDIX: PROOF

Rewrite the data generating process as $u_t = u_0\rho_T^t + \sum_{j=1}^{t} \rho_T^{t-j}\varepsilon_j$. Then for the partial sums of first differences we have that, for $r \in [0,1]$,

$$\begin{aligned}
\omega_\varepsilon^{-1}T^{-\frac{1}{2}}\sum_{j=1}^{[Tr]}\Delta u_j &= \omega_\varepsilon^{-1}T^{-\frac{1}{2}}\left(u_{[Tr]} - u_0\right) \\
&= \omega_\varepsilon^{-1}T^{-\frac{1}{2}}\sum_{j=1}^{[Tr]}\rho_T^{[Tr]-j}\varepsilon_j + M\left(\rho_T^{[Tr]} - 1\right) \\
&\Rightarrow \left(W(r) + c\int_0^r e^{(r-s)c}W(s)ds\right) + M\left(\exp(cr) - 1\right) \\
&\equiv K_0\left(r; c, M\right)
\end{aligned}$$

where, by the near unit root asymptotics of Phillips (1987), $\omega_\varepsilon^{-1}T^{-\frac{1}{2}}\sum_{j=1}^{[Tr]}\rho_T^{[Tr]-j}\varepsilon_j \Rightarrow K_0\left(.; c, 0\right)$, the Ornstein-Uhlenbeck process; the notation $\Rightarrow$ stands for weak convergence. Since, as shown below, $\widehat{\omega}_{\Delta y}^2 \xrightarrow{p} \omega_\varepsilon^2$, the limiting representation (2.5) readily follows by the Continuous Mapping Theorem (CMT).

For the case of linear trend in the levels (2.6), we have that

$$\begin{aligned}
\omega_\varepsilon^{-1}T^{-\frac{1}{2}}\sum_{j=1}^{[Tr]}\left(\Delta y_j - \overline{\Delta y}\right) &= \omega_\varepsilon^{-1}T^{-\frac{1}{2}}\sum_{j=1}^{[Tr]}\left(\Delta u_j - \overline{\Delta u}\right) \\
&\Rightarrow K_0\left(r; c, M\right) - rK_0\left(1; c, M\right) \equiv K_1\left(r; c, M\right),
\end{aligned}$$

and thus the result follows again by the CMT.

Finally, the proof that $\widehat{\omega}^2_{\Delta y} \xrightarrow{p} \omega^2_\varepsilon$ follows from (i) Δu_t is asymptotically equivalent to ε_t and (ii) $\widehat{\omega}^2_\varepsilon = \sum_{j=-m}^{m} w_j \left(T^{-1} \sum_{t=|j|+2}^{T} \varepsilon_t \varepsilon_{t-|j|}\right) \xrightarrow{p} \omega^2_\varepsilon$, by the consistency arguments of long-run variance and spectral estimators. For simplicity of notation, we consider just the case $d_t = \beta_0$ and compute the autocovariance of first differences without demeaning, i.e. we define $\widehat{\omega}^2_{\Delta y} = \sum_{j=-m}^{m} w_j \left(T^{-1} \sum_{t=|j|+2}^{T} \Delta y_t \Delta y_{t-|j|}\right)$, where w_j is a standard lag window (such as the Bartlett kernel $w_j = 1 - |j|/(m+1)$), and $m \to \infty$ such that $m/T \to 0$. Since $\Delta y_t = \Delta u_t = cT^{-1}u_{t-1} + \varepsilon_t$ we just need to show that

$$\widehat{\omega}^2_{\Delta y} - \widehat{\omega}^2_\varepsilon = \frac{1}{T} \sum_{j=-m}^{m} w_j \left(A_{T,j} + B_{T,j} + C_{T,j}\right) \xrightarrow{p} 0,$$

where $A_{T,j} = c^2 T^{-2} \sum_{t=|j|+2}^{T} u_{t-1} u_{t-1-|j|}$, $B_{T,j} = cT^{-1} \sum_{t=|j|+2}^{T} u_{t-1} \varepsilon_{t-|j|}$, $C_{T,j} = cT^{-1} \sum_{t=|j|+1}^{T} u_{t-1-|j|} \varepsilon_t$. Since, by Phillips (1987), $T^{-1} \sum_t u_{t-1} \varepsilon_t$ and $T^{-2} \sum_t u_t^2$ are both of $O_p(1)$, by repeated substitutions of the data generating process it is easy to see that, for $j = o(T)$, the three terms $A_{T,j}$, $B_{T,j}$, $C_{T,j}$, are also bounded in probability. Thus $\widehat{\omega}^2_{\Delta y} - \widehat{\omega}^2_\varepsilon \xrightarrow{p} 0$ since m is of smaller order than T.

Figure 1. Actual size of the test ξ_0 against the initial condition *M* and for a range of persistence parameter c.

