Working Paper 95-40
Statistics and Econometrics Series 12
September 1995

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (341) 624-9849

ON THE EXACT MOMENTS OF NON-STANDARD ASYMPTOTIC DISTRIBUTIONS IN NON STATIONARY AUTOREGRESSIONS WITH DEPENDENT ERRORS

Jesús Gonzalo and Jean-Yves Pitarakis*

Abstract

In this paper we derive the exact moments of the asymptotic distributions of the OLS estimate and t-statistic in an unstable AR(1) with dependent errors. We can therefore establish theoretically and without simulations, the distortions induced by the presence of non iid errors on inferences as judged by their impact on the moments of the limiting distributions. In addition we study the relationship between the number of lagged dependent variables required for matching the moments of the distribution in the "approximately iid errors" model with those occuring in the purely iid case. Our framework allows us to distinguish explicitly between different types of error processes and study their implications for the lag length selection. A very accurate normal approximation also allows us to obtain approximate magnitudes for the size distortions when the iid based distributions are used for inferences.

Key Words

Dependent errors; Dickey & Fuller distribution; Moment Generating Function; Exact Moments.

*Gonzalo, Departamento de Estadística y Econometría, Universidad Carlos III de Madrid; Pitarakis, C.O.R.E. Université Catholique de Louvain. Financial help from the European Union "Human Capital and Mobility Programme". N° ERBCHBICT941677 is gratefully acknowledged.

1 Introduction

Many results regarding the impact of dependent errors on test statistics and coefficient estimates in the context of non stationary time series have been established via Monte-Carlo simulations. In an extensive empirical study for instance, Schwert (1989) considered the distortions induced by the presence of moving average errors on various test statistics within a non stationary AR(1) model. More recently, Agiakoglou and Newbold (1992) studied the size properties of the ADF test under a moving average error structure. Since the relevant asymptotic distributions are expressed in terms of stochastic integrals involving Wiener processes, they are not directly usable for computational purposes. This partly explains why most studies adopted the computationally demanding simulation approach.

However, since most quantities of interest are expressed in terms of ratios of stochastic integrals, which themselves are the limits of properly normalized quadratic forms it is possible to obtain their exact moments using their joint moment generating function (MGF thereafter). The recent unit root literature contains m any attempts to obtain exact distributional results via the relevant characteristic function and the use of Gurland's (1948) inversion theorem (Evans and Savin (1981,1984), Perron (1989), Hisamatsu and Maekawa (1994) among others). Most of these studies however used the iid errors assumption and focused solely on the normalized OLS estimate of the autoregressive coefficient. More recently however, Perron (1994) considers MA and AR error structures in a near integrated framework. His objectives are different from ours since he focuses mainly on obtaining exact distributional results for the normalized OLS coefficient and studying the adequacy of the asymptotic approximations.

Typically, the two most important quantities arising in the non-standard asymptotic distributions are $\int_0^1 W(r)\, dW(r)$ and $\int_0^1 W(r)^2 dr$ where $W(r)$ denotes a standard Brownian Motion. In this paper we will analyze the impact of the presence of dependent errors on autoregressive coefficient estimates and corresponding t-statistics by focusing on the influence of the error process on the exact moments of the asymptotic distributions and indirectly on inferences. Thus, we will establish theoretically and exactly various empirically observed facts with the aim of better understanding the behavior of these asymptotic distributions when the error structure is not iid. Our framework allows us to distinguish explicitely between different types of error processes.

As a byproduct, we extend a result by Le Breton & Pham (1989) regarding the exact asymptotic bias of the OLS estimate in a non stationary AR(1), to the case where the error process is not iid. In addition, we will investigate the connection between the magnitude of the coefficients of the error process and the number of lags of the dependent variable necessary to dampen their effects on the original asymptotic distribution. Our results can also be viewed as a preliminary step for the design of Bartlett corrections to the test statistics in the context of non stationary processes.

The plan of this paper is as follows. Section 2 presents the general framework and the theoretical tools. Section 3 focuses on the quantitative results and Section 4 concludes. All proofs are gathered in the appendix.

2 The Framework and Methodology

We consider the following first order autoregressive process

$$\Delta X_t = \alpha X_{t-1} + u_t \tag{1}$$
$$\phi(L)\, u_t = \theta(L)\, \epsilon_t \tag{2}$$

where $\phi(L) = 1 - \rho L$, $\theta(L) = 1 + \theta L$ and $X_0 = 0$. We further assume that ϵ_t is iid Gaussian with mean zero and variance σ_ϵ^2, and the two polynomials in the lag operator L have no roots in common and satisfy the usual stability assumptions. For notational simplicity, and with no loss of generality we also put $\sigma_\epsilon^2 = 1$. The quantities of interest are the OLS estimate of α in (1) and the corresponding t-statistic denoted $t_{\hat{\alpha}}$. In what follows, we distinguish between an AR(1) process for u_t (ie. $\theta = 0$), an MA(1) and mixed ARMA(1,1) respectively. The various asymptotic distributions of the quantities of interest when $\alpha = 0$ are gathered in the following lemma where "$\Rightarrow$" denotes convergence in distribution.

Lemma 2.1

- Case $\rho = 0$ (*MA*(1) *errors*)

  1. $T(\hat{\alpha} - \alpha) \Rightarrow \frac{\int_0^1 W(r)dW(r)}{\int_0^1 W(r)^2 dr} + \frac{\theta}{(1+\theta)^2} \frac{1}{\int_0^1 W(r)^2 dr}$

2. $t_{\hat{\alpha}} \Rightarrow \frac{(1+\theta)}{(1+\theta^2)^{1/2}} \frac{\int_0^1 W(r)dW(r)}{(\int_0^1 W(r)^2 dr)^{1/2}} + \frac{\theta}{(1+\theta)(1+\theta^2)^{1/2}} \frac{1}{(\int_0^1 W(r)^2 dr)^{1/2}}$

- Case $\theta = 0$ ($AR(1)$ *errors*)

