Econometric Theory, **30**, 2014, 357–371.
doi:10.1017/S0266466613000339

EMPIRICAL LIKELIHOOD TEST FOR CAUSALITY OF BIVARIATE AR(1) PROCESSES

D. Li
Fudan University
N. H. Chan
Chinese University of Hong Kong and Renmin University of China
L. Peng
Georgia Institute of Technology

Testing for causality is of critical importance for many econometric applications. For bivariate AR(1) processes, the limit distributions of causality tests based on least squares estimation depend on the presence of nonstationary processes. When nonstationary processes are present, the limit distributions of such tests are usually very complicated, and the full-sample bootstrap method becomes inconsistent as pointed out in Choi (2005, *Statistics and Probability Letters* 75, 39–48). In this paper, a profile empirical likelihood method is proposed to test for causality. The proposed test statistic is robust against the presence of nonstationary processes in the sense that one does not have to determine the existence of nonstationary processes a priori. Simulation studies confirm that the proposed test statistic works well.

1. INTRODUCTION

Consider the bivariate first-order autoregressive (AR) process

$$\begin{cases} Y_{i,1} = \beta_{11} Y_{i-1,1} + \beta_{12} Y_{i-1,2} + \epsilon_{i,1}, \\ Y_{i,2} = \beta_{21} Y_{i-1,1} + \beta_{22} Y_{i-1,2} + \epsilon_{i,2}, \end{cases} \tag{1}$$

for $i = 1, \ldots, n$ with $Y_{0,1} = Y_{0,2} = 0$, where $(\epsilon_{1,1}, \epsilon_{1,2})^T, \ldots, (\epsilon_{n,1}, \epsilon_{n,2})^T$ are independent and identically distributed random vectors with mean zero and finite variances.

Testing for causality has been studied in economic applications since the seminal work by Granger (1969). For example, if one wants to determine if there is a causality between $Y_{i,2}$ and $Y_{i,1}$, then one can test for $H_0 : \beta_{12} = 0$ and $\beta_{21} = 0$.

We thank the co-editor, Professor Yuichi Kitamura, and two anonymous referees for their helpful comments and constructive suggestions, which led to an improved version of this paper. This research is supported in part by HKSAR-RGC-GRF grants. 400408 and 400410, HKSAR-RGC-CRF grant CityU8/CRF/09, NSF grant DMS-1005336, and NNSFC grant 11171074. Address correspondence to D. Li, Department of Statistics, School of Management, Fudan University, 670 Guoshun Road, Shanghai 200433, P.R. China; e-mail: deyuanli@fudan.edu.cn.

© Cambridge University Press 2013

https://doi.org/10.1017/S0266466613000339 Published online by Cambridge University Press

Naturally, one way to construct a test statistic is by means of the least squares estimation of $(\beta_{12}, \beta_{21})^T$. The limit distribution of such a test statistic depends, however, on whether both processes are stationary or if one of them is nonstationary; see Toda and Phillips (1993, 1994) for details. Because the limit distribution is very different when a nonstationary process presents, it is therefore necessary to determine if a nonstationary process exists before one can apply a causality test. Undoubtedly, distinguishing the different cases of nonstationarity and stationarity poses a serious challenge in practice. To complicate matters further, it is found that bootstrapping a causality test becomes inconsistent when a nonstationary process is present; see Choi (2005a) for details. It is therefore important and practical to develop a causality test that works for both stationary and nonstationary processes. Some unified methods for testing linear constraints and causality include the lag augmented method (LAM) in Toda and Yamamoto (1995), the fully modified method in Phillips (1995), the improved lag augmented method in Kurozumi and Yamamoto (2000) and the blockwise bootstrap method in Choi (2005b). A simulation comparison was conducted in Yamada and Toda (1998), where discussions on advantages and disadvantages of these methods are also given.

Motivated by the recent study on weighted estimation for AR(1) processes in Chan, Li, and Peng (2012), we propose an alternative way to test causality for both stationary and nonstationary processes by applying the profile empirical likelihood method to some weighted score equations. The new method is easy to implement and has good finite-sample performance. For an overview on empirical likelihood methods, we refer to Owen (2001). This paper is organized as follows. Section 2 presents the methodology and asymptotic results. A simulation study is given in Section 3. All proofs are in the Appendix.

2. METHODOLOGY

It is known that the least squares estimate for $\beta = (\beta_{11}, \beta_{12}, \beta_{21}, \beta_{22})^T$ is the solution to the following score equations:

$$\begin{cases} \sum_{i=1}^{n} Y_{i-1,1}\{Y_{i,1} - \beta_{11}Y_{i-1,1} - \beta_{12}Y_{i-1,2}\} = 0, \\ \sum_{i=1}^{n} Y_{i-1,2}\{Y_{i,1} - \beta_{11}Y_{i-1,1} - \beta_{12}Y_{i-1,2}\} = 0, \\ \sum_{i=1}^{n} Y_{i-1,1}\{Y_{i,2} - \beta_{21}Y_{i-1,1} - \beta_{22}Y_{i-1,2}\} = 0, \\ \sum_{i=1}^{n} Y_{i-1,2}\{Y_{i,2} - \beta_{21}Y_{i-1,1} - \beta_{22}Y_{i-1,2}\} = 0. \end{cases} \tag{2}$$

Based on these equations, one may employ the profile empirical likelihood method of Qin and Lawless (1994) to test for causality. However, when nonstationary sequences are involved, the Wilks theorem fails; i.e., the asymptotic distribution is no longer a chi-square distribution. Motivated by the method of unifying the stationary case and nonstationary case for AR models in Chan et al. (2012),

https://doi.org/10.1017/S0266466613000339 Published online by Cambridge University Press

one may apply the profile empirical likelihood method to the following weighted score equations:

$$\begin{cases} \sum_{i=1}^{n} \frac{Y_{i-1,1}}{\sqrt{\delta_1+Y_{i-1,1}^2}} \{Y_{i,1} - \beta_{11} Y_{i-1,1} - \beta_{12} Y_{i-1,2}\} = 0, \\ \sum_{i=1}^{n} \frac{Y_{i-1,2}}{\sqrt{\delta_1+Y_{i-1,2}^2}} \{Y_{i,1} - \beta_{11} Y_{i-1,1} - \beta_{12} Y_{i-1,2}\} = 0, \\ \sum_{i=1}^{n} \frac{Y_{i-1,1}}{\sqrt{\delta_1+Y_{i-1,1}^2}} \{Y_{i,2} - \beta_{21} Y_{i-1,1} - \beta_{22} Y_{i-1,2}\} = 0, \\ \sum_{i=1}^{n} \frac{Y_{i-1,2}}{\sqrt{\delta_1+Y_{i-1,2}^2}} \{Y_{i,2} - \beta_{21} Y_{i-1,1} - \beta_{22} Y_{i-1,2}\} = 0, \end{cases} \tag{3}$$

for some $\delta_1 \geq 0$. In practice δ_1 can simply be chosen as one. Unfortunately this idea does not work for the case of cointegration such as when both processes are nearly integrated. The reason is that both

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{Y_{i-1,1}}{\sqrt{\delta_1 + Y_{i-1,1}^2}} \left\{ Y_{i,1} - \beta_{11}^0 Y_{i-1,1} - \beta_{12}^0 Y_{i-1,2} \right\}$$

and

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{Y_{i-1,2}}{\sqrt{\delta_1 + Y_{i-1,2}^2}} \left\{ Y_{i,1} - \beta_{11}^0 Y_{i-1,1} - \beta_{12}^0 Y_{i-1,2} \right\}$$

have the same asymptotic limit as that of $1/\sqrt{n}\sum_{i=1}^{n} \epsilon_{i,1}$ when both processes are nearly integrated; i.e., we have $Y_{n,1} \xrightarrow{p} \infty$ and $Y_{n,2} \xrightarrow{p} \infty$ as $n \to \infty$. Here $\beta^0 = \left(\beta_{11}^0, \beta_{12}^0, \beta_{21}^0, \beta_{22}^0\right)^T$ denotes the true value of β.

