PROPERTIES OF THE ADF UNIT ROOT TEST FOR MODELS WITH TREND AND CYCLES *

Fabrice Barthélémy
GREQAM, Marseille and University of Aix-Marseille II

Michel Lubrano
GREQAM, CNRS, Marseille

December 1994, Revised May 1996

*We would like to thank Mohamed Boutahar, Russell Davidson, James MacKinnon, Pierre Perron, Jean-François Richard and Peter Schotman for useful discussions and comments. Remaining errors are solely ours.

Abstract

A unit root test is usually carried out by using the regression test introduced by Dickey and Fuller (1979). Under the null hypothesis the series should be a random walk. But a non-stationary series can usually be decomposed into a random walk and a stationary component. This is what is done in additive decompositions between a trend and a cycle. The model considered here lies in the class of UC model developed by Harvey (1989) whose reduced form is an ARIMA(0,1,q) model. In this context, testing for a unit root can be compared to testing for a unit root in an ARIMA constrained model with a moving average polynomial. The paper analyses the asymptotic distribution of the Dickey and Fuller (DF) test and of the Augmented Dickey and Fuller (ADF) test under this kind of null hypothesis. It shows graphically the size distortion of the ADF test as a function of the weight of the unit root and of the number of lags. A Monte Carlo experience concludes by showing that usual lag length selection procedure are most of the time not adequate.

Résumé

Un test de racine unitaire s'effectue en général au moyen de la régression de test introduite par Dickey et Fuller (1979). Sous l'hypothèse nulle, la série est censée suivre une marche aléatoire pure. Or, une série non stationnaire se décompose en général sous la forme d'une marche aléatoire à laquelle se rajoute une composante stationnaire. C'est ce qui est fait dans les décompositions additives entre tendance et cycle. Le modèle que l'on considère ici appartient à la classe des modèles à composantes inobservables développés par Harvey (1989) dont la forme réduite est un modèle ARMA(0,1,q). Dans ce contexte, tester la présence d'une racine unitaire revient à tester la présence d'une racine unitaire dans un modèle ARIMA contraint avec un polynôme moyenne mobile. Le papier analyse la distribution asymptotique du test de Dickey et Fuller (DF) et du test augmenté de Dickey et Fuller (ADF) sous ce type d'hypothèse nulle. Il montre graphiquement que la distorsion de taille du test ADF est une fonction de la taille de la racine unitaire et du nombre de retards. Une expérience Monte Carlo conclut en montrant que la procédure habituelle de selection de retards n'est pas valable dans la plupart des cas.

Classification JEL: C11, C12, C22

Key-words: unit root tests, unobserved components, size distortion.

Mots-clés: tests de racines unitaires, composantes inobservables, distorsion de taille.

1 Introduction

The relation between long term economic activity and short term fluctuations depends strongly on how the economy reacts to exogenous shocks. Persistence of shocks in the long term means that the GNP, for instance, contains a unit root and that it can be decomposed into a stochastic trend plus a stationary cycle. This correspondence between the presence of a unit root and of a stochastic trend is formally provided by the Beveridge Nelson decomposition. See their 1981 paper. Starting from the Wold representation of the differenced series $\Delta y_t = C(L)\epsilon_t$, the lag polynomial $C(L)$ can be factorized into $C(L) = C(1) + \Delta C^*(L)$ where $C^*(L)$ has all its roots outside the unit circle[1]. Integrating Δy_t means that $y_t = y_0 + C(1)\sum_{i=1}^t \epsilon_i + C^*(L)\epsilon_t$. In this decomposition, $C^*(L)\epsilon_t$ represent the cycle, $C(1)\sum_{i=1}^t \epsilon_i$ represents the stochastic trend and $C(1)$ is a measure of persistence. In empirical applications based on quarterly data of the US GNP covering the post war period [see for instance Stock and Watson (1988)], the measure of persistence is very often greater than one. This period experienced a very big shock with the 1973 first oil shock which can be thought of having an impact greater than its size. However other explanations could be that the simple Beveridge-Nelson decomposition based on an ARIMA representation is too simple and in fact misspecified first because it ignores the presence of a broken deterministic trend and second because it presumes the presence of a single error term common to the cycle and the stochastic trend. Many other models have been proposed to decompose a series between its permanent and transitory components. They all rely on two errors either considering a single series with the unobserved components models initiated by Harvey (1985) or considering two series with the approach based on structural VAR proposed either by Blanchard and Quah (1989) or by Evans (1989)[2]. In the structural VAR approach, y_t is still I(1) with a stochastic trend, but the additional variable is I(0) and correlated to the cycle. But the marginal distribution of y_t can be shown to have an unobserved component representation. All these models, when applied to the decomposition of the US GNP, have in common that they produce measure of persistence lower than one. When the error terms are independent, they have also in common to produce a reduced form with a negative MA component. We shall detail these assertions in the text.

So y_t is supposed to have a unit root in its AR representation. This assertion must of course be tested. The parametric test proposed by Dickey and Fuller (1981) is certainly the most popular for empirical applications as underlined for instance in Stock (1994). Initially designed for for AR(1) processes, it was extended for AR(p) processes in its ADF version. When the series has an ARMA representation, Said and Dickey (1984) proved that the ADF test has a minimum size distortion provided the number of lags grows at a sufficient speed. However, in Monte Carlo experiments, Schwert (1989) has showed that in small sample, the ADF test presented a serious size distortion when there was a negative

[1]The factorization results from the Euclidian division of $C(L)$ by $1 - L$. See for instance Barthélémy and Lubrano (1998) for a proof or any textbook on time series analysis.

[2]The model proposed by Evans incorporates a broken deterministic trend which appears to be necessary in order to achieve a correct modelling.

MA part in the series. We have motivated above the presence of negative MA part in macroeconomic series and the use of unobserved component models based on two error terms to analyze the business cycle and measure the persistence of exogenous shocks. The central question that will be examined in this paper is what to do with the ADF test in such circumstances.

The paper first details some of the properties of the unobserved component model and shows in section 2 that the reduced form of the model has a negative MA representation. It also shows that the marginal distribution of a bivariate VAR has the form of an unobserved component model. Section 3 derives the asymptotic distribution of the ADF test when the true model is an unobserved component model. In section 4, we simulate that asymptotic distribution and derive an empirical relationship which gives the optimal lag length as a function of the variance ratio in the unobserved component model. We compare this relation to the usual lag length determination methods currently proposed in the literature. We finally propose in section 5 a modification of the usual ADF test which optimally correct for size distortion while keeping a reasonable power.

2 The unobserved component model and its properties

A parsimonious version of the unobserved component model of Harvey (1985) with an I(1) trend can be:

$$\left\{\begin{array}{rcl} y_t & = & \mu_t + c_t \\ \mu_t & = & \mu_{t-1} + \delta + \widetilde{\eta}_t \\ \phi(L)c_t & = & \alpha(L)\,\widetilde{\varepsilon}_t \end{array}\right. \tag{1}$$

The series y_t is formed by the sum of a stochastic trend μ_t and a cycle c_t. The stochastic trend is modeled as a random walk with drift δ and white noise $\widetilde{\eta}_t$ of variance $\sigma^2_{\widetilde{\eta}}$. The stochastic trend is reduced to a deterministic trend of slope δ when $\sigma^2_{\widetilde{\eta}} = 0$. The cycle, which represents fluctuations of the series around its trend, follows an ARMA(p,q) process where $\widetilde{\varepsilon}_t$ is of variance $\sigma^2_{\widetilde{\varepsilon}}$. The two white noises are supposed to be uncorrelated. The roots of the lag polynomials $\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$ and $\alpha(L) = 1 + \alpha_1 L + \cdots + \alpha_q L^q$ are supposed to be outside the unit circle.

This model encompasses the pure random walk and the trend stationary models. When $\sigma^2_{\widetilde{\varepsilon}} = 0$, y_t becomes a pure random walk with drift: $\Delta y_t = \delta + \widetilde{\eta}_t$. When $\sigma^2_{\widetilde{\eta}} = 0$, the series becomes a process with a deterministic trend: $\phi(L)(y_t - y_0 - \delta t) = \alpha(L)\widetilde{\varepsilon}_t$. In order to incorporate in the parameterization the trade off between $\sigma^2_{\widetilde{\varepsilon}} = 0$ and $\sigma^2_{\widetilde{\eta}} = 0$, it is convenient to introduce the following reparameterization:

$$\left\{\begin{array}{lcl} y_t & = & \mu_t + c_t \\ \mu_t & = & \mu_{t-1} + \delta + (1-\theta)\eta_t \\ \phi(L)c_t & = & \alpha(L)\,\theta\varepsilon_t \end{array}\right. \tag{2}$$

where

$$\theta \in [0,1] \quad \tilde{\varepsilon}_t = \theta\varepsilon_t \quad \tilde{\eta}_t = (1-\theta)\eta_t \quad \text{and } \sigma_\eta^2 = \sigma_\varepsilon^2 = \sigma^2 \tag{3}$$

The equality of σ_ε^2 and σ_η^2 is an identification condition. When θ is near to 1, the process is near to the sum of a stationary process and a deterministic trend. When θ tends to 0, the process tends to a pure random walk with drift.