3. $T(\hat{\alpha} - \alpha) \Rightarrow \frac{\int_0^1 W(r)dW(r)}{\int_0^1 W(r)^2 dr} + \frac{\rho}{1+\rho} \frac{1}{\int_0^1 W(r)^2 dr}$

4. $t_{\hat{\alpha}} \Rightarrow \frac{(1+\rho)^{1/2}}{(1-\rho)^{1/2}} \frac{\int_0^1 W(r)dW(r)}{(\int_0^1 W(r)^2 dr)^{1/2}} + \frac{\rho}{(1-\rho)^{1/2}(1+\rho)^{1/2}} \frac{1}{(\int_0^1 W(r)^2 dr)^{1/2}}$

- Case $\theta \neq 0$ and $\rho \neq 0$ ($ARMA(1,1)$ *errors*)

5. $T(\hat{\alpha} - \alpha) \Rightarrow \frac{\int_0^1 W(r)dW(r)}{\int_0^1 W(r)^2 dr} + f_1(\rho, \theta) \frac{1}{\int_0^1 W(r)^2 dr}$

6. $t_{\hat{\alpha}} \Rightarrow f_2(\rho, \theta) \frac{\int_0^1 W(r)dW(r)}{(\int_0^1 W(r)^2 dr)^{1/2}} + f_3(\rho, \theta) \frac{1}{(\int_0^1 W(r)^2 dr)^{1/2}}$

where the functions f_i ($i = 1, 2, 3$) above are defined as follows

$$\begin{aligned}
f_1(\rho, \theta) &= \frac{(\theta+\rho)(1+\rho\theta)}{(1+\rho)(1+\theta)^2}, \\
f_2(\rho, \theta) &= \frac{(1+\rho)^{1/2}}{(1-\rho)^{1/2}} \frac{(1+\theta)}{(1+\theta+2\rho\theta)^{1/2}}, \\
f_3(\rho, \theta) &= \frac{(1+\theta)(1+\theta^2+2\rho\theta)^{1/2}(\rho+\theta)(1+\rho-\rho^2+\theta)}{(1-\rho)^{9/2}(1+ho)^{3/2}}.
\end{aligned}$$

Noting that when the errors are iid the asymptotic distributions are

$$\frac{\int_0^1 W(r)dW(r)}{\int_0^1 W^2(r)dr} \quad and \quad \frac{\int_0^1 W(r)dW(r)}{(\int_0^1 W^2(r)dr)^{1/2}}$$

for $T(\hat{\alpha} - 1)$ and $t_{\hat{\alpha}}$ respectively, the extra components illustrate the impact on inferences of dependent errors, when we mistakenly assume them to be iid. Clearly, the signs and magnitudes of the parameters of the error processes will play a crucial role.

All the above distributions are expressed in terms of ratios of stochastic integrals. Indeed letting $Q_1 = \int_0^1 W(r)\, dW(r)$ and $Q_2 = \int_0^1 W(r)^2 dr$, the ratios of interest are Q_1/Q_2, $Q_1/\sqrt{Q_2}$, $1/Q_2$ and $1/\sqrt{Q_2}$ respectively.

Lemma 2.2

The joint moment generating function $\phi(u,v)$ of Q_1 and Q_2 has the following representation

$$\phi(u,v) = e^{-u/2}[\cosh(-2v)^{1/2} - \frac{u}{(-2v)^{1/2}} \sinh(-2v)^{1/2}]^{-\frac{1}{2}}.$$

We can now apply a result due to Sawa (1958). Assuming $P(Q_2 > 0) = 1$, and letting $\phi(u,v)$ denote the joint MGF of Q_1 and Q_2, the m^{th} order moment of Q_1/Q_2 is given by

$$E\left[\left(\frac{Q_1}{Q_2}\right)^m\right] = \Gamma^{-1}(m) \int_0^\infty v^{m-1} \left[\frac{\partial^m \phi(u,-v)}{\partial u^m}\right]_{u=0} dv$$

provided the moments exist and where $\Gamma(.)$ is the gamma function, and m is a positive integer. Given the form of the ratios of interest to us the above expression is not directly applicable for obtaining the exact moments of quantities such as $Q_1/\sqrt{Q_2}$ or Q_2^{-m} respectively. However, using a result in Davies et al. (1985), we can extend Sawa's result to the above cases as well.

Lemma 2.3

$$E\left[\left(\frac{Q_1}{\sqrt{Q_2}}\right)^m\right] = \Gamma^{-1}(\frac{m}{2}) \int_0^\infty v^{\frac{m}{2}-1} \left[\frac{\partial^m \phi(u,-v)}{\partial u^m}\right]_{u=0} dv$$

$$E\left[\left(\frac{1}{\sqrt{Q_2}}\right)^m\right] = \Gamma^{-1}(\frac{m}{2}) \int_0^\infty v^{\frac{m}{2}-1} [\phi(u,-v)]_{u=0}\, dv.$$

We can therefore use directly the joint moment generating function of Q_1 and Q_2 in Lemma 2.2 to solve the above integrals. This is the objective of the next section, where we also investigate the behavior of the moments with *θ and ρ.*

3 Numerical Results

3.1 Original (Non Augmented) Model

The following lemma presents the exact numerical value of the expectations (moments) of the ratios of interest. We use the following notational convention, letting $\mu^m_{\frac{A}{B}}$ denote $E((\frac{A}{B})^m)$. In what follows, we will focus solely on the first moment and the variance. Higher order moments can also be easily obtained using Lemma 2.3.

