JOURNAL OF TIME SERIES ANALYSIS
J. Time. Ser. Anal. **35:** 282–297 (2014)
Published online 7 February 2014 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1111/jtsa.12064

UNIFIED INTERVAL ESTIMATION FOR RANDOM COEFFICIENT AUTOREGRESSIVE MODELS

JONATHAN HILL[a] AND LIANG PENG[b,*]
[a] *Department of Economics, University of North Carolina, Chapel Hill, NC, USA.*
[b] *School of Mathematics, Georgia Institute of Technology, Atlanta, GA, USA.*

The consistency of the quasi-maximum likelihood estimator for random coefficient autoregressive models requires that the coefficient be a non-degenerate random variable. In this article, we propose empirical likelihood methods based on weighted-score equations to construct a confidence interval for the coefficient. We do not need to distinguish whether the coefficient is random or deterministic and whether the process is stationary or non-stationary, and we present two classes of equations depending on whether a constant trend is included in the model. A simulation study confirms the good finite-sample behaviour of our resulting empirical likelihood-based confidence intervals. We also apply our methods to study US macroeconomic data.

Received 12 July 2013; Revised 29 December 2013; Accepted 6 January 2014

Keywords: Empirical likelihood method; random coefficient autoregression; weighted estimation.

1. INTRODUCTION

Consider the following random coefficient autoregression (AR) model

$$X_j = \theta + (\psi + b_j)X_{j-1} + e_j, \quad j = 1, \dots, n, \tag{1}$$

where θ and ψ are real numbers and $\{(b_j, e_j)^{\mathrm{T}}\}_{j=1}^n$ is a sequence of i.i.d. random vectors with $Eb_j = 0$, $Ee_j = 0$, $Eb_j^2 = w^2 \geq 0$ and $Ee_j^2 = \sigma^2 > 0$. We allow b_j and e_j to be mutually dependent; hence, $\mathrm{cov}(b_j, e_j) \neq 0$ is possible. Throughout, we write $(\theta_0, \psi_0, w_0^2, \sigma_0^2)$ to denote the true value of $(\theta, \psi, w^2, \sigma^2)$ and write $\log^+(x) = \max\{\log x, 0\}$.

Assume for now that the constant trend θ_0 is known to be zero. That is, consider the random coefficient model (e.g. Quinn 1982):

$$X_j = (\psi + b_j)X_{j-1} + e_j. \tag{2}$$

In view of $(b_j, e_j)^{\mathrm{T}}$ being i.i.d., under the general conditions

$$E\log^+|e_1| < \infty \quad \text{and} \quad E\log^+|\psi_0 + b_1| < \infty, \tag{3}$$

it is known that there is a strictly stationary, non-anticipated solution to (2); that is, X_k is measurable with respect to the σ field generated by $\{(b_i, e_i) : i \leq k\}$, *if and only if*

$$-\infty \leq E\log|\psi_0 + b_1| < 0. \tag{4}$$

* Correspondence to: Liang Peng, School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332-0160, USA.
E-mail: peng@math.gatech.edu

See Quinn (1982) and Aue *et al.* (2006). The case $w_0^2 = 0$ corresponds to an AR(1), where (4) of course implies $|\psi_0| < 1$.

There is an increasing interest in studying the non-stationary case, and in allowing for any autoregressive parameter value $\psi_0 \in \mathbb{R}$. In particular, researchers have investigated the asymptotic behaviour of some conventional estimators such as least squares and quasi-maximum likelihood under condition (3), which includes the non-stationary case $E \log|\psi_0 + b_1| \geq 0$. Some details of these estimators follow for model (2).

Consider the strict random coefficient autoregressive (RCA) case: $w_0^2 > 0$. By noticing that $E(X_j|X_{j-1}) = \psi X_{j-1}$, one can estimate ψ in (2) by the least squares estimator

$$\hat{\psi}_{LS} = \frac{\sum_{j=1}^{n} X_j X_{j-1}}{\sum_{j=1}^{n} X_{j-1}^2}.$$

However, as shown in Hwang and Basawa (2005), the least squares estimator of ψ is inconsistent when (2) has a non-stationary solution, that is, $E \log|\psi_0 + b_1| \geq 0$. As an improvement, Hwang and Basawa (2005) and Hwang *et al.* (2006) proposed the weighted least squares estimator

$$\hat{\psi}_{WLS} = \frac{\sum_{j=1}^{n} \frac{X_j X_{j-1}}{\tilde{w}^2 X_{j-1}^2 + \tilde{\sigma}^2}}{\sum_{j=1}^{n} \frac{X_{j-1}^2}{\tilde{w} X_{j-1}^2 + \tilde{\sigma}^2}},$$

where $\tilde{w}^2$ and $\tilde{\sigma}^2$ are consistent estimators for w_0^2 and σ_0^2 respectively. They showed that $\hat{\psi}_{WLS}$ is consistent for ψ_0 and has a normal limit even when the sequence $\{X_j\}$ is non-stationary under the restrictive assumptions such that b_j is a binary random variable and ϵ_j has a symmetric distribution. It is known, however, that the variance σ_0^2 cannot be estimated consistently by the quasi-maximum likelihood estimator in the non-stationary case (Berkes *et al.* 2009).

The quasi-maximum likelihood estimator for the stationary case was studied by Aue *et al.* (2006). Recently, Aue and Horváth (2011) showed that the quasi-maximum likelihood estimator of ψ_0, defined as

$$\hat{\psi}_{QML} = \underset{\psi_1 \leq \psi \leq \psi_2}{\arg\max} \left(\sup_{w_1^2 \leq w^2 \leq w_2^2, \sigma_1^2 \leq \sigma^2 \leq \sigma_2^2} \frac{1}{n} \sum_{j=1}^{n} \left\{ -\frac{(X_j - \psi X_{j-1})^2}{w^2 X_{j-1}^2 + \sigma^2} - \log\left(w^2 X_{j-1}^2 + \sigma^2\right) \right\} \right),$$

is consistent for ψ_0 and asymptotically normal

$$\sqrt{n}\left\{\hat{\psi}_{QML} - \psi_0\right\} \xrightarrow{d} \mathrm{N}(0, \tau^2), \tag{5}$$

where

$$\tau^2 = \begin{cases} w_0^2 & \text{if } E\left(\log|\psi_0 + b_1|\right) \geq 0 \\ \frac{w_0^2 \alpha(4,2) + \sigma_0^2 \alpha(2,2)}{\alpha^2(2,1)} & \text{if } E\left(\log|\psi_0 + b_1|\right) < 0 \end{cases},$$

$\alpha(k,\gamma) = E[X_1^k / (w_0^2 X_1^2 + \sigma_0^2)^\gamma]$, as long as model (2) satisfies (3) and a pre-assigned parameter range is needed. That is, $\hat{\psi}_{QML}$ is consistent for ψ_0 only if we restrict estimation to an interval $[\psi_1, \psi_2]$ known to contain it. Although the quasi-maximum likelihood estimator for σ_0^2 is inconsistent in the case of

DOI: 10.1111/jtsa.12064

non-stationarity, the asymptotic variance τ^2 can still be estimated consistently by using the quasi-maximum likelihood estimator; hence, a confidence interval for ψ_0 can be obtained via (5); see Aue and Horváth (2011) for details.

