Unit Root Testing in ARMA Models: A Likelihood Ratio Approach

Juan R. Hernández
Banco de México[1]

Seminario Aleatorio ITAM, January 2016

[1] The views expressed in this paper reflect those of the author and not necessarily those of Banco de México.

Table of Contents

0. What is this Paper about?

This paper:

- Proposes a one-step Likelihood Ratio unit root tests (*LR*) which deals with dependence in a time series with an $ARMA(1,1)$ model.

- Derives the asymptotic distribution of *LR* showing that:
  - It is independent of the short-run parameters, and
  - Has good size and power properties.

- Shows that the *LR* has higher power than the ADF^* test for several sample sizes and true values of the *MA* parameter through Monte Carlo experiments.

1. Motivation: Why do we test for Unit Roots

When we find a unit root we are facing a nonstationary process. This changes the empirical and theoretical approach:

- We need special models to deal with it (e.g. Cointegrated VAR).
- Consequences for economics are important (e.g. Hall 1978 conclusion about the consumption obeying the PIH).
- Asymptotics are considerably different.

1. Motivation: Available Unit Root Tests

How does unit root tests at hand deal with dependence?

- Long *AR* regression:
  - *ADF* and all its variations.
  - Elliott, Rothenberg and Stock tests: $DF - GLS$ (a two-step unit root test).
  - A number of the M-Tests from Ng and Perron.
- Fourier Analysis (non-parametric estimation):
  - Phillips and Perron.
  - KPSS.

But, what is dependence exactly?

1. Motivation: Building blocks of time series econometrics

The triple $(\Omega, \mathcal{F}, P)$ is a (complete) probability space, where Ω is the event space, $\mathcal{F}$ is a $\sigma-$*field* and P is a probability measure.

Definition 1 (Stochastic Process)

A (discrete) Stochastic Process $u(\omega)$ is a sequence of (product) measurable functions of the form $u : \Omega \times \mathbb{Z} \longrightarrow \mathbb{R}$ with typical element $u_t(\omega) = u_t$.

Definition 2 (Economic Time Series)

An economic time series, $\{y_t\}_{t=0}^{n}$, is composed by n observations each one generated by a rule -or Data Generating Process- $\mathcal{M}$, regarding the stochastic process u,

$$y_t = \mathcal{M}\left(\{u_l\}_{l=-\infty}^{t}, y_j; \theta\right), \; j < t.$$

1. Motivation: Dependence

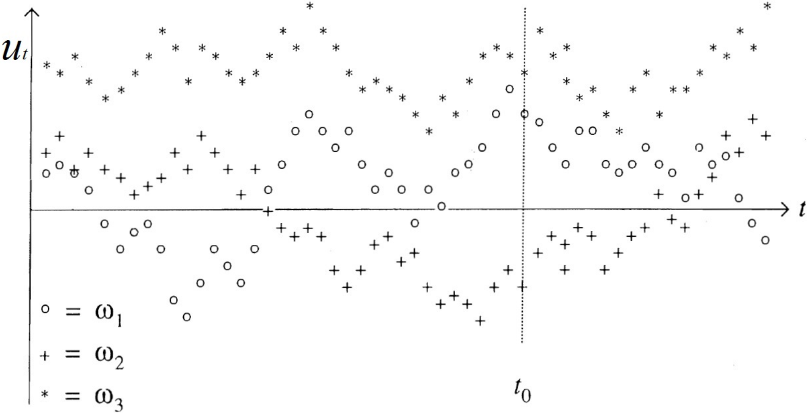


Figure: Dependence illustration. Source: Davidson (1994) Chapter 13.

1. Motivation: Modelling Dependence

Since we do not observe $\mathcal{M}$, we need to approximate it, or model it (i.e. we need to model dependence). Typical (linear) models:

- Long *AR* regression.
- *ARMA*.
- Fourier Analysis.

1. Motivation: Pros and Cons of the Long *AR*

Pros:

- Easy to estimate with OLS.
- Easy to interpret.

Cons:

- Approximates dependence with “long” regressions.
- Low power for one-step estimation unit-root tests.