2.1 Reduced form

Let us take y_t in differences and replace the trend and the cycle by their expressions:

$$\phi(L)(\Delta y_t - \delta) = (1-\theta)\,\phi(L)\,\eta_t + \alpha(L)\,\theta\,\Delta\varepsilon_t \tag{4}$$

Model (**??**) represents the variable an ARMA process where the MA part is the weighted sum of an MA(p) and an MA($q+1$). This is called by Harvey (1989) the UCARIMA form of the model which means unobserved component ARIMA to underline the fact that the MA part is constrained to be the sum of two independent MA processes as the sum of the two independent MA is an MA of order Max($p, q+1$) on a new white noise e_t (see Hamilton 1994 p.106). Its coefficients are constrained and can be found equating the autocorrelations of the two processes. In the general case, a solution can be found using the Kalman filter (see Hamilton 1994, p 391). In the simple case where $p = q = 0$, an imediate solution is available. The autocorrelations of an MA(1)

$$\Delta y_t = \delta + e_t + \beta e_{t-1} \tag{5}$$

are $\beta/(1+\beta^2)$ for the first order and zero after. The autocorrelations of (**??**) are $-\theta^2/((1-\theta)^2+2\theta^2)$ for the first order and zero after. Equating these two autocorrelations gives two roots for β. The following root

$$\beta = \frac{-(1-\theta)^2 - 2\theta^2 + (1-\theta)\sqrt{(1-\theta)^2+4\theta^2}}{2\theta^2} \tag{6}$$

is the only one for which the MA (**??**) is invertible. β takes its values in $[-1,0]$ for $\theta \in [0,1]$. This demonstrate on this example that the UCARIMA is a constrained ARMA process.

2.2 VAR and UCARIMA

As underlined in the introduction, an I(1) variable y_t can also be decomposed into a stochastic trend and a stationary cycle if we introduce an extra stationary variable x which is supposed to be correlated with the cycle and represent the joint process of Δy_t, x_t by a bivariate VAR model. We shall now precise the relationship between these two representations. Using the matrix of lag polynomials $A(L)$ of order r, a bivariate VAR is noted:

$$\begin{pmatrix} a_{11}(L) & a_{12}(L) \\ a_{21}(L) & a_{22}(L) \end{pmatrix} \begin{pmatrix} \Delta y_t \\ x_t \end{pmatrix} = \begin{pmatrix} e_{yt} \\ e_{xt} \end{pmatrix} \tag{7}$$

We shall suppose that the system is stationary and invertible which means that $\mathrm{Det}(A(L)) \neq 0$. The marginal distribution of Δy_t is found by solving the system which gives:

$$\mathrm{Det}(A(L))\Delta y_t = a_{22}(L)e_{yt} - a_{12}(L)e_{xt} \tag{8}$$

We can identify the coefficients of this ARMA(2r,r) process with those of the UCARIMA presented in (**??**) only if the coefficients of the VAR are constrained. This can easily be seen on an example. Let us consider the UCARIMA(1,2) (**??**) obtained with $p = q = 1$. A VAR with one lag does not manage to encompass (**??**). Two lags are necessary for the MA part of the marginal in Δy_t, but constraints on $A(L)$ must be introduced to reduce the degree of the AR part of that marginal. A possible set of restrictions is for instance:

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} a_{11,1} & a_{12,1} \\ 0 & 0 \end{pmatrix} L + \begin{pmatrix} 0 & a_{12,2} \\ 0 & 0 \end{pmatrix} L^2 \tag{9}$$

which gives:

$$(1 + a_{11,1}L)\Delta y_t = e_{yt} - (a_{12,1}L + a_{12,2}L^2)e_{xt} \tag{10}$$

This is an ARMA(1,2) provided e_{yt} and e_{xt} are correlated, which is the case in VAR models.

2.3 Further details for future computations

To compute some of our asymptotic distributions, we shall use another form of the unobserved component model, introducing explicitly the initial conditions. If we set $y_0 = \mu_0 + c_0$, we have:

$$\phi(L)(y_t - y_0 - \delta t) = (1 - \theta)\phi(L)\sum_{i=0}^{t-1} \eta_{t-i} + \theta\alpha(L)\varepsilon_t \tag{11}$$

or alternatively that:

$$\phi(L)(y_t - \mu_t) = \theta\alpha(L)\varepsilon_t \tag{12}$$

with

$$\mu_t = \mu_0 + \delta t + (1 - \theta)\sum_{i=0}^{t-1} \eta_{t-i} \tag{13}$$

This means that the deviation of y_t from a deterministic trend is equal to the weighted sum between random walk and an infinite invertible moving average.

3 Properties of the DF test when the true DGP is the structural model

The object of this section is to determine the asymptotic distribution of the Dickey-Fuller test when the true generating model is the unobserved component model (**??**). We shall show in particular how it differs from the usual result. We know that asymptotic distributions are dependent on the form of the deterministic terms. But for simplicity we shall consider only the case without constant or trend. In the next section, we shall present the same type of result for the Augmented Dickey-Fuller test.

The usual Dickey-Fuller test assumes that the test regression and the true DGP are the same with for instance:

$$(1-\rho L)(y_t-\mu-\delta t)=\eta_t \tag{14}$$

Solving in y_t this recurrence equation gives:

$$y_t=\mu+\delta t+\sum_{i=0}^{t-1}\rho^i\eta_{t-i}+\rho^t\eta_0 \tag{15}$$

which under the null, reduces to:

$$y_t=y_0+\delta t+\sum_{i=0}^{t-1}\eta_{t-i} \tag{16}$$

Expression (**??**) is used to determine the asymptotic distribution of the usual Dickey-Fuller test. It clearly shows that it corresponds to the case $\theta=0$ and $\phi(L)=1$ in (**??**).

3.1 Asymptotic distribution of the DF test

The Dickey Fuller unit root test, in the case of uncorrelated residuals and non-deterministic components, is based on the following test regression:

$$y_t=\rho y_{t-1}+u_t \quad u_t\sim N(0,\sigma_u^2) \tag{17}$$

The test statistics τ is $\tau=(\hat\rho-1)/\hat\sigma_\rho$ and the z statistics $z=T(\hat\rho-1)$. As we are in the structural model, these tests depend on the value of θ. Their complete expression is:

$$\tau(\theta)=\frac{T^{-1}\sum_{t=1}^T y_ty_{t-1}-T^{-1}\sum_{t=1}^T y_{t-1}^2}{\sqrt{s^2\,T^{-2}\sum_{t=1}^T y_{t-1}^2}} \tag{18}$$

and

$$z(\theta)=\frac{T^{-1}\sum_{t=1}^T y_ty_{t-1}-T^{-1}\sum_{t=1}^T y_{t-1}^2}{T^{-2}\sum_{t=1}^T y_{t-1}^2} \tag{19}$$

We have the first theorem:

Theorem 1 *If the unobserved component model represents the null hypothesis, where η_t and ε_t are two independent white noises identically distributed of equal variance σ^2, the test statistics τ of Dickey and Fuller has a limiting distribution which is equal to:*

$$\tau(\theta) \xrightarrow{\mathcal{L}} \frac{(1-\theta)^2\, \frac{1}{2}\, (W(1)^2 - 1) - \theta^2\, \gamma(\alpha)}{[(1-\theta)^2 + 2\theta^2\, \gamma(\alpha)]^{1/2} \left((1-\theta)^2 \int_0^1 W^2(r)\, dr\right)^{1/2}}$$

where

$$\gamma(\alpha) = \left(\sum_{i=0}^{q} \alpha_i^2 - \sum_{i=1}^{q} \alpha_{i-1}\, \alpha_i\right)$$

when there is no constant or trend in the DGP and in the test regression.

Proof: We shall look for the asymptotically dominant components in $\tau(\theta)$ beginning by the denominator. Using (**??**) and summing over T, y_{t-1}^2 can be expanded as follows:

$$\sum_{t=1}^{T} y_{t-1}^2 = \sum_{t=1}^{T} \mu_{t-1}^2 + 2\,\theta \sum_{t=1}^{T} \mu_{t-1}\, \alpha(L)\, \varepsilon_{t-1} + \theta^2 \sum_{t=1}^{T} (\alpha(L)\, \varepsilon_{t-1})^2$$

The first term of this sum is $O(T^2)$ while the two others are $O(T)$ (See respectively Phillips (1987, theorem 3.1a) and proofs 1 and 2 in appendix). So we get:

$$\sum_{t=1}^{T} y_{t-1}^2 \overset{a}{\simeq} \sum_{t=1}^{T} \mu_{t-1}^2 \tag{20}$$

Let's now analyze the numerator of $\tau(\theta)$, expanding first the expression $y_t\, y_{t-1} - y_{t-1}^2$ using model (**??**):

$$\begin{aligned} y_t\, y_{t-1} - y_{t-1}^2 &= (y_t - y_{t-1})\, y_{t-1} \\ &= [(1-\theta)\, \eta_t + \theta\, \alpha(L)\, (\varepsilon_t - \varepsilon_{t-1})]\, [\mu_{t-1} + \alpha(L)\, \theta\, \varepsilon_{t-1}] \end{aligned}$$

Summing over T we get:

$$\begin{aligned} \sum_{t=1}^{T} (y_t y_{t-1} - y_{t-1}^2) = {} & (1-\theta) \sum_{t=1}^{T} \eta_t\, \mu_{t-1} + \theta \sum_{t=1}^{T} \mu_{t-1}\, \alpha(L)(\varepsilon_t - \varepsilon_{t-1}) \\ & + \theta\, (1-\theta) \sum_{t=1}^{T} \eta_t\, \alpha(L)\, \varepsilon_{t-1} + \theta^2 \sum_{t=1}^{T} \alpha(L)\varepsilon_t\, \alpha(L)\varepsilon_{t-1} - \theta^2 \sum_{t=1}^{T} (\alpha(L)\varepsilon_{t-1})^2 \end{aligned}$$

The first and the two last terms are $O(T)$ (See respectively Phillips (1987), theorem 3.1b and proofs 2 and 3 in appendix) while the two other terms are $O(T^{1/2})$ because of the Central Limit Theorem (CLT).