To overcome this degenerate issue, a simple idea is to add some pseudo samples into the estimating equations, say, to consider

$$\begin{cases} \sum_{i=1}^{n} \frac{Y_{i-1,1}}{\sqrt{\delta_1+Y_{i-1,1}^2}} \{Y_{i,1} - \beta_{11} Y_{i-1,1} - \beta_{12} Y_{i-1,2}\} + \sum_{i=1}^{n} W_{i,1} = 0, \\ \sum_{i=1}^{n} \frac{Y_{i-1,2}}{\sqrt{\delta_1+Y_{i-1,2}^2}} \{Y_{i,1} - \beta_{11} Y_{i-1,1} - \beta_{12} Y_{i-1,2}\} = 0, \\ \sum_{i=1}^{n} \frac{Y_{i-1,1}}{\sqrt{\delta_1+Y_{i-1,1}^2}} \{Y_{i,2} - \beta_{21} Y_{i-1,1} - \beta_{22} Y_{i-1,2}\} + \sum_{i=1}^{n} W_{i,2} = 0, \\ \sum_{i=1}^{n} \frac{Y_{i-1,2}}{\sqrt{\delta_1+Y_{i-1,2}^2}} \{Y_{i,2} - \beta_{21} Y_{i-1,1} - \beta_{22} Y_{i-1,2}\} = 0, \end{cases}$$

where $W_{1,1}, \ldots, W_{n,1}, W_{1,2}, \ldots, W_{n,2}$ are independent standard normal random variables. In this case, the asymptotic variance of

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{Y_{i-1,1}}{\sqrt{\delta_1 + Y_{i-1,1}^2}} \left\{ Y_{i,1} - \beta_{11}^0 Y_{i-1,1} - \beta_{12}^0 Y_{i-1,2} \right\} + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_{i,1} \tag{4}$$

becomes larger. Here we replace the weights $\sqrt{\delta_1 + Y_{i-1,1}^2}$ by $\left\{\delta_1 + Y_{i-1,1}^2\right\}^{\delta_2}$ for some $\delta_2 > \frac{1}{2}$ so that the asymptotic variance of (4) is equal to the asymptotic

https://doi.org/10.1017/S0266466613000339 Published online by Cambridge University Press

variance of $1/\sqrt{n}\sum_{i=1}^{n} W_{i,1}$ when the process $\{Y_{i,1}\}_{i=1}^{\infty}$ is nearly integrated. A large δ_2 causes the first sum in (4) to disappear faster in finite samples. On the other hand, by noting that

$$\begin{aligned}&\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\frac{Y_{i-1,1}}{\{\delta_1+Y_{i-1,1}^2\}^{\delta_2}}\{Y_{i,1}-\beta_{11}^0 Y_{i-1,1}-\beta_{12}Y_{i-1,2}\}+\tfrac{1}{n}\textstyle\sum_{i=1}^{n} W_{i,1}\\&\quad=(\beta_{12}^0-\beta_{12})\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\frac{Y_{i-1,1}}{\{\delta_1+Y_{i-1,1}^2\}^{\delta_2}}Y_{i-1,2}+\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\left\{\frac{Y_{i-1,1}}{\{\delta_1+Y_{i-1,1}^2\}^{\delta_2}}\epsilon_{i,1}+W_{i,1}\right\},\end{aligned}$$

we prefer the quantity $1-\delta_2>0$ to be as large as possible so that the proposed test is more powerful when both $\{Y_{i,1}\}_{i=1}^{\infty}$ and $\{Y_{i,2}\}_{i=1}^{\infty}$ are nearly integrated. With these considerations, choose $\delta_1=1$ and $\delta_2=\frac{3}{4}$ and apply the profile empirical likelihood method to the following estimating equations to test $H_0:\beta_{12}=0$ and $\beta_{21}=0$:

$$\begin{cases}\sum_{i=1}^{n}\left\{\frac{Y_{i-1,1}}{\{1+Y_{i-1,1}^2\}^{3/4}}\{Y_{i,1}-\beta_{11}Y_{i-1,1}-\beta_{12}Y_{i-1,2}\}+W_{i,1}\right\}=0,\\ \sum_{i=1}^{n}\frac{Y_{i-1,2}}{\sqrt{1+Y_{i-1,2}^2}}\{Y_{i,1}-\beta_{11}Y_{i-1,1}-\beta_{12}Y_{i-1,2}\}=0,\\ \sum_{i=1}^{n}\left\{\frac{Y_{i-1,1}}{\{1+Y_{i-1,1}^2\}^{3/4}}\{Y_{i,2}-\beta_{21}Y_{i-1,1}-\beta_{22}Y_{i-1,2}\}+W_{i,2}\right\}=0,\\ \sum_{i=1}^{n}\frac{Y_{i-1,2}}{\sqrt{1+Y_{i-1,2}^2}}\{Y_{i,2}-\beta_{21}Y_{i-1,1}-\beta_{22}Y_{i-1,2}\}=0.\end{cases}\tag{5}$$

To reduce the effect of the random seed in generating the aforementioned $W_{i,1}$ and $W_{i,2}$ for $i=1,\ldots,n$, one can use $W_{i,j}=m^{-1/2}\sum_{l=1}^{m}\tilde{W}_{i,j,l}$, where $\tilde{W}_{i,j,l}$ for $i=1,\ldots,n,\ j=1,2,\ l=1,\ldots,m$ are independent standard normal random variables. In the simulation study, we use $m=1{,}000$. Note that such a test is reminiscent of the classical randomized test with which the effect of the random seed is reduced.