There are alternatives to model dependence!

1. Motivation: Why ARMA?

Modelling dependence through an *MA* process has several advantages, *vis-á-vis* long *AR*:

- Has a parsimonious representation.
- Dependence is not approximated, but modelled explicitly.
- *MA* is a mild form of *Mixing*, and buys more generality in modelling asymptotics.

1. Motivation: Why ARMA(1,1)?

To see why, we derive the "Exact Discrete Time" representation of a continuous time DGP in the tradition of Bergstrom (1984 Handbook of Econometrics).

- Many time series are *observed* in given points in time, these series are *generated*, however, continuously in time (e.g. pricing decisions in the stock market, exchange rates, consumption, etc.).
- These series are analized with econometric models for *discrete* data. It is easy to argue that even though we *observe* a data point once a day, say, its value contains some form of accumulative information (from the continuous time DGP).

1. Motivation: Why ARMA(1,1)?

- Let $y(t)$ represent the value of such a time series at *instant* t, then its *observed* value is given by:

$$y_t = \int_{t-1}^{t} y(r)dr. \tag{1}$$

- Bergstrom shows that we can model *exactly* y_t dynamics as follows. The continuous random scalar series $y(t)$ satisfies:

$$dy(t) = [a + bt + Ay(t)]dt + \zeta(dt),\ t > 0, \tag{2}$$

where a, b and A are non-random scalars and $\zeta(dt)$ is a random measure with zero mean and zero covariances among two disjoint time intervals.

1. Motivation: Why ARMA(1,1)?

- Since we are interested in the *level* of $y(t)$, we need to solve the SDE in (2). Given $y(0)$, the solution is

$$y(t) = e^{At}y(0) + \int_0^t e^{A(t-s)}[a+bs]ds + \int_0^t e^{A(t-s)}\zeta(ds), t > 0. \quad (3)$$

- From this expression we can obtain the "exact discrete time representation"

$$\Delta y(t) = (e^A - 1)y(t-1) + \int_{t-1}^t e^{A(t-s)}[a+bs]ds + \int_{t-1}^t e^{A(t-s)}\zeta(ds)$$

and after integrating both sides, the $ARMA(1,1)$

$$\Delta y_t = (e^A - 1)y_{t-1} + \psi_t + u_t, t = 1, 2, \ldots n. \quad (5)$$

since

$$u_t = \int_{t-1}^t \int_{r-1}^r e^{A(r-s)}\zeta(ds)dr. \quad (6)$$

2. DGP and Models

Assumption 3 (Data Generating Process)

The observed time series, $\{y_t\}_{t=0}^n$, is generated by:

$$\begin{aligned} \Delta y_t &= (\rho_0 - 1) y_{t-1} + u_{0,t}, \\ \rho_0 - 1 &= \frac{c}{n},\ c < 0. \end{aligned}$$

Assumption 4 (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

$$\begin{aligned} \varepsilon_t &\sim iid(0, \sigma_0^2), \\ \varepsilon_s &= 0 \text{ for } s \leq 0, \\ u_{0,t} &= \varepsilon_t + \alpha_0 \varepsilon_{t-1}, \\ \sup_t E|u_{0,t}|^\delta &< \infty \text{ for } \delta > 2. \end{aligned}$$

Moreover, the covariance structure is summarised by matrix Γ_0.

2. DGP and Models

- The set of models we estimate is given by:

$$\begin{aligned} \Delta y_t &= (\rho_n - 1)y_{t-1} + u_t, \text{ or} && \text{(A)} \\ \Delta y_t &= \mu_n + (\rho_n - 1)y_{t-1} + u_t, \text{ or} && \text{(B)} \\ \Delta y_t &= \mu_n + \tau_n t_n + (\rho_n - 1)y_{t-1} + u_t, && \text{(C)} \end{aligned}$$

  where i_n is a vector of 1's of dimensions $(n \times 1)$ and t_n is a $(n \times 1)$ vector containing $t_j = j$ in its $j - th$ element.
- We collect the *long-run* parameters to be estimated in the vector $\theta_{1n} = (\rho_n, \mu_n, \tau_n)'$.