The above estimation procedures are only valid for model (2), which presumes a zero constant trend $\theta_0 = 0$, and they require the assumption of a non-trivial random coefficient, that is, $w_0^2 > 0$. In general, these estimators are not robust to the strict AR case $w_0^2 = 0$ and are not appropriate for the general model (1) when θ_0 and $w_0^2 \geq 0$ are unknown. Elsewhere, for testing $H_0 : \psi_0 = 1$ against $H_a : |\psi_0| < 1$ under the general condition (3), Distaso (2008) proposed a Lagrange Multiplier type (LM-type) test, where w_0^2 is allowed to be zero, but the asymptotic limit of the proposed test statistic depends on whether $w_0^2 = 0$ or $w_0^2 > 0$. See also McCabe and Tremaynee (1995), Granger and Swanson (1997), and Nagakura (2009). Moreover, the literature is evidently silent on inference on the RCA parameter ψ_0 when the unknown constant trend θ is present (cf. Aue *et al.* 2006, Hwang and Basawa 2005, Hwang *et al.* 2006, Aue and Horváth 2011).

In this article, we exploit the empirical likelihood method to deliver a unified confidence interval for ψ_0 in the general model (1) under the general assumption (3) without estimating w_0^2 and σ_0^2. The empirical likelihood method has been extended to many different fields including linear and nonlinear time series as a powerful non-parametric method in constructing confidence regions and hypothesis tests. We refer the reader to Owen (1990, 2001) for an overview.

Specifically, first ignoring the constant trend and working with model (2), we propose to apply the empirical likelihood method to some weighted estimating equations as in Chan *et al.* (2012). It turns out that the proposed empirical likelihood method works without distinguishing whether $w_0^2 = 0$ or $w_0^2 > 0$ or whether the process is stationary or non-stationary; it allows for an arbitrary but non-perfect dependence between b_j and e_j and performs well in finite samples. Second, for the general model (1), we exploit a different set of estimating equations for valid empirical likelihood inference on ψ_0 irrespective of the constant trend value θ_0. We do not consider estimation of the variance of the random coefficient b_j. That topic is treated in, among others, Giraitis *et al.* (2010). Consult that source for references.

Further, our methods for constructing valid confidence intervals for ψ_0 for any constant trend case do not extend to a linear time trend model $X_j = \theta_1 + \theta_2 \times j + (\psi + b_j)X_{j-1} + e_j$ except in simple special cases. In general, there are cases where the gradient matrix associated with the three estimating equations is not positive definite; hence, the empirical likelihood method does not apply to our proposed weighted least squares equations. It remains unknown to us whether it is possible to provide a unified interval estimation for RCA models with a possible linear time trend.

We organize this article as follows. In Section 2, we present the new methodologies first for model (2) and then for the general model (1). A simulation study is given in Section 3, and in Section 4, we apply our methods to macroeconomic and financial data. All proofs are put in the Appendix.

2. METHODOLOGY

We first tackle the estimation of the RCA slope ψ_0 for model (2), which does not include a constant trend. We then tackle the more difficult problem of inference on ψ_0 for model (1) with an unknown constant trend θ_0.

2.1. Random coefficient autoregression without a trend

Consider model (2). Since the least squares estimator of ψ_0 is inconsistent in the non-stationary case, we cannot employ the empirical likelihood method with the score equation $\sum_{j=1}^{n}(X_j - \psi X_{j-1})X_{j-1} = 0$. Motivated by the weighted least squares estimator in Hwang and Basawa (2005) and the unified interval estimation for AR models in Chan *et al.* (2012), we propose to employ the empirical likelihood method to the weighted least squares score equation

  J. Time. Ser. Anal. **35:** 282–297 (2014)
DOI: 10.1111/jtsa.12064

$$\sum_{j=1}^{n}(X_j - \psi X_{j-1})\frac{X_{j-1}}{1+X_{j-1}^2}. \tag{6}$$

Note that the weight function $1 + X_{j-1}^2$ can be replaced by any function $g(X_{j-1}^2)$ where $g : [0,\infty) \to (0,\infty)$ is a Borel function satisfying $x^2/g(x) \to 1$ as $|x| \to \infty$. More specifically, we define the empirical likelihood function for ψ as

$$L_n(\psi) = \sup\left\{\prod_{j=1}^{n}(np_j) : p_1 \geq 0, \cdots, p_n \geq 0, \sum_{j=1}^{n} p_j = 1, \sum_{j=1}^{n} p_j \frac{(X_j - \psi X_{j-1})X_{j-1}}{1+X_{j-1}^2} = 0\right\}.$$

By the Lagrange multiplier technique, we obtain the well-known log-empirical likelihood representation

$$l_n(\psi) = -2\log L_n(\psi) = 2\sum_{j=1}^{n}\log\left\{1 + \lambda\frac{(X_j - \psi X_{j-1})X_{j-1}}{1+X_{j-1}^2}\right\},$$

where the multiplier $\lambda = \lambda(\psi)$ satisfies

$$\sum_{j=1}^{n}\frac{\frac{(X_j-\psi X_{j-1})X_{j-1}}{1+X_{j-1}^2}}{\frac{1+\lambda(X_j-\psi X_{j-1})X_{j-1}}{1+X_{j-1}^2}} = 0. \tag{7}$$

The following theorem shows that Wilks' theorem holds for the proposed empirical likelihood method under the condition (3) with exclusion of the non-stationary AR case $w_0^2 = 0$ and $|\psi_0| \geq 1$ but otherwise unifies stationary and non-stationary cases.

Theorem 2.1. Suppose model (2) satisfies (3), and $E|e_1|^{2+r} < \infty$ and $E|b_1|^{2+r} < \infty$ for some $r > 0$. Further, when $w_0^2 > 0$, assume there exists no constant a such that $P(b_1 = ae_1) = 1$, and when $w_0^2 = 0$, assume that $|\psi_0| < 1$. Then $l_n(\psi_0)$ converges in distribution to a chi-square limit with one degree of freedom as $n \to \infty$.

Based on the above theorem, an empirical likelihood confidence interval for ψ_0 with significance level $\beta \in (0,1)$ is

$$I_\beta = \left\{\psi : l_n(\psi) \leq \chi_{1,\beta}^2\right\},$$

where $\chi_{1,\beta}^2$ denotes the βth quantile of a chi-square distribution with one degree of freedom. In view of linearity and exact identification, the maximum empirical likelihood estimator for ψ_0 via minimizing the empirical likelihood function $l_n(\psi)$ is

$$\hat{\psi}_{MEL} = \frac{\sum_{j=1}^{n}\frac{X_j X_{j-1}}{1+X_{j-1}^2}}{\sum_{j=1}^{n}\frac{X_{j-1}^2}{1+X_{j-1}^2}}.$$

Theorem 2.2. Under the conditions of Theorem 2.1, we have $n^{1/2}(\hat{\psi}_{MEL} - \psi_0) \xrightarrow{d} \mathrm{N}(0, \mathcal{V}^2)$, where $\mathcal{V}^2 = w_0^2 E\left[X_{j-1}^4/\left(1+X_{j-1}^2\right)^2\right] + \sigma_0^2 E\left[X_{j-1}^2/\left(1+X_{j-1}^2\right)^2\right]$ if $E\log|\psi_0 + b_1| < 0$ and otherwise $\mathcal{V}^2 = w_0^2 > 0$.

DOI: 10.1111/jtsa.12064

When $w_0^2 = 0$ and $|\psi_0| \geq 1$, we have $|X_n| \xrightarrow{p} \infty$ as $n \to \infty$, which implies

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \frac{(X_j - \psi_0 X_{j-1})X_{j-1}}{1 + X_{j-1}^2} = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \frac{\epsilon_j X_{j-1}}{1 + X_{j-1}^2} = o_p(1).$$

Hence, Wilks' theorem does not apply. This explains why the empirical likelihood method and estimator in Theorems 2.1 and 2.2 have to exclude this non-stationary AR case. To include this case, we now propose to combine the above empirical likelihood method with the one in Chan *et al.* (2012) so as to unify all cases for model (2).