Put

$$\begin{aligned}Z_i(\beta)&=(Z_{i,1}(\beta),Z_{i,2}(\beta),Z_{i,3}(\beta),Z_{i,4}(\beta))^T,\\ Z_{i,1}(\beta)&=\frac{Y_{i-1,1}}{\{1+Y_{i-1,1}^2\}^{3/4}}\{Y_{i,1}-\beta_{11}Y_{i-1,1}-\beta_{12}Y_{i-1,2}\}+W_{i,1},\\ Z_{i,2}(\beta)&=\frac{Y_{i-1,2}}{\sqrt{1+Y_{i-1,2}^2}}\{Y_{i,1}-\beta_{11}Y_{i-1,1}-\beta_{12}Y_{i-1,2}\},\\ Z_{i,3}(\beta)&=\frac{Y_{i-1,1}}{\{1+Y_{i-1,1}^2\}^{3/4}}\{Y_{i,2}-\beta_{21}Y_{i-1,1}-\beta_{22}Y_{i-1,2}\}+W_{i,2},\\ Z_{i,4}(\beta)&=\frac{Y_{i-1,2}}{\sqrt{1+Y_{i-1,2}^2}}\{Y_{i,2}-\beta_{21}Y_{i-1,1}-\beta_{22}Y_{i-1,2}\},\end{aligned}$$

https://doi.org/10.1017/S0266466613000339 Published online by Cambridge University Press

for $i = 1, \ldots, n$. Then define the empirical likelihood function for β as

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

It follows from the Lagrange multiplier method that

$$l_n(\beta) = -2 \log L_n(\beta) = 2 \sum_{i=1}^{n} \log\{1 + \lambda^T Z_i(\beta)\},$$

where $\lambda = \lambda(\beta)$ satisfies

$$\sum_{i=1}^{n} \frac{Z_i(\beta)}{1 + \lambda^T Z_i(\beta)} = 0. \tag{6}$$

Because we are only interested in β_{12} and β_{21}, consider the following profile empirical likelihood function:

$$l_n^P(\beta_{12}, \beta_{21}) = \min_{\beta_{11}, \beta_{22}} l_n(\beta).$$

The following theorem shows that the Wilks theorem holds for the proposed profile empirical likelihood method.

THEOREM 1. *Suppose that model (1) holds with* $\mathbb{E}(|\epsilon_{1,1}|^d + |\epsilon_{1,2}|^d) < \infty$ *for some* $d > 2$, $\beta_{12} = 0$, *and* $\beta_{21} = 0$. *Further we assume one of the following cases:*

(i) both processes are stationary: $|\beta_{11}| < 1$ *and* $|\beta_{22}| < 1$ *are independent of n;*

(ii) only $\{Y_{i,1}\}$ *is near integrated:* $\beta_{11} = 1 - \gamma_{11}/n$ *for some* $\gamma_{11} \in R$ *and* $|\beta_{22}| < 1$ *is independent of n;*

(iii) only $\{Y_{i,2}\}$ *is near integrated:* $|\beta_{11}| < 1$ *is independent of n and* $\beta_{22} = 1 - \gamma_{22}/n$ *for some* $\gamma_{22} \in R$;

(iv) both processes are near integrated: $\beta_{11} = 1 - \gamma_{11}/n$ *for some* $\gamma_{11} \in R$ *and* $\beta_{22} = 1 - \gamma_{22}/n$ *for some* $\gamma_{22} \in R$.

Then $l_n^P(0, 0)$ *converges in distribution to a chi-square distribution with two degrees of freedom.*

Remark 1. When the model has intercepts, i.e.,

$$\begin{cases} Y_{i,1} = \beta_{10} + \beta_{11} Y_{i-1,1} + \beta_{12} Y_{i-1,2} + \epsilon_{i,1} \\ Y_{i,2} = \beta_{20} + \beta_{21} Y_{i-1,1} + \beta_{22} Y_{i-1,2} + \epsilon_{i,2}, \end{cases}$$

one can apply the profile empirical likelihood method to the following weighted score equations:

https://doi.org/10.1017/S0266466613000339 Published online by Cambridge University Press

$$\begin{cases} \sum_{i=1}^{n}\{Y_{i,1}-\beta_{10}-\beta_{11}Y_{i-1,1}-\beta_{12}Y_{i-1,2}\}=0, \\ \sum_{i=1}^{n}\left\{\frac{Y_{i-1,1}}{\{1+Y_{i-1,1}^2\}^{3/4}}\{Y_{i,1}-\beta_{10}-\beta_{11}Y_{i-1,1}-\beta_{12}Y_{i-1,2}\}+W_{i,1}\right\}=0, \\ \sum_{i=1}^{n}\left\{\frac{Y_{i-1,2}}{\{1+Y_{i-1,2}^2\}^{3/4}}\{Y_{i,1}-\beta_{10}-\beta_{11}Y_{i-1,1}-\beta_{12}Y_{i-1,2}\}+W_{i,2}\right\}=0, \\ \sum_{i=1}^{n}\left\{Y_{i,2}-\beta_{20}-\beta_{21}Y_{i-1,1}-\beta_{22}Y_{i-1,2}\right\}=0, \\ \sum_{i=1}^{n}\left\{\frac{Y_{i-1,1}}{\{1+Y_{i-1,1}^2\}^{3/4}}\{Y_{i,2}-\beta_{20}-\beta_{21}Y_{i-1,1}-\beta_{22}Y_{i-1,2}\}+W_{i,3}\right\}=0, \\ \sum_{i=1}^{n}\left\{\frac{Y_{i-1,2}}{\{1+Y_{i-1,2}^2\}^{3/4}}\{Y_{i,2}-\beta_{20}-\beta_{21}Y_{i-1,1}-\beta_{22}Y_{i-1,2}\}+W_{i,4}\right\}=0, \end{cases}$$

where $W_{i,1}, W_{i,2}, W_{i,3}, W_{i,4}$ for $i=1,\ldots,n$ are independent standard normal random variables. Similar procedures can be developed for bivariate AR(p) models.

Next, consider testing the causality $H_0:\beta_{12}=0$. Obviously one can apply the profile empirical likelihood method to the preceding four equations (5) with nuisance parameters $\beta_{11},\beta_{21},\beta_{22}$. To reduce the computation in the profile empirical likelihood method, we propose employing the profile empirical likelihood method to the first two equations in (5). More specifically, put $\bar{\beta}=(\beta_{11},\beta_{12})^T$, $\bar{Z}_i(\bar{\beta})=\big(\bar{Z}_{i,1}(\bar{\beta}),\bar{Z}_{i,2}(\bar{\beta})\big)^T$,

$$\bar{Z}_{i,1}(\bar{\beta})=\frac{Y_{i-1,1}}{\{1+Y_{i-1,1}^2\}^{3/4}}\{Y_{i,1}-\beta_{11}Y_{i-1,1}-\beta_{12}Y_{i-1,2}\}+W_{i,1},$$

$$\bar{Z}_{i,2}(\beta)=\frac{Y_{i-1,2}}{\sqrt{1+Y_{i-1,2}^2}}\{Y_{i,1}-\beta_{11}Y_{i-1,1}-\beta_{12}Y_{i-1,2}\},$$

for $i=1,\ldots,n$. Then define the empirical likelihood function for $\bar{\beta}$ as

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

It follows from the Lagrange multiplier method that

$$\bar{l}_n(\bar{\beta})=-2\log\bar{L}_n(\bar{\beta})=2\sum_{i=1}^{n}\log\left\{1+\bar{\lambda}^T\bar{Z}_i(\bar{\beta})\right\},$$

where $\bar{\lambda}=\bar{\lambda}(\bar{\beta})$ satisfies

$$\sum_{i=1}^{n}\frac{\bar{Z}_i(\bar{\beta})}{1+\bar{\lambda}^T\bar{Z}_i(\bar{\beta})}=0. \tag{7}$$

Because we are only interested in β_{12}, consider the following profile empirical likelihood function:

$$\bar{l}_n^P(\beta_{12})=\min_{\beta_{11}}\bar{l}_n(\bar{\beta}).$$

https://doi.org/10.1017/S0266466613000339 Published online by Cambridge University Press

The following theorem shows that the Wilks theorem holds for the proposed profile empirical likelihood method.