Lemma 3.1

Letting $x = (-2v)^{1/2}$, we have

- $\mu^1_{\frac{Q_1}{Q_2}} = -\frac{1}{2}\int_0^\infty \frac{x}{(\cosh x)^{1/2}}dx + \frac{1}{2}\int_0^\infty \frac{\sinh x}{(\cosh x)^{3/2}}dx = -1.78143$
- $\mu^2_{\frac{Q_1}{Q_2}} = \int_0^\infty \frac{x^3}{8(\cosh x)^{1/2}} - \frac{x^2 \sinh x}{4(\cosh x)^{3/2}} + \frac{3x(\sinh x)^2}{8(\cosh x)^{5/2}}\, dx = 13.2857$
- $\mu^1_{\frac{1}{Q_2}} = \int_0^\infty \frac{x}{(\cosh x)^{1/2}}dx = 5.56286$
- $\mu^2_{\frac{1}{Q_2}} = \int_0^\infty \frac{x^3}{2(\cosh x)^{1/2}}dx = 67.8312$
- $\mu^1_{\frac{Q_1}{\sqrt{Q_2}}} = \Gamma^{-1}(\frac{1}{2})\left[-\frac{1}{2}\int_0^\infty \frac{1}{(\cosh x)^{1/2}}dx + \frac{1}{\sqrt{2}}\int_0^\infty \frac{\sinh x}{x(\cosh x)^{3/2}}dx\right] = -0.42310$
- $\mu^1_{\frac{1}{\sqrt{Q_2}}} = \Gamma^{-1}(\frac{1}{2})\int_0^\infty \frac{\sqrt{2}}{(\cosh x)^{1/2}}dx = 2.09211$
- $\mu^2_{\frac{Q_1}{\sqrt{Q_2}}} = \int_0^\infty \frac{x}{4(\cosh x)^{1/2}} - \frac{\sinh x}{2(\cosh x)^{3/2}} + \frac{3(\sinh x)^2}{4x(\cosh x)^{5/2}}dx = 1.1416$
- $\mu^1_{\frac{Q_1}{Q_2^2}} = \int_0^\infty \frac{-x^3}{4(\cosh x)^{1/2}} + \frac{x^2 \sinh x}{2(\cosh x)^{3/2}}dx = -22.7899.$

Using the above Lemma we can assess the exact impact of θ *and* ρ, on the first and second moments as well as the variance of the asymptotic distributions in Lemma 2.1. Letting $g^j_\ell(\rho,\theta)$ denote the j^{th} moment of the distributions in cases $\ell = 1,2,3,4,5,6$ of Lemma 2.1, we have

$$g^1_1(\theta) = -1.78143 + \frac{\theta}{(1+\theta)^2}\, 5.56286,$$

$$g^1_2(\theta) = -0.42310\frac{(1+\theta)}{(1+\theta^2)^{1/2}} + \frac{\theta}{(1+\theta)(1+\theta^2)^{1/2}}\, 2.09211,$$

$$
\begin{aligned}
g_3^1(\rho) &= -1.78143 + \frac{\rho}{(1+\rho)}\,5.56286, \\
g_4^1(\rho) &= -0.42310\,\frac{(1+\rho)^{1/2}}{(1-\rho)^{1/2}} + \frac{\rho}{(1-\rho)^{1/2}(1+\rho)^{1/2}}\,2.09211, \\
g_5^1(\rho,\theta) &= -1.78143 + \frac{(\theta+\rho)(1+\rho\theta)}{(1+\rho)(1+\theta)^2}\,5.56286, \\
g_6^1(\rho,\theta) &= -0.42310\,f_2(\rho,\theta) + f_3(\rho,\theta)\,2.09211
\end{aligned}
$$

for the pure autoregressive, moving average and ARMA processes respectively. The following tables illustrate in more detail the exact quantitative impact of ρ and θ on the location and variance of the asymptotic distributions. It is worth emphasizing the fact that our figures are *exact*. The use of a purely numerical Monte-Carlo approach would have required days of computing time in order to lead to numbers close to ours. We also believe that our approach leads to a better understanding of the distinct impact of the structure of the error process.

Table 3.1.1: $T(\hat{\alpha}-\alpha)$

(ρ,θ)	$g_1^1(\theta)$	$g_3^1(\rho)$	$v_1(\theta)$	$v_3(\rho)$
-0.9	-502.4388	-51.8472	302104.71	3329.83
-0.7	-45.0481	-14.7614	2528.36	297.00
-0.1	-2.4682	-2.3995	15.23	14.67
0.0	-1.7814	-1.7814	10.11	10.11
0.1	-1.3217	-1.2757	7.32	7.06
0.7	-0.4340	0.5092	3.34	1.18
0.9	-0.3946	0.8536	3.21	0.92

Table 3.1.2: $t_{\hat{\alpha}}$

(ρ,θ)	$g_2^1(\theta)$	$g_4^1(\rho)$	$v_2(\theta)$	$v_4(\rho)$
-0.9	-14.0270	-4.4167	53.9669	5.9559
-0.7	-4.1031	-2.2284	5.2340	1.7764
-0.1	-0.6102	-0.5930	0.9640	0.9625
0.0	-0.4231	-0.4231	0.9626	0.9626
0.1	-0.2739	-0.2575	0.9854	0.9893
0.7	0.1165	1.0435	1.1598	2.4115
0.9	0.1391	2.4754	1.1756	7.2124

In tables 3.1.1 and 3.1.2 above, $v_\ell(\theta)$ and $v_\ell(\rho)$ denote the variances in the MA(1) and AR(1) cases respectively. The number -1.78143 appearing in the row corresponding to $\rho = \theta = 0$ (Table 3.1.1) is the asymptotic bias of the OLS estimate in the unit root model with iid errors (Le Breton & Pham (1989)) and -0.4231 (Table 3.1.2) represents the mean of the asymptotic distribution of the t-statistic in the same context. Clearly, both cases illustrate the more frequent occurence of negative values due to the presence of the unit root.

Particularly interesting are the magnitudes of the directions of shift and the differences between the autoregressive and moving average error processes. In Table 3.1.1, we can clearly observe the drastic shift to the left of the asymptotic distribution of $T(\hat{\alpha} - 1)$ as θ and ρ tend towards -1 respectively, the impact being much stronger in the case of moving average errors as judged by *both* the mean and variance. Indeed, we can observe an important influence of the error structure on the distributional shifts and variance changes. An autoregressive structure being much less distortionary than a moving average one when the parameters are negative and large in absolute value.

These shifts explain the severe size distortions occuring in the presence of dependent errors when inferences are based on the asymptotic distributions derived under the iid errors assumption. Both types of error structures will produce severe size distortions, but for realistic parameter values our results suggest that an MA process will be much more distortionary.

