HAL
open science

Testing for unit root processes in random coefficient autoregressive models

Walter Distaso

To cite this version:

Walter Distaso. Testing for unit root processes in random coefficient autoregressive models. Econometrics, 2007, 142 (1), pp.581. 10.1016/j.jeconom.2007.09.002 . hal-00501807

HAL Id: hal-00501807

https://hal.science/hal-00501807v1

Submitted on 12 Jul 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Author's Accepted Manuscript

Testing for unit root processes in random coefficient autoregressive models

Walter Distaso

PII: S0304-4076(07)00176-5
DOI: doi:10.1016/j.jeconom.2007.09.002
Reference: ECONOM 2986

www.elsevier.com/locate/jeconom

To appear in: *Journal of Econometrics*

Cite this article as: Walter Distaso, Testing for unit root processes in random coefficient autoregressive models, *Journal of Econometrics* (2007), doi:10.1016/j.jeconom.2007.09.002

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Testing for unit root processes in random coefficient autoregressive models

Walter Distaso*

Imperial College London

July 2007

Abstract

This paper proposes new tests for simple unit root and unit root with a possibly nonzero drift processes, in the context of a random coefficient autoregressive model. The asymptotic distributions of the tests are derived, and their properties are investigated through a Monte-Carlo experiment. The tests have good power properties, and in many cases they perform better than the competing univariate tests available in the literature, despite testing for a multiple joint hypothesis. In particular, for moderate to large sample sizes, very small values of the variance of the random coefficient variable are needed in order for the tests to reach some power against roots very close to unity. Finally, the proposed tests are applied to the US GDP series.

Keywords: random coefficient autoregressive models, unit root testing, one-sided tests.

JEL classification: C22.

Accepted manuscript

*Tanaka Business School, Imperial College London, London SW7 2AZ, UK, tel:+44 20 7594 3293, fax: +44 20 7594 9189, email: w.distaso@imperial.ac.uk.

ACCEPTED MANUSCRIPT

1 Introduction

A great deal of the econometric literature of the last twenty years has focused on the issue of testing for the unit root hypothesis in economic time series. This has been typically done by using autoregressive models with fixed coefficients and then testing for the autoregressive parameter being equal to one (see, e.g., Dickey and Fuller, 1979, 1981; Evans and Savin, 1981; Phillips, 1987). More recently, some attention has been dedicated to random coefficient autoregressive models. This way of handling the data allows for large shocks in the dynamic structure of the model, and also for some flexibility in the features of the volatility of the series, which are not available in fixed coefficient autoregressive models.

Extensive analysis of the stationary random coefficient autoregressive models has been provided by Nicholls and Quinn (1982); McCabe and Tremayne (1995) and, subsequently, Leybourne et al. (1996) have proposed a test for difference stationarity, given that the series has a randomized unit root. In their case the random autoregressive coefficient has a mean equal to one and the test is based on deciding whether its variance is equal to zero. They justify the randomized unit root approach by arguing that macroeconomic time series result from aggregating different micro-level regimes and that the aggregation process often induces some heterogeneity in the macro series. Their approach then allows to model the described heterogeneity phenomenon. Abadir (2004) has analytically shown that the time-varying aggregation of autoregressive processes gives rise to a random coefficient autoregressive model.

Granger and Swanson (1997) have provided a general analysis of stochastic unit root processes, focusing also on estimation issues. They rely on the testing procedure proposed by Leybourne et al. (1996), showing that the test has power against their class of stochastic unit root processes. They argue that stochastic unit root processes are consistent with economic theory and cite, as a leading example, the theory of the permanent income hypothesis (see, e.g., Hall, 1978).

In the paper, the series $\{y_t\}$ is generated by the mechanism

$$y_t = \rho_t y_{t-1} + \varepsilon_t, \tag{1}$$

where ρ_t is i.i.d. $\left(\rho, \omega^2\right)$, ε_t is $\mathrm{IN}\left(0, \sigma_\varepsilon^2\right)$, $\kappa_\varepsilon^2 \equiv \mathrm{E}\left(\left(\varepsilon_t^2 - \sigma_\varepsilon^2\right)^2\right) = 2\sigma_\varepsilon^4$ and $\{\varepsilon_t\}$, $\{\rho_t\}$ are stochastically independent.

The aim of this paper is to extend the testing procedure proposed by Leybourne et al. (1996), by allowing the mean of the random autoregressive coefficient ρ_t to be unknown (instead of having it fixed to one). Then, tests for the hypotheses that: (i) the variance of ρ_t is zero; (ii) the mean of ρ_t is equal to one; (iii) the mean of ρ_t is equal to one and its variance equal to zero are derived. Under the first hypothesis, the series follows an autoregressive model with a possible unit root; under the second hypothesis, the series follows a possibly stochastic unit root process; finally, when the joint hypothesis is true, the series has an exact unit root. Under the alternative hypothesis the series can be generated by either a stochastic unit

Accepted manuscript

ACCEPTED MANUSCRIPT

root process or a random coefficient process (where the mean of the random autoregressive coefficient is less than one in absolute value).

The joint hypothesis probably represents the most interesting case. In fact, when testing for a unit root in random coefficient autoregressive models, ω^2 represents a typical nuisance parameter. McCabe and Tremayne (1995) and Leybourne et al. (1996) have circumvented this problem by assuming prior knowledge that the autoregressive parameter is either equal to one, or randomly distributed around a mean of unity. But if one does not know whether $\rho = 1$ or $|\rho| < 1$, then standard conventional Dickey-Fuller type tests are not valid, even asymptotically (see Granger and Swanson, 1997). Hence, the usefulness of the joint testing approach.

The paper is organized as follows. In Section 2 the model is explained in some detail. In Section 3 the test statistics and their limiting distribution under the null hypothesis are derived. In Section 4 the properties of the tests are investigated, by means of a Monte-Carlo experiment. Section 5 contains an empirical example and Section 6 concludes. All the proofs are provided in the Appendix.

The following notation and abbreviations will be frequently used throughout the paper. $W(s)$, $B(s)$, $\xrightarrow{w}$, $\xrightarrow{p}$, $\sim$, $\mathcal{F}_t$, i.i.d. $\left(0, \sigma^2\right)$, IN $\left(0, \sigma^2\right)$ and Δ will denote respectively the standard Brownian motion and the Brownian motion with variance σ^2 on $s \in [0, 1]$, weak convergence of the associated probability measure, convergence in probability, equality in distribution, filtration at time t, distributed identically and independently with mean 0 and variance σ^2, normally and independently distributed with mean 0 and variance σ^2, and a generic constant. $\widetilde{\boldsymbol{\theta}}$ will denote the value of the vector of parameters $\boldsymbol{\theta}$ implied by the null hypothesis. The indicator function $\mathbf{1}_{\mathcal{K}}$ gives 1 if condition $\mathcal{K}$ is satisfied and zero otherwise. DGP, AR (p), LM, CMT and OLS will denote, respectively, Data Generating Process, AutoRegressive model of order p, Lagrange Multiplier, Continuous Mapping Theorem and Ordinary Least Squares.

2 The model

This section explains the main analytical features of the chosen random coefficient autoregressive model in detail.

Treating y_0 as equal to zero,[1] model (1) can be solved by backward substitution yielding

$$y_t = \sum_{j=1}^{t-1} \left(\prod_{i=0}^{j-1} \rho_{t-i} \right) \varepsilon_{t-j} + \varepsilon_t.$$

The first two unconditional moments of the series and the formula for the autocovariances are given

[1] This is only needed for simplifying the subsequent exposition. One can allow for $y_0 = O(1)$ or $y_0 = O_p(1)$ without affecting the asymptotic results given in the paper.

Accepted manuscript

ACCEPTED MANUSCRIPT

respectively by

$$\mathrm{E}\left(y_t\right)=0, \qquad \mathrm{var}\left(y_t\right)=\sigma_\varepsilon^2\left[\sum_{j=0}^{t-1}\left(\rho^2+\omega^2\right)^j\right]$$

and

$$\mathrm{cov}\left(y_t, y_{t+l}\right)=\sigma_\varepsilon^2\rho^l\left[\sum_{j=0}^{t-1}\left(\rho^2+\omega^2\right)^j\right]. \tag{2}$$

Therefore the series is not covariance stationary for $\rho^2+\omega^2 \geq 1$. However, the impulse response function is finite when $|\rho|<1$.

Analysis of (2) highlights further features of the model. In fact (2) can be expanded as

$$\mathrm{cov}\left(y_t, y_{t+l}\right)=\begin{cases}\sigma_\varepsilon^2\rho^l\frac{1-\left(\rho^2+\omega^2\right)^t}{1-\left(\rho^2+\omega^2\right)}, & \text{if } \left(\rho^2+\omega^2\right)\neq 1\\ \\ \sigma_\varepsilon^2\rho^l t, & \text{if } \left(\rho^2+\omega^2\right)=1\end{cases}$$

and in the (asymptotically) stationary case

$$\lim_{t\to\infty}\mathrm{cov}\left(y_t, y_{t+l}\right)=\frac{\sigma_\varepsilon^2}{1-\rho^2-\omega^2}\rho^l.$$

Therefore, for small l and $1-\rho^2-\omega^2$ sufficiently close to 0, the model can accomplish large initial values of autocovariances, but the decay is still exponentially fast. This can also be seen by looking at the autocorrelation function, defined by

$$\begin{aligned}\mathrm{corr}\left(y_t, y_{t+l}\right) &\equiv \frac{\mathrm{cov}\left(y_t, y_{t+l}\right)}{\sqrt{\mathrm{var}\left(y_t\right)\mathrm{var}\left(y_{t+l}\right)}}\\ &= \begin{cases}\rho^l\frac{1-\left(\rho^2+\omega^2\right)^t}{\sqrt{\left(1-\left(\rho^2+\omega^2\right)^t\right)\left(1-\left(\rho^2+\omega^2\right)^{t+l}\right)}}, & \text{if } \left(\rho^2+\omega^2\right)\neq 1\\ \\ \rho^l\frac{t}{\sqrt{t(t+l)}}, & \text{if } \left(\rho^2+\omega^2\right)=1.\end{cases}\end{aligned}$$

If $\left(\rho^2+\omega^2\right)\leq 1$, then

$$\lim_{t\to\infty}\mathrm{corr}\left(y_t, y_{t+l}\right)=\rho^l.$$

Instead, if $\left(\rho^2+\omega^2\right)>1$, then

$$\lim_{t\to\infty}\mathrm{corr}\left(y_t, y_{t+l}\right)=\frac{\rho^l}{\sqrt{\left(\rho^2+\omega^2\right)^l}}.$$

A special subcase occurs when the series is not covariance stationary. When $\rho^2+\omega^2\geq 1$ but $|\rho|<1$, the series is (covariance) nonstationary, but its impulse response function is finite and the decay of the autocorrelation function is exponential. Also, covariance nonstationarity is not removed when the series is differenced.

Accepted manuscript

ACCEPTED MANUSCRIPT

Simulations reported by Yoon (2002) show that if one lets $y_t = \log(S_t)$, where S_t represents the price of a financial asset, and $r_t = y_t - y_{t-1}$ the return at time t, then autoregressive random coefficient models are capable of allowing for high persistence in the autocorrelation function of $\{|r_t|\}$ and $\{r_t^2\}$, two stylized facts discovered and analyzed by Ding et al. (1993) and Ding and Granger (1996) (see also Granger, 1980, on the relationship between long memory models and the aggregation of random coefficient autoregressive models).

Another interesting feature of random coefficient autoregressive models is that, similarly to conditional heteroskedastic models, they can be successfully used with financial data, which are characterized by fat tails and volatility clustering. As a matter of fact, random coefficient models can be considered as a particular case of a conditional heteroskedastic model. Indeed, model (1) can be equivalently written as

$$y_t = \rho y_{t-1} + u_t, \tag{3}$$

where[2]

$$\mathrm{E}(u_t | y_{t-1}) = 0, \quad \mathrm{var}(u_t | y_{t-1}) = \sigma_\varepsilon^2 h_t(\boldsymbol{\theta}) \quad \text{and} \quad h_t(\boldsymbol{\theta}) = 1 + \omega^2/\sigma_\varepsilon^2 y_{t-1}^2. \tag{4}$$

For a discussion of the several similarities arising between random coefficient autoregressive and conditional heteroskedastic models, see Tsay (1987) and Leybourne et al. (1996).

In the context of model (1), it can be of interest to test the following set of hypotheses:

(i)

$$\mathrm{H}_0 : \omega^2 = 0, \tag{5}$$

versus

$$\mathrm{H}_1 : \omega^2 \neq 0.$$

Under the hypothesis defined in (5), the series follows a possibly nonstationary autoregressive process;

(ii)

$$\mathrm{H}_0 : \rho = 1, \tag{6}$$

versus

$$\mathrm{H}_1 : |\rho| < 1.$$

Under the null hypothesis defined in (6), the series follows a (possibly stochastic) unit root process;

(iii)

$$\mathrm{H}_0 : \rho = 1 \quad \text{and} \quad \omega^2 = 0, \tag{7}$$

[2]This was kindly pointed out by a referee.

Accepted manuscript

ACCEPTED MANUSCRIPT

versus the general alternative defined by

$$\mathrm{H}_1 : |\rho| < 1 \quad \text{or} \quad \omega^2 \neq 0.$$

The alternative hypothesis above is quite general and allows the series to have a randomized unit root (as in the case investigated by Leybourne et al., 1996) or to be an asymptotically stationary random coefficient autoregressive process, or lastly to be non stationary but with an exponentially decaying autocorrelation function (more precisely when $|\rho| < 1$, $\omega^2 \neq 0$, and $\rho^2 + \omega^2 \geq 1$). Under the null hypothesis in (7), the process generating the time series has an exact unit root and therefore is a simple random walk.

In the next section the relevant test statistics are derived, and their properties assessed.

3 Derivation of the test statistic

3.1 The Gaussian case

Conditioning on the information set up to time $(t-1)$ yields

$$\mathrm{E}\left(y_t|\mathcal{F}_{t-1}\right) = y_{t-1}\,\mathrm{E}\left(\rho_t|\mathcal{F}_{t-1}\right) + \mathrm{E}\left(\varepsilon_t|\mathcal{F}_{t-1}\right),$$

$$\mathrm{var}\left(y_t|\mathcal{F}_{t-1}\right) = y_{t-1}^2\,\mathrm{var}\left(\rho_t|\mathcal{F}_{t-1}\right) + \mathrm{var}\left(\varepsilon_t|\mathcal{F}_{t-1}\right) + 2y_{t-1}\,\mathrm{cov}\left(\varepsilon_t, \rho_t|\mathcal{F}_{t-1}\right).$$

The following assumption specifies the properties of $\{\rho_t\}$ and $\{\varepsilon_t\}$.

Assumption 1.

