Sieve empirical likelihood for unit root tests*

Francesco Bravo†

University of York

First draft January 2005
This draft June 2005

Abstract

This paper develops a new test for a unit root in autoregressive models with serially correlated errors. The test is based on the empirical likelihood method and uses a sieve approximation to eliminate the bias in the asymptotic distribution of the test due to presence of serial correlation. The paper derives the asymptotic distributions of the sieve empirical likelihood ratio under the null hypothesis of a unit root and under a local-to-unity alternative hypothesis, and uses a small Monte Carlo study to assess the finite sample properties of the proposed test statistic. The results of the simulations seem to suggest that the empirical likelihood ratio has in general similar size and in most cases better power properties than those of standard Augmented Dickey-Fuller tests of a unit root.

Keywords and Phrases. Autoregressive approximation, Empirical likelihood, Linear process, Martingale, Unit root test.

*Financial support from ESRC under research grant no. RES-000-22-0980 is gratefully acknowledged.

†I would like to thank partecipants of the C*e*MMAP workshop on empirical likelihood methods for some useful comments. All remaining errors are my own responsibility.

Address correspondence to: Francesco Bravo, Department of Economics and Related Studies, University of York, York, YO10 5DD, United Kingdom. email fb6@york.ac.uk

1 Introduction

The problem of testing for a unit root in autoregressive models -the so-called unit root hypothesis- has drawn considerable interest in economics and finance in the past two decades. The hypothesis is of great importance in both fields because of its practical implications. For example in economics it implies that shocks have permanent effects on the future path of the economy, whereas in finance it implies that (log)prices of assets follow a random walk and thus are not predictable. It is therefore not surprising that a large number of unit root tests have been developed -see Phillips and Xiao (1998) for an updated survey of unit root testing approaches- and applied to real data. The general conclusion seems to be that most macroeconomic and many financial time series are consistent with the unit root hypothesis. However there also exists also a large body of simulation evidence which shows that the finite sample properties of commonly used unit root tests are often very different from the asymptotic properties at sample sizes common in econometric practice. This is particularly evident in the case of autoregressive models with weakly dependent errors.

In situations where asymptotic approximations to the distribution of test statistics are not reliable, one option is to approximate their finite sample distribution using the bootstrap. The validity of bootstrap methods for unit root models was originally investigated by Basawa, Mallikand, McCormick, Reeves and Taylor (1991) and more recently by Psaradakis (2001), Romano and Wolf (2001), Park (2002), and Chang and Park (2003) among others. Another approach is to consider alternative tests for a unit root. In this paper we follow the latter and propose a new test for a unit root based on Owen's (1988) empirical likelihood.

Empirical likelihood is a nonparametric method of statistical inference that allows the researcher to use likelihood methods without having to assume that the observations available come from a parametric family of distributions. It is very general and applications can be found in a number of different areas. Empirical likelihood enjoys a number of desirable statistical properties: it is range preserving, transformation respecting, and does not require estimation of scale or skewness. Moreover it is Bartlett correctable (DiCiccio, Hall and Romano, 1991), optimal under a generalised Neyman-Pearson criterion (Kitamura, 2001), and second-order maximinity

(Bravo, 2003). See the recent monograph of Owen (2001) for an account of some of these properties, recent developments and applications of empirical likelihood.

In this paper we develop an empirical likelihood test for a unit root in an autoregressive model with errors parameterised as a general linear process. The results of the paper are based on the same autoregressive approximation used by Chang and Park (2002) in the context of Augmented Dickey-Fuller unit root tests, and by Psaradakis (2001) and Chang and Park (2003) in the context of bootstrap unit root tests. This approximation captures the dependent structure of the errors by fitting autoregressive models with order increasing with the sample size at an appropriate rate. Since the approximation is sometimes known in time series literature as sieve approximation (see e.g. Bühlmann (1997)), we shall call the resulting empirical likelihood sieve empirical likelihood.

In this paper we establish the asymptotic distributions of the sieve empirical likelihood ratio under the null hypothesis of a unit root and under a sequence of local-to-unity alternatives. The resulting distributions correspond to the square of well-known functionals of a Brownian motion and Ornstein-Uhlenbeck process, respectively. These results extend some of the results obtained by Chuang and Chan (2002) in the case of empirical likelihood for unit root autoregressive model with errors forming a martingale difference sequence. In the paper we also use simulations to assess the finite sample properties of the sieve empirical likelihood ratio. The results of the simulations seem to suggest that although empirical likelihood does not solve the size distortion problem, it does produce a test statistic that can have in most cases better power properties than those of a (squared) Augmented Dickey-Fuller t-statistic.

The rest of the paper is structured as follows: next section first reviews the concept of sieve approximation and shows how it can be used in the context of empirical likelihood inference for a unit root. Then it presents the asymptotic results for the sieve empirical likelihood ratio statistic. Section 3 contains the results of a Monte Carlo study, while Section 4 contains some concluding remarks and indications for future research. An Appendix contains all the proofs.

2 Main results

Consider the following $AR(1)$ process

$$\begin{aligned} y_t &= \beta y_{t-1} + u_t, \quad t = 1, 2, ..., n \\ u_t &= \Psi(L)\varepsilon_t = \sum_{j=1}^{\infty} \psi_j \varepsilon_{t-j} \end{aligned} \tag{1}$$

where, for simplicity, $y_0 = 0$[1], and the innovations ε_t form a martingale difference sequence satisfying the following assumptions:

M (I) $E\left(\varepsilon_t^2 | \mathcal{F}_{t-1}\right) = \sigma^2$ $a.s.$, (II) $E\left(|\varepsilon_t|^\gamma | \mathcal{F}_{t-1}\right) < \infty$ $a.s.$ for some $\gamma \geq 4$,

and $\Psi(L)$ satisfies

LP $\Psi(z) \neq 0$ for all $|z| \leq 1$ and $\sum_{j=1}^{\infty} j^k \left|\psi_j\right| < \infty$ for some $k \geq 1$.

Assumptions **M** and **LP** imply that u_t is a strictly stationary linear process driven by martingale difference innovations. Note also that **LP** includes models with polynomial decay of the coefficients $\left\{\psi_j\right\}_{j=0}^{\infty}$. General $ARMA(p, q)$ processes satisfy **LP** with an exponential decay of $\left\{\psi_j\right\}_{j=0}^{\infty}$.

It is well known (see, for example, Phillips (1987a)) that the dependent structure of the error process u_t introduces a bias term that affects the asymptotic distributions of the t- statistic or coefficient test for the unit root hypothesis $H_0 : \beta = 1$ in (1). One way to solve this problem is to correct the test statistics with nonparametric estimates of the bias as originally proposed by Phillips (1987a) and Phillips and Pierron (1988)[2]. Alternatively one can approximate u_t by a finite autoregressive process of order increasing with the sample size as recently suggested by Chang and Park (2002). As mentioned in the Introduction, this autoregressive approximation is sometimes known in time series literature as sieve approximation (Bühlmann, 1997),

[1] The initial value of y_t does not affect the asymptotic results obtained below as long as $y_t = O_p(1)$.

[2] Note however that the corrections proposed by Phillips (1987a) and Phillips and Pierron (1988) are also valid for other weakly dependent structures of the errors. In particular they are valid if u_t is assumed to be a strong-mixing process with mixing coefficients α_m satisfying $\sum_{=1}^{\infty} \alpha_j^{1-2/\alpha} < \infty$ for some $\alpha > 2$.

since it is based on a sequence of autoregressive processes (i.e. a sequence of finite dimensional parametric models) approximating a linear process which can be thought of as an infinite dimensional nonparametric model. It should be noted that in the context of tests for a unit root, a similar idea was originally proposed by Said and Dickey (1984) and more recently by Xiao and Phillps (1998) as an extension to the traditional augmented Dickey-Fuller (ADF) test for a unit root with weakly dependent errors. Both Said and Dickey (1984) and Xiao and Phillps (1998) however considered only linear processes with geometrically decaying coefficients and i.i.d. innovations, as opposed to the general linear processes considered in Chang and Park (2002) and in this paper.

2.1 Sieve approximation

Let

$$\Phi(L)\, u_t = \varepsilon_t$$

denote the $AR(\infty)$ representation of the linear process u_t, and notice that **LP** implies that $\Phi(z)$ is bounded away from 0 for $|z| \leq 1$ and that $\sum_{j=1}^{\infty} j^k \left|\phi_j\right| < \infty$. Consider the following $AR(p)$ approximation to u_t

$$u_t = \sum_{j=1}^{p} \phi_j u_{t-j} + \varepsilon_{p,t} \tag{2}$$

where

$$\varepsilon_{p,t} = \varepsilon_t + \sum_{j=p+1}^{\infty} \phi_j u_{t-j},$$

and note that by iterated expectations, $\mathbf{M}$(II) and **LP** $E\left|\varepsilon_{p,t} - \varepsilon_t\right|^{\gamma} = o\left(p^{-\gamma k}\right)$(Chang and Park, 2002), implying that the larger p becomes, the smaller is the error in the autoregressive approximation (2). The basic idea in the sieve approximation for u_t is to allow the order p of the autoregressive approximation to increase at an appropriate rate with the sample size, that is $p = p(n) \to \infty$ as $n \to \infty$. Using (2), we can rewrite (1) as

$$y_t = \beta y_{t-1} + \sum_{j=1}^{p(n)} \phi_j \Delta y_{t-j} + \varepsilon_{p(n),t}, \tag{3}$$

where $\Delta y_t = u_t$. Define the sieve (or augmented) "score" function

$$m_{p(n),t}(\beta,\phi) = \begin{bmatrix} y_{t-1} & \Delta y_{t-1} & \dots & \Delta y_{t-p(n)} \end{bmatrix}' \left(y_t - \beta y_{t-1} - \sum_{j=1}^{p(n)} \phi_j \Delta y_{t-j} \right), \quad (4)$$

and let $\begin{bmatrix} \widehat{\beta} & \widehat{\phi}' \end{bmatrix}'$ denote the least square estimator that solves $\sum_{t=1}^{n} m_{p(n),t}\left(\widehat{\beta},\widehat{\phi}\right) = 0$. Then using (3) the unit root hypothesis $H_0 : \beta = 1$ can be tested using the ADF t-statistic

$$ADF_t = \left(\widehat{\beta} - 1\right) / \widehat{\sigma}_{\widehat{\beta}} \quad (5)$$

where $\widehat{\sigma}_{\widehat{\beta}}$ is the standard error for the estimated coefficient $\widehat{\beta}$.

Let $B(r)$ denote a standard Brownian motion on $C[0,1]$, the space of continuous functions on the interval $[0,1]$, and $\xrightarrow{w}$ denote weak convergence in distribution. Chang and Park (2002) show that under the minimal rate condition[3] $p(n) = o\left(n^{1/2}\right)$ the asymptotic distribution of (5) is

$$ADF_t \xrightarrow{w} \left(\int_0^1 B(r)\, dB(r)\right) / \left(\int_0^1 B^2(r)\, dr\right)^{1/2},$$

that is the distribution of ADF_t coincides with that of the t-statistic for a unit root obtained by Dickey and Fuller (1979) in the absence of serially correlated errors.

