# Unit Root Testing in ARMA Models: A Likelihood Ratio Approach

Hernández, Juan R.

Banco de Mexico

2016

# Unit Root Testing in ARMA Models: A Likelihood Ratio Approach

Juan R. Hernández*

April 2016

**Abstract**

In this paper I propose a Likelihood Ratio test for a unit root ($LR$) with a local-to-unity autoregressive parameter embedded in $ARMA(1,1)$ models. By dealing explicitly with dependence in a time series through the Moving Average, as opposed to the long Autorregresive lag approximation, the test shows gains in power and has good small-sample properties. The asymptotic distribution of the test is shown to be independent of the short-run parameters. The Monte Carlo experiments show that the $LR$ test has higher power than the Augmented Dickey Fuller test for several sample sizes and true values of the Moving Average parameter. The exception is the case when this parameter is very close to -1 with a considerably small sample size.

*Correspondence: Juan R. Hernández, Banco de México, Dirección General de Investigación Económica, 5 de Mayo 18, Mexico City, Mexico. E-mail: juan.hernandez@banxico.org.mx

# 1 Introduction

The econometric analysis of an economic time series requires the econometrician to decide how to model two key properties of said series: (i) the inherent dependence between the observations, and (ii) whether the series is stationary or non-stationary by means of a unit root test. Under a small set of assumptions, the Autoregressive Moving Average ($ARMA$) set-up presents itself as a parsimonious and flexible way to address the dependence since it is parametric and, typically, easy to estimate through well known techniques.

Despite the popularity among empiricists of the Autoregressive ($AR$) model, there are a number of arguments in favour of using an $ARMA$ model instead when analysing an error process that features serial correlation. First, modelling explicitly this serial correlation is of value for the econometrician. The temporal aggregation, stemming from the patterns of data collection, has been shown to hide information about the cycle of an economic time series (see Rossana and Seater (1995)); thus creating serial correlation. Second, the same authors conclude that a mixed $ARMA$ model will be better addressing several macroeconomic time series. Third, a mixed $ARMA$ model will yield a parsimonious approximation to the actual process. This in turn improves the power of the unit root test.

A great bulk of available unit root tests, however, is based on estimating "long" $AR$ models, thereby only approximating any dependence of the error process.[1] Among the most popular unit root tests are the Dickey-Fuller ($DF$) test and its augmented version ($ADF$) due to Dickey and Fuller (1979) and Said and Dickey (1984), respectively.

While prominent in its insights and widespread use, these tests present some drawbacks. The $DF$ test is based on the assumption of i.i.d. Gaussian disturbances, but many economic time series feature some form of dependence. The $ADF$ deals with dependence, but this requires the $AR$ order of the linear regression to increase with the sample size. This, however, has a cost: by increasing the $AR$ order, the power of the $ADF$ - that is, the probability of rejecting the null hypothesis- decreases.

The latter introduces a trade-off between testing with low power, on the one hand, and dealing appropriately with the serial correlation on the other. Up to now this issue has been circumvented by choosing the $AR$ order with an information criteria. The test I propose in this paper does not possess said trade-off since dependence is modelled explicitly, and it has better power properties than the $ADF$ test.

[1]See Stock (1994) and Haldrup and Jansson (2007) for a thorough review of the unit root tests available. Semi-parametric tests such as that advocated by Phillips and Perron (1988) are not discussed since the test proposed here is fully parametric.

In their seminal work Elliott, Rothenberg, and Stock (1996), ERS henceforth, proposed the Likelihood Ratio test for the presence of a unit root -the Point Optimal test- and the $DF-GLS$ test. Their tests have better power properties than the $ADF$ test, particularly dealing with deterministic components - a constant mean or a linear trend. ERS worked on a local-to-unity framework, and this allows easy comparisons across models. The framework also allows simpler ways to write both the test statistic and its asymptotic distribution. Through the Neyman-Pearson Lemma, ERS get an upper bound for the power of the test in an $AR(p)$ process, which overlaps the Gaussian Power Envelope. The ERS tests have been part of the unit root testing practice they appeared and are now included in econometrics software as a standard component.

The mechanics in the computation of the $DF-GLS$ test statistic can be summarised in two steps: (i) de-trend the economic time series through Generalised Least Squares ($GLS$) or Maximum Likelihood ($ML$), (ii) test the de-trended series for a unit root with the $ADF$ test. This draws the econometrician to face the same trade-off as in the implementation of the $ADF$ test. The test I propose differs from those in ERS in two respects: (i) the avoidance of the aforementioned trade-off and (ii) the computation is made in one-step.

In this paper I outline the construction of a likelihood ratio unit root test for $ARMA(1,1)$ models with no deterministic terms (i.e. a constant and a trend). Focusing in the particular case of the $ARMA(1,1)$ may seem restrictive, but can be justified on the following basis. If the data generating process behind an economic time series is suspected to be a continuous time process (e.g. consumption or pricing decisions), but the econometrician only has access to data collected in discrete points in time, the *exact discrete* representation of the data generating process will often be an $ARMA(1,1)$ model as shown by Bergstrom (1984) and Chambers (2009). This model also buys a number of benefits. First, the main parameters driving the dynamics are included. Second, consistency results and asymptotic distributions are easier to present. Third, as in Said and Dickey (1985), the $ARMA(1,1)$ case provides a foundation for more general cases.

This paper contains 6 additional sections. Section 2 outlines the DGP and the model to be estimated. In section 3 I present the set-up to estimate through ML and then derive the asymptotic behaviour of the estimates. The $LR$ test is presented in section 4, followed by the empirical analysis in section 5. Section 6 has an application of the test to inflation series in several countries. Finally, section 7 presents some concluding remarks.

## Notation

Throughout this paper I use the following notation: $\longrightarrow$ denotes convergence; $\longrightarrow_p$ convergence in probability; $\longrightarrow_d$ convergence in distribution; $\Rightarrow$ weak convergence. For matrix operators I use: $\|A\| = tr^{1/2}\{A'A\}$; $\|A\|_1 = \sum_t \sum_s |a_{ts}|$; $\det|A|$ for determinant of matrix $A$; $tr\{A\}$ for trace of matrix $A$.

# 2 The Data Generating Process

In this section I introduce the elements on which the rest of the paper is built. In particular, the probability space, the data generating process and the error process are introduced. This will pave the way for the estimation set-up in the following sections.

The framework on which I develop the analysis requires a sample space $\Omega_n$ and $\mathcal{F}_n$, the $\sigma - field$ of $\Omega_n$. Moreover, $\Theta$ is the parameter space with typical element $\theta_n$. $P_{n,\theta_n}$ is a probability measure for $n$ observations of a time series indexed by $\theta_n$. Thus, the triple $(\Omega_n, \mathcal{F}_n, P_{n,\theta_n})$ forms a complete probability space underlying the DGP as stated in the next assumption.

**Assumption 2.A** (Data Generating Process)**.** *The observed time series, $\{y_t\}_{t=0}^n$, is generated by*

$$\Delta y_t = (\rho_0 - 1)y_{t-1} + u_{0,t}, \tag{2.1}$$

$$\rho_0 - 1 = \frac{c}{n},\ c < 0. \tag{2.2}$$

The local-to-unity set-up, given by (2.2), goes back at least to Phillips (1987a) and has been exploited greatly in the unit root testing literature, while the stochastic behaviour of the error process $\{u_{0,t}\}_{t=0}^n$ in (2.1) can be thought as a draw from the probability space showing linear dependence with behaviour summarised by the following assumption.

**Assumption 2.B** (Error Process)**.** *The error process $\{u_{0,t}\}_{t=0}^n$, depends on a parameter vector $\theta_{20} = (\alpha_0, \sigma_0^2)'$ and satisfies: (i) $\varepsilon_t \sim i.i.d(0, \sigma_0^2)$. (ii) $\varepsilon_s = 0$ for $s \leq 0$. (iii) $u_{0,t} = \varepsilon_t - \alpha_0 \varepsilon_{t-1}$. (iv) $\sup_t E|u_{0,t}|^\delta < \infty$ for $\delta > 2$.*

Both (i) and (ii) are standard in econometric analysis, while (iii) sets up the $MA$ and (iv) is necessary in view that a $MA$ process is a special case of a mixing process where $\delta$ controls the dispersion as discussed in Phillips (1987a). Moreover, the assumed value of $\delta$ is needed as the ML estimators will typically involve the second moments of $\sigma_0^2$. The error process sequence $\{u_{0,t}\}_{t=0}^{n}$ can also be written in vector form by stacking its last $n$ elements into the $n \times 1$ vector $u_0$.

Write the observed time series in vector form stacking the observations so that $\Delta y = (\Delta y_1, \ldots, \Delta y_n)'$ is a $n \times 1$ vector, similarly $y_{-1} = (y_0, \ldots, y_{n-1})'$ and $u = (u_1, \ldots, u_n)'$, to get

$$u(\theta_{1n}) = \Delta y - (\theta_{1n} - 1)y_{-1}.$$

Assumption 2.B implies that the error vector, $u_0$, has a covariance matrix $\Gamma(\theta_{20})$ with typical $jk$ element $\gamma(j-k) = Cov(u_{0,t-j}, u_{0,t-k})$. Moreover, $\Gamma(\theta_{20})$ satisfies: (i) $\gamma(0) = (1+\alpha_0^2)\sigma_0^2$. (ii) $\gamma(1) = \gamma(-1) = -\alpha_0\sigma_0^2$. (iii) $\gamma(j-k) = 0$ for all $|j-k| > 1$. The latter implies that $\Gamma(\theta_{20})$ is a band-Toeplitz matrix containing $\gamma(j)$ in its $jth$-band. Finally, define the (finite) long run variance $\sigma_u^2 = \sum_{k=-\infty}^{\infty} \gamma_k = (1-\alpha_0)^2\sigma_0^2$.

# 3 Maximum Likelihood Estimation and Asymptotics

In this section the estimation is presented along with a series of preliminary definitions and Lemmas that will ease the introduction of the main asymptotic results. After the set-up of the ML estimation problem, I show that the obtained estimators are consistent, a task that will require to split the objective function in two parts: one containing long-run parameters and a second part with short-run parameters driving the dependence. The section concludes with the derivation of the asymptotic distribution of the ML estimator.

ML estimation requires the properties of the parameter space to be clearly defined, which I present in the next assumption.

**Assumption 3.A** (Parameter Space)**.** *(a) The parameter space $\Theta$ is (i) convex and (ii) compact. (b) Write $\Theta = \Theta_1 \times \Theta_2$, where $\Theta_1$ contains the long run parameters (i.e. $\theta_{1n} = \rho_n$) and $\Theta_2$ contains only short-run parameters (i.e. $\theta_{2n} = (\alpha_n, \sigma_n^2)'$). (c) $\Theta_2$ contains only elements that ensure $\Gamma^{-1}(\theta_{2n})$ exists.*

To estimate the parameter vector $\theta_n = (\theta_{1n}, \theta_{2n})'$ write the log-likelihood function for $\{u_{0,t}\}_{t=0}^{n}$,[2]

$$l^n(\theta_n) = -\frac{n}{2}\ln|2\pi| - \frac{1}{2}\ln\det|\Gamma(\theta_{2n})| - \frac{1}{2}u(\theta_{1n})'\Gamma^{-1}(\theta_{2n})u(\theta_{1n}), \tag{3.1}$$

[2]The likelihood function is constructed "as if" the disturbances are normally distributed, anticipating an approximation based on a limiting distribution.

or equivalently, define the loss function $Q_n(\theta_n) = -2l^n(\theta_n) - n \ln|2\pi|$ so that the objective is now to find the minimizer of

$$Q_n(\theta_n) = \ln\det|\Gamma(\theta_{2n})| + u(\theta_{1n})'\Gamma^{-1}(\theta_{2n})u(\theta_{1n}). \tag{3.2}$$

Let $\Gamma_n = \Gamma(\theta_{2n})$ to ease notation, and following Saikkonen (1995) add and subtract $\theta_{10}y_{-1}$ from $u_n = u(\theta_{1n})$ to obtain

$$Q_n(\theta_n) = Q_{1n}(\theta_n) + Q_{2n}(\theta_{2n}),$$

where

$$Q_{1n}(\theta_n) = (\theta_{1n} - \theta_{10})^2 y'_{-1}\Gamma_n^{-1}y_{-1} - 2(\theta_{1n} - \theta_{10})y'_{-1}\Gamma_n^{-1}u_0, \tag{3.3}$$

$$Q_{2n}(\theta_{2n}) = \ln\det|\Gamma_n| + u'_0\Gamma_n^{-1}u_0, \tag{3.4}$$

where by making explicit the distance $(\theta_{1n} - \theta_{10})$, consistency results can be clearly presented and additional advantages are presented in the form of two remarks for future reference.

**Remark 3.B.** *The loss function $Q_{1n}(\theta_n)$ depends on both long and short-run parameters. The loss function $Q_{2n}(\theta_{2n})$ as defined in (3.4) depends only on short-run parameters.*

**Remark 3.C.** $Q_{1n}(\theta_0) = 0$. *$Q_{2n}(\theta_{2n})$ is equivalent to the estimation problem of a stationary MA(1) process.*

## 3.1 Preliminaries

To present the main result on consistency of this paper, I need to introduce a number of preliminary results and notation. The Wiener Processes $\mathcal{W}(r)$, as characterised in White (2001), and the Ornstein-Uhlenbeck Process defined by Phillips (1987b).