Positive parameter values imply a shift rightwards of the asymptotic distributions and therefore an easier wrong acceptance of the unit root hypothesis when the iid based distributions are used for inferences. Our results suggest that for the $T(\hat{\alpha} - \alpha)$ statistic, the "undersizeness" will be more severe under an autoregressive error structure, where *both* the mean and variance are more severely modified relative to the MA case (see Table 3.1.1).

When both ρ and θ are close to 0 however their respective effect on the asymptotic distribution is quite comparable in magnitude.

The behavior of the asymptotic distribution of the $t - statistic$ (Table 3.1.2) displays less pronounced displacements with θ or ρ. Although the directions are similar to the ones occuring in the $T(\hat{\alpha} - \alpha)$ case, they are also weaker. Clearly, this supports the view that the $t - statistic$ is more reliable than the normalized OLS estimate for making inferences in such frameworks. Although Dickey and Fuller (1979) suggested that the $T(\hat{\alpha} - \alpha)$

statistic is more powerful than the $t-statistic$, our previous finding supports Schwert's (1989) Monte-Carlo based claim that the latter is more robust to model misspecifications. It will display better size properties when the model contains dependent errors.

This point will be reinforced in the next section where we analyze the impact of augmenting the model on the behaviour of the two test statistics. Regarding the differences in the impact of the two types of error processes, it is also worth observing that when the parameters are positive, their increase leads to a faster shift of the mean (rightwards) as well as a faster increase in the variance under an autoregressive error structure. On the other hand, when negative values are considered, the mean decreases and the variance increases faster under the MA process.

Finally, Tables 3.1.3-3.1.7 below focus on an ARMA(1,1) model for the error process. It is difficult to argue that such a mixed process will necessarily lead to more pronounced changes in the asymptotic distributions than in the pure AR or MA cases. Indeed, this will depend on the mix of values taken by θ and ρ together and especially on their respective signs. It might happen for instance that a large positive ρ considerably dampens the influence of a large negative θ, leading to distributions that remain closer to the iid case than say when both θ and ρ are moderate but negative. Overall, the t-statistic displays less pronounced deviations from the iid error based asymptotic distribution.

Table 3.1.3 Both roots negative and close ($|\theta| > |\rho|$)

(θ,ρ)	$g_5^1(\rho,\theta)$	$g_6^1(\rho,\theta)$	$v_5(\rho,\theta)$	$v_6(\rho,\theta)$
(-0.9,-0.8)	-8134.6828	-169.9493	78878976.00	7831.78
(-0.5,-0.4)	-41.8340	-5.5308	2107.75	9.61
(-0.2,-0.1)	-4.7367	-1.1160	34.2074	1.19
(0.0, 0.0)	-1.78143	0.42310	10.11	0.9626

Table 3.1.4 Both roots negative and close ($|\rho| > |\theta|$)

(θ,ρ)	$g_5^1(\rho,\theta)$	$g_6^1(\rho,\theta)$	$v_5(\rho,\theta)$	$v_6(\rho,\theta)$
(-0.8,-0.9)	-4068.2321	-161.0269	19729168	7036.32
(-0.4,-0.5)	-35.1586	-5.1903	1492.56	8.71
(-0.1,-0.2)	-4.4083	-1.0583	30.5023	1.17
(0.0, 0.0)	-1.78143	0.42310	10.11	0.9626

Table **3.1.5** Both roots positive and close ($|\theta| > |\rho|$)

(θ,ρ)	$g_5^1(\rho,\theta)$	$g_6^1(\rho,\theta)$	$v_5(\rho,\theta)$	$v_6(\rho,\theta)$
(0.9,0.8)	0.7218	4.9240	5.99	35.75
(0.5,0.4)	0.1258	0.7355	5.6161	3.05
(0.2,0.1)	-0.7068	-0.0253	6.5124	1.17
(0.0, 0.0)	-1.78143	0.42310	10.11	0.9626

Table **3.1.6** Roots with opposite signs

(θ,ρ)	$g_5^1(\rho,\theta)$	$g_6^1(\rho,\theta)$	$v_5(\rho,\theta)$	$v_6(\rho,\theta)$
(-0.9,0.8)	-10.4348	-1.8387	139.44	1.66
(-0.5,0.4)	-3.0529	-0.7511	17.93	1.00
(-0.2,0.1)	-2.5558	-0.6350	14.41	0.98
(0.0, 0.0)	-1.78143	0.42310	10.11	0.9626

Table **3.1.7** Roots with opposite signs

(θ,ρ)	$g_5^1(\rho,\theta)$	$g_6^1(\rho,\theta)$	$v_5(\rho,\theta)$	$v_6(\rho,\theta)$
(0.8,-0.9)	-2.2622	-0.5721	12.61	1.01
(0.4,-0.5)	-2.2355	-0.5543	12.46	0.97
(0.1,-0.2)	-2.3446	-0.5820	13.10	0.97
(0.0, 0.0)	-1.78143	0.42310	10.11	0.9626

3.2 Normal Approximations

Given that we obtained the exact mean and variance of the asymptotic distributions of $t_{\hat{\alpha}}$ and $T(\hat{\alpha} - \alpha)$, it is natural to inquire about the quality of a normal approximation to these non-standard distributions. The unit root literature has often raised this question by comparing the left tails of the Dickey-Fuller distributions to the ones of the standard normal. It would be more legitimate however to make the comparison with $N(\mu_\infty, \sigma_\infty)$ where μ_∞ and σ_∞ are the mean and standard deviation of the correct non-standard asymptotic distributions. Indeed, for the $t_{\hat{\alpha}}$ statistic, we have $\mu_{1,\infty} = -0.4231$ and $\sigma^2_{1,\infty} = 0.9626$, and for $T(\hat{\alpha} - \alpha)$, $\mu_{2,\infty} = -1.78143$ and $\sigma^2_{2,\infty} = 10.11$. Another motivation behind these calculations is to obtain approximate numerical values for the magnitudes of the size distortions implied by the shifts in the moments when the iid based asymptotic distributions are used for inferences. Theoretically, one could obtain the exact tail probabilities by inverting the relevant characteristic functions corresponding to MA, AR or ARMA errors. However, this is beyond the scope of this paper and to our knowledge has never been done for the more commonly used t-statistic.