3. ML Estimation: Assumptions

Assumption 5 (Parameter Space)

(i) *The parameter space* Θ *is convex and compact.*

(ii) $\Theta = \Theta_1 \times \Theta_2$. Θ_1 *contains the long run parameters (i.e.* θ_{1n}*) and* Θ_2 *contains only short-run parameters (i.e.* $\theta_{2n} = (\alpha_n, \sigma_n^2)'$*).*

(iii) Θ_2 *contains only elements that ensure* Γ_n^{-1} *exists.*

Remark 6

Letting $\gamma(j) = Cov(u_{0,t}, u_{0,t-j})$*, Assumption 4 implies a finite long-run variance:* $\sigma_u^2 = \sum_{k=-\infty}^{\infty} \gamma_k < \infty$.

Remark 7 (Gaussianity)

Assumption on the error process $\{u_{0,t}\}_{t=0}^{n}$ *allows us to use the Gaussian likelihood for estimation.*

3. ML Estimation: Usual set-up

The vector of estimates $\widehat{\theta}_n$ can be obtained from maximizing the log-likelihood for $\{u_{0,t}\}_{t=0}^{n}$:

$$l^n(\theta_n) = -\frac{n}{2}\ln|2\pi| - \frac{1}{2}\ln\det|\Gamma_n| - \frac{1}{2}u_n'\Gamma_n^{-1}u_n. \quad (8)$$

Define a new loss function $Q_n(\theta_n) = -2l^n(\theta_n) - n\ln|2\pi|$ so that the objective is now to find the minimizer in

$$Q_n(\theta_n) = \ln\det|\Gamma_n| + u_n'\Gamma_n^{-1}u_n,$$

where, in vector notation, and conditional on the estimated model we have

$$\begin{aligned} u_n &= \Delta y - (\rho_n - 1)y_{-1}, \text{ or} \\ u_n &= \Delta y - (\rho_n - 1)y_{-1} - \mu_n i_n, \text{ or} \\ u_n &= \Delta y - (\rho_n - 1)y_{-1} - \mu_n i_n - \tau_n t_n. \end{aligned}$$

3. ML Estimation: A problem

As explained by Saikkonen (2005 Econometric Theory), it is not possible to obtain consistency of $\widehat{\theta}_{1n}$ with the usual steps. Why?

- Consistency proofs in the “textbook way” assume all estimates converge at the same rate.
- This is not the case here: Long-run estimate $\widehat{\theta}_{1n}$ typically converges faster (superconsistency).
- Short-run estimates converge at rate $n^{1/2}$ (Hannan (1973 J. of Applied Probability)).

Solution proposed by Saikkonen: obtain consistency by splitting the ML estimation in two, a short-run and a long-run problem.

3. ML Estimation: Decompose the optimization in a long/short-run problem

Decomposing the ML problem in a short-run part and a long-run part also allows us to:

- Use existent results to deal with the short-run dynamics causing the dependence.
- Obtain tractable closed forms out of the first order conditions.

Following Saikkonen, add and subtract the elements of θ_{10} from u_n to get:

$$\begin{aligned} u_n &= u_0 - (\rho_n - \rho_0)y_{-1}, \text{ or} \\ u_n &= u_0 - (\rho_n - \rho_0)y_{-1} - (\mu_n - \mu_0)i_n, \text{ or} \\ u_n &= u_0 - (\rho_n - \rho_0)y_{-1} - (\mu_n - \mu_0)i_n - (\tau_n - \tau_0)t_n. \end{aligned}$$

How does the new optimization problem looks?

3. ML Estimation: Decompose the optimization in a long/short-run problem

The new optimization problem, conditionally on the estimated model, is for example:

$$
\begin{aligned}
Q_n(\theta_n) &= (\rho_n - \rho_0)^2 y_{-1}' \Gamma_n^{-1} y_{-1} - 2(\rho_n - \rho_0) y_{-1}' \Gamma_n^{-1} u_0 \\
&+ \ln \det |\Gamma_n| + u_0' \Gamma_n^{-1} u_0, \\
&= Q_{1n}(\theta_n) + Q_{2n}(\theta_{2n}). \qquad (9)
\end{aligned}
$$

Remark 8

The optimization problems for models (B) and (C) can be split in the same fashion, isolating $Q_{2n}(\theta_{2n})$.

