

Pentti Saikkonen – Rickard Sandberg

Testing for a unit root in noncausal autoregressive models



EUROJÄRJESTELMÄ
EUROSYSTEMET

Bank of Finland Research
Discussion Papers
26 • 2013

Testing for a Unit Root in Noncausal Autoregressive Models*

by

Pentti Saikkonen[†]

University of Helsinki and Bank of Finland

Rickard Sandberg[‡]

Stockholm School of Economics

September 2013

Abstract

This work develops likelihood-based unit root tests in the noncausal autoregressive (NCAR) model formulated by Lanne and Saikkonen (2011, *Journal of Time Series Econometrics* 3, Iss. 3, Article 2). The possible unit root is assumed to appear in the causal autoregressive polynomial and for reasons of identification the error term of the model is supposed to be non-Gaussian. In order to derive the tests, asymptotic properties of the maximum likelihood estimators are established under the unit root hypothesis. The limiting distributions of the proposed tests depend on a nuisance parameter determined by the distribution of the error term of the model. A simple procedure to handle this nuisance parameter dependence in applications is proposed. Finite sample properties of the tests are examined by means of Monte Carlo simulations. The results show that the size properties of the tests are satisfactory and the power against stationary NCAR alternatives is significantly higher than the power of conventional Dickey-Fuller tests and the M -tests of Lucas (1995, *Econometric Theory* 11, 331-346). In an empirical application to a Finnish interest rate series evidence in favour of a stationary NCAR model with leptokurtic errors is found.

Key words: Maximum likelihood estimation; Noncausal autoregressive model; Non-Gaussian time series; Unit root.

*The first author acknowledges financial support from the Academy of Finland, the OP-Pohjola Group Research Foundation, and the Finnish Cultural Foundation. Part of this research was done during his research visit to the Monetary Policy and Research Department of the Bank of Finland whose hospitality is gratefully acknowledged. The second author acknowledges financial support from Jan Wallander's and Tom Hedelius' Foundation, Grant No. P2012-0085:1.

[†] Department of Mathematics and Statistics, University of Helsinki, P.O. Box 68 (Gustaf Hällströmin katu 2b), FIN-00014 University of Helsinki, e-mail: pentti.saikkonen@helsinki.fi.

[‡] Department of Economics, Center for Economic Statistics, Stockholm School of Economics, P.O. Box 6501 (Sveavägen 65), 113 83 Stockholm, e-mail: rickard.sandberg@hhs.se.

1 Introduction

Testing for the unit root hypothesis is an important step in the analysis of economic time series, and has attracted an enormous amount of interest during the past decades. In this context, the most widely used model is the conventional (causal) autoregressive (AR) model where the current observation is expressed as a weighted average of past observations and an error term. An essential assumption of the conventional AR model is that the error term is unpredictable by the past of the considered time series. However, in (say) economic applications this assumption may break down because the impact of omitted variables, interrelated with the considered (univariate) time series, is ignored. More specifically, if relevant variables are omitted their impact goes (at least partly) to the error term of the model and, as the considered time series may help to predict the omitted variables, the assumed unpredictability condition may break down. As economic variables are typically interrelated, this point appears particularly pertinent in economic applications. In cases like this the noncausal AR (NCAR) model may provide a viable alternative, for it explicitly allows for the predictability of the error term by the past of the considered series.

Early studies of NCAR models and their extensions, noncausal and (potentially) noninvertible autoregressive moving average (ARMA) models, were mainly motivated by applications to natural sciences and engineering (see, e.g., Breidt et al. (1991), Lii and Rosenblatt (1996), Huang and Pawitan (2000), Rosenblatt (2000), Breidt et al. (2001), Wu and Davis (2010), and the references therein). More recently, a slightly different formulation of the NCAR model was considered by Lanne and Saikkonen (2011) (hereafter L&S), and further studied by Lanne et al. (2012a), Lanne et al. (2012b), Lanne et al. (2012c), Lanne and Saikkonen (2013), and Gouriéroux and Zakoian (2013). These papers demonstrate that the NCAR model can successfully describe and forecast many economic time series, and it often outperforms its conventional causal alternative in terms of model fit and forecasting accuracy.

Even though the properties of the stationary NCAR model are by now well understood and asymptotic distribution theory for various parameter estimators (typically maximum likelihood estimators) have been developed, the nonstationary case and tests for a unit root have not yet been studied in the literature. As unit root type nonstationarity appears quite common (particularly) in economic time series, and hence potential applications of the NCAR model, this work aims at proposing unit root tests in the context of the NCAR model of L&S. We develop Wald type unit root tests by assuming that the possible unit root appears in the causal autoregressive polynomial of the model, and to this end we first derive asymptotic properties of a (local) maximum likelihood (ML) estimator of the parameters of the model under the unit root hypothesis. As in the stationary case, a non-Gaussian error term is required to achieve identification (see, e.g., Brockwell and Davis (1987, pp. 124-125) and Rosenblatt (2000, pp. 10-11)). This renders the estimation problem nonlinear which, in turn, makes the derivation of limiting distributions less straightforward than in the context of conventional unit root tests, where estimation is carried out by linear least squares (LS) techniques. To address this issue, we use ideas similar to those used in statistical models whose likelihood ratios satisfy the so-called locally asymptotically mixed normal (LAMN) condition (see Basawa and Scott (1983), Ch. 2). It turns out that the limiting distributions of our tests are not distribution free and appear, in general, very complicated depending on a number of nuisance parameters. To obtain tests with manageable limiting distributions we assume that the error term

of the model has a symmetric distribution. Then the limiting distributions of our tests only depend on a single nuisance parameter determined by the distribution of the error term, and this problem can be rather easily circumvented by using estimated critical values (described in Section 5.1).

We examine the practical relevance of our asymptotic tests by means of Monte Carlo simulations. The results show that our tests perform satisfactorily in terms of size and their power against stationary NCAR alternatives is significantly higher than the power of conventional Dickey-Fuller (DF) tests and the M -tests of Lucas (1995). To illustrate the practical implementation of our tests we present an application to a Finnish interest rate series for which a stationary NCAR model with Student's t -distributed errors is found to provide a good description.

The plan of the paper is as follows. Section 2 defines the considered NCAR model and discusses the testing problem. Parameter estimation and related asymptotic results are presented in Section 3 and used in Section 4 to obtain our unit root tests. Section 5 reports the results of the Monte Carlo simulations and Section 6 presents the empirical application. Section 7 concludes. Three appendices contain mathematical proofs and some technical details.

Finally, the following notation is used throughout the paper. The notation \xrightarrow{p} signifies convergence in probability and \xrightarrow{d} is used for convergence in distribution and also for weak convergence in a function space. We write $B(u) \sim BM(\cdot)$ for a Brownian motion $B(u)$ with indicated variance or covariance matrix. Unless otherwise stated, all vectors will be treated as column vectors and, for notational convenience, we shall write $x = (x_1, \dots, x_n)$ for the (column) vector x where the components x_i may be either scalars or vectors (or both).

2 Model and testing problem

Following L&S we consider the NCAR model

$$\phi(B) \varphi(B^{-1}) y_t = \epsilon_t, \quad t = 1, 2, \dots, \quad (1)$$

where ϵ_t is a sequence of independent and identically distributed (IID) random variables with mean 0 and finite variance $\sigma^2 > 0$, B is the usual backward shift operator ($By_t = y_{t-k}$ for $k = 0, \pm 1, \dots$), and $\phi(B) = 1 - \phi_1 B - \dots - \phi_r B^r$ and $\varphi(B^{-1}) = 1 - \varphi_1 B^{-1} - \dots - \varphi_s B^{-s}$. L&S assume that the polynomials $\phi(z)$ and $\varphi(z)$ ($z \in \mathbb{C}$) have their roots outside the unit circle in which case the difference equation (1) has a stationary solution. In this paper, we allow for the possibility that, due to a unit root in the causal autoregressive polynomial $\phi(B)$, the process y_t is a nonstationary integrated process. Thus, we assume that $r > 0$ and proceed in the conventional way by writing the lag polynomial $\phi(B)$ as

$$\phi(B) = \Delta - \phi B - \pi_1 \Delta B - \dots - \pi_{r-1} \Delta B^{r-1}, \quad (2)$$

where $\Delta = 1 - B$ is the difference operator. Our focus is in testing for the unit root null hypothesis $H_0 : \phi = 0$ against the stationary alternative $H_1 : \phi < 0$. At this point we abstract from any deterministic terms such as a constant term or linear time trend in the process. These extensions will be discussed in Section 4.2.

Unless otherwise stated we assume throughout the paper that the null hypothesis H_0 holds and that the roots of the polynomials $\pi(z) = 1 - \pi_1 z - \dots - \pi_{r-1} z^{r-1}$ and $\varphi(z)$ lie outside the unit circle or, formally, that

$$\pi(z) \neq 0 \quad \text{for } |z| \leq 1 \quad \text{and} \quad \varphi(z) \neq 0 \quad \text{for } |z| \leq 1. \quad (3)$$

Using equation (2) we can write equation (1) as

$$\Delta y_t = \phi y_{t-1} + \pi_1 \Delta y_{t-1} + \cdots + \pi_{r-1} \Delta y_{t-r+1} + v_t, \quad t = 1, 2, \dots, \quad (4)$$

where the process $v_t = \phi(B) y_t = \varphi(B^{-1})^{-1} \epsilon_t$ has the forward moving average representation

$$v_t = \sum_{j=0}^{\infty} \beta_j \epsilon_{t+j}, \quad \beta_0 = 1. \quad (5)$$

Here β_j is the coefficient of z^{-j} in the Laurent series expansion of $\varphi(z^{-1})^{-1}$. By the latter condition in (3) this expansion is well defined for $|z| \geq b_\varphi$ with some $b_\varphi < 1$ and with the coefficients β_j decaying to zero at a geometric rate as $j \rightarrow \infty$. Equation (4) shows that our testing problem can be thought of as testing for a unit root in an $\text{AR}(r)$ process with stationary errors following the purely noncausal $\text{AR}(0, s)$ process $\varphi(B^{-1}) v_t = \epsilon_t$ (as in L&S we use the acronym $\text{AR}(r, s)$ for the model defined in equation (1)). When $r = 1$ the lagged differences vanish from the right hand side of equation (4) which becomes a special case of a first-order autoregression with general stationary (or short-memory) errors. Testing for a unit root in such contexts has been considered in a number of papers since the work of Phillips (1987) and Phillips and Perron (1988). That the errors in (4) are generated by a purely noncausal $\text{AR}(0, s)$ process distinguishes our formulation from its previous counterparts. At this point we also note that the conceivable alternative of including a unit root in the noncausal polynomial $\varphi(\cdot)$ involves the difficulty that there seems to be no meaningful way to define the process in that case.

For later use we also introduce the (causal) $\text{AR}(r)$ process $u_t = \varphi(B^{-1}) y_t$ or $\phi(B) u_t = \epsilon_t$ ($t = 1, 2, \dots$). Under the null hypothesis, $\pi(B) \Delta u_t = \epsilon_t$ and the former condition in (3) yields the conventional backward moving average representation

$$\Delta u_t = \sum_{j=0}^{\infty} \alpha_j \epsilon_{t-j}, \quad \alpha_0 = 1, \quad (6)$$

where the coefficients α_j of the power series representation of $\pi(z)^{-1}$ decay to zero at a geometric rate as $j \rightarrow \infty$ for $|z| \leq b_\phi$ and some $b_\phi > 1$. Thus, u_t is a nonstationary $\text{I}(1)$ process.

Finally, note that equation (1) and the conditions in (3) imply that there exist initial values such that the differenced process Δy_t has the two-sided moving average representation

$$\Delta y_t = \sum_{j=-\infty}^{\infty} \psi_j \epsilon_{t-j}, \quad (7)$$

where ψ_j is the coefficient of z^j in the Laurent series expansion of $\pi(z)^{-1} \varphi(z^{-1})^{-1} \stackrel{\text{def}}{=} \psi(z)$ so that $\psi(z) = \sum_{j=-\infty}^{\infty} \psi_j z^j$ exists for $b_\varphi \leq |z| \leq b_\phi$ with $b_\varphi < 1 < b_\phi$ defined above and with ψ_j decaying to zero at a geometric rate as $|j| \rightarrow \infty$. The representation (7) implies that Δy_t is a stationary and ergodic process with finite second moments. Hence, the invariance principle and weak convergence results of sample covariance matrices given in Phillips (1988) apply to y_t for any (random or nonrandom) initial value y_0 . This implies that the usual asymptotic results needed to develop limit theory for unit root tests are available. To simplify presentation we assume that, under the null hypothesis, the processes Δy_t and Δu_t are stationary and not only asymptotically stationary.

We derive a unit root test in a likelihood framework similar to that in L&S (for the employed assumptions, see also Andrews et al. (2006)). Thus, we impose the following assumption on the error term in (1).

Assumption 1. *The zero mean error term ϵ_t is a sequence of non-Gaussian IID random variables with a (Lebesgue) density $\sigma^{-1}f(\sigma^{-1}x; \lambda)$ which depends on the (finite and positive) error variance σ^2 and (possibly) on the parameter vector λ ($d \times 1$) taking values in an open set $\Lambda \subseteq \mathbb{R}^d$.*

As discussed in Breidt et al. (1991), Rosenblatt (2000, pp. 10-11), L&S, and others, causal and noncausal autoregressions are statistically indistinguishable if the error term (and hence the observed process) is Gaussian. This explains why Assumption 1 includes the requirement of non-Gaussian errors. Further assumptions on the density function $f(x; \lambda)$ will be made later.

3 Parameter estimation

3.1 Approximate likelihood function

To obtain our tests we first discuss the likelihood function based on the observed time series $\{y_1, \dots, y_T\}$ generated by the AR(r, s) process (1). Proceeding in the same way as in Section 3.1 of L&S suggests approximating the log-likelihood function by

$$l_T(\theta) = \sum_{t=r+1}^{T-s} g_t(\theta), \quad (8)$$

where

$$\begin{aligned} g_t(\theta) &= \log f(\sigma^{-1}(\Delta u_t(\varphi) - \phi u_{t-1}(\varphi) - \pi_1 \Delta u_{t-1}(\varphi) - \dots - \pi_{r-1} \Delta u_{t-r+1}(\varphi)); \lambda) - \log \sigma \\ &= \log f(\sigma^{-1}(v_t(\phi, \pi) - \varphi_1 v_{t+1}(\phi, \pi) - \dots - \varphi_s v_{t+s}(\phi, \pi)); \lambda) - \log \sigma. \end{aligned}$$

Here $u_t(\varphi)$ and $v_t(\phi, \pi)$ signify the series $u_t = \varphi(B^{-1})y_t$ and $v_t = \phi(B)y_t$, respectively, treated as functions of the parameters $\varphi = (\varphi_1, \dots, \varphi_s)$ and $(\phi, \pi) = (\phi, \pi_1, \dots, \pi_{r-1})$, and the parameter vector $\theta = (\phi, \pi, \varphi, \sigma, \lambda)$ ($(r + s + 1 + d) \times 1$) contains the parameters of the model. Maximizing $l_T(\theta)$ over permissible values of θ gives an (approximate) ML estimator of θ . In what follows, we drop the word ‘‘approximate’’ from the ML estimator and related quantities.