(i) ρ_t *is* i.i.d. $\left(\rho, \omega^2\right)$*;*

(ii) ε_t *is* IN $\left(0, \sigma_\varepsilon^2\right)$*;*

(iii) $\{\varepsilon_t\}$ *and* $\{\rho_t\}$ *are stochastically independent.*

The Gaussianity assumption is removed in the next subsection, where robust tests are considered. The log-likelihood function takes the form of

$$\ell\left(\boldsymbol{\theta}\right) = \text{const} - \frac{1}{2}\sum_{t=2}^{n}\log\left(y_{t-1}^2\omega^2 + \sigma_\varepsilon^2\right) - \frac{1}{2}\sum_{t=2}^{n}\frac{\left(y_t - \rho y_{t-1}\right)^2}{y_{t-1}^2\omega^2 + \sigma_\varepsilon^2}, \tag{8}$$

where $\boldsymbol{\theta}' \equiv \left[\rho\ \omega^2\ \sigma_\varepsilon^2\right]'$. The score vector is defined by $\boldsymbol{q}\left(\boldsymbol{\theta}\right) \equiv \partial\ell\left(\boldsymbol{\theta}\right)/\partial\boldsymbol{\theta}$, and its components are given by

$$q_1\left(\boldsymbol{\theta}\right) \equiv \frac{\partial\ell\left(\boldsymbol{\theta}\right)}{\partial\rho} = \sum_{t=2}^{n}\frac{y_{t-1}\left(y_t - \rho y_{t-1}\right)}{\varpi_{t-1}^2}, \quad q_2\left(\boldsymbol{\theta}\right) \equiv \frac{\partial\ell\left(\boldsymbol{\theta}\right)}{\partial\omega^2} = \frac{1}{2}\sum_{t=2}^{n}\frac{\left(y_t - \rho y_{t-1}\right)^2\left(y_{t-1}^2\right)}{\left(\varpi_{t-1}^2\right)^2} - \frac{1}{2}\sum_{t=2}^{n}\frac{y_{t-1}^2}{\varpi_{t-1}^2},$$

Accepted manuscript

ACCEPTED MANUSCRIPT

$$q_3(\boldsymbol{\theta}) \equiv \frac{\partial \ell(\boldsymbol{\theta})}{\partial \sigma_\varepsilon^2} = \frac{1}{2}\sum_{t=2}^{n} \frac{(y_t - \rho y_{t-1})^2}{\left(\varpi_{t-1}^2\right)^2} - \frac{1}{2}\sum_{t=2}^{n} \frac{1}{\varpi_{t-1}^2},$$

where $\varpi_{t-1}^2 \equiv y_{t-1}^2\omega^2 + \sigma_\varepsilon^2$. The elements of the Hessian matrix, defined by $\boldsymbol{\mathcal{H}}(\boldsymbol{\theta}) \equiv \partial^2 \ell(\boldsymbol{\theta}) / (\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}')$, take the form of

$$\mathcal{H}_{11}(\boldsymbol{\theta}) \equiv \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \rho^2} = -\sum_{t=2}^{n} \frac{y_{t-1}^2}{\varpi_{t-1}^2}, \quad \mathcal{H}_{12}(\boldsymbol{\theta}) \equiv \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \rho \partial \omega^2} = -\sum_{t=2}^{n} \frac{y_{t-1}^3 (y_t - \rho y_{t-1})}{\left(\varpi_{t-1}^2\right)^2},$$

$$\mathcal{H}_{13}(\boldsymbol{\theta}) \equiv \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \rho \partial \sigma_\varepsilon^2} = -\sum_{t=2}^{n} \frac{y_{t-1}(y_t - \rho y_{t-1})}{\left(\varpi_{t-1}^2\right)^2}, \quad \mathcal{H}_{22}(\boldsymbol{\theta}) \equiv \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \omega^4} = \frac{1}{2}\sum_{t=2}^{n} \frac{y_{t-1}^4}{\left(\varpi_{t-1}^2\right)^2} - \sum_{t=2}^{n} \frac{y_{t-1}^4 (y_t - \rho y_{t-1})^2}{\left(\varpi_{t-1}^2\right)^3},$$

$$\mathcal{H}_{23}(\boldsymbol{\theta}) \equiv \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \omega^2 \partial \sigma_\varepsilon^2} = \frac{1}{2}\sum_{t=2}^{n} \frac{y_{t-1}^2}{\left(\varpi_{t-1}^2\right)^2} - \sum_{t=2}^{n} \frac{y_{t-1}^2 (y_t - \rho y_{t-1})^2}{\left(\varpi_{t-1}^2\right)^3},$$

$$\mathcal{H}_{33}(\boldsymbol{\theta}) \equiv \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \sigma_\varepsilon^4} = \frac{1}{2}\sum_{t=2}^{n} \frac{1}{\left(\varpi_{t-1}^2\right)^2} - \sum_{t=2}^{n} \frac{(y_t - \rho y_{t-1})^2}{\left(\varpi_{t-1}^2\right)^3}.$$

The formulae given above need to be evaluated at particular values of the unknown parameters, in order for the test statistics to be operational. In this paper the score vector and the Hessian matrix are evaluated at the parameter values implied by the null hypotheses. The proposed test statistic is, therefore, an LM type test.

The choice of the LM test is motivated by the fact that in this case it is particularly convenient to work under the null hypothesis, because the score vector and the Hessian matrix are then available in an analytic form. Therefore, when the null hypothesis is true, the limiting distribution of the test statistics can be derived.

It is worth mentioning that, under the unit root hypothesis, there is no uniformly most powerful test (see Anderson, 1948), since the minimal sufficient statistic is not one-dimensional, even asymptotically and in the simplest case of no deterministic components.[3]

The conventional LM test, using the Hessian matrix as a normalizing factor, takes the general form of

$$\text{ALM} = \boldsymbol{q}\left(\widetilde{\boldsymbol{\theta}}\right)' \left(-\boldsymbol{\mathcal{H}}\left(\widetilde{\boldsymbol{\theta}}\right)\right)^{-1} \boldsymbol{q}\left(\widetilde{\boldsymbol{\theta}}\right).$$

Here the suffix 'A' stands for approximate and reflects the fact that the random Hessian matrix has been used for normalizing the score vector, as opposed to its deterministic counterpart, the exact information matrix, which is defined by $\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}) \equiv \mathrm{E}(-\boldsymbol{\mathcal{H}}(\boldsymbol{\theta}))$. This choice has not been arbitrary, and finds some justification in the fact that LM test statistics based on the exact information matrix have low power when testing for a unit root (see Abadir, 1993).

[3]Following related work by Dufour and King (1991), Elliott et al. (1996) and Xiao (2001), in fixed coefficient models it is possible to derive point optimal tests for the unit root hypothesis, namely tests that are tangent to the power envelope at least for that point alternative.

Accepted manuscript

ACCEPTED MANUSCRIPT

The test for hypothesis (5) will be considered first. Its formula is given by

$$\mathrm{ALM}^{(\omega^2)} = \frac{q_2\left(\widetilde{\boldsymbol{\theta}}\right)^2}{-\mathcal{H}_{22}\left(\widetilde{\boldsymbol{\theta}}\right)} = \frac{\left(\sum_{t=2}^{n} y_{t-1}^2\left[\left(y_t-\widetilde{\rho}y_{t-1}\right)^2-\widetilde{\sigma}^2\right]\right)^2}{2\widetilde{\sigma}^2\sum_{t=2}^{n} y_{t-1}^4\left[2\left(y_t-\widetilde{\rho}y_{t-1}\right)^2-\widetilde{\sigma}^2\right]}, \tag{9}$$

where

$$\widetilde{\rho},\widetilde{\sigma}^2 = \arg\max_{\rho,\sigma^2} \ell\left(\boldsymbol{\theta}\right)|_{\omega^2=0}.$$

Theorem 1 shows that the two cases where $\rho = 1$ and $|\rho| < 1$ give rise to two different limiting distributions for the test statistic. Important precedents to the statistic in (9) include the paper by Cai et al. (1998). They consider tests for homoskedasticity in wavelet regression models. Starting from the model formulation given in (4), they construct LM tests for the absence of heteroskedasticity, similarly to the approach followed by Breusch and Pagan (1979). Although very similar in spirit to the LM test for homoscedasticity, the test proposed in this paper differs from the one by Cai et al. (1998) in two respects: first, as shown in (9), only the elements of the score vector corresponding to the parameters of interest are selected; second, the score is standardized by (minus) the Hessian matrix, rather than by the information matrix. This has been done purely for the purpose of maximizing the power of the proposed tests and is justified by previous findings in the literature on low power of tests based on the information matrix when the data are nonstationary. The results of the Monte-Carlo experiment, reported in Section 4, reveal the full extent of the power improvements. Of course, the test in (9) reduces to the one of Cai et al. (1998) if: (i) the full score vector is used, and (ii) the score is standardized by the information matrix.

The statistic for testing hypothesis (6) is given by

$$\mathrm{ALM}^{(\rho)} = \frac{q_1\left(\widetilde{\boldsymbol{\theta}}\right)}{\sqrt{-\mathcal{H}_{11}\left(\widetilde{\boldsymbol{\theta}}\right)}} = \frac{\sum_{t=2}^{n}\frac{y_{t-1}(y_t-y_{t-1})}{\widetilde{\omega}^2 y_{t-1}^2+\widetilde{\sigma}^2}}{\sqrt{\sum_{t=2}^{n}\frac{y_{t-1}^2}{\widetilde{\omega}^2 y_{t-1}^2+\widetilde{\sigma}^2}}}.$$

Theorem 1 shows that the test statistic above has a standard normal limiting distribution if the nuisance parameter $\omega^2 \neq 0$, and a nonstandard limiting distribution if $\omega^2 = 0$.

Finally, for the joint hypothesis (7), a modification to the conventional LM statistic will be proposed, in order to exploit the partially one-sided nature (in the autoregressive parameter ρ) of the alternative hypothesis and derive more powerful tests.

The method applied is the general one derived in Abadir and Distaso (2006). By incorporating the nature of the alternative hypothesis in the construction of the test statistic and therefore providing more information to the statistic, uniform power gains are achieved with respect to the unmodified counterparts. This is shown analytically when the components of the test are (at least asymptotically) normally distributed, and by Monte-Carlo simulations in nonstandard setups, such as the present one. An application to the modification of the LM test is given below.

Accepted manuscript

ACCEPTED MANUSCRIPT

The modified LM test statistic can be derived in the following way. After having partitioned the score vector

$$\boldsymbol{q}(\boldsymbol{\theta}) = \begin{pmatrix} q_1(\boldsymbol{\theta}) \\ \boldsymbol{q}_2(\boldsymbol{\theta}) \end{pmatrix},$$

where $\boldsymbol{q}_2(\boldsymbol{\theta}) = \begin{pmatrix} q_2(\boldsymbol{\theta}) \\ q_3(\boldsymbol{\theta}) \end{pmatrix}$ and the Hessian matrix accordingly

$$\boldsymbol{\mathcal{H}}(\boldsymbol{\theta}) = \begin{bmatrix} \mathcal{H}_{11}(\boldsymbol{\theta}) & \boldsymbol{h}_{21}(\boldsymbol{\theta})' \\ \boldsymbol{h}_{21}(\boldsymbol{\theta}) & \boldsymbol{\mathcal{H}}_{22}(\boldsymbol{\theta}) \end{bmatrix},$$

it is possible to modify the score vector

$$\boldsymbol{q}^{\perp}(\boldsymbol{\theta}) \equiv \begin{pmatrix} q_1(\boldsymbol{\theta}) \\ \boldsymbol{q}_{2_\perp 1}(\boldsymbol{\theta}) \end{pmatrix} = \begin{pmatrix} q_1(\boldsymbol{\theta}) \\ \boldsymbol{q}_2(\boldsymbol{\theta}) - \frac{\boldsymbol{h}_{21}(\boldsymbol{\theta})}{\mathcal{H}_{11}(\boldsymbol{\theta})} q_1(\boldsymbol{\theta}) \end{pmatrix},$$

such that the modified Hessian matrix is given by

$$-\boldsymbol{\mathcal{H}}^{\perp}(\boldsymbol{\theta}) = -\begin{bmatrix} \mathcal{H}_{11}(\boldsymbol{\theta}) & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{\mathcal{H}}_{2_\perp 1}(\boldsymbol{\theta}) \end{bmatrix} = -\begin{bmatrix} \mathcal{H}_{11}(\boldsymbol{\theta}) & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{\mathcal{H}}_{22}(\boldsymbol{\theta}) - \frac{1}{\mathcal{H}_{11}(\boldsymbol{\theta})} \boldsymbol{h}_{21}(\boldsymbol{\theta}) \boldsymbol{h}_{21}(\boldsymbol{\theta})' \end{bmatrix},$$

and implying that $q_1(\boldsymbol{\theta})$ and $\boldsymbol{q}_{2_\perp 1}(\boldsymbol{\theta})$ are orthogonal. Then the modified test statistic is given by

$$\begin{aligned}
\text{AMLM} &= \mathbf{1}_{\frac{q_1(\widetilde{\boldsymbol{\theta}})}{n}<0} \left[\begin{pmatrix} 1 & 0 & 0 \end{pmatrix} \left(-\boldsymbol{\mathcal{H}}\left(\widetilde{\boldsymbol{\theta}}\right)\right) \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}' \right]^{-1} \left(q_1\left(\widetilde{\boldsymbol{\theta}}\right)\right)^2 \\
&\quad + \left(q_2\left(\widetilde{\boldsymbol{\theta}}\right) - \frac{\mathcal{H}_{12}\left(\widetilde{\boldsymbol{\theta}}\right)}{\mathcal{H}_{11}\left(\widetilde{\boldsymbol{\theta}}\right)} q_1\left(\widetilde{\boldsymbol{\theta}}\right) \right)^2 \left[\begin{pmatrix} -\frac{\mathcal{H}_{21}(\widetilde{\boldsymbol{\theta}})}{\mathcal{H}_{11}(\widetilde{\boldsymbol{\theta}})} & 1 & 0 \end{pmatrix} \left(-\boldsymbol{\mathcal{H}}\left(\widetilde{\boldsymbol{\theta}}\right)\right) \begin{pmatrix} -\frac{\mathcal{H}_{21}(\widetilde{\boldsymbol{\theta}})}{\mathcal{H}_{11}(\widetilde{\boldsymbol{\theta}})} & 1 & 0 \end{pmatrix}' \right]^{-1} \\
&= \mathbf{1}_{\frac{\sum y_{t-1}(y_t - y_{t-1})}{n\widetilde{\sigma}^2}<0} \frac{\left(\sum y_{t-1}(y_t - y_{t-1})\right)^2}{\widetilde{\sigma}^2 \sum y_{t-1}^2} + \frac{\left(\sum y_{t-1}^2 \left((y_t - y_{t-1})^2 - \widetilde{\sigma}^2\right)\right)^2}{2\widetilde{\sigma}^2 \left(2\sum y_{t-1}^4 (y_t - y_{t-1})^2 - \widetilde{\sigma}^2 \sum y_{t-1}^4\right)},
\end{aligned}$$

where the 'M' after the 'A' stands for modified. By considering only the cases where the score takes negative values, the test is specifically tailored to incorporate the direction of the alternative hypothesis of interest. This modification leads to power improvements, as shown in Section 4.

The next theorem states the limiting densities of the test statistics under the null hypothesis.

Theorem 1. *Let Assumption 1 hold. In model (1):*

(i) when hypothesis (5) is true and $\rho = 1$, the following holds

$$\text{ALM}^{(\omega^2)} \xrightarrow{w} \frac{\left(\int_0^1 W_1(s)^2 \, \mathrm{d}W_2(s) - \int_0^1 W_1(s)^2 \, \mathrm{d}s \, W_2(1)\right)^2}{\int_0^1 W_1(s)^4 \, \mathrm{d}s},$$

where $W_1(s)$ and $W_2(s)$ are independent standard Brownian motions;

ACCEPTED MANUSCRIPT

(ii) when hypothesis (5) is true and $|\rho| < 1$, the following holds

$$\mathrm{ALM}^{(\omega^2)} \xrightarrow{d} \chi_1^2;$$

(iii) when hypothesis (6) is true and $\omega^2 = 0$, the following holds

$$\mathrm{ALM}^{(\rho)} \xrightarrow{w} \frac{\left(W_1(1)^2 - 1\right)}{2\sqrt{\int_0^1 W_1(s)^2 \,\mathrm{d}s}}; \tag{10}$$

(iv) when hypothesis (6) is true and $\omega^2 \neq 0$, the following holds

$$\mathrm{ALM}^{(\rho)} \xrightarrow{d} \mathrm{N}(0,1);$$

(v) finally, when hypothesis (7) is true, then

$$\mathrm{AMLM} \xrightarrow{w} \mathbf{1}_{\frac{1}{2}\left(W_1(1)^2-1\right)<0} \frac{\left(W_1(1)^2 - 1\right)^2}{4\int_0^1 W_1(s)^2 \,\mathrm{d}s} + \frac{\left(\int_0^1 W_1(s)^2 \,\mathrm{d}W_2(s) - \int_0^1 W_1(s)^2 \,\mathrm{d}s\, W_2(1)\right)^2}{\int_0^1 W_1(s)^4 \,\mathrm{d}s}.$$

The presence of a unit root in the DGP gives rise to a nonstandard limiting distribution of the statistic for testing (5). In addition, when $\omega^2 = 0$, the limiting distribution of the statistic for hypothesis (6) is given in terms of Wiener processes, rather than standard normal.[4] This raises the important point of using the correct critical values in empirical work. For example, if one had *a priori* information that $\omega^2 = 0$, he would use the critical values of the distribution given in (10) for $\mathrm{ALM}^{(\rho)}$. Otherwise, he would use the quantiles of the standard normal. However, this information on ω^2 is not available. There are three possible ways to address this problem: the first method, which is valid for $\mathrm{ALM}^{(\omega^2)}$, consists in estimating ρ, which can be done consistently under both the null and alternative hypothesis (since the information matrix is block diagonal, see Breusch and Pagan, 1979), and then use the critical value corresponding to $\widehat{\rho}$. For $\mathrm{ALM}^{(\rho)}$, one has two available options: the first is to use always the critical values of the nonstandard distributions, which are larger than the normal ones. This implies a conservative testing strategy for hypothesis (6).[5] The second option is to use model selection criteria, such as those by Phillips and Ploberger (1996), which allow for potentially nonstationary data.

Finally, asymptotic critical values can be obtained by simulation methods, using a very large sample size, since the limiting distributions are not known in close form. This approach is followed in Section 4, where asymptotic critical values are obtained by fixing $n = 1000$ (see Tables 6 and 8).

[4]Notice that this result is in contrast with what happens in unit root processes with GARCH innovations. There, unit root tests have the usual Dickey-Fuller distribution. Here, the limiting distribution changes according to the value of ω^2. The reason for this difference lies in the fact that, from (4), the conditional variance is nonstationary when $\omega^2 \neq 0$ and therefore the standard proof for the validity of the Dickey-Fuller distribution in unit root models with GARCH innovations cannot be applied.

[5]The implied size distortion is not severe, as confirmed by the results of Section 4.

Accepted manuscript

ACCEPTED MANUSCRIPT

Inspection of the formulae for the limiting distributions just derived reveals some similarities with results obtained in the literature. The first part of the formula for the limiting density of AMLM is the positively censored and then squared equivalent of the limiting distribution of the LM test for a unit root in a simple AR(1) model (see Phillips, 1987, equation (11), p.283). The numerator in the formula for the second component of AMLM is the same as the one derived by Leybourne et al. (1996); the difference lies in the choice of the normalizing factor, which in their case is a deterministic factor ensuring that the (normalized) score has a non-degenerate limiting distribution, whereas in the present paper it is the Hessian matrix.