Remark 9

(a) $Q_{2n}(\theta_{2n})$ *is the ML problem of a* $MA(1)$ *process. It is thus stationary and we know its asymptotic properties. (b)* $Q_{1n}(\theta_0) = 0$.

3.1 ML Estimation: Consistency

Theorem 10 (Consistency of $\widehat{\theta}_n$)

Let $\underline{\nu} = diag(n^{\nu_1}, n^{\nu_2}, n^{\nu_3})$. If $\{y_t\}_{t=0}^n$, $\{u_{0,t}\}_{t=0}^n$ and Θ are given by assumptions 3, 4 and 5, respectively, then the ML estimate $\widehat{\theta}_n$ is consistent. In particular $\underline{\nu}(\widehat{\theta}_{1n} - \theta_{10}) \longrightarrow_p 0$ and $\widehat{\theta}_{2n} - \theta_{20} \longrightarrow_p 0$.

Proof (Intuition).

For any θ_n write

$$Q_n(\theta_n) - Q_n(\theta_0) = Q_{1n}(\theta_n) + Q_{2n}(\theta_{2n}) - Q_{2n}(\theta_{20}).$$

We want to prove the following: If $d(\theta_n, \theta_0) > 0$ (i.e are not close) then $\inf_{\theta_n : d(\theta_n, \theta_0) > 0} Q_n(\theta_n) - Q_n(\theta_0) > 0$. We do this for Q_{1n} and Q_{2n}. ■

3.2 ML Estimation: Asymptotic Distribution (Preliminaries)

The *normalised* Score vector $s_n(\theta_n)$, corresponding to the optimization problem (9) is as follows. For every estimated model we will have:

$$s_{2n,j}(\theta_n) = n^{-1/2}\frac{\partial Q_{1n}(\theta_n)}{\partial \theta_{2n,j}} + n^{-1/2}\frac{\partial Q_{2n}(\theta_n)}{\partial \theta_{2n,j}},\; j = 1, 2. \tag{10}$$

- If the estimated model is (A):

$$s_{1n}(\theta_n) \quad = \quad 2n^{-1}\left[(\rho_n - \rho_0)y'_{-1}\Gamma_n^{-1}y_{-1} - y'_{-1}\Gamma_n^{-1}u_0\right]. \tag{11}$$

- If the estimated model is (B):

$$s_{1n,\rho}(\theta_n) = 2n^{-1}\left[(\rho_n - \rho_0)y'_{-1}\Gamma_n^{-1}y_{-1} + (\mu_n - \mu_0)i'_n\Gamma_n^{-1}y_{-1} - u'_0\Gamma_n^{-1}y_{-1}\right]. \tag{12}$$

- If the estimated model is (C):

$$s_{1n,\rho}(\theta_n) = 2n^{-1}\left[(\rho_n - \rho_0)y'_{-1}\Gamma_n^{-1}y_{-1} + (\mu_n - \mu_0)i'_n\Gamma_n^{-1}y_{-1} + (\tau_n - \tau_0)t'_n\Gamma_n^{-1}y_{-1} - u'_0\Gamma_n^{-1}y_{-1}\right]. \tag{13}$$

3.2 ML Estimation: Asymptotic Distribution (Preliminaries)

Lemma 11

If $Q_n(\theta_n)$ is given by expression (9), then for each estimated model $Q_{1n}(\theta_n)$ is not relevant to derive the asymptotic distribution of $\widehat{\theta}_{2n}$. Equivalently, if $Q_{1n}(\theta_n)$ is given as in (9), then

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

Definition 12 (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 and $\mathcal{W}(\cdot)$ is a standard Brownian Motion, satisfying the SDE $d\mathcal{J}_c(r) = c\mathcal{J}_c(r)dr + d\mathcal{W}(r)$. To ease notation write $\mathcal{J}_c = \mathcal{J}_c(r)$, $\int_0^1 \mathcal{J}_c = \int_0^1 \mathcal{J}_c(r)dr$, $\mathcal{W} = \mathcal{W}(r)$ and $\int_0^1 \mathcal{W} = \int_0^1 \mathcal{W}(r)dr$.

