Quasi maximum likelihood, indirect inference and bias-corrected pooled least squares estimators for dynamic panels with short T *

In Choi[†] and Sanghyun Jung[‡]

July 30, 2020

Abstract

This paper proposes new estimators for panel autoregressive (PAR) models with short time dimensions (T) and large cross sections. These estimators are based on the cross-sectional regression model using the first time series observations as a regressor and the last as a dependent variable. The regressors and errors of this regression model are correlated. The first estimator is the quasi maximum likelihood estimator (QMLE). The second estimator is the indirect inference estimator (IIE) that attempts to remove the finite-sample bias of the QMLE. The third estimator is the bias-corrected pooled least squares estimator (BCPLSE) that eliminates the asymptotic bias of the pooled least squares estimator by using the QMLE. The QMLE, IIE and BCPLSE are extended to the PAR model with endogenous regressors. The QMLE, IIE and BCPLSE provide consistent estimates of the PAR coefficients for stationary, unit root and explosive PAR models, estimate the coefficients of endogenous regressors consistently and can be computed as long as $T \geq 2$. Their finite sample properties are compared with those of some other estimators for the PAR model of order 1. The estimators of this paper are shown to perform quite well in finite samples.

Keywords: dynamic panels, quasi maximum likelihood estimator, indirect inference estimator, pooled least squares estimator, stationarity, unit root, explosive root

*This paper is an expanded and revised version of the first author's paper entitled as "Cross-sectional maximum likelihood and bias-corrected pooled least squares estimators for dynamic panels with short T." Some results in this paper were presented at the 22nd Panel Data conference held in Perth, Australia in 2016. The author thanks the participants, especially Stéphanne Bonhomme, for helpful comments. This paper was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2015S1A5A2A01009333). We thank Joakim Westerlund and two anonymous referees for constructive comments on this paper.

†Sogang University, Seoul, Korea. Personal Webpage: http://inchoi.sogang.ac.kr. E-mail: inchoi@gmail.com, inchoi@sogang.ac.kr.

‡Sogang University, Seoul, Korea. Personal Webpage: https://sites.google.com/view/sanghyunjung. E-mail: sanghnjung@gmail.com

1 Introduction

Panel autoregressive (PAR) models have been the focus of much research in recent years. When the number of time series observations (T) is small, the PAR model of order 1 (PAR(1) model) is often used in practice and, accordingly, much emphasis has been given to the PAR(1) model. There are several approaches to the estimation of the PAR(1) model. The most popular approach uses generalized methods of moments (GMM) estimators. These estimators improve on Anderson and Hsiao's (1981) instrumental variables (IV) estimator in terms of efficiency. The improved efficiency stems from additional moment conditions these estimators employ. Most notable papers studying the GMM estimators for the PAR(1) model are Arellano and Bond (1991), Ahn and Schmidt (1995) and Blundell and Bond (1998). But when the PAR(1) coefficient is close to unity, Arellano and Bond's and Ahn and Schmidt's GMM estimators are subject to the problem of weak instruments as analyzed in Blundell and Bond. Blundell and Bond's estimator does not share the same problem, is known to be more efficient than the other two, and has been used widely in applications.[1] Hahn (1999) shows that the efficiency gain of Blundell and Bond's estimator comes from the initial condition they use. In addition, Hahn, Hausman and Kuersteiner (2007) introduce a long difference regression model for the PAR(1) process and devise a GMM estimator. They report via simulation that it is sometimes more efficient than Blundell and Bond's GMM estimator. Ashley and Sun (2016) improve the GMM estimators for the case of stationarity using Hansen, Heaton and Yaron's (1996) continuous-updating method. Baltagi (2008) and Bun and Sarafidis (2015) provide nice reviews on the literature related to the GMM estimators.

The second approach to the estimation of the PAR model is the maximum likelihood estimation in differences. As alternatives to GMM estimation, Hsiao, Pesaran and Tahmiscioglu (2002) and Kruiniger (2008, 2013) have proposed maximum likelihood estimators (MLEs) using the differenced model, which are called the first difference MLEs (FDMLEs). Differencing eliminates individual effects and a unit root, if any, so that standard asymptotic theories are applicable. But because the FDMLE employs the differenced PAR(1) model, it becomes impossible to estimate the coefficients of time-invariant regressors.