Above we assumed unrealistically that the orders of the model, r and s , are known. As in Breidt et al. (1991) and L&S we specify these orders in practice as follows. First, we fit a conventional causal AR model by LS and determine its order by using conventional procedures such as model selection criteria and residual diagnostics. We deem a causal model adequate when its residuals show no signs of autocorrelation. Due to the aforementioned identifiability issue we also need to check for the non-Gaussianity of the residuals because otherwise there is no point to consider noncausal models. If non-Gaussianity is supported by the data a non-Gaussian error distribution is adopted and all causal and noncausal models of the selected order are estimated. Of these models the one that maximizes the likelihood function is selected and its adequacy is checked by diagnostic tests.

In practice a purely noncausal model ($r = 0, s > 0$) may turn out to be the most appropriate choice but, due to the assumption $r > 0$, it is not in accordance with the assumed formulation. If one wants to perform a formal test in a case like this one may augment the model with a first-order causal polynomial and base the test on the AR(1, s) model.

3.2 Score vector and Hessian matrix

As our goal is to derive a Wald type test for the unit root hypothesis, we have to assume that the likelihood function satisfies conventional differentiability conditions similar to those used in the related previous work of Andrews et al. (2006) and L&S. Thus, we impose the following assumption.

Assumption 2. For all $(x, \lambda) \in (\mathbb{R}, \Lambda)$, $f(x; \lambda) > 0$ and $f(x; \lambda)$ is twice continuously differentiable with respect to (x, λ) and an even function of x , that is, $f(x; \lambda) = f(-x; \lambda)$.

Unlike the aforementioned previous authors we require that the function $f(\cdot; \lambda)$ is even. As will be discussed in Section 4.1, this assumption is imposed to simplify the limiting distribution of the obtained unit root test.

For the derivation of the Wald type test we need to estimate the unrestricted model and derive the limiting distribution of the ML estimator of ϕ under the null hypothesis. Because the data are assumed to be generated by a nonstationary I(1) process the derivation of the limiting distribution of the ML estimator involves features different from those in the previous literature on stationary NCAR models. Moreover, as the estimation problem is nonlinear, the presence of an I(1) process implies that methods used in the context of conventional unit root tests based on linear LS estimation are not directly applicable. Therefore, we use ideas similar to those developed for likelihood-based statistical models whose estimation theory is nonstandard in the sense that the information matrix is random even asymptotically. Such nonergodic models are discussed in Basawa and Scott (1983) and Jeganathan (1995) amongst others, and to facilitate their treatment we introduce the notation θ_0 for the true value of θ and similarly for its components. As the null hypothesis is assumed to hold, the true value of ϕ is zero.

We shall now derive weak limits of (appropriately standardized versions of) the score vector and Hessian matrix associated with the log-likelihood function evaluated at the true parameter value. We use a subscript to signify a partial derivative indicated by the subscript; for instance $g_{\theta,t}(\theta) = \partial g_t(\theta) / \partial \theta$, $f_x(x; \lambda) = \partial f(x; \lambda) / \partial x$, and $f_\lambda(x; \lambda) = \partial f(x; \lambda) / \partial \lambda$. Denote $V_{t+1} = (v_{t+1}, \dots, v_{t+s})$ and $\Delta U_{t-1} = (\Delta u_{t-1}, \dots, \Delta u_{t-r+1})$ where v_t are Δu_t have the representations (5) and (6) with the coefficients replaced by their true values $\beta_{0,j}$ and $\alpha_{0,j}$ so that the latter, for example, is obtained from $\pi_0(z)^{-1} = \sum_{j=0}^{\infty} \alpha_{0,j} z^j$. The first and second partial derivatives of $g_t(\theta)$, the log-likelihood based on a single observation, are presented in Appendix A. When evaluated at the true parameter value, the vector of first partial derivatives is

$$g_{\theta,t}(\theta_0) = \begin{bmatrix} g_{\phi,t}(\theta_0) \\ g_{\pi,t}(\theta_0) \\ g_{\varphi,t}(\theta_0) \\ g_{\sigma,t}(\theta_0) \\ g_{\lambda,t}(\theta_0) \end{bmatrix} = \begin{bmatrix} -\sigma_0^{-1} e_{x,t} u_{t-1} \\ -\sigma_0^{-1} e_{x,t} \Delta U_{t-1} \\ -\sigma_0^{-1} e_{x,t} V_{t+1} \\ -\sigma_0^{-2} (e_{x,t} \epsilon_t + \sigma_0) \\ e_{\lambda,t} \end{bmatrix}$$

where $e_{x,t} = f_x(\sigma_0^{-1} \epsilon_t; \lambda_0) / f(\sigma_0^{-1} \epsilon_t; \lambda_0)$ and $e_{\lambda,t} = f_\lambda(\sigma_0^{-1} \epsilon_t; \lambda_0) / f(\sigma_0^{-1} \epsilon_t; \lambda_0)$.

To obtain the weak limit of the score, we have to assume that the error density $f(x; \lambda)$ satisfies regularity conditions such as those employed by Andrews et al. (2006) and L&S. Rather than presenting the needed conditions explicitly we simplify the presentation by using suitable ‘‘high level’’ assumptions that can be verified by using the regularity conditions given in the aforementioned papers. To this end, it is convenient to write $\theta = (\phi, \vartheta) = (\phi, \vartheta_1, \vartheta_2)$ where $\vartheta_1 = (\pi, \varphi)$ and

$\vartheta_2 = (\sigma, \lambda)$. The score of ϑ (evaluated at θ_0) is clearly a stationary and ergodic process similar to the score in L&S. We make the following assumption.

Assumption 3. (i) $E[e_{x,t}] = 0$ and $E[e_{x,t}^2] = \mathcal{J}$, where $\mathcal{J} = \int (f_x(x; \lambda_0)^2 / f(x; \lambda_0)) dx > 1$ is finite. Moreover, $Cov[\epsilon_t, e_{x,t}] = -\sigma_0$.

(ii) The score vector $g_{\vartheta,t}(\theta_0) = (g_{\vartheta_1,t}(\theta_0), g_{\vartheta_2,t}(\theta_0))$ has zero expectation and finite positive definite covariance matrix $\Sigma = \text{diag}(\Sigma_1, \Sigma_2)$ where $\Sigma_i = Cov[g_{\vartheta_i,t}(\theta_0)]$ ($i = 1, 2$) and the partition is conformable to that of $g_{\vartheta,t}(\theta_0)$.

Part (i) of this assumption can be verified by using the definition of $e_{x,t}$, the regularity conditions in Andrews et al. (2006) and L&S, and direct calculation. Specifically, the expression of $Cov[\epsilon_t, e_{x,t}]$ is obtained from the definition of $e_{x,t}$ and condition (A2) of these papers, whereas condition (A5) implies that the inequality $\mathcal{J} > 1$ holds if and only if the distribution of ϵ_t is non-Gaussian. This inequality and the explicit expressions of the matrices Σ_1 and Σ_2 obtained from L&S can further be used to verify the positive definiteness of the covariance matrix Σ_1 in part (ii), whereas, due to the generality of the error distribution, the positive definiteness of Σ_2 has to be assumed. The other conditions in part (ii) can be verified by using the regularity conditions imposed on the density function $f(x; \lambda)$ in the aforementioned papers.

Assumption 3(i) and a standard functional central limit theorem for IID sequences yield

$$T^{-1/2} \sum_{t=1}^{\lfloor Tu \rfloor} (e_{x,t}, \epsilon_t) \xrightarrow{d} (B_{e_x}(u), B_\epsilon(u)) \sim BM \left(\begin{bmatrix} \mathcal{J} & -\sigma_0 \\ -\sigma_0 & \sigma_0^2 \end{bmatrix} \right), \quad (9)$$

where the covariance matrix is positive definite when ϵ_t is non-Gaussian. Using Assumptions 1-3 we can further derive the limiting distribution of the score vector of θ . The result is presented in the following lemma.

Lemma 1. *Suppose that Assumptions 1-3 hold. Then,*

$$T^{-1} \sum_{t=r+1}^{T-s} g_{\phi,t}(\theta_0) \xrightarrow{d} Z_1 = -\frac{1}{\sigma_0 \pi_0(1)} \int_0^1 B_\epsilon(u) dB_{e_x}(u) \quad (10)$$

and

$$T^{-1/2} \sum_{t=r+1}^{T-s} g_{\vartheta,t}(\theta_0) \xrightarrow{d} Z_2 \sim N(0, \Sigma). \quad (11)$$

Moreover, joint weak convergence applies with Z_1 and Z_2 independent.

The proof of this lemma is presented in Appendix B. As discussed therein, the requirement that the function $f(\cdot; \lambda)$ is even is needed to establish the independence statement (further discussion on this issue will be given at the end of Section 4.1).

Next consider the Hessian matrix associated with the log-likelihood function $l_T(\theta)$. Expressions for the required second partial derivatives are obtained from Appendix A. Similarly to the first partial derivatives we use notations such as $g_{\theta\theta,t}(\theta) = \partial^2 g_t(\theta) / \partial\theta\partial\theta'$, $f_{xx}(x; \lambda) = \partial^2 f(x; \lambda) / \partial x^2$, and $f_{x\lambda}(x; \lambda) = \partial^2 f(x; \lambda) / \partial x \partial \lambda$. We also define

$$e_{xx,t} = \frac{f_{xx}(\sigma_0^{-1}\epsilon_t; \lambda_0)}{f(\sigma_0^{-1}\epsilon_t; \lambda_0)} - e_{x,t}^2$$

and

$$e_{\lambda x,t} = \frac{f_{\lambda x}(\sigma_0^{-1}\epsilon_t; \lambda_0)}{f(\sigma_0^{-1}\epsilon_t; \lambda_0)} - \frac{f_{\lambda}(\sigma_0^{-1}\epsilon_t; \lambda_0)}{f(\sigma_0^{-1}\epsilon_t; \lambda_0)} e_{x,t},$$

and make the following assumption.

Assumption 4. $E[e_{xx,t}] = -E[e_{x,t}^2]$ and $E[g_{\vartheta\vartheta,t}(\theta_0)] = -\Sigma$ with Σ given in Assumption 3(ii). Moreover, $E[e_{xx,t}\epsilon_t] = 0$ and $E[e_{\lambda x,t}] = 0$.

Similarly to Assumption 3, this assumption can be verified by using the regularity conditions in Andrews et al. (2006) and L&S. The first moment equality is obtained from Assumption (A3) of these papers, whereas the second one states that the negative of the Hessian matrix of the log-likelihood function with respect to the short-run parameter ϑ equals the covariance matrix of the score of ϑ , a fact that can be established by direct calculation (see L&S). As for the last two moment conditions, both $e_{xx,t}\epsilon_t$ and $e_{\lambda x,t}$ are odd functions of ϵ_t so that, given Assumption 2, only finiteness of the expectations is required. This in turn can be obtained from condition (A7) of Andrews et al. (2006) and L&S.

Now we can prove the following lemma.

Lemma 2. *Suppose that Assumptions 1-4 hold. Then,*

$$-T^{-2} \sum_{t=r+1}^{T-s} g_{\phi\phi,t}(\theta_0) \xrightarrow{d} \frac{\mathcal{J}}{\sigma_0^2 \pi_0(1)^2} \int_0^1 B_{\epsilon}^2(u) d(u) \stackrel{def}{=} g_{\phi\phi}(\theta_0), \quad (12)$$

$$-T^{-1} \sum_{t=r+1}^{T-s} g_{\vartheta\vartheta,t}(\theta_0) \xrightarrow{p} \Sigma, \quad (13)$$

and

$$-T^{-3/2} \sum_{t=r+1}^{T-s} g_{\phi\vartheta,t}(\theta_0) \xrightarrow{p} 0. \quad (14)$$

Moreover, the weak convergences in (12) and in Lemma 1 hold jointly, and $g_{\phi\phi}(\theta_0)$ and Z_2 are independent.

Using the limits obtained in Lemmas 1 and 2 we define $Z = (Z_1, Z_2)$ and $G(\theta_0) = \text{diag}(g_{\phi\phi}(\theta_0), \Sigma)$, and we also introduce the matrix $D_T = \text{diag}(T, T^{-1/2}I_{r+s+d})$. The following proposition is an immediate consequence of Lemmas 1 and 2.

Proposition 1. *Suppose that Assumptions 1-4 hold. Then,*

$$S_T(\theta_0) \stackrel{def}{=} D_T^{-1} \sum_{t=r+1}^{T-s} g_{\theta,t}(\theta_0) \xrightarrow{d} Z \quad (15)$$

and

$$G_T(\theta_0) \stackrel{def}{=} -D_T^{-1} \sum_{t=r+1}^{T-s} g_{\theta\theta,t}(\theta_0) D_T^{-1} \xrightarrow{d} G(\theta_0), \quad (16)$$

where the weak convergences in (15) and (16) hold jointly with $(Z_1, G(\theta_0))$ and Z_2 independent.

In the next section we derive the limiting distribution of the ML estimator of the parameter θ by using Proposition 1 and arguments similar to those used by Basawa and Scott (1983, Ch. 2.4) in the context of statistical models whose likelihood ratios satisfy the LAMN condition.

3.3 Limiting distribution of the ML estimator

To obtain the limiting distribution of the ML estimator of the parameter θ we have to supplement the assumptions made so far by conditions on the standardized Hessian matrix $G_T(\theta) \stackrel{\text{def}}{=} -D_T^{-1} \sum_{t=r+1}^{T-s} g_{\theta\theta,t}(\theta) D_T^{-1}$. A sufficient “high level” condition, used by Basawa and Scott (1983, pp. 33-34) in a more general form, requires that, for all $c > 0$,

$$\sup_{\theta \in N_{T,c}} \|G_T(\theta) - G_T(\theta_0)\| \xrightarrow{p} 0, \quad (17)$$

where $N_{T,c} = \{\theta : D_T \|\theta - \theta_0\| \leq c\}$. As discussed in Appendix C, this condition can be verified by using assumptions similar to those used by Lii and Rosenblatt (1996) in the context of (stationary) noncausal and noninvertible ARMA models and by Meitz and Saikkonen (2013) in the context of a (stationary) noninvertible ARMA model with conditionally heteroskedastic errors. Proposition 1 combined with condition (17) enables us to establish the limiting distribution of the ML estimator of θ under the unit root hypothesis.

Proposition 2. *Suppose that Assumptions 1-4 and condition (17) hold. Then, with probability approaching one, there exists a sequence of local maximizers of the log-likelihood function $\hat{\theta}_T = (\hat{\phi}_T, \hat{\vartheta}_T)$ such that*

$$\left(D_T(\hat{\theta}_T - \theta_0), G_T(\theta_0) \right) \xrightarrow{d} \left(G(\theta_0)^{-1} Z, G(\theta_0) \right).$$

Moreover, $G_T(\hat{\theta}_T) - G_T(\theta_0) \xrightarrow{p} 0$.

Proposition 2 can be proved along the same lines as Theorems 1 and 2 of Basawa and Scott (1983, pp. 56-59). An outline of the needed arguments is provided in Appendix B. Now all ingredients for the derivation of our unit root tests are available.

4 Test procedures

4.1 Test statistic

With Proposition 2 at hand it is straightforward to derive Wald type unit root tests. As we are interested in one-sided (stationary) alternatives we use a “*t*-ratio” type test statistic defined as

$$\tau_T \stackrel{\text{def}}{=} \frac{\hat{\phi}_T}{\sqrt{G_T^{1,1}(\hat{\theta}_T)}},$$

where $G_T^{1,1}(\hat{\theta}_T)$ abbreviates the (1,1)-element of $G_T(\hat{\theta}_T)^{-1}$. The following proposition presents the asymptotic distribution of τ_T .