**Definition 3.D** (Wiener Process)**.** *Within $(\Omega_n, \mathcal{F}_n, P_{n,\theta_n})$. Then $\mathcal{W} : [0,\infty) \times \Omega \to \mathbb{R}$ is a Wiener Process if for $r \in [0,\infty)$, $\mathcal{W}(r,\cdot)$ is measurable-$\mathcal{F}_n$, and in addition:*

i. *The process starts at zero: $P_{n,\theta_n}[\mathcal{W}(0,\cdot) = 0] = 1$.*

ii. *The increments are independent: If $0 \leq r_0 \leq r_1 \leq \cdots \leq r_k < \infty$, then $\mathcal{W}(r_i,\cdot) - \mathcal{W}(r_{i-1},\cdot)$ is independent of $\mathcal{W}(r_j,\cdot) - \mathcal{W}(r_{j-1},\cdot)$, $j = 1,\ldots,k, j \neq i$ for all $i = 1,\ldots k$.*

*iii. The increments are normally distributed: For* $0 \leq a \leq b < \infty$ *the increment* $\mathcal{W}(b, \cdot) - \mathcal{W}(a, \cdot)$ *is distributed as* $N(0, b-a)$.

*To ease notation write:* $\mathcal{W} = \mathcal{W}(r)$ *and* $\int_0^\infty \mathcal{W} = \int_0^\infty \mathcal{W}(r)dr$.

**Definition 3.E** (Ornstein-Uhlenbeck Process)**.** *For* $r$ *and* $s$ *real numbers, the functional* $\mathcal{J}_c(r)$ *of the form*

$$\begin{aligned} \mathcal{J}_c(r) &= \int_0^r \exp[(r-s)c]d\mathcal{W}(s) \\ &= \mathcal{W}(r) + c\int_0^r \exp[(r-s)c]\mathcal{W}(s)ds, \end{aligned}$$

*is the Ornstein-Uhlenbeck Process associated with* $c$*. The process satisfies:*

*i. Being the solution to the stochastic differential equation* $d\mathcal{J}_c(r) = c\mathcal{J}_c(r)dr + d\mathcal{W}(r)$.

*ii.* $\mathcal{J}_c(0) = 0$.

*iii.* $\mathcal{J}_c(r)$ *is distributed as* $N\left(0, \frac{\exp[2rc]-1}{2c}\right)$.

*To ease notation write* $\mathcal{J}_c = \mathcal{J}_c(r)$ *and* $\int_0^1 \mathcal{J}_c = \int_0^1 \mathcal{J}_c(r)dr$.

**Remark 3.F.** *With* $\mathcal{J}_c$ *as given in Definition 3.E, in the particular case of* $c = 0$, $\mathcal{J}_c$ *collapses to* $\mathcal{W}$.

Taking advantage of $\Gamma_0^{-1}$ being itself a covariance matrix (i.e. symmetric and positive definite), I introduce the following Definition and Lemma.

**Definition 3.G** (Spectral Density of $\Gamma_0^{-1}$)**.** *Let* $g_{ts}$ *denote the* $ts$ *element of* $\Gamma_0^{-1}$*and with spectral function* $g(\omega)$ *with* $\omega \in (-\pi, \pi]$ *so that each of its elements can be recovered using*

$$\begin{aligned} g_{ts} &= \int_0^1 e^{i(t-s)\omega} g(\omega)d\omega, \\ g(\omega) &= \frac{1}{2\pi}\sum_{j=-\infty}^{\infty} e^{-ij\omega} g_j. \end{aligned}$$

**Lemma 3.H.** *If* $\{y_t\}_{t=0}^n$ *and* $\{u_{0,t}\}_{t=0}^n$ *are given by assumptions 2.A and 2.B, then*

$$n^{-1}y'_{-1}\Gamma_n^{-1}u_0 \Rightarrow 2\pi g(0)\left[\sigma_u^2\int_0^1 \mathcal{J}_c d\mathcal{W} + \frac{\sigma_u^2 - \gamma_0}{2}\right] + \sum_{j=-\infty}^{\infty} g_j\Phi_j, \tag{3.5}$$

$$n^{-2}y'_{-1}\Gamma_n^{-1}y_{-1} \Rightarrow 2\pi g(0)\sigma_u^2\int_0^1 \mathcal{J}_c^2, \tag{3.6}$$

*where* $\Phi_j = 0$ *if* $j = 0$*;* $\Phi_j = \sum_{k=0}^{j-1} \gamma_k$ *if* $j > 0$*; and* $\Phi_j = -\sum_{k=1}^{|j|} \gamma_k$ *if* $j < 0$*.*

**Proof.** I prove first (3.5). Start by writing

$$\begin{aligned} n^{-1} y_{-1}' \Gamma_n^{-1} u_0 &= n^{-1} \sum_{t=1}^{n} \sum_{s=1}^{n} y_{t-1} u_{0,s} g_{ts} \\ &= n^{-1} \sum_{t=1}^{n} \sum_{s=1}^{n} y_{t-1} u_{0,s} \int_0^1 e^{i(t-s)\omega} g(\omega) d\omega \\ &= n^{-1} \sum_{t=1}^{n} \sum_{s=1}^{n} y_{t-1} u_{0,s} \sum_{j=-\infty}^{\infty} g_j \int_0^1 e^{i(t-s-j)\omega} d\omega \\ &= \sum_{j=-\infty}^{\infty} g_j \left[ n^{-1} \sum_{t=1}^{n} y_{t-1} u_{0,t-j} \right], \end{aligned} \tag{3.7}$$

where the second and third lines are obtained by direct substitution, and the fourth line follows from orthogonality of the integral for $s = t - j$. Now I take cases, if $j = 0$ in (3.7)

$$n^{-1} \sum_{t=1}^{n} y_{t-1} u_{0,t-j} \Rightarrow \sigma_u^2 \int_0^1 \mathcal{J}_c d\mathcal{W} + \frac{\sigma_u^2 - \gamma_0}{2}, \tag{3.8}$$

from Lemma 1 in Phillips (1987b). If $j > 0$ I need to use the backward solution of the stochastic difference equation (2.1) at $t - 1$,

$$y_{t-1} = \theta_{10}^j y_{t-j-1} + \sum_{k=0}^{j-1} \theta_{10}^k u_{0,t-1-k},$$

which together with the bracket expression in (3.7) yields

$$n^{-1} \sum_{t=1}^{n} y_{t-1} u_{0,t-j} = \theta_{10}^j n^{-1} \sum_{t=1}^{n} y_{t-j-1} u_{0,t-j} + n^{-1} \sum_{t=1}^{n} \sum_{k=0}^{j-1} \theta_{10}^k u_{0,t-1-k} u_{0,t-j},$$

the first term has as asymptotic limit (3.8) since $\lim_{n\to\infty} \theta_{10}^k = 1$ for all $k$, and the second term converges to

$$n^{-1} \sum_{t=1}^{n} \sum_{k=0}^{j-1} \theta_{10}^k u_{0,t-1-k} u_{0,t-j} \longrightarrow_p \sum_{k=0}^{j-1} \gamma_k.$$

Finally, if $j < 0$ I need the forward solution of the stochastic difference equation, this is

$$y_{t-j} = \theta_{10}^{-j+1} y_{t-1} + \sum_{k=0}^{|j|} \theta_{10}^{k} u_{0,t+k},$$

which makes

$$n^{-1}\sum_{t=1}^{n} y_{t-1}u_{0,t-j} = \theta_{10}^{-(-j+1)} n^{-1}\sum_{t=1}^{n} y_{t-j-1}u_{0,t-j} - \theta_{10}^{-(-j+1)} n^{-1}\sum_{t=1}^{n}\sum_{k=0}^{|j|}\theta_{10}^{k} u_{0,t+k}u_{0,t-j},$$

where, again, the first term converges to (3.8) and the the second has limit

$$-\theta_{10}^{-(-j+1)} n^{-1}\sum_{t=1}^{n}\sum_{k=0}^{|j|}\theta_{10}^{k} u_{0,t+k}u_{0,t-j} \longrightarrow_p -\sum_{k=1}^{|j|}\gamma_k,$$

adding across values of $j$ completes the proof of (3.5).

To prove (3.6) start by noting that

$$\begin{aligned} n^{-2} y_{-1}' \Gamma_0^{-1} y_{-1} &= n^{-2}\sum_{t=1}^{n}\sum_{s=1}^{n} y_{t-1}y_{s-1}g_{ts} \\ &= \sum_{j=-\infty}^{\infty} g_j \left[ n^{-2}\sum_{t=1}^{n} y_{t-1}y_{t-j-1} \right], \end{aligned}$$

if I follow the same steps used to obtain (3.7). If $j = 0$ I use again Lemma 1 from Phillips (1987b) to obtain

$$n^{-2} y_{-1}' \Gamma_0^{-1} y_{-1} \Rightarrow \sigma_u^2 \int_0^1 \mathcal{J}_c^2. \tag{3.9}$$

When $j > 0$, using the backward solution note that I can write

$$n^{-2}\sum_{t=1}^{n} y_{t-1}y_{t-j-1} = \theta_{10}^{-j} n^{-2}\sum_{t=1}^{n} y_{t-1}^2 - n^{-2}\theta_{10}^{-j}\sum_{t=1}^{n} y_{t-1}\sum_{k=0}^{j-1}\theta_{10}^{k} u_{0,t-1-k},$$

the first term converges to (3.9) while the second is $o_p(1)$ in view of (3.5). The same argument shows that if $j < 0$ the limit I get is (3.9), adding across $j$ completes the proof.

■

Now I define a few more objects, following Saikkonen (1995) let $\eta > 0$ (not necessarily the same across results), $\nu > 0$, $\delta_1, \delta_2 > 0$. Define the closed balls

$$B_{1,\nu} = \left\{\theta_{1n} \in \Theta_1 : |\theta_{1n} - \theta_{10}| \leq \frac{\delta_1}{n^{\nu}}\right\}, \tag{3.10}$$

$$\overline{B}_{1,\nu} = \left\{\theta_{1n} \in \Theta_1 : |\theta_{1n} - \theta_{10}| \geq \frac{\delta_1}{n^{\nu}}\right\}, \tag{3.11}$$

$$\overline{B}_2 = \left\{\theta_{2n} \in \Theta_2 : \|\theta_{2n} - \theta_{20}\| \geq \delta_2\right\}. \tag{3.12}$$

Also let

$$\begin{aligned} M_1 &= |y_{-1}'\Gamma_n^{-1}u_0| = O_p(n), \\ M_2 &= |y_{-1}'\Gamma_n^{-1}y_{-1}| = O_p(n^2), \end{aligned}$$

as implied by Lemma 3.H, and

$$\overline{M}_i = \sup_{\theta_{2n}} M_i; \ \underline{M}_i = \inf_{\theta_{2n}} M_i.$$

## 3.2 Consistency of $\widehat{\theta}_n$

Proving consistency of the ML estimator $\widehat{\theta}_{1n}$ in the framework outlined above is a task considerably different from that in the textbook case where $\{y_t\}_{t=0}^{n}$ is stationary. The main reason, as explained by Chambers and McCrorie (2007) who found themselves in the same conundrum, is that $Q_n(\theta_n)$ does not satisfy weak uniform convergence. Specifically, $Q_n(\cdot)$ converges at different rates in different directions of the parameter space. To overcome this difficulty they rely on the results developed by Saikkonen (1995), and I will use them here extensively.

**Proposition 3.I** (Consistency of $\widehat{\theta}_n$)**.** *Let $\nu \in [0, 1)$. Given assumptions 2.A and 2.B, ML estimate $\widehat{\theta}_n$ satisfies $n^{\nu}(\widehat{\theta}_{1n} - \theta_{10}) \longrightarrow_p 0$ and $\widehat{\theta}_{2n} - \theta_{20} \longrightarrow_p 0$.*

The proof of Proposition 3.I is based on Saikkonen (1995) and is divided in several Lemmas for ease of presentation. Lemma 3.J establishes consistency of the long run parameter when $\nu = 0$, Lemma 3.K does so for $\nu \in (0, 1)$, and Lemmas 3.L and 3.M prove the consistency of the short run parameter vector.

**Lemma 3.J.** *If $Q_n(\theta_n)$ is given by (3.2) and $\nu = 0$, then*

$$\lim_{n\to\infty} P_{n,\theta_0}\left\{\inf_{\theta_n \in \overline{B}_{1,0}\times\Theta_2} Q_n(\theta_n) - Q_n(\theta_0) > 0\right\} = 1.$$

*Equivalently,*

$$\inf_{\theta_n \in \overline{B}_{1,0} \times \Theta_2} n^{-1} Q_{1n}(\theta_n) + \inf_{\theta_n \in \overline{B}_{1,0} \times \Theta_2} n^{-1} Q_{2n}(\theta_{2n}) - n^{-1} Q_{2n}(\theta_{20}) > \eta. \tag{3.13}$$

**Proof.** I prove (3.13) in two parts. First I claim $\inf_{\theta_n \in \overline{B}_{1,0} \times \Theta_2} n^{-1} Q_{1n}(\theta_n) > 0$. To see this note

$$n^{-1} Q_{1n}(\theta_n) \geq n^{-1} |\theta_{1n} - \theta_{10}|^2 \underline{M}_2 - 2n^{-1} |\theta_{1n} - \theta_{10}| \overline{M}_1,$$

or

$$\inf_{\theta_n \in \overline{B}_{1,0} \times \Theta_2} n^{-1} Q_{1n}(\theta_n) \geq n^{-1} \delta_1^2 \underline{M}_2 - 2n^{-1} \delta_1 \overline{M}_1,$$

where considering elements of $\Theta_1$ on the set defined in (3.11) the first element is diverging to $+\infty$ despite $|\theta_{1n} - \theta_{10}|^2$ being bounded, since $\underline{M}_2 = O_p(n^2)$. The second element is $O_p(1)$ since $\Theta_1$ is bounded and it can be shown that $\overline{M}_1 = O_p(n)$ from (3.5); thus $n^{-1} Q_{1n}(\theta_n) > 0$.