Let's regroup the terms of dominant order in $\tau(\theta)$:

$$\tau(\theta) \overset{a}{\simeq} \frac{T^{-1}\,(1-\theta)\sum_{t=1}^{T}\eta_t\mu_{t-1} - \theta^2\,T^{-1}\,\sum_{t=2}^{T}(\sum_{i=0}^{q}\alpha_i^2\,\varepsilon_{t-i}^2 - \sum_{i=1}^{q}\alpha_{i-1}\alpha_i\varepsilon_{t-i}^2)}{\left(T^{-2}\sum_{t=1}^{T}\mu_{t-1}^2\right)^{1/2}\;s}$$

To calculate the asymptotic distribution of this statistics, we shall apply formulae (17.3.22) and (17.3.26) of Hamilton (1994) when $\mu_t = \mu_{t-1} + (1-\theta)\eta_t$:

$$\begin{aligned} T^{-2}\sum_{t=1}^{T}\mu_{t-1}^2 &\overset{\mathcal{L}}{\to} (1-\theta)^2\sigma^2\int_0^1 W^2(r)\,dr \\ T^{-1}\sum_{t=1}^{T}\mu_{t-1}\,\eta_t &\overset{\mathcal{L}}{\to} (1-\theta)^2\frac{\sigma^2}{2}(W(1)^2-1) \end{aligned} \tag{21}$$

As under H_0, $u_t = \Delta y_t = (1-\theta)\,\eta_t + (1-L)\,\theta\,\alpha(L)\,\varepsilon_t$, the limiting variance of u_t is:

$$\mathrm{Var}(u_t) = (1-\theta)^2\,\sigma_\eta^2 + 2\theta^2\left(\sum_{i=0}^{q}\alpha_i^2 - \sum_{i=1}^{q}\alpha_i\,\alpha_{i-1}\right)\sigma_\varepsilon^2$$

which simplifies into:

$$\sigma_u^2 = \left\{(1-\theta)^2 + 2\theta^2\,\gamma(\alpha)\right\}\sigma^2$$

as $\sigma_\eta^2 = \sigma_\varepsilon^2 = \sigma^2$. Regrouping all these results, we get:

$$\tau(\theta) \overset{\mathcal{L}}{\to} \frac{(1-\theta)^2\,\frac{\sigma^2}{2}\,(W(1)^2-1) - \theta^2\,\sigma^2\,\gamma(\alpha)}{\left\{(1-\theta)^2 + 2\theta^2\,\gamma(\alpha)\right\}^{1/2}(\sigma^2)^{1/2}\left((1-\theta)^2\,\sigma^2\int_0^1 W^2(r)\,dr\right)^{1/2}}$$

We find the desired result when simplifying by σ^2.

□

We can immediately deduce the following corollary for the z statistics of Dickey and Fuller.

Corollary 1 *Under the null hypothesis of theorem* **??**, *the z statistics of Dickey and Fuller has the following limiting distribution:*

$$z(\theta) \overset{\mathcal{L}}{\to} \frac{(1-\theta)^2\,\frac{1}{2}\,(W(1)^2-1) - \theta^2\,\gamma(\alpha)}{(1-\theta)^2\int_0^1 W^2(r)\,dr}$$

when there is no constant or trend in the DGP and in the test regression.

Let us now comment the asymptotic distribution of the τ statistics under H_0.

- When $\theta = 0$, we have a pure random walk. The asymptotic distribution of $\tau(\theta)$ simplifies to:

$$\tau(\theta = 0) \xrightarrow{\mathcal{L}} \frac{\frac{1}{2}(W(1)^2 - 1)}{\left(\int_0^1 W^2(r)\, dr\right)^{1/2}}$$

  which is the asymptotic distribution of the τ statistics found by Dickey and Fuller (1981).

- When θ is different from 0, the asymptotic distribution is different from that of Dickey and Fuller. Its numerator is a linear combination of the usual distribution weighted by $(1-\theta)^2$, and of a component which depends on the stationary part of the series, weighted by θ^2.

- When θ tends to 1, the test statistics becomes $O(T^{1/2})$. The asymptotic distribution of $\sqrt{T}\,\tau(\theta)$, tends to a degenerate distribution whose probability is totally concentrated in one point:

$$T^{-1/2}\,\tau(\theta) \xrightarrow{p} \gamma(\alpha)$$

- The z statistics presents the same characteristics. However, it can be easily factorized into:

$$z(\theta) \xrightarrow{\mathcal{L}} \frac{\frac{1}{2}(W(1)^2 - 1)}{\int_0^1 W^2(r)\, dr} - \frac{\theta^2 \gamma(\alpha)}{(1-\theta)^2} \frac{1}{\int_0^1 W^2(r)\, dr} \tag{22}$$

  The first part of the factorization corresponds to the distribution of the usual z statistics. This factorization shows that the critical values for $z(\theta)$ are going to be much lower than those for z when θ approaches 1. We can compare this analytical result with that found in the Monte Carlo study of Schwert (1989). When $q = 0$, the reduced form of the model is an MA(1) and $\theta \to 1$ implies that $\beta \to -1$ using (**??**).

- The τ statistics is more difficult to factorize. We cannot obtain an expression where the usual τ statistics is corrected by an additive term, but simply:

$$\tau(\theta) \xrightarrow{\mathcal{L}} \frac{1-\theta}{[(1-\theta)^2 + 2\theta^2\gamma(\alpha)]^{1/2}} \times \frac{\frac{1}{2}(W(1)^2 - 1)}{\left[\int_0^1 W^2(r)\, dr\right]^{1/2}} - \frac{\theta^2\,\gamma(\alpha)}{1-\theta} \times \frac{1}{\left[\int_0^1 W^2(r)\, dr\right]^{1/2}} \tag{23}$$

  Thus the distribution of the $\tau(\theta)$ statistics is equal to the distribution of the τ statistics multiplied by a function of θ and α plus a correcting term which is the same as the correcting term in the case of $z(\theta)$, except that its denominator is now in square root.

- All these results can be extended to the case $q \to \infty$ provided the α_i are absolutely summable. Thus the case where the cycle is an ARMA process is covered.

4 Properties of the ADF test when the true DGP is the structural model

When there remains some autocorrelation in the residuals, the immediate reaction is to use the ADF test instead of the simple DF test. This test usual copes correctly with the problem. Moreover the ADF test τ has the same asymptotic distribution than the DF test. The ADF test z, divided by the sum of the autocorrelation coefficients, has the same limiting distribution as the z statistics. A common intuition would be that the ADF test is going to behave far better than the DF test in our problem. What we show in this section is that this intuition is partially wrong.

4.1 Preliminary discussion

The ADF test regression is:

$$\phi(L)\,\Delta y_t = (\rho - 1)\,y_{t-1} + u_t \quad u_t \sim N(0, \sigma_u^2) \tag{24}$$

where $\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$, is a lag polynomial whose roots are outside the unit circle. When u_t is a white noise, i.e. in the standard case, the z_p statistics which is $T(\hat{\rho} - 1)$ has the same limiting distribution as the z statistics, when divided by $\hat{\phi}(1)$.