The following table displays the relevant normal and "exact" DF critical values

Table 3.2.1: Normal Approximations[2]

	N(-0.4231,0.9626)	DF_1	N(-1.78143,10.11)	DF_2
2.5%	-2.35	-2.23	-8.01	-10.50
5%	-2.04	-1.95	-7.03	-8.10
10%	-1.68	-1.62	-5.85	-5.70

Clearly, a suitable normalization leads to a very accurate normal approximation for the t-statistic at all relevant percentage points. The closeness of these distributions can partly justify the use of the normal approximation in order to obtain approximate estimates of the size distortions under different error structures. However, since the approximation for the normalised OLS coefficient is less accurate, we will concentrate solely on the t-statistic. Obviously, we also need to check the validity of the approximation when the original DF distributions are shifted due to the presence of MA or AR errors. For this reason, the following tables also include direct DF based size estimates obtained via numerical simulations. The previously obtained moments of the various asymptotic distributions were very informative about the directions of shift of these distributions and gave an overall intuition about the seriousness and magnitude of the distortions, however in order to obtain a more precise numerical value of say the probability of rejecting the null when true, we need to compute the relevant tail areas.

Our main objective here is to illustrate the connection between the shifts in the moments and the implied "new tail area". Going back to our previous point, we saw for instance that when $\theta = -0.7$, we have $\mu_\infty = -4.1031$ and $\sigma^2_\infty = 5.2340$ for the standard $t-statistic$. We can therefore compute the implied size distortion via the following probability

$$P[X \leq -1.95 | X \simeq N(-4.1031, 5.2340)]$$

[2] DF_1 and DF_2 denote the left tail 5% critical values of the non-standard (Dickey & Fuller) asymptotic distribution of $t_{\hat{\alpha}}$ and $T(\hat{\alpha} - \alpha)$ respe ctively.

where -1.95 corresponds to the 5% DF critical value (see Table 3.2.1). Similar probabilities can therefore be obtained for a whole range of values for θ. The following table summarizes some of the implied approximate size estimates for the t-statistic under a 5% nominal size. The last two columns display the counterparts of the size estimates obtained by Monte-Carlo simulations with N=10000 replications and a sample size of 5000 observations.

Table 3.2.2: Approximate & Empirical Size Estimates

(ρ, θ)	SIZE (MA(1))	SIZE (AR(1))	DF(MA(1))	DF(AR(1))
-0.9	94.95%	84.38%	100%	86.48%
-0.7	82.64%	58.32%	84.04%	52.86%
-0.1	8.53%	8.69%	8.52%	7.96%
0.0	5.94%	5.94%	5.01%	5.01%
0.1	4.55%	4.46%	2.90%	2.89%
0.7	2.74%	2.68%	0.56%	0.00%
0.9	2.68%	4.95%	0.33%	0.00%

The above numbers confirm that the undersizeness (due to both AR or MA error processes with $(\rho, \theta) > 0$) is more serious when the errors are characterized by an autoregressive structure. On the other hand an MA process leads to greater oversizeness when the parameters are negative. These results unanimously confirm our moments based analysis of section 3.1. In addition our results based on the normal approximation were also able to give an accurate description of the size distortions implied by the shifts in the moments, especially for the most relevant cases (negative or close to zero parameter values).

3.3 Augmented Model

In practice, in order to be able to continue using the the distributions corresponding to $\theta = 0$ *and/or* $\rho = 0$ even when the error process is not iid, one adds lagged changes of the dependent variable to the right hand side of (1). This has the effect of whitening the error process which can then be assumed to be approximatively iid. In applied work, an important issue is then the selection of the truncation lag. Indeed, in order for inferences to be based on Q_1/Q_2 even when say $\theta \neq 0$ or $\rho \neq 0$, the lag length needs to satisfy certain speed conditions (Said and Dickey (1984), Ng and Perron (1994)), and will therefore play a crucial role on the quality of inferences even asymptotically. In order to shed some light on this issue, we computed the asymptotic distributions of $\hat{\alpha}$ and $t_{\hat{\alpha}}$ in (1) when $k = 1$ *and* 2, and where k denotes the truncation lag. We can therefore analyze explicitly the relationship between the lag length and the magnitudes of θ *or* ρ. The estimated model is given by

$$\Delta X_t = \alpha_1 X_{t-1} + \sum_{j=1}^{k} \gamma_j \Delta X_{t-j} + u_t$$

The following lemma first presents the different distributions under the hypothesis that $\alpha_1 = 0$ and fo a given lag length.

Lemma 3.2

- Case $k = 1$ *and* $MA(1)$

1. $T(\hat{\alpha}_1 - \alpha_1) \Rightarrow \left(1 - \frac{\theta}{(1+\theta^2)}\right) \frac{\int_0^1 W(r)dW(r)}{\int_0^1 W(r)^2 dr} - \frac{\theta^2}{(1+\theta)^2(1+\theta^2)} \frac{1}{\int_0^1 W(r)^2}$

2. $t_{\hat{\alpha}_1} \Rightarrow \left(\frac{1-\theta^2}{1+\theta^2}\right)^{1/2} \frac{(1+\theta)(1+\theta^2-\theta)}{(1-\theta^6)^{1/2}} \frac{\int_0^1 W(r)dW(r)}{(\int_0^1 W(r)^2 dr)^{1/2}} - \left(\frac{1-\theta^2}{1+\theta^2}\right)^{1/2} \frac{\theta^2}{(1+\theta)(1-\theta^6)^{1/2}} \frac{1}{(\int_0^1 W(r)^2 dr)^{1/2}}$

- Case $k = 1$ *and* $AR(1)$

3. $T(\hat{\alpha}_1 - \alpha_1) \Rightarrow (1 - \rho) \frac{\int_0^1 W(r)dW(r)}{\int_0^1 W(r)^2 dr}$