The third approach employs least squares estimators (LSEs). Han and Phillips (2010) and Han, Phillips and Sul (2014) introduce transformations of the PAR(1) model that make the regressors and errors uncorrelated. Then, they apply LSEs to estimate the PAR(1) coefficient. Being able to use LSEs is a convenient aspect of this approach, but validity of their transformations depends on their assumptions on the model and may not hold under different assumptions (see Supplement 1 to this paper for more discussion). Hahn and Kuersteiner (2002) propose a bias-corrected Within-OLS estimator of the PAR(1) coefficient. Gourieroux, Phillips and Yu (2010) also introduce a bias-corrected Within-OLS estimator using the indirect inference method of Gourieroux, Monfort and Renault (1993). In Hahn and Kuersteiner's and

[1] But they can also be subject to the weak instrument problem as reported in Hayakawa (2007) and Bun and Windmeijer (2010).

Gourieroux, Phillips and Yu's approaches, it is required that $T \to \infty$ and that the PAR(1) coefficient takes values less than 1, which may limit their use in practice.

The fourth approach is the maximum likelihood estimation in level suggested by Anderson and Hsiao (1982) and Hsiao and Zhou (2018). They assume exogenous regressors. How to extend Anderson and Hsiao's approach to the case of endogenous regressors does not seem to be obvious. In addition, they do not consider the case where the PAR(1) coefficient takes values greater than or equal to 1, and the type of the initial variable considered in this paper. See also Ahn and Thomas (2006) and Alvarez and Arellano (2004) for the maximum likelihood estimation in level.

The other approaches using the level data include Bun and Carree (2005, 2006), Bai (2013) and Hayakawa (2012). Bun and Carree propose bias-corrected within-OLS. Bai introduces MLE utilizing the factor structure of the PAR(1) model. Hayakawa suggests a GMM estimator. All of them assume a stable PAR(1) coefficient.

The purpose of this paper is to suggest a new approach for the estimation of the PAR(1) model. This approach employs the cross-sectional regression model using the first time series observations as a regressor and the last as a dependent variable. Because the initial observation and the individual effect are correlated, the regressors and errors of the regression model are dependent, making LSEs inapplicable. Instead, a quasi maximum likelihood estimator (QMLE) is proposed for the regression model. The estimator is called the cross-sectional QMLE.

MLEs are often subject to finite-sample biases and, so is the QMLE. Thus, we propose to employ the indirect inference estimator (IIE) of Smith (1993) and Gouriéroux, Monfort and Renault (1993) in conjunction with the QMLE. Gouriéroux, Phillips and Yu (2010) also employ the indirect inference estimation method for dynamic panels and report good finite sample properties of their estimator. But the parameter space of the PAR(1) coefficient is restricted to (-1,1) in their paper. The IIE improves on the QMLE according to our simulation results.

The QMLE makes it feasible to estimate the asymptotic bias of the pooled least squares estimator for the dynamic panel regression model. The estimated bias can be used to construct a bias-corrected pooled LSE that will be called the bias-corrected pooled least squares estimator (BCPLSE).

Asymptotic properties of the QMLE, IIE and BCPLSE are studied in this paper. The estimators are also extended to the PAR(1) model with endogenous regressors. In addition, their finite sample properties are compared with those of 7 other estimators. It is found that the new estimators of this paper perform quite well compared to the other estimators.

There are some advantages of this paper's approach. First, there are no restrictions on the parameter space of the PAR(1) coefficient. It can be any compact subset of the real line. By contrast, most papers require the parameter space to be either (-1,1) or (-1,1]. Moreover, it is possible to estimate the PAR(1) coefficient consistently by using the QMLE, IIE and BCPLSE even when it is greater than 1. This explosive case is potentially important for applications to financial data. For example, bubbles of financial markets are modelled using an explosive, univariate AR process in Phillips, Wu and Yu (2011).