It is convenient to rewrite the test regression as:

$$\Delta y_t = X_t \beta + u_t \quad u_t \sim N(0, \sigma_u^2)$$

where $X_t = (y_{t-1}, \Delta y_{t-1}, \Delta y_{t-2}, \ldots, \Delta y_{t-k})$ et $\beta = (\rho - 1, \phi_1, \phi_2, \ldots, \phi_p)$. The OLS estimator is $\hat{\beta} = (X'X)^{-1}\, X'\Delta y$ with $X'X = \sum_{t=1}^{T} X_t X_t'$ and $X'\Delta y = \sum_{t=1}^{T} X_t' \Delta y_t$. Let us detail the matrices $X'X$ and $X'\Delta y$ (omitting the summation for ease of notation):

$$\begin{bmatrix} y_{t-1}^2 & y_{t-1}\Delta y_{t-1} & y_{t-1}\Delta y_{t-2} & & \cdots & y_{t-1}\Delta y_{t-p} \\ & \Delta y_{t-1}^2 & \Delta y_{t-1}\Delta y_{t-2} & \Delta y_{t-1}\Delta y_{t-3} & \cdots & \Delta y_{t-1}\Delta y_{t-p} \\ & & \ddots & \ddots & \ddots & \vdots \\ & & & & & \Delta y_{t-p-2}\Delta y_{t-p} \\ & & & & & \Delta y_{t-p-1}\Delta y_{t-p} \\ & & & & & \Delta y_{t-p}^2 \end{bmatrix}^{-1} \begin{bmatrix} y_{t-1}\Delta y_t \\ \Delta y_{t-1}\Delta y_t \\ \Delta y_{t-2}\Delta y_t \\ \vdots \\ \\ \Delta y_{t-p}\Delta y_t \end{bmatrix}$$

In the standard case (when there is no MA and when there is the right number of lags in the test regression), the use of another test regression of the form:

$$y_t = X_t \beta + u_t \quad u_t \sim N(0, \sigma^2)$$

simplify the calculus of the asymptotic distribution. In this reparameterization, under H_0, $\rho_0 = 1$ and we derive the test statistics from:

$$\hat{\rho} - 1 = \left((X'X)^{-1}X'u\right)_1$$

Using a scaling matrix, we can show that the distribution depends only on the first element in $X'u$ divided by the first element in $X'X$. This simplicity, which is exploited in standard proofs (see e.g. Hamilton 1994 p.519), comes from the fact that the test regression encompasses the true DGP. This is no longer the case here. In the unobserved component model, the reduced form is an MA(1) when $q = 0$ and an MA(q+1) in the general case. In order to encompass the true DGP, we would need in theory an infinite number of lags of Δy_t in the test regression.

We have seen in the previous section that the $\tau(\theta)$ and the $z(\theta)$ statistics had similar limiting distributions, but that it was possible to express the limiting distribution of the $z(\theta)$ statistics as the sum of the usual distribution of the z statistics plus a correcting factor. This decomposition is more difficult for the $\tau(\theta)$ statistics. Consequently we shall first focus our attention on the $z(\theta)$ statistics, now labelled $z(\theta)_p$ as its distribution will depend on the number of correcting lags, treating the case of $\tau(\theta)_p$ in a next subsection.

4.2 Asymptotic distribution of the ADF test in the MA(1) case

Let us start with the simple case where the reduced form of the true DGP is an MA(1), corresponding to $\alpha(L) = 1$. Thus the cycle is just a white noise and $q = 0$. The asymptotic distribution of the $z(\theta)_p$ test can be first apprehended by the following proposition:

Proposition 1: *Let us consider the unobserved component model with $q = 0$. The asymptotic distribution of the $z(\theta)_p$ test is the same as that of the $z(\theta)$ test plus a correcting factor which is a function of the number of lags p and of θ.*

Proof: It is useful to define $A = X'X$ and $b = X'\Delta y$ and to write down the order of these matrices. We have:

$$\begin{bmatrix} T^2 & T & T & & \cdots & T \\ & T & T & T^{1/2} & \cdots & T^{1/2} \\ & & & & & \\ & & \ddots & \ddots & \ddots & \vdots \\ & & & & & \\ & & & & & T^{1/2} \\ & & & & & T \\ & & & & & T \end{bmatrix}^{-1} \begin{bmatrix} T \\ T \\ T^{1/2} \\ \\ \vdots \\ \\ T^{1/2} \end{bmatrix}$$

These order of magnitudes are explained by the fact that $\Delta y_t = (1-\theta)\eta_t + \theta\Delta\varepsilon_t$ which thus contains a unit root in the MA part. As it is easy to show that the first line of A^{-1} is $O(T^{-2})$, the OLS estimator of $(\hat{\rho} - 1)$ is asymptotically equivalent to:

$$(\hat{\rho} - 1) \stackrel{a}{=} a^{11}b_1 + a^{12}b_2 \qquad (25)$$

where superscripts denote elements of the inverse. The first element of the partitioned inverse is quite easy to compute. Let us first write:

$$a^{11} = \frac{|\text{adj } a_{11}|}{|A|} = \frac{|\text{adj } a_{11}|}{a_{11}|\text{adj } a_{11}| + \sum_{j=2}^{p}(-1)^{j+1}a_{1j}|\text{adj } a_{1j}|}$$

In this expression all the determinants of adjoint matrices are $O(T^p)$, a_{11} is $O(T^2)$, while all the other a_{1j} are $O(T)$. Consequently:

$$a^{11} \stackrel{a}{=} \frac{|\text{adj } a_{11}|}{a_{11}|\text{adj } a_{11}|} = 1/a_{11}$$

This first term in (**??**) is equal to $\sum y_{t-1}\Delta y_t / \sum y_{t-1}^2$ which is exactly $z(\theta)/T$.

The asymptotic distribution of the second element $a^{12}b_2$ is more tedious to derive. Let us write:

$$a^{12} = -\frac{|\text{adj } a_{12}|}{|A|} \stackrel{a}{=} -\frac{|\text{adj } a_{12}|}{a_{11}|\text{adj } a_{11}|}$$

There is no further possible simplification here. We shall prove in proposition 2 that $|\text{adj } a_{11}|$ converges in probability to an expression which is a difference equation of order 2 in p and θ. We shall prove in proposition 3 that $|\text{adj } a_{12}|$ converges to the same type of difference equation plus a second member which is a functional of a Brownian motion.

□

Let us define $\mathbf{H_p} = \text{adj } a_{11}$ which is a matrix of dimension $p \times p$ which contains all the cross products $\Delta y_{t-i}\Delta y_{t-j}$. $\mathbf{H_{p-1}}$ is the same matrix, but with only $p-1$ lags of Δy_t. Let us define $\mathbf{N_p} = \text{adj } a_{12}$ which is of dimension $p \times p$. The correcting factor of the statistics $z(\theta)_p$ is now expressed as:

$$a^{12}b_2 \stackrel{a}{=} -\frac{|\mathbf{N_p}|}{a_{11}|\mathbf{H_p}|}b_2 \tag{26}$$

Proposition 2: *Let us define* $\mathbf{a_p}$ *as the scaled determinant of* $\mathbf{H_p}$ *with:*

$$\mathbf{a_p} = \sigma^{-2p}\, T^{-p}\, |\mathbf{H_p}|$$

We have the following asymptotic difference equation of order 2 *in* $\mathbf{a_p}$*:*

$$\mathbf{a_p} \stackrel{a}{=} [(1-\theta)^2 + 2\theta^2]\, \mathbf{a_{p-1}} - \theta^4\, \mathbf{a_{p-2}}$$

with:

$$\begin{aligned} \mathbf{a_0} &= 1 \\ \mathbf{a_1} &\stackrel{a}{=} (1-\theta)^2 + 2\theta^2 \end{aligned}$$

Proof: Let us develop the determinant of $|\mathbf{H_p}|$. The leading elements are:

$$|\mathbf{H_p}| \stackrel{a}{=} h_{p,p}|\mathbf{H_{p-1}}| - h_{p-1,p}|\text{adj } h_{p-1,p}|$$

where $h_{i,j}$ is the i,j element of $\mathbf{H_p}$. It is easy to show that:

$$|\text{adj } \; h_{p-1,p}| \overset{a}{=} h_{p,p-1}|\mathbf{H_{p-2}}|$$

as $h_{p,p-1} = h_{p-1,p}$, we get:

$$|\mathbf{H_p}| \overset{a}{=} h_{p,p}|\mathbf{H_{p-1}}| - h_{p-1,p}^2|\mathbf{H_{p-2}}|$$

Now:

$$\begin{aligned} T^{-1}h_{p,p} &= T^{-1}\sum \Delta y_{t-p}^2 \xrightarrow{p} [(1-\theta)^2 + 2\theta^2]\sigma^2 \\ T^{-2}h_{p-1,p}^2 &= T^{-2}(\sum \Delta y_{t-p}\Delta y_{t-p-1})^2 \xrightarrow{p} \theta^4\sigma^4 \end{aligned}$$

Let us multiply the above difference equation by $T^{-p}\sigma^{-2p}$. Replacing $h_{p,p}$ and $h_{p-1,p}^2$ by their asymptotic limit, we get the desired result. The initial conditions are obtained by direct calculation for $p = 1$, starting with $a_0 = 1$.