Proposition 3. *Suppose that Assumptions 1-4 and condition (17) hold. Then*

$$\tau_T \xrightarrow{d} \left(\mathcal{J} \int_0^1 W_\epsilon^2(u) d(u) \right)^{-1/2} \left(\int_0^1 W_\epsilon(u) dW_\epsilon(u) - (\mathcal{J} - 1)^{1/2} \int_0^1 W_\epsilon(u) dW(u) \right) \stackrel{\text{def}}{=} \tau(\mathcal{J}), \quad (18)$$

where $W_\epsilon(u) = \sigma_0^{-1} B_\epsilon(u) \sim BM(1)$, and $W(u) \sim BM(1)$ is independent of $W_\epsilon(u)$.

To see how this result can be obtained, note that Proposition 2 and the continuous mapping theorem yield

$$\tau_T \xrightarrow{d} - \left(\mathcal{J} \int_0^1 B_\epsilon^2(u) d(u) \right)^{-1/2} \int_0^1 B_\epsilon(u) dB_{e_x}(u).$$

The stated result is obtained by replacing the Brownian motion $B_{e_x}(u)$ on the right hand side by the expression

$$B_{e_x}(u) = -\sigma_0^{-1} B_\epsilon(u) + (\mathcal{J} - 1)^{1/2} W(u) = -W_\epsilon(u) + (\mathcal{J} - 1)^{1/2} W(u),$$

obtained via a Cholesky decomposition of the covariance matrix in (9).

Proposition 3 implies that the limiting distribution of test statistic τ_T is free of nuisance parameters except for the parameter \mathcal{J} . For subsequent analysis and discussions we notice that for Student's t -distributed errors with $\lambda > 2$ degrees of freedom

$$\mathcal{J} = \frac{\lambda(\lambda + 1)}{(\lambda - 2)(\lambda + 3)}. \quad (19)$$

Of course, the obtained limiting distribution is of limited practical use because it depends on the the nuisance parameter \mathcal{J} . Fortunately, this problem is rather easily circumvented and is further discussed in Section 5.1. The distribution of the limiting variable $\tau(\mathcal{J})$ is a weighted average of a standard normal distribution and a Dickey-Fuller type of distribution. More specifically, letting $\mathcal{J} \rightarrow \infty$ in (18) a standard normal distribution is obtained, as

$$\lim_{\mathcal{J} \rightarrow \infty} \tau(\mathcal{J}) = \left(\int_0^1 W_\epsilon^2(u) d(u) \right)^{-1/2} \int_0^1 W_\epsilon(u) dW(u) = \eta \sim N(0, 1),$$

where the second equality holds true because $\int_0^1 W_\epsilon(u) dW(u)$ is a scale mixture of normal distributions and can be written as $\int_0^1 W_\epsilon(u) dW(u) = \left(\int_0^1 W_\epsilon^2(u) d(u) \right)^{1/2} \eta$. On the other hand, letting $\mathcal{J} \rightarrow 1$ in (18) the Dickey-Fuller type of distribution is obtained, as

$$\lim_{\mathcal{J} \rightarrow 1} \tau(\mathcal{J}) = \left(\int_0^1 W_\epsilon^2(u) d(u) \right)^{-1/2} \int_0^1 W_\epsilon(u) dW_\epsilon(u).$$

That the limiting distribution of τ_T is relatively simple, depending only on the nuisance parameter \mathcal{J} , is achieved by assuming that the function $f(\cdot; \lambda)$ is even. This assumption is used to establish the independence of $g_{\phi\phi}(\theta_0)$ and Z_2 in Lemma 2, and further the independence of $(Z_1, G(\theta_0))$ and Z_2 in Proposition 1, and it is also used to justify the block diagonality of $G(\theta_0)$ (see the proof of Lemma 2 for some details). If these results do not hold the limiting distribution of τ_T will be a considerably more complicated function of the short-run parameters of the model, making the implementation of the resulting test very difficult.

4.2 Tests allowing for deterministic terms

The result of Proposition 3 only applies to mean-zero data. To accommodate series with trend components we consider the model

$$x_t = \mu + \delta t + y_t, \quad t = 1, 2, \dots,$$

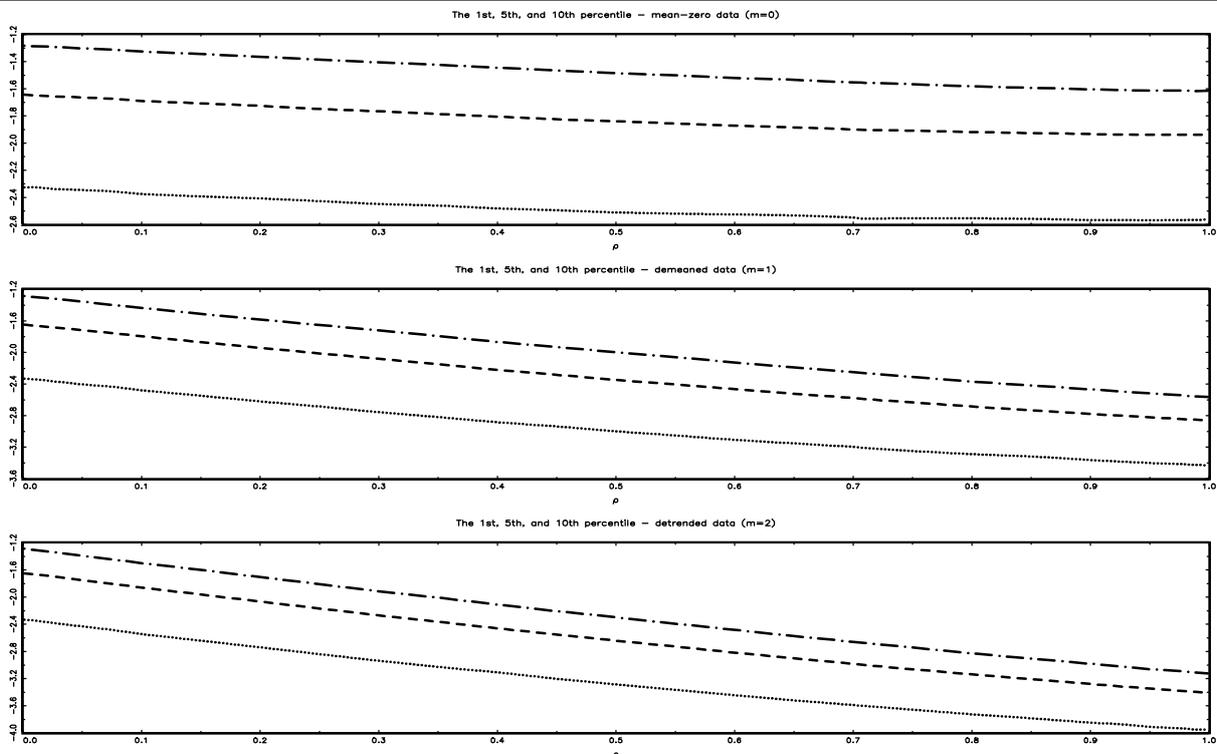
where x_t is the observed time series and y_t is a noncausal $AR(r, s)$ process. The trend coefficients μ and δ are estimated by LS to obtain the estimates $\hat{\mu}$ and $\hat{\delta}$ after which the test statistic τ_T introduced in the preceding section is formed by using $y_t = x_t - \hat{\mu}$ in the case of demeaned data and $y_t = x_t - \hat{\mu} - \hat{\delta}t$ in the case of detrended data. As in other unit root tests, the distribution of the resulting test statistic depends on the trend component chosen, and therefore we denote the test statistic by $\tau_T(m)$, where $m = 0$, $m = 1$, and $m = 2$ refer to mean-zero, demeaned, and detrended data, respectively. The result of Proposition 3 applies even for $\tau_T(1)$ and $\tau_T(2)$ as long as the Brownian motion $W_\epsilon(u)$ is replaced by corresponding detrended Brownian motion (see, e.g., Park and Phillips (1988)).

5 Simulation studies

5.1 Estimated critical values

The problem of the nuisance parameter \mathcal{J} ($\in (0, \infty)$) appearing in the limiting distribution of test statistic $\tau_T(m)$ is addressed next. We shall first illustrate how the value of the parameter \mathcal{J} affects the distribution of $\tau(\mathcal{J})$ (see (18)). It turns out to be convenient to study this effect by using the correlation between the two Brownian motions $B_\epsilon(u)$ and $B_{e_x}(u)$, that is, $\rho = \mathcal{J}^{-1/2} \in (0, 1)$ (see (9)). The following figure displays the 1% (dotted lines), 5% (dashed lines), and 10% (dashed-dotted lines) percentiles of the distribution of $\tau(\mathcal{J})$ as a function of ρ .

Figure 1 Percentiles of the distribution of $\tau(\mathcal{J})$ as a function of $\rho = \mathcal{J}^{-1/2}$



Notes: 1st percentiles (dotted lines), 5th percentiles (dashed lines), and 10th percentiles (dash-dotted lines) for the asymptotic distribution of the $\tau_T(m)$ statistic. The Brownian motions appearing in the limiting distribution of test statistic $\tau_T(m)$ are approximated using (appropriately scaled) sums of normal $IID(0, 1)$ variables with $T = 5,000$ and $500,000$ replications.

In Figure 1 a monotonically decreasing relationship between the percentiles and ρ is seen. As already mentioned, the Dickey-Fuller distributions and the standard normal distribution are obtained as limiting cases by letting $\mathcal{J} \rightarrow 1$ ($\rho \rightarrow 1$) and $\mathcal{J} \rightarrow \infty$ ($\rho \rightarrow 0$), respectively. Thus, in Figure 1 the 1%, 5%, and 10% critical values for the DF-statistics and a standard normal variate are found at the very left and very right, respectively.

Due to the percentiles monotonicity in ρ it is obvious that if the value of \mathcal{J} were known Figure 1 could be used to determine (conventional) critical values. Taking a more rigorous approach we proceed instead with curve estimation of the percentiles by fitting a second-order polynomial $cv_{\alpha,m}(\rho) = b_0 + b_1\rho + b_2\rho^2$ for $\alpha \in \{.01, .05, .10\}$ and $m \in \{0, 1, 2\}$. The curve estimates, obtained by LS, yield the coefficients in Table 1 that can be used to compute asymptotic critical values.

To exemplify how Table 1 can be used, assume that we wish to test the unit root hypothesis in the NCAR model at a 10% significance level in the case of demeaned data with $\mathcal{J} = 2$ ($\rho = 1/\sqrt{2}$). Then, the estimated asymptotic critical value equals $cv_{.10,1}(2) = -1.276 - 1.584 \times (1/\sqrt{2}) + .289 \times (1/\sqrt{2})^2 = -2.252$. To this end, the value of \mathcal{J} is in practice obviously not known and must be estimated. In the case of Student's t -distributed errors we can use equation (19) with the estimator $\hat{\lambda}$ used in place of λ . More generally, in cases where the distribution of the error term comprises less straightforward calculations of \mathcal{J} we may, by virtue of Assumption 3(i), use the estimator

$$\hat{\mathcal{J}} = \frac{1}{T - r - s} \sum_{t=r+1}^{T-s} \left[\frac{f_x(\hat{\sigma}^{-1}\hat{\epsilon}_t; \hat{\lambda})}{f(\hat{\sigma}^{-1}\hat{\epsilon}_t; \hat{\lambda})} \right]^2, \quad (20)$$

where $\hat{\epsilon}_t = \Delta\hat{u}_t - \hat{\phi}\hat{u}_{t-1} - \hat{\pi}_1\Delta\hat{u}_{t-1} - \dots - \hat{\pi}_{r-1}\Delta\hat{u}_{t-r+1}$ with $\hat{u}_t = \hat{\varphi}(B^{-1})y_t$.

Table 1 Coefficients to compute asymptotic critical values $cv_{\alpha,m}(\rho)$ of test statistic $\tau_T(m)$

Case	Significance level (α)	b_0	b_1	b_2	R^2
mean-zero data ($m = 0$)	1%	-2.321	-.492	.251	.998
	5%	-1.639	-.495	.187	.999
	10%	-1.276	-.480	.131	.999
demeaned data ($m = 1$)	1%	-2.322	-1.578	.474	1.00
	5%	-1.639	-1.591	.367	1.00
	10%	-1.276	-1.584	.289	1.00
detrended data ($m = 2$)	1%	-2.324	-2.201	.575	1.00
	5%	-1.640	-2.230	.462	1.00
	10%	-1.276	-2.231	.381	1.00

Notes: For each significance level and each trend specification the coefficients b_0 , b_1 , and b_2 are obtained from the regression of $cv_{\alpha,m}(\rho)$ on $(1, \rho, \rho^2)$ (using LS). R^2 is the regression coefficient of determination.

5.2 Empirical size and power simulations

In this section, we examine finite sample properties of the $\tau_T(m)$ -test for $m \in \{0, 1, 2\}$ by means of simulation experiments. The nominal significance level employed is 5%, and the benchmark process is a noncausal autoregressive process as defined in (1) with $r = s = 1$, and with the independent and identically distributed error term ϵ_t having Student's t -distribution with degrees of freedom λ equal to 3 and standard deviation σ equal to 0.1. A realization $\{y_1, \dots, y_T\}$ from this process is obtained by first generating the “noncausal” part as

$$v_t = \varphi_1 v_{t+1} + \epsilon_t, \quad t = T, T-1, \dots, 1,$$

where $v_{T+1} = 0$, and thereafter the “causal” part as

$$y_t = \phi_1 y_{t-1} + v_t, \quad t = 1, 2, \dots, T,$$

where $y_0 = 0$. To eliminate effects of the terminal and starting values, 100 observations at the end and beginning of each realization are discarded. Finally, in all experiments the true order of the process is assumed known (i.e. $r = s = 1$), and the estimation of the parameter $\hat{\theta} = (\hat{\phi}, \hat{\varphi}_1, \hat{\sigma}, \hat{\lambda})$ is carried out in GAUSS 12 using the BHHH algorithm in the CML library.

In the first experiment the empirical size of the $\tau_T(m)$ -test is examined when the parameter φ_1 is varied and estimated (asymptotic) critical values based on different estimates of \mathcal{J} are used. The parameter values and sample sizes considered are $\phi_1 = 1$ ($\phi = 0$), $\varphi_1 \in \{.10, .50, .90\}$ and $T \in \{100, 250\}$, respectively. Moreover, all the results in this experiment are based on 10,000 realizations of the $\{y_1, \dots, y_T\}$ process, and for each realization 5% critical values are obtained by the second-order polynomials in Table 2 using (19) with $\hat{\lambda}$ ($\hat{\mathcal{J}}_1$), the estimate in (20) ($\hat{\mathcal{J}}_2$), and $\mathcal{J} = 2$ (the true asymptotic value) as estimates. The outcomes of this experiment are reported in Table 2.