Analysis of the probabilistic orders of magnitude of the Hessian matrix suggests a further useful result.

Corollary 1. *Let Assumption 1 hold. In model (1), when hypothesis (7) is true, the test statistics*

$$\mathbf{1}_{\frac{q_1(\widetilde{\boldsymbol{\theta}})}{n}<0}\frac{\left(q_1\left(\widetilde{\boldsymbol{\theta}}\right)\right)^2}{-\mathcal{H}_{11}\left(\widetilde{\boldsymbol{\theta}}\right)}=\mathbf{1}_{\frac{\sum y_{t-1}(y_t-y_{t-1})}{n\widetilde{\sigma}^2}<0}\frac{\left(\sum y_{t-1}\left(y_t-y_{t-1}\right)\right)^2}{\widetilde{\sigma}^2\sum y_{t-1}^2}$$

and

$$\frac{q_2\left(\widetilde{\boldsymbol{\theta}}\right)^2}{-\mathcal{H}_{22}\left(\widetilde{\boldsymbol{\theta}}\right)}=\frac{\left(\sum_{t=2}^n y_{t-1}^2\left[\left(y_t-y_{t-1}\right)^2-\widetilde{\sigma}^2\right]\right)^2}{2\widetilde{\sigma}^2\sum_{t=2}^n y_{t-1}^4\left[2\left(y_t-y_{t-1}\right)^2-\widetilde{\sigma}^2\right]}$$

are asymptotically uncorrelated.

Then, in applied work, provided that the sample size is large, one can use the simpler version of the joint test, given by the sum of the (censored and then squared) test on ρ and the test on ω^2, without orthogonalizing the two components. Of course, in small samples, this simpler test could suffer from some loss of power.

In fact, the nature of the alternative hypothesis regarding ω^2 is not really two-sided, since variances can only take non-negative values. However, there is no need to make any correction or modification for this in the joint test statistic, since the test statistic on ω^2 is virtually one-sided (see McCabe and Tremayne, 1995; Leybourne et al., 1996).

Several extensions to the above setup are possible. The first one is allowing for a non-zero correlation between ρ_t and ε_t (i.e. setting $\operatorname{cov}\left(\varepsilon_t,\rho_t|\mathcal{F}_{t-1}\right)=\varrho\omega^2$). This does not alter the limiting distribution of the test statistics, as the terms involving ϱ turn out to be of a smaller probabilistic order and therefore can be asymptotically neglected (see Leybourne et al., 1996).[6] A possible intuition behind this result is the similarity between random coefficient autoregressive models and conditionally heteroskedastic autoregressive

[6] In this case, the score for the parameter ω^2 would take the form of

$$q_2\left(\widetilde{\boldsymbol{\theta}}\right)=\frac{\sum y_{t-1}^2\left(\varepsilon_t^2-\widetilde{\sigma}^2\right)+2\varrho\sum y_{t-1}\left(\varepsilon_t^2-\widetilde{\sigma}^2\right)}{2\widetilde{\sigma}^4},$$

which is clearly dominated by the first component. A similar result holds for the elements of the Hessian matrix.

Accepted manuscript

ACCEPTED MANUSCRIPT

models, where the parameters of the mean and variance are asymptotically uncorrelated.[7] This property is verified in a small Monte-Carlo exercise in Section 4.

It would certainly be of interest to relax the i.i.d. assumption on $\{\rho_t\}$, by modeling it as an autoregressive process, and then proceed to estimate the model, taking into account serial dependence (e.g. Kalman filter). However, one of the main points of the paper is to provide a test on the mean of the autoregressive random variable being equal to one. Then, it would not be clear what to test in this case, since the mean of the process $\{\rho_t\}$ would depend on the initial value of the process. Therefore, this extension is not considered in the paper.

Also, deterministic components can be added to the model, to cover more general data generating processes. This paper considers a constant and a linear trend, though extension to higher order trends is straightforward.[8] Model (1) can then be extended in the following way

$$\begin{aligned} y_1 &= \alpha + \beta + \varepsilon_1 \\ y_t - \alpha - \beta t &= \rho_t \left(y_{t-1} - \alpha - \beta \left(t-1 \right) \right) + \varepsilon_t, \quad t = 2, \ldots, n. \end{aligned} \tag{11}$$

Notice that the constant and trend term are included in the model following the work of Bhargava (1986); this achieves the advantage that the resulting test statistics are invariant to the values of the nuisance parameters in the DGP, as shown by results from DeJong et al. (1992) and Distaso (2002). This fact sets a marked difference with respect to the widely used linear (Dickey-Fuller) specification, where different limiting distributions arise according to the value of the nuisance parameters in the DGP (for a detailed discussion about the differences about the two models and an intuition about why different distributions arise, see Schmidt and Phillips, 1992).

The log-likelihood in the augmented model takes the form of

$$\ell\left(\boldsymbol{\theta}_a\right) = \text{const} - \frac{1}{2}\log\left(\sigma_\varepsilon^2\right) - \frac{1}{2\sigma_\varepsilon^2}\left(y_1 - \alpha - \beta\right)^2 - \frac{1}{2}\sum_{t=2}^{n}\log\left(y_{t-1}^{*2}\omega^2 + \sigma_\varepsilon^2\right) - \frac{1}{2}\sum_{t=2}^{n}\frac{\left(y_t^* - \rho y_{t-1}^*\right)^2}{y_{t-1}^{*2}\omega^2 + \sigma_\varepsilon^2},$$

where $\boldsymbol{\theta}_a' \equiv \left[\rho\ \omega^2\ \sigma_\varepsilon^2\ \alpha\ \beta\right]'$ and $y_t^* \equiv y_t - \alpha - \beta t$.

Suitable test statistics (denoted by ALM_a and AMLM_a) for hypotheses (5), (6) and (7) can be constructed in a similar way to the case analyzed above.

Under the null hypothesis (5), the series becomes an autoregressive model with a possible unit root; under (6), the series behaves like a (possibly randomized) unit root process plus a drift. Finally, under (7) it simplifies to a random walk with with a possibly nonzero drift. In particular, the test statistics are

[7] This is not the case, though, for some GARCH specifications, such as GARCH in mean.

[8] Similarly to results obtained by Elliott et al. (1996), Phillips and Xiao (1998) and Pere (2003), the limiting distributions of the test statistics on ρ are not affected by the presence of constants or slowly trending regressors. Therefore, the case where only an intercept α is included in the model gives rise to the same limiting distributions as in Theorem 1.

Accepted manuscript

ACCEPTED MANUSCRIPT

given by

$$\mathrm{ALM}_a^{(\omega^2)} = \frac{\left(\sum y_{t-1}^{*2}\left(\left(y_t^* - \widetilde{\rho} y_{t-1}^*\right)^2 - \widetilde{\sigma}^2\right)\right)^2}{2\widetilde{\sigma}^2\left(2\sum y_{t-1}^{*4}\left(y_t^* - \widetilde{\rho} y_{t-1}^*\right)^2 - \widetilde{\sigma}^2 \sum y_{t-1}^{*4}\right)}, \tag{12}$$

$$\mathrm{ALM}_a^{(\rho)} = \frac{\sum \frac{y_{t-1}^*\left(y_t^* - y_{t-1}^*\right)}{\widetilde{\omega}^2 y_{t-1}^{*2} + \widetilde{\sigma}^2}}{\sqrt{\sum \frac{y_{t-1}^{*2}}{\widetilde{\omega}^2 y_{t-1}^{*2} + \widetilde{\sigma}^2}}} \tag{13}$$

and

$$\mathrm{AMLM}_a = \mathbf{1}_{\sum y_{t-1}^*\left(y_t^* - y_{t-1}^*\right) < 0} \frac{\left(\sum y_{t-1}^*\left(y_t^* - y_{t-1}^*\right)\right)^2}{\widetilde{\sigma}^2 \sum y_{t-1}^{*2}} + \frac{\left(\sum y_{t-1}^{*2}\left(\left(y_t^* - y_{t-1}^*\right)^2 - \widetilde{\sigma}^2\right)\right)^2}{2\widetilde{\sigma}^2\left(2\sum y_{t-1}^{*4}\left(y_t^* - y_{t-1}^*\right)^2 - \widetilde{\sigma}^2 \sum y_{t-1}^{*4}\right)}. \tag{14}$$

In order to use the statistics above in practical work, one needs to substitute in (12), (13) and (14) some consistent estimators for α and β.

Analytical results and simulations reported in Distaso (2002) reveal that for testing hypothesis (7), if one chooses the fully restricted estimators (i.e. obtained imposing the restriction implied by the hypotheses $\omega^2 = 0$ and $\rho = 1$) for α and β, then the power of the test is substantially (often dramatically) reduced. Therefore the partially restricted estimators, namely $\widehat{\alpha}$ and $\widehat{\beta}$, resulting from

$$\widehat{\alpha}, \widehat{\beta}, \widehat{\rho}, \widehat{\sigma}_\varepsilon^2 = \arg \max_{\alpha,\beta,\rho,\sigma_\varepsilon^2} \ell\left(\boldsymbol{\theta}_a\right)|_{\omega^2=0}$$

will be used in the construction of the statistic. Since the function to be minimized is nonlinear in the parameters, the following stepwise (iterative) least squares estimation procedure is used:

1) Start from an estimator of ρ and then estimate α and β from

$$\begin{aligned} y_1 &= \alpha + \beta + \varepsilon_1, \\ y_t - \widehat{\rho} y_{t-1} &= \alpha\left(1 - \widehat{\rho}\right) + \beta\left[t - \widehat{\rho}\left(t-1\right)\right] + \varepsilon_t, \quad t = 2, \ldots, n \end{aligned}$$

by OLS.

2) Having obtained an estimate of α and β, the estimate of ρ can be updated from

$$y_t - \widehat{\alpha} - \widehat{\beta} t = \rho\left(y_{t-1} - \widehat{\alpha} - \widehat{\beta}\left(t-1\right)\right) + \varepsilon_t, \quad t = 2, \ldots, n$$

again using OLS.

3) Return to step 1 and repeat until convergence.

Convergence of the estimation procedure is established in Distaso (2002). The limiting distribution of the test statistics are derived in the following theorem.

Theorem 2. *Let Assumption 1 hold. In model (11):*

Accepted manuscript

ACCEPTED MANUSCRIPT

(i) when hypothesis (5) is true and $\rho = 1$, the following holds

$$\mathrm{ALM}_a^{(\omega^2)} \xrightarrow{w} \frac{\left(\int_0^1 (W_1(s) - s\xi)^2 \, \mathrm{d}W_2(s) - \int_0^1 (W_1(s) - s\xi)^2 \, \mathrm{d}s \, W_2(1)\right)^2}{\int_0^1 (W_1(s) - s\xi)^4 \, \mathrm{d}s},$$

where ξ denotes the limiting distribution of $n^{\frac{1}{2}}\left(\widehat{\beta} - \beta\right)/\sigma_\varepsilon$. The asymptotic behaviour of ξ and η, where η denotes the limiting distribution of $n(\widehat{\rho} - 1)$, is determined jointly by the following stochastic differential equations

$$\begin{aligned}
\xi &= \frac{\int_0^1 (1 - \eta s) \, \mathrm{d}W_1(s) - \eta \int_0^1 (1 - \eta s) W_1(s) \, \mathrm{d}s}{\int_0^1 (1 - \eta s)^2 \, \mathrm{d}s}, \\
\eta &= \frac{1/2\left(W_1(1)^2 - 1\right) - \xi\left(\int_0^1 s \, \mathrm{d}W_1(s) + \int_0^1 (W_1(s) - \xi s) \, \mathrm{d}s\right)}{\int_0^1 (W_1(s) - \xi s)^2 \, \mathrm{d}s}.
\end{aligned}$$

There is no analytic solution to these equations, and the limiting distributions are obtained by numerical methods;

(ii) when hypothesis (5) is true and $|\rho| < 1$, the following holds

$$\mathrm{ALM}_a^{(\omega^2)} \xrightarrow{d} (\mathrm{N}(0, V_a))^2,$$

where

$$V_a = \mathrm{plim} \frac{\mathrm{var}\left(1/\sqrt{n} \sum y_{t-1}^{*2} \left(\varepsilon_t^2 - \sigma_\varepsilon^2\right)\right)}{2\sigma_\varepsilon^4 \sum y_{t-1}^{*4}/n};$$

(iii) when hypothesis (6) is true and $\omega^2 = 0$, the following holds

$$\mathrm{ALM}_a^{(\rho)} \xrightarrow{w} \frac{\int_0^1 (W_1(s) - s\xi) \, \mathrm{d}W_1(s) - \xi \int_0^1 (W_1(s) - s\xi) \, \mathrm{d}s}{\sqrt{\int_0^1 (W_1(s) - s\xi)^2 \, \mathrm{d}s}};$$

(iv) when hypothesis (6) is true and $\omega^2 \neq 0$, the following holds

$$\mathrm{ALM}_a^{(\rho)} \xrightarrow{d} \mathrm{N}(0, 1);$$

(v) finally, when hypothesis (7) is true the following holds

$$\begin{aligned}
\mathrm{AMLM}_a \xrightarrow{w} \; & \mathbf{1}_{\left[\int_0^1 (W_1(s) - s\xi) \, \mathrm{d}W_1(s) - \xi \int_0^1 (W_1(s) - s\xi) \, \mathrm{d}s\right] < 0} \\
& \times \frac{\left(\int_0^1 (W_1(s) - s\xi) \, \mathrm{d}W_1(s) - \xi \int_0^1 (W_1(s) - s\xi) \, \mathrm{d}s\right)^2}{\int_0^1 (W_1(s) - s\xi)^2 \, \mathrm{d}s} \\
& + \frac{\left(\int_0^1 (W_1(s) - s\xi)^2 \, \mathrm{d}W_2(s) - \int_0^1 (W_1(s) - s\xi)^2 \, \mathrm{d}s \, W_2(1)\right)^2}{\int_0^1 (W_1(s) - s\xi)^4 \, \mathrm{d}s}.
\end{aligned}$$

Accepted manuscript

ACCEPTED MANUSCRIPT

3.2 Robust tests

In this subsection, the Gaussianity assumption is removed and some dependence is allowed for $\{\varepsilon_t\}$. In particular, ρ_t and $\boldsymbol{\zeta}_t \equiv \left(\varepsilon_t, \varepsilon_t^2 - \sigma_\varepsilon^2\right)'$ satisfy the following assumption.

Assumption 2.

(i) ρ_t *is* i.i.d. $\left(\rho, \omega^2\right)$*;*

(ii) $\mathrm{E}\left(\varepsilon_t\right) = 0$*;*

(iii) $\sup_t \mathrm{E}\left|\varepsilon_t\right|^{\gamma+\nu} < \infty$*, for some* $\gamma > 4$ *and* $\nu > 0$*;*

(iv) $\lim_{n\to\infty} n^{-1} \operatorname{var}\left(\sum_{t=1}^n \boldsymbol{\zeta}_t\right) = \boldsymbol{\Sigma}_{\boldsymbol{\zeta}}$*, having elements equal respectively to* σ^2*,* $\psi\sigma\kappa$ *and* κ^2 *and admitting a decomposition of the kind* $\boldsymbol{\Sigma}_{\boldsymbol{\zeta}}^{-1/2}\left(\boldsymbol{\Sigma}_{\boldsymbol{\zeta}}^{-1/2}\right)' = \boldsymbol{\Sigma}_{\boldsymbol{\zeta}}^{-1}$*;*