□

Proposition 3: *Let us define* $\mathbf{b_p}$ *as the scaled determinant of* $\mathbf{N_p}$*with:*

$$\mathbf{b_p} = \sigma^{-2p}\,T^{-p}\,|\mathbf{N_p}|$$

We have the following asymptotic difference equation of order 2 on $\mathbf{b_p}$

$$\mathbf{b_p} \overset{a}{=} [(1-\theta)^2 + 2\theta^2]\,\mathbf{b_{p-1}} - \theta^4\,\mathbf{b_{p-2}} + \theta^{2p-2}\left[\mathbf{b_1} - \theta^2\right]$$

with:

$$\begin{aligned} \mathbf{b_0} &= 0 \\ \mathbf{b_1} &\overset{a}{=} (1-\theta)^2\,\frac{1}{2}\,(W(1)^2 - 1) + (1-\theta)^2 + \theta^2 \end{aligned}$$

The difference equation is conveniently factorized into:

$$\mathbf{b_p} = \mathbf{c_p} + \mathbf{d_p} \times \frac{1}{2}(W^2(1) - 1)$$

where

$$\begin{aligned} \mathbf{c_p} &\overset{a}{=} [(1-\theta)^2 + 2\theta^2]\,\mathbf{c_{p-1}} - \theta^4\mathbf{c_{p-2}} + \theta^{2(p-1)}\,(1-\theta)^2 \\ \mathbf{d_p} &\overset{a}{=} [(1-\theta)^2 + 2\theta^2]\,\mathbf{d_{p-1}} - \theta^4\mathbf{d_{p-2}} + \theta^{2(p-1)}\,(1-\theta)^2 \end{aligned}$$

with initial conditions

$$\begin{aligned} \mathbf{c_0} &= \mathbf{d_0} = 0 \\ \mathbf{c_1} &\overset{a}{=} (1-\theta)^2 + \theta^2 \\ \mathbf{d_1} &\overset{a}{=} (1-\theta)^2 \end{aligned}$$

Proof: Let us develop the determinant of $\mathbf{N_p}$. The leading terms are:

$$|\mathbf{N_p}| \stackrel{a}{=} n_{p,p}|\mathbf{N_{p-1}}| - n_{p-1,p}|\text{adj } n_{p-1,p}|$$

where $n_{i,j}$ is the i,j element of $\mathbf{N_p}$. We can develop the last determinant according to its leading terms:

$$|\text{adj } n_{p-1,p}| \stackrel{a}{=} n_{p,p-1}|\mathbf{N_{p-2}}| - n_{p-2,p-1}|\text{adj } n_{p-2,p-1}|$$

The last determinant is equivalent to a difference equation of order 1. Noting $|\text{adj } n_{p-2,p-1}| = |\mathbf{J_{p-2}}|$, the recurrence is:

$$|\mathbf{J_{p-2}}| \stackrel{a}{=} n_{p-3,p-2}|\mathbf{J_{p-3}}|$$

where $\mathbf{J_1} = n_{p,1}$. Its solution is thus:

$$|\mathbf{J_{p-2}}| \stackrel{a}{=} n_{p,1}\prod_{j=2}^{p-2} n_{j-1,j}$$

The complete recurrence equation becomes:

$$|\mathbf{N_p}| \stackrel{a}{=} n_{p,p}|\mathbf{N_{p-1}}| - n_{p-1,p}\left[n_{p,p-1}|\mathbf{N_{p-2}}| - n_{p-2,p-1}n_{p,1}\prod_{j=2}^{p-2} n_{j-1,j}\right]$$

We must now study the convergence of all the $n_{i,j}$. From proofs 4,5 and 6 in the appendix, we have:

$$\begin{aligned}
T^{-1}n_{p-i,p-i+1} &= T^{-1}\sum \Delta y_{t-p+i}\Delta y_{t-p+i-1} \xrightarrow{p} -\theta^2\sigma^2 \\
T^{-1}n_{p,p} &= T^{-1}\sum \Delta y_{t-p}^2 \xrightarrow{p} [(1-\theta)^2 + 2\theta^2]\sigma^2 \\
T^{-1}n_{p,1} &= T^{-1}\sum y_{t-1}\Delta y_{t-p} \xrightarrow{\mathcal{L}} \left[(1-\theta)^2\tfrac{1}{2}(W^2(1)-1) + (1-\theta)^2\right]\sigma^2
\end{aligned}$$

Let us multiply the above difference equation by $T^{-p}\sigma^{-2p}$. Replacing the $n_{i,j}$ by their asymptotic limit, we get the desired result. The initial conditions are obtained by direct calculation for $p = 1$, starting with $b_0 = 0$. The decomposition of the recurrence equation is immediate.

□

We can now regroup our results in the following theorem which expresses the asymptotic distribution of the $z(\theta)_p$ statistics as the sum of the usual distribution of the z statistics plus a correcting factor which is a function of θ and p:

Theorem 2 *Let us consider the structural unobserved component model where the cycle is just a white noise. The asymptotic distribution of the $z(\theta)_p$ statistics is given by:*

$$\begin{aligned}
z(\theta)_p = T(\hat\rho - 1) \xrightarrow{\mathcal{L}} & \frac{\frac{1}{2}(W^2(1)-1)}{\int_0^1 W^2(r)\,dr}\left(1 + \frac{\mathbf{d_p}}{\mathbf{a_p}}\frac{\theta^2}{(1-\theta)^2}\right) \\
& + \frac{\theta^2}{(1-\theta)^2}\left(\frac{\mathbf{c_p}}{\mathbf{a_p}} - 1\right)\frac{1}{\int_0^1 W^2(r)\,dr}
\end{aligned}$$

Proof: We simply regroup our results and use the fact that:

$$\begin{aligned} T^{-1}b_2 &= T^{-1}\sum \Delta y_{t-1}\Delta y_t \xrightarrow{p} -\theta^2\sigma^2 \\ T^{-2}a_{11} &= T^{-2}\sum y_{t-1}^2 \xrightarrow{\mathcal{L}} (1-\theta)^2\sigma^2 \int_0^1 W^2(r)\,dr \end{aligned}$$

□

Remarks:

- The first term in this distribution is that of the z statistics.
- It is easy to verify that

$$1 + \frac{\mathbf{d_p}}{\mathbf{a_p}}\frac{\theta^2}{(1-\theta)^2} = \hat{\phi}(1)$$

Consequently

$$z(\theta)_p/\hat{\phi}(1) \xrightarrow{\mathcal{L}} \frac{\frac{1}{2}(W^2(1)-1)}{\int_0^1 W^2(r)\,dr} + \frac{\theta^2(\mathbf{c_p}-\mathbf{a_p})}{\mathbf{a_p}(1-\theta)^2+\mathbf{d_p}\theta^2}\frac{1}{\int_0^1 W^2(r)\,dr} \tag{27}$$

This expression gives the distribution of the corrected $z(\theta)_p$ statistics which is the same as that of the usual ADF z statistics when $\theta = 0$.

4.3 the $\tau(\theta)_p$ statistics

It is quite easy to derive the asymptotic distribution of the $\tau(\theta)_p$ when we know that of $z(\theta)_p$ as $\tau(\theta)_p = (\hat{\rho}-1)/\hat{\sigma}_\rho = T^{-1}z(\theta)_p/\hat{\sigma}_\rho$. Thus, we simply have to determine the limiting value of $\hat{\sigma}_\rho$ under H_0:

Proposition 4: *The variance of the OLS residuals in an AR(p) regression without a constant or a trend when the true model is the unobserved component model is given by:*

$$s^2 \xrightarrow{p} [(1-\theta)^2 + 2\,\theta^2]\sigma^2 \sum_{i=0}^{p}\left[\frac{\mathbf{a_{p-i}}}{\mathbf{a_p}}\right]^2\theta^{4i} - 2\sigma^2\sum_{i=0}^{p-1}\frac{\mathbf{a_{p-i}}\,\mathbf{a_{p-i-1}}}{\mathbf{a_p}^2}\theta^{4(i+1)}$$

Proof: Under H_0, $\mathrm{Var}(u_t) = \mathrm{Var}(\phi(L)\Delta y_t)$.

$$\begin{aligned} \mathrm{Var}(\phi(L)\Delta y_t) &= \mathrm{E}(\Delta y_t - \textstyle\sum_{i=1}^{p}\phi_i\Delta y_{t-i})^2 \\ &= \mathrm{E}(\Delta y_t^2) + \textstyle\sum_{i=1}^{p}\phi_i^2\mathrm{E}(\Delta y_{t-i}^2) - 2\phi_1\mathrm{E}(\Delta y_t\Delta y_{t-1}) + 2\textstyle\sum_{i=1}^{p-1}\phi_i\phi_{i+1}\mathrm{E}(\Delta y_{t-i}\Delta y_{t-i-1}) \end{aligned}$$

Knowing that $\phi_i \xrightarrow{p} \mathbf{a_{p-i}}/\mathbf{a_p}(-\theta^{2i})$ with $\phi_0 = 1$, we get the desired result.

□

Theorem 3 *If the unobserved component model represents the null hypothesis, the ADF test statistics $\tau(\theta)_p$ when there is no constant or trend has limiting distribution:*

$$\tau(\theta)_p \xrightarrow{\mathcal{L}} \frac{(1-\theta)^2\mathbf{a_p} + \theta^2\mathbf{d_p}}{s(\theta)^{1/2}(1-\theta)\mathbf{a_p}} \times \frac{\frac{1}{2}(W(1)^2 - 1)}{\left[\int_0^1 W^2(r)\,dr\right]^{1/2}} + \frac{\theta^2}{s(\theta)^{1/2}(1-\theta)}\frac{\mathbf{c_p}-1}{\mathbf{a_p}} \times \frac{1}{\left[\int_0^1 W^2(r)\,dr\right]^{1/2}}$$

where $s(\theta) \stackrel{a}{=} \mathrm{Var}(u_t)$.