For the second part, let $\widetilde{\theta}_{2n}$ be the (infeasible) ML estimate for the stationary MA(1) process in $Q_{2n}(\theta_{2n})$. Note that standard theory on ML estimation now applies and $\widetilde{\theta}_{2n} - \theta_{20} = O_p(n^{-1/2})$. I claim that $\inf_{\theta_n \in \overline{B}_{1,0} \times \Theta_2} n^{-1} Q_{2n}(\widetilde{\theta}_{2n}) - n^{-1} Q_{2n}(\theta_{20}) = o_p(1)$. To see this, write

$$\begin{aligned} \inf_{\theta_n \in \overline{B}_{1,0} \times \Theta_2} n^{-1} Q_{2n}(\widetilde{\theta}_{2n}) - n^{-1} Q_{2n}(\theta_{20}) \quad &\geq \quad \inf_{\theta_n \in \overline{B}_{1,0} \times \Theta_2} n^{-1} \ln \det |\widetilde{\Gamma}_n| - n^{-1} \ln \det |\Gamma_0| \\ &+ \quad \inf_{\theta_n \in \overline{B}_{1,0} \times \Theta_2} n^{-1} u_0' \widetilde{\Gamma}_n^{-1} u_0 - n^{-1} u_0' \Gamma_0^{-1} u_0. \end{aligned}$$

The first term in the right of the last expression is $o_p(1)$ as proven in Lemma 1 of Yao and Brockwell (2006). The same authors show that the second term is $o_p(1)$ in their Theorem 1. ■

**Lemma 3.K.** *If $Q_n(\theta_n)$ is given by (3.2) and $\nu \in (0,1)$, then*

$$\lim_{n \to \infty} P_{n,\theta_0} \left\{ \inf_{\theta_n \in \overline{B}_{1,\nu} \times \Theta_2} Q_n(\theta_n) - Q_n(\theta_0) > 0 \right\} = 1.$$

*Equivalently,*

$$\inf_{\theta_n \in \overline{B}_{1,\nu} \times \Theta_2} Q_{1n}(\theta_n) + \inf_{\theta_n \in \overline{B}_{1,\nu} \times \Theta_2} Q_{2n}(\theta_{2n}) - Q_{2n}(\theta_{20}) > \eta. \tag{3.14}$$

**Proof.** Here I prove (3.14), starting with the claim: $\inf_{\theta_n \in \overline{B}_{1,\nu} \times \Theta_2} Q_{1n}(\theta_n) > 0$. To see this write

$$Q_{1n}(\theta_n) \geq |\theta_{1n} - \theta_{10}|^2 \underline{M}_2 - 2|\theta_{1n} - \theta_{10}|\overline{M}_1$$

$$= |\theta_{1n} - \theta_{10}|^2 \underline{M}_2 \left[1 - \frac{2\overline{M}_1}{|\theta_n - \theta_{10}|\underline{M}_2}\right].$$

Considering $\theta_n \in \overline{B}_{1,\nu}$ as defined in (3.11), I can now write

$$\inf_{\theta_n \in \overline{B}_{1,\nu} \times \Theta_2} Q_{1n}(\theta_n) \geq \delta_1^2 n^{-2\nu} \underline{M}_2 \left[1 - \frac{2n^{-1}\overline{M}_1}{n^{-(1+\nu)}\delta_1 \underline{M}_2}\right].$$

The term within brackets in the last expression converges to 1 since $\nu \in (0,1)$, while $\delta_1^2 n^{-2\nu} \underline{M}_2$ is diverging to $+\infty$ proving the claim.

Regarding the short-term part of (3.14), again, let $\widetilde{\theta}_{2n}$ be the (infeasible) ML estimate for the stationary $MA(1)$ process, so that standard results on ML estimation of stationary process apply (e.g. from Hannan (1973)) and I have $\widetilde{\theta}_{2n} - \theta_{20} = O_p(n^{-1/2})$ and $\inf_{\theta_n \in \overline{B}_{1,\nu} \times \Theta_2} Q_{2n}(\widetilde{\theta}_{2n}) - Q_{2n}(\theta_{20}) = O_p(1)$. ■

**Lemma 3.L.** *If $Q_{1n}(\theta_n)$ is given by (3.3) and $\nu \in (1/2, 1)$, then*

$$\lim_{n\to\infty} P_{n,\theta_0} \left\{ \sup_{\theta_n \in B_{1,\nu} \times \overline{B}_2} \left|n^{-1} Q_{1n}(\theta_n)\right| \leq \eta \right\} = 1.$$

**Proof.** Write directly

$$\sup_{\theta_n \in B_{1,\nu} \times \overline{B}_2} |n^{-1} Q_{1n}(\theta_n)| \leq \sup_{\theta_{1n} \in B_{1,\nu} \times \overline{B}_2} |\theta_n - \theta_{10}|^2 n^{-1} \overline{M}_2 + 2 \sup_{\theta_{1n} \in B_{1,\nu} \times \overline{B}_2} |\theta_{1n} - \theta_{10}| n^{-1} \overline{M}_1.$$

Recall $\nu \in (1/2, 1)$ and, from Lemma 3.H, $M_1 = O_p(n)$ and $M_2 = O_p(n^2)$. Since $\theta_n \in B_{1,\nu} \times \overline{B}_2$, with these sets defined in (3.10) and (3.12), note that both terms on the right are $o_p(1)$. Start with the first term,

$$\sup_{\theta_{1n} \in B_{1,\nu} \times \overline{B}_2} |\theta_n - \theta_{10}|^2 n^{-1} \overline{M}_2 \leq \delta_1^2 n^{-(1+2\nu)} \overline{M}_2 = o_p(1).$$

Finally,

$$\sup_{\theta_{1n} \in B_{1,\nu} \times \overline{B}_2} |\theta_{1n} - \theta_{10}| n^{-1} \overline{M}_1 \leq \delta_1 n^{-(1+\nu)} \overline{M}_1 = o_p(1).$$

■

**Lemma 3.M.** *If $Q_{2n}(\theta_{2n})$ is given by (3.4), then*

$$\lim_{n\to\infty} P_{n,\theta_0}\left\{\inf_{\theta_{2n}\in\overline{B}_2} n^{-1}Q_{2n}(\theta_n) - n^{-1}Q_{2n}(\theta_0) > \eta\right\} = 1.$$

**Proof.** Write

$$\begin{aligned}\inf_{\theta_{2n}\in\overline{B}_2} n^{-1}Q_{2n}(\theta_n) - n^{-1}Q_{2n}(\theta_0) \;\geq\; & \inf_{\theta_{2n}\in\overline{B}_2} n^{-1}\ln\det|\Gamma_n| - n^{-1}\ln\det|\Gamma_0| \\ &+ \inf_{\theta_{2n}\in\overline{B}_2} n^{-1}u_0'\Gamma_n^{-1}u_0 - n^{-1}u_0'\Gamma_0^{-1}u_0.\end{aligned}$$

The first term on the right is $o_p(1)$ from Lemma 1 of Yao and Brockwell (2006). The second term satisfies standard ML theory arguments for stationary time series. In particular, for $\theta_{2n}\in\overline{B}_2$ defined in (3.12), I have $n^{-1}u_0'\Gamma_n^{-1}u_0 > n^{-1}u_0'\Gamma_0^{-1}u_0$ as proved by the same authors in their Theorem 1. This proves the Lemma. ■

**Proof of Proposition 3.I.** From Lemmas 3.J and 3.K,

$$\lim_{n\to\infty} P_{n,\theta_0}\left\{\inf_{\theta_n\in\overline{B}_{1,\nu}\times\Theta_2} Q_n(\theta_n) - Q_n(\theta_0) > 0\right\} = 1,$$

which together with a standard convergence result such as Theorem 21.6 from Davidson (1994) implies $n^{\nu}(\widehat{\theta}_{1n} - \theta_{10}) \longrightarrow_p 0$.

Also, from Lemmas 3.L and 3.M it follows that

$$\lim_{n\to\infty} P_{n,\theta_0}\left\{\inf_{\theta_n\in\Theta_1\times\overline{B}_2} Q_n(\theta_n) - Q_n(\theta_0) > 0\right\} = 1,$$

and again with the aid of Theorem 21.6 from Davidson (1994) I conclude $\widehat{\theta}_{2n} - \theta_{20} \longrightarrow_p 0$. ■

## 3.3 Asymptotic Distribution of $\widehat{\theta}_n$

After I have proved consistency of $\widehat{\theta}_n$, in this section I present its asymptotic distribution. The textbook way to obtain the asymptotic distribution of the ML estimate is through a mean value expansion and the asymptotics of the Score vector and the sample information matrix. To follow this road, I would need to prove that the sample information matrix is stochastic equicontinous as explained by Saikkonen (1995). The Score vector of the ML problem in this paper yields closed forms for $\widehat{\theta}_{1n} - \theta_{10}$ and $\widehat{\theta}_{2n} - \theta_{20}$. Thus, instead of proving stochastic equicontinuity, I apply the expansions proposed by Phillips (1991) and Saikkonen (2001) and used by Chambers and McCrorie (2007).

The *normalised* Score vector $s_n(\theta_n)$, corresponding to the optimization problem (3.2) is partitioned accordingly to the previous section into $\theta_n = (\theta_{1n}, \theta_{2n}')'$,

$$s_{1n}(\theta_n) = 2n^{-1}\left[(\theta_{1n} - \theta_{10})y_{-1}'\Gamma_n^{-1}y_{-1} - y_{-1}'\Gamma_n^{-1}u_0\right], \tag{3.15}$$

$$s_{2n,j}(\theta_n) = n^{-1/2}\frac{\partial Q_{1n}(\theta_n)}{\partial \theta_{2n,j}} + n^{-1/2}\frac{\partial Q_{2n}(\theta_n)}{\partial \theta_{2n,j}}, \; j = 1, 2, \tag{3.16}$$

where

$$\begin{aligned}
\frac{\partial Q_{1n}(\theta_n)}{\partial \theta_{2n,j}} &= (\theta_{1n} - \theta_{10})^2 y_{-1}'\frac{\partial \Gamma_n^{-1}}{\partial \theta_{2n,j}}y_{-1} - 2(\theta_{1n} - \theta_{10})y_{-1}'\frac{\partial \Gamma_n^{-1}}{\partial \theta_{2n,j}}u_0, \\
\frac{\partial Q_{2n}(\theta_n)}{\partial \theta_{2n,j}} &= tr\left\{\Gamma_n^{-1}\frac{\partial \Gamma_n}{\partial \theta_{2n,j}}\right\} + u_0'\frac{\partial \Gamma_n^{-1}}{\partial \theta_{2n,j}}u_0,
\end{aligned}$$

$j = 1, 2$ and, as usual, solving the system $s_n(\widehat{\theta}_n) = 0$ yields the ML estimate for a given $n$, $\widehat{\theta}_n$. I present the closed form of $\widehat{\theta}_1$ for future reference, and then its asymptotic distribution. I write $\widehat{\Gamma}_n = \Gamma(\widehat{\theta}_{2n})$ in the rest of the paper.

**Remark 3.N.** *$s_{1n}(\theta_n)$ as given in (3.15) implies a closed form for $\widehat{\theta}_{1n}$,*

$$\widehat{\theta}_{1n} - \theta_{10} = \frac{y_{-1}'\widehat{\Gamma}_n^{-1}u_0}{y_{-1}'\widehat{\Gamma}_n^{-1}y_{-1}}.$$

**Proposition 3.O** (Asymptotic Distribution of $\widehat{\theta}_{1n}$)**.** *Given $s_{1n}(\theta_n)$ from (3.15), then*

$$n(\widehat{\theta}_{1n} - \theta_{10}) \Rightarrow \frac{\int_0^1 \mathcal{J}_c d\mathcal{W}}{\int_0^1 \mathcal{J}_c^2}.$$

An important feature of the result presented in this Proposition is the absence of short-term parameters, $\theta_{20}$. The following Lemmas and Definitions shed light on why this is so, and will ease the presentation of the proof of the Proposition.