Note that when $w_0^2 = 0$ and $|X_n| \xrightarrow{p} \infty$ as $n \to \infty$,

$$\frac{1}{n} \sum_{j=1}^{n} \frac{\left(X_j - \hat{\psi}_{MEL} X_{j-1}\right)^2 X_{j-1}^2}{\left(1 + X_{j-1}^2\right)^2} \leq \left\{ \frac{1}{n} \sum_{j=1}^{n} \frac{X_{j-1}^2}{\left(1 + X_{j-1}^2\right)^2} \right\}^{\frac{1}{2}} \leq \frac{1}{n^{\frac{1}{4}}} \tag{8}$$

holds with probability tending to 1. For other cases, (8) holds with probability tending to 0. Let $I(\cdot)$ denote the indicator function. Then we propose to define the empirical likelihood function for ψ as

$$L_n^*(\psi) = \sup \left\{ \prod_{j=1}^{n} (np_j) : p_1 \geq 0, \cdots, p_n \geq 0, \ \sum_{j=1}^{n} p_j = 1, \sum_{j=1}^{n} p_j Z_{n,j}(\psi) = 0 \right\},$$

where

$$Z_{n,j}(\psi) = \frac{(X_j - \psi X_{j-1})X_{j-1}}{1 + X_{j-1}^2}(1 - \Delta_n) + \Delta_n \frac{(X_j - \psi X_{j-1})X_{j-1}}{\left(1 + X_{j-1}^2\right)^{\frac{1}{2}}} \tag{9}$$

and

$$\Delta_n = I\left(\frac{1}{n} \sum_{j=1}^{n} \frac{\left(X_j - \hat{\psi}_{MEL} X_{j-1}\right)^2 X_{j-1}^2}{\left(1 + X_{j-1}^2\right)^2} \leq \left\{ \frac{1}{n} \sum_{j=1}^{n} \frac{X_{j-1}^2}{\left(1 + X_{j-1}^2\right)^2} \right\}^{\frac{1}{2}} \leq \frac{1}{n^{\frac{1}{4}}} \right).$$

As before, by the Lagrange multiplier technique, we obtain

$$l_n^*(\psi) = -2 \log L_n^*(\psi) = 2 \sum_{j=1}^{n} \log \left\{ 1 + \lambda^* Z_{n,j}(\psi) \right\},$$

where $\lambda^* = \lambda^*(\psi)$ satisfies

$$\sum_{j=1}^{n} \frac{Z_{n,j}(\psi)}{1 + \lambda^* Z_{n,j}(\psi)} = 0. \tag{10}$$

The next result shows that Wilks' theorem holds for the above proposed empirical likelihood method, even in the strict AR case $w_0^2 = 0$.

Theorem 2.3. Suppose model (2) satisfies (3), and $E|e_1|^{2+r} < \infty$ and $E|b_1|^{2+r} < \infty$ for some $r > 0$. Further, when $w_0^2 > 0$, assume there exists no constant a such that $P(b_1 = ae_1) = 1$. Then $l_n^*(\psi_0)$ converges in distribution to a chi-square limit with one degree of freedom as $n \to \infty$.

DOI: 10.1111/jtsa.12064

Based on the above theorem, an empirical likelihood confidence interval for ψ_0 with level β is $I_\beta^* = \left\{\psi : l_n^*(\psi) \le \chi^2_{1,\beta}\right\}$. The maximum empirical likelihood estimator for ψ_0 via minimizing $l_n^*(\psi)$ is

$$\hat{\psi}_{MEL*} = (1 - \Delta_n)\hat{\psi}_{MEL} + \Delta_n \hat{\psi}_{CLP},$$

where

$$\hat{\psi}_{CLP} = \frac{\sum_{j=1}^{n} \frac{X_j X_{j-1}}{\sqrt{1+X_{j-1}^2}}}{\sum_{j=1}^{n} \frac{X_{j-1}^2}{\sqrt{1+X_{j-1}^2}}}.$$

Theorem 2.4. Let the conditions of Theorem 2.3 hold. If $w_0^2 = 0$ and $|\psi_0| \ge 1$, then

$$\frac{1}{n}\sum_{i=1}^{n} \frac{X_{j-1}^2}{\sqrt{1+X_{j-1}^2}} \sqrt{n}\left\{\hat{\psi}_{MEL*} - \psi_0\right\} \xrightarrow{d} N\left(0, \sigma_0^2\right),$$

and otherwise $\sqrt{n}\{\hat{\psi}_{MEL*} - \psi_0\} \xrightarrow{d} \mathrm{N}(0, \mathcal{V}^2)$, where $\mathcal{V}^2$ is given in Theorem 2.2.

Remark 1. In an RCA setting, note that Berkes *et al.* (2009) assumed serial independence and $\mathrm{cov}(b_1, e_1) = 0$, and Aue and Horváth (2011) assumed serial and mutual independence for b_j and e_j. Both imply that b_j and e_j are not perfectly dependent: there exists no constant a such that $P(b_1 = ae_1) = 1$. However, our simulation study below indicates that the quasi-maximum likelihood estimator for ψ_0 still works for the linear dependence case $\mathrm{cov}(b_1, e_1) \ne 0$. When $w_0^2 > 0$, Lemma A.1 in the Appendix may be employed to extend the results in Berkes *et al.* (2009) to the case that there exists no constant a such that $P(b_1 = ae_1) = 1$ instead of $\mathrm{cov}(b_1, e_1) = 0$, allowing arbitrary but not perfect dependence.

Remark 2. Since we do not study the asymptotic normality of estimators for σ^2 and w^2, the moment conditions in Theorems 2.1 and 2.3 are sufficient. Otherwise, one may need at least a finite fourth moment. Although the asymptotic variances given in Theorems 2.2 and 2.4 have a different formula for the stationary case and non-stationary case, Theorems 2.1 and 2.3 can be employed to construct a confidence interval for ψ since the empirical likelihood method does not explicitly require a case-dependent consistent estimator for the asymptotic variance (cf. Owen 1990).

2.2. Random coefficient autoregression with a constant trend

Our next task is to treat the general model (1). As before, one may apply the profile empirical likelihood method to some similar weighted-score equations. Unfortunately, Wilks' theorem does not hold for a direct application. To see this clearly, consider at first an AR model with a unit root; hence, $w_0 = 0$ and $|\psi_0| = 1$. We use the following estimating instrument throughout:

$$\xi_j(\gamma) := \frac{X_j}{\left(1 + X_j^2\right)^\gamma} \text{ for } \gamma > 0 \text{ and } \xi_j := \xi_j\left(\frac{1}{2}\right).$$

DOI: 10.1111/jtsa.12064

In this case, θ and ψ can in principle be estimated by solving least squares score equations:

$$\begin{cases} \sum_{j=1}^{n}(X_j - \theta - \psi X_{j-1}) = 0, \\ \sum_{j=1}^{n}(X_j - \theta - \psi X_{j-1})X_{j-1} = 0. \end{cases} \tag{11}$$

It can be easily shown that the joint limit of $\frac{1}{\sqrt{n}}\sum_{j=1}^{n}(X_j - \theta_0 - \psi_0 X_{j-1})$ and $\sum_{j=1}^{n}(X_j - \theta_0 - \psi_0 X_{j-1})X_{j-1}/\sqrt{\sum_{j=1}^{n} X_{j-1}^2}$ cannot be normally distributed; hence Wilks' theorem does not hold for a direct application of the empirical likelihood method to the score equations (11). Similarly, Wilks' theorem fails for an application of the empirical likelihood method to the score equations

$$\begin{cases} \sum_{j=1}^{n}(X_j - \theta - \psi X_{j-1}) = 0, \\ \sum_{j=1}^{n}(X_j - \theta - \psi X_{j-1})\xi_{j-1} = 0. \end{cases}$$