4. $t_{\hat{\alpha}_1} \Rightarrow \frac{\int_0^1 W(r)dW(r)}{(\int_0^1 W(r)^2 dr)^{1/2}}$

- Case $k = 2$ *and* $MA(1)$

5. $T(\hat{\alpha}_1 - \alpha_1) \Rightarrow \frac{1+\theta^2}{1+0a^2+\theta} \frac{\int_0^1 W(r)dW(r)}{\int_0^1 W(r)^2 dr} + \frac{\theta^3}{(1+\theta)^2(1+\theta^2-\theta)(1+\theta^2+\theta)} \frac{1}{\int_0^1 W(r)^2}$

6. $t_{\hat{\alpha}_1} \Rightarrow \frac{(1+\theta)(1+\theta^2)(1-\theta^6)^{1/2}}{(1+\theta^2+\theta)(1-\theta^8)^{1/2}} \frac{\int_0^1 W(r)dW(r)}{(\int_0^1 W(r)^2 dr)^{1/2}} + \frac{(\theta^3)(1-\theta^6)^{1/2}}{(1+\theta^2+\theta)(1+\theta^2-\theta)(1+\theta)(1-\theta^8)^{1/2}} \frac{1}{(\int_0^1 W(r)^2)^{1/2}}$

- Case $k = 2$ *and* $AR(1)$

7. $T(\hat{\alpha}_1 - \alpha_1) \Rightarrow (1 - \rho) \frac{\int_0^1 W(r)dW(r)}{\int_0^1 W(r)^2 dr}$

8. $t_{\hat{\alpha}_1} \Rightarrow \frac{\int_0^1 W(r)dW(r)}{(\int_0^1 W(r)^2 dr)^{1/2}}$.

We can now compare the moments of the above distributions in order to quantify the strength of the lag length for a given magnitude of θ and ρ. The following table displays the first moments of the distributions under no lags added, one lag and two lags respectively in both the MA and AR cases.

Table 3.3.1: $T(\hat{\alpha}_1 - \alpha_1)$

(ρ,θ)	$g_1^1(\theta)_{k=0}$	$g_1^1(\theta)_{k=1}$	$g_1^1(\theta)_{k=2}$	$g_3^1(\rho)_{k=0}$	$g_3^1(\rho)_{k=1}$	$g_3^1(\rho)_{k=2}$
-0.9	-502.45	-251.61	-167.99	-51.85	-3.38	-3.38
-0.7	-45.04	-22.94	-15.61	-14.76	-3.03	-3.03
-0.1	-2.47	-2.03	-1.98	-2.40	-1.96	-1.96
0.0	-1.78	-1.78	-1.78	-1.78	-1.78	-1.78
0.1	-1.32	-1.65	-1.62	-1.28	-1.60	-1.60
0.7	-0.43	-1.58	-0.83	0.51	-0.53	-0.53
0.9	-0.39	-1.59	-0.73	0.85	-0.18	-0.18

The first three columns of the above table display the moments of the asymptotic distribution of $T(\hat{\alpha}_1 - \alpha_1)$ under MA(1) errors when k=0,1 and 2 respectively. When θ is negative and as we increase k, we get closer and closer to the moment of the iid errors based asymptotic distribution. The picture is very different when we focus on positive values for θ. Indeed, in this latter case there is an initial improvement as we move from k=0 to k=1, but as we go further to k=2, the situation deteriorates and hence adding lags will not always be beneficial when inferences are based on $T(\hat{\alpha}_1 - \alpha_1)$. For instance when $\theta > 0$, the optimal number of lags seems to be k=1. It is important to observe that this does not mean that usig k=3 or further will not lead to any improvement. The point is that an increase in k does not yield to a *strict* improvement. This phenomenon has also occured in Monte-Carlo results where the interest lied in determining the empirical size of the tests under MA(1) errors and as k was being increased (see Agiakoglou and Newbold (1992) pp. 474-475 for instance). Indeed, an increase in the number of lags from say k=1 to k=8 always improved the size but not always when we move from k=1 to k=2 or 3 for instance.

The last three colums above focus on the AR(1) errors case. Clearly, much less lags are required for whitening the error process in this case for similar magnitudes of ρ. The following table displays the equivalent numbers for the t statistic.

Table 3.3.2: $t_{\hat{\alpha_1}}$

(ρ,θ)	$g_2^1(\theta)_{k=0}$	$g_2^1(\theta)_{k=1}$	$g_2^1(\theta)_{k=2}$	$g_4^1(\rho)_{k=0}$	$g_4^1(\rho)_{k=1}$	$g_4^1(\rho)_{k=2}$
-0.9	-14.0270	-8.075	-5.6858	-4.4167	-0.4231	-0.4231
-0.7	-4.1031	-2.3010	-1.5695	-2.2284	-0.4231	-0.4231
-0.1	-0.6102	-0.4410	-0.4249	-0.5930	-0.4231	-0.4231
0.0	-0.4231	-0.4231	-0.4231	-0.4231	-0.4231	-0.4231
0.1	-0.2739	-0.4380	-0.4216	-0.2575	-0.4231	-0.4231
0.7	0.1165	-0.729	-0.2374	1.0435	-0.4231	-0.4231
0.9	0.1391	-0.768	-0.1918	2.4754	-0.4231	-0.4231

Again the three first columns above focus on the MA(1) case and the remaining on AR(1). Noting that if the true process is a simple random walk with iid errors and we instead fit the random walk by also adding one lagged change of the dependent variable to the right hand side, we implicitely have a random walk DGP with AR(1) errors. This is the reason why the column corresponding to AR(1) errors with $k = 1$ shows a perfect match of the first moment for any value of ρ. The same also happens with k=2, since asymptotically the inclusion of extra lags (beyond the true number) does not affect the asymptotic distribution of the t statistic. An interesting point also arises by looking at the asymptotic distribution of $T(\hat{\alpha_1} - \alpha_1)$ under AR(1) errors when k=1 or 2 (cases 3 and 7 in Lemma 3.2). Indeed, we can notice that although one lag is enough for the t-statistic to be brought to the iid distribution case (cases 4 and 8), the $T(\hat{\alpha_1} - \alpha_1)$ statistic will always remain displaced with respect to the iid distribution no matter how many lags we use. This can be intuitively explained by the fact that the latter statistic does not take the variance into account and clearly therefore its use will always lead to more severely distorted inferences.