2.2 Sieve empirical likelihood

The basic idea behind empirical likelihood is that without restrictions on the joint distribution of the observations the empirical distribution function is an optimal estimator (i.e. it is the nonparametric maximum likelihood estimator) for the unknown distribution function. In the presence of a set of restrictions however this is typically not true. Empirical likelihood exploits this fact and estimates among all the distributions supported on the sample the one consistent with the same set of restrictions. Then, in analogy with parametric likelihood methods, one way to assess whether the set of restrictions is compatible with the observations available is to use an empirical likelihood ratio statistic based on the log-difference between the unconstrained and constrained distribution of the observations.

[3] Chang and Park (2002) note that this rate is not sufficient for consistency of the estimates of the ϕ_js.

If the error process u_t in (1) was a martingale difference sequence, the martingale property of the score function $m_t(\beta) = y_{t-1}(y_t - \beta y_{t-1})$ under the true value of β, say β_0, implicitly imposes the moment restriction $E[m_t(\beta_0)] = 0$ that, as shown by Chuang and Chan (2002), can be used to build an empirical likelihood ratio test for the hypothesis $H_0 : \beta = \beta_0$. In the case of weakly dependent errors the same restriction typically does not hold and thus cannot be used by empirical likelihood. On the other hand using iterated expectations it is easy to see that under $H_0 : \left[\begin{array}{cc} \beta & \phi' \end{array}\right]' = \left[\begin{array}{cc} \beta_0 & \phi_0' \end{array}\right]'$ the sieve score $m_{p(n),t}(\beta_0, \phi)$ defined in (4) give rise to a set of $p(n)$ moment restrictions that can be tested using empirical likelihood. To be specific let $\{w_t\}_{t=1}^n$ denote a set of weights such that $w_t > 0$ and $\sum_{t=1}^n w_t = 1$; then the sieve empirical likelihood finds the w_t's consistent with $E\left[m_{p(n),t}(\beta_0, \phi_0)\right] = 0$ by solving

$$\max_{w_t} \left\{ \prod_{t=1}^n w_t \,|\, \sum_{t=1}^n w_t m_{p(n),t}(\beta_0, \phi_0) = 0, \sum_{t=1}^n w_t = 1 \right\}. \tag{6}$$

As discussed by Owen (1990) a unique solution to (6) exists provided that the following condition holds[4]

$$\Pr\left(0 \in ch\left\{ \begin{array}{ccc} m_{p(n),1}(\beta_0, \phi_0) & \cdots & m_{p(n),t}(\beta_0, \phi_0)\, m_n(1) \end{array} \right\}\right) \to 1 \text{ as } n \to \infty.$$

This convex hull (ch) condition implies that asymptotically the mean of the distribution lies within the convex hull $ch\{\cdot\}$ of the $m_{p(n),t}(\beta_0, \phi_0)$'s. Otherwise it would be impossible to reweight the $m_{p(n),t}(\beta_0, \phi_0)$'s so that the weighted mean is 0. A Lagrange multiplier argument shows that the unique solution to (6) is

$$w_t^{-1} = n\left(1 + \hat{\lambda}' m_{p(n),t}(\beta_0, \phi_0)\right)$$

where $\hat{\lambda} := \hat{\lambda}(\beta, \phi)$ is a Lagrange multiplier that solves

$$0 = \sum_{t-1}^n m_{p(n),t}(\beta_0, \phi_0) / \left(1 + \lambda' m_{p(n),t}(\beta_0, \phi_0)\right).$$

Since the empirical distribution function puts equal probability mass $1/n$ on each of the $m_t(\beta, \phi)$'s it follows that an empirical likelihood ratio statistic for the null

[4] In the Appendix we show that in an appropriate neighbourhood of zero the empirical likelihood estimator exists $a.s.$ implying that the convex hull condition is automatically satisfied. I am indebited to Whitney Newey for pointing this out to me.

hypothesis $H_0 : \left[\begin{array}{cc} \beta & \phi' \end{array} \right]' = \left[\begin{array}{cc} \beta_0 & \phi_0' \end{array} \right]'$ is

$$W(\beta_0, \phi_0) = 2\sum_{t=1}^{n} (\log(1/n) - \log(w_t)) = 2\sum_{t=1}^{n} \log\left(1 + \hat{\lambda}' m_{p(n),t}(\beta_0, \phi_0)\right).$$

The unit root hypothesis $H_0 : \beta = 1$ is a composite hypothesis with β the parameter of interest and ϕ a vector of nuisance parameters. Empirical likelihood deals with nuisance parameters by profiling (i.e. concentrating them out). Let

$$W\left(1, \widehat{\phi}\right) := 2\min_{\phi} \sum_{t=1}^{n} \log\left(1 + \hat{\lambda}' m_{p(n),t}(1, \phi)\right)$$

denote the profile sieve empirical likelihood ratio for $H_0 : \beta = 1$. In addition to **M**(I),(II) and **LP** we assume that

MU $\sup_t E\left(|\varepsilon_t|^{\alpha} |\mathcal{F}_{t-1}\right) < \infty$ *a.s.* for some $\alpha > 2$,

and that as $p(n) \to \infty$

P (I) $p(n) = o\left(n^{1/2-\beta}\right)$ for some $\beta > 1/\alpha$, (II) $p(n)$ satisfies $\lim_{n\to\infty} n^{1/2} \sum_{j=p(n)+1}^{\infty} \left|\phi_j\right| = 0$.

Assumption **MU** is a uniform integrability condition, which is in general stronger than the one used by Chuang and Chan (2002) in the case of unstable autoregressive processes with martingale difference innovations. In this paper it is used to impose a restriction, as specified in assumption **P**(I), on the growth rate of $p(n)$ which is the weaker the stronger **MU** is. Assumption **P**(II) is as in Berk (1974) and Said and Dickey (1984) and ensures the $n^{1/2}$ consistency of the estimates of the ϕ_j's. At the same time **P**(II) rules out the logarithmic rate for $p(n)$, which implies that information-based selection rules for $p(n)$, such as the Akaike information criterion (AIC) or the Bayesian (Schwartz) criterion (BIC) are not allowed. It should be noted that **P**(II) is stronger than what is needed for the validity of this paper's results. Indeed Theorems 1 and 2 below are still valid if **P**(II) is replaced by the weaker assumption

P (II)' $p(n)$ satisfies $\lim_{n\to\infty} p(n)^{1/2} \sum_{j=p(n)+1}^{\infty} \left|\phi_j\right| = 0$.

On the other hand $\mathbf{P}$(II) allows using the sequential tests method of Ng and Perron (1995, p. 270) for the choice of $p(n)$. This method seems to work well in practice -see next section for more details- and provides consistent estimates (the least squares estimates) that can be used as initial values for the ϕ_j's in the numerical algorithm used to compute $W\left(1, \widehat{\phi}\right)$.

Theorem 1 *Assume that* ***M****(I),(II)* ***MU****,* ***LP*** *and* ***P****(I),(II) hold. Then for* $p(n) = o\left(n^{1/4}\right)$

$$W\left(1, \widehat{\phi}\right) \xrightarrow{w} \left(\int_0^1 B(r)\, dB(r)\right)^2 / \int_0^1 B^2(r)\, dr. \tag{7}$$

Theorem 1 shows that $W\left(1, \widehat{\phi}\right)$ converges to the square of the ADF t-statistic for a unit root defined in (5). Thus Theorem 1 provides an asymptotic justification for the test that rejects the unit root hypothesis at α level, when $W\left(1, \widehat{\phi}\right) > u_{1-\alpha}$ where $u_{1-\alpha}$ is the $1-\alpha$ quantile of (7) which can be found by simulation. Note that the rate $o\left(n^{1/4}\right)$ is the same as that of Donald, Imbens and Newey (2003) in the context of consistent empirical likelihood tests for conditional moment restrictions.

To assess the power properties of $W\left(1, \widehat{\phi}\right)$ we consider the same type of (local) alternative hypotheses considered by Chan and Wei (1987) and Phillips (1987*b*). It is important to note that in the context of empirical likelihood inference the distribution functions under the null and the local alternative hypotheses are related in the sense that they are both assumed to belong to the same class of multinomial distributions supported on the sample indexed by the parameter β. Bearing this in mind, the sequence of alternatives we consider is $H_n : \beta_n = 1 - \gamma/n$ for some $-\infty < \gamma < \infty$. Let $J_\gamma(r) = \int_0^r \exp\left[-(r-s)\gamma\right] dB(s)$ denote an Ornstein-Uhlenbeck process.

Theorem 2 *Under the same assumptions of Theorem 1,*

$$W\left(\beta_n, \widehat{\phi}\right) \xrightarrow{w} \left(\int_0^1 J_\gamma(r)\, dB(r)\right)^2 / \int_0^1 J_\gamma{}^2(r)\, dr. \tag{8}$$

Theorem 2 shows that $W\left(\beta_n, \widehat{\phi}\right)$ has the same local asymptotic power of ADF_t^2. Since the rate of divergence under the alternative of the latter is $O_p(n)$ we may expect $W\left(\beta_n, \widehat{\phi}\right)$ to have good power properties and in particular greater power than standard t-ratio test (5). On the other hand the profile sieve empirical likelihood ratio is a nondirectional test whereas the standard ADF t-ratio is directional against stationary or explosive alternatives. To obtain a directional profile sieve empirical likelihood ratio test one can consider its squared root, that is

$$R\left(1,\widehat{\phi}\right) := sign\left(\widehat{\beta}-1\right)\left[W\left(1,\widehat{\phi}\right)\right]^{1/2},$$

where $sign(\cdot)$ is the sign function. A straightforward application of the continuos mapping theorem shows that $R\left(1,\widehat{\phi}\right)$ converges weakly to the same distribution as that of ADF_t.

Remark. Models with deterministic trends x_t can be analysed similarly. Suppose that z_t is generated by

$$z_t = \gamma' x_t + y_t \tag{9}$$

and y_t is given by (1). Then the unit root hypothesis can still be tested as in (3) using the residuals $\widehat{y}_t$ of a preliminary regression (9). It can be shown that the distribution of the resulting profile sieve empirical likelihood ratio is

$$W\left(1,\widehat{\phi}\right) \xrightarrow{w} \left(\int_0^1 B_X(r)\, dB(r)\right)^2 / \int_0^1 B_X^2(r)\, dr$$

where $B_X(r) = B(r) - \left[\int_0^1 B(r) X(r)'\right]\left[\int_0^1 X(r) X(r)'\right]^{-1} X(r)$ is a detrended Brownian motion and depends on the limit trend function $X(r)$.

3 Monte Carlo evidence

In this section we use simulations to investigate the performance of the profile sieve empirical likelihood ratio test for a unbit root in finite samples. We consider the model

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

with $\beta = \{1, 0.95, 0.90\}$, and u_t is either $u_t = \theta u_{t-1} + \varepsilon_t$ or $u_t = \varepsilon_t + \theta\varepsilon_{t-1}$. In the simulations we use $\theta = \{-.8, -.5, -.2, 0, .2, .5, .8\}$ and three different specifications for the distribution of the error ε_t, namely standard normal, t_5 (t-distribution with five degrees of freedom), and $\chi_4^2 - 4$ (centred chi-squared distribution with four degrees of freedom). The sample size n is set to 200.

