UNIFORM INTERVAL ESTIMATION FOR AN AR(1) PROCESS WITH AR ERRORS

Jonathan Hill, Deyuan Li and Liang Peng

University of North Carolina, Fudan University and Georgia State University

Abstract: An empirical likelihood method was proposed in Hill and Peng (2014) to construct a unified interval estimation for the coefficient in an AR(1) model, regardless of whether the sequence was stationary or near integrated. The error term, however, was assumed independent, and this method fails when the errors are dependent. Testing for a unit root in an AR(1) model has been studied in the literature for dependent errors, but existing methods cannot be used to test for a near unit root. In this paper, assuming the errors are governed by an AR(p) process, we exploit the efficient empirical likelihood method to give a unified interval for the coefficient by taking the structure of errors into account. Furthermore, a jackknife empirical likelihood method is proposed to reduce the computation of the empirical likelihood method when the order in the AR errors is not small. A simulation study is conducted to examine the finite sample behavior of the proposed methods.

Key words and phrases: AR model, empirical likelihood, jackknife empirical likelihood method, weighted score.

1. Introduction

Consider the autoregressive time series model

$$X_t = \beta_0 X_{t-1} + \epsilon_t \quad \text{for} \quad t = 1, \dots, n, \tag{1.1}$$

with n the sample size, and the $\epsilon_t's$ identically distributed with zero mean and finite variance. A commonly estimator for the true parameter value β_0 is the least squares estimator $\hat{\beta}_{LS} = \sum_{t=1}^n X_t X_{t-1} / \sum_{t=1}^n X_{t-1}^2$ which minimizes $\sum_{t=1}^n \{X_t - \beta X_{t-1}\}^2$. As is well known, the asymptotic distribution of $\hat{\beta}_{LS}$ is normal for the stationary case, $|\beta_0| < 1$, and non-normal for the nearly integrated case, $\beta_0 = 1 - \delta/n$ for some $\delta \in \mathbb{R}$, cf., Phillips (1987), making interval estimation quite challenging since inference by least squares is not unified.

In a more general environment, when $\{\epsilon_t\}$ may be a martingale difference sequence, Chan, Li, and Peng (2012) applied the empirical likelihood method to the weighted score equation

$$\sum_{i=1}^n \{X_t - \beta X_{t-1}\} \frac{X_{t-1}}{\sqrt{1 + X_{t-1}^2}} = 0, \tag{1.2}$$

so that a unified interval for β_0 can be obtained, regardless of whether $\{X_t\}$ is stationary or nearly integrated. Other weighted functions can be employed in order to achieve weighted score equations similar to those in Ling (2007) and Pan, Wang, and Yao (2007). The empirical likelihood method based on estimating equations was studied by Qin and Lawless (1994) amongst many others, and an overview of the method can be found in Owen (2001). References for more computationally-involved methods for constructing unified intervals for β_0 can be found in Chan, Li, and Peng (2012).

In practice, assuming independent, or even martingale difference, errors is restrictive since this requires (1.1) to be the true data generating process. In many cases, however, we simply want to know if X_t is difference stationary or not, in which case naturally ϵ_t may be dependent (Phillips (1987), Phillips and Perron (1988)). Some recent references on processes with dependent errors include Wu (2005), Wu and Min (2005), and Chen, Min, and Chen (2013). By assuming that $\{\epsilon_t\}$ is an α-mixing sequence, Phillips and Perron (1988) proposed statistics for testing a unit root under model (1.1) with an intercept. However, these statistics cannot be employed to construct a confidence interval for β_0 in the case of a near unit root. Therefore, it remains unknown how to construct a unified interval for β_0 under the assumption that (1.1) holds with possibly dependent ϵ_t. The procedure in Chan, Li, and Peng (2012) does not work since it does not account for general weak dependence, but the blockwise empirical likelihood method has the ability to deal with mixing data (see Kitamura (1997)). One can attempt to apply the blockwise empirical likelihood method to some weighted score equations for obtaining a unified interval. But there are serious issues in doing so.

We investigate ways to construct a unified interval for β_0 when $\{\epsilon_t\}$ in (1.1) is governed by an AR(p) process. The least squares estimator $\hat{\beta}_{LS}$ is clearly inefficient since it does not take the structure of errors into account, and the empirical likelihood method based on (1.2) does not lead to a chi-squared limit due to the dependent errors. We propose a unified empirical likelihood method based on some different weighted score equations that take the structure of $\{\epsilon_t\}$ into account. We then propose a jackknife empirical likelihood method to reduce the computation of the proposed empirical likelihood method when the order in the errors is not small. Our unified intervals can be immediately applied to a unit root test with a general alternative.

To keep arguments short, we assume the AR(p) error has an iid term $\{e_t\}$. The theory developed here can be extended to the case where e_t is a martingale difference, covering conditional volatility of unknown form. A simulation study in Section 3 confirms this argument.

We organize the remainder of the paper as follows. Section 2 presents the methodologies for an AR(1) model without or with a time trend, when the errors follow an AR process. Simulation results are presented in Section 3. Proofs are in Section 4.

2. Methodologies

2.1. Profile empirical likelihood method

Consider model (1.1) with $\{\epsilon_t\}$ following from an AR(p) process:

$$e_t = \epsilon_t + \sum_{j=1}^{p} \gamma_{0,j}\epsilon_{t-j} \quad \text{for} \quad t = 1, \ldots, n, \tag{2.1}$$

where e_t is iid with zero mean and finite variance $\sigma^2 > 0$, and $\gamma_0 = (\gamma_{0,1}, \ldots, \gamma_{0,p})^T$ is such that $\{\epsilon_t\}$ is a stationary process.

One can efficiently estimate β_0 and γ_0 simultaneously by minimizing

$$\sum_{t=1}^{n} \left\{ X_t - \beta X_{t-1} + \sum_{j=1}^{p} \gamma_j \left(X_{t-j} - \beta X_{t-j-1} \right) \right\}^2,$$

hence the proposed estimators of $(\beta_0, \gamma_0)^T$ solve the score equations

$$\sum_{t=1}^{n} \left\{ X_t - \beta X_{t-1} + \sum_{k=1}^{p} \gamma_k(X_{t-k} - \beta X_{t-k-1}) \right\} (X_{t-j} - \beta X_{t-j-1}) = 0$$
$$\text{for } j = 1, \ldots, p, \tag{2.2}$$

$$\sum_{t=1}^{n} \left\{ X_t - \beta X_{t-1} + \sum_{j=1}^{p} \gamma_j(X_{t-j} - \beta X_{t-j-1}) \right\} \left\{ X_{t-1} + \sum_{j=1}^{p} \gamma_j X_{t-j-1} \right\} = 0. \tag{2.3}$$

Intuitively this new estimator for β_0 should be more efficient than the least squares estimator $\hat{\beta}_{LS} = \sum_{t=1}^{n} X_t X_{t-1} / \sum_{t=1}^{n} X_{t-1}^2$ since it takes the error structure into account.