3.2 ML Estimation: Asymptotic Distribution

Theorem 13 (Asymptotic Distribution of $\widehat{\theta}_{1n}$)

Let $\{y_t\}_{t=0}^n$, $\{u_{0,t}\}_{t=0}^n$ and Θ be given by assumptions 3, 4 and 5, respectively,

(i) *If the estimated model is (A), and $s_{1n}(\theta_n)$ is given by (11), then*

$$n(\widehat{\rho}_n - \rho_0) \Rightarrow \frac{\int_0^1 \mathcal{J}_c d\mathcal{W}}{\int_0^1 \mathcal{J}_c^2}.$$

(ii) *If the estimated model is (B), and $s_{1n}(\theta_n)$ is given by (12), then*

$$n(\widehat{\rho}_n - \rho_0) \quad \Rightarrow \quad \Delta_\mu^{-1} \left[\int_0^1 \mathcal{J}_c d\mathcal{W} - \mathcal{W}(1) \int_0^1 \mathcal{J}_c \right].$$

(iii) *If the estimated model is (C), and $s_{1n}(\theta_n)$ is given by (13), then*

$$n(\widehat{\rho}_n - \rho_0) \Rightarrow \Delta_{\mu\tau}^{-1} \left[\int_0^1 \mathcal{J}_c d\mathcal{W} + \mathcal{W}(1) \left(6 \int_0^1 r\mathcal{J}_c - 4 \int_0^1 \mathcal{J}_c \right) + \int_0^1 rd\mathcal{W} \left(6 \int_0^1 \mathcal{J}_c - 12 \int_0^1 r\mathcal{J}_c \right) \right],$$

where
$\Delta_\mu = \int_0^1 \mathcal{J}_c^2 - \left(\int_0^1 \mathcal{J}_c \right)^2$ *and* $\Delta_{\mu\tau} = \int_0^1 \mathcal{J}_c^2 - 12 \left(\int_0^1 r\mathcal{J}_c \right)^2 - 4 \left(\int_0^1 \mathcal{J}_c \right)^2 + 12 \int_0^1 r\mathcal{J}_c \int_0^1 \mathcal{J}_c$.

3.2 ML Estimation: Asymptotic Distribution

Theorem 14 (Asymptotic Distribution of $\widehat{\theta}_{2n}$ (Hannan 1973))

Let $\{y_t\}_{t=0}^n$, $\{u_{0,t}\}_{t=0}^n$ and Θ be given by assumptions 3, 4 and 5, respectively. Given Lemma 11, and first order condition (10), 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 $\{u_{0,t}\}_{t=0}^n$.

4. *LR* Class of Unit Root Tests

Definition 15 (*LR* Test Statistic)

Let $\widehat{\theta}_n$ and $\widetilde{\theta}_n$ be the unrestricted and restricted ML estimates, respectively. For a given sample size, n, and the log-likelihood function defined by (8), the Likelihood Ratio Statistic for testing the Null Hypothesis $\rho_0 = 1$, against the Alternative Hypothesis $\rho_0 < 1$, is given by

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

In particular:

(i) If the estimated model is (A), $\widetilde{\theta}_{1n} = 1$ and $LR_c = LR$.

(ii) If the estimated model is (B), $\widetilde{\theta}_{1n} = (1, 0)'$ and $LR_c^{\mu} = LR$.

(iii) If the estimated model is (C), $\widetilde{\theta}_{1n} = (1, 0, 0)'$ and $LR_c^{\mu\tau} = LR$.