A practical issue that arises in calculating both $W\left(1,\widehat{\phi}\right)$ and ADF_t^2 (or their square roots) is the choice of the lag $p(n)$ in the sieve approximation. There is a large body of simulation evidence showing how this choice has important implications for the finite sample properties of standard ADF tests for a unit root. We investigated this issue in a preliminary Monte Carlo study using different specifications of $p(n)$. The simulations[5] clearly indicated that the choice of $p(n)$ has bearing on the finite sample properties of both $W\left(\beta,\widehat{\phi}\right)$ and ADF_t^2, and that among various data dependent criteria for choosing $p(n)$ the one based on sequential testing suggested by Ng and Perron (1995) produced tests with better finite sample properties. Accordingly the Monte Carlo evidence reported in Tables 1-3 below is based on the lag $p(n)$ chosen by means of a sequential 0.10 level two-sided t- test for the significance of the coefficient of the longest lag.

Tables 1-3 report the finite sample size and power of $W\left(\beta,\widehat{\phi}\right)$ and ADF_t^2 for the unit root hypothesis $H_0 : \beta = 1$ at the $\{0.10, 0.05\}$ nominal level. For $\beta = 1$ the empirical size is obtained form 5000 replications using simulated asymptotic critical values[6]. For $\beta < 1$ the empirical power is calculated from 1000 replications using empirical critical values obtained from the simulations under the null hypothesis, and thus represents size-corrected power.

Tables 1-3 here

The results of Tables 1-3 can be summarised as follows: First the empirical likelihood ratio is characterised by similar size properties to those of the square of the ADF t- statistic. Second the empirical likelihood ratio seem to have in most cases better power than ADF_t^2, particularly when the alternative hypothesis is closer to the null.

[5] Available upon request.

[6] The critical values were obtained by approximating $B(r)$ with partial sums of $N(0,1)$ random variables with 5,000 steps and 99,999 replications.

4 Conclusion

In this paper we have proposed an empirical likelihood ratio test for a unit root in an autoregressive model with serially correlated errors. Our approach is based on a sieve approximation that uses autoregressive models of appropriate order to capture the dependent structure of the errors. We have derived the asymptotic distribution of the resulting sieve empirical likelihood ratio and assessed its finite sample properties using a small Monte Carlo study. The results of the Monte Carlo study indicate that empirical likelihood ratio has, in general, similar size and in most cases better power properties than those of a squared Augmented Dickey-Fuller t- statistic. These results suggest some directions for future research. First the scale of the Monte Carlo study of the paper is small. It would useful to extend the simulations by considering different sample sizes and other values of the alternative hypothesis. It would also be useful to analyse the finite sample properties of the signed squared root of the sieve empirical likelihood ratio and compare its properties with the standard ADF t statistic.

Second the issue of selecting $p(n)$ requires further investigation. In the paper we have considered the method suggested by Ng and Perron (1995). This method seems to work well in practice, however it needs not to be optimal.

Third the empirical likelihood ratio is one of the possible test statistics that can be used. There are in fact other asymptotically equivalent statistics that can be used, like, for example, the Cressie-Read power-divergence statistic (Read and Cressie, 1988), or the generalised empirical likelihood statistic of Newey and Smith (2004). This possibility is certainly of interest and is left for future research.

References

Basawa, I., Mallikand, A., McCormick, W., Reeves, J. and Taylor, L. (1991), 'Bootstrapping unstable first-order autoregressive processes', *Annals of Statistics* **19**, 1098–1101.

Berk, N. (1974), 'Consistent autoregressive spectral estimates', *Annals of Statistics* **2**, 489–502.

Beveridge, S. and Nelson, C. (1981), 'A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the "business cycle"', *Journal of Monetary Economics* **7**, 151–174.

Bravo, F. (2003), 'Second-order power comparisons for a class of nonparametric likelihood-based tests', *Biometrika* **90**, 881–890.

Bühlmann, P. (1997), 'Sieve bootstrap for time series', *Bernoulli* **3**, 123–148.

Chan, N. and Wei, C. (1987), 'Asymptotic inference of nearly nonstationary AR(1) process', *Annals of Statistics* **15**, 1050–1063.

Chan, N. and Wei, C. (1988), 'Limiting distributions of least squares estimates of unstable autoregressive processes', *Annals of Statistics* **16**, 367–401.

Chang, Y. and Park, J. (2002), 'On the asymptotics of ADF tests for unit roots', *Econometric Reviews* **21**, 449–476.

Chang, Y. and Park, J. (2003), 'A sieve bootstrap for the test of a unit root', *Journal of Time Series Analysis* **24**, 379–400.

Chuang, C. and Chan, N. (2002), 'Empirical likelihood for autoregressive models, with applications to unstable time series', *Statistica Sinica* **12**, 387–408.

DiCiccio, T., Hall, P. and Romano, J. (1991), 'Empirical likelihood is Bartlett-correctable', *Annals of Statistics,* **19**, 1053–1061.

Dickey, D. and Fuller, W. (1979), 'Distribution of the estimators for autoregressive time series with a unit root', *Journal of the American Statistical Association* **74**, 427–431.

Donald, S., Imbens, G. and Newey, W. (2003), 'Empirical likelihood estimation and consistent tests with conditional moment restrictions', *Journal of Econometrics* **117**, 55–93.

Kitamura, Y. (2001), 'Asymptotic optimality of empirical likelihood for testing moment restrictions', *Econometrica* **69**, 1661–1672.

Lai, T. and Wei, C. (1982), 'Least squares estimates in stochastic regression models with applications to identification and control of dynamic system', *Annals of Statistics* **10**, 154–166.

Newey, W. and Smith, R. (2004), 'Higher order properties of GMM and generalized empirical likelihood estimators', *Econometrica* **72**, 219–256.

Ng, S. and Perron, P. (1995), 'Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag', *Journal of the American Statistical Association* **90**, 268–281.

Owen, A. (1988), 'Empirical likelihood ratio confidence intervals for a single functional', *Biometrika* **36**, 237–249.

Owen, A. (1990), 'Empirical likelihood ratio confidence regions', *Annals of Statistics* **18**, 90–120.

Owen, A. (2001), *Empirical Likelihood*, Chapman and Hall.

Park, J. (2002), 'An invariance principle for sieve bootstrap in time series', *Econometric Theory* **18**, 469–490.

Phillips, P. (1987*a*), 'Time series regression with a unit root', *Econometrica* **55**, 277–301.

Phillips, P. (1987*b*), 'Towards a unified asymptotic theory for autoregression', *Biometrika* **74**, 535–547.

Phillips, P. and Pierron, P. (1988), 'Testing for a unit root in time series regressions', *Biometrika* **75**, 335–346.

Phillips, P. and Xiao, Z. (1998), 'A primer on unit root testing', *Journal of Economics Surveys* **12**, 423–470.

Psaradakis, Z. (2001), 'Bootstrap tests for an autoregressive unit root in the presence of weakly dependent errors', *Journal of Time Series Analysis* **22**, 577–594.

Read, T. R. C. and Cressie, N. (1988), *Goodness-of-Fit Statistics for Discrete Multivariate Data*, New York: Springer-Verlag.

Romano, J. and Wolf, M. (2001), 'Subsampling intervals in autoregressive models with linear time trend', *Econometrica* **69**, 1283–1314.

Said, S. and Dickey, D. (1984), 'Testing for unit roots in autoregressive-moving average models of unknown order', *Biometrika* **71**, 599–608.

Xiao, Z. and Phillps, P. (1998), 'An ADF coefficient test for a unit root in ARMA models with unknown order with empirical applications to the US economy', *Econometrics Journal* **1**, 27–43.

Appendix

In the Appendix C denotes a generic positive constant, $\sum = \sum_{t=1}^{n}$, unless otherwise stated, and "CMT" stands for continuous mapping theorem. Also if M is a symmetric matrix and v is a vector the matrix norm $\|M\|$ is defined as $\sup_{(v'v)^{1/2}\leq 1} \|Mv\|$, whereas $\varsigma_{\max}(M)$ and $\varsigma_{\min}(M)$ denote the largest and smallest eigenvalue of M. For notational convenience, let $p(n) := p$ and let

$$\begin{aligned} \Delta \mathbf{y}_{p,t} &= \left[\begin{array}{ccc} \Delta y_t & \ldots & \Delta y_{t-p+1} \end{array} \right]', \quad S_n = diag\left[\begin{array}{cc} n^{-1} & n^{-1/2} I_p \end{array} \right], \\ Y_{t-1,n} &= S_n \left[\begin{array}{cc} y_{t-1} & \Delta \mathbf{y}'_{p,t-1} \end{array} \right]', \quad m_{tn} = Y_{t-1,n}\varepsilon_{p,t}, \\ M_{nn} &= E\left(\sum m_{tn} m'_{tn} | \mathcal{F}_{t-1} \right), \quad \Sigma_{\omega p} = diag\left(\sigma^2 \left[\begin{array}{cc} \omega & \Sigma_p \end{array} \right] \right) \end{aligned}$$

where $\varepsilon_{p,t} = \varepsilon_t + \sum_{j=p+1}^{\infty} \phi_j u_{t-j}$, ω is as random variable such that $\Pr(\omega > 0) \to 1$, and the $\Re^{p\times p}$ valued matrix $\Sigma_p = O(1)$ has typical (i,j) element $E(\Delta y_{t-i} \Delta y_{t-j})$.

Lemma 3 *Let* $y_t = y_{t-1} + u_t$ *where* $u_t = \Psi(L)\varepsilon_t = \sum_{j=1}^{\infty} \psi_j \varepsilon_{t-j}$, ε_t *is a martingale difference sequence satisfying* ***M(I)*** *and* ***MU,*** $\Psi(L)$ *satisfies* ***LP*** *and* $y_0 = 0$. *Then* $\max_t |y_t/n| = O_{a.s.}\left(n^{-1/2} (\log\log n)^{1/2} \right)$ *and* $\max_t |u_t| = o_{a.s.}(n^{\beta})$ *for any* $\beta > 1/\alpha$.