(v) $\{\varepsilon_t\}$ *is strong mixing with mixing coefficients* ι_m *satisfying*

$$\sum_{m=1}^{\infty} \iota_m^{1-2/\gamma} < \infty.$$

(vi) $\{\varepsilon_t\}$ *and* $\{\rho_t\}$ *are stochastically independent.*

It is clear that, under Assumption 2, the martingale difference sequence hypothesis $\mathrm{E}\left(\varepsilon_t|\mathcal{F}_{t-1}\right) = 0$ does not hold any more. The parameters of the model can still be estimated consistently maximizing (8), using a quasi maximum likelihood argument. However, in order to conduct correct inference, the standardizing factor needs to be changed. In particular, the general form of the robust approximate LM statistic becomes

$$\mathrm{ALMR} = \boldsymbol{q}\left(\widetilde{\boldsymbol{\theta}}\right)'\left(-\boldsymbol{\mathcal{H}}\left(\widetilde{\boldsymbol{\theta}}\right)\right)^{-1}\left(\boldsymbol{\mathcal{V}}\left(\widetilde{\boldsymbol{\theta}}\right)\right)^{-1}\left(-\boldsymbol{\mathcal{H}}\left(\widetilde{\boldsymbol{\theta}}\right)\right)^{-1}\boldsymbol{q}\left(\widetilde{\boldsymbol{\theta}}\right),$$

where 'R' stands for robust, $\boldsymbol{\mathcal{V}}\left(\widetilde{\boldsymbol{\theta}}\right) = \left(-\boldsymbol{\mathcal{H}}\left(\widetilde{\boldsymbol{\theta}}\right)\right)^{-1}\boldsymbol{\mathcal{A}}\left(\widetilde{\boldsymbol{\theta}}\right)\left(-\boldsymbol{\mathcal{H}}\left(\widetilde{\boldsymbol{\theta}}\right)\right)^{-1}$ and $\boldsymbol{\mathcal{A}}\left(\widetilde{\boldsymbol{\theta}}\right)$ denotes the outer product of the gradient matrix. Define $m_n = o\left(n^{1/4}\right)$, the truncation lag. Let $m_n \to \infty$ as $n \to \infty$, and the corresponding weights satisfy $|w_{n\tau}| \leq \Delta$, $n = 2, \ldots,$ $\tau = 1, \ldots, m_n$, such that $w_{n\tau} \to 1$ as $n \to \infty$, for each τ. Then, the generic elements of $\boldsymbol{\mathcal{A}}\left(\boldsymbol{\theta}\right)$ are given by

$$\mathcal{A}_{11}\left(\boldsymbol{\theta}\right) = \sum_{t=2}^{n} \frac{y_{t-1}^2\left(y_t - \rho y_{t-1}\right)^2}{\left(\varpi_{t-1}^2\right)^2} + 2\sum_{\tau=1}^{m_n} w_{n\tau} \sum_{t=\tau+2}^{n} \frac{y_{t-1}\left(y_t - \rho y_{t-1}\right)}{\varpi_{t-1}^2} \frac{y_{t-\tau-1}\left(y_{t-\tau} - \rho y_{t-\tau-1}\right)}{y_{t-\tau-1}^2\omega^2 + \sigma_\varepsilon^2},$$

$$\begin{aligned}\mathcal{A}_{12}\left(\boldsymbol{\theta}\right) \quad = \quad & \sum_{t=2}^{n} \frac{y_{t-1}^3\left(y_t - \rho y_{t-1}\right)\left[\left(y_t - \rho y_{t-1}\right)^2 - \left(\varpi_{t-1}^2\right)\right]}{2\left(\varpi_{t-1}^2\right)^3} \\ & + \sum_{\tau=1}^{m_n} w_{n\tau} \sum_{t=\tau+2}^{n} \frac{y_{t-1}\left(y_t - \rho y_{t-1}\right)}{\varpi_{t-1}^2} \frac{y_{t-\tau-1}^2\left[\left(y_{t-\tau} - \rho y_{t-\tau-1}\right)^2 - \left(\varpi_{t-\tau-1}^2\right)\right]}{2\left(\varpi_{t-\tau-1}^2\right)^2}\end{aligned}$$

ACCEPTED MANUSCRIPT

$$+\sum_{\tau=1}^{m_n} w_{n\tau} \sum_{t=\tau+2}^{n} \frac{y_{t-\tau-1}\left(y_{t-\tau}-\rho y_{t-\tau-1}\right)}{\varpi_{t-\tau-1}^2} \frac{y_{t-1}^2\left[\left(y_t-\rho y_{t-1}\right)^2-\left(\varpi_{t-1}^2\right)\right]}{2\left(\varpi_{t-1}^2\right)^2},$$

$$\begin{aligned}
\mathcal{A}_{13}(\boldsymbol{\theta}) &= \sum_{t=2}^{n} \frac{y_{t-1}\left(y_t-\rho y_{t-1}\right)\left[\left(y_t-\rho y_{t-1}\right)^2-\left(\varpi_{t-1}^2\right)\right]}{2\left(\varpi_{t-1}^2\right)^3} \\
&\quad +\sum_{\tau=1}^{m_n} w_{n\tau} \sum_{t=\tau+2}^{n} \frac{y_{t-1}\left(y_t-\rho y_{t-1}\right)}{\varpi_{t-1}^2} \frac{\left(y_{t-\tau}-\rho y_{t-\tau-1}\right)^2-\left(\varpi_{t-\tau-1}^2\right)}{2\left(\varpi_{t-\tau-1}^2\right)^2} \\
&\quad +\sum_{\tau=1}^{m_n} w_{n\tau} \sum_{t=\tau+2}^{n} \frac{y_{t-\tau-1}\left(y_{t-\tau}-\rho y_{t-\tau-1}\right)}{\varpi_{t-\tau-1}^2} \frac{\left(y_t-\rho y_{t-1}\right)^2-\left(\varpi_{t-1}^2\right)}{2\left(\varpi_{t-1}^2\right)^2},
\end{aligned}$$

$$\begin{aligned}
\mathcal{A}_{22}(\boldsymbol{\theta}) &= \sum_{t=2}^{n} \frac{y_{t-1}^4\left[\left(y_t-\rho y_{t-1}\right)^2-\left(\varpi_{t-1}^2\right)\right]^2}{4\left(\varpi_{t-1}^2\right)^4} \\
&\quad +2\sum_{\tau=1}^{m_n} w_{n\tau} \sum_{t=\tau+2}^{n} \frac{y_{t-1}^2\left[\left(y_t-\rho y_{t-1}\right)^2-\left(\varpi_{t-1}^2\right)\right]}{2\left(\varpi_{t-1}^2\right)^2} \frac{y_{t-\tau-1}^2\left[\left(y_{t-\tau}-\rho y_{t-\tau-1}\right)^2-\left(\varpi_{t-\tau-1}^2\right)\right]}{2\left(\varpi_{t-\tau-1}^2\right)^2},
\end{aligned}$$

$$\begin{aligned}
\mathcal{A}_{23}(\boldsymbol{\theta}) &= \sum_{t=2}^{n} \frac{y_{t-1}^2\left[\left(y_t-\rho y_{t-1}\right)^2-\left(\varpi_{t-1}^2\right)\right]^2}{4\left(\varpi_{t-1}^2\right)^4} \\
&\quad +\sum_{\tau=1}^{m_n} w_{n\tau} \sum_{t=\tau+2}^{n} \frac{y_{t-1}^2\left[\left(y_t-\rho y_{t-1}\right)^2-\left(\varpi_{t-1}^2\right)\right]}{2\left(\varpi_{t-1}^2\right)^2} \frac{\left(y_{t-\tau}-\rho y_{t-\tau-1}\right)^2-\left(\varpi_{t-\tau-1}^2\right)}{2\left(\varpi_{t-\tau-1}^2\right)^2} \\
&\quad +\sum_{\tau=1}^{m_n} w_{n\tau} \sum_{t=\tau+2}^{n} \frac{y_{t-\tau-1}^2\left[\left(y_{t-\tau}-\rho y_{t-\tau-1}\right)^2-\left(\varpi_{t-\tau-1}^2\right)\right]}{2\left(\varpi_{t-\tau-1}^2\right)^2} \frac{\left(y_t-\rho y_{t-1}\right)^2-\left(\varpi_{t-1}^2\right)}{2\left(\varpi_{t-1}^2\right)^2},
\end{aligned}$$

$$\mathcal{A}_{33}(\boldsymbol{\theta}) = \sum_{t=2}^{n} \frac{\left[\left(y_t-\rho y_{t-1}\right)^2-\left(\varpi_{t-1}^2\right)\right]^2}{4\left(\varpi_{t-1}^2\right)^4} + 2\sum_{\tau=1}^{m_n} w_{n\tau} \sum_{t=\tau+2}^{n} \frac{\left(y_t-\rho y_{t-1}\right)^2-\left(\varpi_{t-1}^2\right)}{2\left(\varpi_{t-1}^2\right)^2} \frac{\left(y_{t-\tau}-\rho y_{t-\tau-1}\right)^2-\left(\varpi_{t-\tau-1}^2\right)}{2\left(\varpi_{t-\tau-1}^2\right)^2}.$$

The outer product of the score matrix has been constructed in order to take into account the amount of serial dependence in $\{\varepsilon_t\}$. Under Assumption 2, the limiting distributions of the proposed test statistics change, since they depend on some unknown nuisance parameters, namely σ_ε^2, σ^2, κ^2, κ_ε^2 and $\upsilon = \mathrm{E}\left(\varepsilon_t^3\right)$. The following Lemma reports convergence results for the elements of the score, the Hessian and the outer product of the score matrix, under hypothesis (7). Convergence results under hypotheses (5) and (6) (and for the augmented model (11)) can be obtained using similar arguments and are omitted for space reasons.

Lemma 1. *Let Assumption 2 hold. Then in model (1) the following holds*

$$\frac{q_1\left(\widetilde{\boldsymbol{\theta}}\right)}{n} \xrightarrow{w} \frac{1}{\sigma_\varepsilon^2} \int_0^1 B_1(s)\, \mathrm{d}B_1(s);$$

Accepted manuscript

ACCEPTED MANUSCRIPT

$$\begin{aligned}\frac{q_2\left(\widetilde{\boldsymbol{\theta}}\right)}{n^{3/2}} &\xrightarrow{w} \frac{\kappa\sigma^2}{2\sigma_\varepsilon^4}\int_0^1\left(W_1(s)^2-\int_0^1 W_1(s)^2\,\mathrm{d}s\right)\mathrm{d}W^*(s)\\ &= \frac{\kappa\sigma^2}{2\sigma_\varepsilon^4}\left(\int_0^1 W_1(s)^2\,\mathrm{d}W^*(s)-\int_0^1 W_1(s)^2\,\mathrm{d}s\,W^*(1)\right),\end{aligned}$$

where $W^*(s)=\psi W_1(s)+\left(1-\psi^2\right)^{1/2}W_2(s)$, $W_1(s)$ *and* $W_2(s)$ *are independent standard Brownian motions;*

$$q_3\left(\widetilde{\boldsymbol{\theta}}\right)=0;\quad -\frac{\mathcal{H}_{11}\left(\widetilde{\boldsymbol{\theta}}\right)}{n^2}\xrightarrow{w}\frac{1}{\sigma_\varepsilon^2}\int_0^1 B_1(s)^2\,\mathrm{d}s=\frac{\sigma^2}{\sigma_\varepsilon^2}\int_0^1 W_1(s)^2\,\mathrm{d}s;$$

$$-\frac{\mathcal{H}_{12}\left(\widetilde{\boldsymbol{\theta}}\right)}{n^2}\xrightarrow{w}\frac{1}{\sigma_\varepsilon^4}\int_0^1 B_1(s)^3\,\mathrm{d}B_1(s);\quad -\frac{\mathcal{H}_{13}\left(\widetilde{\boldsymbol{\theta}}\right)}{n}\xrightarrow{w}\frac{1}{\sigma_\varepsilon^4}\int_0^1 B_1(s)\,\mathrm{d}B_1(s);$$

$$-\frac{\mathcal{H}_{22}\left(\widetilde{\boldsymbol{\theta}}\right)}{n^3}\xrightarrow{w}\frac{1}{2\sigma_\varepsilon^4}\int_0^1 B_1(s)^4\,\mathrm{d}s=\frac{\sigma^4}{2\sigma_\varepsilon^4}\int_0^1 W_1(s)^4\,\mathrm{d}s;$$

$$-\frac{\mathcal{H}_{23}\left(\widetilde{\boldsymbol{\theta}}\right)}{n^2}\xrightarrow{w}\frac{1}{2\sigma_\varepsilon^4}\int_0^1 B_1(s)^2\,\mathrm{d}s=\frac{\sigma^2}{2\sigma_\varepsilon^4}\int_0^1 W_1(s)^2\,\mathrm{d}s;$$

$$-\frac{\mathcal{H}_{33}\left(\widetilde{\boldsymbol{\theta}}\right)}{n}\xrightarrow{p}\frac{1}{2\sigma_\varepsilon^4};\quad \frac{\mathcal{A}_{11}\left(\widetilde{\boldsymbol{\theta}}\right)}{n^2}\xrightarrow{w}\frac{1}{\sigma_\varepsilon^2}\int_0^1 B_1(s)^2\,\mathrm{d}s=\frac{\sigma^2}{\sigma_\varepsilon^2}\int_0^1 W_1(s)^2\,\mathrm{d}s;$$

$$\frac{\mathcal{A}_{12}\left(\widetilde{\boldsymbol{\theta}}\right)}{n^{5/2}}\xrightarrow{w}\frac{\upsilon}{2\sigma_\varepsilon^6}\int_0^1 B_1(s)^3\,\mathrm{d}s;\quad \frac{\mathcal{A}_{22}\left(\widetilde{\boldsymbol{\theta}}\right)}{n^3}\xrightarrow{w}\frac{\kappa_\varepsilon^2}{4\sigma_\varepsilon^8}\int_0^1 B_1(s)^4\,\mathrm{d}s.$$

Then, the limiting distributions of the robust versions of the tests can be obtained in a straightforward way just applying the continuous mapping theorem.

The Lemma reveals some interesting results. First, the order of magnitude of $\mathcal{A}_{12}\left(\widetilde{\boldsymbol{\theta}}\right)$ differs from its corresponding element of the Hessian matrix. This is due to allowing for possible asymmetry in the marginal distribution of $\{\varepsilon_t\}$. Also, under (7), in the outer product of the score matrix, the cross terms (for capturing the serial dependence in $\{\varepsilon_t\}$) are of a smaller order then the contemporaneous elements, and the effect of allowing for some serial correlation in $\{\varepsilon_t\}$ comes from the long run variance parameter σ^2. The parameter ψ measures the limiting correlation between $\{\varepsilon_t\}$ and $\left\{\varepsilon_t^2-\sigma_\varepsilon^2\right\}$, and is zero when $\{\varepsilon_t\}$ is symmetric, as in the simulation study of Section 4.

From the limit results above it is immediate to see that, if the conditional innovation error is Gaussian, then $\boldsymbol{\mathcal{A}}(\boldsymbol{\theta})=-\boldsymbol{\mathcal{H}}(\boldsymbol{\theta})$ and robust tests simplify to those obtained in the previous section. As mentioned above, the limit theory of Lemma 1 depends on the unknown nuisance parameters σ_ε^2, σ^2, κ^2, κ_ε^2 and υ. In order to obtain a feasible limit theory, the following correction is proposed. As for the first component of the score, it is possible to use the nonparametric correction proposed by Phillips (1987) (see also Phillips and Perron, 1988) and consider the modified score

$$\frac{\widetilde{\sigma}^2}{\widehat{\sigma}^2}\frac{q_1\left(\widetilde{\boldsymbol{\theta}}\right)}{n}-\frac{1}{2}\frac{\widetilde{\sigma}^2-\widehat{\sigma}^2}{\widehat{\sigma}^2}\xrightarrow{w}\int_0^1 W_1(s)\,\mathrm{d}W_1(s),$$

Accepted manuscript

ACCEPTED MANUSCRIPT

where $\widehat{\sigma}^2$ is a consistent estimator of σ^2, which has a nuisance parameters free limiting distribution.

As for the second component, the modification is more straightforward. It suffices to use

$$\frac{\widetilde{\sigma}^4}{\widehat{\kappa}\widehat{\sigma}^2}\frac{q_2\left(\widetilde{\boldsymbol{\theta}}\right)}{n^{3/2}},$$

where $\widehat{\kappa}^2$ is any consistent estimator of and κ^2. Similar standardizations are possible for the components of the Hessian and outer product of the score matrix. Then one obtains nuisance parameters free limiting distributions, under the more stringent assumption that $\sup_t \mathrm{E}\left|\varepsilon_t\right|^{\gamma+\nu} < \infty$, for some $\gamma > 8$, $\nu > 0$, which ensures a consistent estimation of κ^2.

Derivation of an analytic formula for the exact densities of the proposed tests does not seem to be tractable; therefore in small samples the quantiles and power properties of the proposed test statistics have to be simulated. That is the content of the next section.

4 Simulation results

A Monte-Carlo experiment, based on 100,000 replications, has been conducted for different values of the parameters and sample size, assuming that $\alpha, \beta = 0$, $\varepsilon_t \sim \mathrm{IN}(0,1)$, $\rho_t \sim \mathrm{IN}\left(\rho, \omega^2\right)$, with $\{\varepsilon_t\}$, $\{\rho_t\}$ independent.

Tables 1 and 2 report the simulated quantiles of $\mathrm{ALM}^{(\omega^2)}$ and $\mathrm{ALM}_a^{(\omega^2)}$ under (5), for different values of ρ. For small to moderate sample sizes, critical values do not vary too much as ρ changes. Therefore, in practical work, using the critical value obtained for $\rho = 1$ would not produce a severe size distortion.

Tables 3 and 4 report quantiles of $\mathrm{ALM}^{(\rho)}$ and $\mathrm{ALM}_a^{(\rho)}$ under (6), for different values of ω^2. Starting from Table 3, one can see that, when $\omega^2 = 0$, the resulting critical values are virtually identical to those of the standard t-test for a unit root in linear models (see, e.g., Fuller, 1995). As ω^2 increases, critical values start to converge to those of a standard normal. Unreported simulations for larger (but unrealistic) values of ω^2 show an almost perfect correspondence to the standard normal quantiles. Again, for small to moderate sample sizes the critical values of the test when $\omega^2 = 0$ are similar to those when $\omega^2 \neq 0$. The same pattern can be observed for Table 4, with the warning remark that critical values of the test when $\omega^2 = 0$ cannot be compared with those of a t-test for a unit root in the linear model with a trend, but rather with those of the t-test in the (nonlinear) Bhargava model. Also, here the difference between critical values when $\omega^2 = 0$ and those when $\omega^2 \neq 0$ becomes more marked as n increases.