THEOREM 2. *Suppose that model (1) holds with* $\mathbb{E}(|\epsilon_{1,1}|^d + |\epsilon_{1,2}|^d) < \infty$ *for some* $d > 2$ *and* $\beta_{12} = 0$. *Further we assume one of the following cases:*

(i) both processes are stationary: $|\beta_{11}| < 1$ *and* $|\beta_{22}| < 1$ *are independent of* n;

(ii) only $\{Y_{i,1}\}$ *is near integrated:* $\beta_{11} = 1 - \gamma_{11}/n$ *for some* $\gamma_{11} \in R$, $\beta_{21} = 0$, *and* $|\beta_{22}| < 1$ *is independent of* n;

(iii) only $\{Y_{i,1}\}$ *is stationary:* $|\beta_{11}| < 1$ *is independent of* n *and* $\beta_{22} = 1 - \gamma_{22}/n$ *for some* $\gamma_{22} \in R$;

(iv) both are nonstationary with $\beta_{21} = 0$: $\beta_{11} = 1 - \gamma_{11}/n$ *for some* $\gamma_{11} \in R$, $\beta_{21} = 0$, *and* $\beta_{22} = 1 - \gamma_{22}/n$ *for some* $\gamma_{22} \in R$;

(v) both are nonstationary with a nonzero β_{21}: $\beta_{11} = 1 - \gamma_{11}/n$ *for some* $\gamma_{11} \in R$, $\beta_{21} \neq 0$, *and either* $\beta_{22} = 1 - \gamma_{22}/n$ *for some* $\gamma_{22} \in R$ *or* $|\beta_{22}| < 1$ *is independent of* n.

Then $\bar{l}_n^P(0)$ *converges in distribution to a chi-square distribution with one degree of freedom.*

Remark 2. Unlike the cointegration conditions (3a)–(3c) in Toda and Phillips (1993), conditions in Theorem 2 allow β_{11} and β_{22} to depend on the sample size n. For testing $H_0 : \beta_{21} = 0$, one can apply the same profile empirical likelihood method to the last two equations in (5).

3. SIMULATIONS

In this section, we examine the finite-sample behavior of the proposed empirical likelihood tests for testing $H_0 : \beta_{12} = 0$ and $\beta_{21} = 0$ against H_a: at least one of them is nonzero and for testing $H_0 : \beta_{12} = 0$ against $H_a : \beta_{12} \neq 0$. We also compare them with the LAM in Toda and Yamamoto (1995) in terms of size and power.

We draw 10,000 random samples with sizes $n = 50$ and $n = 200$ from model (1), where $\epsilon_{i,1} \sim t(3)$, $\epsilon_{i,2} \sim t(5)$, and $\epsilon_{i,1}$ and $\epsilon_{i,2}$ are independent. Consider $\beta_{11} = 0.9, 1$, $\beta_{22} = 0.9, 0.99, 1$, $\beta_{12} = a/\sqrt{n}$, and $\beta_{21} = -a/\sqrt{n}$ for $a = 0, 0.5, 1$. To implement the proposed empirical likelihood tests, we use the "emplik" package in R to compute $l_n((\beta_{11}, 0, 0, \beta_{22})^T)$ and $\bar{l}_n((\beta_{11}, 0)^T)$, and the "nlm" package in R to calculate $l_n^P(0, 0)$ and $\bar{l}_n^P(0)$, with initial values given by the least squares estimates. Moreover, the added pseudo samples are taken as $W_{i,j} = \frac{1}{\sqrt{1{,}000}} \sum_{l=1}^{1{,}000} \tilde{W}_{i,j,l}$, where $\tilde{W}_{i,j,l}$ for $i = 1, \ldots, n$, $j = 1, 2$, $l = 1, \ldots, 1{,}000$ are independent standard normal random variables.

Tables 1–4 report the sizes and the powers of the proposed profile empirical likelihood method (PELM) and the LAM in Toda and Yamamoto (1995). Note that results for $a = 0$ correspond to the size of the proposed test. Tables 1–4 show

https://doi.org/10.1017/S0266466613000339 Published online by Cambridge University Press

TABLE 1. Sizes and powers of the proposed PELM test and the LAM in Toda and Yamamoto (1995) for testing $H_0 : \beta_{12} = 0$ and $\beta_{21} = 0$ against H_a: at least one of them is nonzero. Here $\beta_{11} = 0.9, 1, \beta_{22} = 0.9, 0.99, 1$, $\beta_{12} = a/\sqrt{n}$, and $\beta_{21} = -a/\sqrt{n}$, and the sample size $n = 50$

$(a, \beta_{11}, \beta_{22})$	PELM 5%	LAM 5%	PELM 5%	LAM 5%
(0, 0.9, 0.9)	0.0676	0.0603	0.1186	0.1120
(0.5, 0.9, 0.9)	0.1440	0.1048	0.2302	0.1741
(1, 0.9, 0.9)	0.4064	0.2582	0.5224	0.3625
(0, 0.9, 0.99)	0.0644	0.0615	0.1133	0.1099
(0.5, 0.9, 0.99)	0.1704	0.1062	0.2646	0.1777
(1, 0.9, 0.99)	0.5230	0.2585	0.6362	0.3622
(0, 0.9, 1)	0.0633	0.0609	0.1127	0.1100
(0.5, 0.9, 1)	0.1712	0.1058	0.2624	0.1781
(1, 0.9, 1)	0.5298	0.2579	0.6428	0.3628
(0, 1, 0.9)	0.0704	0.0619	0.1248	0.1105
(0.5, 1, 0.9)	0.1541	0.1049	0.2460	0.1787
(1, 1, 0.9)	0.4761	0.2598	0.5857	0.3624
(0, 1, 0.99)	0.0684	0.0637	0.1255	0.1152
(0.5, 1, 0.99)	0.3189	0.1056	0.4300	0.1753
(1, 1, 0.99)	0.7910	0.2574	0.8544	0.3614
(0, 1, 1)	0.0683	0.0650	0.1251	0.1164
(0.5, 1, 1)	0.3437	0.1055	0.4555	0.1750
(1, 1, 1)	0.8159	0.2578	0.8739	0.3604

TABLE 2. Sizes and powers of the proposed PELM test and the LAM test in Toda and Yamamoto (1995) for testing $H_0 : \beta_{12} = 0$ and $\beta_{21} = 0$ against H_a: at least one of them is nonzero. Here $\beta_{11} = 0.9, 1, \beta_{22} = 0.9, 0.99, 1$, $\beta_{12} = a/\sqrt{n}$, and $\beta_{21} = -a/\sqrt{n}$, and the sample size $n = 200$