Second, coefficients of time-invariant, endogenous regressors can be estimated along with other parameters. Because most of the procedures discussed so far use differencing or within-group demeaning, it is impossible to estimate the coefficients of time-invariant regressors.

Third, the estimators of this paper require $T \geq 2$, while most of the aforementioned estimators require at least 3 or 4 time series observations. For new panel data sets, this is an important advantage.

This paper is planned as follows. Section 2 introduces the model and basic assumptions. Section 3 introduces the QMLE and studies its asymptotic properties. Section 4 studies the IIE. Section 5 proposes the BCPLSE and studies its asymptotic properties. Section 6 extends the QMLE, IIE and BCPLSE to the PAR(1) model with endogenous regressors. Section 7 reports simulation results. Section 8 provides summary and further remarks. Proofs are relegated to Appendix.

A few words on our notation. $\mathbb{R}$ and $\mathbb{R}^+$ denote the set of real numbers and the set of positive real numbers, respectively. All the limits are taken as $N \to \infty$. Convergence in probability and weak convergence are denoted by $\xrightarrow{p}$ and $\xrightarrow{d}$, respectively.

2 The model and basic assumptions

We are concerned with the unobserved components model for the panel data $\{y_{it}\}$

$$y_{it} = \mu_i + x_{it},\ (i = 1, \ldots, N;\ t = 1, \ldots, T;\ T \geq 2), \tag{1}$$

where $\{x_{it}\}$ is unobserved and follows the AR(1) model

$$x_{it} = \alpha x_{i,t-1} + u_{it}. \tag{2}$$

As usual, i and t are indices for individuals and time, respectively, and $\{\mu_i\}$ denotes the unobserved individual effects. In Model (1), $\{x_{it}\}$ brings dynamics to the evolvement of $\{y_{it}\}$. Model (1) can be written in a more familiar format as

$$y_{it} = \mu_i(1 - \alpha) + \alpha y_{i,t-1} + u_{it}. \tag{3}$$

This model has been used in various works for the estimation of dynamic panels. The reader is referred to Hsiao (2003, p.76) for the comparison of Models (1) and (3).

Regarding the individual effects variable μ_i, let

$$\mu_i = \mu + m_i,$$

where μ is a fixed constant and m_i is a random variable. Using this relation, Model (3) can be written as

$$y_{it} = \mu(1 - \alpha) + \alpha y_{i,t-1} + u_{it} + m_i(1 - \alpha) \tag{4}$$

The following assumption introduces the basic characteristics of the individual effects $\{m_i\}$, the error terms $\{u_{it}\}$ and the initial variables $\{x_{i1}\}$.

Assumption 1 *Let* $u_i = [u_{i2}, ..., u_{iT}]'$. *Assume*

$$\begin{pmatrix} m_i \\ u_i \\ x_{i1} \end{pmatrix} \sim \text{i. i. d.} \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{\text{m}} & 0 & 0 \\ 0 & \sigma^2_{\text{u}} I_{T-1} & 0 \\ 0 & 0 & \sigma^2_{\text{x}_1} \end{bmatrix} \right),$$

where $\sigma^2_{\text{m}} > 0$, $\sigma^2_{\text{u}} > 0$ *and* $\sigma^2_{\text{x}_1} > 0$.

Under this assumption, the variance of x_{i1} is not assumed to be equal to $\sigma^2_{\text{u}}/(1-\alpha^2)$, which is often assumed in the literature on time series and dynamic panels when $\alpha \in (-1, 1)$. In this paper, we just assume that parameter α belongs to a set of real numbers. It is even allowed to be greater than 1. Under Assumption 1, note also that the initial observation y_{i1} and the individual effect variable μ_i are correlated. In fact, $Cov(y_{i1}, \mu_i) = \sigma^2_{\text{m}}$.