Table 2 Empirical size of the $\tau_T(m)$ -test

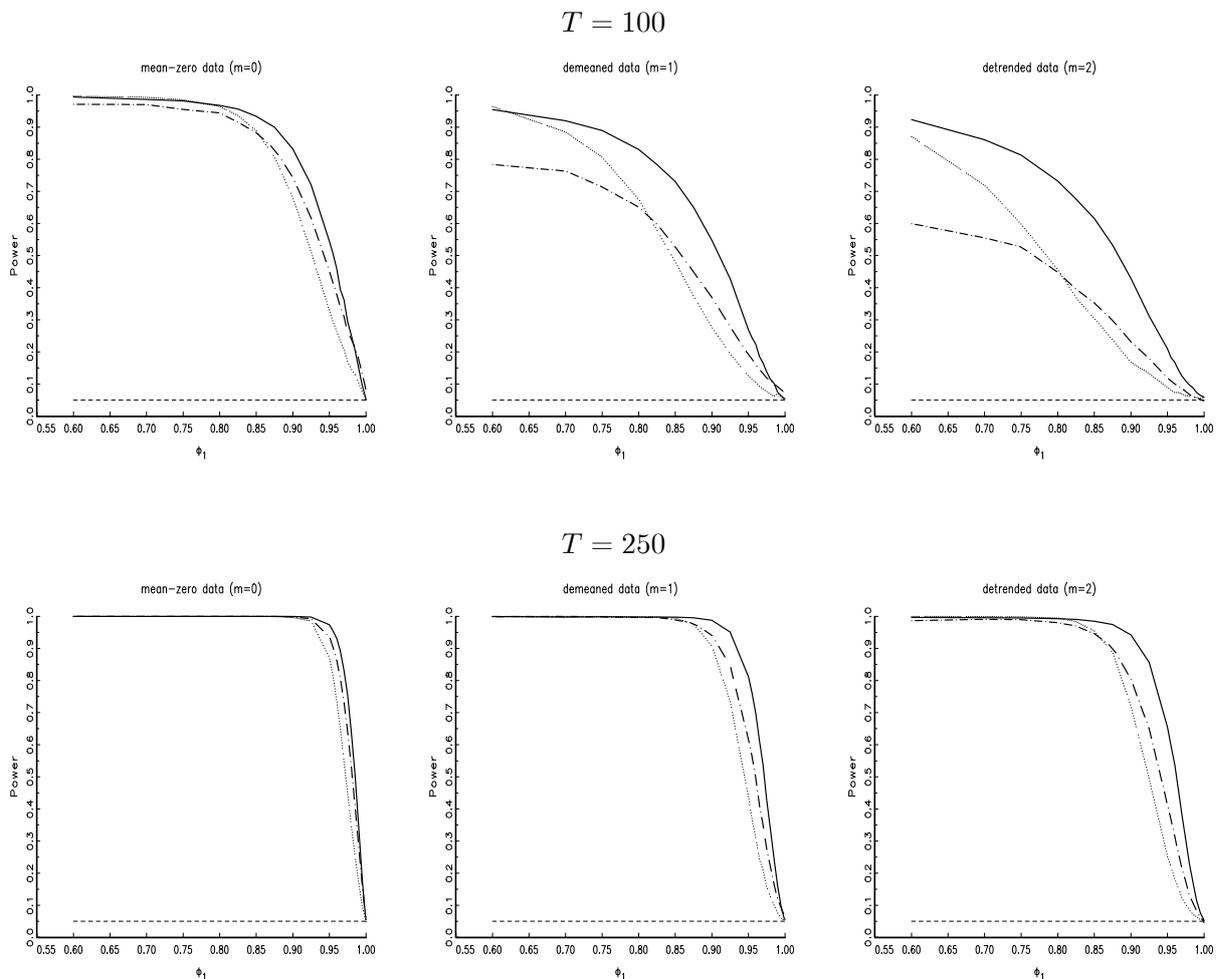
Sample Size		mean-zero data ($m = 0$)			demeaned data ($m = 1$)			detrended data ($m = 2$)		
		φ_1			φ_1			φ_1		
T		0.1	0.5	0.9	0.1	0.5	0.9	0.1	0.5	0.9
100	$\mathcal{J} = 2$.053	.053	.083	.059	.059	.086	.058	.056	.098
	$\hat{\mathcal{J}}_1 \approx 2.127$.052	.052	.083	.054	.054	.080	.056	.059	.097
	$\hat{\mathcal{J}}_2 \approx 2.122$.052	.052	.083	.054	.053	.080	.055	.059	.097
250	$\mathcal{J} = 2$.054	.047	.045	.063	.061	.055	.057	.052	.056
	$\hat{\mathcal{J}}_1 \approx 2.183$.053	.046	.044	.058	.058	.054	.056	.052	.057
	$\hat{\mathcal{J}}_2 \approx 2.181$.053	.046	.044	.058	.058	.054	.056	.052	.057

Notes: The results are based on 10,000 replications, and the nominal size of the tests is 5%. Reported estimated values for \mathcal{J} are based on the average value over the number of replications for each sample size in the case of demeaned data with $\varphi_1 = 0.5$.

In Table 2, the reported estimates for \mathcal{J} are (for each sample size) based on the average number of replications in the case of demeaned data with $\varphi_1 = .5$. It is seen that these estimates are close to the true value even for moderate sample sizes. For the other cases the estimates of \mathcal{J} are similar and therefore omitted. It is further noticed that the empirical size is close to the nominal size for most of the cases considered, and the influence of the parameter φ_1 appears to be modest. One exception, though, is for $T = 100$ and $\varphi_1 = .90$, where the test is somewhat over-sized so that some cautiousness is required. Taken the results in Table 2 together, it appears that the asymptotic distributions of the $\tau_T(m)$ -test, also with estimated \mathcal{J} -values, yield reasonable approximations to the finite sample distributions even for relatively small sample sizes, various trend components, and a wide range of parameter values for φ_1 .

In our second Monte Carlo experiment the power of the $\tau_T(m)$ -test is examined when $\hat{\mathcal{J}}_2$ is used to obtain critical values. The data are generated as described above with $\varphi_1 = .50$ and $\phi_1 \in [0.6, 1.0]$ ($\phi \in [-.4, 0]$). The sample sizes considered are $T \in \{100, 250\}$. For comparison we

Figure 2 Empirical power of the tests $\tau_T(m)$, $\tau_{DF}(m)$ and $M(m)$



Notes: $\tau_T(m)$ -test solid line, $\tau_{DF}(m)$ -test dotted line, and $M(m)$ -test dashed line, nominal size short-dashed line. The results are based on 10,000 replications and the nominal size of the tests is 5%.

we also choose to report the outcomes of the conventional Dickey-Fuller unit root t -test based on an AR(2) process, as well as the t -type unit root test of Lucas (1995) based M -estimation in AR(1) model assuming strictly stationary strong-mixing errors. The former and the latter tests are denoted by $\tau_{DF}(m)$ and $M(m)$, respectively.¹ The $\tau_{DF}(m)$ -test is a natural competitor to our test in that it is widely used among practitioners, and it has also been shown to be rather robust against various misspecifications. The $M(m)$ -test could also be viewed as a natural competitor, for it is designed to be robust against innovation outliers (fat-tailed distributions). The results of this experiment are summarized in Figure 2.

Figure 2 shows that the $\tau_T(m)$ -test is in general the most preferable test, and it performs significantly better than the $\tau_{DF}(m)$ and $M(m)$ -tests for most of the cases considered. For instance, in the case of detrended data with $T = 250$ and $\phi_1 = 0.95$, the differences in power between the $\tau_T(2)$ -test and the $\tau_{DF}(2)$ and $M(2)$ -tests are (approximately) as large as .40 and .25 units, respectively.

6 Empirical application

In this section, we provide an empirical illustration of our test by analyzing a Finnish interest rate series (Government bonds). These data range from 1988:Q1 to 2012:Q4 (quarterly observations) and yield a sample size of 100 observations.² The Government bond series is shown in Figure 3.

For interest rate series (in general) it is most natural to use demeaned data. But, as the Finnish interest rate series is trending in the sample we will also consider the case of detrended data. As a first step in our analysis we fit an AR(p) model to the data by LS and thereafter check if the residual series appears non-Gaussian. For the case of demeaned data both AIC and BIC select an AR(3) model, whereas for the case of detrended data an AR(2) model is selected by both AIC and

Figure 3 Finnish Government bonds



¹Following Lucas (1995) we use the Huber ψ -function $\psi(x) = \min\{c, \max(-c, x)\}$ with $c = 1.345$ to obtain the M -estimator. Furthermore, to operationalize the $M(m)$ -test, nuisance parameters are estimated by the Newey-West estimator with the lag-truncation parameter set at $\lceil 4(T/100)^{2/9} \rceil$. Finally, in the computations of the M -estimator a scale free version is used (see Lucas, 1995, p. 337), and an iterative weighted LS algorithm (similar to the one described in Van Dijk, Franses, and Lucas, 1999, p. 219) is applied.

²The series are obtained from IMF's *International Financial Statistics*.

BIC.³ Even though the null hypothesis of no 4th-order remaining serial correlation is not rejected by the Ljung-Box (LB) test for the two residual series (p -values: .492 and .811 for demeaned and detrended residual series, respectively), we find that the normality assumption is strongly rejected by the Lomnicki, Jarque, and Bera (LJB) test (p -values: $< .001$ and $< .001$ for demeaned and detrended residual series, respectively), and some evidence of 4th-order ARCH effects are also found by the McLeod-Li (McL) test (p -values: .089 and .250 for demeaned and detrended residual series, respectively).⁴ In addition, quantile-quantile plots of the residuals of the AR(3) and AR(2) models (not shown here) indicate that a normal distribution is not appropriate because excess kurtosis in the data is left unexplained. Taken these results together it seems worthwhile to proceed with estimation and unit root testing of NCAR specifications, and to capture the leptokurtic behavior of the residuals series we will adopt t -distributed errors. More specifically, for demeaned and detrended data we consider an AR(r, s) model with $r + s = p = 3$ and $r + s = p = 2$, respectively, and conduct unit root testing for the AR(1, 2) and AR(2, 1) specifications in the former case and for the AR(1, 1) specification in the latter case. As will be discussed below, these noncausal models are supported by the specification strategy discussed in Section 3.1. For comparison we also employ the M -test and the τ_{DF} -test based on AR(3) and AR(2) models in the case of demeaned and detrended data, respectively. The outcomes of these unit root tests as well as various estimation results and LB, LJB, and McL misspecification tests are reported in Tables 3 and 4 below.⁵

Table 3 Unit root testing for demeaned and detrended Finnish interest rate series

$T^* = 97$			$m = 1$		
Model	Test	Outcome	$cv_{.05,1}$	$cv_{.05,1}(\hat{\rho}_1)$	$cv_{.05,1}(\hat{\rho}_2)$
AR(3)	τ_{DF}	-1.041	-2.860		
AR(1)	M	.038	-3.060		
AR(2, 1)- t	τ_T	-5.101***		-2.558	-2.577
AR(1, 2)- t	τ_T	-5.997***		-2.542	-2.531
$T^* = 98$			$m = 2$		
Model	Test	Outcome	$cv_{.05,2}$	$cv_{.05,2}(\hat{\rho}_1)$	$cv_{.05,2}(\hat{\rho}_2)$
AR(2)	τ_{DF}	-3.213*	-3.410		
AR(1)	M	-1.435	-3.660		
AR(1, 1)- t	τ_T	-5.379***		-3.072	-3.048

Notes: T^* is the effective sample size. AR(r, s) abbreviates an autoregressive model with r th-order and s th-order polynomials $\phi(B)$ and $\varphi(B^{-1})$, respectively. N and t refer to Gaussian and t -distributed errors. *** and * denote significance at the 1% and the 10% level, respectively. $cv_{.05,m}$ ($m = 1, 2$) is the 5% critical value for the τ_{DF} -test (see Fuller (1976, p. 373)) and the M -test (see Lucas (1995, p. 339)), and $cv_{.05,m}(\hat{\rho}_1)$ and $cv_{.05,m}(\hat{\rho}_2)$ ($m = 1, 2$) are the 5% estimated critical values obtained by letting $\hat{\rho}_1 = \hat{\mathcal{J}}_1^{-1/2}$ and $\hat{\rho}_2 = \hat{\mathcal{J}}_2^{-1/2}$ ($\hat{\mathcal{J}}_1$ and $\hat{\mathcal{J}}_2$ are estimators of \mathcal{J} using (19) and (20), respectively).

³In this exercise we choose the maximum lag considered as $p_{\max} = \lceil 4(T/100)^{2/9} \rceil = 4$.

⁴The skewness part of the LJB-test is significant at 7.6% and 8.8 % levels for demeaned and detrended data, respectively, indicating that the rejection of Gaussian errors mainly stems from the kurtosis part of the LJB-test.

⁵As discussed in L&S (see p. 12), we use least absolute deviation estimators to find starting values for ϕ and φ_1 ($\tilde{\phi}$ and $\tilde{\varphi}_1$, say), and thereafter maximize $l_T(\tilde{\phi}, \tilde{\varphi}_1, \sigma, \lambda)$ to also find starting values for σ and λ .

Table 4 NCAR estimation results for demeaned and detrended Finnish interest rate series

$m = 1$		AR(r, s) specifications				
$T^* = 97$	AR(3, 0)- N	AR(3, 0)- t	AR(2, 1)- t	AR(1, 2)- t	AR(0, 3)- t	
ϕ	-.015 (.013)	.003 (.013)	-.472 (.093)	-.425 (.071)		
π_1	.568 (.099)	.590 (.091)	.056 (.091)			
π_2	-.180 (.099)	-.084 (.104)				
φ_1			.941 (.025)	.873 (.068)	1.550 (.118)	
φ_2				.056 (.067)	-.695 (.160)	
φ_3					.130 (.093)	
σ	.447 (.170)	.457 (.085)	.477 (.126)	.498 (.155)	.461 (.069)	
λ		3.585 (1.356)	3.017 (.987)	2.831 (.868)	4.129 (1.731)	
LL	-59.556	-49.111	-46.052	-45.637	-53.437	
LB(4)	.492	.324	.674	.348	.522	
McL(4)	.089	.233	.256	.306	< .001	
LJB	< .001					
$m = 2$		AR(r, s) specifications				
$T^* = 98$	AR(2, 0)- N	AR(2, 0)- t	AR(1, 1)- t	AR(0, 2)- t		
ϕ	-.104 (.032)	-.052 (.032)	-0.418 (.078)			
π_1	.505 (.086)	.563 (.085)				
φ_1			.806 (.057)	1.445 (.108)		
φ_2				-.521 (.103)		
σ	.433 (.031)	.442 (.073)	.440 (.091)	.451 (.071)		
λ		3.861 (1.581)	3.293 (1.058)	3.966 (1.596)		
LL	-57.055	-48.998	-43.480	-51.528		
LB(4)	.811	.358	.503	.584		
McL(4)	.250	.257	.338	< .001		
LJB	< .001					

Notes: T^* designates the effective sample. AR(r, s) abbreviates an autoregressive model with r th-order and s th-order polynomials $\phi(B)$ and $\varphi(B^{-1})$, respectively. N and t refer to Gaussian and t -distributed errors, respectively. The figures in parentheses are standard errors, and they are computed using the square root of the diagonal elements of the matrix $D_T^{-1}G_T(\hat{\theta}_T)^{-1}D_T^{-1}$. LL denotes the value of the log-likelihood evaluated at the ML estimates. LB(4) and McL(4) signify the Ljung-Box test and the McLeod-Li test with four lags, and LJB is the Lomnicki, Jarque, and Bera normality test, and the p -value of these tests are reported.