Finally, Table 5 reports the quantiles of AMLM and AMLM_a under (7). The simulated quantiles of AMLM appear to be quite stable for different values of the sample size; there is very little change (especially for the middle quantiles) even for substantially different values of n.

Accepted manuscript

ACCEPTED MANUSCRIPT

Table 6 reports the power of $\mathrm{ALM}^{(\omega^2)}$ and $\mathrm{ALM}_a^{(\omega^2)}$ for $\rho = 1$.[9] The results of the simulation exercise highlight the potential benefit of using sample information when constructing test statistics. In fact, $\mathrm{ALM}^{(\omega^2)}$ and $\mathrm{ALM}_a^{(\omega^2)}$ dominate almost everywhere the competing tests Z_T and $\widehat{H}_T$ proposed respectively by McCabe and Tremayne (1995) and Leybourne et al. (1996) (whose power figures are reported in columns 5 and 6), based on the deterministic normalization factor $n^{3/2}$. Being a locally best invariant test, Z_T does better than $\mathrm{ALM}^{(\omega^2)}$ in the vicinity of the null hypothesis and for small to moderate sample sizes (for $\omega^2 = 0.001$, 0.072 vs. 0.058 when $n = 50$; 0.107 vs. 0.073 when $n = 100$; 0.207 vs. 0.157 when $n = 200$), but then it is clearly dominated, the difference in power being sometimes considerable (e.g. when $\omega^2 = 0.1$ and $n = 100$ the power of Z_T is 0.43 and that of $\mathrm{ALM}^{(\omega^2)}$ is 0.976). $\mathrm{ALM}_a^{(\omega^2)}$ seems to dominate everywhere $\widehat{H}_T$, but the differences in power are less marked. Also, the Monte-Carlo experiment reveals the extent of the power gains obtained by selecting only the components of the score vector corresponding to the parameter of interest and by using the Hessian matrix. In fact, just to cite a few examples, if one constructs the test statistic using the full score vector and the information matrix, then he gets a power of 0.084 for $\omega^2 = .1$ and $n = 25$ (compared with 0.429 of $\mathrm{ALM}^{(\omega^2)}$); a power of 0.189 for $\omega^2 = .1$ and $n = 50$ (compared with 0.787 of $\mathrm{ALM}^{(\omega^2)}$); finally, a power of 0.440 for $\omega^2 = .05$ and $n = 100$ (compared with 0.910 of $\mathrm{ALM}^{(\omega^2)}$). Of course, in finite samples, the Hessian matrix is not guaranteed to be positive definite, and therefore there might be problems when trying to invert it. But this case never occurred in the simulations and in the empirical example of the next Section.

Table 7 reports the size distortions of $\mathrm{ALM}^{(\omega^2)}$ induced by conditional heteroskedasticity; data are generated by a unit root process with GARCH(1,1) innovations, through

$$y_t = y_{t-1} + e_t, \quad e_t \sim \mathrm{N}(0, h_t),$$

where

$$h_t = (1 - \vartheta - \varphi) + \varphi h_{t-1} + \vartheta e_{t-1}^2,$$

with $\vartheta,\ \varphi > 0$ and $\vartheta + \varphi < 1$. The unconditional variance is, therefore, always equal to one. In the exercise, three cases are considered: (i) almost no GARCH: $\vartheta + \varphi = .1$; (ii) moderate GARCH: $\vartheta + \varphi = .5$; (iii) strong GARCH: $\vartheta + \varphi = .9$. In case (i), empirical sizes are very close to correct ones; there is a small distortion in case (ii) and seems to be decreasing in φ. Finally, in case (iii) size distortions appear to be more pronounced. This seems to confirm the discussion of Section 2 about the similarities between random coefficient and conditional heteroskedastic models.

Table 8 reports the power of $\mathrm{ALM}_a^{(\rho)}$ for $\omega^2 = 0$. Figures relative to $\mathrm{ALM}^{(\rho)}$ are equal to those of the standard t-test for a unit root in simple autoregressive models and have been tabulated already by Dickey

[9] With a slight abuse of notation, the subscript $_{(\infty)}$ denotes tests whose powers are based on asymptotic critical values. Also, all the powers are calculated at the 5% level.

Accepted manuscript

ACCEPTED MANUSCRIPT

and Fuller (1979).

Tables 9 and 10 contain the powers of of AMLM and AMLM$_a$, for different values of the sample size and for values of the parameters of interest typically used in similar work in the literature (e.g. Dickey and Fuller, 1979; Leybourne et al., 1996). Analysis of the power figures reveals some interesting properties of the considered tests.

First, the tests are not biased, in contrast to some evidence to the contrary reported in the literature on unit root testing (e.g. Dickey and Fuller, 1979). The justification for the presence of bias, and more generally for the poor power figures against alternative values of ρ very close to one, is the stochastic discontinuity between the null and alternative hypothesis.

The power of AMLM and AMLM$_a$ increases when violation of the null occurs in both directions (namely ρ and ω^2), but not uniformly. In fact, for n greater than 50, the power is higher when violation of the null hypothesis is on the ω^2 component with $\rho = 1$, than in cases when ρ is close to one. This may be due to the stochastic discontinuity between the null and alternative hypothesis described above.

Interestingly, for alternatives quite distant from the null hypothesis, AMLM and AMLM$_a$ reach higher power figures than those of Z_T and $\widehat{H}_T$, despite the fact that the tests used here are inefficient, since they do not make use of the information that $\rho = 1$.

If one knew that the DGP is a simple asymptotically stationary AR(1), the powers of AMLM and AMLM$_a$ display very similar values respectively to the Dickey-Fuller type t-test and to the t-test in Bhargava's model, although the former tests are designed for an (inefficient) joint restriction. Apart from small difference due to rounding, there are no substantial discrepancies between the two pair of tests. This is an indication of the gain (in terms of power) provided by the modification to the conventional LM statistic; the proposed test recovers the loss of power involved in using a two-sided test, when in fact the direction of the alternative hypothesis is one-sided.

Again, just to cite a few examples, if one constructs the test statistic using the full score vector and the information matrix, then he gets a power of 0.120 for $\omega^2 = .01$, $\rho = .8$ and $n = 25$ (compared with 0.323 of AMLM); a power of 0.283 for $\omega^2 = .1$, $\rho = .8$ and $n = 50$ (compared with 0.886 of AMLM); finally, a power of 0.333 for $\omega^2 = .05$, $\rho = .99$ and $n = 100$ (compared with 0.869 of AMLM).

It is worth noting the potential power gains provided by the present testing procedure even for very small values taken by ω^2. For example for the AMLM test, when $n = 200$ and $\rho = 0.99$, it suffices a value of ω^2 equal to 0.01 to get a satisfactory power value (0.660), compared to the low value of 0.130 when ω^2 takes the value of zero. Given the evidence reported by Leybourne et al. (1996), who estimated a large sample of macroeconomic time series frequently used in applied work and found that the majority of them had a value of $\widehat{\omega}^2$ between 0.01 and 0.1, one can expect some power gains in using the suggested testing procedure.

Accepted manuscript

ACCEPTED MANUSCRIPT

The powers of tests based on asymptotic critical values, reported in Tables 6 and 8, are generally higher than those based on exact critical values, since the distribution of the tests tend to shrink when n increases. However, in several cases using asymptotic critical values induces severe size distortions. This is not the case for AMLM, which remains fairly constant as the sample size changes. The power results of AMLM and AMLM_a based on asymptotic critical values are not reported for space reasons and are available upon request.

Since the asymptotic distributions of the proposed tests do not have a close form, in order to compute approximate P values for these tests one can follow the approach of MacKinnon (1994).

5 Empirical Example

The tests derived in Section 3 have been applied to the quarterly series of US GDP. Data refer to seasonally adjusted quarterly GDP values, expressed in billions of dollars at 2000 prices. The series is available from the first quarter of 1947 to the fourth quarter of 2003, for a total of 228 observations. The obtained series has been transformed taking logarithms and then tests for hypotheses (5), (6) and (7) have been performed.

The plot of the series is reported in Figure 1 and suggest that the suitable version of the tests to use are those with deterministic components. In particular, in order to allow for non-normality, conditional

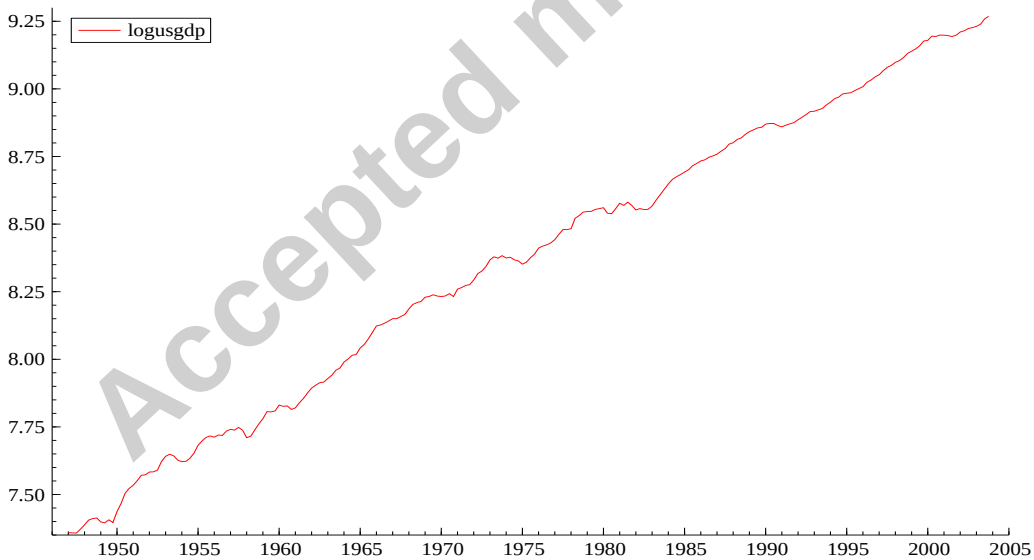

Figure 1: Actual values of US log(GDP)

heteroskedasticity and serial dependence in $\{\varepsilon_t\}$, the robust versions of $\text{ALM}_a^{(\omega^2)}$, $\text{ALM}_a^{(\rho)}$ and and AMLM_a have been calculated. Furthermore, the tests the tests have been modified according to the discussion in Section 3, to remove the effect of the unknown nuisance parameters.

Accepted manuscript

ACCEPTED MANUSCRIPT

The parameters of the conditional mean of the model have been estimated using the method described in Section 3. As for the parameters entering the conditional variance, they have been estimated according to equation (5) in Amemiya (1977). When testing hypotheses (5) and (7), σ_ε^4 has been estimated by the robust method proposed in Cai et al. (1998). Point estimates for the parameters of interest are given below

$$\widehat{\alpha} = 7.352, \quad \widehat{\beta} = 0.0085, \quad \widehat{\rho} = 0.982, \quad \widehat{\omega}^2 = 0.0005.$$

The obtained values of the statistics are

$$\mathrm{ALMR}_a^{(\omega^2)} = 1.89\mathrm{e} - 10, \quad \mathrm{ALMR}_a^{(\rho)} = -1.048, \quad \mathrm{ALMR}_a = 1.194.$$

In order to make an asymptotically valid inference, the subsampling method of Romano and Wolf (2001) has been used, with the block size determined through their data dependent procedure for each (1%, 5% and 10%) considered level of confidence. To comply with the assumptions of Romano and Wolf (2001), serial dependence in $\{\varepsilon_t\}$ has been accounted for parametrically, using augmented regressions containing lagged values of first differences of data. Finally, σ^2 and κ^2 have been estimated using, respectively, an heteroskedasticity-consistent and an heteroskedasticity and autocorrelation-consistent estimator with 8 lags. The null hypotheses are never rejected.

Therefore, the outcome of the testing procedure is that the hypothesis of US log(GDP) evolving as a unit root process with a drift is consistent with data, at least when the null hypothesis is directed against short memory trending processes.

6 Concluding remarks

This paper has proposed new tests for unit root processes with a possibly nonzero drift, in the context of a random coefficient autoregressive model. The novelty of the approach lies in the fact that here the mean of the random coefficient variable is not fixed to unity, as in previous papers, but it is left unknown. Then joint and single restrictions on the mean and variance of the random coefficient are tested. The limiting distributions of the tests are derived and their properties are studied through the results of a Monte-Carlo experiment.

The proposed tests have good power properties and in several cases perform much better than the univariate tests available in the literature. This highlights the potential benefit of using sample information in the construction of tests statistics.

Finally, the proposed testing procedure has been applied to the US GDP quarterly series.

Accepted manuscript

ACCEPTED MANUSCRIPT

Table 1: Simulated quantiles of $\mathrm{ALM}^{(\omega^2)}$ under hypothesis (5) for different n and ρ

ρ	1%	2.5%	5%	10%	50%	90%	95%	97.5%	99%
	$n = 25$								
.8	1e-4	0.001	0.003	0.010	0.260	1.158	1.558	1.989	2.969
.9	8e-5	0.001	0.002	0.010	0.259	1.193	1.616	2.075	2.969
.95	8e-5	0.001	0.002	0.009	0.242	1.168	1.604	2.035	2.771
.99	1e-4	0.001	0.003	0.009	0.253	1.181	1.623	2.060	2.831
1	8e-5	0.001	0.002	0.009	0.251	1.180	1.620	2.143	2.921
	$n = 50$								
.8	9e-5	0.001	0.002	0.009	0.258	1.286	1.844	2.361	3.534
.9	8e-5	0.001	0.002	0.010	0.268	1.332	1.840	2.433	3.727
.95	8e-5	0.001	0.002	0.009	0.258	1.311	1.816	2.416	3.594
.99	6e-5	4e-4	0.002	0.008	0.244	1.300	1.830	2.506	3.597
1	9e-5	0.001	0.002	0.008	0.241	1.270	1.759	2.338	3.280
	$n = 100$								
.8	1e-4	0.001	0.003	0.010	0.277	1.456	2.046	2.746	3.863
.9	8e-5	0.001	0.002	0.010	0.272	1.465	2.092	2.836	3.973
.95	1e-4	0.001	0.002	0.010	0.261	1.391	1.973	2.584	3.657
.99	7e-5	4e-4	0.002	0.008	0.229	1.314	1.877	2.424	3.339
1	6e-5	4e-4	0.002	0.008	0.230	1.336	1.903	2.565	3.531
	$n = 200$								
.8	1e-4	0.001	0.003	0.011	0.275	1.550	2.245	2.940	4.036
.9	9e-5	0.001	0.003	0.010	0.281	1.549	2.155	2.932	4.224
.95	1e-4	0.001	0.002	0.009	0.270	1.560	2.199	2.902	4.101
.99	8e-5	5e-4	0.002	0.009	0.238	1.384	2.013	2.663	3.791
1	1e-4	0.001	0.002	0.008	0.229	1.332	1.894	2.505	3.384
	$n = 500$								
.8	1e-4	0.001	0.003	0.012	0.291	1.718	2.474	3.244	4.336
.9	1e-4	0.001	0.003	0.011	0.283	1.619	2.302	3.054	4.510
.95	1e-4	0.001	0.003	0.010	0.292	1.646	2.299	3.091	4.340
.99	9e-5	0.001	0.002	0.010	0.265	1.531	2.221	2.854	3.870
1	5e-5	5e-4	0.002	0.008	0.221	1.378	1.964	2.632	3.360
	$n = 1000$								
.8	7e-5	0.001	0.003	0.011	0.298	1.717	2.426	3.257	4.475
.9	1e-4	0.001	0.003	0.011	0.297	1.729	2.411	3.118	4.226
.95	1e-4	0.001	0.003	0.010	0.289	1.724	2.482	3.272	4.460
.99	9e-5	0.001	0.002	0.010	0.286	1.648	2.285	3.130	4.172
1	8e-5	5e-4	0.002	0.008	0.212	1.367	1.964	2.618	3.634

Notes: Results are based on 100,000 replications using Gauss.
The simulated model is

$$y_t = \rho y_{t-1} + \varepsilon_t, \quad t = 1, 2, \ldots, n.$$

Accepted manuscript

ACCEPTED MANUSCRIPT

Table 2: Simulated quantiles of $\mathrm{ALM}_a^{(\omega^2)}$ under hypothesis (5) for different n and ρ