Proof: It suffices to note that $\hat{\sigma}_\rho = s^2(\theta)(X'X)^{11}$ and that:

$$(X'X)^{11} \stackrel{a}{=} \sum y_{t-1}^2$$

and from (**??**) and (**??**)

$$T^{-2}\sum y_{t-1}^2 \xrightarrow{\mathcal{L}} (1-\theta)^2\,\sigma^2 \int_0^1 W^2(r)\,dr$$

□

5 Monte Carlo evidences

The asymptotic distribution of the ADF test, when the true DGP is the unobserved component model, can be expressed as a non-linear combination of its usual textbook asymptotic distribution and of a correcting term which is a function of the number of lags p and of the parameter θ. Many questions are interesting:

- What is the optimal lag length p in the ADF test in order to minimize the size distortion?
- Are the usual empirical methods such as the t statistics or the Schwarz criterion efficient to determine p?
- How large is the loss of power when p is optimal to minimize the size distortion?
- Is it possible to modify the ADF test in order to minimize the loss of power?

Some of these questions have already received an answer in the literature using Monte Carlo experiments. Schwert (1989) is the first and most famous reference. Many other followed, see Stock (1994) for a list. In this section, we shall try to find answers to these questions using simulations of the true asymptotic distribution as given in the previous section.

5.1 Size distortion and optimal lag selection

The question is to determine if the ADF test can cope with a negative MA component in the reduced form of the unobserved component model and what is the number p of lags needed in order to get a satisfactory size correction. We shall simulate the asymptotic distribution of the $z(\theta)_p$ statistics (for $q = 0$, which corresponds to an MA(1) in the reduced form) for various values of p and of $theta$ and compare it to the standard distribution of the DF test. In this experiment, one draw of the Brownian motion $W(.)$ is obtained with 400 draws of a $N(0, 1)$.

Davidson and MacKinnon (1994) have developed a nice graphical method to compare two distributions and report the results of Monte Carlo experiments. Let us consider a reference distribution corresponding to the random variable X with its quantiles and the studied distribution corresponding to the random variable $\tilde{X}$, also with its quantiles. If the two distributions are the same, the quantiles will be the same. If the distributions are different, the 10% quantiles of the reference distribution will correspond to, say, the 20% quantiles of the studied distribution. These discrepancies can be visualized in the two dimensional simplex. More precisely, with a set α of levels of critical values[3], we can define the reference vector x such as:

$$\forall i = 1, \dots, n \quad P(X < x_i) = \alpha_i \tag{28}$$

Then we define the $\tilde{\alpha}_i$ sequence by:

$$\forall i = 1, \dots, n \quad P(\tilde{X} < x_i) = \tilde{\alpha}_i \tag{29}$$

The graph of $\tilde{\alpha}_i$ against α_i gives the desired figure where the reference distribution is the 45° line. In our case, the reference function is the asymptotic repartition function of the τ or z statistics.

In figure **??** we present the discrepancy between the quantiles of the distribution of $z(\theta)_p$ and those of z as a function of p. The ADF manages to correct easily the discrepancy for low values of θ. But for high values of θ, a quite high number of lags is necessary, slightly over 10 (with $p = 10$ and $\theta = 0.9$, the true size is 18% instead of 10%).

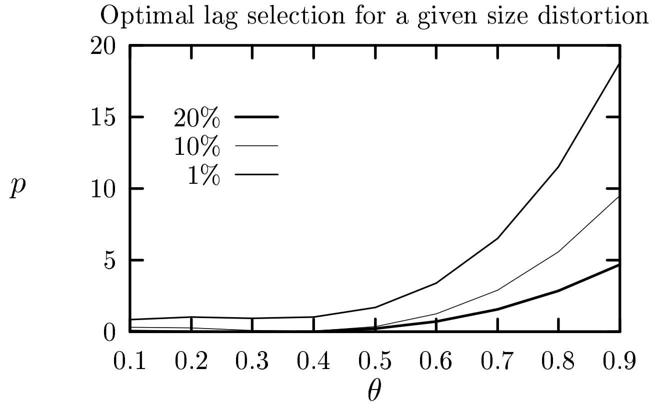
It is convenient to summarize these findings into an empirical function giving the upper value of p which insures a maximum of size distortion of α% when $\theta = \theta_0$. Formally, we have:

$$p\mathrm{max} = f_{\mathrm{ds}<\alpha\%}(\theta) \tag{30}$$

We have adjusted a polynomial function of degree three in θ for $\alpha = 20\%$, 10% and 1%. These three functions are graphed in figure **??**. This figure summarize the previous findings that size distortion is very low for $\theta < 0.5$ and that consequently $p = 0$ or $p = 1$ is enough to guaranty a size distortion less than 1%. However, the optimum value of p grows rather quickly for $\theta > 0.5$ and is around 20 for $\theta = 0.9$.

[3]The vector α we have chosen is $0.001, 0.002, \dots, 0.01, 0.015, 0.020, 0.025, \dots, 0.99, 0.991, \dots, 0.999$.

Figure 1: Optimal lag selection as for a given size distortion



5.2 Comparison with usual lag selection methods

The question of lag selection for an ADF test was addressed by various authors in the literature. Dod and Giles (1995) recommend using four lags in most cases. Clearly this strategy does not work here. Harris (1992) recommends to use the formula in Schwert (1989) which is $p = 12 \times \text{int}(T/100)^{0.25}$. But this formula does not depend of θ and thus is of little use here. Ng and Perron (1995) favor a general to specific method for selecting p using t or F test for the significance of the longest lag. They underline that information criteria usually select a too low number of lags. We shall see, by mean of a Monte Carlo experiment based on the following generating process:

$$y_t = (1-\theta)\sum_{i=1}^{t}\eta_i + \theta\varepsilon_t \qquad \eta_t \sim N(0,1), \quad \varepsilon_t \sim N(0,1), \quad \text{Cov}(\eta_t, \varepsilon_s) = 0 \tag{31}$$

if the general to specific method of Ng and Perron works in our case.

We have designed two Monte Carlo experiments, one with $T = 100$, the other with $T = 1000$. We used 1000 replications with DGP given in (**??**) and θ varying between 0 and 1. We estimate an ADF regression starting with $p = 20$ and then determine the optimal lag length either with the t statistics or with the Schwarz criterion. Figure 3 presents the average optimal p as a function of θ for both criteria. In small samples ($T = 100$), the t statistics manages to select a maximum number of 8 lags, which is not too bad when θ is close to one. But it does not manage to discriminate the cases where no correction is needed ($\theta = 0$). Consequently it produces a test which terribly lacks power. The Schwarz criterion is best for small values of θ, but selects a too parsimonious model for θ close to one. The test produced has good power, but may present strong size distortion.

Figure 2: Size distortion of the $z(\theta)_p$ statistics for 10 000 replications (direct simulation)