In Table 3 we see that the τ_T -tests based on NCAR specifications give rise to strong rejections (in the case of demeaned data the strongest rejection arises from the NCAR(1, 2) model). On the other hand, the τ_{DF} -test fails to reject the unit root hypothesis in the case of demeaned data and yields a rejection only at the 10% level using detrended data, whereas the M -test strongly supports the unit root hypothesis irrespective of detrending procedure.

By the results in Table 4 we conclude that in the case of demeaned data the log-likelihood (LL) is maximized for an AR(1, 2) model with t -distributed errors, albeit closely followed by an AR(2, 1) model with t -distributed errors. In the case of detrended data the LL is maximized for an AR(1, 1) model with t -distributed errors. The maximized LL's of causal and pure noncausal models are substantially lower than those of the aforementioned three NCAR models. Moreover, according to the LB test and the McL test these NCAR models are satisfactory, and an inspection of their quantile-quantile plots (not shown here) lends support to the choice of t -distributed errors. The estimation accuracy of the parameters appears reasonable perhaps with the exception of the degrees of freedom parameter λ whose small point estimates still clearly point to a non-Gaussian error distribution. Altogether, the results in Tables 3 and 4 suggest that a stationary NCAR process with non-Gaussian leptokurtic errors provides a reasonable approximation for the Finnish interest rate series.

7 Concluding Remarks

In this paper, we consider likelihood-based unit root tests in the NCAR model of L&S. In order to derive such tests, we derive asymptotic properties of the ML estimator under the unit root hypothesis. Most of the employed assumptions are common to those in the previous literature on noncausal autoregressions, but to obtain unit root tests with feasible limiting distributions a symmetric error distribution is adopted. With this assumptions the limiting distributions of our tests are shown to depend on a single nuisance parameter the effect of which can readily be eliminated by using estimated critical values.

In our simulation study we demonstrate that the size properties of our tests are satisfactory and that the use of estimated critical values has no appreciable effect on the size properties of the tests. Our simulations also show that, compared to conventional DF-tests and the M -tests of Lucas (1995), the power of our tests against stationary NCAR alternatives is very good.

In an application to a Finnish interest rate series we demonstrate how the tests are implemented in practice. In this application we find that NCAR models with t -distributed errors provide better descriptions to the data than causal AR models and that our tests clearly reject the unit root hypothesis, whereas no evidence or only weak evidence against a unit root is obtained by using DF-tests and the M -tests.

Appendix A: partial derivatives of the log-likelihood function

First partial derivatives. As in Section 3.2, we use a subscript to denote a partial derivative indicated by the subscript. For notational brevity, denote

$$e_{x,t}(\theta) = \frac{f_x(\sigma^{-1}\epsilon_t(\theta); \lambda)}{f(\sigma^{-1}\epsilon_t(\theta); \lambda)} \quad \text{and} \quad e_{\lambda,t}(\theta) = \frac{f_\lambda(\sigma^{-1}\epsilon_t(\theta); \lambda)}{f(\sigma^{-1}\epsilon_t(\theta); \lambda)}$$

where $\epsilon_t(\theta) = \phi(B)\varphi(B^{-1})y_t$ is treated as a function of the parameters ϕ and φ (for notational convenience we do not make explicit that $\epsilon_t(\theta)$ is independent of the parameters σ and λ). Note also the alternative expressions

$$\begin{aligned} \epsilon_t(\theta) &= \Delta u_t(\varphi) - \phi u_{t-1}(\varphi) - \pi_1 \Delta u_{t-1}(\varphi) - \cdots - \pi_{r-1} \Delta u_{t-r+1}(\varphi) \\ &= (v_t(\phi, \pi) - \varphi_1 v_{t+1}(\phi, \pi) - \cdots - \varphi_s v_{t+s}(\phi, \pi)), \end{aligned}$$

where $u_t(\varphi) = \varphi(B^{-1})y_t$ and $v_t(\phi, \pi) = \phi(B)y_t$ (see (2)). We also set $U_{t-1}(\varphi) = (u_{t-1}(\varphi), \dots, u_{t-r+1}(\varphi))$ and $V_{t+1}(\phi, \pi) = (v_{t+1}(\phi, \pi), \dots, v_{t+s}(\phi, \pi))$. With straightforward differentiation one now obtains from (8)

$$g_{\theta,t}(\theta) = \begin{bmatrix} g_{\phi,t}(\theta) \\ g_{\pi,t}(\theta) \\ g_{\varphi,t}(\theta) \\ g_{\sigma,t}(\theta) \\ g_{\lambda,t}(\theta) \end{bmatrix} = \begin{bmatrix} -\sigma^{-1}e_{x,t}(\theta)u_{t-1}(\varphi) \\ -\sigma^{-1}e_{x,t}(\theta)\Delta U_{t-1}(\varphi) \\ -\sigma^{-1}e_{x,t}(\theta)V_{t+1}(\phi, \pi) \\ -\sigma^{-2}(e_{x,t}(\theta)\epsilon_t(\theta) + \sigma) \\ e_{\lambda,t}(\theta) \end{bmatrix}$$

Second partial derivatives. To simplify notation, define

$$\begin{aligned} e_{xx,t}(\theta) &= \frac{f_{xx}(\sigma^{-1}\epsilon_t(\theta); \lambda)}{f(\sigma^{-1}\epsilon_t(\theta); \lambda)} - \left(\frac{f_x(\sigma^{-1}\epsilon_t(\theta); \lambda)}{f(\sigma^{-1}\epsilon_t(\theta); \lambda)} \right)^2 \\ e_{\lambda x,t}(\theta) &= \frac{f_{\lambda x}(\sigma^{-1}\epsilon_t(\theta); \lambda)}{f(\sigma^{-1}\epsilon_t(\theta); \lambda)} - \frac{f_\lambda(\sigma^{-1}\epsilon_t(\theta); \lambda)}{f(\sigma^{-1}\epsilon_t(\theta); \lambda)} \frac{f_x(\sigma^{-1}\epsilon_t(\theta); \lambda)}{f(\sigma^{-1}\epsilon_t(\theta); \lambda)} \\ e_{\lambda\lambda,t}(\theta) &= \frac{f_{\lambda\lambda}(\sigma^{-1}\epsilon_t(\theta); \lambda)}{f(\sigma^{-1}\epsilon_t(\theta); \lambda)} - \frac{f_\lambda(\sigma^{-1}\epsilon_t(\theta); \lambda)}{f(\sigma^{-1}\epsilon_t(\theta); \lambda)} \frac{f_\lambda(\sigma^{-1}\epsilon_t(\theta); \lambda)'}{f(\sigma^{-1}\epsilon_t(\theta); \lambda)}, \end{aligned}$$

and let Y_t stand for the $(r-1) \times s$ matrix with elements y_{t-i+j} ($i = 1, \dots, r-1, j = 1, \dots, s$) whereas Y'_{1t} ($1 \times s$) is used to signify the first row of Y_t .

The following five second partial derivatives involve the long-run parameter ϕ :

$$\begin{aligned} g_{\phi\phi,t}(\theta) &= \sigma^{-2}e_{xx,t}(\theta)u_{t-1}^2(\varphi) \\ g_{\pi\phi,t}(\theta) &= \sigma^{-2}e_{xx,t}(\theta)u_{t-1}(\varphi)\Delta U_{t-1}(\varphi) \\ g_{\varphi\phi,t}(\theta) &= \sigma^{-2}e_{xx,t}(\theta)u_{t-1}(\varphi)V_{t+1}(\phi, \pi) + \sigma^{-1}e_{x,t}(\theta)Y_{1t} \\ g_{\sigma\phi,t}(\theta) &= \sigma^{-3}e_{xx,t}(\theta)\epsilon_t(\theta)u_{t-1}(\varphi) + \sigma^{-2}e_{x,t}(\theta)u_{t-1}(\varphi) \\ g_{\lambda\phi,t}(\theta) &= -\sigma^{-1}u_{t-1}(\varphi)e_{\lambda x,t}(\theta). \end{aligned}$$

The remaining partial derivatives are as in L&S expect that r is replaced by $r-1$ and differences

of $U_{t-1}(\varphi)$ and Y_t are used. We have

$$\begin{aligned}
g_{\pi\pi,t}(\theta) &= \sigma^{-2}e_{xx,t}(\theta)\Delta U_{t-1}(\varphi)\Delta U'_{t-1}(\varphi) \\
g_{\varphi\varphi,t}(\theta) &= \sigma^{-2}e_{xx,t}(\theta)V_{t+1}(\phi,\pi)V'_{t+1}(\phi,\pi) \\
g_{\sigma\sigma,t}(\theta) &= 2\sigma^{-3}e_{x,t}(\theta)\epsilon_t(\theta) + \sigma^{-4}e_{xx,t}(\theta)\epsilon_t^2(\theta) + \sigma^{-2} \\
g_{\lambda\lambda,t}(\theta) &= e_{\lambda\lambda,t}(\theta) \\
g_{\pi\varphi,t}(\theta) &= \sigma^{-2}e_{xx,t}(\theta)\Delta U_{t-1}(\varphi)V'_{t+1}(\phi,\pi) + \sigma^{-1}e_{x,t}(\theta)\Delta Y_t \\
g_{\pi\sigma,t}(\theta) &= \sigma^{-3}e_{xx,t}(\theta)\epsilon_t(\theta)\Delta U_{t-1}(\varphi) + \sigma^{-2}e_{x,t}(\theta)\Delta U_{t-1}(\varphi) \\
g_{\varphi\sigma,t}(\theta) &= \sigma^{-3}e_{xx,t}(\theta)\epsilon_t(\theta)V_{t+1}(\phi,\pi) + \sigma^{-2}e_{x,t}(\theta)V_{t+1}(\phi,\pi) \\
g_{\lambda\pi,t}(\theta) &= -\sigma^{-1}e_{\lambda x,t}(\theta)\Delta U'_{t-1}(\varphi) \\
g_{\lambda\varphi,t}(\theta) &= -\sigma^{-1}e_{\lambda x,t}(\theta)V'_{t+1}(\phi,\pi) \\
g_{\lambda\sigma,t}(\theta) &= -\sigma^{-2}e_{\lambda x,t}(\theta)\epsilon_t(\theta).
\end{aligned}$$

These partial derivatives form the matrix $g_{\vartheta\vartheta,t}(\theta)$.

Appendix B: proofs for Section 3

Proof of Lemma 1. Recall the notation $\Delta u_t = \sum_{j=0}^{\infty} \alpha_{0,j}\epsilon_{t-j}$ and $v_t = \sum_{j=0}^{\infty} \beta_{0,j}\epsilon_{t-j}$. Using the expression of the components of $g_{\theta,t}(\theta_0)$ one obtains (10) as a straightforward application of the theorem in Phillips (1988) whereas (11) can be justified by a central limit theorem for stationary ergodic processes (see Breidt et al. (1991) or L&S). The stated joint convergence holds because both $g_{\phi,t}(\theta_0)$ and $g_{\vartheta,t}(\theta_0)$ are generated by the same IID sequence ϵ_t .

To establish the independence of the limits, let \mathcal{F}_t^ϵ signify the the σ -algebra generated by $(\epsilon_t, \epsilon_{t-1}, \dots)$ and consider the random vector $w_t = (e_{x,t}, \epsilon_t, g_{\vartheta,t}(\theta_0))$. Using the definitions of $e_{x,t}$ and $g_{\vartheta,t}(\theta_0)$, and Assumption 3 it is straightforward to check that $(w_t, \mathcal{F}_t^\epsilon)$ is an L_2 -mixingale with size -1 (for the definition of an L_2 -mixingale and its size, see Davidson (1994, p. 247)). Thus, Theorem 3 of Scott (1973) shows that a functional central limit theorem applies to w_t and the resulting limiting Brownian motion, $Q(\cdot)$ say, determines the limits in (10) and (11). Thus, it suffices to show that the first two components of $Q(\cdot)$ are independent of the remaining components, which in turn follows if the long-run covariance matrix between the component vectors $(e_{x,t}, \epsilon_t)$ and $g_{\vartheta,t}(\theta_0)$ of w_t is block diagonal.

That the long-run covariance matrix between $(e_{x,t}, \epsilon_t)$ and $(g_{\pi,t}(\theta_0), g_{\varphi,t}(\theta_0))$ is block diagonal is easy to see. For instance, $g_{\varphi,t}(\theta_0) = -\sigma_0^{-1}e_{x,t}V_{t+1}$ is uncorrelated with $(e_{x,t}, \epsilon_t)$ because V_{t+1} is independent of $(e_{x,t}, \epsilon_t)$ and has zero mean, and $g_{\pi,t}(\theta_0)$ is also uncorrelated with $(e_{x,t+k}, \epsilon_{t+k})$, $k \neq 0$, because $e_{x,t}$ is independent of $(V_{t+1}, e_{x,t+k}, \epsilon_{t+k})$ and has zero mean. This implies the desired result and a similar reasoning applies to $g_{\pi,t}(\theta_0) = -\sigma_0^{-1}e_{x,t}\Delta U_{t-1}$.

The long-run covariance matrix between $(e_{x,t}, \epsilon_t)$ and $(g_{\sigma,t}(\theta_0), g_{\lambda,t}(\theta_0))$ reduces to the ordinary covariance matrix. Assumption 2 and the definition of $e_{x,t}$ imply that $e_{x,t}$ is an odd function of ϵ_t . Thus, as $E[g_{\sigma,t}(\theta_0)\epsilon_t] = -\sigma_0^{-3}E[e_{x,t}\epsilon_t^2]$, the fact that $e_{x,t}\epsilon_t^2$ is an odd function of ϵ_t yields $E[g_{\sigma,t}(\theta_0)\epsilon_t] = 0$. In the same way it is seen that $E[g_{\sigma,t}(\theta_0)e_{x,t}] = -\sigma_0^{-3}E[e_{x,t}^2\epsilon_t] = 0$. Furthermore, $g_{\lambda,t}(\theta_0)$ is an even function of ϵ_t so that $E[g_{\lambda,t}(\theta_0)\epsilon_t] = 0$ and $E[g_{\lambda,t}(\theta_0)e_{x,t}] = 0$. Thus, we have shown that the long-run covariance matrix between (e_t, ϵ_t) and $g_{\vartheta,t}(\theta_0)$ is block diagonal.

We still note that in some of the foregoing derivations the assumption that the function $f(x; \lambda)$ is even is needed. For instance, consider the expectation

$$\begin{aligned} E [g_{\sigma,t}(\theta_0) e_{x,t}] &= -\sigma_0^{-3} \int x \left(\frac{f_x(\sigma_0^{-1}x; \lambda_0)}{f(\sigma_0^{-1}x; \lambda_0)} \right)^2 f(\sigma_0^{-1}x; \lambda_0) dx \\ &= -\sigma_0^{-2} \int z \frac{(f_x(z; \lambda_0))^2}{f(z; \lambda_0)} dz, \end{aligned}$$

which need not vanish if the function $f(\cdot; \lambda)$ is not even.