Proof. By the Beveridge-Nelson decomposition (Beveridge and Nelson, 1981) $y_t = \Psi(1) S_t + \eta_t$ where $S_t = \sum_{j=1}^{t} \varepsilon_j$, $\eta_t = \sum_{j=0}^{\infty} \alpha_j \varepsilon_{t-j}$ and $\alpha_j = -\sum_{k=1}^{\infty} \psi_{j+k}$. By Lai and Wei (1982) $\max_t |S_t| = O_{a.s.}\left(n^{1/2} (\log\log n)^{1/2} \right)$ and $\max_t |\varepsilon_t| = o_{a.s.}(n^{\beta})$, so that the conclusion follows by CMT noting that $\max_t |u_t| \leq \sum_{j=}^{\infty} |\psi_j| \max_t |\varepsilon_t| = o_{a.s.}(n^{\beta})$. ■

Lemma 4 *Assume that* ***M(I),(II)*** *and* ***LP*** *hold. Then*

$$\left\| \sum m_{tn} m'_{tn} - M_{nn} \right\| = O_p\left(p/n^{1/2} \right). \tag{A1}$$

Proof. Note that $E\left(\|\sum m_{tn} m'_{tn} - M_{nn}\|^2 |\mathcal{F}_{t-1} \right) \leq \sum \|Y_{t-1,n}\|^4 E\left(|\varepsilon_{p,t}|^4 |\mathcal{F}_{t-1} \right) \leq Cp^2/n$ since $y_{t-1}/n^{1/2} = O_p(1)$ (Chan and Wei, 1988), which implies $(A1)$ by the conditional Chebychev inequality. ■

Lemma 5 *Assume that* ***M*** *and* ***LP*** *hold. Then*

$$\|M_{nn} - \Sigma_{\omega p}\| = O_p\left(p/n^{1/2} \right). \tag{A2}$$

Proof. Note that $E\left[(\varepsilon_{p,t}-\varepsilon_t)^2|\mathcal{F}_{t-1}\right]\leq E\left(u_t^2|\mathcal{F}_{t-1}\right)\left(\sum_{j=p+1}^{\infty}|\phi|\right)^2=o_{a.s.}(1)$, so by triangle inequality $E\left(\varepsilon_{p,t}^2|\mathcal{F}_{t-1}\right)=\sigma^2+o_{a.s.}(1)$, which implies that $E\left(\sum m_{tn}m_{tn}'|\mathcal{F}_{t-1}\right)=\sigma^2\sum Y_{t-1,n}Y_{t-1,n}'+o_{a.s.}(1)$. By Berk (1974) $\sum\Delta\mathbf{y}_{p,t-1}\Delta\mathbf{y}_{p,t-1}'/n\xrightarrow{p}\Sigma_p$ and by Phillips (1987*b*) $\sum y_{t-1}^2/n^2\xrightarrow{w}\sigma^2\Psi(1)\int B^2(r)\,dr:=\sigma^2\omega$. Furthermore, since $\left\|\sum y_{t-1}\Delta\mathbf{y}_{p,t-1}'/n\right\|=O_p\left(p^{1/2}\right)$ (Chang and Park, 2002, Lemma 3.2), $\sum y_{t-1}\Delta\mathbf{y}_{p,t-1}'/n^{3/2}\xrightarrow{p}0$. Thus by iterated expectations it can be shown that $E\left\|M_{nn}-\Sigma_{\omega p}\right\|^2=Cp^2/n$ which implies $(A2)$. ■

Lemma 6 *Assume that* ***M****(I),(II) and* ***LP*** *hold. Then for* $p=o\left(n^{1/2}\right)$

$$\left\|\sum m_{tn}m_{tn}'-\Sigma_{\omega p}\right\|=o_p(1)\tag{A3}$$

and

$$\left|\varsigma\left(\sum m_{tn}m_{tn}'\right)-\varsigma\left(\Sigma_{\omega p}\right)\right|=o_p(1)\tag{A4}$$

where $\varsigma(\cdot)$ *is either* $\varsigma_{\min}(\cdot)$ *or* $\varsigma_{\max}(\cdot)$.

Proof. $(A1)$, $(A2)$ and the triangle inequality imply $(A3)$. Since for any symmetric matrices M_1,M_2 $|\varsigma(M_1)-\varsigma(M_2)|\leq\|M_1-M_2\|$, $(A4)$ follows by $(A3)$. ■

Lemma 7 *Assume that* ***M****(I),(II),* ***MU****,* ***LP*** *hold. Let* $R_n=o\left(diag\begin{pmatrix}n^{-1} & n^{-\beta}I_p\end{pmatrix}\right)$ *for* $\beta>1/\alpha$, $\Lambda_{pn}=\left\{\lambda:p^{1/2}\left\|R_n^{-1}\lambda\right\|\leq\epsilon\right\}$ *where* $\epsilon\leq 1$. *Then*

$$\max_t\sup_{\lambda\in\Lambda_{pn}}|\lambda'm_t|=o_{a.s.}(1),\tag{A5}$$

and

$$\Lambda_{pn}\subset\Lambda_0\quad a.s.\tag{A6}$$

where Λ_0 *is an open interval containing 0.*

Proof. By Lemma 3 $\max_t|y_{t-1}/n|=o_{a.s.}(1)$ and $\max_{j,t}|\Delta y_{t-j}|=o_{a.s}\left(n^{\beta}\right)$ for $j=1,...,p$ so that $\max_t\|R_nm_t\|=o_{a.s.}\left(p^{1/2}\right)$. Then on Λ_{pn} $\max_t\sup_{\lambda\in\Lambda_{pn}}|\lambda'm_t|\leq\epsilon/p^{1/2}\max_t\|R_nm_t\|=o_{a.s.}(1)$ which implies $(A6)$. ■

Lemma 8 *Assume that* ***M****(I),(II),* ***MU****,* ***LP****,* ***P****(I),(II) hold, and let*

$$\widehat{\lambda}:=\arg\max_{\lambda\in\Lambda_0}\sum\log\left(1+\lambda'm_t\right).$$

Then $\widehat{\lambda}$ *exists a.s. and* $\left\|S_n^{-1}\widehat{\lambda}\right\|=O_p\left(p^{1/2}\right)$.

Proof. On Λ_{pn} $\sum \log(1+\lambda' m_t)$ is twice continuously differentiable so that $\widehat{\lambda} := \arg\max_{\Lambda_{pn}} \sum \log(1+\lambda' m_t)$ exists $a.s.$ Let $\widehat{\lambda}_n = \rho\theta$ where $\rho = \left\| \widehat{\lambda}_n \right\|$, $\widehat{\lambda}_n = S_n^{-1}\widehat{\lambda}$ and $\|\theta\| = 1$, and note that $\widehat{\lambda}_n$ solves $0 = \sum m_{tn} / \left(1 + \widehat{\lambda}_n' m_{tn}\right)$. A Taylor expansion shows that

$$\rho = \left[\theta' \sum m_{tn} m_{tn}' \theta / (1 + \rho\theta' m_{tn})\right]^{-1} \theta' \sum m,$$

and note that for $\|\theta\| = 1$ $\theta' \sum m_{tn} m_{tn}' \theta \geq \varsigma_{\min}(\Sigma_{\omega p}) > 0$ $a.s.$by $(A4)$. Furthermore Lemma 7 implies that $1 + \lambda' m_t > 0$ $a.s.$ and thus

$$\begin{aligned} \rho &\leq \left(\theta' \sum m_{tn} m_{tn}' \theta\right)^{-1} \theta' \sum m_{tn} \left(1 + \rho \max_t \|m_{tn}\|\right) \\ &\leq C\theta' \sum m_{tn} \leq C \|\theta\| \left\| \sum m_{tn} \right\| = O_p\left(p^{1/2}\right) \end{aligned}$$

where the last equality follows by

$$\begin{aligned} &E \left\| \sum \Delta \mathbf{y}_{p,t-1}' \varepsilon_{p,t} / n \right\|^2 \leq E \left\| \sum \Delta \mathbf{y}_{p,t-1}' / n \right\|^2 E\left(\varepsilon_t^2 | \mathcal{F}_{t-1}\right) + \\ &E \left\| \sum \Delta \mathbf{y}_{p,t-1}' \right\|^2 E\left(\left(\varepsilon_{p,t} - \varepsilon_t\right)^2 | \mathcal{F}_{t-1}\right) \leq O(p) + o\left(p^{1-2k}\right) \to 0. \end{aligned}$$

Note that for any $\delta_n = O(1)$ $\left\| S_n^{-1} \widehat{\lambda} \right\| \leq \delta_n p^{1/2}$ implies that

$$p^{1/2} \left\| R_n^{-1} \widehat{\lambda} \right\| \leq \varsigma_{\max}\left(S_n R_n^{-1}\right) \delta_n p = O\left(n^{-1/2+\beta}\right) p \to 0$$

for $p = o\left(n^{1/2-\beta}\right)$. Thus $p^{1/2} \left\| S_n^{-1} \widehat{\lambda} \right\| < \epsilon$ $a.s.$ that is $\left\| \widehat{\lambda} \right\| \in int\{\Lambda_{pn}\}$ and hence $\left\| \widehat{\lambda} \right\| \in int\{\Lambda_0\}$. By concavity of $\sum \log(1+\lambda' m_t)$ and convexity of Λ_0 it then follows that $\widehat{\lambda} := \arg\max_{\Lambda_0} \sum \log(1+\lambda' m_t)$ exists $a.s.$ ■

Lemma 9 *Assume that* ***M****(I),(II),* ***MU****,* ***LP****,* ***P****(I),(II) hold. Then on the set* Λ_{pn} *as defined in Lemma 7*

$$\widehat{\lambda}_n = \left(\sum m_{tn} m_{tn}'\right)^{-1} \sum m_{tn} + \zeta_n \tag{A7}$$

where $\|\zeta_n\| = o_p(1)$ *for* $p = o\left(n^{1/3}\right)$.

Proof. By Lemmae 7 and 8 $\widehat{\lambda} \in int\{\Lambda_{pn}\}$ where $\widehat{\lambda}$ solves $0 = \sum m_t / \left(1 + \widehat{\lambda}' m_t\right)$ $a.s.$ and $\widehat{\lambda}' m_t$ is uniformly small on Λ_{pn}. Thus by Taylor expansion

$$0 = \sum m_{tn} \left[1 - \widehat{\lambda}_n' m_{tn}\right] + \eta_n / 2$$

with $\eta_n := \sum m_{tn}\left(\widehat{\lambda}_n' m_{tn}\right)^2 / \left(1+\widehat{\lambda}_n' m_{tn}\right)$. Note that $\|\eta_{tn}\|$ is bounded by

$$\max_t \left|\widehat{\lambda}_n' m_{tn}\right| \left\|\widehat{\lambda}_n\right\| \varsigma_{\max}\left(\sum m_{tn}m_{tn}'\right) / \left|1+\lambda_n' m_{tn}\right| \le C o_p(1)$$

as $(A8)$ implies that if $p = o\left(n^{1/3}\right)$ then $p^{1/2}\left\|\sum m_{tn}m_{tn}' - \Sigma_{\omega p}\right\| = o_p(1)$ hence $\varsigma_{\max}\left(p^{1/2}\sum m_{tn}m_{tn}'\right) < C$. Also by $(A4)$ $\varsigma_{\min}\left(\sum m_{tn}m_{tn}'\right) \ge C$ and thus $\left(\sum m_{tn}m_{tn}'\right)^{-1}$ exists $a.s.$. Then $\widehat{\lambda}_n = \left(\sum m_{tn}m_{tn}'\right)^{-1}\sum m_{tn} + \zeta_n$ where $\zeta_n := \left(\sum m_{tn}m_{tn}'\right)^{-1}\eta_n$ and $(A7)$ follows by CMT since $\left\|\left(\sum m_{tn}m_{tn}'\right)^{-1}\right\| \le C$ $a.s.$ by $(A4)$. ■

Lemma 10 *Assume that **M**(I),(II), **MU**, **LP**, **P**(I),(II) hold. Then on the set Λ_{pn} as defined in Lemma 9*

$$2\sum \log\left(1+\widehat{\lambda}' m_t\right) = \sum m_{tn}'\left(\sum m_{tn}m_{tn}'\right)^{-1}\sum m_{tn} + \xi_n$$

where $|\xi_n| = o_p(1)$ for $p = o\left(n^{1/4}\right)$.