In order to construct a confidence interval for β_0 without estimating the asymptotic variance, one can apply the profile empirical likelihood method to (2.2) and (2.3). However, when $\{X_t\}$ is nearly integrated, Wilks theorem fails. As in Chan, Li, and Peng (2012), we propose to apply the profile empirical likelihood method to (2.2), and to a weighted version of (2.3),

$$\sum_{t=1}^{n} \left\{ X_t - \beta X_{t-1} + \sum_{j=1}^{p} \gamma_j(X_{t-j} - \beta X_{t-j-1}) \right\} \left\{ \frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}} + \sum_{j=1}^{p} \gamma_j \frac{X_{t-j-1}}{\sqrt{1+X_{t-j-1}^2}} \right\} = 0. \tag{2.4}$$

In particular, write $Z_t(\beta, \gamma) \equiv (Z_{t,1}(\beta, \gamma), \ldots, Z_{t,p+1}(\beta, \gamma))^T \in \mathbb{R}^{p+1}$, where

$$Z_{t,j}(\beta, \gamma) = \left\{ X_t - \beta X_{t-1} + \sum_{k=1}^{p} \gamma_k(X_{t-k} - \beta X_{t-k-1}) \right\} (X_{t-j} - \beta X_{t-j-1})$$
$$\text{for } j = 1, \ldots, p,$$

$$Z_{t,p+1}(\beta, \gamma) = \left\{ X_t - \beta X_{t-1} + \sum_{k=1}^{p} \gamma_k (X_{t-k} - \beta X_{t-k-1}) \right\}$$

$$\left\{ \frac{X_{t-1}}{\sqrt{1 + X_{t-1}^2}} + \sum_{k=1}^{p} \gamma_k \frac{X_{t-k-1}}{\sqrt{1 + X_{t-k-1}^2}} \right\},$$

and define the empirical likelihood function for (β_0, γ_0) as

$$L(\beta, \gamma) = \sup \left\{ \prod_{t=1}^{n} (np_t) : p_1, \ldots, p_n \geq 0, \ \sum_{t=1}^{n} p_t = 1, \ \sum_{t=1}^{n} p_t Z_t(\beta, \gamma) = 0 \right\}.$$

Since we are only interested in β_0, we consider the profile empirical likelihood function for β_0:

$$L^P(\beta) = \max_{\gamma \in \mathbb{R}^p} L(\beta, \gamma).$$

Theorem 1. *Suppose* (1.1) *holds with* (2.1), *and* $E|e_t|^{2+d} < \infty$ *for some* $d > 0$. *Then* $-2 \log L^P(\beta_0) \overset{d}{\to} \chi^2(1)$ *as* $n \to \infty$ *whenever* $|\beta_0| < 1$, *independent of* n, *or* $\beta_0 = 1 - \delta/n$ *for some* $\delta \in \mathbb{R}$.

Consider the same model with a linear time trend

$$X_t = \alpha_{0,1} + \alpha_{0,2} t + \beta_0 X_{t-1} + \epsilon_t \text{ for } t = 1, \ldots, n, \tag{2.5}$$

where ϵ_t follows the AR(p) process in (2.1). In this case, take $\bar{Z}_t(\beta, \alpha_1, \alpha_2, \gamma) \equiv [\bar{Z}_{t,1}(\beta, \alpha_1, \alpha_2, \gamma), \ldots, \bar{Z}_{t,p+3}(\beta, \alpha_1, \alpha_2, \gamma)]^T \in \mathbb{R}^{p+3}$, where

$$\bar{Z}_{t,j}(\beta, \alpha_1, \alpha_2, \gamma)$$
$$= \left\{ X_t - \alpha_1 - \alpha_2 t - \beta X_{t-1} + \sum_{k=1}^{p} \gamma_k (X_{t-k} - \alpha_1 - \alpha_2(t - k) - \beta X_{t-k-1}) \right\}$$
$$\times (X_{t-j} - \alpha_1 - \alpha_2(t - j) - \beta X_{t-j-1})$$

for $j = 1, \ldots, p$, and

$$\bar{Z}_{t,p+1}(\beta, \alpha_1, \alpha_2, \gamma)$$
$$= X_t - \alpha_1 - \alpha_2 t - \beta X_{t-1} + \sum_{k=1}^{p} \gamma_k (X_{t-k} - \alpha_1 - \alpha_2(t - k) - \beta X_{t-k-1}),$$

$$\bar{Z}_{t,p+2}(\beta, \alpha_1, \alpha_2, \gamma)$$
$$= \left\{ X_t - \alpha_1 - \alpha_2 t - \beta X_{t-1} + \sum_{k=1}^{p} \gamma_k (X_{t-k} - \alpha_1 - \alpha_2(t - k) - \beta X_{t-k-1}) \right\}$$
$$\times \left\{ t + \sum_{k=1}^{p} \gamma_k (t - k) \right\},$$

$$\bar{Z}_{t,p+3}(\beta, \alpha_1, \alpha_2, \gamma)$$

$$= \left\{ X_t - \alpha_1 - \alpha_2 t - \beta X_{t-1} + \sum_{k=1}^{p} \gamma_k \left(X_{t-k} - \alpha_1 - \alpha_2(t-k) - \beta X_{t-k-1} \right) \right\}$$

$$\times \left\{ \frac{X_{t-1}}{(1+X_{t-1}^2)^q} + \sum_{j=1}^{p} \gamma_j \frac{X_{t-j-1}}{(1+X_{t-j-1}^2)^q} \right\} + W_t \text{ for some } q > 0.$$

The $W_t's$ are simulated iid random variables, distributed as $N(0, \bar{\sigma}^2)$, and $\bar{\sigma}^2 > 0$ is chosen to be smaller than $E(e_t^2)$. Here $E(e_t^2)$ can be estimated by first estimating the errors $\epsilon_t's$ based on the least squares estimate for $\alpha_1, \alpha_2, \beta$, and then using these estimated errors to obtain the least squares estimate for γ, and finally estimating $e_t's$ by these least squares estimates. In order to avoid the effect of a random seed in generating $W_t's$, we use $W_t = 1/\sqrt{1,000} \sum_{i=1}^{1,000} W_{t,i}$ in our simulation study, where the $W_{t,i}'s$ are iid $N(0, \bar{\sigma}^2)$ draws for $t = 1, \ldots, n$ and $i = 1, \ldots, 1,000$.

When $\bar{\sigma} = 0$, $q = 1/2$, and $|X_t| \xrightarrow{p} \infty$, the joint limit of $1/\sqrt{n} \sum_{t=1}^{n} \bar{Z}_{t,p+3}(\beta, \alpha_1, \alpha_2, \gamma)$ and $1/\sqrt{n} \sum_{t=1}^{n} \bar{Z}_{t,p+1}(\beta, \alpha_1, \alpha_2, \gamma)$ is no longer normal, which makes the application of the empirical likelihood method fail. This is why we need to add the pseudo sample W_t to achieve uniform inference. Based on arguments in Li, Chan, and Peng (2014, Sec. 2), and Hill and Peng (2014, p.288), in the nonstationary case a choice of $q > 1/2$ makes $\sum_{t=1}^{n} \bar{Z}_{t,p+3}(\beta_0, \alpha_{0,1}, \alpha_{0,2}, \gamma_0)$ asymptotically equivalent to $\sum_{t=1}^{n} W_t$, while small $q \leq 1$ allows $\sum_{t=1}^{n} \bar{Z}_{t,p+3}(\beta, \alpha_{0,1}, \alpha_{0,2}, \gamma_0)$ to better detect departures from β_0. We therefore enforce $q \in (1/2, 1]$ to balance power and size, and in practice simply use $q = 0.75$.