Continuing with the simple AR model with a unit root, here, we propose to exploit the following score equations:

$$\begin{cases} \sum_{j=1}^{n}(X_j - \theta - \psi X_{j-1}) = 0, \\ \sum_{j=1}^{n}\{(X_j - \theta - \psi X_{j-1})\xi_{j-1}(\gamma) + W_j\} = 0, \end{cases}$$

where $W_1, \ldots, W_n$ are i.i.d. with $N(0, \tilde{\sigma}^2)$ and $\tilde{\sigma}^2 > 0$. The data-dependent choice of $\tilde{\sigma}$ is given in formula (13) for the estimating equations $Z_{n,j}$ in (9). Note that when $|\psi_0| = 1$, we have $|X_j| \xrightarrow{p} \infty$ as $j \to \infty$. Therefore, the second equation asymptotically equals $\sum_{j=1}^{n} W_j$ in the case of $|\psi_0| = 1$ and $\gamma > 1/2$, and the joint limit of $\frac{1}{\sqrt{n}}\sum_{j=1}^{n}(X_j - \theta_0 - \psi_0 X_{j-1})$ and $\frac{1}{\sqrt{n}}\sum_{j=1}^{n} W_j$ is a normal distribution. A large γ makes the term $(X_j - \theta_0 - \psi X_{j-1})\xi_{j-1}(\gamma)$ disappear faster. On the other hand, writing

$$\sum_{j=1}^{n}\frac{(X_j - \theta_0 - \psi X_{j-1})X_{j-1}}{\left(1 + X_{j-1}^2\right)^{\gamma}} = \sum_{j=1}^{n}(X_j - \theta_0 - \psi_0 X_{j-1})\xi_{j-1}(\gamma) + \sum_{j=1}^{n}\frac{(\psi_0 - \psi)X_{j-1}^2}{\left(1 + X_{j-1}^2\right)^{\gamma}}$$

clearly shows we need $\gamma \le 1$ to be as small as possible to better detect the departure of ψ from the true value ψ_0 in the non-stationary case. That is, $\gamma \in (1/2, 1]$. Here, we propose to choose the middle value $\gamma = 0.75$, which works well in simulations when X_j is a non-stationary AR.

In the stationary AR or RCA case, trivially, the joint limit of $\frac{1}{\sqrt{n}}\sum_{j=1}^{n}(X_j - \theta_0 - \psi_0 X_{j-1})$ and $\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\{(X_j - \theta_0 - \psi_0 X_{j-1})\xi_{j-1}(\gamma) + W_j\}$ is again normal. This is true for any value of γ.

DOI: 10.1111/jtsa.12064

Finally, in the non-stationary RCA case, hence $w_0^2 > 0$ and $|X_n| \xrightarrow{p} \infty$, a large $\gamma > 1$ makes the term

$$(X_j - \theta_0 - \psi_0 X_{j-1})\xi_{j-1}(\gamma) = e_j \frac{X_{j-1}}{\left(1 + X_{j-1}^2\right)^{\gamma}} + b_j \frac{X_{j-1}^2}{\left(1 + X_{j-1}^2\right)^{\gamma}}$$

disappear faster. Now, if we write the equation as

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{j=1}^{n}(X_j - \theta_0 - \psi X_{j-1})\xi_{j-1}(\gamma) &= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}(X_j - \theta_0 - (\psi_0 + b_j)X_{j-1})\xi_{j-1}(\gamma)\\ &\quad + (\psi_0 - \psi)\frac{1}{\sqrt{n}}\sum_{j=1}^{n}X_{j-1}\xi_{j-1}(\gamma) + \frac{1}{\sqrt{n}}\sum_{j=1}^{n}b_j X_{j-1}\xi_{j-1}(\gamma),\end{aligned}$$

this reveals we still need a small γ to better detect the departure of ψ from its true value. When $\theta_0 = 0$ and $w_0^2 = 0$, we have $|X_n| = O_p(n^{1/2})$, which implies that we need $\gamma \le 3/2$. That is, $\gamma \in (1, 3/2]$. So we propose to choose the middle value $\gamma = 1.25$.

Given the above considerations, we are ready to modify the Section 2.1 empirical likelihood method such that it extends to model (1). Let $\tilde{\theta}$ and $\tilde{\psi}$ satisfy

$$\begin{cases}\displaystyle\sum_{j=1}^{n}\frac{X_j - \tilde{\theta} - \tilde{\psi}X_{j-1}}{1 + X_{j-1}^2} = 0\\ \displaystyle\sum_{j=1}^{n}\left(X_j - \tilde{\theta} - \tilde{\psi}X_{j-1}\right)\xi_{j-1}(1) = 0\end{cases} \tag{12}$$

and define sample variances

$$\tilde{\sigma}_1^2 = \frac{1}{n}\sum_{j=1}^{n}\left(X_j - \tilde{\theta} - \tilde{\psi}X_{j-1}\right)^2 \text{ and } \tilde{\sigma}_2^2 = \frac{1}{n}\sum_{j=1}^{n}\left(X_j - \tilde{\theta} - \tilde{\psi}X_{j-1}\right)^2\frac{1}{1 + X_{j-1}^2} \tag{13}$$

and an indicator

$$\tilde{\Delta}_n = I\left(\frac{1}{n}\sum_{j=1}^{n}\left(X_j - \tilde{\theta} - \tilde{\psi}X_{j-1}\right)^2\xi_{j-1}^2(1) \le \left\{\frac{1}{n}\sum_{j=1}^{n}\xi_{j-1}^2(1)\right\}^{\frac{1}{2}} \le \frac{1}{n^{\frac{1}{4}}}\right).$$

As in Section 2.1, $\tilde{\Delta}_n \xrightarrow{p} 1$ if X_j is non-stationary AR, and otherwise $\tilde{\Delta}_n \xrightarrow{p} 0$. Now define estimating equations $Z_{n,j}(\theta,\psi) = (Z_{n,j}^{(1)}(\theta,\psi), Z_{n,j}^{(2)}(\theta,\psi))^{\mathrm{T}}$ with

$$\begin{aligned}Z_{n,j}^{(1)}(\theta,\psi) &= \left(1 - \tilde{\Delta}_n\right)\frac{X_j - \theta - \psi X_{j-1}}{1 + X_{j-1}^2} + \tilde{\Delta}_n\left(X_j - \theta - \psi X_{j-1}\right),\\ Z_{n,j}^{(2)}(\theta,\psi) &= \left(1 - \tilde{\Delta}_n\right)\{(X_j - \theta - \psi X_{j-1})\xi_{j-1}(1.25) + \tilde{\sigma}_2 W_j\}\\ &\quad + \tilde{\Delta}_n\{(X_j - \theta - \psi X_{j-1})\xi_{j-1}(0.75) + \tilde{\sigma}_1 W_j\},\end{aligned}$$

where W_j's are i.i.d. random variables with $N(0, 1)$.