ρ	1%	2.5%	5%	10%	50%	90%	95%	97.5%	99%
					$n = 25$				
.8	8e-5	0.001	0.001	0.008	0.252	1.128	1.502	1.836	2.549
.9	9e-5	0.001	0.002	0.009	0.253	1.154	1.526	1.950	2.781
.95	1e-4	0.001	0.002	0.010	0.258	1.142	1.503	1.898	2.697
.99	9e-5	0.001	0.002	0.010	0.257	1.145	1.521	1.914	2.676
1	1e-4	0.001	0.002	0.010	0.253	1.137	1.507	1.942	2.759
					$n = 50$				
.8	1e-4	0.001	0.002	0.009	0.273	1.333	1.798	2.391	3.256
.9	8e-5	0.001	0.002	0.009	0.262	1.328	1.790	2.364	3.387
.95	1e-4	0.001	0.002	0.011	0.266	1.318	1.830	2.408	3.509
.99	1e-4	0.001	0.002	0.009	0.251	1.305	1.781	2.369	3.405
1	9e-5	0.001	0.001	0.008	0.262	1.317	1.800	2.362	3.313
					$n = 100$				
.8	6e-5	0.001	0.002	0.009	0.271	1.457	2.049	2.661	3.823
.9	1e-4	0.001	0.002	0.010	0.281	1.502	2.069	2.705	3.897
.95	9e-5	0.001	0.002	0.009	0.266	1.420	1.998	2.644	3.854
.99	8e-5	0.001	0.002	0.009	0.262	1.409	2.024	2.715	3.955
1	8e-5	0.001	0.002	0.010	0.268	1.383	1.936	2.602	3.512
					$n = 200$				
.8	1e-4	0.001	0.002	0.011	0.292	1.555	2.209	2.920	4.019
.9	9e-5	0.001	0.002	0.009	0.278	1.583	2.233	2.962	4.108
.95	1e-4	0.001	0.002	0.009	0.271	1.554	2.192	2.942	4.252
.99	9e-5	0.001	0.002	0.008	0.255	1.504	2.158	2.834	3.864
1	1e-4	0.001	0.002	0.009	0.262	1.510	2.143	2.824	3.931
					$n = 500$				
.8	1e-4	0.001	0.002	0.009	0.284	1.680	2.424	3.117	4.262
.9	1e-4	0.001	0.002	0.009	0.286	1.702	2.388	3.153	4.250
.95	1e-4	0.001	0.002	0.009	0.277	1.588	2.329	3.057	4.101
.99	8e-5	0.001	0.002	0.009	0.279	1.627	2.284	3.007	4.033
1	8e-5	0.001	0.002	0.009	0.256	1.537	2.199	2.827	3.804
					$n = 1000$				
.8	1e-4	0.001	0.002	0.010	0.293	1.693	2.381	3.131	4.151
.9	1e-4	0.001	0.002	0.009	0.288	1.719	2.488	3.305	4.628
.95	1e-4	0.001	0.003	0.011	0.288	1.709	2.484	3.202	4.271
.99	9e-5	0.001	0.002	0.009	0.279	1.653	2.355	3.081	4.128
1	1e-4	0.001	0.002	0.009	0.271	1.579	2.251	2.894	3.774

Notes: Results are based on 100,000 replications using Gauss.
The simulated model is

$$y_t = \rho y_{t-1} + \varepsilon_t, \quad t = 1, 2, \ldots, n.$$

Accepted manuscript

ACCEPTED MANUSCRIPT

Table 3: Simulated quantiles of ALM$^{(\rho)}$ under hypothesis (6) for different n and ω^2

ω^2	1%	2.5%	5%	10%	50%	90%	95%	97.5%	99%
	$n = 25$								
0	-2.479	-2.150	-1.891	-1.560	-0.478	0.903	1.321	1.620	2.004
.001	-2.505	-2.144	-1.870	-1.553	-0.452	0.910	1.317	1.667	2.080
.005	-2.434	-2.102	-1.853	-1.530	-0.447	0.933	1.336	1.665	2.047
.01	-2.446	-2.129	-1.819	-1.531	-0.443	0.935	1.357	1.707	2.128
.05	-2.383	-2.066	-1.791	-1.473	-0.345	1.095	1.493	1.867	2.239
.1	-2.494	-2.120	-1.803	-1.457	-0.275	1.115	1.535	1.896	2.251
	$n = 50$								
0	-2.505	-2.166	-1.876	-1.561	-0.472	0.870	1.233	1.576	1.994
.001	-2.513	-2.169	-1.895	-1.596	-0.476	0.905	1.323	1.667	2.038
.005	-2.475	-2.194	-1.911	-1.586	-0.452	0.926	1.352	1.667	2.139
.01	-2.494	-2.147	-1.880	-1.566	-0.406	0.989	1.364	1.742	2.123
.05	-2.460	-2.145	-1.859	-1.507	-0.284	1.119	1.488	1.806	2.222
.1	-2.449	-2.107	-1.794	-1.449	-0.234	1.139	1.527	1.857	2.290
	$n = 100$								
0	-2.536	-2.203	-1.901	-1.587	-0.480	0.892	1.261	1.592	1.952
.001	-2.464	-2.162	-1.885	-1.573	-0.453	0.944	1.334	1.651	2.115
.005	-2.472	-2.170	-1.881	-1.542	-0.421	0.977	1.409	1.740	2.145
.01	-2.555	-2.204	-1.918	-1.577	-0.385	1.053	1.468	1.842	2.277
.05	-2.469	-2.111	-1.808	-1.460	-0.238	1.142	1.554	1.931	2.307
.1	-2.469	-2.102	-1.766	-1.415	-0.162	1.151	1.530	1.900	2.236
	$n = 200$								
0	-2.510	-2.216	-1.949	-1.616	-0.507	0.878	1.245	1.588	2.029
.001	-2.539	-2.149	-1.893	-1.582	-0.452	0.971	1.382	1.701	2.049
.005	-2.530	-2.201	-1.923	-1.567	-0.382	1.063	1.496	1.826	2.175
.01	-2.553	-2.183	-1.900	-1.581	-0.373	1.075	1.493	1.833	2.228
.05	-2.534	-2.118	-1.810	-1.454	-0.180	1.137	1.556	1.886	2.210
.1	-2.349	-2.036	-1.737	-1.395	-0.128	1.162	1.554	1.859	2.275
	$n = 500$								
0	-2.563	-2.217	-1.905	-1.596	-0.496	0.916	1.302	1.645	1.991
.001	-2.588	-2.247	-1.915	-1.592	-0.430	1.016	1.406	1.759	2.154
.005	-2.534	-2.190	-1.888	-1.536	-0.313	1.103	1.528	1.920	2.280
.01	-2.509	-2.149	-1.858	-1.501	-0.229	1.127	1.497	1.853	2.194
.05	-2.430	-2.073	-1.773	-1.426	-0.139	1.141	1.506	1.825	2.185
.1	-2.386	-2.016	-1.710	-1.359	-0.069	1.176	1.550	1.849	2.246
	$n = 1000$								
0	-2.641	-2.290	-1.965	-1.621	-0.512	0.867	1.301	1.649	2.042
.001	-2.479	-2.165	-1.898	-1.586	-0.420	1.027	1.461	1.815	2.190
.005	-2.478	-2.163	-1.860	-1.487	-0.249	1.097	1.504	1.825	2.183
.01	-2.543	-2.158	-1.811	-1.460	-0.184	1.141	1.494	1.854	2.239
.05	-2.404	-2.044	-1.738	-1.393	-0.091	1.191	1.547	1.860	2.223
.1	-2.398	-2.035	-1.685	-1.323	-0.040	1.207	1.579	1.924	2.279

Notes: Results are based on 100,000 replications using Gauss.
The simulated model is

$$y_t = \rho_t y_{t-1} + \varepsilon_t, \quad \rho_t \sim \mathrm{IN}(1, \omega^2), \quad t = 1, 2, \ldots, n.$$

Accepted manuscript

ACCEPTED MANUSCRIPT

Table 4: Simulated quantiles of $\mathrm{ALM}_a^{(\rho)}$ under hypothesis (6) for different n and ω^2

ω^2	1%	2.5%	5%	10%	50%	90%	95%	97.5%	99%
					$n = 25$				
0	-3.518	-3.225	-2.972	-2.689	-1.677	-0.890	-0.753	-0.657	-0.556
.001	-3.542	-3.257	-3.015	-2.702	-1.697	-0.899	-0.751	-0.657	-0.556
.005	-3.605	-3.296	-3.067	-2.749	-1.731	-0.893	-0.742	-0.643	-0.519
.01	-3.713	-3.384	-3.138	-2.824	-1.730	-0.863	-0.701	-0.590	-0.484
.05	-4.638	-3.927	-3.459	-3.019	-1.718	-0.670	-0.450	-0.269	0.155
.1	-4.652	-3.997	-3.471	-3.006	-1.607	-0.428	-0.139	0.250	1.628
					$n = 50$				
0	-3.672	-3.338	-3.046	-2.715	-1.666	-0.925	-0.777	-0.679	-0.587
.001	-3.654	-3.341	-3.040	-2.746	-1.674	-0.917	-0.761	-0.651	-0.544
.005	-3.841	-3.428	-3.126	-2.793	-1.700	-0.882	-0.711	-0.602	-0.486
.01	-4.077	-3.668	-3.284	-2.919	-1.706	-0.820	-0.642	-0.520	-0.381
.05	-4.920	-4.062	-3.520	-3.002	-1.525	-0.367	-0.086	0.270	1.454
.1	-4.641	-3.774	-3.291	-2.780	-1.280	0.030	0.491	1.111	1.674
					$n = 100$				
0	-3.568	-3.242	-2.961	-2.643	-1.626	-0.900	-0.772	-0.666	-0.577
.001	-3.633	-3.316	-3.037	-2.713	-1.667	-0.894	-0.741	-0.646	-0.546
.005	-4.080	-3.621	-3.241	-2.863	-1.686	-0.825	-0.645	-0.523	-0.386
.01	-4.568	-3.885	-3.435	-2.979	-1.682	-0.701	-0.488	-0.327	-0.132
.05	-4.979	-3.960	-3.370	-2.788	-1.218	0.033	0.463	1.032	1.572
.1	-4.091	-3.424	-2.928	-2.469	-0.960	0.453	0.927	1.311	1.867
					$n = 200$				
0	-3.566	-3.206	-2.881	-2.585	-1.633	-0.921	-0.785	-0.697	-0.609
.001	-3.678	-3.333	-3.040	-2.709	-1.663	-0.896	-0.731	-0.624	-0.509
.005	-4.454	-3.809	-3.374	-2.896	-1.618	-0.681	-0.467	-0.282	-0.077
.01	-4.987	-4.031	-3.443	-2.927	-1.496	-0.438	-0.174	0.057	0.701
.05	-4.275	-3.489	-2.934	-2.420	-0.907	0.501	0.985	1.368	2.121
.1	-3.578	-3.010	-2.601	-2.159	-0.692	0.685	1.101	1.457	2.143
					$n = 500$				
0	-3.496	-3.179	-2.900	-2.605	-1.640	-0.915	-0.776	-0.681	-0.598
.001	-3.953	-3.505	-3.170	-2.767	-1.659	-0.820	-0.647	-0.509	-0.362
.005	-4.984	-4.004	-3.431	-2.869	-1.455	-0.351	-0.071	0.265	1.370
.01	-4.586	-3.737	-3.189	-2.619	-1.147	0.120	0.600	1.223	1.815
.05	-3.358	-2.767	-2.383	-1.966	-0.582	0.826	1.228	1.650	2.225
.1	-3.019	-2.532	-2.161	-1.793	-0.444	0.870	1.223	1.552	2.092
					$n = 1000$				
0	-3.400	-3.131	-2.852	-2.564	-1.622	-0.929	-0.793	-0.693	-0.594
.001	-4.407	-3.710	-3.237	-2.807	-1.613	-0.683	-0.474	-0.315	-0.134
.005	-4.438	-3.666	-3.068	-2.610	-1.184	0.090	0.535	1.107	1.607
.01	-4.163	-3.289	-2.781	-2.308	-0.860	0.549	1.047	1.504	2.239
.05	-3.040	-2.529	-2.153	-1.757	-0.392	0.899	1.274	1.631	2.149
.1	-2.819	-2.369	-2.002	-1.633	-0.304	0.896	1.224	1.516	1.915

Notes: Results are based on 100,000 replications using Gauss.
The simulated model is

$$y_t = \rho_t y_{t-1} + \varepsilon_t, \quad \rho_t \sim \mathrm{IN}(1, \omega^2), \quad t = 1, 2, \ldots, n.$$

Accepted manuscript

ACCEPTED MANUSCRIPT

Table 5: Simulated quantiles of AMLM and AMLM$_a$ under hypothesis (7) for different n

n	1%	2.5%	5%	10%	50%	90%	95%	97.5%	99%
	AMLM								
25	0.001	0.004	0.015	0.054	0.769	3.227	4.564	6.147	8.524
50	0.001	0.003	0.015	0.053	0.796	3.400	4.725	6.222	8.454
100	0.001	0.004	0.014	0.051	0.805	3.459	4.769	6.196	8.224
200	0.001	0.003	0.013	0.049	0.805	3.470	4.749	6.107	7.931
500	0.001	0.003	0.014	0.051	0.818	3.426	4.681	5.975	7.719
1000	0.001	0.003	0.014	0.049	0.815	3.485	4.755	6.004	7.674
	AMLM$_a$								
25	0.532	0.697	0.898	1.223	3.741	9.364	11.742	14.392	18.507
50	0.517	0.677	0.872	1.187	3.479	8.932	11.299	13.744	17.460
100	0.504	0.669	0.867	1.187	3.355	8.376	10.565	12.886	15.924
200	0.511	0.672	0.875	1.186	3.303	7.887	9.826	11.869	14.540
500	0.503	0.672	0.877	1.191	3.289	7.634	9.337	11.086	13.448
1000	0.494	0.666	0.873	1.181	3.263	7.490	9.205	10.847	13.006

Notes: Results are based on 100,000 replications using Gauss.
The simulated model is

$$y_t = y_{t-1} + \varepsilon_t, \quad t = 1, 2, \ldots, n.$$

Accepted manuscript

ACCEPTED MANUSCRIPT

Table 6: Simulated power values of $\mathrm{ALM}^{(\omega^2)}$, $\mathrm{ALM}^{(\omega^2)}_{(\infty)}$, $\mathrm{ALM}^{(\omega^2)}_{a}$, $\mathrm{ALM}^{(\omega^2)}_{a,(\infty)}$, Z_T and $\widehat{H}_T$ for $\rho = 1$, different values of ω^2 and n

ω^2	$\mathrm{ALM}^{(\omega^2)}$	$\mathrm{ALM}^{(\omega^2)}_{(\infty)}$	$\mathrm{ALM}^{(\omega^2)}_{a}$	$\mathrm{ALM}^{(\omega^2)}_{a,(\infty)}$	Z_T	$\widehat{H}_T$
			$n = 25$			
0	0.051	0.349	0.051	0.140	0.052	0.050
.001	0.055	0.358	0.051	0.145	0.057	0.049
.005	0.063	0.377	0.061	0.149	0.084	0.051
.01	0.084	0.392	0.067	0.167	0.115	0.057
.05	0.265	0.619	0.155	0.309	0.251	0.131
.1	0.429	0.778	0.252	0.411	0.300	0.231
			$n = 50$			
0	0.050	0.356	0.049	0.102	0.052	0.051
.001	0.058	0.366	0.053	0.111	0.072	0.052
.005	0.109	0.423	0.072	0.143	0.128	0.062
.01	0.183	0.522	0.106	0.191	0.197	0.086
.05	0.603	0.895	0.373	0.490	0.343	0.341
.1	0.787	0.967	0.565	0.676	0.362	0.538
			$n = 100$			
0	0.050	0.368	0.050	0.076	0.049	0.050
.001	0.073	0.383	0.064	0.091	0.108	0.063
.005	0.279	0.612	0.147	0.195	0.233	0.132
.01	0.482	0.799	0.281	0.333	0.354	0.213
.05	0.910	0.996	0.754	0.783	0.475	0.587
.1	0.979	0.999	0.906	0.921	0.424	0.737
			$n = 200$			
0	0.050	0.359	0.050	0.065	0.048	0.052
.001	0.157	0.512	0.102	0.116	0.215	0.101
.005	0.618	0.905	0.377	0.406	0.422	0.295
.01	0.885	0.982	0.599	0.624	0.508	0.454
.05	0.999	1.000	0.959	0.961	0.580	0.819
.1	1.000	1.000	0.989	0.987	0.497	0.908
			$n = 500$			
0	0.050	0.363	0.050	0.056	0.053	0.051
.001	0.556	0.856	0.329	0.331	0.428	0.232
.005	0.968	1.000	0.810	0.815	0.643	0.641
.01	1.000	1.000	0.940	0.945	0.711	0.811
.05	1.000	1.000	0.997	0.998	0.679	0.969
.1	1.000	1.000	1.000	1.000	0.495	0.908
			$n = 1000$			
0	0.051	0.050	0.050	0.050	0.051	0.049
.001	0.876	0.876	0.637	0.637	0.612	0.450
.005	1.000	1.000	0.967	0.967	0.813	0.912
.01	1.000	1.000	0.989	0.989	0.853	0.986
.05	1.000	1.000	1.000	1.000	0.745	1.000
.1	1.000	1.000	1.000	1.000	0.568	1.000

Notes: Results are based on 100,000 replications using Gauss.
The simulated model is

$$y_t = \rho_t y_{t-1} + \varepsilon_t, \quad \rho_t \sim \mathrm{IN}(1, \omega^2), \quad t = 1, 2, \ldots, n.$$

Accepted manuscript

ACCEPTED MANUSCRIPT

Table 7: Empirical size of $\mathrm{ALM}^{(\omega^2)}$ against a GARCH(1,1) for $\rho = 1$, different values of the parameters and n