The empirical likelihood function for $(\beta, \alpha_1, \alpha_2, \gamma)$ based on $\{\bar{Z}_t(\beta, \alpha_1, \alpha_2, \gamma)\}_{t=1}^{n}$ is defined as

$$\bar{L}(\beta, \alpha_1, \alpha_2, \gamma)$$

$$= \sup \left\{ \prod_{t=1}^{n}(np_t) : p_1, \ldots, p_n \geq 0, \ \sum_{t=1}^{n} p_t = 1, \ \sum_{t=1}^{n} p_t \bar{Z}_t(\beta, \alpha_1, \alpha_2, \gamma) = 0 \right\}$$

and, as before, we only consider the profile empirical likelihood function

$$\bar{L}^P(\beta) = \max_{(\alpha_1, \alpha_2, \gamma^T)^T \in \mathbb{R}^{p+2}} \bar{L}(\beta, \alpha_1, \alpha_2, \gamma).$$

Theorem 2. *Suppose* (2.5) *holds with* $E|e_t|^{2+d} < \infty$ *for some constant* $d > 0$. *Then* $-2 \log \bar{L}^P(\beta_0) \xrightarrow{d} \chi^2(1)$ *as* $n \to \infty$ *whenever* $|\beta_0| < 1$, *independent of* n, *or* $\beta_0 = 1 - \delta/n$ *for some* $\delta \in \mathbb{R}$.

Remark 1. When we consider (1.1) with a constant trend, $\alpha_{0,2} = 0$ in (2.5) is known, Theorem 2 still holds if the term $\bar{Z}_{t,p+2}$ is removed and α_2 is replaced by zero.

2.2. Jackknife empirical likelihood method

The profile empirical likelihood methods become computationally intensive when p is large. In order to reduce computation time, one can estimate γ_0 first by solving (2.2), which results in an explicit function of β, and then apply the empirical likelihood method to (2.4) with γ replaced by this estimator. However, this does not lead to a chi-squared limit due to the plug-in estimator, but rather a weighted sum of independent chi-squared variables, see Chen and Van Keilegom (2009). Recently a jackknife empirical likelihood method was proposed by Jing, Yuan, and Zhou (2009) to deal with non-linear functionals, and Li, Peng, and Qi (2011) employed this idea to reduce the computation of the empirical likelihood method based on estimating equations. Here, we employ the jackknife empirical likelihood method.

Consider model (1.1) with the error ϵ_t following the autoregressive process (2.1). Let $\hat\gamma(\beta) = (\hat\gamma_1(\beta), \ldots, \hat\gamma_p(\beta))^T$ be, for arbitrary β, the solution to (2.2), and for each $i = 1, \ldots, n$ let $\hat\gamma^{(i)}(\beta) = (\hat\gamma_1^{(i)}(\beta), \ldots, \hat\gamma_p^{(i)}(\beta))^T$ be the solution to

$$\sum_{t=1,t\neq i}^{n} \left\{ X_t - \beta X_{t-1} + \sum_{k=1}^{p} \gamma_k(X_{t-k} - \beta X_{t-k-1}) \right\}(X_{t-j} - \beta X_{t-j-1}) = 0$$

$$\text{for} \quad j = 1, \ldots, p. \qquad (2.6)$$

Define the pseudo sample as

$$Z_{n,j}^*(\beta) = \sum_{t=1}^{n} \left\{ X_t - \beta X_{t-1} + \sum_{k=1}^{p} \hat\gamma_k(\beta)(X_{t-k} - \beta X_{t-k-1}) \right\}$$

$$\times \left\{ \frac{X_{t-1}}{\sqrt{1 + X_{t-1}^2}} + \sum_{k=1}^{p} \hat\gamma_k(\beta) \frac{X_{t-k-1}}{\sqrt{1 + X_{t-k-1}^2}} \right\}$$

$$- \sum_{t=1,t\neq j}^{n} \left\{ X_t - \beta X_{t-1} + \sum_{k=1}^{p} \hat\gamma_k^{(j)}(\beta)(X_{t-k} - \beta X_{t-k-1}) \right\}$$

$$\times \left\{ \frac{X_{t-1}}{\sqrt{1 + X_{t-1}^2}} + \sum_{k=1}^{p} \hat\gamma_k^{(j)}(\beta) \frac{X_{t-k-1}}{\sqrt{1 + X_{t-k-1}^2}} \right\}$$

for $j = 1, \ldots, n$. Based on this pseudo sample, the jackknife empirical likelihood function for β is

$$L^*(\beta) = \sup \left\{ \prod_{i=1}^{n}(np_i) : p_1 \geq 0, \ldots, p_n \geq 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i Z_{n,i}^*(\beta) = 0 \right\}.$$

Theorem 3. *Suppose (1.1) holds with (2.1) and $E|e_t|^{4+d} < \infty$ for some $d > 0$. Then $-2\log L^*(\beta_0) \xrightarrow{d} \chi^2(1)$ as $n \to \infty$ whenever $|\beta_0| < 1$, independent of n, or $\beta_0 = 1 - \delta/n$ for some $\delta \in \mathbb{R}$.*

Remark 2. The jackknife empirical likelihood methods in Theorem 3 require a higher moment for e_t than the profile empirical likelihood methods in Theorems 1 and 2. The reason is that we need to derive an accurate approximate distance between $\hat{\gamma}(\beta_0)$ and $\hat{\gamma}^{(i)}(\beta_0)$, and this ultimately requires a central limit theory for cross products of the $e_t's$.

Now consider this model with a linear time trend, model (2.5). Define the parameter subset $\theta = [\alpha_1, \alpha_2, \beta]^T$. Let $\bar{\gamma}(\theta) = (\bar{\gamma}_1(\theta), \ldots, \bar{\gamma}_p(\theta))^T$ denote, for arbitrary θ, the solution to

$$\sum_{t=1}^{n} \bar{Z}_{t,j}(\theta, \gamma) = 0 \quad \text{for} \quad j = 1, \ldots, p, \tag{2.7}$$

and let $\bar{\gamma}^{(i)}(\theta) = (\bar{\gamma}_1^{(i)}(\theta), \ldots, \bar{\gamma}_p^{(i)}(\theta))^T$ for $i = 1, \ldots, n$ denote the solution to

$$\sum_{t=1,t\neq i}^{n} \bar{Z}_{t,j}(\theta, \gamma) = 0 \quad \text{for} \quad j = 1, \ldots, p. \tag{2.8}$$

Next, define the pseudo sample as

$$\bar{Z}_{n,i,k}^*(\theta) = \sum_{t=1}^{n} \bar{Z}_{t,p+k}(\theta, \bar{\gamma}) - \sum_{t=1,t\neq i}^{n} \bar{Z}_{t,p+k}(\theta, \bar{\gamma}^{(i)}),$$

where $i = 1, \ldots, n$ and $k = 1, 2, 3$. Take $\bar{Z}_{n,i}^*(\theta) = (\bar{Z}_{n,i,1}^*(\theta), \bar{Z}_{n,i,2}^*(\theta), \bar{Z}_{n,i,3}^*(\theta))$, and define the jackknife empirical likelihood function for $\theta = (\alpha_1, \alpha_2, \beta)^T$ as

$$\bar{L}^*(\beta, \alpha_1, \alpha_2) = \bar{L}^*(\theta)$$
$$= \sup\left\{\prod_{i=1}^{n}(np_i) : p_1 \geq 0, \ldots, p_n \geq 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i \bar{Z}_{n,i}^*(\theta) = 0\right\}.$$

Since we are only interested in β, we consider the profile jackknife empirical likelihood function

$$\bar{L}^{*P}(\beta) = \max_{(\alpha_1, \alpha_2)^T \in \mathbb{R}^2} \bar{L}^*(\beta, \alpha_1, \alpha_2).$$

Theorem 4. *Suppose* (2.5) *holds with* $E|e_t|^{4+d} < \infty$ *for some constant* $d > 0$. *Then* $-2\log \bar{L}^{*P}(\beta_0) \xrightarrow{d} \chi^2(1)$ *as* $n \to \infty$ *whenever* $|\beta_0| < 1$, *independent of* n, *or* $\beta_0 = 1 - \delta/n$ *for some* $\delta \in \mathbb{R}$.