Proof. Similarly to the proof of the previous Lemma, a Taylor expansion about 0 gives

$$2\sum \log\left(1+\widehat{\lambda}' m_t\right) = 2\sum\left[\widehat{\lambda}' m_t - \left(\widehat{\lambda}' m_t\right)^2/2\right] + \xi_n$$

where

$$|\xi_n| \le \sum\left|\widehat{\lambda}_n' m_{tn}\right|^3 \le \max_t\left|\widehat{\lambda}_n' m_{tn}\right|\sum\left(\widehat{\lambda}_n' m_{tn}\right)^2 \le C o_p(1)$$

as long as $p\left\|\sum m_{tn}m_{tn}' - \Sigma_{\omega p}\right\| = o_p(1)$. The latter follows by $(A4)$ with $p = o\left(n^{1/4}\right)$. The conclusion follows using $(A7)$ and CMT. ■

In the following let $\Phi_{\tau_{pn}}$ denote a sphere centred at ϕ_0 with radius $\tau_n(p) := \tau_{pn}$ such that $\tau_{pn}p \to 0$ as $n \to \infty$, $\widetilde{m}_t$ denote m_t evaluated at some $\widetilde{\phi} \in \Phi_{\tau_{pn}}$ and $l(\lambda,\phi) := \sum\left(\log 1 + \lambda' m_t\right)$.

Lemma 11 *Let $\widetilde{\phi} \in \Phi_{\tau_{pn}}$. Then*

$$\left\|\sum \widetilde{m}_{tn}\widetilde{m}_{tn}' - \sum m_{tn}m_{tn}'\right\| = o_p(1). \qquad \text{(A8)}$$

Proof. By triangle and Cauchy-Schwartz inequalities

$$\begin{aligned}\left\|\sum \widetilde{m}_{tn}\widetilde{m}_{tn}' - \sum m_{tn}m_{tn}'\right\| &\le \sum\left|\widetilde{\varepsilon}_{t,p}^2 - \varepsilon_{t,p}^2\right| \left\|Y_{t-1,n}\right\|^2 \\ \le\ & \sum\left|\left(\widetilde{\varepsilon}_{t,p} - \varepsilon_{t,p}\right)^2 - \left(\widetilde{\varepsilon}_{t,p} - \varepsilon_{t,p}\right)\varepsilon_{t,p}\right| \left\|Y_{t-1,n}\right\|^2 \\ \le\ & C\left\|\widetilde{\phi} - \phi_0\right\|\sum\left\|Y_{t-1,n}\right\|^2 = O_p\left(\tau_{pn}p\right) \to 0.\end{aligned}$$

■

Lemma 12 *Assume that* ***M****(I),(II),* ***MU****, and* ***P****(I) hold. Then on the set* Λ_{pn} *as defined in Lemma 7 for any* $\widetilde{\phi} \in \Phi_{\tau_{pn}}$

$$\max_t \sup_{\widetilde{\phi}\in\Phi_{pn}} \sup_{\lambda\in\Lambda_{pn}} |\lambda' \widetilde{m}_t| = o_p(1), \quad \text{and } \Lambda_{pn} \subset \Lambda_0(\phi) \quad a.s. \tag{A9}$$

where $\Lambda_0(\phi)$ *is an open interval (depending on* ϕ*) containing 0.*

Proof. Note that $\widetilde{m}_t = \begin{bmatrix} y_{t-1} & \Delta\mathbf{y}'_{p,t-1} \end{bmatrix}' \left(\varepsilon_{t,p} - \Delta\mathbf{y}'_{p,t-1}\left(\widetilde{\phi} - \phi_0\right)\right)$. By Lemma 4 $\max_t \|\Delta\mathbf{y}_{p,t-1}\| = o_{a.s.}\left(p^{1/2}n^{\beta}\right)$ and thus by triangle inequality

$$\max_t \sup_{\widetilde{\phi}\in\Phi_{pn}} \sup_{\lambda\in\Lambda_{pn}} |\lambda' \widetilde{m}_t| \le \epsilon/p^{1/2} \max_t \left(\|R_n m_t\| + \left\|R_n \begin{bmatrix} y_{t-1} & \Delta\mathbf{y}'_{p,t-1} \end{bmatrix}' \Delta\mathbf{y}'_{p,t-1}\right\| \tau_{pn}\right)$$
$$= \epsilon/p^{1/2}\left(o_{a.s.}\left(p^{1/2}\right) + o_{a.s.}\left(\tau_{pn}p\right)\right),$$

so that $A9$ follows as in Lemma 7. ■

Lemma 13 *Assume that* ***M****(I),(II),* ***MU****,* ***LP****,* ***P****(I),(II) hold. Let* $\widetilde{\phi} \in \Phi_{\tau_{pn}}$, $\widehat{\lambda} := \arg\max_{\lambda\in\Lambda_0(\widetilde{\phi})} l\left(\lambda, \widetilde{\phi}\right)$ *and suppose that* $\|\widetilde{m}_{tn}\| = O_p\left(p^{1/2}\right)$*. Then* $\widehat{\lambda}$ *exists a.s. and* $\left\|S_n^{-1}\widehat{\lambda}\right\| = O_p\left(p^{1/2}\right)$.

Proof. As in the proof of Lemma 8 $l\left(\lambda, \widetilde{\phi}\right)$ is twice differentiable on Λ_{pn} and thus $\widehat{\lambda} := \arg\max_{\lambda\in\Lambda_{pn}} l\left(\lambda, \widetilde{\phi}\right)$ exists *a.s.* Using the same argument and notation of Lemma 8 we get

$$\rho = \left[\theta' \sum \widetilde{m}_{tn}\widetilde{m}'_{tn}\theta / \left(1 + \rho\theta'\widetilde{m}_{tn}\right)\right]^{-1} \theta' \sum \widetilde{m}_{tn}.$$

Then by $(A4)$, $(A8)$, $(A9)$ and CMT $\varsigma_{\max}\left(\sum \widetilde{m}_{tn}\widetilde{m}'_{tn}\right) \le C$ and $\max_t |\lambda'_n \widetilde{m}_{tn}| = o_p(1)$, and thus

$$\rho \le C\|\theta\| \left\|\sum \widetilde{m}_{tn}\right\| = O_p\left(p^{1/2}\right),$$

so that $\left\|S_n^{-1}\widehat{\lambda}\right\| = O_p\left(p^{1/2}\right)$, and the conclusion follows as in Lemma 8. ■

Lemma 14 *Assume that* ***M****(I),(II),* ***MU****,* ***LP****,* ***P****(I),(II) hold. Let* $\widetilde{\phi} \in \Phi_{\tau_{pn}}$ *and let* $\widehat{\phi} := \arg\min_{\widetilde{\phi}\in\Phi_{\tau_{pn}}} l\left(\lambda, \tilde{\phi}\right)$*. Then* $\|\widehat{m}_{tn}\| = O_p\left(p^{1/2}\right)$.

Proof. Let $\overline{\lambda} = R_n \widehat{m}_{tn} / \|\widehat{m}_{tn}\| p^{1/2}$ so that $\overline{\lambda} \in \Lambda_{np}$ and hence $\max_t \left|\overline{\lambda}' \widehat{m}_t\right| = o_{a.s.}(1)$ by $(A9)$. By $(A4)$, $(A8)$, $(A9)$ and CMT $\varsigma_{\max}\left(\sum \widehat{m}_{tn}\widehat{m}_{tn}' / (1+\lambda'\widehat{m}_{tn})^2\right) \leq C$ so that similarly to Lemma 10 a Taylor expansion shows

$$\begin{aligned} l\left(\overline{\lambda}, \widehat{\phi}\right) &= \overline{\lambda}' S_n^{-1} \widehat{m}_{tn} - \overline{\lambda}' S_n^{-1} \sum \widehat{m}_{tn} \widehat{m}_{tn}' S_n^{-1} \overline{\lambda} / 2 \left(1 + \overline{\lambda}' \widehat{m}_{tn}\right)^2 + o_p(1) \\ &\geq \overline{\lambda}' S_n^{-1} \widehat{m}_{tn} - C \overline{\lambda}' S_n^{-1} S_n^{-1} \overline{\lambda} + o_p(1) \\ &= \varsigma_{\max}\left(R_n S_n^{-1}\right) \|\widehat{m}_{tn}\| / p^{1/2} - C \varsigma_{\max}\left(R_n S_n^{-1}\right)^2 / p. \end{aligned}$$

Note that by the definition of $\widehat{\phi}$ and Lemma 10 $\sup_{\lambda \in \Lambda_{pn}(\widehat{\phi})} l\left(\lambda, \widehat{\phi}\right) \leq \sup_{\lambda \in \Lambda_{pn}} l(\lambda, \phi_0) = O_p(p)$ and thus

$$O_p(p) \geq l\left(\overline{\lambda}, \widehat{\phi}\right) \geq \varsigma_{\max}\left(R_n S_n^{-1}\right) \|\widehat{m}_{tn}\| / p^{1/2} - C O_p\left(\varsigma_{\max}\left(R_n S_n^{-1}\right)^2\right) / p$$

which implies that $\|\widehat{m}_{tn}\| \leq O_p\left(p^{3/2} / n^{1/2-\beta}\right) + O_p\left(n^{1/2-\beta} / p^{1/2}\right) = o_p\left(p^{1/2}\right) + O_p\left(p^{1/2}\right).$
■

Lemma 15 *Assume that **M**(I),(II), **MU**, **LP**, **P**(I),(II) hold. Let* $\widetilde{\phi} \in \Phi_{\tau_{pn}}$ *and suppose that* $\tau_{pn} = o\left(1/\left(n^\beta p^{1/2}\right)\right)$*. Then there exists a* $\widehat{\phi} \in int\left\{\Phi_{\tau_{pn}}\right\}$ *minimising* $l\left(\lambda, \widetilde{\phi}\right)$ *a.s. satisfying* $0 = \partial l\left(\lambda, \widetilde{\phi}\right) / \partial \phi$

Proof. By $(A9)$ and a Taylor expansion we have uniformly on $\Lambda_0\left(\widetilde{\phi}\right)$

$$l\left(\lambda, \tilde{\phi}\right) = \sum \widetilde{m}_{tn}' \left(\sum \widetilde{m}_{tn} \widetilde{m}_{tn}'\right)^{-1} \sum \widetilde{m}_{tn} / 2 + \tilde{\xi}_n$$

where $\left|\tilde{\xi}_n\right| \leq o_{a.s}\left(\left|\sum \widetilde{m}_{tn}' \left(\sum \widetilde{m}_{tn} \widetilde{m}_{tn}'\right)^{-1} \sum \widetilde{m}_{tn}\right|\right)$.