Proof of Lemma 2. First conclude from the definitions that the elements of the matrix $g_{\theta\theta,t}(\theta_0)$ (see Appendix A) are (jointly) stationary and ergodic so that an application of a law of large numbers yields (13). Thus, it suffices to consider (12) and (14). We use the well-known fact that a functional central limit theorem applies to the process Δu_t (see (6)) and that $T^{-1/2}u_{[T\cdot]} \xrightarrow{d} \pi_0(1)^{-1} B_\epsilon(\cdot)$ (see the discussion at the beginning of the proof of the theorem in Phillips (1988)). Similarly it is seen that $T^{-1/2}y_{[T\cdot]}$ obeys a functional central limit theorem.

As $e_{xx,t} = e_{xx,t}(\theta_0)$ the expression of $g_{\phi\phi,t}(\theta)$ in Appendix A yields

$$g_{\phi\phi,t}(\theta_0) = \sigma_0^{-2} \frac{f_{xx}(\sigma_0^{-1}\epsilon_t; \lambda_0)}{f(\sigma_0^{-1}\epsilon_t; \lambda_0)} u_{t-1}^2 - \sigma_0^{-2} e_{x,t}^2 u_{t-1}^2,$$

where

$$E \left(\frac{f_{xx}(\sigma_0^{-1}\epsilon_t; \lambda_0)}{f(\sigma_0^{-1}\epsilon_t; \lambda_0)} \right) = 0$$

by Assumption 4(i). As $T^{-1/2}u_{[T\cdot]}$ converges weakly, the continuous mapping theorem implies that the same is true for $(T^{-1/2}u_{[T\cdot]})^2$, so that from Theorem 3.3 of Hansen (1992) we obtain

$$T^{-2} \sum_{t=r+1}^{T-s} \frac{f_{xx}(\sigma_0^{-1}\epsilon_t; \lambda_0)}{f(\sigma_0^{-1}\epsilon_t; \lambda_0)} u_{t-1}^2 = o_p(1).$$

Hence,

$$\begin{aligned} T^{-2} \sum_{t=r+1}^{T-s} g_{\phi\phi,t}(\theta_0) &= -\sigma_0^{-2} \mathcal{J} T^{-2} \sum_{t=r+1}^{T-s} u_{t-1}^2 - \sigma_0^{-2} T^{-2} \sum_{t=r+1}^{T-s} (e_{x,t}^2 - \mathcal{J}) u_{t-1}^2 + o_p(1) \\ &= -\sigma_0^{-2} \mathcal{J} T^{-2} \sum_{t=r+1}^{T-s} u_{t-1}^2 + o_p(1), \end{aligned}$$

where the second equality can again be justified by Theorem 3.3 of Hansen (1992) and the assumption $E[e_{x,t}^2] = \mathcal{J}$. Thus, as $T^{-1/2}u_{[T\cdot]} \xrightarrow{d} \pi_0(1)^{-1} B_\epsilon(\cdot)$, an application of the continuous mapping theorem yields (12).

To establish (14), first consider

$$g_{\pi\phi,t}(\theta_0) = \sigma_0^{-2} \frac{f_{xx}(\sigma_0^{-1}\epsilon_t; \lambda_0)}{f(\sigma_0^{-1}\epsilon_t; \lambda_0)} u_{t-1} \Delta U_{t-1} - e_{x,t}^2 u_{t-1} \Delta U_{t-1},$$

where

$$E \left(\frac{f_{xx}(\sigma_0^{-1}\epsilon_t; \lambda_0)}{f(\sigma_0^{-1}\epsilon_t; \lambda_0)} \Delta U_{t-1} \right) = 0 \quad \text{and} \quad E(e_{x,t}^2 \Delta U_{t-1}) = 0$$

because $E(\Delta U_{t-1}) = 0$ and ΔU_{t-1} is independent of $(\epsilon_t, e_{x,t})$. Thus, Theorem 3.3 of Hansen (1992) can again be used to obtain

$$T^{-3/2} \sum_{t=r+1}^{T-s} g_{\pi\phi,t}(\theta_0) = o_p(1).$$

As $T^{-1/2}y_{[T\cdot]}$ converges weakly, arguments similar to those already used yield

$$\begin{aligned} T^{-3/2} \sum_{t=r+1}^{T-s} g_{\varphi\phi,t}(\theta) &= \sigma_0^{-2} T^{-3/2} \sum_{t=r+1}^{T-s} e_{xx,t} u_{t-1} V_{t+1} + \sigma_0^{-1} T^{-3/2} \sum_{t=r+1}^{T-s} e_{x,t} Y_{1t} \\ &= o_p(1). \end{aligned}$$

Note, however, that the summand in the first term on the right hand side of the former equation differs from its counterpart in $g_{\pi\phi,t}(\theta_0)$ in that, unlike ΔU_{t-1} , the zero mean stationary process V_{t+1} depends on ϵ_{t+j} , $j > 0$. Hence, the process $e_{xx,t} V_{t+1}$ is not adapted to \mathcal{F}_t^ϵ and does not satisfy the assumptions imposed on the sequence $\{e_i\}$ in Hansen's (1992) Theorem 3.3. However, an inspection of the proof of that theorem reveals that this theorem applies even if $\{e_i\}$ is a general mixingale such as $e_{xx,t} V_{t+1}$ instead of the assumed special type of mixingale (that $e_{xx,t} V_{t+1}$ is a mixingale, see Davidson (1994, p. 247)).

Finally, arguments similar to those used for $g_{\pi\phi,t}(\theta_0)$ apply to $g_{\sigma\phi,t}(\theta_0)$ and $g_{\lambda\phi,t}(\theta_0)$. Because $E[e_{xx,t}\epsilon_t] = 0$, $E[e_{x,t}] = 0$, and $E[e_{\lambda x,t}] = 0$ by Assumption 4, Hansen's (1992) Theorem 3.3 can be used to obtain

$$T^{-3/2} \sum_{t=r+1}^{T-s} g_{\sigma\phi,t}(\theta_0) = o_p(1) \quad \text{and} \quad T^{-3/2} \sum_{t=r+1}^{T-s} g_{\lambda\phi,t}(\theta_0) = o_p(1).$$

Thus, we have established (14). Note that here we also need the assumption that the function $f(x; \lambda)$ is even which, for instance, guarantees that $E[e_{xx,t}\epsilon_t] = 0$, and hence the applicability of Hansen's (1992) Theorem 3.3.

The stated independence follows in the same way as in the proof of Lemma 1 because the long-run covariance matrix between ϵ_t and $g_{\theta,t}(\theta_0)$ is block diagonal (here the assumption $E[e_{xx,t}\epsilon_t] = 0$ or that the function $f(\cdot; \lambda)$ is even is also needed). This completes the proof of the lemma.

Proof of Proposition 2. As indicated after Theorem 1, the proof follows the arguments in Basawa and Scott (1983, pp. 56-59) with suitable modifications. In particular, our probability statements assume that the true parameter value θ_0 is a fixed point in the parameter space so that, instead of uniform and continuous weak convergence employed by Basawa and Scott (1983, pp. 56-59), we employ ordinary weak convergence. A minor modification is that in place of the scaling matrix $I_n(\theta)$ in Assumptions (B.2) and (B.3) of Basawa and Scott (1983, pp. 33-34) we have the matrix D_T that is independent of the parameter θ . It is straightforward to check that, as far as the proof is concerned, this replacement has no essential effect. The same applies to the fact that in our case the limits Z and $G(\theta_0)$ (see Proposition 1) are not independent, implying that the LAMN condition does not hold.

Now, to see how the proof proceeds, note first that, with the preceding modifications, we can repeat the arguments in the proof of Theorem 1 of Basawa and Scott (1983, pp. 56-58) and conclude that, with probability approaching one, there exists a local ML estimator $\hat{\theta}_T$ such that

$D_T(\hat{\theta}_T - \theta_0) = O_p(1)$. Next, as in the proof of Theorem 2 of Basawa and Scott (1983, pp. 58-59) we find that

$$D_T(\hat{\theta}_T - \theta_0) - G_T^{-1}(\theta_0) S_T(\theta_0) = o_p(1).$$

The first result of the theorem now follows from Proposition 1 and the continuous mapping theorem. The second one can be established by using arguments similar to those used to prove Proposition 3.2 of Saikkonen (1993, p. 161) modified as above to concern ordinary weak convergence instead of continuous weak convergence (see also Remark 3.1 of Saikkonen (1993, p. 161)).

Appendix C: sufficient conditions for assumption (17)

To provide a discussion on sufficient conditions for the ‘‘high level’’ condition (17) we first note that the regularity conditions employed by Andrews et al. (2006) and L&S can be used to show that this condition holds for the lower right hand block of the Hessian, $\sum_{t=r+1}^{T-s} g_{\vartheta\vartheta,t}(\theta)$. Thus, assuming these conditions, it suffices to consider the blocks involving the parameter ϕ , that is $g_{\phi\phi,t}(\theta)$ and $g_{\vartheta\phi,t}(\theta)$ (in what follows the null hypothesis of a unit root will also be assumed)

We denote by Λ_0 a neighborhood of λ_0 and assume that, for all $x \in \mathbb{R}$, $\Delta x \in \mathbb{R}$, and $\lambda \in \Lambda_0$, and for some $C < \infty$ and $d_1, d_2 > 0$,

$$|w(x + \Delta x; \lambda) - w(x; \lambda)| \leq C \left[(1 + |x|^{d_1}) |\Delta x| + |\Delta x|^{d_2} \right] \quad (21)$$

for the following choices of the function $w(x; \lambda)$:

$$w(x; \lambda) = \begin{cases} \frac{f_x(x; \lambda)}{f(x; \lambda)} \\ \frac{f_{xx}(x; \lambda)}{f(x; \lambda)} - \left(\frac{f_x(x; \lambda)}{f(x; \lambda)} \right)^2 \\ \frac{f_{\lambda x}(x; \lambda)}{f(x; \lambda)} - \frac{f_\lambda(x; \lambda)}{f(x; \lambda)} \frac{f_x(x; \lambda)}{f(x; \lambda)}. \end{cases}$$

This assumption is an analogue of Assumption B of Lii and Rosenblatt (1996) who used it with the first two choices of $w(x; \lambda)$ and with $f(x; \lambda)$ independent of λ in the context of (stationary) noncausal and noninvertible ARMA models. The third choice is a simplified analog of Assumption 7 of Meitz and Saikkonen (2013) who developed an estimation theory for a (stationary) noninvertible ARMA model with conditionally heteroskedastic errors. Note that replacing the argument x with $\sigma^{-1}\epsilon_t(\theta)$ the three choices of $w(x; \lambda)$ become $e_{x,t}(\theta)$, $e_{xx,t}(\theta)$ and $e_{\lambda x,t}(\theta)$ (see Appendix A), and for the second one, for instance, condition (21) implies

$$|e_{xx,t}(\theta) - e_{xx,t}(\theta_0)| \leq C \left[\left(1 + |\sigma_0^{-1}\epsilon_t(\theta_0)|^{d_1} \right) |\sigma^{-1}\epsilon_t(\theta) - \sigma_0^{-1}\epsilon_t(\theta_0)| + |\sigma^{-1}\epsilon_t(\theta) - \sigma_0^{-1}\epsilon_t(\theta_0)|^{d_2} \right], \quad (22)$$

where $\epsilon_t(\theta_0) = \epsilon_t$, and $E[|\epsilon_t|^{2+d_1}] < \infty$ and $E[|\epsilon_t|^{1+d_2}] < \infty$ is assumed.

It suffices to establish analogs of condition (17) for $g_{\kappa\phi,t}(\theta)$, $\kappa = \phi, \pi, \varphi, \sigma, \lambda$, or to show that, for all $c > 0$,

$$\sup_{\theta \in N_{T,c}} \left\| T^{-b_\kappa} \sum_{t=r+1}^{T-s} [g_{\kappa\phi,t}(\theta) - g_{\kappa\phi,t}(\theta_0)] \right\| \xrightarrow{P} 0, \quad \kappa = \phi, \pi, \varphi, \sigma, \lambda, \quad (23)$$

where $b_\kappa = 2$ for $\kappa = \phi$ and $b_\kappa = 3/2$ otherwise. Unless otherwise stated, $\theta \in N_{T,c}$ will be assumed. We only establish (23) for $\kappa = \phi, \varphi$. The other cases can be handled with similar arguments.

Case $\kappa = \phi$. First note that (see Appendix A)

$$\begin{aligned}
T^{-2} \sum_{t=r+1}^{T-s} [g_{\phi\phi,t}(\theta) - g_{\phi\phi,t}(\theta_0)] &= T^{-2} \sum_{t=r+1}^{T-s} [\sigma^{-2} e_{xx,t}(\theta) u_{t-1}^2(\varphi) - \sigma_0^{-2} e_{xx,t}(\theta_0) u_{t-1}^2(\varphi_0)] \\
&= T^{-2} \sum_{t=r+1}^{T-s} \sigma_0^{-2} e_{xx,t}(\theta_0) [u_{t-1}^2(\varphi) - u_{t-1}^2(\varphi_0)] \\
&\quad + T^{-2} \sum_{t=r+1}^{T-s} [\sigma^{-2} e_{xx,t}(\theta) - \sigma_0^{-2} e_{xx,t}(\theta_0)] u_{t-1}^2(\varphi) \\
&\stackrel{def}{=} A_T^{(1)}(\theta) + A_T^{(2)}(\theta).
\end{aligned}$$

Now consider $A_T^{(1)}(\theta)$ and note that

$$u_{t-1}^2(\varphi) - u_{t-1}^2(\varphi_0) = 2u_{t-1}(\varphi_0)(u_{t-1}(\varphi) - u_{t-1}(\varphi_0)) + (u_{t-1}(\varphi) - u_{t-1}(\varphi_0))^2,$$

where

$$u_{t-1}(\varphi) - u_{t-1}(\varphi_0) = \varphi(B^{-1})y_t - \varphi_0(B^{-1})y_t = Y'_{1,t+1}(\varphi - \varphi_0)$$

with $Y_{1,t+1} = [y_{t+1} \cdots y_{t+s}]'$. Thus,

$$\begin{aligned}
\sup_{\theta \in N_{T,c}} \left| A_T^{(1)}(\theta) \right| &\leq \sup_{\theta \in N_{T,c}} T^{-2} \sum_{t=r+1}^{T-s} |\sigma_0^{-2} e_{xx,t}(\theta_0)| |u_{t-1}^2(\varphi) - u_{t-1}^2(\varphi_0)| \\
&\leq \frac{2}{\sigma_0^2} \sup_{\theta \in N_{T,c}} T^{-2} \sum_{t=r+1}^{T-s} |e_{xx,t}(\theta_0)| |u_{t-1}(\varphi_0)| \|Y_{1,t+1}\| \|\varphi - \varphi_0\| \\
&\quad + \frac{1}{\sigma_0^2} \sup_{\theta \in N_{T,c}} T^{-2} \sum_{t=r+1}^{T-s} |e_{xx,t}(\theta_0)| \|Y_{1,t+1}\|^2 \|\varphi - \varphi_0\|^2 \\
&\stackrel{def}{=} A_T^{(1,1)} + A_T^{(1,2)}.
\end{aligned}$$