Under Assumption 1, running OLS or Within-OLS on Model (4) does not provide a consistent estimator of the parameter α because $Cov(y_{i,t-1}, u_{it} + m_i(1-\alpha)) = Cov(m_i + x_{i,t-1}, u_{it} + m_i(1-\alpha)) = (1-\alpha)\sigma^2_{\text{m}}$ is not zero unless α is equal to one.

3 Cross-sectional quasi maximum likelihood estimation

This section introduces a cross-sectional quasi maximum likelihood estimator (QMLE) for Model (1). Model (1) can be written as

$$y_{it} = \mu + \alpha^{t-1} x_{i1} + w_{it} + m_i, \ (t = 2, ..., T), \tag{5}$$

where $w_{it} = \sum_{j=0}^{t-2} \alpha^j u_{i,t-j}$. Assume $\mu = 0$ for simplicity from now on.[2] Because $x_{i1} = y_{i1} - \mu_i$, relation (5) gives[3] for $t = T$

$$\begin{aligned} y_{iT} &= \alpha^{T-1} x_{i1} + w_{iT} + m_i \\ &= \alpha^{T-1} y_{i1} + w_{iT} + (1-\alpha^{T-1}) m_i. \end{aligned} \tag{6}$$

Running OLS on this equation does not yield a consistent estimator of the regression coefficients because y_{i1} and m_i are correlated unless $\alpha = 1$. As an alternative, we consider quasi maximum likelihood estimation in this section. For the quasi maximum likelihood estimation, it is required that $T \geq 2$. Even at $T = 2$, the QMLE of α can be obtained. This feature is not shared with extant estimation methods in dynamic panel data analysis.

Let $v_{iT} = w_{iT} + (1-\alpha^{T-1}) m_i$. Then, Assumption 1 gives

$$\begin{pmatrix} v_{iT} \\ y_{i1} \end{pmatrix} \sim \text{i. i. d.} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \omega_{11} & \omega_{12} \\ \omega_{12} & \omega_{22} \end{bmatrix} \right), \tag{7}$$

[2] In practice, we can demean $\{y_{iT}\}$ and $\{y_{1i}\}$ to make this assumption plausible.

[3] Model (6) is also used in Choi (2019) to devise unit root tests for dependent and heterogeneous micro-panels.

where

$$\omega_{11} = \sigma_{\rm u}^2 \sum_{j=0}^{T-2} \alpha^{2j} + (1-\alpha^{T-1})^2 \sigma_{\rm m}^2,\ \omega_{12} = (1-\alpha^{T-1})\sigma_{\rm m}^2 \text{ and } \omega_{22} = \sigma_{\rm x_1}^2 + \sigma_{\rm m}^2.$$

Suppose that we project $\mathbf{v}_T = [v_{1T}, ..., v_{NT}]'$ onto the space spanned by $\mathbf{y}_1 = [y_{11}, ..., y_{N1}]'$ such that

$$\mathbf{v}_T = \mathbf{y}_1(\mathbf{y}_1'\mathbf{y}_1)^{-1}\mathbf{y}_1'\mathbf{v}_T + \left(I - \mathbf{y}_1(\mathbf{y}_1'\mathbf{y}_1)^{-1}\mathbf{y}_1'\right)\mathbf{v}_T.$$

Replacing $(\mathbf{y}_1'\mathbf{y}_1)^{-1}\mathbf{y}_1'\mathbf{v}_T$ with its probability limit $\frac{\omega_{12}}{\omega_{22}}$, this relation gives

$$v_{iT} = \frac{\omega_{12}}{\omega_{22}} y_{i1} + r_{iT}.$$

Relation (7) gives

$$\begin{pmatrix} r_{iT} \\ y_{i1} \end{pmatrix} \sim \text{i. i. d.} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \omega_{11.2} & 0 \\ 0 & \omega_{22} \end{bmatrix} \right), \tag{8}$$

where $\omega_{11.2} = \omega_{11} - \frac{\omega_{12}^2}{\omega_{22}}$.