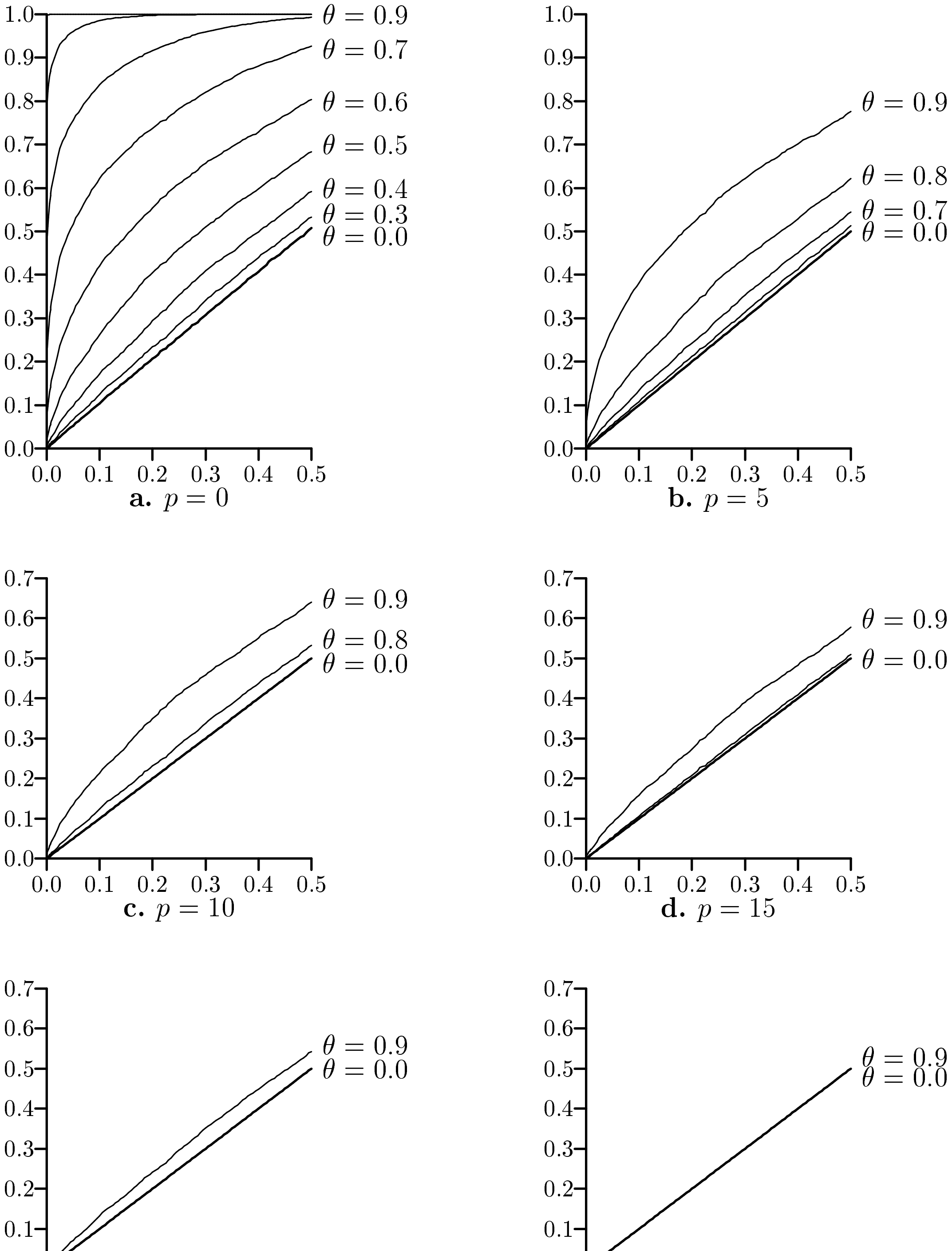


Figure 3: Lag selection for $T = 100$

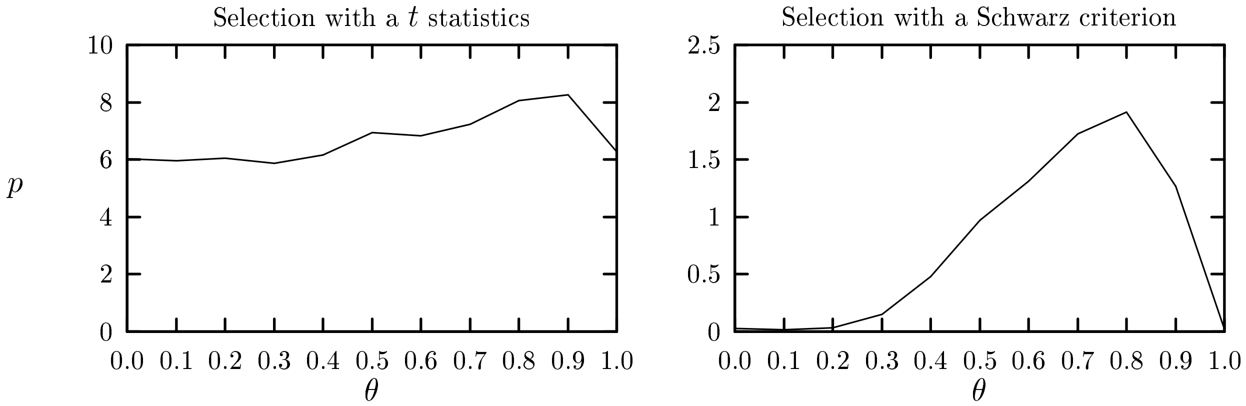
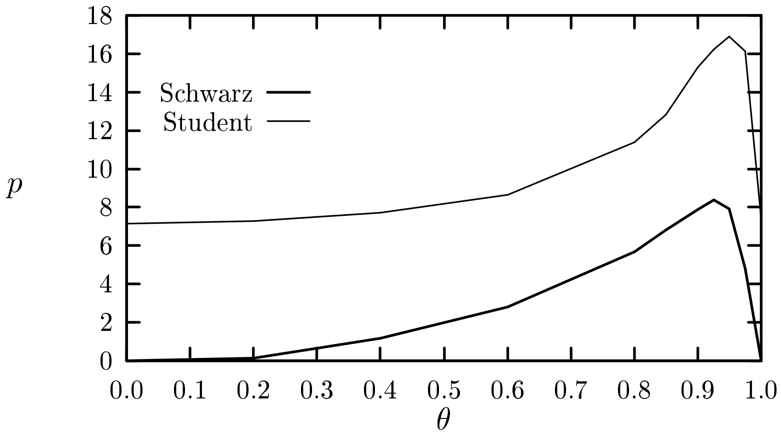


Figure 4: Lag selection for $T = 1000$



For larger samples ($T = 1000$), the Figure 4 shows that the t selection procedure presents similar defaults. The number of lags selected is always too high. The Schwarz selection procedure could give a more reasonable answer, but still present a slight size distortion for θ close to one as the maximum number of lags selected is 8. On Figure 2, this represents a size of 25% instead of 10% when $\theta = 0.9$.

6 Modification of the ADF test for negative MA

- estimation of θ with a non parametric measure

- modify the statistics of theorem 3

6.1 Loss of power

Il faudrait la power function analytique pour faire bien. Sinon, il faut utiliser la simulation indirecte avec comme GDP:

$$(1 - \rho L)y_t = (1 + \beta L)\epsilon_t \tag{32}$$

comme c'est fait dans Stock. Mais lui le fait avec p variable. Il faudrait refaire avec p fixe. Trade off between size and power.

7 Conclusion

When the data are thought to be the sum of a stochastic trend and a cycle, the unobserved component model initiated by Harvey (1985) is a natural candidate for modeling the data. This choice has drastic implications for the type of unit root tests to be made on the data. We have shown, by analytical asymptotic arguments that the usual DF test was not valid. This is not new, but we have also shown that the ADF test was not reliable, at least in small samples. Specific tests have been proposed in the literature such as Franzini and Harvey (1983), Nabeya and Tanaka (1988) and Nyblom (1986). As the reduced form of the unobserved component model is a constrained MA(q+1), other tests are of interest such as Arellano and Pantula (1995), Kwiatkowski *et al* (1992), Phillips and Ouliaris (1988), Saikkonen and Luukkonen (1993a, b) and Tanaka (1990) among others. It is on our agenda to investigate the relative power of these tests.

These tests are in general not very easy to apply. As a bivariate VAR ecompases the UC model, a final and simple possibility to test for unit roots would be to use this representation. But the question is of course to have an auxuliary variable correlated with the stationnary part of the variable under test.

Appendix

A Calculations of convergence orders

A.1 For the DF test

Proof 1:

$$\begin{aligned} \textstyle\sum_{t=1}^T \mu_{t-1}\,\alpha(L)\,\varepsilon_t &= \textstyle\sum_{t=1}^T \sum_{s=1}^t \eta_s \sum_{i=0}^q \alpha_i\,\varepsilon_{t-i} \\ &= \textstyle\sum_{i=0}^q \left(\sum_{t=1}^T \alpha_i\,\varepsilon_{t-i} \sum_{s=1}^t \eta_s\right) \end{aligned} \tag{33}$$

We have the sum of q terms which are of the same asymptotic order. When $i = 0$, we get for the variance:

$$\mathrm{Var}\left(\sum_{t=1}^T \varepsilon_t \sum_{s=1}^t \eta_s\right) = \sum_{t=1}^T \sum_{s=1}^t \sum_{t'=1}^T \sum_{s'=1}^{t'} \mathrm{E}\left(\varepsilon_t \varepsilon_{t'} \eta_s \eta_{s'}\right)$$

As ε_t and η_t are two white noises of variance 1, we get:

$$\mathrm{Var}\left(\sum_{t=1}^{T}\varepsilon_t \sum_{s=1}^{t}\eta_s\right) = \frac{1}{2}T(T+1)$$

As its variance is $O(T^2)$, this term is $O(T)$. Thus (**??**) is the sum of q terms $O(T)$.

■

Proof 2:

$$\sum_{t=1}^{T}(\alpha(L)\,\varepsilon_t)^2 = \sum_{i=0}^{q}\sum_{t=1}^{T}\alpha_i^2\,\varepsilon_{t-i}^2 + 2\sum_{i=0}^{q}\left(\sum_{j=i+1}^{q}\sum_{t=1}^{T}\alpha_i\,\alpha_j\,\varepsilon_{t-i}\,\varepsilon_{t-j}\right) \tag{34}$$

Applying the LLN to the first term gives q terms, each being $O(T)$. Applying the CLT to the second term, we get $q(q+1)/2$ terms which are $O(T^{1/2})$.