**Lemma 3.P.** *If* $\widehat{\theta}_{2n} - \theta_{20} \longrightarrow_p 0$ *, then*

$$n^{-2} y_{-1}' \widehat{\Gamma}_n^{-1} y_{-1} - n^{-2} y_{-1}' \Gamma_0^{-1} y_{-1} \longrightarrow_p 0,$$

$$n^{-1} y_{-1}' \widehat{\Gamma}_n^{-1} u_0 - n^{-1} y_{-1}' \Gamma_0^{-1} u_0 \longrightarrow_p 0.$$

**Proof.** The proof is contained in Chambers and Hernandez (2015) Lemma 4. ■

**Definition 3.Q** (Spectral Density of $\Gamma_0$)**.** *Denote the spectral density of* $\{u_{0,t}\}_{t=0}^n$ *by* $f(\omega;\theta_{20})$*, with* $\omega \in (-\pi, \pi]$*, the* $ts$ *element of* $\Gamma_0$*,* $\gamma_{t-s}$*, can be recovered by*

$$\begin{aligned} \gamma_{t-s} &= \int_{-\pi}^{\pi} e^{i(t-s)\omega} f(\omega;\theta_{20}) d\omega, \\ f(\omega;\theta_{20}) &= \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} e^{-ij\omega} \gamma_{t-j}, \end{aligned}$$

*consequently,* $2\pi f(0;\theta_{20}) = \sigma_u^2$*.*

**Definition 3.R** (Spectral Density of $\Psi$)**.** *Let* $\Psi$ *be a* $n \times n$ *matrix with* $ts$ *element,* $\psi_{t-s}$*, satisfying*

$$\begin{aligned} \psi_{t-s} &= \int_{-\pi}^{\pi} e^{i(t-s)\omega} \left[4\pi^2 f(\omega;\theta_{20})\right]^{-1} d\omega, \\ \left[4\pi^2 f(\omega;\theta_{20})\right]^{-1} &= \sum_{j=-\infty}^{\infty} e^{-ij\omega} \psi_{t-j}. \end{aligned}$$

**Lemma 3.S.** *If* $\Gamma_0$ *is the true covariance matrix of the error process* $\{u_{0,t}\}_{t=0}^n$ *and if* $\Psi$ *is given by Definition 3.R , then*

$$n^{-2} y_{-1}' \Gamma_0^{-1} y_{-1} - n^{-2} y_{-1}' \Psi y_{-1} \longrightarrow_p 0,$$

$$n^{-1} y_{-1}' \Gamma_0^{-1} u_0 - n^{-1} y_{-1}' \Psi u_0 \longrightarrow_p 0.$$

**Proof.** See Lemma 3 of Chambers and Hernandez (2015).

■

**Lemma 3.T.** *If $\widehat{\theta}_{2n} - \theta_{20} \longrightarrow_p 0$ and if $\{y_t\}_{t=0}^n$ and $\{u_{0,t}\}_{t=0}^n$ are given by assumptions 2.A and 2.B, then*

$$n^{-2} y_{-1}' \widehat{\Gamma}_n^{-1} y_{-1} \Rightarrow \int_0^1 \mathcal{J}_c^2,$$

$$n^{-1} y_{-1}' \widehat{\Gamma}_n^{-1} u_0 \Rightarrow \int_0^1 \mathcal{J}_c d\mathcal{W}.$$

The following proof is based on the proof of Lemma 2 of Chambers and Hernandez (2015) and is pivotal for the main result.

**Proof.** To prove the first item, note

$$\begin{aligned} n^{-2} y_{-1}' \widehat{\Gamma}_n^{-1} y_{-1} &= n^{-2} y_{-1}' (\widehat{\Gamma}_n^{-1} - \Gamma_0^{-1}) y_{-1} + n^{-2} y_{-1}' (\Gamma_0^{-1} - \Psi) y_{-1} + n^{-2} y_{-1}' \Psi y_{-1} \\ &= n^{-2} y_{-1}' \Psi y_{-1} + o_p(1), \end{aligned}$$

where the first line is direct and the second follows from Lemmas 3.P and 3.S. Now, I claim that $n^{-2} y_{-1}' \Psi y_{-1} \Rightarrow \int_0^1 \mathcal{J}_c^2$. To see this, note that the same steps of the proof of Lemma 3.H yield

$$n^{-2} y_{-1}' \Psi y_{-1} \Rightarrow 2\pi \sum_{j=-\infty}^{\infty} \psi_j \left[ \sigma_u^2 \int_0^1 \mathcal{J}_c^2 \right],$$

but $\sigma_u^2 = 2\pi f(0; \theta_{20})$ from Definition 3.Q and $\sum_{j=-\infty}^{\infty} \psi_j = [4\pi^2 f(0; \theta_{20})]^{-1}$ from Definition 3.R; thus the claim follows.

Addressing the second item, I use again Lemmas 3.P and 3.S so that

$$n^{-1} y_{-1}' \widehat{\Gamma}_n^{-1} u_0 = n^{-1} y_{-1}' \Psi u_0 + o_p(1),$$

and claim $n^{-1} y_{-1}' \Psi u_0 \Rightarrow \int_0^1 \mathcal{J}_c d\mathcal{W}$. I prove this claim in two steps. First, note how the logic of proof of Lemma 3.H implies

$$n^{-1} y_{-1}' \Psi u_0 \Rightarrow 2\pi \sum_{j=-\infty}^{\infty} \psi_j \left[ \sigma_u^2 \int_0^1 \mathcal{J}_c d\mathcal{W} + \frac{\sigma_u^2 - \gamma_0}{2} + \Phi_j \right].$$

The second step requires me to prove that $2\pi \sum_{j=-\infty}^{\infty} \psi_j \left[ \frac{\sigma_u^2 - \gamma_0}{2} + \Phi_j \right] = 0$. Recall the definition of $\Phi_j$ from Lemma 3.H and write

$$2\pi \sum_{j=-\infty}^{\infty} \psi_j \left[ \frac{\sigma_u^2 - \gamma_0}{2} + \Phi_j \right] = \pi(\sigma_u^2 - \gamma_0) \sum_{j=-\infty}^{\infty} \psi_j + 2\pi\gamma_0 \sum_{j=1}^{\infty} \psi_j - 2\pi \sum_{j=1}^{\infty} \psi_j \gamma_j, \tag{3.17}$$

where the latter expression follows from symmetry of $\Gamma_0$. From the symmetry of the Fourier coefficients of $\Psi$ note that

$$\sum_{j=1}^{\infty} \psi_j = \frac{1}{2} \left[ \sum_{j=-\infty}^{\infty} \psi_j - \psi_0 \right] = \frac{1}{2} \left[ 4\pi^2 f(0; \theta_{20}) \right]^{-1} - \frac{1}{2}\psi_0,$$

which together with Definition 3.Q allows me to write the right side of (3.17) as

$$\pi(\sigma_u^2 - \gamma_0) \left[ 4\pi^2 f(0; \theta_{20}) \right]^{-1} + \pi\gamma_0 \left[ \left[ 4\pi^2 f(0; \theta_{20}) \right]^{-1} - \psi_0 \right] - 2\pi \sum_{j=1}^{\infty} \psi_j \gamma_j,$$

or

$$\frac{1}{2} - \pi\psi_0\gamma_0 - 2\pi \sum_{j=1}^{\infty} \psi_j \gamma_j = 0.$$

To see the last equality, from Zygmund (1959) I know that for the Fourier series, $2\pi \sum_{j=-\infty}^{\infty} \psi_j \gamma_j e^{ijx}$, there exist a function

$$h(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[ 2\pi f(x - s; \theta_{20}) \right]^{-1} 2\pi f(s; \theta_{20}) ds,$$

implying $h(0) = 1$ and thus $2\pi \sum_{j=-\infty}^{\infty} \psi_j \gamma_j = 1$, or

$$2\pi \sum_{j=1}^{\infty} \psi_j \gamma_j = \frac{1}{2}(1 - 2\pi\psi_0\gamma_0),$$

where the equality is explained by the symmetry of the Fourier coefficients. ■

**Proof of Proposition 3.O.** Following Chambers and McCrorie (2007), I start by expanding $s_{1n}(\widehat{\theta}_n)$

$$\begin{aligned} s_{n1}(\widehat{\theta}_n) &= n^{-1}(\widehat{\theta}_{1n} - \theta_{10})y'_{-1}(\widehat{\Gamma}_n^{-1} - \Gamma_0^{-1})y_{-1} - n^{-1}y'_{-1}(\widehat{\Gamma}_n^{-1} - \Gamma_0^{-1})u_0 \\ &+ n^{-1}(\widehat{\theta}_{1n} - \theta_{10})y'_{-1}(\Gamma_0^{-1} - \Psi)y_{-1} - n^{-1}y'_{-1}(\Gamma_0^{-1} - \Psi)u_0 \\ &+ n^{-1}(\widehat{\theta}_{1n} - \theta_{10})y'_{-1}\Psi y_{-1} - n^{-1}y'_{-1}\Psi u_0. \end{aligned}$$

The first line of the right side is $o_p(1)$ from Lemma 3.P and $n(\widehat{\theta}_{1n} - \theta_{10}) = O_p(1)$. The second line is also $o_p(1)$ by direct application of Lemma 3.S. This allows me to write $s_{1n}(\widehat{\theta}_n) = 0$ as

$$n(\widehat{\theta}_{1n} - \theta_{10}) = \frac{n^{-1}y'_{-1}\Psi u_0}{n^{-2}y'_{-1}\Psi y_{-1}} + o_p(1),$$

finally, Lemma 3.T together with the Continuous Mapping Theorem yield the result. ■

**Corollary 3.U.** *From $s_{n1}(\widehat{\theta}_n)$, I can also obtain the closed form,*

$$\widehat{\theta}_{1n} - 1 = \frac{y'_{-1}\Psi \Delta y}{y'_{-1}\Psi y_{-1}} + o_p(1),$$

*with corresponding asymptotic behaviour,*

$$n(\widehat{\theta}_{1n} - 1) \Rightarrow \frac{\int_0^1 \mathcal{J}_c d\mathcal{J}_c}{\int_0^1 \mathcal{J}_c^2}.$$

**Proof.** The closed form for $\widehat{\theta}_{1n} - 1$ is obtained by subtracting one on both sides of the expression in Remark 3.N and using the definition of $u_0$ with Lemmas 3.P and 3.S. The asymptotic behaviour is analogous to that in Proposition 3.O. ■

Now I present the following Lemma that allows me to ignore $Q_{1n}(\theta_n)$ in derivations of the asymptotic distribution of $\widehat{\theta}_{2n}$.

**Lemma 3.V.** *$Q_{1n}(\theta_n)$ as defined by (3.3) is not relevant to derive the asymptotic distribution of $\widehat{\theta}_{2n}$. Equivalently,*

$$n^{-1/2}Q_{1n}(\theta_n) \longrightarrow_p 0.$$

**Proof.** From (3.3) note that

$$\begin{aligned} n^{-1/2}Q_{1n}(\theta) &= n^{2\nu}(\theta_{1n}-\theta_{10})^2 n^{-(2\nu+1/2)} y_{-1}'\Gamma_n^{-1}y_{-1} \\ &- 2n^{\nu}(\theta_{1n}-\theta_{10}) n^{-(\nu+1/2)} y_{-1}'\Gamma_n^{-1}u_0. \end{aligned}$$

The first term on the right is $o_p(1)$, $y_{-1}'\Gamma_n^{-1}y_{-1}$ is only bounded in probability for $\nu = 3/4$ while $n^{3/2}(\theta_{1n}-\theta_{10})^2 = o_p(1)$. The second term is also $o_p(1)$, $y_{-1}'\Gamma_n^{-1}u_0$ is bounded in probability only if $\nu = 1/2$, but $n^{1/2}(\theta_{1n}-\theta_{10}) = o_p(1)$ from Proposition 3.I. ■

**Proposition 3.W** (Asymptotic Distribution of $\widehat{\theta}_{2n}$)**.** *Given Lemma 3.V, and (3.16) the asymptotic distribution of $\widehat{\theta}_{2n}$ is that of a ML estimate for a stationary MA(1) process. This is,*

$$n^{1/2}\left(\widehat{\theta}_{2n}-\theta_{20}\right) \longrightarrow_d N\left(0, V^{-1}(\theta_{20})\right),$$

*where the $kl$ element of $V(\theta_{20})$ is given by*

$$V_{kl}(\theta_{20}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f^{-1}(\omega;\theta_{20})\frac{\partial f(\omega;\theta_{20})}{\partial\theta_{20,k}} f^{-1}(\omega;\theta_{20})\frac{\partial f(\omega;\theta_{20})}{\partial\theta_{20,l}},$$

*and $f(\omega;\theta_{20})$ is the spectral density of $\{u_{0,t}\}_{t=0}^{n}$.*

**Proof.** See Hannan (1973). ■

# 4 Unit Root Test

In this section I introduce the likelihood ratio unit root test and its asymptotic properties. With the ML framework presented in the previous section and the asymptotic behaviour of $\widehat{\theta}_n$ at hand, I can now present the $LR$ test statistic and set up the hypothesis testing problem in the next Definition and Proposition.