4.1 *LR* Class of Unit Root Tests: Asymptotic Distribution

Theorem 16 (Asymptotic Distribution of *LR*)

For the test given in Definition 15, let $\underline{\nu}(\widehat{\theta}_{1n} - \theta_{10}) \longrightarrow_p 0$ *and* $\widehat{\theta}_{2n} - \theta_{20} \longrightarrow_p 0$*:*

(i) *If the estimated model is (A), then*

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

(ii) *If the estimated model is (B), then*

$$LR_c^{\mu} \Rightarrow \Delta_{\mu}^{-1} \left[\int_0^1 \mathcal{J}_c d\mathcal{J}_c \left(\int_0^1 \mathcal{J}_c d\mathcal{J}_c - 2\mathcal{W}(1)\int_0^1 \mathcal{J}_c\right) + \mathcal{W}^2(1)\int_0^1 \mathcal{J}_c^2\right] .$$

(iii) *If the estimated model is (C), then*

$$\begin{aligned} LR_c^{\mu\tau} \Rightarrow \Delta_{\mu\tau}^{-1} \Bigg\{ &\left(\int_0^1 \mathcal{J}_c d\mathcal{J}_c\right)^2 + \mathcal{W}^2(1)\left(4\int_0^1 \mathcal{J}_c^2 - 12\left(\int_0^1 r\mathcal{J}_c\right)^2\right) \\ &+ 12\left(\int_0^1 rd\mathcal{W}\right)^2 \left(\int_0^1 \mathcal{J}_c^2 - \left(\int_0^1 \mathcal{J}_c\right)^2\right) \\ &+ 12\int_0^1 \mathcal{J}_c d\mathcal{J}_c \left[\mathcal{W}(1)\left(\int_0^1 r\mathcal{J}_c - \frac{2}{3}\int_0^1 \mathcal{J}_c\right) + \int_0^1 rd\mathcal{W}\left(\int_0^1 \mathcal{J}_c - 2\int_0^1 r\mathcal{J}_c\right)\right] \\ &+ \mathcal{W}(1)\int_0^1 rd\mathcal{W}\left(24\int_0^1 \mathcal{J}_c\int_0^1 r\mathcal{J}_c - 12\int_0^1 \mathcal{J}_c^2\right)\Bigg\} . \end{aligned}$$

4. *LR* Class of Unit Root Tests: Asymptotic Distribution

Remark 17

The asymptotic distribution of LR_c shown in Theorem 16 coincides with that found in:

- *Johansen (1988 J. of Econ. Dynamics and Control)*
- *Rothenberg and Stock (1997 J. of Econometrics)*

Remark 18

Stock (1994 Handbook of Econometrics) lists the characteristics that good unit root tests have:

(i) *Free from parameters for the constant, the trend or serial correlation;*

(ii) *Good power in large samples; and*

(iii) *Both good power and small size distortions when computed over different models and samples.*

5. Empirical Properties: Asymptotic Size

Test/$(1-\bar{\alpha})$	Sample n	85%	90%	95%	97.5%	99%
LR_0	10^4	2.31	2.98	4.11	5.43	7.17
LR_0^{μ}	10^4	6.61	7.54	9.22	10.73	12.93
$LR_0^{\mu\tau}$	10^4	10.82	11.97	13.94	15.96	18.46

Table: Critical Values from Simulations of *LR*.

Test/$(1-\bar{\alpha})$	Sample n	85%	90%	95%	97.5%	99%
No mean	400	2.32	2.98	4.14	5.30	7.02
Mean	400	6.54	7.50	9.13	10.73	12.73
Trend	400	9.43	10.56	12.39	14.13	16.39

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

5. Empirical Properties: Asymptotic Power Envelope

With these critical values, we simulate the Asymptotic expressions derived above for LR. We run 10^4 repetitions and get the rate of rejection for values of $c \in \{0, 2, ...30\}$. This yields the Asymptotic Power Envelope (APE):

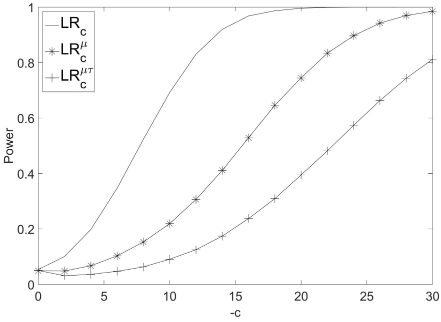


Figure: Asymptotic Power Envelope.