Remark 3. In Theorem 4 we do not estimate $\alpha_1, \alpha_2, \gamma$ simultaneously in the first step since this doesn't lead to explicit formulas. If one wants to estimate them simultaneously, then the approximate jackknife empirical likelihood method in Peng (2012) can be employed, where explicit formulas for estimating nuisance parameters are not required.

Remark 4. When the process is explosive, $|\beta_0| > 1$ and independent of n, the proposed methods are still valid as in Chan, Li, and Peng (2012). The near unit root case includes the slightly explosive case when $\delta < 0$.

Remark 5. When $\epsilon_t = \sum_{j=0}^{\infty} c_j e_{t-j}$, where $\{e_t\}$ is a sequence of independent and identically distributed random variables, similar jackknife empirical likelihood methods may be developed by noting $C^{-1}(B)X_t = \beta_0 C^{-1}(B)X_{t-1} + e_t$ with $C(B) = \sum_{j=0}^{\infty} c_j B^j$, B denoting the usual backward operator, and using the truncation techniques in Xiao et al. (2003). Detailed investigation is a future project.

3. Simulation Study

In this section we examine the finite sample behavior of the proposed methods for models (1.1), (2.5), and (2.5) with known $\alpha_{0,2} = 0$.

We drew 10,000 random samples with size $n = 50$ and 200 from either model (1.1) or (2.5), with $\beta_0 \in \{0.9, 0.99, 1\}$, the error AR order is $p = 2$ or 5 with $\gamma_0 = (\binom{p}{1}0.1, \ldots, \binom{p}{p}0.1^p)$, and $\{e_t\}$ is a sequence of independent standard normals. Under (2.5) we use $\alpha_{0,1} = 0.5$ and $\alpha_{0,2} \in \{0, 0.2\}$. The added pseudo sample $\{W_t\}_{t=1}^n$ is computed using $W_t = 1/\sqrt{1,000} \sum_{i=1}^{1,000} W_{t,i}$, where $W'_{t,i}s$ are iid $N(0, .5)$. We employ the R package 'emplik' to compute the empirical likelihood function and then use the R package 'nlm' to calculate the profile empirical likelihood function. Initial values are computed by least squares estimates. For example, in order to compute coverage probabilities for (2.5), we minimize $\sum_{i=1}^n \{X_t - \alpha_1 - \alpha_2 t - \beta X_{t-1}\}^2$ with respect to α_1 and α_2 to obtain functions $\hat{\alpha}_i(\beta)$, and then minimize

$$\sum_{i=1}^n \Big\{ X_i - \hat{\alpha}_1(\beta) - \hat{\alpha}_2(\beta) t - \beta X_{t-1}$$
$$+ \sum_{k=1}^p \gamma_k (X_{t-k} - \hat{\alpha}_1(\beta) - \hat{\alpha}_2(\beta)(t-k) - \beta X_{t-k-1}) \Big\}^2$$

with respect to $\gamma = (\gamma_1, \ldots, \gamma_p)^T$ to achieve $\hat{\gamma}(\beta)$. Coverage probabilities are computed using simply $\hat{\alpha}_i(\beta_0)$ and $\hat{\gamma}(\beta_0)$.

We consider (1.1) with $e_t = \sqrt{0.1 + 0.5e_{t-1}^2}\,\xi_t$, where ξ_t is a sequence of independent standard normals.

Coverage probabilities are reported in Tables 1−4. We observe from these tables that (i) when the model has no trend, the jackknife empirical likelihood method performs better than the empirical likelihood method for $n = 50$, but they are comparable for $n = 200$; (ii) when the model has a linear trend, the jackknife empirical likelihood method is worse than the empirical likelihood method

Table 1. Model (1): no trend. Coverage probabilities based on Theorem 1 (ELM1 for level 0.9 and ELM2 for level 0.95) and Theorem 3 (JELM1 for level 0.9 and JELM2 for level 0.95) are reported for model (1.1).

(β_0, n)	ELM1 $p = 2$	JELM1 $p = 2$	ELM2 $p = 2$	JELM2 $p = 2$	ELM1 $p = 5$	JELM1 $p = 5$	ELM2 $p = 5$	JELM2 $p = 5$
$(0.9, 50)$	0.8900	0.9022	0.9398	0.9548	0.8887	0.9268	0.9427	0.9711
$(0.99, 50)$	0.8763	0.8889	0.9333	0.9453	0.8623	0.8962	0.9203	0.9545
$(1, 50)$	0.8703	0.8809	0.9300	0.9403	0.8511	0.8881	0.9193	0.9523
$(0.9, 200)$	0.8931	0.8959	0.9434	0.9482	0.8963	0.9097	0.9438	0.9571
$(0.99, 200)$	0.8909	0.8938	0.9445	0.9481	0.8884	0.8977	0.9401	0.9493
$(1, 200)$	0.8856	0.8898	0.9417	0.9464	0.8787	0.8896	0.9387	0.9490

Table 2. Model (7): a linear time trend. Coverage probabilities based on Theorem 2 (ELM1 for level 0.9 and ELM2 for level 0.95) and Theorem 4 (JELM1 for level 0.9 and JELM2 for level 0.95) are reported for model (2.5).

(β_0, n)	ELM1 $p = 2$	JELM1 $p = 2$	ELM2 $p = 2$	JELM2 $p = 2$	ELM1 $p = 5$	JELM1 $p = 5$	ELM2 $p = 5$	JELM2 $p = 5$
$(0.9, 50)$	0.8897	0.8903	0.9449	0.9455	0.8967	0.8154	0.9480	0.9012
$(0.99, 50)$	0.8887	0.8892	0.9447	0.9454	0.8962	0.8173	0.9479	0.9027
$(1, 50)$	0.8890	0.8896	0.9446	0.9453	0.8948	0.8170	0.9479	0.9017
$(0.9, 200)$	0.8961	0.8985	0.9469	0.9475	0.9037	0.8977	0.9492	0.9485
$(0.99, 200)$	0.8969	0.8992	0.9467	0.9473	0.9035	0.8981	0.9489	0.9485
$(1, 200)$	0.8968	0.8991	0.9464	0.9473	0.9038	0.8976	0.9496	0.9493

Table 3. Model (7): constant trend. Coverage probabilities based on Theorem 2 (ELM1 for level 0.9 and ELM2 for level 0.95) and Theorem 4 (JELM1 for level 0.9 and JELM2 for level 0.95) are reported for model (2.5) with $\alpha_{0,2} = 0$ known.