Let $\xi = (\alpha, \sigma_{\rm m}^2, \sigma_{\rm u}^2, \sigma_{\rm x_1}^2)'$, $z_i = (y_{iT}, y_{i1})'$, $\Sigma_{zz} = E(z_1 z_1')$ and $M_{zz} = \frac{1}{n}\sum_{i=1}^n z_i z_i'$. The objective function we adopt[4] is

$$\begin{aligned} L(\xi) &= -\ln\det(\Sigma_{zz}) - tr(\Sigma_{zz}^{-1} M_{zz}) \\ &= -\ln(\omega_{11.2}) - \ln(\omega_{22}) \\ &\quad -\omega_{11.2}^{-1}\left(M_{TT} - 2\chi M_{T1} + \chi^2 M_{11}\right) - \omega_{22}^{-1} M_{11}, \end{aligned} \tag{9}$$

where $\chi = \alpha^{T-1} + \frac{\omega_{12}}{\omega_{22}}$, $M_{TT} = \frac{1}{N}\sum_{i=1}^N y_{iT}^2$, $M_{T1} = \frac{1}{N}\sum_{i=1}^N y_{iT} y_{i1}$ and $M_{11} = \frac{1}{N}\sum_{i=1}^N y_{i1}^2$. This is the likelihood function under the assumption of normality for $\{z_i\}$. But we do not assume normality for the data. Let Ξ be any compact subset of $\{v = (v_\alpha, v_m, v_u, v_x)' \mid v \in \mathbb{R}\times\mathbb{R}^+\times\mathbb{R}^+\times\mathbb{R}^+\}$, where the partition of v is conformable to that of ξ. The QMLE of ξ is defined as

$$\hat{\xi} = \arg\max_{\xi\in\Xi} l(\xi).$$

The following theorem reports the consistency properties of the QMLE. The true value of any parameter a is denoted as a^*.

Theorem 1 *Suppose that Assumption 1 holds.*
(i) If $\alpha^* \neq 1$, $\hat{\xi} \xrightarrow{p} \xi^*$.

[4] In the previous versions of this paper, we used the objective function $l^{\blacklozenge}(\xi) = -\ln(\omega_{11.2}) - \frac{1}{N}\sum_{i=1}^N \omega_{11.2}^{-1}(y_{iT} - \alpha^{T-1}y_{i1} - \frac{\omega_{12}}{\omega_{22}} y_{i1})^2$. This is related to $l(\xi)$ as follows: $l^{\blacklozenge}(\xi) = l(\xi) + \ln(\omega_{22}) + \omega_{22}^{-1} M_{11}$. The objective function $l(\xi)$ provides estimators with better finite sample properties according to our simulation results.

(ii) If $\alpha^* = 1$, $\hat{\alpha} \xrightarrow{p} 1$, $\hat{\sigma}^2_{\mathrm{u}} \xrightarrow{p} \sigma^{*2}_{\mathrm{u}}$ *and* $\hat{\sigma}^2_{\mathrm{x}_1} + \hat{\sigma}^2_{\mathrm{m}} \xrightarrow{p} \sigma^{*2}_{\mathrm{x}_1} + \sigma^{*2}_{\mathrm{m}}$.

If $\alpha^* \neq 1$, $\hat{\xi}$ is consistent for ξ^*. But if $\alpha^* = 1$, only α^*, σ^{*2}_{u} and $\sigma^{*2}_{\mathrm{x}_1} + \sigma^{*2}_{\mathrm{m}}$ can be estimated consistently. This does not affect asymptotic properties of BCPLSE of Section 5 because the bias correction term becomes $o_p(1)$ when $\alpha^* = 1$. Gourieroux, Monfort and Trognon (1984) report that QMLEs are consistent under certain conditions when the objective function is formulated by using a density belonging to linear exponential family. The result reported in this theorem accords well with their finding, although their estimator is a conditional QMLE and our method of proof is different from theirs. The method of our proof was inspired by Bai and Li (2012).

One may hope that the limiting distribution of the QMLE $\hat{\xi}$ can be derived using the usual method relying on the first and second derivatives of the objective function. Against this hope, the first derivatives are linearly related,[5] which results in a singular information matrix. Thus, the conventional method will only yield the limiting distributions of some linear combinations of the QMLEs.