$(a, \beta_{11}, \beta_{22})$	PELM 5%	LAM 5%	PELM 5%	LAM 5%
(0, 0.9, 0.9)	0.0484	0.0581	0.1006	0.1067
(0.5, 0.9, 0.9)	0.1304	0.0994	0.2158	0.1766
(1, 0.9, 0.9)	0.3922	0.2634	0.5119	0.3743
(0, 0.9, 0.99)	0.0458	0.0600	0.0929	0.1073
(0.5, 0.9, 0.99)	0.1825	0.0997	0.2925	0.1764
(1, 0.9, 0.99)	0.5776	0.2638	0.7035	0.3727
(0, 0.9, 1)	0.0442	0.0595	0.0913	0.1062
(0.5, 0.9, 1)	0.1806	0.0994	0.2912	0.1763
(1, 0.9, 1)	0.5632	0.2637	0.6931	0.3727

Table continues on overleaf

https://doi.org/10.1017/S0266466613000339 Published online by Cambridge University Press

TABLE 2. *continued*

$(a, \beta_{11}, \beta_{22})$	PELM 5%	LAM 5%	PELM 5%	LAM 5%
(0, 1, 0.9)	0.0535	0.0574	0.1051	0.1074
(0.5, 1, 0.9)	0.1145	0.0994	0.1924	0.1740
(1, 1, 0.9)	0.3295	0.2641	0.4436	0.3733
(0, 1, 0.99)	0.0484	0.0595	0.1033	0.1088
(0.5, 1, 0.99)	0.5869	0.0973	0.6867	0.1760
(1, 1, 0.99)	0.9661	0.2648	0.9800	0.3715
(0, 1, 1)	0.0522	0.0601	0.1014	0.1095
(0.5, 1, 1)	0.7528	0.0982	0.8187	0.1758
(1, 1, 1)	0.9867	0.2640	0.9933	0.3711

TABLE 3. Sizes and powers of the proposed PELM test and the LAM test in Toda and Yamamoto (1995) for testing $H_0 : \beta_{12} = 0$ against $H_a : \beta_{12} \neq 0$. Here $\beta_{11} = 0.9, 1, \beta_{22} = 0.9, 0.99, 1$, $\beta_{12} = a/\sqrt{n}$, and $\beta_{21} = -a/\sqrt{n}$, and the sample size $n = 50$

$(a, \beta_{11}, \beta_{22})$	PELM 5%	LAM 5%	PELM 5%	LAM 5%
(0, 0.9, 0.9)	0.0567	0.0557	0.1075	0.1078
(0.5, 0.9, 0.9)	0.1339	0.0777	0.2150	0.1327
(1, 0.9, 0.9)	0.3872	0.1330	0.5076	0.2106
(0, 0.9, 0.99)	0.0525	0.0576	0.0998	0.1072
(0.5, 0.9, 0.99)	0.1587	0.0783	0.2543	0.1340
(1, 0.9, 0.99)	0.5079	0.1333	0.6284	0.2124
(0, 0.9, 1)	0.0513	0.0576	0.0992	0.1067
(0.5, 0.9, 1)	0.1596	0.0785	0.2512	0.1352
(1, 0.9, 1)	0.5111	0.1341	0.6327	0.2120
(0, 1, 0.9)	0.0608	0.0569	0.1140	0.1098
(0.5, 1, 0.9)	0.1550	0.0772	0.2433	0.1357
(1, 1, 0.9)	0.4471	0.1341	0.5533	0.2111
(0, 1, 0.99)	0.0614	0.0609	0.1157	0.1141
(0.5, 1, 0.99)	0.3224	0.0764	0.4286	0.1352
(1, 1, 0.99)	0.7763	0.1362	0.8474	0.2111
(0, 1, 1)	0.0607	0.0614	0.1154	0.1132
(0.5, 1, 1)	0.3455	0.0757	0.4548	0.1349
(1, 1, 1)	0.8057	0.1354	0.8672	0.2106

that (i) both tests have similar sizes and the accuracy is improved when the sample size becomes large, and (ii) the PELM is much more powerful than the LAM, especially when $a = 1$.

https://doi.org/10.1017/S0266466613000339 Published online by Cambridge University Press

TABLE 4. Sizes and powers of the proposed PELM test and the LAM test in Toda and Yamamoto (1995) for testing $H_0 : \beta_{12} = 0$ against $H_a : \beta_{12} \neq 0$. Here $\beta_{11} = 0.9, 1, \beta_{22} = 0.9, 0.99, 1$, $\beta_{12} = a/\sqrt{n}$, and $\beta_{21} = -a/\sqrt{n}$, and the sample size $n = 200$

$(a, \beta_{11}, \beta_{22})$	PELM 5%	LAM 5%	PELM 5%	LAM 5%
(0, 0.9, 0.9)	0.0525	0.0542	0.0981	0.1042
(0.5, 0.9, 0.9)	0.1305	0.0733	0.2175	0.1267
(1, 0.9, 0.9)	0.3734	0.1218	0.4948	0.2009
(0, 0.9, 0.99)	0.0457	0.0544	0.0917	0.1050
(0.5, 0.9, 0.99)	0.1964	0.0733	0.3091	0.1275
(1, 0.9, 0.99)	0.5905	0.1226	0.7268	0.2009
(0, 0.9, 1)	0.0446	0.0551	0.0881	0.1042
(0.5, 0.9, 1)	0.2001	0.0730	0.3171	0.1275
(1, 0.9, 1)	0.5790	0.1229	0.7135	0.2002
(0, 1, 0.9)	0.0551	0.0548	0.1067	0.1047
(0.5, 1, 0.9)	0.1057	0.0711	0.1753	0.1271
(1, 1, 0.9)	0.2740	0.1231	0.3752	0.2005
(0, 1, 0.99)	0.0534	0.0540	0.1062	0.1044
(0.5, 1, 0.99)	0.5902	0.0716	0.6882	0.1255
(1, 1, 0.99)	0.9585	0.1226	0.9748	0.2016
(0, 1, 1)	0.0525	0.0546	0.1042	0.1062
(0.5, 1, 1)	0.7593	0.0718	0.8243	0.1244
(1, 1, 1)	0.9833	0.1230	0.9926	0.2020

REFERENCES

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

Chan, N.H. & C.Z. Wei (1987) Asymptotic inference for nearly nonstationary AR(1) processes. *Annals of Statistics* 15, 1050–1063.

Choi, I. (2005a) Inconsistency of bootstrap for nonstationary, vector autoregressive processes. *Statistics and Probability Letters* 75, 39–48.

Choi, I. (2005b) Subsampling vector autoregressive tests of linear constraints. *Journal of Econometrics* 124, 55–89.

Granger, C.W.J. (1969) Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37, 424–438.

Kurozumi, E. & T. Yamamoto (2000) Modified lag augmented vector autoregressions. *Econometric Review* 19, 207–231.

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

Phillips, P.C.B. (1995) Fully modified least squares and vector autoregression. *Econometrica* 63, 1023–1078.

Qin, J. & J. Lawless (1994) Empirical likelihood and general estimating functions. *Annals of Statistics* 22, 300–325.

Toda, H.Y. & P.C.B. Phillips (1993) Vector autoregressions and causality. *Econometrica* 61, 1367–1393.

https://doi.org/10.1017/S0266466613000339 Published online by Cambridge University Press

Toda, H.Y. & P.C.B. Phillips (1994) Vector autoregression and causality: A theoretical overview and simulation study. *Econometric Reviews* 13, 259–285.

Toda, H.Y. & Yamamoto (1995) Statistical inference in vector autoregressions with possibly integrated processes. *Journal of Econometrics* 66, 225–250.