(β_0, n)	ELM1 $p = 2$	JELM1 $p = 2$	ELM2 $p = 2$	JELM2 $p = 2$	ELM1 $p = 5$	JELM1 $p = 5$	ELM2 $p = 5$	JELM2 $p = 5$
$(0.9, 50)$	0.8986	0.9005	0.9473	0.9491	0.8967	0.9027	0.9474	0.9509
$(0.99, 50)$	0.8977	0.8969	0.9492	0.9497	0.8959	0.8975	0.9470	0.9500
$(1, 50)$	0.8985	0.8979	0.9497	0.9504	0.8961	0.8971	0.9471	0.9492
$(0.9, 200)$	0.8954	0.8951	0.9482	0.9479	0.8961	0.8968	0.9484	0.9485
$(0.99, 200)$	0.8965	0.8963	0.9470	0.9472	0.8965	0.8967	0.9454	0.9451
$(1, 200)$	0.8966	0.8966	0.9467	0.9462	0.8991	0.8991	0.9468	0.9473

for $n = 50$ and $p = 5$, but they are comparable for $n = 200$; (iii) when the model has a constant trend, the methods are comparable; (iv) the methods work when $\{e_t\}$ is a martingale difference.

The proposed methods perform quite well. The jackknife empirical likelihood method is preferred when the model has no trend or a constant trend, or the

Table 4. Model (1): no trend, but heteroscedastic $\{e_t\}$. Coverage probabilities based on Theorem 1 (ELM1 for level 0.9 and ELM2 for level 0.95) and Theorem 3 (JELM1 for level 0.9 and JELM2 for level 0.95) are reported for model (1.1) when $e_t = \sqrt{0.1 + 0.5 e_{t-1}^2}\, \xi_t$, $\xi_t' s$ are iid with $N(0,1)$.

(β_0, n)	ELM1	JELM1	ELM2	JELM2
$(0.9, 50)$	0.8845	0.9063	0.9368	0.9573
$(0.99, 50)$	0.8625	0.8796	0.9256	0.9414
$(1, 50)$	0.8575	0.8741	0.9214	0.9377
$(0.9, 200)$	0.8872	0.8955	0.9423	0.9483
$(0.99, 200)$	0.8857	0.8921	0.9380	0.9456
$(1, 200)$	0.8778	0.8860	0.9380	0.9436

sample size is not too small, e.g., $n = 200$. When the model has a linear trend, the jackknife empirical likelihood is worse than the empirical likelihood method for a large order p and a small n, although its computation is less intensive.

4. Proofs

The proof of Theorem 1 exploits the following results.

Lemma 1. *Let (1.1) hold with (2.1) and $E|e_t|^{2+d} < \infty$ for some $d > 0$. Let either $|\beta_0| < 1$ independent of n, or $\beta_0 = 1 - \delta/n$ for some $\delta \in \mathbb{R}$. Then $1/\sqrt{n} \sum_{t=1}^{n} Z_t(\beta_0, \gamma_0) \xrightarrow{d} N(0, \Sigma)$ as $n \to \infty$, where $\Sigma = (\sigma_{i,j})_{1 \leq i,j \leq p+1}$ with $\sigma_{i,j} = E(e_t^2) E(\epsilon_{t-i} \epsilon_{t-j})$ for $i, j = 1, \ldots, p,$*

$$\sigma_{i,p+1} = \begin{cases} E(e_t^2) E\left\{ \epsilon_{t-i}\left(\frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}} + \sum_{j=1}^{p} \gamma_{0,j} \frac{X_{t-j-1}}{\sqrt{1+X_{t-j-1}^2}} \right)\right\} & \text{when } |\beta_0| < 1, \\ 0 & \text{when } \beta_0 = 1 - \frac{\delta}{n} \end{cases}$$

for $i = 1, \ldots, p,$ and

$$\sigma_{p+1,p+1} = \begin{cases} E(e_t^2) E\left\{ \frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}} + \sum_{j=1}^{p} \gamma_{0,j} \frac{X_{t-j-1}}{\sqrt{1+X_{t-j-1}^2}} \right\}^2 & \text{when } |\beta_0| < 1, \\ E(e_t^2) \times \left(1 + \sum_{j=1}^{p} \gamma_{0,j} \right)^2 & \text{when } \beta_0 = 1 - \frac{\delta}{n}. \end{cases}$$

Proof. Note that

$$\begin{cases} Z_{t,j}(\beta_0, \gamma_0) = e_t \epsilon_{t-j} \text{ for } j = 1, \ldots, p, \\ Z_{t,p+1}(\beta_0, \gamma_0) = e_t\left\{ \frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}} + \sum_{j=1}^{p} \gamma_{0,j} \frac{X_{t-j-1}}{\sqrt{1+X_{t-j-1}^2}} \right\}. \end{cases} \tag{4.1}$$

In the local to unity case $\beta_0 = 1 - \delta/n$, $|X_t| \overset{p}{\to} \infty$ as $t \to \infty$, hence

$$\left(\frac{X_{t-1}}{\sqrt{1 + X_{t-1}^2}} + \sum_{j=1}^{p} \gamma_{0,j} \frac{X_{t-j-1}}{\sqrt{1 + X_{t-j-1}^2}} \right)^2 \overset{p}{\to} \left(1 + \sum_{j=1}^{p} \gamma_{0,j} \right)^2 \text{ as } t \to \infty. \quad (4.2)$$

The result follows from (4.1), (4.2), and the central limit theorem for martingale differences. See McLeish (1974) and Hall and Heyde (1980), and see arguments in the appendices of Chan, Li, and Peng (2012) and Hill and Peng (2014).

Lemma 2. *Under the conditions of Lemma 1, $n^{-1} \sum_{t=1}^{n} Z_t^T(\beta_0, \gamma_0) Z_t(\beta_0, \gamma_0) \overset{p}{\to} \Sigma$ as $n \to \infty$.*

Proof. The claim follows from (4.1)−(4.2) and the weak law of large numbers for martingale differences (see Hall and Heyde (1980)).

Lemma 3. *Under the conditions of Lemma 1, as $n \to \infty$, with probability one $L(\beta_0, \gamma)$ attains its maximum or local maximum value at some point $\tilde{\gamma}$ in the interior of the ball $||\gamma - \gamma_0|| \le n^{-d/3}$, and $\tilde{\gamma}$ and $\tilde{\lambda}$ satisfy $Q_{1n}(\tilde{\gamma}, \tilde{\lambda}) = 0$ and $Q_{2n}(\tilde{\gamma}, \tilde{\lambda}) = 0$, where*

$$Q_{1n}(\gamma, \lambda) \equiv \frac{1}{n} \sum_{i=1}^{n} \frac{Z_i(\beta_0, \gamma)}{1 + \lambda^T Z_i(\beta_0, \gamma)}$$

and

$$Q_{2n}(\gamma, \lambda) \equiv \frac{1}{n} \sum_{i=1}^{n} \frac{1}{1 + \lambda^T Z_i(\beta_0, \gamma)} \left(\frac{\partial Z_i(\beta_0, \gamma)}{\partial \gamma} \right)^T \lambda.$$

Proof. The proof is similar to the proof of Lemma 1 in Qin and Lawless (1994) by using Lemmas 1 and 2.