**Definition 4.A** (LR Test Statistic)**.** *For a given sample size, $n$, and the log-likelihood function defined by (3.1), the Likelihood Ratio Statistic for testing the Null Hypothesis $\theta_{10} = 1$, against the Alternative Hypothesis $\theta_{10} < 1$, is given by*

$$LR_c = 2\left[l^n(\widehat{\theta}_n) - l^n(\widetilde{\theta}_n)\right],$$

*where $\widehat{\theta}_n$and $\widetilde{\theta}_n$ are the unrestricted and restricted ML estimates, respectively.*

**Proposition 4.B** (Asymptotic Distribution of $LR_c$)**.** *If* $n^{\nu}(\widehat{\theta}_{1n} - \theta_{10}) \longrightarrow_p 0$ *for* $\nu \in [0, 1)$, *and* $\widehat{\theta}_{2n} - \theta_{20} \longrightarrow_p 0$, *and the Null Hypothesis to test is that of Definition 4.A, then*

$$LR_c \Rightarrow \frac{\left[\int_0^1 \mathcal{J}_c d\mathcal{J}_c\right]^2}{\int_0^1 \mathcal{J}_c^2}.$$

**Proof of Proposition 4.B .** The proof is based on one fact, that $LR_c$ can be written as a second order mean value expansion of the log-likelihood function $l^n(\widetilde{\theta})$ about $\widehat{\theta}_n$ (see Davidson (2000) p.290)

$$LR_c = -(\widehat{\theta}_n - \widetilde{\theta}_n)' \overline{\mathbb{H}}_n (\widehat{\theta}_n - \widetilde{\theta}_n),$$

where $\overline{\mathbb{H}}_n$ is the Hessian of $l^n(\theta_n)$ evaluated at $\overline{\theta}_n$, a convex combination of $\widehat{\theta}_n$ and $\widetilde{\theta}_n$. A convenient feature of $\overline{\mathbb{H}}_n$ is that it can be divided in blocks according to the rate of convergence of its elements. In particular

$$\overline{\mathbb{H}}_n = -\begin{bmatrix} y'_{-1}\overline{\Gamma}_n^{-1} y_{-1} & \overline{A}_1' \\ \\ \overline{A}_1 & \overline{H} \end{bmatrix}$$

where $\overline{A}_1$ is a $2 \times 1$ vector with elements

$$\overline{A}_{1,[i,1]} = \frac{\partial^2 l^n(\theta_n)}{\partial\theta_{1n}\partial\theta_{2n,i}} = y'_{-1}\frac{\partial\Gamma_n^{-1}}{\partial\theta_{2n,i}} u_0, \; i = 1, 2. \tag{4.1}$$

and $\overline{H}$ is the $2 \times 2$ Hessian of the short-run parameters.

$$\overline{H}_{[i,j]} = -\frac{1}{2}\frac{\partial}{\partial\theta_{2n,i}} tr\left\{\Gamma_n^{-1}\frac{\partial\Gamma_n}{\partial\theta_{2n,j}}\right\} - u'_n\frac{\partial^2\Gamma_n^{-1}}{\partial\theta_{2n,i}\partial\theta_{2n,j}} u_0 \; i, j = 1, 2. \tag{4.2}$$

Thus, I get

$$\begin{aligned} LR_c &= \left[n(\widehat{\theta}_{1n} - 1)\right]^2 \frac{1}{n^2} y'_{-1}\overline{\Gamma}_n^{-1} y_{-1} \\ &- 2n(\widehat{\theta}_{1n} - 1)\left(n^{1/2}(\widehat{\theta}_{2n} - \theta_{20}) - n^{1/2}(\widetilde{\theta}_{2n} - \theta_{20})\right)' \frac{1}{n^{3/2}}\overline{A}_1 \\ &+ n^{1/2}\left((\widehat{\theta}_{2n} - \theta_{20}) - (\widetilde{\theta}_{2n} - \theta_{20})\right)' \frac{1}{n}\overline{H} n^{1/2}\left((\widehat{\theta}_{2n} - \theta_{20}) - (\widetilde{\theta}_{2n} - \theta_{20})\right), \end{aligned}$$

where the second and third terms on the right are $o_p(1)$ since the elements of $\overline{A}_1$ are $O_p(n)$, and $\overline{H}$ has only $O_p(n^{1/2})$ elements. After some algebra and using $n^{-2}y'_{-1}(\overline{\Gamma}_n^{-1} - \Gamma_0^{-1})y_{-1} = o_p(1)$, I have the expression

$$LR_c = \frac{1}{n}\Delta y' \overline{\Gamma}_n^{-1} y_{-1} n(\widehat{\theta}_{1n} - 1) + o_p(1),$$

thus, I only need to invoke Lemma 3.T, Proposition 3.O, and the Continuous Mapping Theorem to obtain the result. ■

**Remark 4.C.** *Stock (1994) lists the characteristics that good unit root tests have: (i) the test is independent of the parameters for the constant, the trend or serial correlation; (ii) the test has good power in large samples; and (iii) the test has both good power and small size distortions when computed over different models and samples.*[3]

**Remark 4.D.** *The asymptotic distribution of $LR_c$ presented in Proposition 4.B coincides with that found in Johansen (1988), Larsson (1998) and Rothenberg and Stock (1997). Moreover it is independent of serial correlation parameters. Thus, it satisfies the first of Stock's requirements for a good test of Remark 4.C (i).*

**Corollary 4.E** (Asymptotic Distribution of $LR_c$ under the Null Hypothesis)**.** *If $c = 0$, then I get the squared of the Dickey-Fuller t-statistic's asymptotic distribution*

$$LR_0 \Rightarrow \frac{\left[\int_0^1 \mathcal{W} d\mathcal{W}\right]^2}{\int_0^1 \mathcal{W}^2}.$$

**Proof.** The proof is straight forward from Remark 3.F. ■

# 5 Empirical Analysis

Once I have the $LR_c$ test, the next step entails the evaluation of its empirical properties. The analytical tool to do so is the Power Envelope, and its derivation is presented in this section. It includes a subsection with preliminary definitions that should help the exposition of the power envelope, both in finite samples and its definition for the large sample case. The empirical properties of the $LR_c$ are then presented along with the algorithm to obtain them. The section concludes with a comparison of the power properties of $LR_c$ with the Augmented Dickey Fuller test.

[3] Stock (1994) pp 2764. The Null and alternative hypothesis he considers are standard.

## 5.1 Preliminaries

A few preliminary definitions and notation are necessary here, in particular, recall I make a *Type I* Error if I incorrectly reject the Null Hypothesis. Additionally, I make a *Type II* Error if I fail to reject the Null Hypothesis when the true parameter is not satisfying it. Thus, I can introduce a straightforward definition of the power the test and Power Function.

**Definition 5.A** (Power Function). *The power of the test is defined as the probability of rejecting the Null Hypothesis, and $\phi(\theta_n)$ is the Power Function satisfying*
*(i) $\phi(\theta_n) = Pr(\text{Type I error})$, or*
*(ii) $\phi(\theta_n) = 1 - Pr(\text{Type II error})$.*

**Remark 5.B.** *Given definitions 4.A and 5.A, a unit root test that has low power will under-reject the Null Hypothesis of a unit root.*

## 5.2 Power Envelope: Finite Samples

I will derive the Power Envelope following the exposition of Jansson and Nielsen (2012). Define the rejection region for the test statistic, $\mathcal{R}$, and the size of the test, $\overline{\alpha}$. Recall the $LR_c$ statistic is given by Definition (4.A) and define the Power Envelope for each $\mathcal{R}$, $\bar{\alpha}$, $n$, and $c$ as

$$\Pi_n^{\bar{\alpha}}(\theta_{1n}(c)) = \max_{Pr(LR_c \in \mathcal{R}):\phi(\theta_{1n}(0))=\bar{\alpha}} \phi\left(\theta_{1n}(c)\right).$$

An application of the Neyman-Pearson Lemma yields,

$$\Pi_n^{\bar{\alpha}}(\theta_{1n}(c)) = P_{\theta_{1n}(c)}\left(LR_c > k_n^{\bar{\alpha}}\right), \tag{5.1}$$

where $P_{\theta_{1n}(c)}(\cdot)$ is the probability measure underlying $\theta_{1n}(c)$, and $k_n^{\bar{\alpha}}$ satisfies

$$P_{\theta_{1n}(0)}\left(LR_c > k_n^{\bar{\alpha}}\right) = \bar{\alpha}. \tag{5.2}$$

This is, $k_n^{\bar{\alpha}}$ is the critical value obtained from the frequency distribution of $LR_0$ as given by Proposition 4.B at the $(1-\bar{\alpha})$ quantile and sample size $n$.

**Remark 5.C.** *(i) $k_n^{\bar{\alpha}}$ does not depend on $c$. (ii) By construction, the Power Envelope as described by (5.1) corresponds to the family of most powerful unit root tests.*[4] *(iii) This family of tests has an Asymptotic Power Envelope.*

[4] Stock (1994) p. 2771.

### 5.3 Asymptotic Power Envelope

An additional tool for the evaluation of the test is the Asymptotic Power Envelope, which provides a benchmark for the asymptotic behaviour of the family of tests and is given by

$$\lim_{n\to\infty} \Pi_n^{\bar{\alpha}}(\theta_{1n}(c)) = \max_{Pr(LR_c \in \mathcal{R}): \lim_{n\to\infty} \phi(\theta_{1n}(0)) = \bar{\alpha}} \lim_{n\to\infty} \phi\left(\theta_{1n}(c)\right).$$

Let $\lim_{n\to\infty} k_n^{\bar{\alpha}} = k^{\bar{\alpha}}$, taking limits to (5.1) and using the Neyman-Pearson Lemma

$$\lim_{n\to\infty} \Pi_n^{\bar{\alpha}}(\theta_{1n}(c)) = \lim_{n\to\infty} P_{\theta_{1n}(c)}\left[LR_c > k^{\bar{\alpha}}\right] = P_1\left[\frac{\left[\int_0^1 \mathcal{J}_c d\mathcal{J}_c\right]^2}{\int_0^1 \mathcal{J}_c^2} > k^{\bar{\alpha}}\right], \tag{5.3}$$

where, $k^{\bar{\alpha}}$ satisfies $P_1\left[\frac{\left[\int_0^1 \mathcal{W} d\mathcal{W}\right]^2}{\int_0^1 \mathcal{W}^2} > k^{\bar{\alpha}}\right] = \bar{\alpha}$.

**Remark 5.D.** *The Asymptotic Power Envelope as defined by equation (5.3) does not depend on the short run parameters $\theta_{2n}$.*

### 5.4 Empirical Properties

I conducted Monte Carlo experiments to obtain the empirical properties of the test statistic in Definition 4.A. Recall the asymptotic expression for $LR_c$ in Proposition 4.B corresponds to that found in previous work, in particular in Johansen (1988). This means there are critical values that can be used as a benchmark, and they are a useful guide. Regarding the discrete approximations to the Wiener processes and the Ornstein -Uhlenbeck processes, I have followed the methods outlined in Johansen (1995) chapter 15 and Saikkonen and Lutkepohl (1999) p. 61, respectively. Finally, selection of the sample size, $n$, and true values for the short run parameters related to the error process correspond to that of the recent literature (c.f. Jansson and Nielsen (2012)).

### Asymptotic Size

To obtain the asymptotic distribution of $LR_c$ under the Null Hypothesis, I simulate the asymptotic expression in Proposition 4.B through discrete approximations to Wiener Processes. Results are contained in Table 1 for $10^6$ replications and the distribution is shown in Figure 1.

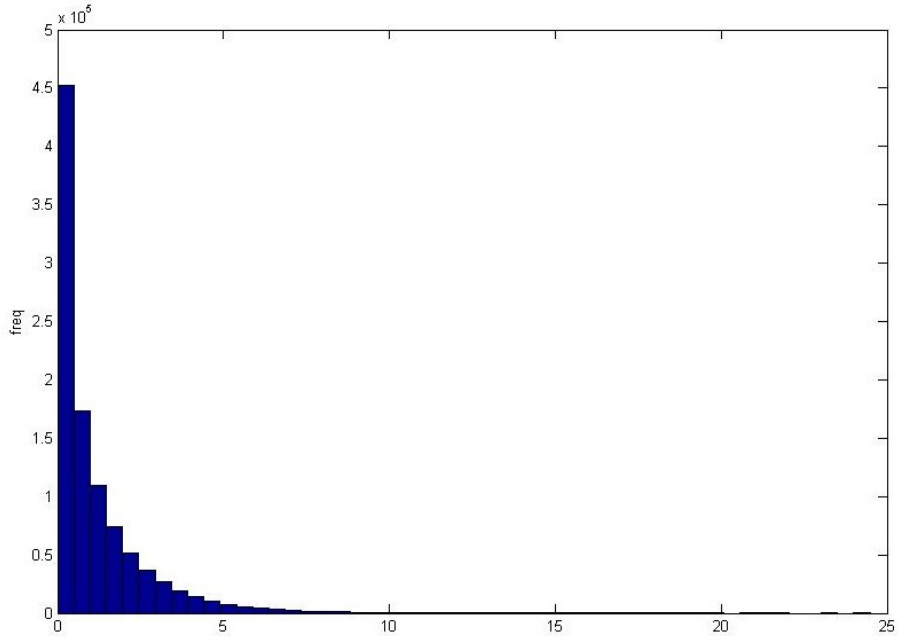


Figure 1: Asymptotic Distribution of $LR_0$.

| Sample/$(1-\bar{\alpha})$ | $85\%$ | $90\%$ | $95\%$ | $97.5\%$ | $99\%$ | $99.5\%$ |
|---|---|---|---|---|---|---|
| $n = 10^4$ | 2.3345 | 2.9857 | 4.1332 | 5.3210 | 6.9288 | 8.1821 |

Note: Wiener Processes were simulated in Matlab® with variance $1/n$ in $10^6$ Monte Carlo replications.