5.1 Empirical Properties: Cancelling Roots

Recall the $ARMA(1,1)$ DGP we are using here,

$$\begin{aligned} y_t &= \rho_0 y_{t-1} + u_{0,t} \\ u_{0,t} &= \varepsilon_t + \alpha_0 \varepsilon_t. \end{aligned}$$

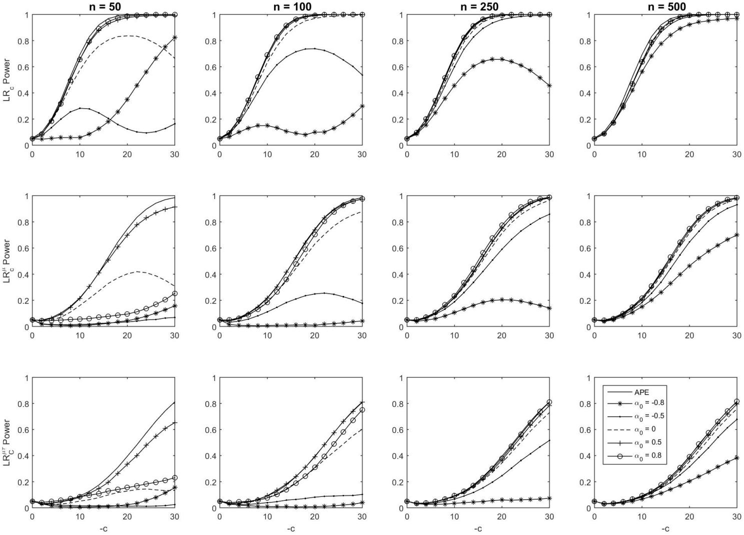
When we write the problem in polynomial-lag form, we get

$$(1 - \rho_0 L) y_t = (1 + \alpha_0 L) \varepsilon_t.$$

But note that if $-\rho_0 = \alpha_0$, $y_t = \varepsilon_t$ (i.e. y_t is white noise).

Testing for a unit root in this case becomes more cumbersome. So we need to be careful!

5.1 Empirical Properties: Power Comparison *LR*



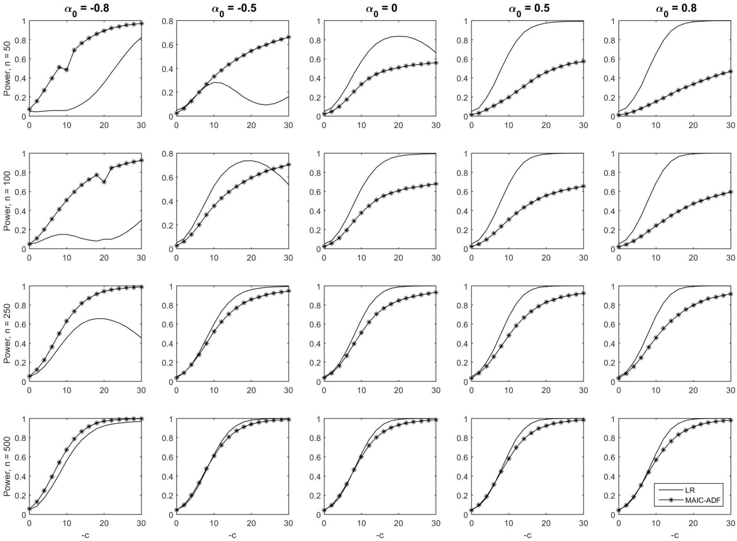
5.2 Empirical Properties: Comparing *LR* with *ADF**

- Having established the empirical properties of the power of the *LR* test, we compare its performance with the Augmented Dickey Fuller test, ADF^*.
- This test is a good benchmark since it is still popular among practitioners.
- The $DF - GLS$ and the point optimal tests from ERS are not good benchmarks as they deal with the deterministic parameters before carrying out the unit root testing (i.e. the tests are computed in two steps whereas the test proposed in this paper is computed in one-step).
- ADF^* is a *t-ratio* test on the whether $(\widehat{\rho}_n^* - 1) = 0$ in the following model