As a functional central limit theorem applies to $T^{-1/2}y_{[T]}$, we have $\max_{1 \leq t \leq T} |T^{-1/2}y_t| = O_p(1)$. Thus, $\max_{1 \leq t \leq T} \|T^{-1/2}Y_{1,t+1}\| = O_p(1)$ and, as $u_t(\varphi_0) = y_t - \varphi_{0,1}y_{t+1} - \cdots - \varphi_{0,s}y_{t+s}$, also $\max_{1 \leq t \leq T} |T^{-1/2}u_t(\varphi_0)| = O_p(1)$. Because $\sup_{\theta \in N_{T,c}} \|\varphi - \varphi_0\| \leq c/T^{1/2}$ we get, for all c ,

$$A_T^{(1,1)} \leq \frac{2c}{\sigma_0^2} \max_{1 \leq t \leq T} |T^{-1/2}u_t(\varphi_0)| \max_{1 \leq t \leq T} \|T^{-1/2}Y_{1,t+1}\| T^{-3/2} \sum_{t=r+1}^{T-s} |e_{xx,t}(\theta_0)| = o_p(1).$$

Here the equality follows because $|e_{xx,t}(\theta_0)|$ has finite expectation by Assumptions 3(i) and 4. Similar arguments also yield $A_T^{(1,2)} = o_p(1)$ so that altogether we have shown that

$$\sup_{\theta \in N_{T,c}} \left| A_T^{(1)}(\theta) \right| = o_p(1).$$

To obtain a similar result for $A_T^{(2)}(\theta)$ note first that, by the definition of $u_t(\varphi)$, $\max_{1 \leq t \leq T} |T^{-1/2}u_t(\varphi)| \leq C_1 \max_{1 \leq t \leq T+s} |T^{-1/2}y_t| = O_p(1)$ for a finite C_1 independent of φ . Thus,

$$\sup_{\theta \in N_{T,c}} \left| A_T^{(2)}(\theta) \right| \leq C_1^2 \max_{1 \leq t \leq T} (T^{-1}y_t^2) \sup_{\theta \in N_{T,c}} T^{-1} \sum_{t=r+1}^{T-s} |\sigma^{-2} e_{xx,t}(\theta) - \sigma_0^{-2} e_{xx,t}(\theta_0)|,$$

so that it suffices to consider

$$\begin{aligned} \sup_{\theta \in N_{T,c}} T^{-1} \sum_{t=r+1}^{T-s} |\sigma^{-2} e_{xx,t}(\theta) - \sigma_0^{-2} e_{xx,t}(\theta_0)| &\leq \sup_{\theta \in N_{T,c}} |\sigma^{-2} - \sigma_0^{-2}| T^{-1} \sum_{t=r+1}^{T-s} |e_{xx,t}(\theta_0)| \quad (24) \\ &+ \sup_{\theta \in N_{T,c}} \sigma^{-2} T^{-1} \sum_{t=r+1}^{T-s} |e_{xx,t}(\theta) - e_{xx,t}(\theta_0)| \\ &\stackrel{def}{=} A_T^{(2,1)} + A_T^{(2,2)}. \end{aligned}$$

We need to establish that $A_T^{(2,i)} = o_p(1)$, $i = 1, 2$. As $\sup_{\theta \in N_{T,c}} |\sigma^{-2} - \sigma_0^{-2}| = o(1)$, $A_T^{(2,1)} = o_p(1)$ immediately follows from the aforementioned fact that $E[|e_{xx,t}(\theta_0)|] < \infty$. A proof for $A_T^{(2,2)} = o_p(1)$ is obtained by using the inequality (22) and arguments similar to those above. To provide an idea of the needed arguments, first notice that (see Appendix A)

$$\epsilon_t(\theta) = \Delta u_t(\varphi) - \sum_{j=1}^{r-1} \pi_j \Delta u_{t-j}(\varphi) - \phi u_{t-1}(\varphi) \stackrel{def}{=} \tilde{\epsilon}_t(\theta) - \phi u_{t-1}(\varphi),$$

where $u_t(\varphi) = y_t - \varphi_1 y_{t+1} - \dots - \varphi_s y_{t+s}$ with $\varphi_1, \dots, \varphi_s$ belonging to a bounded set. Thus, as $|\phi| \leq c/T$ for $\theta \in N_{T,c}$, $\max_{1 \leq t \leq T} |\phi u_t(\varphi)| = O_p(T^{-1/2})$ holds uniformly in $N_{T,c}$ and

$$|\sigma^{-1} \epsilon_t(\theta) - \sigma_0^{-1} \epsilon_t(\theta_0)| \leq \sigma^{-1} |\tilde{\epsilon}_t(\theta) - \epsilon_t(\theta_0)| + |\sigma^{-1} - \sigma_0^{-1}| |\epsilon_t(\theta_0)| + O_p(T^{-1/2}),$$

where the term $O_p(T^{-1/2})$ is uniform over $1 \leq t \leq T$ and $\theta \in N_{T,c}$. As $|a_1 + a_2|^d \leq M_d (|a_1|^d + |a_2|^d)$, $M_d < \infty$, for any real numbers a_1, a_2 and $d > 0$, the desired result $A_T^{(2,2)} = o_p(1)$ can be obtained by applying the inequality (22) with the term $|\sigma^{-1} \epsilon_t(\theta) - \sigma_0^{-1} \epsilon_t(\theta_0)|$ on the right hand side replaced by $\sigma^{-1} |\tilde{\epsilon}_t(\theta) - \epsilon_t(\theta_0)| + |\sigma^{-1} - \sigma_0^{-1}| |\epsilon_t(\theta_0)|$. As $E[|\epsilon_t(\theta_0)|] = E[|\epsilon_t|] < \infty$, arguments similar to those used for the first term on the right hand side of (24) show that here the contribution of latter term is of order $o_p(1)$ so that we can replace $|\sigma^{-1} \epsilon_t(\theta) - \sigma_0^{-1} \epsilon_t(\theta_0)|$ on the right hand side of (22) by $\sigma^{-1} |\tilde{\epsilon}_t(\theta) - \epsilon_t(\theta_0)|$ and obtain

$$A_T^{(2,2)} \leq C \sup_{\theta \in N_{T,c}} \sigma^{-2} T^{-1} \sum_{t=r+1}^{T-s} \left[\left(1 + |\sigma_0^{-1} \epsilon_t(\theta_0)|^{d_1}\right) \sigma^{-1} |\tilde{\epsilon}_t(\theta) - \epsilon_t(\theta_0)| + \sigma^{-d_2} |\tilde{\epsilon}_t(\theta) - \epsilon_t(\theta_0)|^{d_2} \right] + o_p(1).$$

Here $\tilde{\epsilon}_t(\theta) - \epsilon_t(\theta_0) = -\sum_{j=1}^{r-1} (\pi_j - \pi_{0,j}) \Delta u_{t-j}(\varphi)$ and $\sup_{\theta \in N_{T,c}} \|\pi - \pi_0\| \leq c/T^{-1/2}$. Thus, as $\Delta u_t(\varphi) = \Delta y_t - \varphi_1 \Delta y_{t+1} - \dots - \varphi_s \Delta y_{t+s}$ with $\varphi_1, \dots, \varphi_s$ belonging to a bounded set, it follows that the majorant of $A_T^{(2,2)}$ is of order $o_p(1)$ (note that here use is also made of the assumptions $E[|\epsilon_t|^{1+d_1}] < \infty$ and $E[|\epsilon_t|^{d_2}] < \infty$ implied by their more stringent counterparts mentioned below (22)). Thus, we have established (23) with $\kappa = \phi$.

Case $\kappa = \varphi$. By the definition of $g_{\varphi\phi,t}(\theta)$ (see Appendix A)

$$\begin{aligned} &T^{-3/2} \sum_{t=r+1}^{T-s} [g_{\varphi\phi,t}(\theta) - g_{\varphi\phi,t}(\theta_0)] \\ &= T^{-3/2} \sum_{t=r+1}^{T-s} [\sigma^{-2} e_{xx,t}(\theta) u_{t-1}(\varphi) V_{t+1}(\phi, \pi) - \sigma_0^{-2} e_{xx,t}(\theta_0) u_{t-1}(\varphi_0) V_{t+1}(\phi_0, \pi_0)] \\ &\quad + T^{-3/2} \sum_{t=r+1}^{T-s} [\sigma^{-1} e_{x,t}(\theta) - \sigma_0^{-1} e_{x,t}(\theta_0)] Y_{1t} \\ &\stackrel{def}{=} B_T^{(1)}(\theta) + B_T^{(2)}(\theta). \end{aligned}$$

We need to show that $\sup_{\theta \in N_{T,c}} |B_T^{(i)}(\theta)| = o_p(1)$, $i = 1, 2$. First consider

$$\begin{aligned} \sup_{\theta \in N_{T,c}} |B_T^{(2)}(\theta)| &\leq \sup_{\theta \in N_{T,c}} T^{-3/2} \sum_{t=r+1}^{T-s} |\sigma^{-1}e_{x,t}(\theta) - \sigma_0^{-1}e_{x,t}(\theta_0)| \|Y_{1,t+1}\| \\ &\leq \max_{1 \leq t \leq T} \|T^{-1/2}Y_{1,t+1}\| \sup_{\theta \in N_{T,c}} T^{-1} \sum_{t=r+1}^{T-s} |\sigma^{-1}e_{x,t}(\theta) - \sigma_0^{-1}e_{x,t}(\theta_0)|, \end{aligned}$$

where $\max_{1 \leq t \leq T} \|T^{-1/2}Y_{1,t+1}\| = O_p(1)$, as already noted. That the supremum in the last expression is of order $o_p(1)$ can be established by using arguments similar to those used for $A_T^{(2)}(\theta)$ above, so details are omitted.

As for $B_T^{(1)}(\theta)$, we write

$$\begin{aligned} B_T^{(1)}(\theta) &= T^{-3/2} \sum_{t=r+1}^{T-s} \sigma_0^{-2}e_{xx,t}(\theta_0) [u_{t-1}(\varphi) V_{t+1}(\phi, \pi) - u_{t-1}(\varphi_0) V_{t+1}(\phi_0, \pi_0)] \\ &\quad + T^{-3/2} \sum_{t=r+1}^{T-s} [\sigma^{-2}e_{xx,t}(\theta) - \sigma_0^{-2}e_{xx,t}(\theta_0)] u_{t-1}(\varphi) V_{t+1}(\phi, \pi) \\ &\stackrel{def}{=} B_T^{(1,1)}(\theta) + B_T^{(1,2)}(\theta), \end{aligned}$$

and demonstrate that $\sup_{\theta \in N_{T,c}} |B_T^{(1,i)}(\theta)| = o_p(1)$, $i = 1, 2$.

First consider $B_T^{(1,1)}(\theta)$ and recall that $V_{t+1}(\phi, \pi) = (v_{t+1}(\phi, \pi), \dots, v_{t+s}(\phi, \pi))$ with $v_t(\phi, \pi) = \phi(B)y_t$ or

$$v_t(\phi, \pi) = \Delta y_t - \sum_{j=1}^{r-1} \pi_j \Delta y_{t-j} - \phi y_{t-1} \stackrel{def}{=} \tilde{v}_t(\pi) - \phi y_{t-1}$$

(see (4)). Define $\tilde{V}_{t+1}(\pi) = (\tilde{v}_{t+1}(\pi), \dots, \tilde{v}_{t+s}(\pi))$ so that $V_{t+1}(\phi, \pi) = \tilde{V}_{t+1}(\pi) - \phi y_{t-1} \mathbf{1}_s$ where $\mathbf{1}_s = (1, \dots, 1)$ ($s \times 1$). Clearly, $\max_{1 \leq t \leq T} \|V_t(\phi, \pi) - \tilde{V}_t(\pi)\| = \sqrt{s} \max_{1 \leq t \leq T} |\phi y_t| = O_p(T^{-1/2})$ uniformly in $N_{T,c}$, and $\tilde{V}_{t+1}(\pi_0) = V_{t+1}(\phi_0, \pi_0)$ is stationary. As we also have $E[\|\epsilon_{xx,t}(\theta_0)\|] < \infty$ and $\max_{1 \leq t \leq T} |T^{-1/2}u_t(\varphi)| = O_p(1)$ uniformly in $N_{T,c}$, we can replace $V_{t+1}(\phi, \pi)$ in $B_T^{(1,1)}(\theta)$ by $\tilde{V}_{t+1}(\pi)$ and consider

$$\begin{aligned} \sup_{\theta \in N_{T,c}} \|\tilde{B}_T^{(1,1)}(\theta)\| &= \sup_{\theta \in N_{T,c}} \left\| T^{-3/2} \sum_{t=r+1}^{T-s} \sigma_0^{-2}e_{xx,t}(\theta_0) \left[u_{t-1}(\varphi) \tilde{V}_{t+1}(\pi) - u_{t-1}(\varphi_0) V_{t+1}(\phi_0, \pi_0) \right] \right\| \\ &\leq \sup_{\theta \in N_{T,c}} \left\| T^{-3/2} \sum_{t=r+1}^{T-s} \sigma_0^{-2}e_{xx,t}(\theta_0) u_{t-1}(\varphi_0) \left[\tilde{V}_{t+1}(\pi) - V_{t+1}(\phi_0, \pi_0) \right] \right\| \\ &\quad + \sup_{\theta \in N_{T,c}} \left\| T^{-3/2} \sum_{t=r+1}^{T-s} \sigma_0^{-2}e_{xx,t}(\theta_0) [u_{t-1}(\varphi) - u_{t-1}(\varphi_0)] \tilde{V}_{t+1}(\pi) \right\| \\ &\stackrel{def}{=} \tilde{B}_T^{(1,3)} + \tilde{B}_T^{(1,4)}. \end{aligned}$$

First consider $\tilde{B}_T^{(1,4)}$ and recall that $u_{t-1}(\varphi) - u_{t-1}(\varphi_0) = Y'_{1,t+1}(\varphi - \varphi_0)$ and $\sup_{\theta \in N_{T,c}} \|\varphi - \varphi_0\| \leq$

$c/T^{1/2}$ (see the treatment of $A_T^{(1)}(\theta)$ in case $\kappa = \phi$). Thus,

$$\begin{aligned}\tilde{B}_T^{(1,4)} &\leq \sigma_0^{-2} \sup_{\theta \in N_{T,c}} T^{-3/2} \sum_{t=r+1}^{T-s} |e_{xx,t}(\theta_0)| \|Y_{1,t+1}\| \|\varphi - \varphi_0\| \left\| \tilde{V}_{t+1}(\pi) \right\| \\ &\leq c\sigma_0^{-2} \max_{1 \leq t \leq T} \left\| T^{-1/2} Y_{1,t+1} \right\| \sup_{\theta \in N_{T,c}} T^{-3/2} \sum_{t=r+1}^{T-s} |e_{xx,t}(\theta_0)| \left\| \tilde{V}_{t+1}(\pi) \right\|,\end{aligned}$$

where $\max_{1 \leq t \leq T} \left\| T^{-1/2} Y_{1,t+1} \right\| = O_p(1)$. That the last term is of order $o_p(1)$ now follows from the following two facts: (i) The components of $\tilde{V}_{t+1}(\pi)$ satisfy $|\tilde{v}_{t+i}(\pi)| \leq C_2 (|\Delta y_{t+i}| + \dots + |\Delta y_{t+i-r+1}|)$ for some finite constant C_2 independent of π and (ii) using condition (A7) in L&S one can show that $E[|e_{xx,t}(\theta_0) \Delta y_{t+i-j}|] < \infty$ ($i = 1, \dots, s, j = 1, \dots, r-1$).