Let

$$\begin{aligned} \sum \Delta \mathbf{y}_{p,t-1} Y_{t-1,n}' \left(\sum m_{tn} m_{tn}'\right)^{-1} \sum Y_{t-1,n} \Delta \mathbf{y}_{p,t-1}' &:= P_n^2, \\ \sum \Delta \mathbf{y}_{p,t-1} Y_{t-1,n}' \left(\sum m_{tn} m_{tn}'\right)^{-1} \sum m_{tn} &:= Q_n \end{aligned}$$

so that using $l(\lambda, \phi_0) = \sum m_{tn}' \left(\sum m_{tn} m_{tn}'\right)^{-1} \sum m_{tn} + \xi_n$ and $\widetilde{m}_{tn} = m_{tn} - Y_{t-1,n} \Delta \mathbf{y}_{p,t-1}' \times \left(\widetilde{\phi} - \phi_0\right)$

$$l\left(\lambda, \tilde{\phi}\right) - l(\lambda, \phi_0) = -Q_n\left(\widetilde{\phi} - \phi_0\right) + \left(\widetilde{\phi} - \phi_0\right)' P_n^2 \left(\widetilde{\phi} - \phi_0\right) / 2 + \zeta_n \qquad \text{(A10)}$$

where $|\zeta_n| \leq o_{a.s}\left(\left|\left(\widetilde{\phi}-\phi_0\right)' \widetilde{R}_n^2\left(\widetilde{\phi}-\phi_0\right)\right|\right)$ where

$$\widetilde{R}_n^2 := \sum \Delta \mathbf{y}_{p,t-1} Y_{t-1,n}' \left[\left(\sum \widetilde{m}_{tn}\widetilde{m}_{tn}'\right)^{-1} - \left(\sum m_{tn}m_{tn}'\right)^{-1}\right] \sum Y_{t-1,n}\Delta \mathbf{y}_{p,t-1}'.$$

By CMT and $(A4)$ $\varsigma_{\min}\left(P_n^2/n\right) \geq C$, which implies $\left(P_n^2\right)^{-1}$ exists $a.s$. Define $\overline{\phi} = \phi_0 - \left(P_n^2\right)^{-1} Q_n$ and note that $(A10)$ can be written as

$$l\left(\lambda,\tilde{\phi}\right) - l\left(\lambda,\overline{\phi}\right) = \left(\overline{\phi}-\widetilde{\phi}\right)' P_n^2 \left(\overline{\phi}-\widetilde{\phi}\right) - \left(\overline{\phi}-\phi_0\right)' \overline{R}_n^2 \left(\overline{\phi}-\phi_0\right) + \zeta_n \qquad \text{(A11)}$$

where $\overline{R}_n^2 := \sum \Delta \mathbf{y}_{p,t-1} Y_{t-1,n}' \left[\left(\sum \overline{m}_{tn}\overline{m}_{tn}'\right)^{-1} - \left(\sum m_{tn}m_{tn}'\right)^{-1}\right] \sum Y_{t-1,n}\Delta \mathbf{y}_{p,t-1}'$. Since

$$\left\|\overline{R}_n^2\right\| \leq \left\|\sum \Delta \mathbf{y}_{p,t-1} Y_{t-1,n}'\right\|^2 \left\|\left(\sum \overline{m}_{tn}\overline{m}_{tn}'\right)^{-1} - \left(\sum m_{tn}m_{tn}'\right)^{-1}\right\|$$

where $\left\|\left(\sum \overline{m}_{tn}\overline{m}_{tn}'\right)^{-1} - \left(\sum m_{tn}m_{tn}'\right)^{-1}\right\| = o_p(1)$ using $(A3)$, Lemma 11, triangle inequality and the same arguments of Berk (1974, p. 493), and that

$$E\left\|\sum \Delta \mathbf{y}_{p,t-1} Y_{t-1,n}'/n^{1/2} - \begin{bmatrix} 0 & \Sigma_p \end{bmatrix}'\right\|^2 = O_p(p/n),$$

which implies that $\left\|\sum \Delta \mathbf{y}_{p,t-1} Y_{t-1,n}'\right\|^2 = O_p(p)$, it follows by CMT that $\left\|\overline{R}_n^2\right\|$ and $\left\|\widetilde{R}_n^2\right\| = \delta_n$ where $\delta_n = o_p(1)$ for $p = o\left(n^{1/4}\right)$. Note that for any θ such that $\|\theta\| \leq 1$ $\tau_{pn}^2 \theta' P_n^2 \theta \geq \varsigma_{\min}\left(P_n^2\right)$, $n^{1/2}\left\|\left(\overline{\phi}-\widetilde{\phi}\right)\right\| \leq \varsigma_{\max}\left(\left(P_n^2\right)^{-1}\right)\left\|Q_n/n^{1/2}\right\| = O_p\left(p^{1/2}\right) = o\left(\tau_{pn}\right)$ and hence $\left(\overline{\phi}-\widetilde{\phi}\right) \in \Phi_{\tau_{pn}}$, and that by triangle inequality $\widetilde{\phi}-\phi_0 \in 2\Phi_{\tau_{pn}}$. Thus since $\sup_{\widetilde{\phi}\in\Phi_{\tau_{pn}}}\left|\left(\overline{\phi}-\phi_0\right)' \overline{R}_n^2\left(\overline{\phi}-\phi_0\right)\right| \leq 4\delta_n\tau_{pn}^2$, and $\sup_{\widetilde{\phi}\in\Phi_{\tau_{pn}}}|\zeta_n| \leq \delta_n\tau_{pn}^2$, we then have that

$$\min \sup_{\widetilde{\phi}\in\Phi_{\tau_{pn}}} l\left(\lambda,\tilde{\phi}\right) - l\left(\lambda,\bar{\phi}\right) \geq \left(\varsigma_{\min}\left(P_n^2\right)/2 - 5\delta_n\right)\tau_{pn}^2.$$

Because $\delta_n = o_p(1)$ and $\tau_{pn} > 0$ for each n it follows that $l\left(\lambda,\tilde{\phi}\right)$ attains $a.s.$ its minimum value at some point $\widehat{\phi} \in int\left\{\Phi_{\tau_{pn}}\right\}$ and since $l\left(\lambda,\tilde{\phi}\right)$ is continuous on $\Phi_{\tau_{pn}}$ it follows that $\widehat{\phi}$ satisfies $0 = \partial l\left(\lambda,\tilde{\phi}\right)/\partial\phi$ $a.s.$ ■

Lemma 16 *Assume that **M**(I),(II), **MU**, **LP**, **P**(I),(II) hold, and let*

$$\widehat{\phi} := \arg\min_{\widetilde{\phi}\in\Phi_{\tau_{pn}}} l\left(\lambda,\tilde{\phi}\right).$$

Then $\left\|\widehat{\phi}-\phi_0\right\| = O_p\left((p/n)^{1/2}\right)$.

Proof. As in the proof of the previous lemma, using a Taylor expansion and $\widehat{m}_{tn} = m_{tn} - Y_{t-1,n}\Delta\mathbf{y}'_{p,t-1}\left(\widehat{\phi} - \phi_0\right)$ it follows that

$$\sum \widehat{m}'_{tn}\left(\sum \widehat{m}_{tn}\widehat{m}'_{tn}\right)^{-1}\sum \widehat{m}_{tn} = \sum m'_{tn}\left(\sum m_{tn}m'_{tn}\right)^{-1}\sum m_{tn} -$$
$$2\sum m'_{tn}\left(\sum m_{tn}m'_{tn}\right)^{-1}\sum Y_{t-1,n}\Delta\mathbf{y}'_{p,t-1}\left(\widehat{\phi} - \phi_0\right) +$$
$$\left(\widehat{\phi} - \phi_0\right)'\sum \Delta\mathbf{y}_{p,t-1}Y'_{t-1,n}\left(\sum m_{tn}m'_{tn}\right)^{-1}\sum Y_{t-1,n}\Delta\mathbf{y}'_{p,t-1}\left(\widehat{\phi} - \phi_0\right).$$

Let

$$\left(\widehat{\phi} - \phi_0\right)'\sum \Delta\mathbf{y}_{p,t-1}Y'_{t-1,n}\left(\sum m_{tn}m'_{tn}\right)^{-1}\sum Y_{t-1,n}\Delta\mathbf{y}'_{p,t-1}\left(\widehat{\phi} - \phi_0\right) := F_n^2,$$
$$\sum \widehat{m}'_{tn}\left(\sum \widehat{m}_{tn}\widehat{m}'_{tn}\right)^{-1}\sum \widehat{m}_{tn} := \widehat{M}^2, \quad \sum m'_{tn}\left(\sum m_{tn}m'_{tn}\right)^{-1}\sum m_{tn} := M_n^2.$$

By repeated use of the triangle inequality we have that

$$F_n^2 \leq \widehat{M}_n^2 + M_n^2 + 2F'_nM_n \leq \widehat{M}_n^2 + M_n^2 + 2F'_n\left(\widehat{M}_n + M_n\right)$$

where $F'_n = \left(\widehat{\phi} - \phi_0\right)'\sum \Delta\mathbf{y}_{p,t-1}Y'_{t-1,n}\left(\sum m_{tn}m'_{tn}\right)^{-1/2}$, $M'_n = \sum m'_{tn}\left(\sum m_{tn}m'_{tn}\right)^{-1/2}$ and $\widehat{M}_n$ is defined similarly. Subtracting $2F'_n\left(\widehat{M}_n + M_n\right)$, adding $\left(\widehat{M}_n + M_n\right)^2$ and taking the square roots from both sides yields $\left\|F_n - \left(M_n + \widehat{M}_n\right)\right\| \leq 2^{1/2}\left\|\widehat{M}_n + M_n\right\|$. Note that $M_n^2 \leq C\left\|m_{tn}\right\|^2$ and $\widehat{M}_n^2 \leq C\left\|\widehat{m}_{tn}\right\|^2$ so that by Lemmae 12, 16 and CMT both M_n^2 and $\widehat{M}_n^2$ are $O_p(p)$. Then again by triangle inequality $\left\|F_n - \left(M_n + \widehat{M}_n\right)\right\| \geq \left\|F_n\right\| - \left\|M_n + \widehat{M}_n\right\|$ hence

$$\left\|F_n\right\|^2 \leq C\left\|\widehat{M}_n + M_n\right\|^2 = O_p(p).$$

Since $\sum \Delta\mathbf{y}_{p,t-1}Y'_{t-1,n}\left(\sum m_{tn}m'_{tn}\right)^{-1}\sum Y_{t-1,n}\Delta\mathbf{y}'_{p,t-1}/n \xrightarrow{p} \Sigma_p$ we have that $F_n^2 \geq \varsigma_{\min}(\Sigma_p)\, n\left\|\widehat{\phi} - \phi_0\right\|^2 \geq n\left\|\widehat{\phi} - \phi_0\right\|^2 C$ and thus

$$\left\|\widehat{\phi} - \phi_0\right\|^2 C \leq F_n^2 \leq \widehat{M}_n^2 + M_n^2 = O_p(p/n).$$