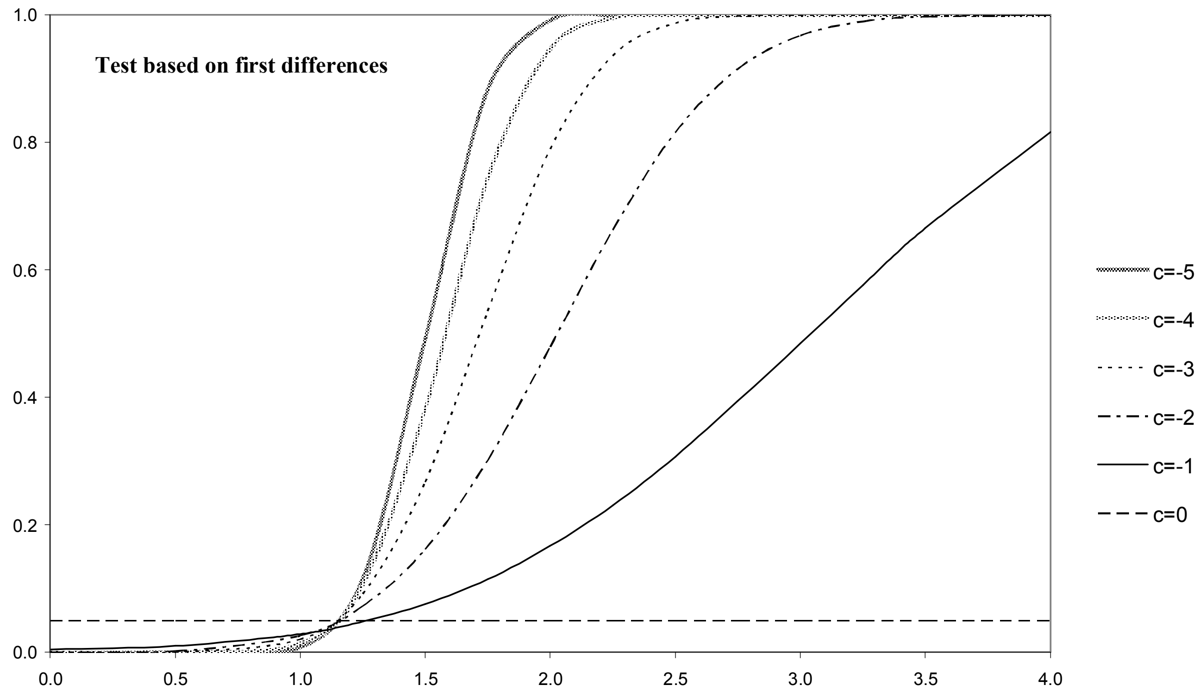

Figure 2. Actual size of the test ξ_1 against the initial condition *M* and for a range of persistence parameter c.

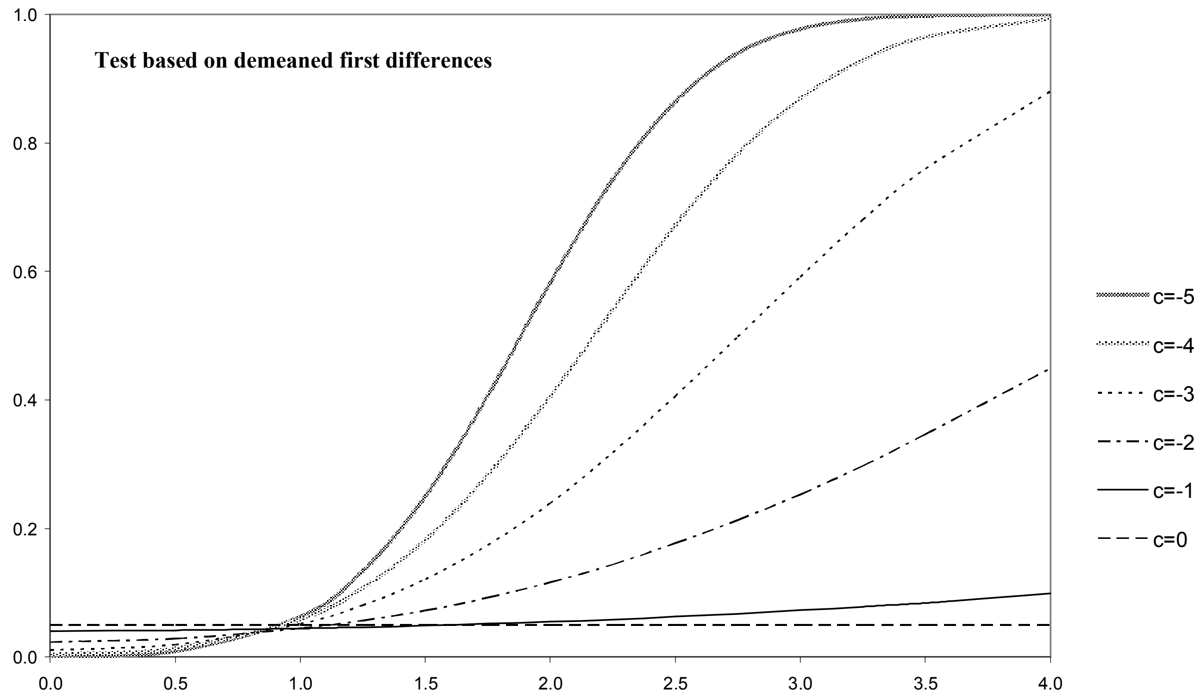

Figure 3. Difference of the logarithm of the Consumer Price Index between Milan and Ancona (base year 2002).