Now consider $\tilde{B}_T^{(1,3)}$ for which we have

$$\begin{aligned}\tilde{B}_T^{(1,3)} &\leq \sigma_0^{-2} \sup_{\theta \in N_{T,c}} T^{-3/2} \sum_{t=r+1}^{T-s} |e_{xx,t}(\theta_0)| |u_{t-1}(\varphi_0)| \left\| \tilde{V}_{t+1}(\pi) - V_{t+1}(\phi_0, \pi_0) \right\| \\ &\leq \sigma_0^{-2} \max_{1 \leq t \leq T} \left| T^{-1/2} u_t(\varphi_0) \right| \sup_{\theta \in N_{T,c}} T^{-1} \sum_{t=r+1}^{T-s} |e_{xx,t}(\theta_0)| \left\| \tilde{V}_{t+1}(\pi) - V_{t+1}(\phi_0, \pi_0) \right\|,\end{aligned}$$

where $\max_{1 \leq t \leq T} \left| T^{-1/2} u_t(\varphi_0) \right| = O_p(1)$ as seen in the case $\kappa = \phi$. That the last term is of order $o_p(1)$ follows from the fact that the components of the vector $\tilde{V}_{t+1}(\pi) - V_{t+1}(\phi_0, \pi_0)$ are $\sum_{j=1}^{r-1} (\pi_j - \pi_{0,j}) \Delta y_{t+i-j}$ with $\sup_{\theta \in N_{T,c}} \|\pi - \pi_0\| \leq c/T^{1/2}$, and $E[|e_{xx,t}(\theta_0) \Delta y_{t+i-j}|] < \infty$ ($i = 1, \dots, s, j = 1, \dots, r-1$), as noted above. Thus, we have established $\sup_{\theta \in N_{T,c}} \left| B_T^{(1,1)}(\theta) \right| = o_p(1)$, and we still need to obtain a similar result for $B_T^{(1,2)}(\theta)$.

By using arguments similar to those used for $B_T^{(1,1)}(\theta)$ it can first be seen that we can replace σ_0^{-2} in the definition of $B_T^{(1,2)}(\theta)$ by σ^{-2} , so that denoting the resulting quantity by $\tilde{B}_T^{(1,2)}(\theta)$ we can consider

$$\begin{aligned}\sup_{\theta \in N_{T,c}} \left\| \tilde{B}_T^{(1,2)}(\theta) \right\| &= \sup_{\theta \in N_{T,c}} \left\| \sigma^{-2} T^{-3/2} \sum_{t=r+1}^{T-s} [e_{xx,t}(\theta) - e_{xx,t}(\theta_0)] u_{t-1}(\varphi) V_{t+1}(\phi, \pi) \right\| \\ &\leq C_3 \max_{1 \leq t \leq T} \left| T^{-1/2} y_t \right| \sup_{\theta \in N_{T,c}} T^{-1} \sum_{t=r+1}^{T-s} |e_{xx,t}(\theta) - e_{xx,t}(\theta_0)| \|V_{t+1}(\phi, \pi)\|,\end{aligned}$$

where $C_3 < \infty$ is independent of θ . Here the inequality is obtained by dominating $|u_{t-1}(\varphi)|$ in the same way as in handling $A_T^{(2)}(\theta)$. As $\max_{1 \leq t \leq T} \left| T^{-1/2} y_t \right| = O_p(1)$ it suffices to show that the supremum in the last expression is of order $o_p(1)$. As seen above, $\max_{1 \leq t \leq T} \left\| V_t(\phi, \pi) - \tilde{V}_t(\pi) \right\| = O_p(T^{-1/2})$ which in conjunction with condition (A7) of L&S can be used to show that we can here replace $V_t(\phi, \pi)$ with the (stationary) variable $\tilde{V}_t(\pi)$ whose components, $\tilde{v}_{t+i}(\pi)$, satisfy $|\tilde{v}_{t+i}(\pi)| \leq C_2 (|\Delta y_{t+i}| + \dots + |\Delta y_{t+i-r+1}|)$ ($C_2 < \infty, i = 1, \dots, s$). Thus, for $i \in \{1, \dots, s\}$ it suffices to consider

$$\begin{aligned}&\sup_{\theta \in N_{T,c}} T^{-1} \sum_{t=r+1}^{T-s} |e_{xx,t}(\theta) - e_{xx,t}(\theta_0)| |\Delta y_{t+i}| \\ &\leq C \sup_{\theta \in N_{T,c}} T^{-1} \sum_{t=r+1}^{T-s} \left[\left(1 + |\sigma_0^{-1} \epsilon_t(\theta_0)|^{d_1}\right) |\sigma^{-1} \epsilon_t(\theta) - \sigma_0^{-1} \epsilon_t(\theta_0)| + |\sigma^{-1} \epsilon_t(\theta) - \sigma_0^{-1} \epsilon_t(\theta_0)|^{d_2} \right] |\Delta y_{t+i}| \\ &\leq C \sup_{\theta \in N_{T,c}} \sigma^{-2} T^{-1} \sum_{t=r+1}^{T-s} \left[\left(1 + |\sigma_0^{-1} \epsilon_t(\theta_0)|^{d_1}\right) \sigma^{-1} |\tilde{\epsilon}_t(\theta) - \epsilon_t(\theta_0)| + \sigma^{-d_2} |\tilde{\epsilon}_t(\theta) - \epsilon_t(\theta_0)|^{d_2} \right] |\Delta y_{t+i}| + o_p(1).\end{aligned}$$

Here the first inequality is based on (22) and the second one is obtained in the same way as its analog for $A_T^{(2,2)}$ in case $\kappa = \phi$. That the first term in the last expression is of order $o_p(1)$ can be seen in the same way as in the case of $A_T^{(2,2)}$ (see the end of case $\kappa = \phi$ and note that here we need the more stringent assumptions $E[|\epsilon_t|^{2+d_1}] < \infty$ and $E[|\epsilon_t|^{1+d_2}] < \infty$).

References

- Andrews, B., R.A. Davis, and F.J. Breidt (2006). Maximum likelihood estimation for all-pass time series models. *Journal of Multivariate Analysis* 97, 1638-1659.
- Basawa, I.V. and D.J. Scott (1983). *Asymptotic Optimal Inference for Nonergodic Models*. Springer-Verlag. New York.
- Breidt, J., R.A. Davis, K.S. Lii, and M. Rosenblatt (1991). Maximum likelihood estimation for noncausal autoregressive processes. *Journal of Multivariate Analysis* 36, 175-198.
- Breidt, J., R.A. Davis, and A.A. Trindade (2001). Least absolute deviation estimation for all-pass time series models. *The Annals of Statistics* 29, 919-946.
- Brockwell, P.J. and R.A. Davis (1987). *Time Series: Theory and Methods*. Springer-Verlag. New York.
- Davidson, J. (1994). *Stochastic Limit Theory*. Oxford University Press, Oxford.
- Fuller, W.A. (1976). *Introduction to Statistical Time Series*. Wiley. New York.
- Gouriéroux, C. and J.M. Zakoian (2013). Explosive bubble modelling by noncausal process. CREST Working Paper 2013-04.
- Hansen, B.E. (1992). Convergence to stochastic integrals for dependent heterogeneous processes. *Econometric Theory* 8, 489-500.
- Huang, J. and Y. Pawitan (2000). Quasi-likelihood estimation of noninvertible moving average processes. *Scandinavian Journal of Statistics* 27, 689-710.
- Jeganathan, P. (1995). Some aspects of asymptotic theory with applications to time series models. *Econometric Theory* 11, 818-887.
- Lanne, M. and P. Saikkonen (2011). Modeling expectations with noncausal autoregressions. *Journal of Time Series Econometrics* 3, Iss. 3, Article 2.
- Lanne, M., J. Luoto, and P. Saikkonen (2012a). Optimal forecasting of noncausal autoregressive time series. *International Journal of Forecasting* 28, 623-631.
- Lanne, M., A. Luoma, and J. Luoto (2012b). Bayesian model selection and forecasting in noncausal autoregressive models. *Journal of Applied Econometrics* 27, 812-830.
- Lanne, M., H. Nyberg, and E. Saarinen (2012c). Forecasting U.S. macroeconomic and financial time series with noncausal and causal AR models: a comparison. *Economics Bulletin* 32, 2849-2859.

- Lanne, M. and P. Saikkonen (2013). Noncausal vector autoregression. *Econometric Theory* 29, 447-481.
- Lii, K.-S. and M. Rosenblatt (1996). Maximum likelihood estimation for non-Gaussian non-minimum phase ARMA sequences. *Statistica Sinica* 6, 1-22.
- Lucas, A. (1995). Unit root tests based on M estimators. *Econometric Theory* 11, 331-346.
- Meitz, M. and P. Saikkonen (2013). Maximum likelihood estimation of a noninvertible ARMA model with autoregressive conditional heteroskedasticity. *Journal of Multivariate Analysis* 114, 227-255.
- Park, J.Y. and P.C.B. Phillips (1988). Statistical inference in regressions with integrated processes: part 1. *Econometric Theory* 4, 468-497.
- Phillips, P.C.B. (1987). Time series regression with a unit root. *Econometrica* 55, 277-301.
- Phillips, P.C.B. (1988). Weak convergence of sample covariance matrices to stochastic integrals via martingale approximations. *Econometric Theory* 4, 528-533.
- Phillips, P.C.B. and P. Perron (1988). Testing for a unit root in time series regression. *Biometrika* 75, 335-346.
- Rosenblatt, M. (2000). *Gaussian and Non-Gaussian Linear Time Series and Random Fields*. Springer-Verlag, New York.
- Saikkonen, P. (1993). Continuous weak convergence and stochastic equicontinuity results for integrated processes with an application to the estimation of a regression model. *Econometric Theory* 9, 155-188.
- Scott, D.J. (1973). Central limit theorems for martingales and for processes with stationary increments using a Skorokhod representation approach. *Advances in Applied Probability* 5, 119-137.
- Van Dijk, D., P.H. Franses., and A. Lucas (1999). Testing for smooth transition nonlinearity in the presence of outliers. *Journal of Business and Economic Statistics* 17, 217-235.
- Wu, R. and R.A. Davis (2010). Least absolute deviation estimation for general autoregressive moving average time-series models. *Journal of Time Series Analysis* 31, 98-112.

BANK OF FINLAND RESEARCH DISCUSSION PAPERS

ISSN 1456-6184, online

- 1/2013 Tuomas Takalo **Rationales and instruments for public innovation policies.** 2013. 29 p. ISBN 978-952-462-000-0, online.
- 2/2013 Tuomas Takalo – Tanja Tanayama – Otto Toivanen **Market failures and the additionality effects of public support to private R&D: Theory and empirical implications.** 2013. 40 p. ISBN 978-952-462-001-7, online.
- 3/2013 Martin Ellison – Andreas Tischbirek **Unconventional government debt purchases as a supplement to conventional monetary policy.** 2013. 26 p. ISBN 978-952-462-002-4, online.
- 4/2013 Fabio Verona – Manuel M. F. Martins – Inês Drumond **(Un)anticipated monetary policy in a DSGE model with a shadow banking system.** 2013. 40 p. ISBN 978-952-6699-04-2, online.
- 5/2013 Fabio Verona – Maik H. Wolters **Sticky information models in Dynare.** 2013. 17 p. ISBN 978-952-6699-08-0, online.
- 6/2013 Sami Oinonen – Maritta Paloviita – Lauri Vilmi **How have inflation dynamics changed over time? Evidence from the euro area and USA.** 2013. 36 p. ISBN 978-952-6699-09-7, online.
- 7/2013 Iftekhar Hasan – Matej Marinč **Should competition policy in banking be amended during crises? Lessons from the EU.** 2013. 35 p. ISBN 978-952-6699-10-3, online.
- 8/2013 Bill Francis – Iftekhar Hasan – Qiang Wu **The benefits of conservative accounting to shareholders: Evidence from the financial crisis.** 2013. 44 p. ISBN 978-952-6699-09-7, online.
- 9/2013 Juha Kilponen – Jouko Vilmunen – Oskari Vähämaa **Estimating intertemporal elasticity of substitution in a sticky price model.** 2013. 27 p. ISBN 978-952-6699-12-7, online.
- 10/2013 Owain ap Gwilym – Qingwei Wang – Iftekhar Hasan – Ru Xie **In search of concepts: The effects of speculative demand on returns and volume.** 2013. 35 p. ISBN 978-952-6699-13-4, online.
- 11/2013 Matti Viren **Sensitivity of fiscal-policy effects to policy coordination and business cycle conditions.** 2013. 21 p. ISBN 978-952-6699-20-2, online.
- 12/2013 Seppo Honkapohja **The euro crisis: a view from the North.** 2013. 28 p. ISBN 978-952-6699-23-3, online.
- 13/2013 Stergios Leventis – Iftekhar Hasan – Emmanouil Dedoulis **The cost of sin: The effect of social norms on audit pricing.** 2013. 58 p. ISBN 978-952-6699-27-1, online.
- 14/2013 Chrysovalantis Gaganis – Iftekhar Hasan – Fotios Pasiouras **Efficiency and stock returns: Evidence from the insurance industry.** 2013. 33 p. ISBN 978-952-6699-28-8, online.

- 15/2013 Chung-Hua Shen – Iftekhar Hasan – Chih-Yung Lin **The government's role in government-owned banks.** 2013. 51 p. ISBN 978-952-6699-29-5, online.
- 16/2013 Fabio Verona **Lumpy investment in sticky information general equilibrium.** 2013. 37 p. ISBN 978-952-6699-30-1, online.
- 17/2013 Jouko Vilmunen – Peter Palmroos **Closed form solution of correlation in doubly truncated or censored sample of bivariate log-normal distribution.** 2013. 9 p. ISBN 978-952-6699-31-8, online.
- 18/2013 Fabio Verona **Investment dynamics with information costs.** 2013. 38 p. ISBN 978-952-6699-34-9, online.
- 19/2013 Mervi Toivanen **Contagion in the interbank network: an epidemiological approach.** 2013. 41 p. ISBN 978-952-6699-35-6, online.
- 20/2013 Jagjit S. Chadha – Elisa Newby **Midas, transmuting all, into paper: The Bank of England and the Banque de France during the Revolutionary and Napoleonic Wars.** 2013. 54 p. ISBN 978-952-6699-38-7, online.
- 21/2013 Kimmo Ollikka – Janne Tukiainen **Central bank liquidity auction mechanism design and the interbank market.** 2013. 47 p. ISBN 978-952-6699-39-4, online.
- 22/2013 Daron Acemoglu – Ufuk Akcigit – Nicholas Bloom – William Kerr **Innovation, reallocation and growth.** 2013. 47 p. ISBN 978-952-6699-40-0, online.
- 23/2013 Iftekhar Hasan – Jonathan O'Brien – Pengfei Ye **What types of bondholders impede corporate innovative activities?** 2013. 51 p. ISBN 978-952-6699-42-4, online.
- 24/2013 Jussi Lintunen – Lauri Vilmi **On optimal emission control – Taxes, substitution and business cycles.** 2013. 30 p. ISBN 978-952-6699-43-1, online.
- 25/2013 Ilkka Kiema – Esa Jokivuolle **99.9% – really?** 2013. 42 p. ISBN 978-952-6699-44-8, online.
- 26/2013 Pentti Saikkonen – Rickard Sandberg **Testing for a unit root in noncausal autoregressive models.** 2013. 29 p. ISBN 978-952-6699-45-5, online.