DOI: 10.1111/jtsa.12064

As before, we define the empirical likelihood function for $(\theta, \psi)^{\mathrm{T}}$ as

$$L_n(\theta, \psi) = \sup\left\{\prod_{j=1}^{n}(np_j) : p_1 \geq 0, \ldots, p_n \geq 0, \sum_{j=1}^{n} p_j = 1, \sum_{j=1}^{n} p_j Z_{n,j}(\theta, \psi) = 0\right\},$$

and it follows from the Lagrange multiplier technique that

$$\hat{l}_n(\theta, \psi) = -2\log L_n(\theta, \psi) = 2\sum_{j=1}^{n}\log\left\{1 + \lambda^{\mathrm{T}} Z_{n,j}(\theta, \psi)\right\},$$

where $\lambda = \lambda(\theta, \psi) \in \mathbb{R}^2$ satisfies

$$\sum_{j=1}^{n}\frac{Z_{n,j}(\theta, \psi)}{1 + \lambda^{\mathrm{T}} Z_{n,j}(\theta, \psi)} = 0.$$

Since we are interested in a confidence interval for ψ, we consider the profile empirical likelihood ratio function

$$\hat{l}_n^P(\psi) = \min_{\theta \in \mathbb{R}} \hat{l}_n(\theta, \psi).$$

The following theorem shows that Wilks' theorem holds for this profile empirical likelihood method.

Theorem 2.5. Suppose model (1) satisfies (3), and $E|e_1|^{2+r} < \infty$ and $E|b_1|^{2+r} < \infty$ for some $r > 0$. Further, when $w_0^2 > 0$, assume there exists no constant a such that $P(b_1 = ae_1) = 1$. Then $\hat{l}_n^P(\psi_0)$ converges in distribution to a chi-square limit with one degree of freedom as $n \to \infty$.

Based on the above theorem, an empirical likelihood confidence interval for ψ_0 with level β is $\hat{I}_\beta = \left\{\psi : \hat{l}_n^P(\psi) \leq \chi^2_{1,\beta}\right\}$.

Remark 3. Similar to Theorem 2.4, we can obtain the maximum empirical likelihood estimator for ψ_0 by minimizing $\hat{l}_n^P(\psi)$.

Remark 4. If $w_0 = 0$ were known, then one can replace the above Δ_n by 1. In this case, Theorem 2.5 generalizes the method of Chan *et al.* (2012) to the case of having a constant trend in an AR(1) model.

Remark 5. To reduce the effect of the random seed for the added pseudo-sample W_j, one can use $W_j = m^{-1/2}\sum_{l=1}^{m} W_{j,l}$ for a large m, where $W_{j,l}$'s are i.i.d. random variables with N(0, 1). In the simulation study below, for example, we use $m = 1000$ to generate the pseudo-sample W_j's.

3. SIMULATION STUDY

First, we consider model (2), that is, model (1) with known $\theta_0 = 0$, and compare the proposed empirical likelihood methods in Section 2.1 with the normal approximation method based on (5) and the estimator for the asymptotic variance τ^2 given by Aue and Horváth (2011) in terms of coverage probability and interval length. We draw 10,000 random samples with size $n = 50$ and 200 from model (2), with (b_j, e_j) having a bivariate normal distribution with $\sigma_0^2 = E\left[e_1^2\right] = 1$ and a correlation coefficient $\rho = 1/\sqrt{2}$. Consider the cases

  J. Time. Ser. Anal. **35:** 282–297 (2014)
DOI: 10.1111/jtsa.12064

$$\left(\psi_0, w_0^2\right) = (0.5, 0.25), (0.5, 1.5), (1, 3), (1.5, 1), (0.9, 0), (0.99, 0), (1, 0), (1.1, 0).$$

Notice that the last four cases are AR(1) models (i.e. $w_0^2 = 0$) and the first four cases are studied in the simulation section of Aue and Horváth (2011). For computing the coverage probability of the empirical likelihood confidence intervals (I_β and I_β^*), we employ the R package 'emplik'. For calculating the interval I_β^N based on (5), we choose the parameter range as

$$\psi_1 = -5|\psi_0|, \psi_2 = 5|\psi_0|, w_1^2 = w_0^2/5, w_2^2 = 5w_0^2 I(w_0 > 0) + I(w_0 = 0), \sigma_1^2 = \sigma_0^2/5, \sigma_2^2 = 5\sigma_0$$

and use the R package 'nlm' to maximize the likelihood function with initial values $\left(\hat{\psi}_{LS}, w_0^2, \sigma_0^2\right)$.

Table I. Random coefficient autoregression *without* trend

(ψ_0, w_0^2)	n	$I_{0.9}$	$I_{0.9}^*$	$I_{0.9}^N$	$I_{0.95}$	$I_{0.95}^*$	$I_{0.95}^N$
$(0.5, 0.25)$	50	−0.0151	−0.0129	−0.0134	−0.0072	−0.0069	−0.0067
$(0.5, 1.5)$	50	−0.0102	−0.0103	−0.0167*	−0.0073	−0.0072	−0.0073*
$(1.0, 3.0)$	50	−0.0128	−0.0128	−0.0232*	−0.0095	−0.0095	−0.0188*
$(1.5, 1.0)$	50	−0.0123	−0.0122	−0.0263*	−0.0107	−0.0105	−0.0211*
$(0.9, 0.0)$	50	−0.0172*	−0.0078	−0.0404*	−0.0125*	−0.0028	−0.0269*
$(0.99, 0.0)$	50	−0.0440*	0.0027	−0.0862*	−0.0283*	0.0010	−0.0664*
$(1.0, 0.0)$	50	−0.0427*	0.0050	−0.0993*	−0.0294	0.0009	−0.0821*
$(1.1, 0.0)$	50	−0.0419*	−0.0079	−0.1918*	−0.0369*	−0.0065	−0.1781*
$(0.5, 0.25)$	200	−0.0001	−0.0001	−0.0103*	0.0002	0.0002	−0.0062*
$(0.5, 1.5)$	200	0.0015	0.0015	−0.0010	0.0008	0.0008	−0.0007
$(1.0, 3.0)$	200	0.0007	0.0007	−0.0025	−0.0004	−0.0004	−0.0029
$(1.5, 1.0)$	200	−0.0026	−0.0026	−0.0048	−0.0028	−0.0028	−0.0045
$(0.9, 0.0)$	200	−0.0075	−0.0075	−0.0524*	−0.0020	−0.0020	−0.0442*
$(0.99, 0.0)$	200	−0.0287*	0.0042	−0.1258*	−0.0199*	0.0011	−0.1102*
$(1.0, 0.0)$	200	−0.0442*	0.0128	−0.1798*	−0.0289*	0.0049	−0.1604*
$(1.1, 0.0)$	200	−0.0484*	−0.0009	−0.0221*	−0.0400*	−0.0005	−0.0323*

Differences between the coverage probability and the nominal level are reported for the proposed empirical likelihood methods (I_β, I_β^*) and the normal approximation method based on $\hat{\psi}_{ML}(I_\beta^N)$ for levels $\beta = 0.9, 0.95$.
* p-value of the corresponding test is less than 0.1%.

Table II. Random coefficient autoregression *without* trend

(ψ_0, w_0^2)	$I_{0.9}$ $n = 50$	$I_{0.9}^*$ $n = 50$	$I_{0.9}^N$ $n = 50$	$I_{0.90}$ $n = 200$	$I_{0.90}^*$ $n = 200$	$I_{0.90}^N$ $n = 200$
$(0.5, 0.25)$	0.5267*	0.5242	0.4883*	0.2604	0.2604	0.2496*
$(0.5, 1.5)$	0.7053*	0.7068	0.6757*	0.3531	0.3531	0.3479*
$(1.0, 3.0)$	0.8498	0.8498	0.8175*	0.4115	0.4115	0.4073*
$(1.5, 1.0)$	0.4803	0.4806	0.4646*	0.2338	0.2338	0.2325*
$(0.9, 0.0)$	0.3184*	0.2905	0.2466*	0.1485	0.1485	0.1088*
$(0.99, 0.0)$	0.2258*	0.1744	0.1552*	0.0843*	0.0754	0.0480*
$(1.0, 0.0)$	0.2123*	0.1586	0.1417*	0.0687*	0.0528	0.0352*
$(1.1, 0.0)$	0.1193*	0.0287	0.0669*	0.0252*	0.0000	0.0196*

Interval lengths are reported for the proposed empirical likelihood methods (I_β, I_β^*) and the normal approximation method based on $\hat{\psi}_{ML}(I_\beta^N)$ for level $\beta = 0.9$.
* p-value of the corresponding test is less than 0.1%.