Table 1: Critical Values from Simulations of Wiener Process

| Sample/$(1-\bar{\alpha})$ | $50\%$ | $75\%$ | $80\%$ | $85\%$ | $90\%$ | $95\%$ | $97.5\%$ | $99\%$ |
|---|---|---|---|---|---|---|---|---|
| $n = 400$ | 0.60 | 1.56 | 1.89 | 2.32 | 2.98 | 4.14 | 5.30 | 7.02 |

Table 2: Critical Values from Johansen's Test out of 5,000 Monte Carlo replications.

Given that the asymptotic distribution of $LR_c$ coincides with that of Johansen (1988) rank test as mentioned, I can use it as a benchmark to check its validity. In Table 2 I reproduce the critical values given in Johansen (1995) Table 15.1. The outcome of the comparison is contained in the following Remark for future reference.

**Remark 5.E.** *The similarity between the values for* $(1-\bar{\alpha}) = 95\%$ *in Table 1 and Table 2 guarantees the experiments are designed correctly.*

## Asymptotic Power Envelope

To simulate the Asymptotic Power Envelope (5.3), I could have used Johansen's, but the critical values presented in Table 1 enjoy a computational advantage as they were obtained by a considerably larger number of repetitions, $10^6$ against $5000$. I use the following algorithm to get the simulations.

1. Set initial parameters: $n = 10^4$ for the sample size as an approximation to infinite sample size and choose $k^{\bar{\alpha}}$ for $\bar{\alpha} = 5\%$, this is the 4th column of Table 1 and the true $\sigma_0^2 = 1$. Moreover $maxit$ is the total number of repetitions and $maxj$ as the number of values $c$ can take.

2. Define a count $M$ of dimensions $maxit, maxj$.

3. Define a sensible grid for the local-to-unity parameter $c$. I use $c_j = -j$ with $j \in \{0, 1, 2, \ldots, 30\}$ which is the same as in Jansson and Nielsen (2012).

4. Set $it = 1$ for the first repetition.

5. Obtain an i.i.d sequence of standard normal errors $\{\epsilon\}_{t=0}^{n}$ with the random number generator.

6. Set $\epsilon_1 = 0$, and construct a *discretized* Wiener Process defined as: $W_t = \sum_{s=1}^{t} \epsilon_s$.

7. For $j = 0$, construct a *discretized* Ornstein-Uhlenbeck Process: $J_{c_j,t} = (1+c_j/n)J_{c,t-1}+ \epsilon_t$ where $J_{c_j,0} = 0$ for all $j$.[5]

8. Compute $dJ_{c_j,t} = J_{c_j,t} - J_{c_j,t-1}$.

9. Compute and save the asymptotic test statistic from Proposition 4.B
$LR_{c_j} = \left[n^{-2} \sum_{s=1}^{n} J_{c_j,s}^2\right]^{-1} \left[n^{-1} \sum_{s=1}^{n} J_{c_j,s} dJ_{c_j,s}\right]^2$.

10. Change the value of $j$ to $j + 1$ and go back to Step 7. Repeat the loop until $j = maxj$.

11. Once the loop on $j$ is finished, change $it$ to $it + 1$ and go back to Step 4. Repeat the loop until $it = maxit$. I repeated the experiment $10^6$ times; thus $maxit = 10^6$.

[5]The discrete approximations to the Ornstein-Uhlenbeck Processes are obtained following Saikkonen and Lutkepohl (1999) pp 61.

Finally, compute the Empirical Asymptotic Power

$$\widehat{\Pi}_{\infty}^{\bar{\alpha}}(\theta_{1n}(c_j)) \quad = \quad \frac{1}{10^6}\sum_{s=1}^{10^6}\mathbb{I}\{LR_{c_{j,s}} > k_n^{\bar{\alpha}}\},$$

for each $j$ where $\mathbb{I}\{\cdot\}$ is the indicator function. The outcome of the simulations is presented in Figure 2.

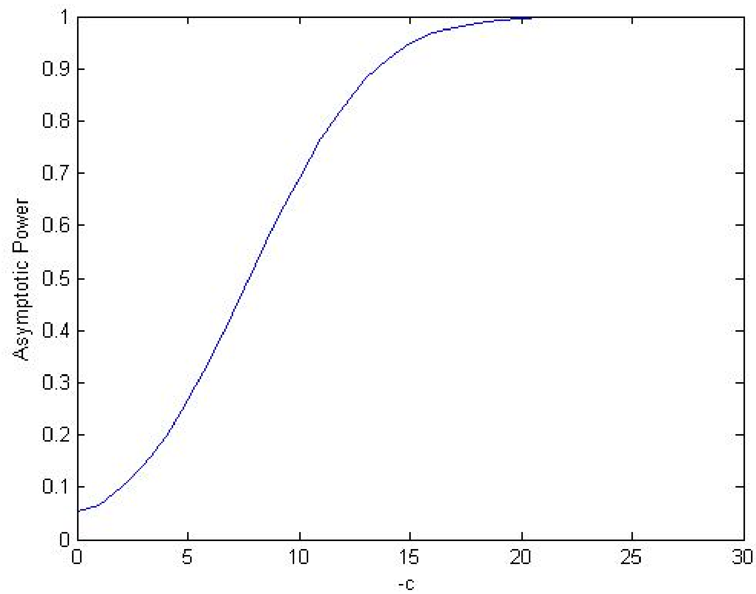


Figure 2: Asymptotic Power Envelope for $LR_c$.

## Power Function

In this subsection I outline the algorithm to construct the Power Function, $\phi(\theta_{1n}(c))$, and present the results from the Monte Carlo experiments. These include results for (small) sample sizes $n \in \{50, 100, 250, 500\}$ and a set of Moving Average ($MA$) parameters $\alpha_0 \in \{\pm 0.8, \pm 0.5, 0\}$ with a grid for $c$ as previously defined.

## Algorithm

To get the Power Function for each value $\alpha_0$ and each $n$, I use the following algorithm.

1. Set initial parameters: $n \in \{50, 100, 250, 500\}$ for the sample size, and the "true" $a \in \{\pm 0.8, \pm 0.5, 0\}$. Set the total number of repetitions $maxit$, and the number of

values $c$ can take, $maxj = 30$.

2. Define the grid for the local-to-unity parameter $c$. I use $c_j = -j$ with $j \in \{0, 2, 4, \ldots, 28, 30\}$.

3. Define a vector of "true" values of $r$ with element $j$: $r_j = 1 - c_j/n$.

4. Set $it = 1$ for the first repetition.

5. Obtain an i.i.d. sequence of standard normal errors $\{\epsilon_t\}_{t=0}^{n}$.

6. Create the "true" DGP for $c_j$ given by $\{y_t\}_{t=0}^{n}$, where $y_t = r_j y_{t-1} + \epsilon_t + a\epsilon_{t-1}$ and $\epsilon_0 = y_0$.

7. Estimate the unrestricted and the restricted ($c = 0$) models with the following sub-algorithm:

   (a) For $j = 1$, write the model as: $\epsilon_t = y_t - r_j y_{t-1} - a\epsilon_{t-1}$, where $r_j$ and $a$ are the parameters to be estimated.

   (b) Since in practice I only observe the sequence $\{y_t\}_{t=0}^{n}$ I need to get a sequence of estimated errors, $\{\widehat{\epsilon}_t\}_{t=0}^{n}$, to estimate $r_j$ and $a$. Note that, for given $y_0 = \widehat{\epsilon}_0 = 0$ the process can be written in iterative form:

$$\begin{aligned}
\widehat{\epsilon}_1 &= y_1 \\
\widehat{\epsilon}_2 &= y_2 - r_j y_1 - a\widehat{\epsilon}_1 \\
&\vdots \\
\widehat{\epsilon}_n &= y_n - r_j y_{n-1} - a\widehat{\epsilon}_{n-1}
\end{aligned}$$

   to use Maximum Likelihood on

$$-\frac{n}{2}\ln|s^2| - \frac{1}{2s^2}\sum_{t=1}^{n}\widehat{\epsilon}_t^2.$$

   (c) Obtain $\widehat{\theta} = (\widehat{r}_j, \widehat{a}, \widehat{s}^2)'$ and $\widetilde{\theta} = (1, \widetilde{a}, \widetilde{s}^2)'$.

8. Compute and save $LR_c$ from Definition 4.A. Compute and save the *Modified Akaike Criteria Augmented Dickey Fuller*, $ADF^*$, test statistic as detailed in Ng and Perron (2001).

9. Change $j = 1$ to $j = j + 1$ and go back to step 6. When the case $j = maxj$ is finish, change $it = 1$ to $it = it + 1$ and go back to step 5. Repeat until $it = maxit$.

This allows me to obtain sequences $\left\{LR_{c_{j,s}}\right\}_{s=1}^{10^4}$ and $\left\{ADF^*_{c_{j,s}}\right\}_{s=1}^{10^4}$. Finally, I compute the (empirical) power for each $c_j$ as:

$$\widehat{\Pi}_n^{\bar{\alpha}}(c_j) = \frac{1}{10^4}\sum_{s=1}^{10^4}\mathbb{I}\{LR_{c_{j,s}} > k_n^{\bar{\alpha}}\}, \tag{5.4}$$

$$\widehat{\Pi}_{ADF^*,n}^{\bar{\alpha}}(c_j) = \frac{1}{10^4}\sum_{s=1}^{10^4}\mathbb{I}\left\{ADF^*_{c_{j,s}} > k_{ADF^*,n}^{\bar{\alpha}}\right\}, \tag{5.5}$$

where $\mathbb{I}\{\cdot\}$ in (5.4) is the indicator function that is equal to 1 if $LR_{c_{j,s}} > k_n^{\bar{\alpha}}$ is true, and 0 otherwise. Likewise, $\mathbb{I}\{\cdot\}$ is 1 if $ADF^*_{c_{j,s}} > k_{ADF^*,n}^{\bar{\alpha}}$ and 0 otherwise in (5.5), $k_{ADF^*,n}^{\bar{\alpha}}$ is the critical value for the $ADF$ at $k^*$ lags obtained with the Modified Akaike Criteria.

## Monte Carlo Results

Figure 3 shows the Asymptotic Power Envelope from Figure 2 and the Power Function for each value of $\alpha_0$ and $n$. Several general conclusions emerge from the inspection of the Monte Carlo experiments. First, for all values of $n$, the simulations show negligible size-bias. This is a desirable property in view of Remark 4.C. Second, for a large sample, $n = 500$, all power functions are considerably close to the asymptotic power envelope. Third, in line with the derived asymptotic behaviour in this paper, as $n$ increases, the Power Functions for all values of $\alpha_0$ converge towards the Asymptotic Power Envelope (c.f. lower-right plot in Figure 3).

Particular conclusions, can also be drawn from the empirical exercise. First, for each value of $n$, if $\alpha_0 > 0$, the test shows power levels close to the asymptotic power envelope. When $n \geq 100$, the power is close to optimal for $\alpha_0 = 0$ as well. I conclude that when $\alpha_0 \geq 0$ I can test for a unit root with $LR_c$ and expect to have close to optimal power properties and negligible size-bias.

For the case of $\alpha_0 < 0$ I need to be more careful. There is a caveat when analysing an $ARMA(1,1)$ model. It is important to keep track of the values of the roots for the lag polynomials, this is, if $\theta_0 = -\alpha_0$, then the stochastic process $\{y_t\}_{t=0}^n$ degenerates into a white noise process. To see why, write the model as

$$(1-\theta_{10}L)y_t = (1+\alpha_0 L)\varepsilon_t.$$
$$\theta_{10} = 1 + \frac{c}{n}, c < 0.$$

With $\theta_{10} = -\alpha_0$, note how the previous expression reduces to $y_t = \varepsilon_t$. In view of the definition of $\theta_{10}$, and given the choice of the values for the parameter $c$, I should expect that the roots *nearly* cancel depending on the value of $n$.

In particular, I should observe the following "slumped" power functions. First, the combination $n = 50$, $c \in \{-9, -10, -11\}$, and $\alpha_0 = -0.8$. Second, the combination $n = 50$, $c \in \{-24, -26, -27\}$, and $\alpha_0 = -0.5$. Third, the case of $n = 100$, $c \in \{-19, -20, -21\}$, and $\alpha_0 = -0.8$. Fourth, if $n = 250$, $c \in \{-29, -30\}$, and $\alpha_0 = -0.8$. Finally, no loss of power is expected for $n = 500$.

Figure 3 corroborate the expectations on the cancelling roots where is clear that as $n$ increases the "slumps" correct gradually. The experiments with $n = 50$ have three "slumps" in $\alpha_0 \in \{-0.8, -0.5, 0\}$, whereas that for $n = 100$ only has two in $\alpha_0 \in \{-0.8, -0.5\}$ and for $n = 250$ only one in $\alpha_0 = -0.8$. Finally, at $n = 500$ there is no loss of power derived by cancelling roots.

I conclude from the analysis that for a large sample, $n = 500$, all desirable properties proposed by Stock (1994) given in Remark 4.C are satisfied. For smaller sample sizes the negative values of the $MA$ parameter may fail to have a good power while retaining negligible size-bias.