■

Proof 3:

$$\sum_{t=1}^{T}\alpha(L)\varepsilon_t\,\alpha(L)\varepsilon_{t-1} = \sum_{j=0}^{q}\sum_{i=j}^{q}\sum_{t=1}^{T}\alpha_j\alpha_i\varepsilon_{t-j-1}\varepsilon_{t-i} + \sum_{i=1}^{q}\sum_{t=1}^{T}\alpha_{i-1}\alpha_i\varepsilon_{t-i}^2 + \sum_{i=2}^{q}\sum_{j=0}^{i-2}\sum_{t=1}^{T}\alpha_i\alpha_j\varepsilon_{t-i}\varepsilon_{t-j-1}$$

For the same reasons as before this is $O(T)$ because:

$$\sum_{t=1}^{T}\alpha(L)\varepsilon_t\,\alpha(L)\varepsilon_{t-1} \overset{a}{\simeq} \sum_{i=1}^{q}\sum_{t=1}^{T}\alpha_{i-1}\alpha_i\varepsilon_{t-i}^2$$

■

A.2 For the ADF test

We have, for all i:

$$\begin{aligned} y_{t-i} &= (1-\theta)\,\textstyle\sum_{t=1}^{t-i}\eta_s + \theta\,\varepsilon_{t-i} \\ \Delta y_{t-i} &= (1-\theta)\,\eta_{t-i} + \theta\,\varepsilon_{t-i} - \theta\,\varepsilon_{t-i-1} \end{aligned}$$

From Hamilton (1994) p 486:

$$\begin{aligned} \forall i,\ (1-\theta)\,T^{-1}\sum_{t=1}^{T}\mu_{t-i-1}\,\eta_{t-i} &\overset{\mathcal{L}}{\to} (1-\theta)^2\,\frac{\sigma^2}{2}\,(W(1)^2-1) \\ \forall i,\ T^{-2}\sum_{t=1}^{T}\mu_{t-i}^2 &\overset{\mathcal{L}}{\to} (1-\theta)^2\,\sigma^2\int_0^1 W(r)^2\,dr \end{aligned}$$

Proof 4: As

$$\sum_{t=1}^{T}\Delta y_{t-i}\Delta y_{t-i-1} \overset{a}{=} -\theta^2\sum_{t=1}^{T}\varepsilon_{t-i-1}^2$$

thus

$$T^{-1} \sum_{t=1}^{T} \Delta y_{t-i} \Delta y_{t-i-1} \xrightarrow{\mathcal{L}} -\theta^2 \, \sigma^2$$

$\blacksquare$

Proof 5: As

$$\sum_{t=1}^{T} \Delta y_{t-i}^2 \;\overset{a}{=}\; (1-\theta)^2 \sum_{t=1}^{T} \eta_{t-i}^2 + \theta^2 \sum_{t=1}^{T} \varepsilon_{t-i}^2 + \theta^2 \sum_{t=1}^{T} \varepsilon_{t-i-1}^2$$

we have

$$T^{-1} \sum_{t=1}^{T} \Delta y_{t-i}^2 \xrightarrow{\mathcal{L}} \left[(1-\theta)^2 + 2\theta^2 \right] \sigma^2$$

$\blacksquare$

Proof 6: As

$$\sum_{t=1}^{T} \Delta y_{t-1} \, y_{t-1} \;\overset{a}{=}\; (1-\theta) \sum_{t=1}^{T} \mu_{t-2} \, \eta_{t-1} + (1-\theta)^2 \sum_{t=1}^{T} \eta_{t-1}^2 + \theta^2 \sum_{t=1}^{T} \varepsilon_{t-1}^2$$

thus

$$T^{-1} \sum_{t=1}^{T} \Delta y_{t-1} \, y_{t-1} \xrightarrow{\mathcal{L}} \left[(1-\theta)^2 \, \frac{1}{2} \left(W(1)^2 - 1 \right) + (1-\theta)^2 + \theta^2 \right] \sigma^2$$

$\blacksquare$

Proof 7: As

$$\sum_{t=1}^{T} \Delta y_t \, y_{t-1} \overset{a}{=} (1-\theta) \sum_{t=1}^{T} \mu_{t-1} \, \eta_t - \theta^2 \sum_{t=1}^{T} \varepsilon_{t-1}^2$$

thus

$$T^{-1} \sum_{t=1}^{T} \Delta y_{t-1} \, y_{t-1} \xrightarrow{\mathcal{L}} \left[(1-\theta)^2 \, \frac{1}{2} \left(W(1)^2 - 1 \right) - \theta^2 \right] \sigma^2$$

$\blacksquare$

Proof 8: As

$$\sum_{t=1}^{T} \Delta y_{t-i} \, y_{t-1} \;\overset{a}{=}\; (1-\theta) \sum_{t=1}^{T} \mu_{t-i-1} \, \eta_{t-i} + (1-\theta)^2 \sum_{t=1}^{T} \eta_{t-i}^2$$

thus

$$T^{-1} \sum_{t=1}^{T} \Delta y_{t-i} \, y_{t-1} \xrightarrow{\mathcal{L}} \left[(1-\theta)^2 \, \frac{1}{2} \left(W(1)^2 - 1 \right) + (1-\theta)^2 \right] \sigma^2$$

$\blacksquare$

REFERENCES

Arellano, C. and S.G. Pantula (1995), "Testing for Trend Stationarity versus Difference Stationarity", *Journal of Time Series Analysis*, 16, 147-164.

Campbell, J.Y. and G. Mankiw (1987a), "Are Output Fluctuations Transitory ?", *Quarterly Journal of Economics*, 102, 857-880.

Campbell, J.Y. and G. Mankiw (1987b), "Permanent and Transitory Components in Macroeconomic Fluctuations", *AEA Papers and Proceedings*, 77, 111-117.

Cochrane, J.H. (1988), "How Big is the Random Walk in Output", *Journal of Political Economy*, 96, 893-920.

Davidson, R. and J.G. MacKinnon (1994), "Graphical Methods for Investigating the Size and Power of Hypothesis Tests", *G.R.E.Q.A.M. Working Paper* No 94A23.

Dickey, D.A. and W.A. Fuller (1979), "Distribution of the Estimators for Autoregressive Time Series With a Unit Root", *Journal of the American Statistical Association*, 74, 427-431.

Dickey, D.A. and W.A. Fuller (1981), "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root", *Econometrica*, 49, 1057-1072.

Dods, J.L. and D.E.A. Giles (1995) Alternative Strategies for Augmenting the Dickey-Fuller Test: Size Robustness in the Face of Pretesting, *Journal of Statistical Computation and Simulation*, 53, 243-258.

Franzini, L. and A.C. Harvey (1983), "Testing for Deterministic Trend and Seasonal Components in Time Series Models", *Biometrika*, 70, 673-682.

Hamilton, J.D. (1994), "Time Series Analysis", *Princeton University Press*.

Harris, R.I.D. (1992), "Testing for Unit Roots using the Augmented Dickey- Fuller Test", *Economics Letters*, 38, 381-386.

Harvey, A.C. (1985), "Trends and Cycles in Macroeconomic Time Series", *American Statistical Association*, 3, 216-27.

Harvey, A.C. (1989), "Forecasting, Structural Time Series Models and the Kalman Filter", *Cambridge University Press*.

Kwiatkowski, D., P.C.B. Phillips, P. Schmidt and Y. Shin (1992), "Testing the Null Hypothesis of Stationarity Against the Alternative for Unit Root: How Sure are we that Economic Time Series Have a Unit Root?", *Journal of Econometrics*, 54, 159-178.

Nabeya, S. and K. Tanaka (1988), "Asymptotic Theory of a Test for the Constancy of Regression Coefficients Against the Random Walk Alternative", *The Annals of Statistics*, 16, 218-235.

Nelson, C.R. and C.I. Plosser (1982), "Trends and Random Walks in Macroeconomic Time Series", *Journal of Monetary Economics*, 10, 139-162.

Ng, S. and P. Perron (1995), "Unit Root Tests in ARMA Models with Data Dependent Methods for the Selection of the Truncation Lag", *Journal of the American Statistical Association*, 90, 268-281.

Nyblom, J. (1986), "Testing for Deterministic Linear Trend in Time Series", *Journal of the American Statistical Association*, 81, 545-549.

Phillips, P.C.B. (1987), "Time Series Regression with a Unit Root", *Econometrica*, 55,277-301.

Phillips, P.C.B. and S. Ouliaris (1988) "Testing for Cointegration Using Principals Components Methods", *Journal of Economic Dynamics and Control*, 12, 205-230.

Saikkonen, P. and R. Luukkonen (1993a), "Point Optimal Tests for Testing the Order of Differencing in ARIMA Models", *Econometric Theory*, 9, 343-362.

Saikkonen, P. and R. Luukkonen (1993b), "Testing for a Moving Average Unit Root in Autoregressive Integrated Moving Average Models", *Journal of the American Statistical Association*, 88, 596-601.

Schwert, G.W. (1989), "Tests for Unit Roots: A Monte Carlo Investigation", *Journal of Business and Economic Statistics*, 7, 147-159.

Stock, J. and M. Watson (1988), "Variable Trends in Economic Time Series", *Journal of Economic Perspectives*, 2, 147-174.

Tanaka, K. (1990), "Testing for Moving Average Unit Root", *Econometric Theory*, 6, 433-444.