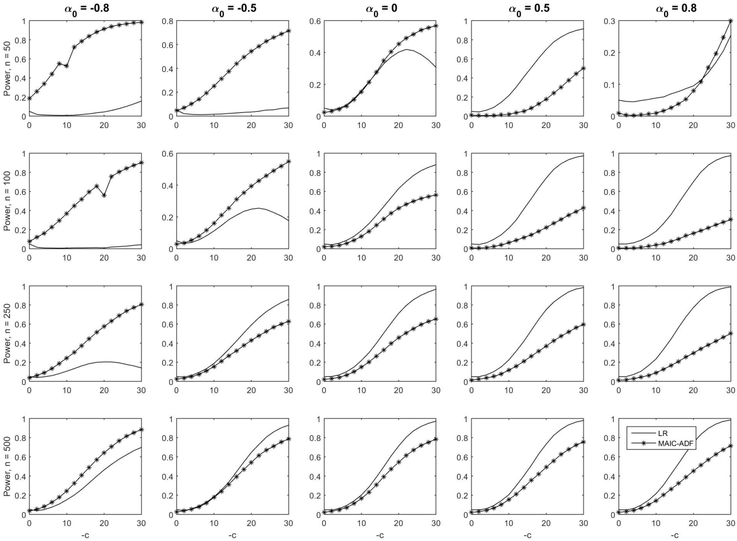
$$\Delta y_t = (\widehat{\rho}_n^* - 1) y_{t-1} + \sum_{j=1}^{k} \widehat{\zeta}_j \Delta y_{t-j} + \widehat{v}_t,$$

where the lag-length, $k > 0$, is chosen using the Modified Akaike Information Criteria proposed by Ng and Perron (2001 Econometrica).

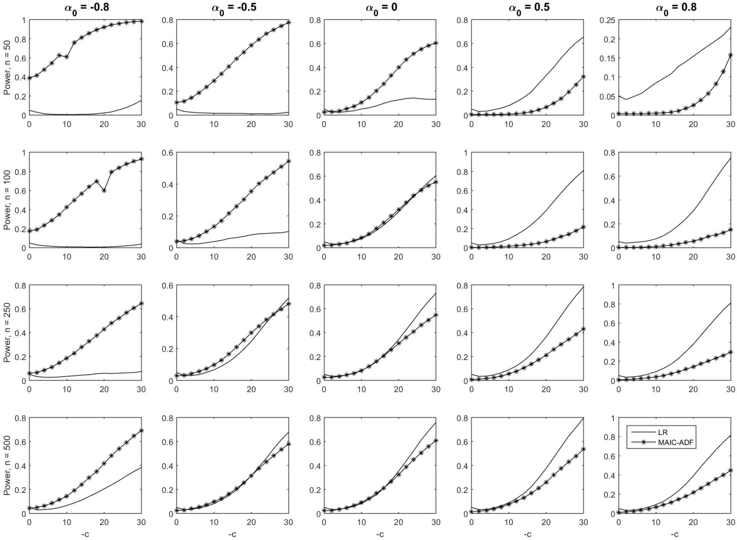
5.2 Empirical Properties: Comparing LR_c with ADF^*



5.2 Empirical Properties: Comparing LR_c^μ with ADF^*



5.2 Empirical Properties: Comparing $LR_c^{\mu\tau}$ with ADF^*



6. Concluding Remarks

- This paper contributes to the unit root testing literature with a fully parametric unit root test within the *LR* class of tests modelling dependence explicitly with a $MA(1)$ process.
- It also introduces a number of limiting results that involve modelling dependence explicitly. Results lay down the path for future work within the *LR* class of tests with more general $ARMA(p,q)$ models.
  - In Chambers and Hernandez (2015) we propose to estimate the parameters in two steps *á la* ERS in an $ARMA(p,q)$ model.
- Empirical analysis showed good power properties when compared to the Asymptotic Power Envelope, particularly for large enough samples.
- Compared with the ADF^*, the *LR* test displayed higher power consistently, as long as there is no root-cancelling and the *MA* parameter is larger than -0.5.