## Comparison with the Benchmark

Having established the empirical properties of the power of the $LR_c$, I compare its performance to that of the $ADF^*$ test. This test is a good benchmark since it is still popular among practitioners in cases with no deterministic terms. Moreover, the work from Ng and Perron (2001) has reduced considerably the size distortion caused by the serial correlation.

Table 3 displays the rejection rates for both the $LR_c$ and the $ADF^*$ tests for all values of $\alpha_0$ and a set of values of the local-to-unity parameter, including the Null Hypothesis. In particular, the table contains $\rho_0 \in \{1, 0.99, 0.98, 0.96, 0.94\}$. The size bias for $LR_c$ is smaller in the majority of cases than that of the $ADF^*$, as shown in the column corresponding to $\rho_0 = 1$. The rest of the Table shows that, except for the case of $\alpha_0 = -0.8$, $LR_c$ has higher power, even for small sample sizes.

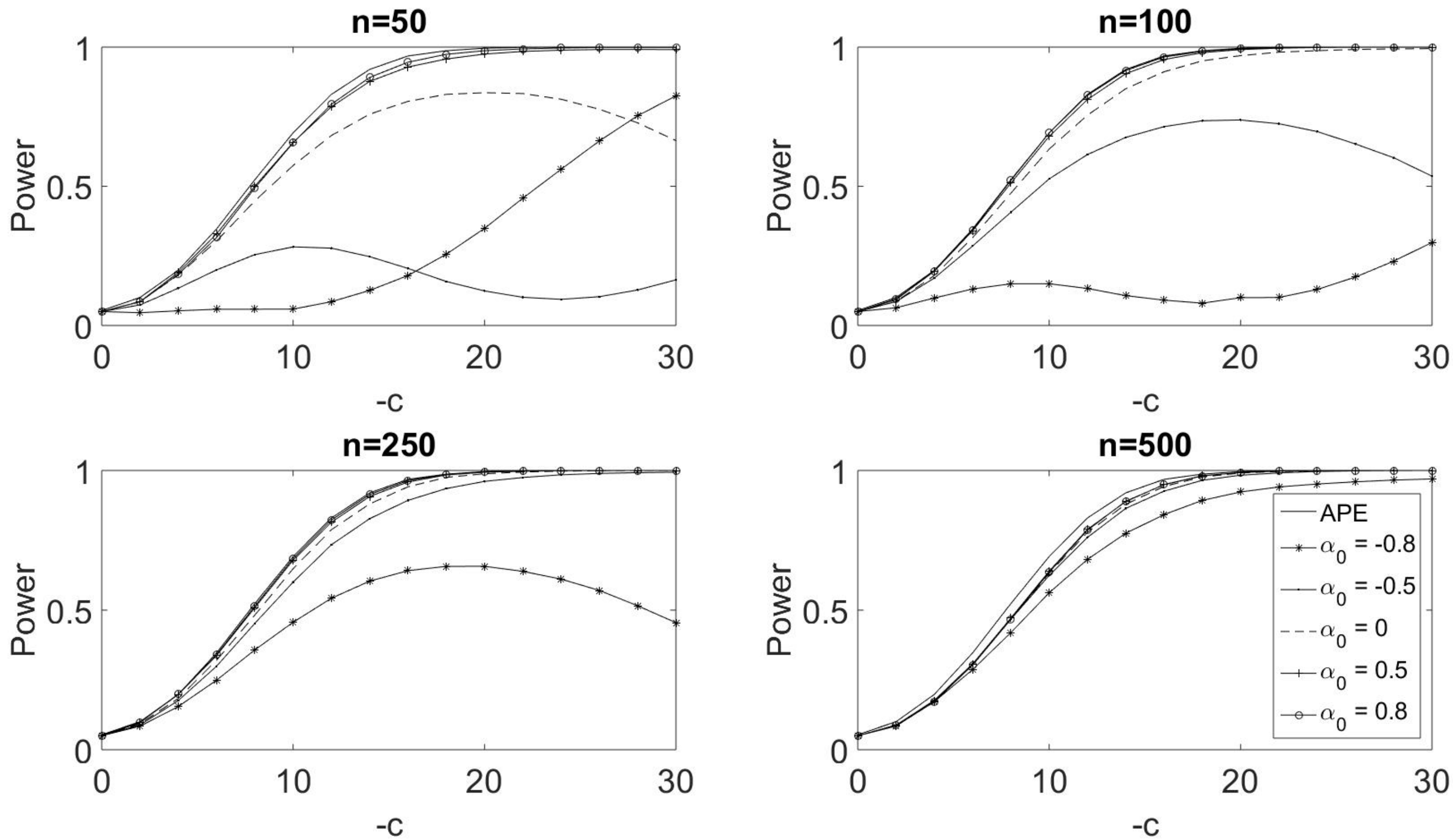


Figure 3: Asymptotic Power Envelope and Power for $LR_c$.

For several combinations of parameters the rejection rate of the $LR_c$ test is more than two times that of the $ADF^*$, this is true for example when $\rho_0 = 0.96$ and $n \leq 100$. Another feature displayed in the Table corresponds to the stark differences in rejection rates between the tests when $\rho_0 = 0.94$ and $n = 250$. The latter is significant as the true local-to-unity parameter is well off from the Null Hypothesis value and the $ADF^*$ test is still failing to reject in more than 25 per cent of the cases. For a large sample size, in this case $n = 500$, the differences in power between the tests for $\alpha_0 = -0.8$ decrease considerably.

Figure 4 shows the Power Function of the $LR_c$ and $ADF^*$ with $k^*$ lags determined by the Modified Akaike Criterion. Each sub-plot corresponds to a value of $\alpha_0$ and each Figure to a sample size. There are features that can be observed across the figure. First, for all combinations of $n$ and $\alpha_0 \geq 0$, $LR_c$ has better power properties than $ADF^*$. Second, as $c$ moves away from zero, the differences in power between the tests grow whenever $\alpha_0 > 0$.

Particular conclusions are, first, gains in power are largest for small sample sizes $n \in \{50, 100\}$ and $\alpha_0 \geq 0$, as shown in the first and second rows of Figure 4. Second, there is no misspecification to worry about when the true process resembles an $ARMA(1, 1)$, even in small samples. This conclusion does not hold, however, for the case of $\alpha_0 = -0.8$. When the true value of $\theta_{1n}(c)$ approximates $-0.8$ the test has lower power than the $ADF^*$. Finally, if $n = 500$, the $LR_c$ test will perform better for each value of $\alpha_0$. Note that even though the $ADF^*$ test converges to the Asymptotic Power Envelope as $n$ increases, the $LR_c$ has higher power in the majority of the cases.

# 6 Empirical Application

For the empirical application I borrow that from Ng and Perron (2001), and test the performance of $LR_c$ against $ADF^*$ on inflation from the GDP Deflator for the G7 countries -U.S., Canada(CAN), Japan (JPN), Great Britain (GBR), Germany (GER), Italy (ITA) and France (FRA). The term structure literature provides a good motivation for testing inflation for a unit root when parameters are estimated with a VAR model. In particular, Ang and Piazzesi (2003) and Kim (2009) use inflation as a factor in their models for the term structure of the interest rate.

I obtained the data from the FRED database. The largest data span starts in the first quarter of 1955 through the fourth quarter of 2014 for the U.S. and GBR, while the shortest, that of Japan, starts in the second quarter of 1994. As Ng and Perron, I compute the inflation rates as 400 times the log-differences of successive quarters. In order to have mean-zero processes, I

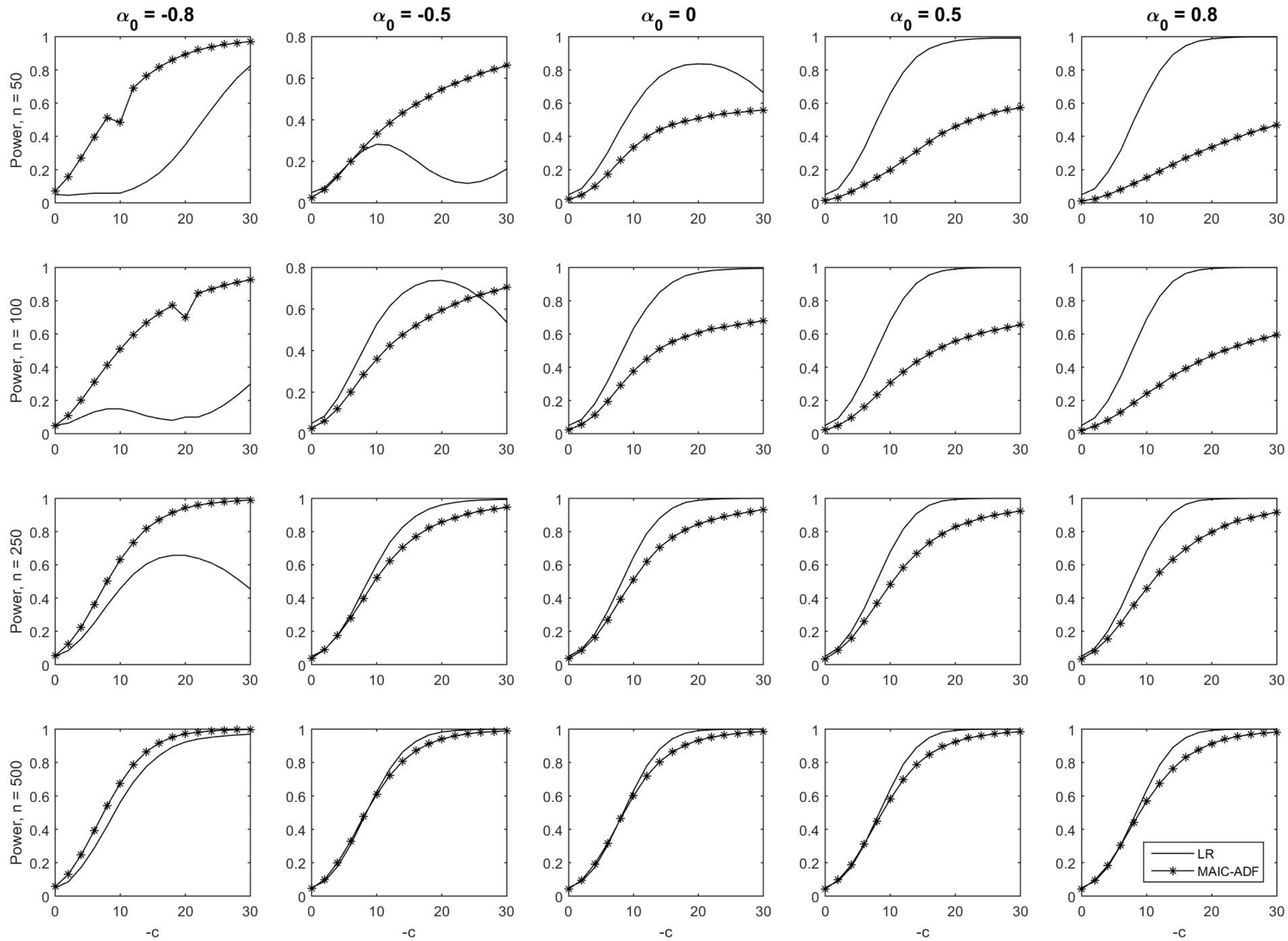


Figure 4: Power Comparison of $LR_c$ with $ADF^*$.

| $n$ | $\alpha_0$ | $\rho_0 = 1$ $LR_c$ | $ADF^*$ | $\rho_0 = 0.99$ $LR_c$ | $ADF^*$ | $\rho_0 = 0.98$ $LR_c$ | $ADF^*$ | $\rho_0 = 0.96$ $LR_c$ | $ADF^*$ | $\rho_0 = 0.94$ $LR_c$ | $ADF^*$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 50 | -0.8 | 0.065 | 0.070 | | | | | 0.061 | 0.156 | | |
| | -0.5 | 0.050 | 0.026 | | | | | 0.073 | 0.064 | | |
| | 0 | 0.050 | 0.021 | | | | | 0.086 | 0.046 | | |
| | 0.5 | 0.051 | 0.015 | | | | | 0.087 | 0.033 | | |
| | 0.8 | 0.055 | 0.012 | | | | | 0.095 | 0.025 | | |
| 100 | -0.8 | 0.056 | 0.050 | | | 0.071 | 0.108 | 0.109 | 0.201 | 0.147 | 0.309 |
| | -0.5 | 0.050 | 0.027 | | | 0.085 | 0.062 | 0.168 | 0.120 | 0.283 | 0.199 |
| | 0 | 0.051 | 0.024 | | | 0.090 | 0.056 | 0.187 | 0.112 | 0.325 | 0.196 |
| | 0.5 | 0.051 | 0.022 | | | 0.094 | 0.049 | 0.197 | 0.097 | 0.344 | 0.160 |
| | 0.8 | 0.051 | 0.020 | | | 0.096 | 0.045 | 0.198 | 0.082 | 0.348 | 0.129 |
| 250 | -0.8 | 0.055 | 0.053 | 0.090 | 0.122 | 0.169 | 0.224 | 0.480 | 0.630 | 0.625 | 0.817 |
| | -0.5 | 0.055 | 0.038 | 0.097 | 0.090 | 0.192 | 0.174 | 0.631 | 0.523 | 0.846 | 0.706 |
| | 0 | 0.052 | 0.035 | 0.101 | 0.085 | 0.199 | 0.161 | 0.678 | 0.512 | 0.896 | 0.705 |
| | 0.5 | 0.052 | 0.034 | 0.101 | 0.084 | 0.206 | 0.157 | 0.694 | 0.481 | 0.913 | 0.666 |
| | 0.8 | 0.052 | 0.034 | 0.101 | 0.083 | 0.207 | 0.156 | 0.697 | 0.457 | 0.920 | 0.633 |
| 500 | -0.8 | 0.055 | 0.056 | 0.185 | 0.247 | 0.588 | 0.675 | 0.932 | 0.972 | 0.974 | 0.998 |
| | -0.5 | 0.057 | 0.044 | 0.194 | 0.200 | 0.660 | 0.609 | 0.987 | 0.941 | 1.000 | 0.990 |
| | 0 | 0.058 | 0.043 | 0.199 | 0.192 | 0.682 | 0.600 | 0.993 | 0.934 | 1.000 | 0.987 |
| | 0.5 | 0.058 | 0.042 | 0.199 | 0.185 | 0.690 | 0.581 | 0.996 | 0.924 | 1.000 | 0.984 |
| | 0.8 | 0.058 | 0.042 | 0.200 | 0.183 | 0.693 | 0.568 | 0.996 | 0.914 | 1.000 | 0.983 |

Table 3: Rejection rates for a set of values for $\rho_0$. The rejection rates for the parameter combinations in blank were not computed as these values of $\rho_0$ were not feasible given the set of values for $n$ and $c$.

subtract the mean of each inflation series.