Yamada, H. & H.Y. Toda (1998) Inference in possibly integrated vector autoregressive models: Finite sample evidence. *Journal of Econometrics* 86, 55–95.

APPENDIX: Proofs

Let $\theta = (\theta_1, \theta_2)^T$,

$$\theta_1 = \begin{cases} \sqrt{n}\beta_{11}, & \text{if} \quad \beta_{11} = 1 - \gamma_{11}/n, \\ \beta_{11}, & \text{if} \quad |\beta_{11}| < 1 \quad \text{is independent of } n, \end{cases}$$

$$\theta_2 = \begin{cases} \sqrt{n}\beta_{22}, & \text{if} \quad \beta_{22} = 1 - \gamma_{22}/n, \\ \beta_{22}, & \text{if} \quad |\beta_{22}| < 1 \quad \text{is independent of } n, \end{cases}$$

$\tilde{Z}_i(\theta) = Z_i\left((\theta_1, 0, 0, \theta_2)^T\right)$, $\tilde{l}_n(\theta) = l_n\left((\theta_1, 0, 0, \theta_2)^T\right)$ and denote the true value of θ by θ_0. Before proving Theorem 1, we need the following lemma.

LEMMA 1. *Under conditions of Theorem 1, with probability tending to one, $\tilde{l}_n(\theta)$ attains its minimum at some point $\tilde{\theta}$ in the interior of the ball $||\theta - \theta_0|| \le n^{-1/d_0}$ for some $d_0 \in (2, d)$, and $\tilde{\theta}$ and $\tilde{\lambda} = \lambda((\tilde{\theta}_1, 0, 0, \tilde{\theta}_2)^T)$ satisfy*

$$Q_{1n}(\tilde{\theta}, \tilde{\lambda}) = 0 \quad \textit{and} \quad Q_{2n}(\tilde{\theta}, \tilde{\lambda}) = 0, \tag{A.1}$$

where

$$Q_{1n}(\theta, \lambda) = \frac{1}{n}\sum_{i=1}^{n} \frac{\tilde{Z}_i(\theta)}{1 + \lambda^T \tilde{Z}_i(\theta)} \quad \textit{and}$$

$$Q_{2n}(\theta, \lambda) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{1 + \lambda^T \tilde{Z}_i(\theta)} \left\{\frac{\partial}{\partial \theta}\tilde{Z}_i(\theta)\right\}^T \lambda. \tag{A.2}$$

Proof. We only consider case (iv) in Theorem 1, i.e., both processes are nearly integrated; the other cases can be shown similarly. As $\beta_{12} = \beta_{21} = 0$, $\beta_{11} = 1 - \gamma_{11}/n$, and $\beta_{22} = 1 - \gamma_{22}/n$, it follows from Lemma 2.1 of Chan and Wei (1987) that

$$|Y_{n,1}| \xrightarrow{p} \infty, \quad \max_{1 \le i \le n} |Y_{i,1}| = O_p\left(n^{1/2}\right),$$

$$|Y_{n,2}| \xrightarrow{p} \infty, \quad \max_{1 \le i \le n} |Y_{i,2}| = O_p\left(n^{1/2}\right). \tag{A.3}$$

Write

$$\begin{aligned} \tilde{Z}_{i,1}(\theta) &= \frac{Y_{i-1,1}}{\left\{1 + Y_{i-1,1}^2\right\}^{3/4}}\{Y_{i,1} - \beta_{11}Y_{i-1,1}\} + W_{i,1}, \\ &= \frac{Y_{i-1,1}}{\left\{1 + Y_{i-1,1}^2\right\}^{3/4}}\epsilon_{i,1} + \frac{Y_{i-1,1}^2}{\left\{1 + Y_{i-1,1}^2\right\}^{3/4}}\left(\beta_{11}^0 - \beta_{11}\right) + W_{i,1}. \end{aligned}$$

https://doi.org/10.1017/S0266466613000339 Published online by Cambridge University Press

Note that we have $\theta_1 = \sqrt{n}\beta_{11}$ and $\theta_2 = \sqrt{n}\beta_{22}$. It is easy to verify that

$$\frac{1}{n}\sum_{i=1}^{n}\frac{Y_{i-1,1}}{\{1+Y_{i-1,1}^2\}^{3/4}}\epsilon_{i,1} = o_p\big(n^{-1/2}\big), \qquad \max_{1\le i\le n}\left|\frac{Y_{i-1,1}}{\{1+Y_{i-1,1}^2\}^{3/4}}\epsilon_{i,1}\right| = O_p\big(n^{1/d}\big),$$

$$\frac{1}{n}\sum_{i=1}^{n} W_{i,1} = O_p\big(n^{-1/2}\big), \qquad \max_{1\le i\le n}|W_{i,1}| = O_p\big(n^{1/d}\big),$$

$$\sup_{1\le i\le n}\sup_{||\theta-\theta_0||\le n^{-1/d_0}}\left|\frac{Y_{i-1,1}^2}{\{1+Y_{i-1,1}^2\}^{3/4}}\left(\beta_{11}^0-\beta_{11}\right)\right|$$

$$= O_p\left(n^{-1/2-1/d_0}\max_{1\le i\le n}|Y_{i-1,1}|^{1/2}\right) = o_p\big(n^{-1/d_0}\big),$$

which imply that

$$\sup_{||\theta-\theta_0||\le n^{-1/d_0}}\left|\frac{1}{n}\sum_{i=1}^{n}\tilde{Z}_{i,1}(\theta)\right| = o_p\big(n^{-1/d_0}\big) \quad \text{and}$$

$$\sup_{1\le i\le n}\sup_{||\theta-\theta_0||\le n^{-1/d_0}}|\tilde{Z}_{i,1}(\theta)| = O_p\big(n^{1/d}\big).$$

Using the same arguments, it can be shown that

$$\begin{cases} \sup_{||\theta-\theta_0||\le n^{-1/d_0}}\left\|\frac{1}{n}\sum_{i=1}^{n}\tilde{Z}_i(\theta)\right\| = O_p\big(n^{-1/d_0}\big), \\ \sup_{1\le i\le n}\sup_{||\theta-\theta_0||\le n^{-1/d_0}}\left\|\tilde{Z}_i(\theta)\right\| = O_p\big(n^{1/d}\big), \end{cases} \tag{A.4}$$

and

$$\sup_{||\theta-\theta_0||\le n^{-1/d_0},||x||=1}\frac{1}{n}\sum_{i=1}^{n}x^T\tilde{Z}_i(\theta)\tilde{Z}_i^T(\theta)x > 0 \tag{A.5}$$

with probability tending to one. Therefore, the lemma follows from (A.4), (A.5) and the same arguments used in proving Lemma 1 of Qin and Lawless (1994). ■