It is most natural to maximize the objective function (9) with respect to the 4 parameters to obtain their QMLEs. However, we will reparametrize the objective function using some information from method-of-moments estimators. According to our experiments, the QMLE of α using the method-of-moments estimators have better finite sample performance than those based on the full maximization of the objective function (9). To reduce the number of parameters in the objective function (9), assume that $\omega_{22} = \sigma^2_{\mathrm{x}_1} + \sigma^2_{\mathrm{m}}$ is known. The parameter ω_{22} can be estimated consistently by $\frac{1}{N}\sum_{i=1}^{N} y_{i1}^2$. Next, assume $\lambda = Var(\Delta y_{i2}) = (1-\alpha)^2\sigma^2_{\mathrm{x}_1} + \sigma^2_{\mathrm{u}}$ is known.[6] In practice, λ can be estimated consistently by $\frac{1}{N}\sum_{i=1}^{N} (\Delta y_{i2})^2$. Then, because $\sigma^2_{\mathrm{u}} = \lambda - (1-\alpha)^2(\omega_{22} - \sigma^2_{\mathrm{m}})$ and $\sigma^2_{\mathrm{x}_1} = \omega_{22} - \sigma^2_{\mathrm{m}}$, there remain only α and σ^2_{m} in the objective function. Now, we estimate σ^2_{m} by maximizing the objective function (9). Then, plug the QMLE of σ^{*2}_{m} and the method-of moments estimators of σ^2_{u} and $\sigma^2_{\mathrm{x}_1}$ into the objective function (9) and maximize it with respect to α. The resulting QMLE of α^* is consistent as can be deduced from the proof of Theorem 1. We will use this QMLE in Sections 4 and 7.

The QMLE considered so far can be extended to the case of time-series heteroskedasticity. To this ends, let $E(u_{it}^2) = \sigma^2_{\mathrm{u,t}}$ for all i and assume

$$\begin{pmatrix} m_i \\ u_i \\ x_{i1} \end{pmatrix} \sim \mathrm{i.i.d.} \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{\mathrm{m}} & 0 & 0 \\ 0 & \Gamma & 0 \\ 0 & 0 & \sigma^2_{\mathrm{x}_1} \end{bmatrix} \right),$$

where $\Gamma = diag(\sigma^2_{\mathrm{u},2}, ..., \sigma^2_{\mathrm{u,T}})$. The objective function is constructed in a similar manner except that $\omega_{11} = \sum_{j=0}^{T-2} \alpha^{2j}\sigma^2_{\mathrm{u,T-j}} + (1-\alpha^{T-1})^2\sigma^2_{\mathrm{m}}$. Using the same methods as for Theorem 1, it is straightforward to show that the QMLE is consistent. The details are not deemed to be worth reporting here.

The nuisance parameters in the objective function are eliminated similarly. In

[5] The relationship is reported in Supplement 2 to this paper.

[6] One may wonder immediately why $\sum_{t=2}^{T} Var(\Delta y_{ti})$ is not used instead. But it turns out that this brings higher asymptotic variance of the MLE of α, so we chose to use $Var(\Delta y_{2i})$.

order to eliminate $\sigma^2_{\mathrm{u},2},...,\sigma^2_{\mathrm{u,T}}$, let

$$\begin{aligned}
\lambda_2 &= Var(\Delta y_{i2}) = (\alpha-1)^2\sigma^2_{\mathrm{x}_1}+\sigma^2_{\mathrm{u},2} \\
\lambda_3 &= Var(\Delta y_{i3}) = (\alpha^2-\alpha)^2\sigma^2_{\mathrm{x}_1}+\sigma^2_{\mathrm{u},3}+(\alpha-1)^2\sigma^2_{\mathrm{u},2} \\
&\vdots \\
\lambda_T &= Var(\Delta y_{iT}) = (\alpha^{T-1}-\alpha^{T-2})^2\sigma^2_{\mathrm{x}_1}+\sigma^2_{\mathrm{u,T}}+(\alpha-1)^2\sigma^2_{\mathrm{u,T-1}}+...+(\alpha^{T-2}-\alpha^{T-3})^2\sigma^2_{\mathrm{u},2}.
\end{aligned}$$

Since $\lambda_2,...,\lambda_T$ are estimable, we assume than they are known. Then, $\sigma^2_{\mathrm{u},2},...,\sigma^2_{\mathrm{u,T}}$ can be written as functions of only α and σ^2_{m} $(=\omega_{22}-\sigma^2_{\mathrm{x}_1})$. Thus, the objective function can be written as a function of only α and σ^2_{m}, and the QMLE of α^* can be calculated as before.