Again, for negative values of ρ the distortions are much less pronounced than in the MA errors case, but when θ and ρ are positive, it is the autoregressive structure that leads to higher displacements of the asymptotic distributions. Focusing specifically on the MA(1) case, when θ is not so large, the inclusion of even one lag seems to adjust the distribution quite efficiently towards the iid one. The following table display the variances of the asymptotic distributions of the t-statistic in the MA(1) case for the augmented model

Table 3.3.3: $V_\infty(t_{\hat{\alpha}_1})$

(ρ,θ)	$v(\theta)_{k=0}$	$v(\theta)_{k=1}$	$v(\theta)_{k=2}$
-0.9	53.967	18.332	9.422
-0.7	5.234	2.136	1.397
-0.1	0.964	0.961	0.963
0.0	0.963	0.963	0.963
0.1	0.985	0.962	0.963
0.7	1.160	0.982	0.995
0.9	1.176	0.990	1.008

We can now compute the implied size distortions using the normal approximation. We focus solely on MA errors since we previously showed that the distortions induced by the presence of AR errors are neutralized when the number of lags is equal to or greater than one. The first three columns of Table 3.3.4 below display the size estimates obtained via the normal approximation using the relevant mean and variance of the asymptotic distributions for various values of the lag length. The last three columns are again the "exact" counterparts obtained via Monte-Carlo simulations using the correct Dickey & Fuller distribution.

Table 3.3.4: Size Estimates (Normal vs Monte Carlo when MA(1))

θ	$N_{k=0}$	$N_{k=1}$	$N_{k=2}$	$\mathrm{DF}_{k=0}$	$\mathrm{DF}_{k=1}$	$\mathrm{DF}_{k=2}$
-0.9	94.95%	92.36%	88.88%	100%	99.54%	94.64%
-0.7	82.64%	59.48%	37.45%	84.04%	54.32%	35.46%
-0.1	8.53%	6.18%	6.06%	8.52%	5.42%	5.24%
0.0	5.94%	5.94%	5.94%	5.01%	5.00%	5.00%
0.1	4.55%	6.18%	5.94%	2.90%	5.10%	4.80%
0.7	2.74%	10.93%	4.27%	0.56%	10.70%	2.57%
0.9	2.68%	11.70%	4.01%	0.33%	11.56%	2.26%

It is again interesting to observe the evolution of the size estimates when $\theta > 0$, where an increase from k=0 to k=1 seriously deteriorate the size properties of the t-statistic. This clearly highlights the importance of the lag length selection for carrying on ADF based unit root tests and reinforces our argument that an increase of the lag length does not always lead to a *strict* improvement of the size probabilities.

4 Conclusion

In this paper, our objective was to offer a complementary analysis to the usual Monte-Carlo simulations for evaluating the properties of distributions arising in non stationary autoregressions. More specifically, we investigated the impact of the presence of dependent errors on the asymptotic distributions of the two most important quantities used for testing. We then focused on the properties of the standard method used for whitening the error process by analyzing the connection between the magnitude and sign of the parameters and the number of lags necessary in order to legitimately use the iid errors based distributions. Our framework allows us to distinguish specifically between the different types of error processes. In addition, we showed that a proper normal approximation to the non-standard asymptotic distributions could give valuable hints on the magnitudes of the size distortions. Our results can easily be generalized to the multivariate framework using a multivariate analogue of $\phi(u, v)$ recently obtained by Abadir and Larsson (1994). This can also open the way to multivariate Edgeworth type asymptotic expansions as in Knight and Satchell (1993). Finally, our results can also be used as a starting point for constructing Bartlett corrections for unit root tests. This could be particularly fruitful since the asymptotic distributions are often a poor approximation in moderately sized sample sizes.

Appendix

Proof of Lemma 2.1

We consider the distribution of $T(\hat{\alpha}-1)$ and $t_{\hat{\alpha}}$ in a non stationary AR(1) framework where the error process is given by $\phi(L)u_t = \theta(L)\epsilon_t$ with $\phi(L) = 1-\rho L$ and $\theta(L) = 1+\theta L$. We assume that the ϵ_t process is iid (0,1) and focus on the following statistics

- $T(\hat{\alpha}-\alpha) = \frac{\frac{\sum x_{t-1}u_t}{T}}{\frac{\sum x_{t-1}^2}{T^2}}$
- $t_{\hat{\alpha}} = \frac{\frac{\sum x_{t-1}u_t}{T}}{\left(\frac{\sum x_{t-1}^2}{T^2}\right)^{1/2}(s^2)^{1/2}}$

where $s^2 = \frac{\sum \hat{u}^2}{T}$. From Phillips (1987) we have,

- $\frac{\sum x_{t-1}u_t}{T} \Rightarrow \frac{\sigma^2}{2}[W(1)^2 - \frac{\sigma_u^2}{\sigma^2}]$
- $\frac{\sum x_{t-1}^2}{T^2} \Rightarrow \sigma^2 \int W(r)^2 dr$

where $\sigma_u^2 = \lim_{T\to\infty} \frac{1}{T}\sum E(u_t^2)$ and $\sigma^2 = \lim_{T\to\infty}\frac{1}{T}E(S_T^2)$ with $S_t = \sum_{j=1}^t u_j$ and W(r) is a standard Wiener process. Using the definition of σ^2 and σ_u^2 for our specific error processes (ie. MA(1), AR(1) and ARMA(1,1)) and the continuous mapping theorem, the results follow.