DOI: 10.1111/jtsa.12064

Table III. Random coefficient autoregression *with* a constant trend

(ψ_0, w_0^2)	$\tilde{I}_{0.9}$ $n = 50$	$\tilde{I}_{0.95}$ $n = 50$	$\tilde{I}_{0.90}$ $n = 200$	$\tilde{I}_{0.95}$ $n = 200$
(0.5, 0.25)	−0.0138	−0.0133	−0.0043	−0.0023
(0.5, 1.5)	−0.0137	−0.0108	−0.0014	−0.0017
(1.0, 3.0)	−0.0092	−0.0105	−0.0076	−0.0027
(1.5, 1.0)	−0.0117	−0.0075	−0.0080	−0.0045
(0.9, 0.0)	−0.0074	−0.0071	−0.0064	−0.0041
(0.99, 0.0)	−0.0083	−0.0078	−0.0063	−0.0031
(1.0, 0.0)	−0.0081	−0.0073	−0.0061	−0.0030
(1.1, 0.0)	−0.0113	−0.0118	−0.0169	−0.0133

Differences between the coverage probability and the nominal level are reported for the proposed empirical likelihood method ($\hat{I}_\beta$) for levels $\beta = 0.9, 0.95$.

In Table I, we report the differences between the coverage probability and the nominal level for these three intervals with level $\beta = 0.9, 0.95$. The average interval lengths for level $\beta = 0.9$ are reported in Table II. We observe first that the normal approximation method does not work for the case $w_0^2 = 0$. Second, the first empirical likelihood method fails for the case $w_0^2 = 0$ and $|\psi_0| \geq 1$. Third, the second empirical likelihood method works exceptionally well under any case, stationary or non-stationary and AR or RCA. Fourth, in the stationary AR case with $w_0^2 = 0$ and $|\psi_0| < 1$, the bands I_β and I_β^* should be similar and are equivalent asymptotically with probability approaching 1. Our results show that they are in fact identical for $n = 200$. Fifth, for the RCA case $w_0^2 > 0$, the empirical likelihood-based intervals are more accurate and have a slightly larger length than the normal approximation method for the small sample size $n = 50$, but all three methods are quite comparable when $n = 200$.

Recall that coverage probability is simply the simulation average $1/R \sum_{r=1}^{R} \mathcal{I}_r$ where $\mathcal{I}_r$ is the rth sample's i.i.d. Bernoulli random variable with $\mathcal{I}_r = 1$ when the computed interval contains the true ψ_0 (note that $R = 10{,}000$ samples). We also test whether the difference of coverage probability and nominal level among these three methods is significant by using the Bernoulli variables $\mathcal{I}_r$ generated from each sample. Since the second empirical likelihood method works for all cases, we test whether the first empirical likelihood method is significantly different from the second empirical likelihood method and whether the normal approximation method is significantly different from the second empirical likelihood method. In Tables I and II, a number with an asterisk means the p-value of the corresponding test is less than 0.1%; hence, at the 99.9% significance level, we reject the hypothesis that the corresponding coverage probability is equal to the coverage probability for the second empirical likelihood method. These numbers with an asterisk further confirm the above conclusions: although in some non-trivial cases, the coverage probabilities are similar across estimation methods, in general, the first empirical likelihood method and the normal approximation method are significantly different from the second empirical likelihood method with an exceptionally high degree of confidence.

Next, we consider model (1). Because the results in Aue and Horváth (2011) do not apply here, there is no other method to compare with the proposed empirical likelihood method in Theorem 2.5. We draw 10,000 random samples from model (1) with $\theta_0 = 1$, and all other parameters are the same as for model (2). Coverage probabilities are reported in Table III for various cases, which show that the proposed method works well for all considered cases.

4. DATA ANALYSIS

We now apply the proposed methods to various US macroeconomic and financial time series: log-real gross domestic product (GDP), log-M2 where M2 is a measure of the aggregate money supply, log-SP500, 3-month Treasury bill rate, and log-Industrial Production Index. All data are monthly, except GDP, which is quarterly, with samples starting prior to 1960 and ending in Oct. 2011. See Table IV for data descriptions, sample periods, and sample sizes.

DOI: 10.1111/jtsa.12064

Table IV. US macroeconomic data descriptions

Data	Details	n	SA	Source
Log-real GDP	Chained 2005 dollars	228	Yes	FSB
Log-M2	Aggregate money supply	637	Yes	FSB
Log-SP500	Open/adjusted close average	682	No	Yahoo
T-bill	3-month Treasury bill rate	682	No	FSB
Log-IPI	Industrial production index	682	Yes	FSB

SA, seasonally adjusted; FSB, Saint Louis Federal Reserve Bank; Yahoo, finance.yahoo.com.
GDP is quarterly for Jan. 1955–Oct. 2011; all other variables are monthly. M2 is for Jan. 1959–Oct. 2011 since previous months are not available at the FSB; all other variables are for Jan. 1955–Oct. 2011.

Table V. Estimation results: random coefficient autoregression without trend, weighted estimation and intervals

	$\hat{\psi}_{MEL*}$	$I^*_{0.9}$	$\hat{\psi}_{QML}$	$\hat{w}^2_{QML}$	$\hat{\sigma}^2_{QML}$	$I^N_{0.9}$
GDP	1.000850	(1.000734, 1.000963)	1.000862	3.9e−06	0.0987	(0.996984, 1.004739)
M2	1.000734	(1.000702, 1.000767)	1.000760	9.9e−07	0.0981	(0.998069, 1.003452)
SP500	1.000941	(1.000568, 1.001299)	1.001005	2.3e−06	0.0962	(0.997480, 1.004530)
T-bill	1.001173	(0.993558, 1.005207)	1.000722	0.0035	0.0217	(0.996240, 1.005204)
IPI	1.000631	(1.000450, 1.000737)	1.000630	1.8e−06	0.0933	(0.995769, 1.005491)

M2, aggregate money supply; T-bill, 3-month treasury bill rate; IPI, industrial production index; GDP, gross domestic product.

Table VI. Estimation results: random coefficient autoregression *with* a constant trend

	$\tilde{\theta}$	$\tilde{\psi}$	$\tilde{\theta}^*$	$\tilde{\psi}^*$	$\hat{I}_{0.9}$
GDP	0.02903	0.99756	1.89201	0.78637	(0.69410, 0.81163)
M2	0.00947	0.99948	0.02212	0.99780	(0.99629, 1.00646)
SP500	0.00779	0.99951	0.07432	0.98723	(0.95865, 1.01861)
T-bill	0.00159	1.00077	−0.01229	1.00878	(0.98483, 1.03279)
IPI	0.00988	0.99808	−0.02969	1.00847	(0.99022, 1.01759)

M2, aggregate money supply; T-bill, 3-month treasury bill rate; IPI, industrial production index; GDP, gross domestic product.
The weighted least squares estimator $(\tilde{\theta}, \tilde{\psi})$ solves equations (12) and the weighted least squares estimator $(\tilde{\theta}^*, \tilde{\psi}^*)$ solves $\sum_{j=1}^{n} Z_{n,j}(\theta, \psi) = \mathbf{0}$ in Section 2.2.