■

Proof of Theorem 1. Lemma 16 shows that $\left\|\widehat{\phi} - \phi_0\right\| = O_p\left((p/n)^{1/2}\right)$, which implies that $\widehat{\phi} - \phi_0 \in int\{\Phi\}$ since $(p/n)^{1/2} = o\left(p^{-1/2}n^{-\beta}\right)$ and hence by Lemma 15 there exists a $\widehat{\phi}$ such that $0 = \partial l\left(\lambda, \tilde{\phi}\right)/\partial\phi$ *a.s.* Then by Lemma 14$\|\widehat{m}_{tn}\| = O_p\left(p^{1/2}\right)$ and hence by Lemma 13 $\widetilde{\lambda} = \arg\max_{\lambda\in\Lambda_0} l\left(\lambda,\phi\right)$ exists *a.s.* Lemma 12 $\max_t\left|\widehat{\lambda}'\widehat{m}_t\right| = o_{a.s.}\left(1\right)$ so that for all λ and ϕ in a neighbourhood of $\left(\widehat{\lambda},\widehat{\phi}\right)$ $l\left(\lambda,\phi\right)$ is twice continuously differentiable and $\partial^2\left(\widehat{\lambda},\widehat{\phi}\right)/\partial\lambda\partial\lambda' = -\sum\widehat{m}_t\widehat{m}_t'/\left(1+\widehat{\lambda}'\widehat{m}_t\right)^2$ and note that $S_n\left[\partial^2\left(\widehat{\lambda},\widehat{\phi}\right)/\partial\lambda\partial\lambda'\right]S_n$ is nonsingular *a.s.*. By the implicit function theorem there is a continuously differentiable function $\lambda\left(\phi\right)$ such that $0 = \partial l\left(\lambda\left(\phi\right),\phi\right)/\partial\phi$ so that by the envelope theorem

$$0 = n^{-1}\partial l\left(\widehat{\lambda},\widehat{\phi}\right)/\partial\phi = n^{-1}\sum\left(\partial\widehat{m}_{tn}/\partial\phi\right)'\widehat{\lambda}_n/\left(1+\widehat{\lambda}_n'\widehat{m}_{tn}\right). \tag{A12}$$

Then similarly to Lemma 9 $\widehat{\lambda}_n = \left(\sum\widehat{m}_{tn}\widehat{m}_{tn}'\right)^{-1}\sum\widehat{m}_{tn} + o_p\left(1\right)$ and using $\widehat{m}_{tn} = m_{tn} - Y_{t-1,n}\Delta\mathbf{y}_{p,t-1}'\left(\widehat{\phi}-\phi_0\right)$ in $(A12)$ gives

$$\begin{aligned} 0 &= \sum\left(n^{-1/2}\partial\widehat{m}_{tn}/\partial\phi\right)'\left[\sum\widehat{m}_{tn}\widehat{m}_{tn}'/\left(1+\widehat{\lambda}'\widehat{m}_t\right)^2\right]^{-1} \\ &\quad \sum\left[m_{tn} - Y_{t-1,n}\Delta\mathbf{y}_{p,t-1}'\left(\widehat{\phi}-\phi_0\right)/n^{1/2}\right]\left(1+\widehat{\lambda}_n'\widehat{m}_{tn}\right)^{-1} + o_p\left(1\right). \end{aligned}$$

Note that $n^{-1/2}\sum\partial\widehat{m}_{tn}/\partial\phi = \sum Y_{t-1,n}\Delta\mathbf{y}_{p,t-1}'/n^{1/2} \xrightarrow{p} \begin{bmatrix} 0 & \Sigma_p \end{bmatrix}'$ whereas by $(A8)$, $(A9)$ and CMT $\left[\sum\widehat{m}_{tn}\widehat{m}_{tn}'/\left(1+\widehat{\lambda}'\widehat{m}_t\right)^2\right]^{-1} \xrightarrow{p} \Sigma_{\omega p}^{-1}$, so that

$$\begin{aligned} &\sum\left(n^{-1/2}\partial\widehat{m}_{tn}/\partial\phi\right)'\left[\sum\widehat{m}_{tn}\widehat{m}_{tn}'/\left(1+\widehat{\lambda}'\widehat{m}_t\right)^2\right]^{-1}\sum Y_{t-1,n}\Delta\mathbf{y}_{p,t-1}'/n^{1/2} \xrightarrow{p} \\ &\begin{bmatrix} 0 & \Sigma_p \end{bmatrix}\Sigma_{\omega p}^{-1}\begin{bmatrix} 0 & \Sigma_p \end{bmatrix}' \end{aligned}$$

and hence $\left(\widehat{\phi}-\phi_0\right) = \Sigma_p^{-1}\sum\Delta\mathbf{y}_{p,t-1}\varepsilon_{tp}$. The latter implies that $\sum\widehat{m}_{tn} = \sum m_{tn} - Y_{t-1,n}\Delta\mathbf{y}_{p,t-1}'\left(\widehat{\phi}-\phi_0\right) = \sum\begin{bmatrix} y_{t-1}\varepsilon_{tp}/n & 0' \end{bmatrix}'$ and hence by $(A8)$, $(A9)$ and Taylor expansion

$$\begin{aligned} 2l\left(\widehat{\lambda},\widehat{\phi}\right) &= \sum\widehat{m}_{tn}'\Sigma_{\omega p}^{-1}\sum\widehat{m}_{tn} + o_p\left(1\right) = \left(\sum y_{t-1}\varepsilon_{pt}\right)^2/\left(\sigma^2\sum y_{t-1}^2\right) \\ &\xrightarrow{w} \left(\int_0^1 B\left(r\right)dB\left(r\right)\right)^2/\int_0^1 B^2\left(r\right)dr. \end{aligned}$$

■

Lemma 17 *Let* $y_t = (1-\gamma/n)\, y_{t-1} + u_t$ *where* u_t *is as in Lemma 3 and* $y_0 = 0$*. Then for any* $1 \geq \delta > 1/2$ $\max_t |y_t| = o_{a.s.}\left(n^{\delta}\right)$.

Proof. Let $\beta^{\gamma}_{j,t} = (1-\gamma/n)^{t-j}$; by recursive substitution $y_t = \sum_{j=1}^{t} \beta^{\gamma}_{j,t} u_j$. Using the same notation of Lemma 4, the Beveridge-Nelson decomposition and summation by parts give

$$y_t = \beta^{\gamma}_{t,t}\left[\Psi(1)\, S_t + \eta_t\right] - \sum_{j=1}^{t} S_{j-1}\left(\beta^{\gamma}_{j,t} - \beta^{\gamma}_{j-1,t}\right). \tag{A13}$$

By Chow's strong law of large number $S_t = o_{a.s.}\left(n^{\delta}\right)$ so that by Lemma 3 and triangle inequality

$$\max_t |y_t| \leq \left|\beta^{\gamma}_{t,t}\right| \max_t |S_t + \zeta_t| + \max_t |S_{t-1}| \sum_{j=1}^{t} \left|\beta^{\gamma}_{j,t} - \beta^{\gamma}_{j-1,t}\right| = o_{a.s.}\left(n^{\delta}\right)$$

since $\sum_{j=1}^{t} \left|\beta^{\gamma}_{j,t} - \beta^{\gamma}_{j-1,t}\right| < \infty$. ■

Lemma 18 *Let* $y_t = (1-\gamma/n)\, y_{t-1} + u_t$ *where* u_t *is as in Lemma 4 and* $y_0 = 0$*. Then*

$$\sum y_{t-1}\varepsilon_{p,t} = \sum y_{t-1}\varepsilon_t + o_p(n).$$

Proof. The proof is similar to the one given in Lemma 3.1 of Chang and Park (2002). For notational convenience let $\varepsilon_{p,t} - \varepsilon_t = \epsilon_{pt}$, and note that $\sum y_{t-1}\varepsilon_{p,t} = \sum y_{t-1}\varepsilon_t + \sum y_{t-1}\epsilon_{pt}$. By $(A13)$

$$\sum y_{t-1}\left(\varepsilon_{p,t} - \varepsilon_t\right) = \left[\beta^{\gamma}_{t-1,t-1} \sum \left[\Psi(1)\, S_{t-1} + \zeta_{t-1}\right] + \sum \sum_{j=2}^{t-1} S_{j-2}\left(\beta^{\gamma}_{j,t-1} - \beta^{\gamma}_{j-1,t-1}\right)\right] \epsilon_{pt}.$$

Consider first $\sum \beta^{\gamma}_{t-1,t-1} S_t \epsilon_{pt}$ and note that $\epsilon_{pt} = \sum_{l=p+1}^{\infty} \gamma_{p,l}\varepsilon_{t-l}$ where $\sum_{l=p+1}^{\infty} \gamma^2_{p,l} \leq C \sum_{l=p+1}^{\infty} \alpha_l^2 = o\left(p^{-2k}\right)$. Let δ_{jk} denote the Kronecker delta; then

$$\begin{aligned}
\sum \beta^{\gamma}_{t-1,t-1} S_{t-1}\epsilon_{pt} &= \sum_{l=p+1}^{\infty} \gamma_{p,l}\beta^{\gamma}_{t-1,t-1} \sum \sum_{i=1}^{t-1} \varepsilon_{t-i}\varepsilon_{t-l} \\
&= \beta^{\gamma}_{t-1,t-1}\left[n\sigma^2 \sum_{l=p+1}^{\infty} \left|\gamma_{p,l}\right| + \sum_{l=p+1}^{\infty} \gamma_{p,l} \sum\sum_{i=1}^{t-1}\left(\varepsilon_{t-i}\varepsilon_{t-l} - \sigma^2\delta_{ij}\right)\right] \\
&\leq \beta^{\gamma}_{t-1,t-1}\left\{n\sigma^2 \sum_{l=p+1}^{\infty} \left|\gamma_{p,l}\right| + \sum_{l=p+1}^{\infty} \left|\gamma_{p,l}\right| E^{1/2}\left[\sum\sum_{i=1}^{t-1}\left(\varepsilon_{t-i}\varepsilon_{t-l} - \sigma^2\delta_{il}\right)\right]^2\right\} \\
&\leq Co\left(np^{-k}\right).
\end{aligned}$$

Next

$$\begin{aligned}\sum \beta^{\gamma}_{t-1,t-1}\zeta_{t-1}\epsilon_{pt} &= \sum_{j=0}^{\infty}\alpha_j \sum_{l=p+1}^{\infty}\gamma_{p,l}\sum \beta^{\gamma}_{t-1,t-1}\varepsilon_{t-i-1}\varepsilon_{t-l} \\ &= \beta^{\gamma}_{t-1,t-1}\left[n\sigma^2\sum_{l=p+1}^{\infty}\alpha_{l-1}\gamma_{p,l}+\sum_{j=0}^{\infty}\alpha_j\sum_{l=p+1}^{\infty}\gamma_{p,l}\sum\left(\varepsilon_{t-i-1}\varepsilon_{t-l}-\sigma^2\delta_{i+1l}\right)\right] \\ &= C\left(o\left(np^{-k}\right)+o_{a.s}\left(n^{\eta}p^{-k}\right)\right)\end{aligned}$$

for any $1 \geq \eta > 1/2$ by Chow's strong law. Similarly for the last the term it can be shown that $\sum\sum_{j=2}^{t-1} S_{j-2}\left(\beta^{\gamma}_{j,t-1}-\beta^{\gamma}_{j-1,t-1}\right)\epsilon_{pt} = o_p\left(np^{-k}\right)$ using $\sum\left|\left(\beta^{\gamma}_{j,n}-\beta^{\gamma}_{j-1,n}\right)\right| < \infty$. ■