	$\vartheta = 0.1 - \varphi$			$\vartheta = 0.5 - \varphi$			$\vartheta = 0.9 - \varphi$		
n	$\varphi = 0.01$	$\varphi = 0.05$	$\varphi = 0.09$	$\varphi = 0.1$	$\varphi = 0.25$	$\varphi = 0.4$	$\varphi = 0.1$	$\varphi = 0.45$	$\varphi = 0.8$
25	0.064	0.055	0.050	0.108	0.092	0.069	0.163	0.139	0.070
50	0.062	0.055	0.052	0.120	0.106	0.070	0.196	0.206	0.099
100	0.071	0.064	0.055	0.140	0.123	0.081	0.227	0.273	0.147
200	0.068	0.052	0.047	0.153	0.125	0.075	0.278	0.313	0.189
500	0.068	0.062	0.051	0.167	0.142	0.088	0.284	0.353	0.236
1000	0.067	0.063	0.055	0.186	0.149	0.092	0.305	0.389	0.266

Notes: Results are based on 100,000 replications using Gauss.
The simulated model is

$$y_t = y_{t-1} + e_t, \quad e_t \sim \mathrm{N}(0, h_t), \quad h_t = (1 - \vartheta - \varphi) + \varphi h_{t-1} + \vartheta e_{t-1}^2.$$

Table 8: Simulated power values of $\mathrm{ALM}_a^{(\rho)}$ and $\mathrm{ALM}_{a,(\infty)}^{(\rho)}$ for $\omega^2 = 0$, different values of ρ and n

ρ	$\mathrm{ALM}_a^{(\rho)}$	$\mathrm{ALM}_{a,(\infty)}^{(\rho)}$	$\mathrm{ALM}_a^{(\rho)}$	$\mathrm{ALM}_{a,(\infty)}^{(\rho)}$	$\mathrm{ALM}_a^{(\rho)}$	$\mathrm{ALM}_{a,(\infty)}^{(\rho)}$
	$n = 25$		$n = 50$		$n = 100$	
.8	0.117	0.166	0.284	0.399	0.832	0.880
.9	0.080	0.110	0.107	0.172	0.303	0.383
.95	0.070	0.089	0.072	0.109	0.122	0.163
.99	0.056	0.074	0.051	0.075	0.054	0.071
1	0.052	0.069	0.047	0.075	0.051	0.064
	$n = 200$		$n = 500$		$n = 1000$	
.8	1.000	1.000	1.000	1.000	1.000	1.000
.9	0.835	0.876	1.000	1.000	1.000	1.000
.95	0.318	0.353	0.956	0.963	1.000	1.000
.99	0.068	0.073	0.119	0.126	0.327	0.327
1	0.056	0.060	0.048	0.056	0.051	0.051

Notes: Results are based on 100,000 replications using Gauss.
The simulated model is

$$y_t = \rho y_{t-1} + \varepsilon_t, \quad t = 1, 2, \ldots, n.$$

Accepted manuscript

ACCEPTED MANUSCRIPT

Table 9: Simulated power values of AMLM for different values of the parameters and n

ρ	$\omega^2 = 0$	$\omega^2 = 0.001$	$\omega^2 = 0.005$	$\omega^2 = 0.01$	$\omega^2 = 0.05$	$\omega^2 = 0.1$
			$n = 25$			
.8	0.306	0.311	0.315	0.323	0.389	0.474
.9	0.145	0.156	0.155	0.158	0.232	0.336
.95	0.093	0.092	0.095	0.099	0.173	0.291
.99	0.055	0.055	0.059	0.066	0.141	0.261
1	0.049	0.050	0.055	0.061	0.137	0.255
			$n = 50$			
.8	0.747	0.745	0.751	0.760	0.820	0.886
.9	0.315	0.317	0.331	0.353	0.549	0.747
.95	0.151	0.153	0.166	0.186	0.436	0.694
.99	0.064	0.066	0.083	0.094	0.400	0.676
1	0.051	0.051	0.061	0.087	0.399	0.675
			$n = 100$			
.8	0.997	0.997	0.996	0.996	0.996	0.997
.9	0.745	0.748	0.765	0.786	0.947	0.996
.95	0.318	0.323	0.361	0.420	0.881	0.988
.99	0.078	0.091	0.131	0.233	0.869	0.986
1	0.050	0.053	0.119	0.243	0.878	0.987
			$n = 200$			
.8	1.000	1.000	1.000	1.000	1.000	1.000
.9	0.997	0.997	0.997	0.997	1.000	1.000
.95	0.753	0.758	0.809	0.882	0.998	1.000
.99	0.130	0.142	0.342	0.660	0.998	1.000
1	0.051	0.070	0.395	0.709	0.998	1.000
			$n = 500$			
.8	1.000	1.000	1.000	1.000	1.000	1.000
.9	1.000	1.000	1.000	1.000	1.000	1.000
.95	1.000	1.000	1.000	1.000	1.000	1.000
.99	0.319	0.399	0.944	1.000	1.000	1.000
1	0.050	0.329	0.964	1.000	1.000	1.000
			$n = 1000$			
.8	1.000	1.000	1.000	1.000	1.000	1.000
.9	1.000	1.000	1.000	1.000	1.000	1.000
.95	1.000	1.000	1.000	1.000	1.000	1.000
.99	0.755	0.894	1.000	1.000	1.000	1.000
1	0.051	0.806	1.000	1.000	1.000	1.000

Notes: Results are based on 100,000 replications using Gauss.
The simulated model is

$$y_t = \rho_t y_{t-1} + \varepsilon_t, \quad \rho_t \sim \mathrm{IN}(\rho, \omega^2), \quad t = 1, 2, \ldots, n.$$

Accepted manuscript

ACCEPTED MANUSCRIPT

Table 10: Simulated power values of AMLM_a for different values of the parameters and n

ρ	$\omega^2 = 0$	$\omega^2 = 0.001$	$\omega^2 = 0.005$	$\omega^2 = 0.01$	$\omega^2 = 0.05$	$\omega^2 = 0.1$
			$n = 25$			
.8	0.095	0.093	0.095	0.098	0.121	0.154
.9	0.062	0.065	0.065	0.065	0.086	0.123
.95	0.054	0.055	0.054	0.056	0.079	0.117
.99	0.051	0.051	0.053	0.054	0.077	0.117
1	0.050	0.051	0.052	0.054	0.078	0.116
			$n = 50$			
.8	0.257	0.255	0.262	0.270	0.345	0.453
.9	0.102	0.106	0.110	0.117	0.300	0.388
.95	0.064	0.067	0.070	0.077	0.250	0.363
.99	0.053	0.052	0.055	0.062	0.251	0.369
1	0.050	0.050	0.053	0.063	0.265	0.414
			$n = 100$			
.8	0.796	0.800	0.805	0.812	0.872	0.936
.9	0.276	0.282	0.295	0.317	0.573	0.815
.95	0.108	0.113	0.125	0.142	0.471	0.774
.99	0.055	0.055	0.067	0.093	0.485	0.775
1	0.050	0.051	0.068	0.130	0.584	0.775
			$n = 200$			
.8	1.000	1.000	1.000	1.000	1.000	1.000
.9	0.818	0.821	0.831	0.854	0.981	1.000
.95	0.294	0.300	0.340	0.424	0.944	1.000
.99	0.061	0.065	0.119	0.298	0.925	0.981
1	0.051	0.055	0.243	0.418	0.958	0.988
			$n = 500$			
.8	1.000	1.000	1.000	1.000	1.000	1.000
.9	1.000	1.000	1.000	1.000	1.000	1.000
.95	0.959	0.960	0.978	0.993	1.000	1.000
.99	0.118	0.136	0.548	0.891	0.995	1.000
1	0.050	0.165	0.648	0.875	0.993	1.000
			$n = 1000$			
.8	1.000	1.000	1.000	1.000	1.000	1.000
.9	1.000	1.000	1.000	1.000	1.000	1.000
.95	1.000	1.000	1.000	1.000	1.000	1.000
.99	0.302	0.404	0.980	1.000	1.000	1.000
1	0.050	0.384	0.926	0.977	1.000	1.000

Notes: Results are based on 100,000 replications using Gauss.
The simulated model is

$$y_t = \rho_t y_{t-1} + \varepsilon_t, \quad \rho_t \sim \text{IN}(\rho, \omega^2), \quad t = 1, 2, \ldots, n.$$

Accepted manuscript

ACCEPTED MANUSCRIPT

A Appendix

A.1 Proof of Theorem 1

(i) The numerator of the test statistic can be expanded as follows

$$\begin{aligned}
&\frac{1}{2\widetilde{\sigma}^4}\sum_{t=2}^{n} y_{t-1}^2\left[\left(y_t-\widetilde{\rho}y_{t-1}\right)^2-\widetilde{\sigma}^2\right] \\
=\ &\underbrace{\frac{1}{2\widetilde{\sigma}^4}\sum_{t=2}^{n} y_{t-1}^2\left(\varepsilon_t^2-\widetilde{\sigma}^2\right)}_{A_n^{(1)}}-\underbrace{\frac{1}{\widetilde{\sigma}^4}\left(\widetilde{\rho}-1\right)\sum_{t=2}^{n} y_{t-1}^3\varepsilon_t}_{A_n^{(2)}}+\underbrace{\frac{1}{2\widetilde{\sigma}^4}\left(\widetilde{\rho}-1\right)^2\sum_{t=2}^{n} y_{t-1}^4}_{A_n^{(3)}}.
\end{aligned}$$

Now, given that in this case

$$\widetilde{\rho}=\frac{\sum y_t y_{t-1}}{\sum y_{t-1}^2}$$

and therefore $\left(\widetilde{\rho}-1\right)=O_p\left(n^{-1}\right)$, it is straightforward to show that

$$A_n^{(1)}=O_p\left(n^{3/2}\right),\quad A_n^{(2)}=O_p\left(n\right),\quad A_n^{(3)}=O_p\left(n\right).$$

Therefore the dominant term of the numerator is $A_n^{(1)}$. Similarly, the denominator is given by

$$\begin{aligned}
&\frac{1}{2\widetilde{\sigma}^6}\sum_{t=2}^{n} y_{t-1}^4\left[2\left(y_t-\widetilde{\rho}y_{t-1}\right)^2-\widetilde{\sigma}^2\right] \\
=\ &\underbrace{\frac{1}{\widetilde{\sigma}^6}\sum_{t=2}^{n} y_{t-1}^4\left(\varepsilon_t^2-\widetilde{\sigma}^2\right)}_{B_n^{(1)}}-\underbrace{\frac{2}{\widetilde{\sigma}^6}\left(\widetilde{\rho}-1\right)\sum_{t=2}^{n} y_{t-1}^5\varepsilon_t}_{B_n^{(2)}}+\underbrace{\frac{1}{\widetilde{\sigma}^6}\left(\widetilde{\rho}-1\right)^2\sum_{t=2}^{n} y_{t-1}^6}_{B_n^{(3)}}+\underbrace{\frac{1}{2\widetilde{\sigma}^4}\sum_{t=2}^{n} y_{t-1}^4}_{B_n^{(4)}}.
\end{aligned}$$

Again, it is straightforward to show that

$$B_n^{(1)}=O_p\left(n^{5/2}\right),\quad B_n^{(2)}=O_p\left(n^2\right),\quad B_n^{(3)}=O_p\left(n^2\right),\quad B_n^{(4)}=O_p\left(n^3\right).$$

Therefore the dominant term of the denominator is $B_n^{(4)}$. The asymptotic distribution of the test statistic follows by the CMT and by (16) and (17) below.

(ii) Immediate, given normality of $\{\varepsilon_t\}$.

(iii) Start from

$$q_1\left(\widetilde{\boldsymbol{\theta}}\right)=\sum_{t=2}^{n}\frac{y_{t-1}\left(y_t-y_{t-1}\right)}{\widetilde{\omega}^2 y_{t-1}^2+\widetilde{\sigma}^2}.$$

Using simple algebra and noting that, when $\omega^2=0$,

$$y_t-y_{t-1}=\varepsilon_t$$

Accepted manuscript

ACCEPTED MANUSCRIPT

the score can be rewritten as

$$q_1\left(\widetilde{\boldsymbol{\theta}}\right)=\sum_{t=2}^{n}\frac{y_{t-1}\varepsilon_t}{\left(\widetilde{\omega}^2-\omega^2\right)y_{t-1}^2+\omega^2y_{t-1}^2+\left(\widetilde{\sigma}^2-\sigma^2\right)+\sigma^2}.$$

There are different methods to estimate the parameter ω^2 (see, e.g., Amemiya, 1977). The different estimators have the same rate of convergence but different asymptotic variance. Concentrating on the estimator given in Amemiya (1977, equation (5)), it is easy to show that $\left(\widetilde{\omega}^2-\omega^2\right)=O_p\left(n^{-3/2}\right)$. Also, by consistency of $\widetilde{\sigma}^2$, $\left(\widetilde{\sigma}^2-\sigma_\varepsilon^2\right)=o_p\left(1\right)$ and

$$\frac{q_1\left(\widetilde{\boldsymbol{\theta}}\right)}{n}\xrightarrow{w}\int_0^1 W_1\left(s\right)\,\mathrm{d}W_1\left(s\right).$$

The Hessian matrix can be analyzed similarly. In fact,

$$-\mathcal{H}_{11}\left(\widetilde{\boldsymbol{\theta}}\right)=\sum_{t=2}^{n}\frac{y_{t-1}^2}{\left(\widetilde{\omega}^2-\omega^2\right)y_{t-1}^2+\omega^2y_{t-1}^2+\left(\widetilde{\sigma}^2-\sigma^2\right)+\sigma^2}$$

and, using the same argument as above,

$$-\frac{\mathcal{H}_{11}\left(\widetilde{\boldsymbol{\theta}}\right)}{n^2}\xrightarrow{w}\int_0^1 W_1\left(s\right)^2\,\mathrm{d}s.$$

The limit distribution is then obtained by applying the CMT and (18).

(iv) Start from

$$q_1\left(\widetilde{\boldsymbol{\theta}}\right)=\sum_{t=2}^{n}\frac{y_{t-1}\left(y_t-y_{t-1}\right)}{\widetilde{\omega}^2y_{t-1}^2+\widetilde{\sigma}^2}.$$

Noting that

$$y_t-y_{t-1}=\left(\rho_t-1\right)y_{t-1}+\varepsilon_t$$

the score can be rewritten as

$$q_1\left(\widetilde{\boldsymbol{\theta}}\right)=\underbrace{\sum_{t=2}^{n}\frac{y_{t-1}^2\left(\rho_t-1\right)}{\widetilde{\omega}^2y_{t-1}^2+\widetilde{\sigma}^2}}_{C_n^{(1)}}+\underbrace{\sum_{t=2}^{n}\frac{y_{t-1}\varepsilon_t}{\widetilde{\omega}^2y_{t-1}^2+\widetilde{\sigma}^2}}_{C_n^{(2)}}.$$

Clearly, $C_n^{(1)}$ is of a larger order of probability than $C_n^{(2)}$. Then, using simple algebra

$$C_n^{(1)}=\underbrace{\frac{1}{\widetilde{\omega}^2}\sum_{t=2}^{n}\left(\rho_t-1\right)}_{C_n^{(3)}}-\underbrace{\frac{\widetilde{\sigma}^2}{\widetilde{\omega}^2}\sum_{t=2}^{n}\frac{\left(\rho_t-1\right)}{\widetilde{\omega}^2y_{t-1}^2+\widetilde{\sigma}^2}}_{C_n^{(4)}}.$$

The dominant term is given by $C_n^{(3)}$ and it is easy to show that

$$\frac{\widetilde{\omega}}{\sqrt{n-1}}C_n^{(3)}\xrightarrow{d}\mathrm{N}(0,1).$$

Accepted manuscript

ACCEPTED MANUSCRIPT

The Hessian matrix can be analyzed similarly.

$$-\mathcal{H}_{11}\left(\widetilde{\boldsymbol{\theta}}\right)=\sum_{t=2}^{n}\frac{y_{t-1}^{2}}{\widetilde{\omega}^{2}y_{t-1}^{2}+\widetilde{\sigma}^{2}}=\underbrace{\sum_{t=2}^{n}\frac{y_{t-1}^{2}+\widetilde{\sigma}^{2}/\widetilde{\omega}^{2}}{\widetilde{\omega}^{2}y_{t-1}^{2}+\widetilde{\sigma}^{2}}}_{D_{n}^{(1)}}-\underbrace{\frac{\widetilde{\sigma}^{2}}{\widetilde{\omega}^{2}}\sum_{t=2}^{n}\frac{1}{\widetilde{\omega}^{2}y_{t-1}^{2}+\widetilde{\sigma}^{2}}}_{D_{n}^{(2)}}.$$

The dominant term is $D_n^{(1)}=\frac{n-1}{\widetilde{\omega}^2}$. This completes the proof.