Proof of Theorem 1. Apply Lemmas 1−3 and arguments in Qin and Lawless (1994).

Proof of Theorem 2. The proof is similar to the proof of Theorem 1.

Before proving Theorems 3 and 4, we need some notations and lemmas. Put, for $i = 1, \ldots, n$,

$$A = \begin{pmatrix} \frac{1}{n} \sum_{t=1}^{n} e_t \epsilon_{t-1} \\ \cdot \\ \cdot \\ \cdot \\ \frac{1}{n} \sum_{t=1}^{n} e_t \epsilon_{t-p} \end{pmatrix}, \quad A^{(i)} = \begin{pmatrix} \frac{1}{n-1} \sum_{t=1, t \ne i}^{n} e_t \epsilon_{t-1} \\ \cdot \\ \cdot \\ \cdot \\ \frac{1}{n-1} \sum_{t=1, t \ne i}^{n} e_t \epsilon_{t-p} \end{pmatrix},$$

$$B = \begin{pmatrix} \frac{1}{n}\sum_{t=1}^{n}\epsilon_{t-1}^2 & \cdots & \frac{1}{n}\sum_{t=1}^{n}\epsilon_{t-p}\epsilon_{t-1} \\ & \cdot & \\ & \cdot & \\ & \cdot & \\ \frac{1}{n}\sum_{t=1}^{n}\epsilon_{t-p}\epsilon_{t-1} & \cdots & \frac{1}{n}\sum_{t=1}^{n}\epsilon_{t-p}^2 \end{pmatrix},$$

$$B^{(i)} = \begin{pmatrix} \frac{1}{n-1}\sum_{t=1,t\neq i}^{n}\epsilon_{t-1}^2 & \cdots & \frac{1}{n-1}\sum_{t=1,t\neq i}^{n}\epsilon_{t-p}\epsilon_{t-1} \\ & \cdot & \\ & \cdot & \\ & \cdot & \\ \frac{1}{n-1}\sum_{t=1,t\neq i}^{n}\epsilon_{t-p}\epsilon_{t-1} & \cdots & \frac{1}{n-1}\sum_{t=1,t\neq i}^{n}\epsilon_{t-p}^2 \end{pmatrix},$$

$$\bar{\Sigma} = E(B), \quad D = (\bar{\Sigma} - B)\bar{\Sigma}^{-1}A, \quad D^{(i)} = (\bar{\Sigma} - B^{(i)})\bar{\Sigma}^{-1}A^{(i)}.$$

Lemma 4. *Under the conditions of Theorem 3,*

$$\hat{\gamma}(\beta_0) - \gamma_0 + \bar{\Sigma}^{-1}A = O_p(n^{-1}), \tag{4.3}$$

$$\max_{1\leq i\leq n}|\hat{\gamma}^{(i)}(\beta_0) - \gamma_0 + \bar{\Sigma}^{-1}A^{(i)}| = O_p(n^{-1}), \tag{4.4}$$

$$\max_{1\leq i\leq n}|B(\hat{\gamma}^{(i)}(\beta_0) - \hat{\gamma}(\beta_0)) + B\bar{\Sigma}^{-1}(A^{(i)} - A) + D^{(i)} - D| = o_p(n^{-3/2}). \tag{4.5}$$

Proof. Equation (4.3) follows from

$$0 = A + B(\hat{\gamma}(\beta_0) - \gamma_0) = D + B(\hat{\gamma}(\beta_0) - \gamma_0 + \bar{\Sigma}^{-1}A), \tag{4.6}$$

$A = O_p(n^{-1/2})$, and $\bar{\Sigma} - B = O_p(n^{-1/2})$. Write

$$B^{(i)} = \frac{n}{n-1}B - \frac{B_i^*}{n-1}, \quad A^{(i)} = \frac{n}{n-1}A - \frac{A_i^*}{n-1}, \tag{4.7}$$

where

$$B_i^* = \begin{pmatrix} \epsilon_{i-1}^2 & \cdots & \epsilon_{i-p}\epsilon_{i-1} \\ & \cdot & \\ & \cdot & \\ & \cdot & \\ \epsilon_{i-p}\epsilon_{i-1} & \cdots & \epsilon_{i-p}^2 \end{pmatrix}, \quad A_i^* = \begin{pmatrix} e_i\epsilon_{i-1} \\ \cdot \\ \cdot \\ \cdot \\ e_i\epsilon_{i-p} \end{pmatrix}.$$

Since $\max_{1 \leq i \leq n, 1 \leq j \leq p} |\epsilon_{i-1}\epsilon_{i-j}| = o_p(n^{1/2})$, it follows from (4.7) that

$$B^{(i)} = O_p(1) \quad \text{and} \quad B^{(i)} - B = o_p(n^{-1/2}) \quad \text{uniformly in} \quad i = 1, \ldots, n. \quad (4.8)$$

Similarly,

$$\bar{\Sigma} - B^{(i)} = O_p(n^{-1/2}), A^{(i)} = O_p(n^{-1/2}) \text{ and } A^{(i)} - A = o_p(n^{-1/2})$$
$$\text{uniformly in } i = 1, \ldots, n. \quad (4.9)$$

Therefore, (4.4) follows from (4.9) and $0 = D^{(i)} + B^{(i)}(\hat{\gamma}^{(i)}(\beta_0) - \gamma_0 + \bar{\Sigma}^{-1}A^{(i)})$. By writing $0 = D^{(i)} + (B^{(i)} - B)(\hat{\gamma}^{(i)}(\beta_0) - \gamma_0 + \bar{\Sigma}^{-1}A^{(i)}) + B(\hat{\gamma}^{(i)}(\beta_0) - \gamma_0 + \bar{\Sigma}^{-1}A^{(i)})$, (4.5) follows from (4.4), (4.6), and (4.8).

Lemma 5. *Under the conditions of Theorem 3, we have*

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} Z_{n,t}^*(\beta_0) \stackrel{d}{=} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} e_t \tilde{e}_t + o_p(1) \stackrel{d}{\to} N(0, E(e_t^2 \tilde{e}_t^2))$$

and

$$\frac{1}{n} \sum_{t=1}^{n} Z_{n,t}^{*2}(\beta_0) \stackrel{p}{\to} E(e_t^2 \tilde{e}_t^2)$$

as $n \to \infty$, where

$$\tilde{e}_t = \frac{X_{t-1}}{\sqrt{1+X_{t-1}^2}} + \sum_{k=1}^{p} \gamma_{0,k} \frac{X_{t-k-1}}{\sqrt{1+X_{t-k-1}^2}}$$

$$- \begin{pmatrix} \epsilon_{t-1} \\ . \\ . \\ . \\ \epsilon_{t-p} \end{pmatrix}^T \begin{pmatrix} E(\epsilon_p^2) & \cdots & E(\epsilon_p \epsilon_1) \\ . & & \\ . & & \\ . & & \\ E(\epsilon_p \epsilon_1) & \cdots & E(\epsilon_1^2) \end{pmatrix} \begin{pmatrix} E\left\{\epsilon_p\left(\frac{X_p}{\sqrt{1+X_p^2}} + \sum_{k=1}^{p} \gamma_{0,k} \frac{X_{p-k}}{\sqrt{1+X_{p-k}^2}}\right)\right\} \\ . \\ . \\ . \\ E\left\{\epsilon_1\left(\frac{X_p}{\sqrt{1+X_p^2}} + \sum_{k=1}^{p} \gamma_{0,k} \frac{X_{p-k}}{\sqrt{1+X_{p-k}^2}}\right)\right\} \end{pmatrix}.$$