Proof of Theorem 1. We only consider case (iv), i.e., both processes are nearly integrated. Put $\sigma_1^2 = \mathbb{E}\epsilon_{1,1}^2$, $\sigma_2^2 = \mathbb{E}\epsilon_{1,2}^2$, $\sigma_{12} = \mathbb{E}\epsilon_{1,1}\epsilon_{1,2}$, $A_{n,1} = -n^{-3/2}\sum_{i=1}^{n}\left(Y_{i-1,2}/\sqrt{1+Y_{i-1,2}^2}\right)Y_{i-1,1}$, and $A_{n,2} = -n^{-3/2}\sum_{i=1}^{n}|Y_{i-1,2}|$. By (A.3), it is straightforward to show that

$$\frac{\partial Q_{1n}(\theta_0,0)}{\partial\theta} = \begin{pmatrix} 0 & 0 \\ A_{n,1} & 0 \\ 0 & 0 \\ 0 & A_{n,2} \end{pmatrix} + o_p(1)$$

$$= \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} A_{n,1} & 0 \\ 0 & A_{n,2} \end{pmatrix} + o_p(1) =: S_1S_2 + o_p(1),$$

https://doi.org/10.1017/S0266466613000339 Published online by Cambridge University Press

$$\frac{\partial}{\partial\lambda}Q_{1n}(\theta_0,0) = -\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \sigma_1^2 & 0 & \sigma_{12} \\ 0 & 0 & 1 & 0 \\ 0 & \sigma_{12} & 0 & \sigma_2^2 \end{pmatrix} + o_p(1)$$

$$= S_3 + o_p(1),$$

$$Q_{1n}(\theta_0,0) = O_p\left(n^{-1/2}\right), \quad \frac{\partial}{\partial\theta}Q_{1n}(\theta_0,0) = O_p(1),$$

$$\frac{\partial}{\partial\theta}Q_{2n}(\theta_0,0) = 0, \quad \frac{\partial}{\partial\lambda}Q_{2n}(\theta_0,0) = O_p(1).$$

Expanding $Q_{1n}(\tilde{\theta},\tilde{\lambda})$ and $Q_{2n}(\tilde{\theta},\tilde{\lambda})$ at $(\theta_0,0)$, we have

$$0 = Q_{1n}(\theta_0,0) + \frac{\partial Q_{1n}(\theta_0,0)}{\partial\theta}\left(\tilde{\theta}-\theta_0\right) + S_3\tilde{\lambda} + o_p\left(||\tilde{\lambda}|| + ||\tilde{\theta}-\theta_0||\right) \tag{A.6}$$

and

$$0 = Q_{2n}(\theta_0,0) + \frac{\partial Q_{2n}(\theta_0,0)}{\partial\theta}\left(\tilde{\theta}-\theta_0\right) + \frac{\partial Q_{2n}(\theta_0,0)}{\partial\lambda}\tilde{\lambda} + o_p\left(||\tilde{\lambda}|| + ||\tilde{\theta}-\theta_0||\right). \tag{A.7}$$

It follows from (A.6) and (A.7) that

$$\left\{\frac{\partial Q_{2n}(\theta_0,0)}{\partial\lambda}\right\}^T S_3^{-1}\frac{\partial Q_{1n}(\theta_0,0)}{\partial\theta}\sqrt{n}\left(\tilde{\theta}-\theta_0\right)$$

$$= -\left\{\frac{\partial Q_{2n}(\theta_0,0)}{\partial\lambda}\right\}^T S_3^{-1}\sqrt{n}Q_{1n}(\theta_0,0) + o_p(1) \tag{A.8}$$

and

$$\sqrt{n}\tilde{\lambda} = -S_3^{-1}\sqrt{n}Q_{1n}(\theta_0,0) - S_3^{-1}\frac{\partial Q_{1n}(\theta_0,0)}{\partial\theta}\sqrt{n}\left(\tilde{\theta}-\theta_0\right) + o_p(1). \tag{A.9}$$

Because

$$\frac{\partial}{\partial\lambda}Q_{2n}(\theta_0,0) = \left\{\frac{\partial}{\partial\theta}Q_{1n}(\theta_0,0)\right\}^T,$$

equation (A.8) is equivalent to

$$\left\{\frac{\partial Q_{1n}(\theta_0,0)}{\partial\theta}\right\}^T S_3^{-1}\frac{\partial Q_{1n}(\theta_0,0)}{\partial\theta}\sqrt{n}\left(\tilde{\theta}-\theta_0\right)$$

$$= -\left\{\frac{\partial Q_{1n}(\theta_0,0)}{\partial\theta}\right\}^T S_3^{-1}\sqrt{n}Q_{1n}(\theta_0,0) + o_p(1),$$

which implies

$$S_1^T S_3^{-1} S_1 S_2\sqrt{n}\left(\tilde{\theta}-\theta_0\right) = -S_1^T S_3^{-1}\sqrt{n}Q_{1n}(\theta_0,0) + o_p(1),$$

i.e.,

$$S_2\sqrt{n}\left(\tilde{\theta}-\theta_0\right) = -S^{-1}S_1^T S_3^{-1}\sqrt{n}Q_{1n}(\theta_0,0) + o_p(1), \tag{A.10}$$

https://doi.org/10.1017/S0266466613000339 Published online by Cambridge University Press

where $S = S_1^T S_3^{-1} S_1$. By (A.9) and (A.10), we have

$$\sqrt{n}\tilde{\lambda} = \left\{ S_3^{-1} S_1 S^{-1} S_1^T S_3^{-1} - S_3^{-1} \right\} \sqrt{n} Q_{1n}(\theta_0, 0) + o_p(1). \tag{A.11}$$

Using (A.9), (A.11), and Taylor expansion, we have

$$\begin{aligned}
l_n^P(0,0) &= \bar{l}_n(\tilde{\theta}) \\
&= 2\sum_{i=1}^{n} \tilde{\lambda}^T \tilde{Z}_i(\tilde{\theta}) - \sum_{i=1}^{n} \tilde{\lambda}^T \tilde{Z}_i(\tilde{\theta}) \tilde{Z}_i^T(\tilde{\theta}) \tilde{\lambda} + o_p(1) \\
&= 2n\tilde{\lambda}^T Q_{1n}(\tilde{\theta}, 0) + n\tilde{\lambda}^T S_3 \tilde{\lambda} + o_p(1) \\
&= 2n\tilde{\lambda}^T Q_{1n}(\theta_0, 0) + 2n\tilde{\lambda}^T \frac{\partial Q_{1n}(\theta_0, 0)}{\partial \theta} \left(\tilde{\theta} - \theta_0\right) + n\tilde{\lambda}^T S_3 \tilde{\lambda} + o_p(1) \\
&= -2n\tilde{\lambda}^T S_3 \tilde{\lambda} + n\tilde{\lambda}^T S_3 \tilde{\lambda} + o_p(1) \\
&= -n\tilde{\lambda}^T S_3 \tilde{\lambda} + o_p(1) \\
&= \sqrt{n}\left\{ S_3^{-1/2} Q_{1n}(\theta_0, 0) \right\}^T \left\{ I_4 - S_3^{-1/2} S_1 S^{-1} S_1^T S_3^{-1/2} \right\} \\
&\quad \times \sqrt{n}\left\{ S_3^{-1/2} Q_{1n}(\theta_0, 0) \right\} + o_p(1),
\end{aligned}$$

where I_4 denotes the 4×4 identity matrix. Thus, the theorem follows from the facts that

$$\left\{ I_4 - S_3^{-1/2} S_1 S^{-1} S_1^T S_3^{-1/2} \right\}^2 = \left\{ I_4 - S_3^{-1/2} S_1 S^{-1} S_1^T S_3^{-1/2} \right\},$$

$$\operatorname{tr}\left(I_4 - S_3^{-1/2} S_1 S^{-1} S_1^T S_3^{-1/2} \right) = 4 - \operatorname{tr}\left(S^{-1} S_1^T S_3^{-1} S_1 \right) = 4 - \operatorname{tr}(I_2) = 2,$$

and

$$\sqrt{n} Q_{1n}(\theta_0, 0) \xrightarrow{d} N((0,0,0,0)^T, S_3). \qquad \blacksquare$$

Proof of Theorem 2. We only prove the statement (v) with $\beta_{22} = 1 - \gamma_{22}/n$, because the other cases can be shown in the same way.