The natural log of income, production, money supply, and stock indices is traditionally assumed to be difference stationary in view of evidence by Dicky–Fuller unit root tests (see, e.g. Granger and Swanson 1997 for references). Real GDP, for example, is historically argued to have a stochastic trend component, in particular to have one positive unit root and therefore be difference stationary (e.g. Nelson and Plosser 1982). Despite a tradition of well-grounded criticisms against the random walk model of GDP, we assume GDP and all other variables satisfy model (1) for the sake of discussion. See Christiano and Eichenbaum (1990), Rudebusch (1993), and Gaffeo *et al.* (2005).

First, we apply both the quasi-maximum likelihood estimator and the empirical likelihood method based on Theorem 2.3 to model (2) for each data set. Corresponding estimators and intervals $I^*_{0.9}$ and $I^N_{0.9}$ are reported in Table V. From Table V, we observe that all data sets except the T-bill have a very small $\hat{w}_{MLE}$ and the resulting empirical likelihood intervals prefer a slightly explosive AR model.

Next, we apply model (1) to each data set. Since the quasi-maximum likelihood estimator was only derived for model (1) under the assumption that $\theta_0 = 0$ is known, in Table VI, we only report the weighted least squares estimator $(\tilde{\theta}, \tilde{\psi})$ solving equation (12), the weighted least squares estimator $(\tilde{\theta}^*, \tilde{\psi}^*)$ solving $\sum_{j=1}^{n} Z_{n,j}(\theta, \psi) = 0$ in Section 2.2, and the empirical likelihood confidence interval $\hat{I}_{0.9}$ based on Theorem 2.5. We observe that $\tilde{\theta}$ is quite small and close to zero and the interval $\hat{I}_{0.9}$ includes both $\tilde{\psi}$ and $\tilde{\psi}^*$ for all data sets except GDP. In the GDP case, $\tilde{\theta} \approx 0$ while $\tilde{\theta}^*$ is quite large, and $\tilde{\psi}$ is close to but less than unity while $\tilde{\psi}^*$ is roughly 0.79. Therefore, for all data sets except GDP, one may simply employ model (2), which does not have a constant trend, instead of

DOI: 10.1111/jtsa.12064

the general model (1). The strange behaviour of GDP may be due to the small sample size and large variability incurred by using the larger and therefore more general model (1).

APPENDIX A: PROOFS

First, we extend Lemma 1 in Berkes *et al.* (2009) to the case that there exists no constant a such that $P(b_1 = ae_1) = 1$.

Lemma A.1. Suppose model (2) satisfies (3) and $E\log|\psi_0 + b_1| \geq 0$ with $w_0^2 > 0$ and $\sigma_0^2 > 0$. Further assume that there exists no constant a such that $P(b_1 = ae_1) = 1$. Then $|X_n| \xrightarrow{p} \infty$ as $n \to \infty$.

Proof
If $P(\theta_0 + c(\psi_0 + b_1) + e_1 = c) = 1$ for some c, then $\theta_0 Ee_1 + c\psi_0 Ee_1 + cEb_1e_1 + Ee_1^2 = cEe_1$ and $\theta_0 Eb_1 + Ee_1b_1 + c\psi_0 Eb_1 + cEb_1^2 = cEb_1$. Since $Eb_1 = 0$, $Ee_1 = 0$, $Eb_1^2 > 0$, and $Ee_1^2 > 0$, we have $(Eb_1e_1)^2 = Eb_1^2 Ee_2^2$; hence, $|\mathrm{corr}(b_1, e_1)| = 1$, which contradicts with the fact that there exists no constant a such that $P(b_1 = ae_1) = 1$. Hence, we conclude $P(\theta_0 + e_1 + c(\psi_0 + b_1) = c) < 1$ for any constant c. Since $E\log|\psi_0 + b_1| \geq 0$ implies $P(\psi_0 + b_1 = 0) = 0$, the lemma now follows from Remark 2.8 and Corollary 4.1 of Goldie and Maller (2000). □

Proof of Theorem 2.1
Put

$$Z_j := \frac{(X_j - \psi_0 X_{j-1})X_{j-1}}{1 + X_{j-1}^2} = \frac{b_j X_{j-1}^2}{1 + X_{j-1}^2} + \frac{e_j X_{j-1}}{1 + X_{j-1}^2}.$$

We shall only prove the theorem for the case when $E\log|\psi_0 + b_1| \geq 0$ and $w_0^2 > 0$ since other cases can be shown easily.

Using Lemma A.1, the fact that $(b_j, e_j)^{\mathrm{T}}$ is independent of $\{X_i : i < j\}$, and $b_j X_{j-1}^2 / \left(1 + X_{j-1}^2\right)$ is therefore L_{2+r} bounded for some $r > 2$ and a martingale difference with respect to $\sigma(X_j, X_{j-1}, ...)$, we can easily show that (Berkes *et al.* 2009: Section 4, Chan *et al.* 2012: Appendix)

$$\frac{1}{\sqrt{n}}\sum_{j=1}^n Z_j = \frac{1}{\sqrt{n}}\sum_{j=1}^n \frac{b_j X_{j-1}^2}{1 + X_{j-1}^2} + o_p(1) = \frac{1}{\sqrt{n}}\sum_{j=1}^n b_j + o_p(1) \xrightarrow{d} N\left(0, w_0^2\right), \tag{A.1}$$

$$\frac{1}{n}\sum_{j=1}^n Z_j^2 = \frac{1}{n}\sum_{j=1}^n b_j^2 + o_p(1) \xrightarrow{p} w_0^2, \tag{A.2}$$

and

$$\max_{1\leq j\leq n}|Z_j| \leq \max_{1\leq j\leq n}|b_j| + \max_{1\leq j\leq n}|e_j| = o_p(n^{\frac{1}{2}}). \tag{A.3}$$

It follows from (7) that

$$\frac{1}{n}\sum_{j=1}^n Z_j = \frac{1}{n}\sum_{j=1}^n \frac{\lambda Z_j^2}{1 + \lambda Z_j},$$

DOI: 10.1111/jtsa.12064

which implies

$$\left|\frac{1}{n}\sum_{j=1}^{n} Z_j\right| \geq |\lambda|\frac{1}{n}\sum_{j=1}^{n} Z_j^2 \left\{1+|\lambda| \max_{1\leq j\leq n} |Z_j|\right\}^{-1}.$$

Hence, by (A.1)–(A.3), we have

$$|\lambda| = O_p\left(n^{-\frac{1}{2}}\right) \quad \text{and} \quad \lambda = \frac{\sum_{j=1}^{n} Z_j}{\sum_{j=1}^{n} Z_j^2} + o_p\left(n^{-\frac{1}{2}}\right). \tag{A.4}$$

It follows from (A.1), (A.2), (A.4), and the Taylor expansion that

$$l_n(\psi_0) = 2\sum_{j=1}^{n} \lambda Z_j - \sum_{j=1}^{n} \lambda^2 Z_j^2 + o_p(1) = \frac{\left(\sum_{j=1}^{n} Z_j\right)^2}{\sum_{j=1}^{n} Z_j^2} + o_p(1) \xrightarrow{d} \chi^2(1)$$

as $n \to \infty$. □

Proof of Theorem 2.2

The stationary AR case is easily proven, so consider the non-stationary RCA case $E \log|\psi_0 + b_1| \geq 0$ and $w_0^2 > 0$. We first prove $1/n \sum_{j=1}^{n} X_{j-1}^2/(\delta + X_{j-1}^2) \xrightarrow{p} 1$. We have

$$E\left|\frac{1}{n}\sum_{j=1}^{n}\left\{\frac{X_{j-1}^2}{\delta + X_{j-1}^2} - 1\right\}\right| \leq \delta\frac{1}{n}\sum_{j=1}^{n} E\left|\frac{1}{1+X_{j-1}^2}\right|. \tag{A5}$$