We have used only $\{y_{i1}\}$ and $\{y_{iT}\}$ to formulate the QMLE. As one of the reviewers on this paper suggested, we may consider choosing any two of the cross-sectional data $\{y_{i1}\}$, $\{y_{i2}\}$, ..., $\{y_{iT}\}$ to calculate the QMLE. According to small-scale, unreported simulation results for $T=4$, we find that using $\{y_{i1}\}$ and $\{y_{i4}\}$ tends to give the best results among the six combinations and that the sample mean of the six QMLEs provides the best results without exception. Although we could not use the sample mean of the QMLEs for the simulation exercise of this paper due to its enormous computational need, we advise using it when dealing with real data in practice.

4 Indirect inference estimation

The QMLE may be subject to bias (cf. Cox and Snell, 1968). Bias-corrections for the QMLE of the last section deserve a serious consideration for this reason. There have been various methods for bias-corrections for dynamic panels (e.g., Kiviet, 1995; Bun and Carree, 2005; Hahn and Kuersteiner, 2002; Gouriéroux, Phillips and Yu, 2010). In this section, we apply the principle of indirect inference estimation originally proposed by Smith (1993) and Gouriéroux, Monfort and Renault (1993) to Model (4) in conjunction with the QMLE of the last section. A distinct advantage of indirect inference is that it does not require explicit calculation of bias terms, which is quite complicated to obtain for the QMLE. Resampling methods such as jackknife and bootstrapping have the same advantage, but are not considered here. Median-unbiased estimation also eliminates finite sample bias as shown in Andrews (1993), but it is not pursed here. Gouriéroux, Phillips and Yu (2010) also employ the indirect inference estimation method for dynamic panels, but the parameter space of the PAR(1) coefficient is restricted to (-1,1).

Suppose that we can generate simulated data by using Model (4) once the value of α is given, and denote the h-th simulated data as $\{\mathbf{y}^h_1(\alpha),...,\ \mathbf{y}^h_T(\alpha)\}$, where $\mathbf{y}^h_t(\alpha)$ denotes the vector of $\{y_{it}\}_{i=1,...,N}$ at given α. The details of data generation will be discussed later. Let $\hat{\alpha}^h(\alpha)$ be the QMLE of α using the h-th simulated data

$\{\mathbf{y}_1^h(\alpha), ..., \ \mathbf{y}_T^h(\alpha)\}$. The indirect inference estimator of α^* is defined by

$$\hat{\alpha}_{IIE} = \arg\min_{\alpha\in\Lambda} \left\| \hat{\alpha} - \frac{1}{H}\sum_{h=1}^{H} \hat{\alpha}^h(\alpha) \right\|,$$

where Λ, the parameter space of α, is a compact subset of $\mathbb{R}$ and $\|\cdot\| : \mathbb{R} \to [0, \infty)$ is a norm. In practice, the IIE can be obtained by solving the equation

$$\hat{\alpha} - \frac{1}{H}\sum_{h=1}^{H} \hat{\alpha}^h(\alpha) = 0 \tag{10}$$

for α. Letting $b_N^H(\alpha) = \frac{1}{H}\sum_{h=1}^{H} \hat{\alpha}^h(\alpha)$, assume

Assumption 2 $b_N^H(\alpha) : \Lambda \to \Lambda$ *is an one-to-one, continuous function of* α *for all* N *and* H.

A unique solution of equation (10) exists under this assumption. That is, $\hat{\alpha}_{IIE}^H = b_N^{H-1}(\hat{\alpha})$.