Following Ng and Perron and Kim (2009), I estimate an $ARMA(1,1)$ model for each country and obtain values of the estimate of the $AR$ ranging from $-0.6767$ (JAP) to $0.9786$ (FRA). The estimates for the $MA$ parameter are negative except for JPN and ITA. The results are summarized in Table 4. Several interesting features emerge from the application. First, the range of values for $\widehat{\alpha}_n$ covers that of the simulation. Second, the cancelling root issue discussed above is present for the cases of JPN and ITA. Third, $\widetilde{\alpha}_n$ is negative in all cases and close to -1 for GER, an issue that has been widely discussed in the literature as in Ng and Perron (2001).

| | U.S. | CAN | JPN | GBR | GER | ITA | FRA |
|---|---|---|---|---|---|---|---|
| Start | 1955Q1 | 1961Q2 | 1994Q2 | 1955Q2 | 1970Q2 | 1991Q2 | 1960Q2 |
| (Obs) | (239) | (215) | (83) | (239) | (179) | (95) | (219) |
| $AR(\widehat{\theta}_{1n})$ | 0.9622 | 0.9213 | -0.6767 | 0.9584 | 0.9702 | -0.8749 | 0.9786 |
| (S.E.) | (0.0188) | (0.0427) | (0.2631) | (0.0235) | (0.0206) | (0.0674) | (0.0210) |
| $MA(\widehat{\alpha}_n)$ | -0.4292 | -0.5474 | 0.6495 | -0.6908 | -0.8217 | 0.7331 | -0.7078 |
| (S.E.) | (0.0428) | (0.1110) | (0.2664) | (0.0377) | (0.0507) | (0.0744) | (0.0483) |
| $MA(\widetilde{\alpha}_n)$ | -0.4540 | -0.6541 | -0.7848 | -0.7310 | -0.8439 | -0.8115 | -0.7269 |
| (S.E.) | (0.0543) | (0.0032) | (0.0653) | (0.0492) | (0.0754) | (0.0516) | (0.0443) |

Note: De-mean Quarterly Inflation: 400 times the log-differences of GDP Deflator about the mean. All series end in 2014Q4.

Table 4: ML estimates for the De-mean GDP Deflator Inflation in G7 countries. $\widehat{\theta}_n$ and $\widetilde{\theta}_n$ are the unrestricted and restricted parameter vector estimates, respectively.

| | U.S. | CAN | JPN | GBR | GER | ITA | FRA |
|---|---|---|---|---|---|---|---|
| $k^{MAIC}$ | 4 | 4 | 5 | 3 | 4 | 4 | 5 |
| Implicit $c$ | 8.9890 | 16.8322 | 137.4927 | 9.8991 | 5.3106 | 176.2438 | 4.6607 |
| $LR_c$ | 4.0937 | 6.6095 | 10.7454 | 4.2118 | 2.7690 | 31.8083 | 1.9978 |
| (p-value) | (0.0550) | (0.0150) | (0.0050) | (0.0350) | (0.1450) | (0.0000) | (0.1850) |
| $ADF^*$ | -2.2467 | -2.3678 | -1.7026 | -2.6946 | -2.8920 | -1.6835 | -1.6428 |
| (p-value) | (0.0242) | (0.0178) | (0.0836) | (0.0075) | (0.0044) | (0.0871) | (0.0944) |

Table 5: De-mean GDP Deflator Inflation in G7 countries unit root tests. $ADF^*$ is the Augmented Dickey-Fuller test with lag structure chosen from MAIC.

Table 5 contains the outcome of the unit root tests. In particular, the first row displays $k^{MAIC}$, the number of autoregressive parameters in addition to the unit root parameter and the variance estimate, that the $MAIC$ suggests (i.e. the $ADF^*$ test for the U.S. requires to estimate 6 parameters). This confirms that for all countries the computation of the $LR_c$ test requires a more parsimonious model. The second row computes the implicit value of the local-to-unity parameter, given the number of observations and the estimate for $\widehat{\theta}_{1n}$, this should provide a basis for comparison of the power of each test, in view of the Monte Carlo experiments above. Both the $LR_c$ and the $ADF^*$ tests reject the null hypothesis for the U.S., CAN and GBR, but the former was found to have higher power (see the third row in Figure 4).

The tests yield different results for the rest of the countries. For JPN and ITA the $LR_c$ rejects the null hypothesis whereas the $ADF^*$ fails to reject. As shown in the second row of Figure 4, the $LR_c$ has a higher power (i.e. a lower probability of making a Type II Error).

Finally, the tests for GER and FRA, which need to deal with $\hat{\alpha}_n$ close to -1, are computed with a sample size that is not large enough for $LR_c$ to have higher power than $ADF^*$, as shown in the third row of Figure 4.

# 7 Concluding Remarks

Testing the nonstationarity of a particular time series with $AR$ based tools has been the norm in parametric unit root testing. But this is not the only way to model dependence. The $LR_c$ I propose here provides an alternative to the $AR$ paradigm, and has more flexibility in dealing with dependence. This is true, in particular, when the error process shows dependence in the form of a $MA$ process since the asymptotic distribution of $LR_c$ is independent of the short run parameters. The $LR_c$ test proves to be just as good in power terms as the $ADF^*$ test for true values of the $MA$ parameter $\alpha_0 \in \{\pm 0.8, \pm 0.5, 0\}$. For a small sample size of $n \leq 100$, the test has close to optimal power only if $\alpha_0 > 0$, but as the sample size increases optimality is gained for the cases $\alpha_0 \in \{-0.8, -0.5, 0\}$.

Comparing the $LR_c$ test with the $ADF^*$ for sample size $n = 250$, Monte Carlo simulations show that $LR_c$ has higher power than $ADF^*$, except for the case of $\alpha_0 = -0.8$. The low power is explained by the caveat detailed in Section 5.4, when $\theta_{10}$ is close to $\alpha_0 = -0.8$ the test can not distinguish between the time series and a pure white noise process. For a sample size of $n = 500$ $LR_c$ has close to optimal power properties for all values of $\alpha_0$. The empirical application shows both the advantages and shortcomings of the $LR_c$ test. The model analysed in this paper can be extended to account for deterministic components, in particular a polynomial trend of order one or zero. I leave this task for a follow-up paper. The model can also be extended to the more general $ARMA(p, q)$, a work which is already in progress in Chambers and Hernandez (2015). The latter in turn will possibly require an information criterion to choose the lag lengths, and will be pursued in future research.


# Acknowledgements

I thank comments from Rod McCrorie and Maria Kyriacou and two anonymous referees on earlier drafts. I owe special thanks to Marcus Chambers for his comments and guidance through the writing of this article which is the first chapter of my PhD thesis. Andrea Miranda provided excellent research assistance. All remaining errors are my own.

## References

ANG, A. AND M. PIAZZESI (2003): "A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables," *Journal of Monetary Economics*, 50, 745 – 787.

BERGSTROM, A. R. (1984): "Continuous time stochastic models and issues of aggregation over time," Elsevier, vol. 2 of *Handbook of Econometrics*, 1145 – 1212.

CHAMBERS, M. J. (2009): "Discrete Time Representations of Cointegrated Continuous Time Models with Mixed Sample Data," *Econometric Theory*, 25, pp. 1030–1049.

CHAMBERS, M. J. AND J. R. HERNANDEZ (2015): "Likelihood-Based Tests for a Unit Root in a Near-Integrated ARMA Model," *Mimeo*.

CHAMBERS, M. J. AND J. R. MCCRORIE (2007): "Frequency domain estimation of temporally aggregated Gaussian cointegrated systems," *Journal of Econometrics*, 136, 1 – 29.

DAVIDSON, J. (1994): *Stochastic Limit Theory: An Introduction for Econometricians*, Oxford University Press.

——— (2000): *Econometric Theory*, Blackwell Publishing.

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, pp. 427–431.

ELLIOTT, G., T. J. ROTHENBERG, AND J. H. STOCK (1996): "Efficient Tests for an Autoregressive Unit Root," *Econometrica*, 64, pp. 813–836.

HALDRUP, N. AND M. JANSSON (2007): "Improving Size and Power in Unit Root Testing," in *Handbook of Econometrics*, ed. by T. C. Mills and K. Patterson, Palgrave Macmillan, vol. 1 of *Handbook of Econometrics*, 865 – 934.

HANNAN, E. J. (1973): "The Asymptotic Theory of Linear Time-Series Models," *Journal of Applied Probability*, 10, pp. 130–145.

JANSSON, M. AND M. O. NIELSEN (2012): "Nearly efficient likelihood ratio tests of the unit root hypothesis," *Econometrica*, 80, 2321–2332.

JOHANSEN, S. (1988): "Statistical analysis of cointegration vectors," *Journal of Economic Dynamics and Control*, 12, 231 – 254.

——— (1995): *Likelihood Based Inference in Cointegrated Vector Autoregresive Models*, Advanced Texts in Econometrics, Oxford University Press.

KIM, D. H. (2009): "Challenges in macro-finance modeling," *Federal Reserve Bank of St. Louis Review*, 519–544.

LARSSON, R. (1998): "Bartlett Corrections for Unit Root Test Statistics," *Journal of Time Series Analysis*, 19, 425–438.

NG, S. AND P. PERRON (2001): "Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power," *Econometrica*, 69, pp. 1519–1554.

PHILLIPS, P. C. B. (1987a): "Time Series Regression with a Unit Root," *Econometrica*, 55, pp. 277–301.

——— (1987b): "Towards a Unified Asymptotic Theory for Autoregression," *Biometrika*, 74, pp. 535–547.

——— (1991): "Optimal Inference in Cointegrated Systems," *Econometrica*, 59, pp. 283–306.

PHILLIPS, P. C. B. AND P. PERRON (1988): "Testing for a Unit Root in Time Series Regression," *Biometrika*, 75, pp. 335–346.

ROSSANA, R. J. AND J. J. SEATER (1995): "Temporal Aggregation and Economic Time Series," *Journal of Business & Economic Statistics*, 13, pp. 441–451.

ROTHENBERG, T. J. AND J. H. STOCK (1997): "Inference in a nearly integrated autoregressive model with nonnormal innovations," *Journal of Econometrics*, 80, 269 – 286.

SAID, S. E. AND D. A. DICKEY (1984): "Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order," *Biometrika*, 71, pp. 599–607.

——— (1985): "Hypothesis Testing in ARIMA(p, 1, q) Models," *Journal of the American Statistical Association*, 80, pp. 369–374.

SAIKKONEN, P. (1995): "Problems with the Asymptotic Theory of Maximum Likelihood Estimation in Integrated and Cointegrated Systems," *Econometric Theory*, 11, pp. 888–911.

——— (2001): "Statistical Inference in Cointegrated Vector Autoregressive Models with Nonlinear Time Trends in Cointegrating Relations," *Econometric Theory*, 17, pp. 327–356.

SAIKKONEN, P. AND H. LUTKEPOHL (1999): "Local Power of Likelihood Ratio Tests for the Cointegrating Rank of a VAR Process," *Econometric Theory*, 15, pp. 50–78.

STOCK, J. H. (1994): "Unit roots, structural breaks and trends," in *Handbook of Econometrics*, ed. by R. F. Engle and D. L. McFadden, Elsevier, vol. 4, 2739 – 2841.

WHITE, H. (2001): *Asymptotic Theory for Econometricians. Revised Edition*, Emerald.

YAO, Q. AND P. J. BROCKWELL (2006): "Gaussian Maximum Likelihood Estimation for ARMA Models. I. Time Series," *Journal of Time Series Analysis*, 857–875.

ZYGMUND, A. (1959): *Trigonometric Series*, no. v. 1 in Cambridge Mathematical Library, Cambridge University Press.