Proof of Lemma 2.2

Assuming ϵ_t to be iid, the joint moment generating function of $\sum x_{t-1}\epsilon_t$ and $\sum x_{t-1}^2$ is $\phi_T(u,v) = E[e^{u\sum x_{t-1}\epsilon_t + v\sum x_{t-1}^2}]$ where $(\sum x_{t-1}\epsilon_t)/T \Rightarrow \int W dW(r)$ and $(\sum x_{t-1}^2)/T^2 \Rightarrow \int W^2$. Furthermore, from White (1958) $\lim_{T\to\infty}\phi_T(u/T, v/T^2) = \psi(u,v)$ where $\psi(u,v)$ is the joint moment generating function of $\int_0^1 W dW$ and $\int_0^1 W^2$. Using the expression for ϕ_T in White (1958) and letting $T\to\infty$ in it, the expression for $\psi(u,v)$ follows immediately.

Proof of Lemma 2.3

We assume $P[Q_2 > 0] = 1$ and follow the same line of proof as in Magnus (1986). We focus on $E[Q_1^m Q_2^{-m/2}]$ only since the proof for $E[Q_2^{-m/2}]$ will follow immediately. Letting $\phi(u,v)$ denote the joint moment generating function of Q_1 and Q_2. We first have,

(i) $Q_1^m = [\frac{\partial^m e^{uQ_1}}{\partial u^m}]_{u=0}$,

(ii) $Q_2^{-m/2} = \Gamma^{-1}(\frac{m}{2}) \int_0^{inf ty} v^{\frac{m}{2}-1-vQ_2} dv.$

The result in (i) is straightforward. The expression for $Q_2^{-m/2}$ in (ii) follows from the definition of the Gamma function $\Gamma(\alpha)$. Indeed,

$$\Gamma(\alpha) \;=\; \int_0^\infty y^{\alpha-1} e^{-y} dy,$$

then making the change of variable $y = vQ_2$ we have,

$$\Gamma(\alpha) \;=\; Q_2^\alpha \int_0^\infty v^{\alpha-1} e^{-vQ_2} dv,$$

leading to

$$Q_2^{-\alpha} \;=\; \Gamma^{-1}(\alpha) \int_0^\infty v^{\alpha-1} e^{-vQ_2} dv.$$

Putting $\alpha = \frac{m}{2}$, we have the desired preliminary result in (ii). Therefore,

$$E[Q_1^m Q_2^{-\frac{m}{2}}] \;=\; \Gamma^{-1}(\frac{m}{2}) E[Q_1^m \int_0^\infty v^{\frac{m}{2}-1} e^{-vQ_2} dv].$$

By Fubini's theorem, we have

$$E[Q_1^m Q_2^{-\frac{m}{2}}] \;=\; \Gamma^{-1}(\frac{m}{2}) \int_0^\infty v^{\frac{m}{2}-1} E[Q_1^m e^{-vQ_2}] dv.$$

Now $E[Q_1^m e^{-vQ_2}] \;=\; E[Q_1^m e^{uQ_1 - vQ_2}]_{u=0} = E[\frac{\partial^m e^{uQ_1-vQ_2}}{\partial u^m}]_{u=0} = \left(\frac{\partial^m \phi(u,-v)}{\partial u^m}\right)_{u=0}$, leading to the desired result.

Proof of Lemma 3.1

Follows directly from Lemm[illegible] been obtained by numerical integratio[illegible] ematica.

Proof of Lemma 3.2

Similar to Lemma 2.1.

References

1. Abadir, K. and Larsson, R. (1994), "The Joint Moment Generating Function of Quadratic Forms in Multivariate Autoregressive Series," *University of Exeter, DP in Economics 94/04.*

2. Agiakoglou, C. and Newbold, P. (1992), "Empirical Evidence on Dickey-Fuller Type Tests," *Journal of Time Series Analysis, 13, 471-483.*

3. Davies, N., M.B Pate and J.D Petrucelli (1985), "Exact Moments of the Sample Cross Correlations of Multivariate Autoregressive Moving Average Time Series," *Sankhya: The Indian Journal of Statistics 47, Series B, 325-337.*

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

5. Evans, G.B.A. and Savin, N.E. (1981), "Testing for Unit Roots:1," *Econometrica 49, 753-779.*

6. Evans, G.B.A. and Savin, N.E. (1984), "Testing for Unit Roots:2," *Econometrica 52, 1241-1269.*

7. Gurland, J. (1948), "Inversion Formulae for the Distribution of Ratios," *Annals of Mathematical Statistics, 19, 228-237.*

8. Hisamatsu, H. and Maekawa, K. (1994), "The Distribution of the Durbin Watson Statistic in Integrated and Near Integrated Models," *Journal of Econometrics, 61, 367-382.*

9. Knight, J.L. and Satchell, S.E. (1993), "Asymptotic Expansions for Random Walks with Normal Errors," *Econometric Theory, 9, 363-376.*

10. Le Breton, A. and Pham, D.T (1989), "On the Bias of the Least Squares Estimator for the First Order Autoregressive Process," *Annals of the Institute of Statistical Mathematics, 41, 555-563.*

11. Magnus, J.R (1986), "The Exact Moments of a Ratio of Quadratic Forms in Normal Variables," *Annales d'Economie et de Statistique, 4, 95-109.*

12. Ng, S. and Perron, P. (1994), "Unit Root Tests in ARMA Models with Dependent Methods for the Selection of the Truncation Lag," *Manuscript, Université de Montréal.*

13. Perron, P. (1989) "The Calculation of the Limiting Distribution of the Least Squares Estimator in a Near Integrated Model," *Econometric Theory, 5, 241-255.*

14. Perron, P. (1994) "The Adequacy of Limiting Distributions in the AR(1) Model with Dependent Errors," *CRDE Working Paper, December 1994, Université de Montréal.*

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

16. Said, S. and Dickey, D.A (1984), "Testing for Unit Roots in Autoregressive Moving Average Models of Unknown Order," *Biometrika, 71, 599-607.*

17. Sawa, T. (1978), "The Exact Moments of the Least Squares Estimator for the Autoregressive Model," *Journal of Econometrics, 8, 159-172.*

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

19. White, J.S. (1958), "The Limiting Distribution of the Serial Coefficient in the Explosive Case," *Biometrika, 74, 535-547.*