(v) The elements of the score vector, evaluated at the values of the parameters implied by hypothesis (7) are given by

$$\begin{aligned}
q_1\left(\widetilde{\boldsymbol{\theta}}\right)&=\frac{1}{\widetilde{\sigma}^2}\sum y_{t-1}\left(y_t-y_{t-1}\right),\\
q_2\left(\widetilde{\boldsymbol{\theta}}\right)=\frac{1}{2\widetilde{\sigma}^4}\sum y_{t-1}^2\left(y_t-y_{t-1}\right)^2-\frac{1}{2\widetilde{\sigma}^2}\sum y_{t-1}^2&=\frac{1}{2\widetilde{\sigma}^4}\sum y_{t-1}^2\left(\varepsilon_t^2-\widetilde{\sigma}^2\right),\\
q_3\left(\widetilde{\boldsymbol{\theta}}\right)&=0,
\end{aligned}\tag{15}$$

where hereafter the sums go from 2 to n and $\widetilde{\sigma}^2=\sum\left(y_t-y_{t-1}\right)^2/\left(n-1\right)$. When H_0 is true, $\widetilde{\sigma}^2\xrightarrow{p}\sigma_\varepsilon^2$ and

$$y_t=\sum_{j=0}^{t-1}\varepsilon_{t-j}.$$

With appropriate standardization, and applying the functional central limit theorem, the following convergence results can be established

$$\frac{q_1\left(\widetilde{\boldsymbol{\theta}}\right)}{n}\xrightarrow{w}\int_0^1 W_1\left(s\right)\,\mathrm{d}W_1\left(s\right)$$

and

$$\begin{aligned}
\frac{q_2\left(\widetilde{\boldsymbol{\theta}}\right)}{n^{3/2}}&\xrightarrow{w}\frac{\kappa_\varepsilon\sigma_\varepsilon^2}{2\sigma_\varepsilon^4}\int_0^1\left(W_1\left(s\right)^2-\int_0^1 W_1\left(s\right)^2\,\mathrm{d}s\right)\mathrm{d}W^*\left(s\right)\\
&=\frac{1}{\sqrt{2}}\left(\int_0^1 W_1\left(s\right)^2\,\mathrm{d}W^*\left(s\right)-\int_0^1 W_1\left(s\right)^2\,\mathrm{d}s\,W^*\left(1\right)\right),
\end{aligned}\tag{16}$$

where the last result is due to McCabe and Tremayne (1995).

Similarly the following convergence results can be established for the elements of the Hessian matrix, appropriately standardized

$$-\frac{\mathcal{H}_{11}\left(\widetilde{\boldsymbol{\theta}}\right)}{n^2}=\frac{1}{n^2\widetilde{\sigma}^2}\sum y_{t-1}^2\xrightarrow{w}\int_0^1 W_1\left(s\right)^2\,\mathrm{d}s,$$

$$-\frac{\mathcal{H}_{12}\left(\widetilde{\boldsymbol{\theta}}\right)}{n^2}=\frac{1}{n^2\widetilde{\sigma}^4}\sum y_{t-1}^3\left(y_t-y_{t-1}\right)\xrightarrow{w}\int_0^1 W_1\left(s\right)^3\,\mathrm{d}W_1\left(s\right),$$

$$-\frac{\mathcal{H}_{13}\left(\widetilde{\boldsymbol{\theta}}\right)}{n}=\frac{1}{n\widetilde{\sigma}^4}\sum y_{t-1}\left(y_t-y_{t-1}\right)\xrightarrow{w}\frac{1}{\sigma_\varepsilon^2}\int_0^1 W_1\left(s\right)\,\mathrm{d}W_1\left(s\right),$$

Accepted manuscript

ACCEPTED MANUSCRIPT

$$-\frac{\mathcal{H}_{22}\left(\widetilde{\boldsymbol{\theta}}\right)}{n^3}=\frac{1}{n^3}\left(\frac{1}{\widetilde{\sigma}^6}\sum y_{t-1}^4\left(y_t-y_{t-1}\right)^2-\frac{1}{2\widetilde{\sigma}^4}\sum y_{t-1}^4\right)\xrightarrow{w}\frac{1}{2}\int_0^1 W_1\left(s\right)^4\,\mathrm{d}s, \tag{17}$$

$$-\frac{\mathcal{H}_{23}\left(\widetilde{\boldsymbol{\theta}}\right)}{n^2}=\frac{1}{n^2}\left(\frac{1}{\widetilde{\sigma}^6}\sum y_{t-1}^2\left(y_t-y_{t-1}\right)^2-\frac{1}{2\widetilde{\sigma}^4}\sum y_{t-1}^2\right)\xrightarrow{w}\frac{1}{2\sigma_\varepsilon^2}\int_0^1 W_1\left(s\right)^2\,\mathrm{d}s,$$

$$-\frac{\mathcal{H}_{33}\left(\widetilde{\boldsymbol{\theta}}\right)}{n}=\frac{1}{n}\left(\frac{1}{\widetilde{\sigma}^6}\sum\left(y_t-y_{t-1}\right)^2-\frac{n-1}{2\widetilde{\sigma}^4}\right)\xrightarrow{p}\frac{1}{2\sigma_\varepsilon^4}.$$

Then it suffices to apply the CMT and to use the following result

$$\int_0^1 W_1\left(s\right)\,\mathrm{d}W_1\left(s\right)=\frac{1}{2}\left(W_1\left(1\right)^2-1\right) \tag{18}$$

to obtain the stated formula. ■

A.2 Proof of Corollary 1

In order to prove the Corollary, it suffices to look at the probabilistic orders of magnitude of the components of the test statistic.

Then

$$\frac{\mathcal{H}_{12}\left(\widetilde{\boldsymbol{\theta}}\right)}{\mathcal{H}_{11}\left(\widetilde{\boldsymbol{\theta}}\right)}q_1\left(\widetilde{\boldsymbol{\theta}}\right)=o_p\left(q_2\left(\widetilde{\boldsymbol{\theta}}\right)\right),$$

$\boldsymbol{\mathcal{H}}_{22}\left(\widetilde{\boldsymbol{\theta}}\right)=\boldsymbol{\mathcal{H}}_{2_\perp 1}\left(\widetilde{\boldsymbol{\theta}}\right)=O_p\left(n^3\right)$, and finally $\left(\mathcal{H}_{21}\left(\widetilde{\boldsymbol{\theta}}\right)\right)^2/\mathcal{H}_{11}\left(\widetilde{\boldsymbol{\theta}}\right)=o_p\left(n^3\right)$.

■

A.3 Proof of Theorem 2

(i) Follows immediately, by a similar argument to that in Theorem 1, part (i) and by part (v) below.

(ii) Follows immediately, by a similar argument to that in Theorem 1, part (ii).

(iii) Follows immediately, by a similar argument to that in Theorem 1, part (iii) and by part (v) below.

(iv) Follows immediately, by a similar argument to that in Theorem 1, part (iv).

(v) In order to prove this part, recourse is made to the result of Corollary 1, which holds trivially when deterministic components are included in the model.

In fact

$$\frac{\mathcal{H}_{12}\left(\widetilde{\boldsymbol{\theta}}_a\right)}{\mathcal{H}_{11}\left(\widetilde{\boldsymbol{\theta}}_a\right)}q_1\left(\widetilde{\boldsymbol{\theta}}_a\right)=o_p\left(q_2\left(\widetilde{\boldsymbol{\theta}}_a\right)\right)$$

and

$$\frac{\left(\mathcal{H}_{21}\left(\widetilde{\boldsymbol{\theta}}_a\right)\right)^2}{\mathcal{H}_{11}\left(\widetilde{\boldsymbol{\theta}}_a\right)}=o_p\left(\mathcal{H}_{22}\left(\widetilde{\boldsymbol{\theta}}_a\right)\right).$$

Accepted manuscript

ACCEPTED MANUSCRIPT

Then the limiting distribution of AMLM_a is given by the sum of the squared test statistics on ρ and on ω^2. The limiting distribution of the first component of the test statistic can be readily obtained using the result of Phillips and Xiao (1998, p.433), by substituting for c (the local to unity parameter) equal to zero and specializing the reported formulae to the case of a constant and linear trend.

The score and the Hessian for ω^2 evaluated at the value of the parameters implied by H_0 and using the unrestricted estimators of the nuisance parameters α and β are given respectively by

$$q_2\left(\widetilde{\boldsymbol{\theta}}_a\right) = \frac{1}{2\widetilde{\sigma}^4}\sum\left(\widehat{y}_t^* - \widehat{y}_{t-1}^*\right)^2 \widehat{y}_{t-1}^{*2} - \frac{1}{2\widetilde{\sigma}^2}\sum \widehat{y}_{t-1}^{*2} = \frac{1}{2\widetilde{\sigma}^4}\sum\left[\widehat{y}_{t-1}^{*2}\left(\widehat{\varepsilon}_t^2 - \widetilde{\sigma}^2\right)\right] \tag{19}$$

and

$$-\mathcal{H}_{22}\left(\widetilde{\boldsymbol{\theta}}_a\right) = \frac{1}{\widetilde{\sigma}^6}\sum \widehat{y}_{t-1}^{*4}\left(\widehat{y}_t^* - \widehat{y}_{t-1}^*\right)^2 - \frac{1}{2\widetilde{\sigma}^4}\sum \widehat{y}_{t-1}^{*4} = \frac{1}{2\widetilde{\sigma}^6}\left[2\sum \widehat{y}_{t-1}^{*4}\left(\widehat{\varepsilon}_t^2 - \widetilde{\sigma}^2\right) + \widetilde{\sigma}^2\sum \widehat{y}_{t-1}^{*4}\right],$$

where $\widehat{y}_t^* = y_t - \widehat{\alpha} - \widehat{\beta}t$ and $\widehat{\varepsilon}_t = \widehat{y}_t^* - \widehat{y}_{t-1}^*$. When H_0 is true, then

$$y_t = \alpha + \beta t + \sum_{j=0}^{t-1}\varepsilon_{t-j}.$$

Therefore the partial sum process of

$$\sum \widehat{y}_{t-1}^* = \sum\left[(\alpha - \widehat{\alpha}) + \left(\beta - \widehat{\beta}\right)(t-1) + \sum_{j=0}^{t-2}\varepsilon_{t-j-1}\right],$$

appropriately standardized by $n^{3/2}$, converges weakly to

$$\sigma_\varepsilon \int_0^1 \left(W_1(s) - s\xi\right)\mathrm{d}s,$$

since $(\alpha - \widehat{\alpha}) = O_p(1)$ (for more details, see Distaso, 2002).

Applying the functional central limit theorem, the following convergence results can be established

$$\begin{aligned}\frac{q_2\left(\widetilde{\boldsymbol{\theta}}_a\right)}{n^{3/2}} &\xrightarrow{w} \frac{\kappa_\varepsilon\sigma_\varepsilon^2}{2\sigma_\varepsilon^4}\int_0^1\left(\left(W_1(s) - s\xi\right)^2 - \int_0^1\left(W_1(s) - s\xi\right)^2\mathrm{d}s\right)\mathrm{d}W^*(s)\\ &= \frac{1}{\sqrt{2}}\left(\int_0^1\left(W_1(s) - s\xi\right)^2\mathrm{d}W^*(s) - \int_0^1\left(W_1(s) - s\xi\right)^2\mathrm{d}s\, W^*(1)\right),\end{aligned}$$

and

$$-\frac{\mathcal{H}_{22}\left(\widetilde{\boldsymbol{\theta}}_a\right)}{n^3} \xrightarrow{w} \frac{1}{2}\int_0^1\left(W_1(s) - s\xi\right)^4\mathrm{d}s.$$

Then it suffices to apply the CMT to obtain the stated formula. ■

Accepted manuscript

ACCEPTED MANUSCRIPT

A.4 Proof of Lemma 1

Immediate, by Herrndorff's functional central limit theorem (Herrndorff, 1984). ■

Acknowledgements. This paper is the basis of one of the chapter of my Doctoral dissertation at the University of York. I wish to thank the Editor, Peter Robinson, an Associate Editor and two anonymous referees for comments which led to a substantially improved version of the paper. Comments and suggestions from Karim Abadir, Valentina Corradi, Liudas Giraitis, Brendan McCabe and Peter Phillips proved most helpful. The usual disclaimer applies. Financial support from the ESRC (UK) grant RES-062-23-0311 is gratefully acknowledged.

Accepted manuscript

ACCEPTED MANUSCRIPT

References

Abadir, K. M., 1993, On the Asymptotic Power of Unit Root Tests, Econometric Theory 9, 189–221.

Abadir, K. M., 2004, Cointegration Theory, Equilibrium and Disequilibrium Economics, Manchester School 72, 60–71.

Abadir, K. M., Distaso, W., 2006, Testing Joint Hypotheses when One of the Alternatives is One-Sided, Journal of Econometrics, forthcoming.

Amemiya, T., 1977, A Note on a Heteroscedastic Model, Journal of Econometrics 6, 365–370.

Anderson, T. W., 1948, On the Theory of Testing Serial Correlation, Skandinavisk Aktuarietidskrift 31, 88–116.

Bhargava, A., 1986, On the Theory of Testing for Unit Roots in Observed Time Series, Review of Economic Studies 53, 369–384.

Breusch, T. S., Pagan, A. R., 1979, A Simple Test for Heteroscedasticity and Random Coefficient Variation, Econometrica 47, 1287–1294.

Cai, Z., Hurvich, C. M., Tsai, C., 1998, Score Tests for Heteroscedasticity in Wavelet Regression, Biometrika 85, 229–234.

DeJong, D. N., Nankervis, J. C., Savin, N. E., Whiteman, C. H., 1992, Integration versus Trend Stationarity in Time Series, Econometrica 60, 423–433.

Dickey, D. A., Fuller, W. A., 1979, Distribution of the Estimators for Autoregressive Time Series with a Unit Root, Journal of the American Statistical Association 74, 427–431.

Dickey, D. A., Fuller, W. A., 1981, Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root, Econometrica 49, 1057–1072.

Ding, Z., Granger, C. W. J., 1996, Modeling Volatility Persistence of Speculative returns: A New Approach, Journal of Econometrics 73, 185–215.

Ding, Z., Granger, C. W. J., Engle, R. F., 1993, A Long Memory Property of Stock Returns and a New Model, Journal of Empirical Finance 1, 83–106.

Distaso, W., 2002, Improved Inference in Unit Root Models, PhD thesis, University of York.

Dufour, J. M., King, M. L., 1991, Optimal Invariant Tests for the Autocorrelation Coefficient in Linear Regressions with Stationary or Nonstationary AR(1) Errors, Journal of Econometrics 47, 115–143.

Accepted manuscript

ACCEPTED MANUSCRIPT

Elliott, G., Rothenberg, T. J., Stock, J. H., 1996, Efficient Tests for an Autoregressive Unit Root, Econometrica 64, 813–836.

Evans, G. B. A., Savin, N. E., 1981, Testing for Unit Roots: 1, Econometrica 49, 753–779.

Fuller, W. A., 1995, Introduction to Statistical Time Series, Second edn, John Wiley & Sons, New York.

Granger, C. W. J., 1980, Long Memory Relationships and the Aggregation of Dynamic Models, Journal of Econometrics 14, 227–238.

Granger, C. W. J., Swanson, N. R., 1997, An Introduction to Stochastic Unit-Root Processes, Journal of Econometrics 80, 35–62.

Hall, R. E., 1978, Stochastic Implications of the Life Cycle-Permanent Income Hypothesis, Journal of Political Economy 96, 971–987.

Herrndorff, N., 1984, A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables, Annals of Probability 12, 141–153.

Leybourne, S. J., McCabe, B. P. M., Tremayne, A. R., 1996, Can Economic Time Series Be Differenced to Stationarity ?, Journal of Business & Economic Statistics 14, 435–446.

MacKinnon, J. G., 1994, Approximate Asymptotic Distribution Functions for Unit-Root and Cointegration Tests, Journal of Business & Economic Statistics 12, 167–176.

McCabe, B. P. M., Tremayne, A. R., 1995, Testing a Time Series for Difference Stationarity, Annals of Statistics 23, 1015–1028.

Nicholls, D. F., Quinn, B. G., 1982, Random Coefficient Autoregressive Models: an Introduction, Springer-Verlag, New York.

Pere, P., 2003, AR(1) Models, Unit Roots, and Adjusted Profile Likelihood, Econometric Theory 19, 885–922.

Phillips, P. C. B., 1987, Time Series Regression with a Unit Root, Econometrica 55, 277–301.

Phillips, P. C. B., Perron, P., 1988, Testing for a Unit Root in Time Series Regression, Biometrika 75, 335–346.

Phillips, P. C. B., Ploberger, W., 1996, An Asymptotic Theory of Bayesan Inference for Time Series, Econometrica 64, 381–412.

Phillips, P. C. B., Xiao, Z., 1998, A Primer on Unit Root Testing, Journal of Economic Surveys 12, 423–469.

Accepted manuscript

ACCEPTED MANUSCRIPT

Romano, J. P., Wolf, M., 2001, Subsampling Intervals in Autoregressive Models with Linear Time Trend, Econometrica 69, 1283–1314.

Schmidt, P., Phillips, P. C. B., 1992, LM Tests for a Unit Root in the Presence of Deterministic Time Trends, Oxford Bulletin of Economics and Statistics 54, 257–287.

Tsay, R. S., 1987, Conditional Heteroskedastic Time Series Models, Journal of the American Statistical Association 82, 590–604.

Xiao, Z., 2001, Likelihood-based Inference in Trending Time Series with a Root near Unity, Econometric Theory 17, 1082–1112.

Yoon, G., 2002, A Simple Model that Generates Stylized Facts of Returns, University of California, San Diego.

Accepted manuscript