It follows from Lemma 2.1 of Chan and Wei (1987) that there exist standard Brownian motions $B_1(t)$ and $B_2(t)$ such that

$$\begin{cases}
n^{-1/2} \beta_{11}^{[nt]-n} Y_{[nt],1} \xrightarrow{D} B_1(e^{-2\gamma_{11}} (e^{2t\gamma_{11}} - 1)/(2\gamma_{11})) \\
n^{-1/2} \beta_{22}^{[nt]-n} \sum_{j=1}^{[nt]} \beta_{22}^{j-1} \epsilon_{i-j+1,1} \xrightarrow{D} B_2\left(e^{-2\gamma_{22}} \left(e^{2t\gamma_{22}} - 1\right)/\left(2\gamma_{22}\right)\right)
\end{cases} \tag{A.12}$$

in $D[0,1]$. Note that the second equation in (1) gives

$$Y_{i,2} = \beta_{21} \sum_{j=1}^{i} \beta_{22}^{j-1} Y_{i-j,1} + \sum_{j=1}^{i} \beta_{22}^{j-1} \epsilon_{i-j+1,2}. \tag{A.13}$$

Hence, by (A.12) and (A.13) we have

$$\begin{aligned}
&|Y_{n,1}| \xrightarrow{p} \infty, \qquad \max_{1 \le i \le n} |Y_{i,1}| = O_p\left(n^{1/2}\right), \\
&|Y_{n,2}| \xrightarrow{p} \infty, \qquad \max_{1 \le i \le n} |Y_{i,2}| = O_p\left(n^{3/2}\right).
\end{aligned} \tag{A.14}$$

https://doi.org/10.1017/S0266466613000339 Published online by Cambridge University Press

By (A.14), the weak law of large numbers, and the central limit theorem for martingales, we have

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\bar{Z}_i\left(\bar{\beta}^0\right)\xrightarrow{d}N\left((0,0)^T,\begin{pmatrix}1 & 0\\ 0 & \mathbb{E}\epsilon_{1,1}^2\end{pmatrix}\right) \tag{A.15}$$

and

$$\frac{1}{n}\sum_{i=1}^{n}\bar{Z}_i\left(\bar{\beta}^0\right)\bar{Z}_i^T\left(\bar{\beta}^0\right)\xrightarrow{p}\begin{pmatrix}1 & 0\\ 0 & \mathbb{E}\epsilon_{1,1}^2\end{pmatrix}, \tag{A.16}$$

where $\bar{\beta}_0=(\beta_{11}^0,0)^T$ and β_{11}^0 denotes the true value of β_{11}. Put $\theta=\sqrt{n}\beta_{11}$ and $\theta^0=\sqrt{n}\beta_{11}^0$. By (A.14)–(A.16) and the same arguments as in the proof of Lemma 1, we can show that, with probability tending to one, $\bar{l}_n\left((\theta,0)^T\right)$ attains its minimum at some point $\bar{\theta}$ such that $|\bar{\theta}-\theta^0|\le n^{-1/d_0}$ for some $d_0\in(2,d)$, and $\bar{\theta}$ and $\bar{\lambda}=\bar{\lambda}(\bar{\theta})$ satisfy

$$\bar{Q}_{1n}(\bar{\theta},\bar{\lambda})=0\quad\text{and}\quad\bar{Q}_{2n}(\bar{\theta},\bar{\lambda})=0, \tag{A.17}$$

where

$$\begin{cases}\bar{Q}_{1n}(\theta,\lambda)=\frac{1}{n}\sum_{i=1}^{n}\frac{\bar{Z}_i\left((\theta,0)^T\right)}{1+\lambda\bar{Z}_i\left((\theta,0)^T\right)}\\ \bar{Q}_{2n}(\theta,\lambda)=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{1+\lambda^T\bar{Z}_i\left((\theta,0)^T\right)}\left\{\frac{\partial}{\partial\theta}\bar{Z}_i\left((\theta,0)^T\right)\right\}^T\lambda.\end{cases} \tag{A.18}$$

Put

$$A_n=-n^{-3/2}\sum_{i=1}^{n}\frac{Y_{i-1,2}Y_{i-1,1}}{\sqrt{1+Y_{i-1,2}^2}},\qquad\bar{S}_1=\begin{pmatrix}0\\ A_n\end{pmatrix},\quad\text{and}\quad\bar{S}_2=\begin{pmatrix}1 & 0\\ 0 & \mathbb{E}\epsilon_{1,1}^2\end{pmatrix}.$$

It follows from (A.14)–(A.16) that

$$\begin{cases}\frac{\partial\bar{Q}_{1n}(\theta^0,0)}{\partial\theta}=\bar{S}_1+o_p(1),\quad\frac{\partial\bar{Q}_{1n}(\theta^0,0)}{\partial\lambda}=-\bar{S}_2+o_p(1),\\ \frac{\partial\bar{Q}_{2n}(\theta^0,0)}{\partial\theta}=0,\quad\frac{\partial\bar{Q}_{2n}(\theta^0,0)}{\partial\lambda}=\bar{S}_1^T+o_p(1).\end{cases}$$

Use the same arguments in deriving (A.11), we can show that

$$\sqrt{n}\bar{\lambda}=\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}\sqrt{n}\bar{Q}_{1n}(\theta^0,0)+o_p(1).$$

Hence it follows from Taylor expansion, (A.15) and (A.16) that, as $n\to\infty$

$$\begin{aligned}\bar{l}_n^P(0)&=2\sum_{i=1}^{n}\bar{\lambda}^T\bar{Z}_i\left(\bar{\beta}^0\right)-\sum_{i=1}^{n}\bar{\lambda}^T\bar{Z}_i\left(\bar{\beta}^0\right)\bar{Z}_i^T\left(\bar{\beta}^0\right)\bar{\lambda}+o_p(1)\\ &=2n\bar{\lambda}^T\bar{Q}_{1n}(\theta^0,0)-n\bar{\lambda}^T\bar{S}_2\bar{\lambda}+o_p(1)\\ &=\sqrt{n}\bar{Q}_{n1}^T(\theta^0,0)\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}\sqrt{n}\bar{Q}_{n1}(\theta^0,0)+o_p(1)\\ &=\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\bar{Z}_{i,1}(\bar{\beta}^0)\right\}^2+o_p(1)\\ &\xrightarrow{d}\chi^2(1).\end{aligned}$$

■

https://doi.org/10.1017/S0266466613000339 Published online by Cambridge University Press