Proof of Theorem 2. Note that

$$\max_t |y_{t-1}(\varepsilon_{p,t}-\varepsilon_t)/n| \leq \max_t |y_{t-1}\varepsilon_t/n| \left|\sum_{j=p+1}^{\infty}\psi_j\right| = o_{a.s.}(1)\, o\left(p^{-s}\right) = o_{a.s.}(1)$$

by Lemma 17 and **MU**, so that

$$\max_t |y_{t-1}\varepsilon_{p,t}/n| \leq \max_t |y_{t-1}\varepsilon_t/n| + \max_t |y_{t-1}(\varepsilon_{p,t}-\varepsilon_t)/n| = o_{a.s.}(1). \qquad \text{(A14)}$$

Note that by Phillips (1987*b*) $\sum y^2_{t-1}/n^2 \xrightarrow{w} \sigma^2\Psi(1)^2\int_0^1 J_{\gamma}{}^2(r)\,dr := \sigma^2\omega_{\gamma}$ whereas for each $j = 1, ..., p$ it is possible to show that $\sum\left(y_{t-1}u_{t-j-1}/n\right)^2 = O_p(1)$ so that $E\left\|\sum y_{t-1}\Delta\mathbf{y}'_{p,t-1}/n^{3/2}\right\|^2 = O(p/n)$. Since $y_t = O_p\left(n^{1/2}\right)$, $\sum \Delta\mathbf{y}_{p,t-1}\Delta\mathbf{y}'_{p,t-1}/n \xrightarrow{p} \Sigma_p$ it then follows that the conclusions of Lemmae 4-5 and hence of Lemma 6 are still valid. Furthermore in view of $(A14)$ Lemma 7 is also valid. Thus by the same arguments of Lemmae 8-10 it is possible to show that

$$2W(\beta_n,\phi_0) = \sum m_{tn}(\beta_n,\phi_0)'\left(\sum m_{tn}(\beta_n,\phi_0)\, m_{tn}(\beta_n,\phi_0)'\right)^{-1}\sum m_{tn}(\beta_n,\phi_0) + o_p(1)$$

Note that $\sum m_{tn}(\beta_n,\phi_0)\, m_{tn}(\beta_n,\phi_0)' \xrightarrow{w} diag\left(\sigma^2\left[\begin{array}{cc}\omega_{\gamma} & \Sigma_p\end{array}\right]\right)$ and hence ϕ_0 and β are asymptotically independent under the sequence of local alternatives β_n. Thus Lemmae 11-16 can be used exactly as in the proof of Theorem 1 to show that there exists a $\widehat{\phi} \in int\{\Phi_{pn}\}$ that solves $0 = \partial W\left(\beta_n,\widetilde{\phi}\right)/\partial\phi$. Let $m_{tn}\left(\beta,\widehat{\phi}\right) = Y_{t-1,n}\left(y_t - \beta_n y_{t-1} - \Delta\mathbf{y}'_{p,t-1}\widehat{\phi}\right)$ and note that $\sum m_{tn}\left(\beta_n,\widehat{\phi}\right) = \sum m_{tn}(\beta_n,\phi_0) - Y_{t-1,n}\Delta\mathbf{y}'_{p,t-1}\left(\widehat{\phi}-\phi_0\right)$. The block diagonality of $\sum m_{tn}(\beta_n,\phi_0)\, m_{tn}(\beta_n,\phi_0)$ implies that $\left(\widehat{\phi}-\phi_0\right)=\Sigma_p^{-1}\Delta\mathbf{y}_{p,t-1}\varepsilon_{tp}$ and hence $\sum m_{tn}\left(\beta_n,\widehat{\phi}\right)=\sum\left[\begin{array}{cc} y_{t-1}\varepsilon_{tp}/n & 0'\end{array}\right]'$ as in Theorem 1.Thus by Taylor expansion

$$\begin{aligned}
2W\left(\beta_n,\widehat{\phi}\right) &= 2\sum\log\left(1+\widehat{\lambda}'m_t\left(\beta_n,\widehat{\phi}\right)\right)=\sum m_{tn}\left(\beta_n,\widehat{\phi}\right)'M_{nn}^{-1}\sum m_{tn}\left(\beta_n,\widehat{\phi}\right)+o_p(1)\\
&= \sum m_{tn}\left(\beta_n,\widehat{\phi}\right)\left[diag\left(\sigma^2\left[\begin{array}{cc}\omega_\gamma & \Sigma_p\end{array}\right]\right)\right]^{-1}\sum m_{tn}\left(\beta_n,\widehat{\phi}\right)+o_p(1)\\
&= \sum y_{t-1}\varepsilon_{pt}^2/\left(\sigma^2\sum y_{t-1}^2\right)+o_p(1)\,.
\end{aligned}$$

By Lemma 18 $\sum y_{t-1}\varepsilon_{pt}/n=\sum y_{t-1}\varepsilon_t/n+o_p(1)$ so that the conclusion follows by Phillips (1987*b*) and CMT. ■

5 Tables

Table 1. Finite sample size and power of $W\left(1,\widehat{\phi}\right) := W$ and ADF_t^2 with $N\left(0,1\right)$ errors at 0.10 and 0.05 nominal level

		AR				MA			
β	θ	W		ADF_t^2		W		ADF_t^2	
1	−.8	.116	.053	.113	.054	.127	.080	.123	.064
1	−.5	.108	.057	.109	.052	.120	.069	.115	.054
1	−.2	.108	.047	.110	.048	.118	.069	.120	.063
1	0	.105	.052	.107	.053	.120	.073	.117	.062
1	.2	.110	.058	.108	.064	.131	.075	.125	.064
1	.5	.112	.059	.109	.056	.128	.073	.122	.066
1	.8	.106	.057	.107	.055	.132	.082	.124	.064
.95	−.8	.805	.643	.811	.617	.605	.373	.524	.343
.95	−.5	.832	.666	.836	.652	.408	.232	.368	.207
.95	−.2	.800	.623	.791	.601	.357	.208	.378	.228
.95	0	.804	.603	.795	.591	.399	.217	.368	.199
.95	.2	.806	.618	.791	.613	.289	.144	.397	.174
.95	.5	.799	.539	.752	.554	.367	.216	.339	.198
.95	.8	.687	.508	.689	.492	.274	.204	.311	.178
.90	−.8	.988	.948	.996	.972	.889	.690	.880	.687
.90	−.5	.988	.970	.994	.964	.627	.485	.688	.503
.90	−.2	.990	.956	.987	.943	.615	.433	.606	.416
.90	0	.984	.955	.985	.947	.594	.404	.590	.388
.90	.2	.991	.964	.989	.956	.493	.399	.568	.376
.90	.5	.989	.934	.981	.916	.580	.396	.568	.397
.90	.8	.990	.968	.994	.964	.465	.304	.457	.298

Table 2. Finite sample size and power of $W\left(1, \widehat{\phi}\right) := W$ and ADF_t^2 with $\chi_4^2 - 4$ errors at 0.10 and 0.05 nominal level

		AR				MA			
β	θ	W		ADF_t^2		W		ADF_t^2	
1	$-.8$	.107	.057	.102	.053	.123	.070	.124	.068
1	$-.5$	.107	.061	.106	.053	.124	.086	.120	.065
1	$-.2$	.105	.060	.104	.056	.118	.077	.122	.067
1	0	.106	.057	.102	.052	.115	.068	.111	.065
1	.2	.110	.058	.104	.055	.128	.075	.125	.070
1	.5	.107	.058	.104	.052	.114	.077	.109	.060
1	.8	.112	.062	.107	.059	.131	.089	.117	.069
.95	.-8	.832	.650	.812	.628	.598	.487	.589	.448
.95	-.5	.822	.612	.792	.595	.393	.285	.377	.224
.95	-.2	.805	.633	.797	.616	.400	.215	.397	.205
.95	0	.799	.610	.778	.582	.385	.199	.355	.195
.95	.2	.800	.603	.781	.581	.379	.194	.343	.178
.95	.5	.764	.576	.750	.549	.363	.201	.358	.197
.95	.8	.696	.496	.683	.479	.388	.213	.342	.187
.90	-.8	.976	.942	.994	.957	.843	.675	.863	.669
.90	-.5	.965	.932	.995	.955	.712	.423	.703	.446
.90	-.2	.945	.940	.994	.959	.604	.378	.587	.345
.90	0	.958	.930	.984	.938	.543	.367	.567	.344
.90	.2	.945	.921	.988	.948	.586	.377	.575	.369
.90	.5	.968	.925	.981	.928	.592	.392	.580	.362
.90	.8	.914	.815	.921	.804	.525	.354	.530	.336

Table 3. Finite sample size and power of $W\left(1, \widetilde{\phi}\right) := W$ and ADF_t^2 with t_5 errors at 0.10 and 0.05 nominal level

		AR				*MA*			
β	θ	W		ADF_t^2		W		ADF_t^2	
1	$-.8$	.143	.105	.118	.064	.123	.076	.118	.064
1	$-.5$	.109	.060	.108	.058	.118	.084	.113	.059
1	$-.2$	.104	.059	.103	.056	.125	.074	.114	.060
1	0	.109	.056	.105	.054	.136	.079	.113	.065
1	.2	.111	.055	.105	.053	.122	.073	.112	.065
1	.5	.108	.059	.104	.052	.114	.087	.119	.073
1	.8	.113	.065	.111	.057	.129	.78	.123	.076
.95	-.8	.875	.655	.831	.635	.597	.393	.593	.383
.95	-.5	.850	.649	.809	.625	.402	.294	.389	.275
.95	-.2	.847	.605	.804	.591	.383	.212	.343	.194
.95	0	.821	.610	.802	.602	.368	.207	.339	.199
.95	.2	.831	.623	.800	.597	.34	.23	.364	.207
.95	.5	.745	.588	.739	.532	.394	.163	.379	.152
.95	.8	.708	.513	.686	.485	.387	.173	.389	.162
.90	-.8	.980	.945	.991	.963	.742	.594	.896	.679
.90	-.5	.990	.966	.992	.958	.605	.472	.630	.458
.90	-.2	.983	.976	.949	.987	.573	.405	.529	.396
.90	0	.967	.960	.983	.959	.582	.397	.599	.384
.90	.2	.967	.932	.985	.947	.567	.374	.572	.371
.90	.5	.974	.921	.988	.939	.554	.342	.574	.355
.90	.8	.915	.788	.922	.779	.543	.304	.549	.315