In view of Lemma A.1, $1/(\delta + X_{j-1}^2) \xrightarrow{p} 0$ as $j \to \infty$. By dominated convergence, therefore, $E\left[1/\left(1+X_{j-1}^2\right)\right] \to 0$ as $j \to \infty$; hence, the Cesàro sum $1/n \sum_{j=1}^{n} E\left[1/\left(1+X_{j-1}^2\right)\right] \to 0$. This proves $1/n \sum_{j=1}^{n} X_{j-1}^2/\left(1+X_{j-1}^2\right) \xrightarrow{p} 1$ by (A5) and Markov's inequality. Therefore, by (A.1), it follows

$$n^{\frac{1}{2}}\left(\hat{\psi}_{MEL} - \psi_0\right) = \frac{n^{-\frac{1}{2}}\sum_{j=1}^{n} Z_j}{\frac{\frac{1}{n}\sum_{j=1}^{n} X_{j-1}^2}{\delta + X_{j-1}^2}} = \frac{1}{n^{\frac{1}{2}}}\sum_{j=1}^{n} Z_j\,(1+o_p(1)) \xrightarrow{d} N\left(0, w_0^2\right).$$

□

Proof of Theorem 2.3

When $w_0^2 = 0$ and $X_n \xrightarrow{p} \infty$, (8) holds with probability tending to 1. For other cases, (8) holds with probability tending to 0. Hence, the theorem follows from Theorem 2.1 and the result for non-stationary AR(1) processes in Chan *et al.* (2012). □

DOI: 10.1111/jtsa.12064

Proof of Theorem 2.4
When $w_0^2 = 0$ and $E \log|\psi_0 + b_1| < 0$, (8) holds with probability tending to 1, which implies that $\hat{\psi}_{MEL*} = \hat{\psi}_{CLP}$ with probability tending to 1. Hence, the result follows from that in Chan *et al.* (2012). For the rest of the cases, $\hat{\psi}_{MEL*} = \hat{\psi}_{MEL}$ with probability tending to 1. Hence, the desired result follows from Theorem 2.2. □

Proof of Theorem 2.5
First, it is easy to show that $|\tilde{\theta} - \theta_0| + |X_n||\tilde{\psi} - \psi_0| = o_p(1)$ when either $w_0 = 0$ or $\{X_t\}$ is stationary and $|\tilde{\theta} - \theta_0| + |X_n||\tilde{\psi} - \psi_0| = O_p(1)$ when $w_0^2 > 0$ and $|X_n| \xrightarrow{p} \infty$. Hence, $\tilde{\sigma}_1^2 \xrightarrow{p} \sigma_0^2$ if $w_0 = 0$, $\tilde{\sigma}_2^2 \xrightarrow{p} w_0^2$ if $w_0^2 > 0$ and $|X_n| \xrightarrow{p} \infty$, and $\tilde{\sigma}_2^2 \xrightarrow{p} E[(e_2 + b_2 X_1)^2/(1 + X_1^2)]$ if $w_0^2 > 0$ and $\{X_t\}$ is stationary. It is also straightforward to show that $P(\tilde{\Delta}_n = 1) \to 1$ when $w_0 = 0$ and $|X_n| \xrightarrow{p} \infty$, and $P(\tilde{\Delta}_n = 1) \to 0$ for other cases. Further, by working with $r^{\mathrm{T}} Z_{n,j}(\theta_0, \psi_0)$ for conformable r, $r^{\mathrm{T}} r = 1$, and using the martingale difference central limit theorem arguments from Berkes *et al.* (2009) and Chan *et al.* (2012), and the Cramér-Wold theorem, it follows that $1/\sqrt{n} \sum_{j=1}^n Z_{n,j}(\theta_0, \psi_0)$ has a trivariate normal distribution limit with covariance matrix depending on the case whether $w_0^2 = 0$ or $w_0^2 > 0$ and $\{X_t\}$ is stationary or $w_0^2 > 0$ and $|X_n| \xrightarrow{p} \infty$. Using these facts, the rest follows from the standard approach in Qin and Lawless (1994). □

ACKNOWLEDGEMENTS

We thank two reviewers for their helpful comments, which improved the results and presentation of the article. Peng's research was supported by the NSF grant NSF DMS-1005336. Peng appreciates support from the NCTS Mathematics Division (Taipei Office) during his visit of Professor Ming-Yen Cheng at National Taiwan University in September 2013.

REFERENCES

Aue A, Horváth L. 2011. Quasi-likelihood estimation in stationary and nonstationary autoregressive models with random coefficients. *Statistica Sinica* **21**: 973–999.

Aue A, Horváth L, Steinebach J. 2006. Estimation in random coefficient autoregressive models. *Journal of Time Series Analysis* **27**: 61–76.

Berkes I, Horváth L, Ling S. 2009. Estimation in nonstationary random coefficient autoregressive models. *Journal of Time Series Analysis* **30**: 395–416.

Chan NH, Li D, Peng L. 2012. Toward a unified interval estimation of autoregressions. *Econometric Theory* **28**: 705–717.

Christiano LJ, Eichenbaum M. 1990. Unit roots in real GNP: do we know, and do we care? *Carnegie-Rochester Conference Series on Public Policy,* Elsevier, **32**(1): 7–61.

Distaso W. 2008. Testing for unit root processes in random coefficient autoregressive models. *Journal of Econometrics* **142**: 581–609.

Gaffeo E, Gallegati M, Gallegati M. 2005. Requiem for the unit root in per capita real GDP? Additional evidence from historical data. *Empirical Economics* **30**: 37–63.

Giraitis L, Kapetanios G, Yates T 2010.,'Inference on stochastic time-varying coefficient models,' Working Paper, Queen Mary, University of London.

Goldie CM, Maller RA. 2000. Stability of perpetuities. *Annals of Probability* **28**: 1195–1218.

Granger CWJ, Swanson NR. 1997. An introduction to stochastic unit-root processes. *Journal of Econometrics* **80**: 35–62.

Hwang SY, Basawa IV. 2005. Explosive random-coefficient AR(1) processes and related asymptotics for least-squares estimation. *Journal of Time Series Analysis* **26**: 807–824.

Hwang SY, Basawa IV, Kim TY. 2006. Least squares estimation for critical random coefficient first-order autoregressive processes. *Statistics & Probability Letters* **76**: 310–317.

McCabe BPM, Tremaynee AR. 1995. Testing a time series for difference stationarity. *Annals of Statistics* **23**: 1015–1028.

DOI: 10.1111/jtsa.12064

Nagakura D. 2009. Asymptotic theory for explosive random coefficient autoregressive models and inconsistency of a unit root test against a stochastic unit root process. *Statistics & Probability Letters* **79**: 2476–2483.
Nelson CR, Plosser C. 1982. Trends and random walks in macroeconomic time series: some evidence and implications. *Journal of Monetary Economics* **10**: 139–162.
Owen A. 1990. Empirical likelihood confidence regions. *Annals of Statistics* **18**: 90–120.
Owen A. 2001. *Empirical Likelihood*. Boca Raton, Florida, USA: Chapman & Hall.
Qin J, Lawless J. 1994. Empirical likelihood and general estimating equations. *Ann. Statist.* **22**: 300–325.
Quinn BG. 1982. A note on the existence of strictly stationary solutions to bilinear equations. *Journal of Time Series Analysis* **3**: 249–252.
Rudebusch GD. 1993. The uncertain unit root in real GNP. *American Economic Review* **83**: 264–272.

DOI: 10.1111/jtsa.12064