Proof. By (4.7)−(4.9) and

$$\sum_{i=1}^{n}(A^{(i)} - A) = 0 \quad \text{and} \quad \sum_{i=1}^{n}(B^{(i)} - B) = 0, \quad (4.10)$$

we have

$$\sum_{i=1}^{n}\{D^{(i)} - D\} = \sum_{i=1}^{n}(B - B^{(i)})\bar{\Sigma}^{-1}(A^{(i)} - A)$$

$$= \sum_{i=1}^{n}(\frac{n}{n-1}B - B^{(i)})\bar{\Sigma}^{-1}(A^{(i)} - \frac{n}{n-1}A) - \frac{n}{(n-1)^2}B\bar{\Sigma}^{-1}A$$

$$= -\frac{1}{(n-1)^2}\sum_{i=1}^{n}B_i^*\bar{\Sigma}^{-1}A_i^* + O_p(n^{-1})$$

$$= O_p(n^{-1}), \tag{4.11}$$

$$D^{(i)} - D = A^{(i)} - A - (B^{(i)} - B)\bar{\Sigma}^{-1}A^{(i)} - B\bar{\Sigma}^{-1}(A^{(i)} - A)$$

$$= o_p(n^{-1/2}) \quad \text{uniformly in} \quad i = 1, \ldots, n. \tag{4.12}$$

Using (4.5), (4.10), (4.11), and (4.12), we can show that, for any $p \times p$ matrix Δ,

$$no_p(n^{-3/2}) = \sum_{i=1}^{n}\{\Delta(\hat{\gamma}(\beta_0) - \hat{\gamma}^{(i)}(\beta_0)) - \Delta\bar{\Sigma}^{-1}(A^{(i)} - A) - \Delta B^{-1}(D^{(i)} - D)\}$$

$$= \sum_{i=1}^{n}\Delta(\hat{\gamma}(\beta_0) - \hat{\gamma}^{(i)}(\beta_0)) + o_p(n^{-1/2}),$$

$$\sum_{i=1}^{n}(A^{(i)} - A)^T\bar{\Sigma}^{-1}\Delta\bar{\Sigma}^{-1}(A^{(i)} - A)$$

$$= \sum_{i=1}^{n}(A^{(i)} - \frac{n}{n-1}A)^T\bar{\Sigma}^{-1}\Delta\bar{\Sigma}^{-1}(A^{(i)} - \frac{n}{n-1}A) + O_p(n^{-1})$$

$$= \frac{1}{(n-1)^2}\sum_{i=1}^{n}A_i^{*T}\bar{\Sigma}^{-1}\Delta\bar{\Sigma}^{-1}A_i^* + O_p(n^{-1})$$

$$= O_p(n^{-1}),$$

$$\sum_{i=1}^{n}(A^{(i)} - A)^T\bar{\Sigma}^{-1}\Delta B^{-1}(D^{(i)} - D)$$

$$= \sum_{i=1}^{n}(A^{(i)} - \frac{n}{n-1}A)^T\bar{\Sigma}^{-1}\Delta B^{-1}(D^{(i)} - D) + O_p(n^{-2})$$

$$= -\frac{1}{n-1}\sum_{i=1}^{n}A_i^{*T}\bar{\Sigma}^{-1}\Delta B^{-1}(D^{(i)} - D) + O_p(n^{-2})$$

$$= O_p(n^{-1/2})o_p(n^{-1/2}) + O_p(n^{-2}) = O_p(n^{-1}),$$

$$\sum_{i=1}^{n}(D^{(i)} - D)^T B^{-1}\Delta B^{-1}(D^{(i)} - D)$$

$$= \sum_{i=1}^{n}\{A^{(i)} - A\}^T B^{-1}\Delta B^{-1}(D^{(i)} - D)$$

$$-\sum_{i=1}^{n}\{(B^{(i)} - B)\bar{\Sigma}^{-1}A^{(i)}\}^{T}B^{-1}\Delta B^{-1}(D^{(i)} - D)$$

$$-\{B\bar{\Sigma}^{-1}(A^{(i)} - A)\}^{T}B^{-1}\Delta B^{-1}(D^{(i)} - D)$$

$$= O_{p}(n^{-1}) + no_{p}(n^{-1/2})O_{p}(n^{-1/2})o_{p}(n^{-1/2}) + O_{p}(n^{-1})$$

$$= o_{p}(n^{-1/2}),$$

$$\sum_{i=1}^{n}(A^{(i)} - A)^{T}\bar{\Sigma}^{-1}\Delta(\hat{\gamma}(\beta_0) - \hat{\gamma}^{(i)}(\beta_0))$$

$$= \sum_{i=1}^{n}(A^{(i)} - A)^{T}\bar{\Sigma}^{-1}\Delta\{\hat{\gamma}(\beta_0) - \hat{\gamma}^{(i)}(\beta_0) - \bar{\Sigma}^{-1}(A^{(i)} - A) - B^{-1}(D^{(i)} - D)\}$$

$$+ O_{p}(n^{-1})$$

$$= no_{p}(n^{-1/2})o_{p}(n^{-3/2}) + O_{p}(n^{-1}) = O_{p}(n^{-1}),$$

$$\sum_{i=1}^{n}(D^{(i)} - D)^{T}B^{-1}\Delta(\hat{\gamma}(\beta_0) - \hat{\gamma}^{(i)}(\beta_0)) = O_{p}(n^{-1}),$$

$$no_{p}(n^{-3/2})o_{p}(n^{-3/2}) = \sum_{i=1}^{n}\left\{\hat{\gamma}(\beta_0) - \hat{\gamma}^{(i)}(\beta_0) - \bar{\Sigma}^{-1}(A^{(i)} - A) - B^{-1}(D^{(i)} - D)\right\}^{T}\Delta$$

$$\times \left\{\hat{\gamma}(\beta_0) - \hat{\gamma}^{(i)}(\beta_0) - \bar{\Sigma}^{-1}(A^{(i)} - A) - B^{-1}(D^{(i)} - D)\right\}$$

$$= \sum_{i=1}^{n}(\hat{\gamma}(\beta_0) - \hat{\gamma}^{(i)}(\beta_0))^{T}\Delta(\hat{\gamma}(\beta_0) - \hat{\gamma}^{(i)}(\beta_0)) + o_{p}(n^{-1/2}),$$

which imply that

$$\begin{cases} \sum_{i=1}^{n}\Delta(\hat{\gamma}(\beta_0) - \hat{\gamma}^{(i)}(\beta_0)) = o_{p}(n^{-1/2}), \\ \sum_{i=1}^{n}(\hat{\gamma}(\beta_0) - \hat{\gamma}^{(i)}(\beta_0))^{T}\Delta(\hat{\gamma}(\beta_0) - \hat{\gamma}^{(i)}(\beta_0)) = o_{p}(n^{-1/2}). \end{cases} \tag{4.13}$$

For $j = 1, \ldots, n$, put

$$W_{j1} = \sum_{l=1}^{p}(\hat{\gamma}_l(\beta_0) - \hat{\gamma}_l^{(j)}(\beta_0))\sum_{t=1}^{n}\epsilon_{t-l}\left(\frac{X_{t-1}}{\sqrt{1 + X_{t-1}^2}} + \sum_{k=1}^{p}\gamma_{0,k}\frac{X_{t-k-1}}{\sqrt{1 + X_{t-k-1}^2}}\right)$$

$$+ \sum_{l=1}^{p}(\hat{\gamma}_l(\beta_0) - \hat{\gamma}_l^{(j)}(\beta_0))\sum_{t=1}^{n}e_t\frac{X_{t-l-1}}{\sqrt{1 + X_{t-l-1}^2}}$$

$$+\frac{1}{2}\sum_{k=1}^{p}\sum_{l=1}^{p}(\hat{\gamma}_k(\beta_0) - \hat{\gamma}_k^{(j)}(\beta_0))(\hat{\gamma}_l(\beta_0) - \hat{\gamma}_l^{(j)}(\beta_0))\sum_{t=1}^{n}\epsilon_{t-k}\frac{X_{t-l-1}}{\sqrt{1+X_{t-l-1}^2}},$$

$$W_{j2} = \left(\epsilon_{j-1}\left\{\frac{X_{j-1}}{\sqrt{1+X_{j-1}^2}} + \sum_{k=1}^{p}\gamma_{0,k}\frac{X_{j-k-1}}{\sqrt{1+X_{j-k-1}^2}}\right\} + e_j\frac{X_{j-1-1}}{\sqrt{1+X_{j-1-1}^2}},\right.$$

$$\left.\cdots, \epsilon_{j-p}\left\{\frac{X_{j-1}}{\sqrt{1+X_{j-1}^2}} + \sum_{k=1}^{p}\gamma_{0,k}\frac{X_{j-k-1}}{\sqrt{1+X_{j-k-1}^2}}\right\} + e_j\frac{X_{j-p-1}}{\sqrt{1+X_{j-p-1}^2}}\right)^T.$$

Then it follows from Lemma 4, (4.13) and Taylor expansions that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Z_{n,i}^*(\beta_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}W_{i1} + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\hat{\gamma}^{(i)}(\beta_0) - \gamma_0)^T W_{i2}$$

$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}e_i\left(\frac{X_{i-1}}{\sqrt{1+X_{i-1}^2}} + \sum_{k=1}^{p}\gamma_{0,k}\frac{X_{i-k-1}}{\sqrt{1+X_{i-k-1}^2}}\right) + o_p(1)$$

$$= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(A^{(i)})^T\Sigma^{-1}W_{i2}$$

$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}e_i\left(\frac{X_{i-1}}{\sqrt{1+X_{i-1}^2}} + \sum_{k=1}^{p}\gamma_{0,k}\frac{X_{i-k-1}}{\sqrt{1+X_{i-k-1}^2}}\right) + o_p(1)$$

$$= -(\sqrt{n}A)^T\Sigma^{-1}\frac{1}{n}\sum_{i=1}^{n}W_{i2}$$

$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}e_i\left(\frac{X_{i-1}}{\sqrt{1+X_{i-1}^2}} + \sum_{k=1}^{p}\gamma_{0,k}\frac{X_{i-k-1}}{\sqrt{1+X_{i-k-1}^2}}\right) + o_p(1)$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}e_i\tilde{e}_i + o_p(1).$$

Now apply a martingale central limit theorem argument as in Lemma 1 to $(1/\sqrt{n})\sum_{i=1}^{n}e_i\tilde{e}_i$ to achieve $(1/\sqrt{n})\sum_{i=1}^{n}Z_{n,i}^*(\beta_0) \xrightarrow{d} N(0, E(e_1^2\tilde{e}_1^2))$. Similarly, we can show that

$$\frac{1}{n}\sum_{i=1}^{n}Z_{n,i}^{*2}(\beta_0) = \frac{1}{n}\sum_{i=1}^{n}e_i^2\tilde{e}_i^2 + o_p(1) \xrightarrow{p} E(e_1^2\tilde{e}_1^2).$$

This completes the proof.

Proof of Theorem 3. The claim can be proven by using Lemma 5, and arguments in Qin and Lawless (1994).

Proof of Theorem 4. The argument is similar to the proof of Theorem 3.

Acknowledgement

References

Chan, N. H., Li, D. and Peng, L. (2012). Toward a unified interval estimation of autoregressions. *Econom. Theory* **28**, 705-717.

Chen, S., Min, W. and Chen, R. (2013). Model identification for time series with dependent innovations. *Statist. Sinica* **23**, 873-899.

Chen, S. and Van Keilegom, I. (2009). A review on empirical likelihood methods for regression. *Test* **18**, 415-447.

Hall, P. and Heyde, C. (1980). *Martingale Limit Theory and Its Applications.* Academic Press.

Hill, J. B. and Peng, L. (2014). Unified interval estimation for random coefficient autoregressive models. *J. Time Ser. Anal.* **35**, 282-297.

Jing, B. Y., Yuan, J. and Zhou, W. (2009). Jackknife empirical likelihood. *J. Amer. Statist. Assoc.* **104**, 1224-1232.

Kitamura, Y. (1997). Empirical likelihood methods with weakly dependent processes. *Ann. Statist.* **25**, 2084-2102.

Li, D., Chan, N. H. and Peng, L. (2014). Empirical likelihood test for causality for bivariate AR(1) processes. *Econom. Theory* **30**, 357-371.

Li, M., Peng, L. and Qi, Y. (2011). Reduce computation in profile empirical likelihood method. *Canad. J. Statist.* **39**, 370-384.

Ling, S. (2007). Self-weighted and local quasi-maximum likelihood estimators for ARMA-GARCH/IGARCH models. *J. Econom.* **140**, 849-873.

McLeish, D. (1974). Dependent central limit theorems and invariance principles. *Ann. Probab.* **2**, 620-628.

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

Pan, J., Wang, H. and Yao, Q. (2007). Weighted least absolute deviations estimation for ARMA models with infinite variance. *Econom. Theory* **23**, 852-879.

Peng, L. (2012). Approximate jackknife empirical likelihood method for estimating equations. *Canad. J. Statist.* **40**, 110-123.

Phillips, P. C. B. (1987). Time series regression with a unit root. *Econometrica* **55**, 277-301.

Phillips, P. C. B. and Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika* **75**, 335-346.

Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating functions. *Ann. Statist.* **22**, 300-325.

Wu, W. B. (2005). Unit root testing for functionals of linear processes. *Econom. Theory* **22**, 1-14.

Wu, W. B. and Min, W. L. (2005). On linear processes with dependent innovations. *Stochastic Process. Appl.* **115**, 939–958.

Xiao, Z., Linton, O. B., Carroll, R. J. and Mammen, E. (2003). More efficient local polynomial
    estimation in nonparametric regression with autocorrelated errors. *J. Amer. Statist. Assoc.*
    98, 980-992.

Department of Economics, University of North Carolina, Chapel Hill, NC 27599-3305, USA.

E-mail: jbhill@email.unc.edu

School of Management, Fudan University, Shanghai 200433, P.R. China.

E-mail: deyuanli@fudan.edu.cn

Department of Risk Management and Insurance, Georgia State University, Atlanta, GA 30303,
USA.

E-mail: lpeng@gsu.edu