How can we know that the IIE alleviates the finite-sample bias problem of $\hat{\alpha}$ and what is its limiting distribution? To tackle these questions, assume $H = \infty$. Then, we have by the law of large numbers

$$b_N^\infty(\alpha) = E(\hat{\alpha}^1(\alpha)).$$

In addition, it follows that

$$b_N^\infty(\alpha^*) = E(\hat{\alpha}^1(\alpha^*)) = E(\hat{\alpha}) = E(b_N^\infty(\hat{\alpha}_{IIE}^\infty)),$$

where the last equality holds by the definition of $b_N^\infty(\cdot)$. The standard theory of MLE (cf. Cox and Snell, 1968) yields

$$E(\hat{\alpha}) = \alpha^* + N^{-1}c(\alpha^*).$$

Thus, is is reasonable to assume

Assumption 3 $b_N^\infty(\alpha) = \alpha + N^{-1}c(\alpha)$, *where* $c(\alpha) : \Lambda \to R$ *is a continuously differentiable function of* α.

Properties of $\hat{\alpha}_{IIE}^\infty$ are reported in the following theorem.

Theorem 2 *Strengthen Assumption 1 such that the random vector there is normally distributed, and suppose that the strengthened Assumption 1, Assumption 2 and Assumption 3 hold. Then, if* $\alpha^* \neq 1$,
(i) $\hat{\alpha}_{IIE}^\infty \xrightarrow{p} \alpha^*$.
(ii) $\sqrt{N}(\hat{\alpha}_{IIE}^\infty - \alpha^*)$ *has the same asymptotic distribution as* $\sqrt{N}(\hat{\alpha} - \alpha^*)$.
(iii) $E(\hat{\alpha}_{IIE}^\infty) = \alpha^* + E(\zeta_N)$, *where* $\zeta_N = O_p(N^{-3/2})$.

We need the assumption of normality because we generate normal numbers to calculate the IIE. Additionally, we exclude the case $\alpha^* = 1$ because σ_{m}^{*2} cannot be identified under that case. Note that estimates of this parameter are required to generate $\mathbf{y}_t^h(\alpha)$. This theorem shows that $\hat{\alpha}_{IIE}^{\infty}$ is consistent for α^* and has the same asymptotic distribution as $\hat{\alpha}$. Furthermore, while the bias of QMLE is $O(N^{-1})$, that of $\hat{\alpha}_{IIE}^{\infty}$ is the mean of an $O_p(N^{-3/2})$ random variable. We conjecture from this that $\hat{\alpha}_{IIE}^{\infty}$ is less biased than $\hat{\alpha}$, which is confirmed by simulation in Section 7.

Now, we discuss how $\hat{\alpha}^h(\alpha)$ is calculated using simulated data $\{\mathbf{y}_1^h(\alpha), ..., \mathbf{y}_T^h(\alpha)\}$.
Step 1: Estimate σ_{m}^{*2} and σ_{u}^{*2} using the observed data.
Step 2: Using a normal number generator along with the estimators from Step 1, generate $\{\mathbf{y}_t^h(\alpha)\}_{t=1,..,T}$ for a fixed value of α.
Step 3: Calculate $\hat{\alpha}^h(\alpha)$ using the simulated data $\{\mathbf{y}_1^h(\alpha)\}$ and $\{\mathbf{y}_T^h(\alpha)\}$ from Step 2 and the methods of Section 3.

5 Bias-corrected pooled least squares estimator

This section proposes a bias-corrected pooled least squares estimator (BCPLSE) of α^* for Model (4). We are interested in the PLSE rather than the least squares dummy variable estimator because time-invariant regressors are not eliminated by the PLSE procedure when it is extended to the model with time-invariant regressors.

The PLSE of α^* is defined as $\hat{\alpha}_{PLSE} = \frac{\sum_{i=1}^N \sum_{t=2}^T (y_{i,t-1}-\bar{y}_{-1})y_{it}}{\sum_{i=1}^N \sum_{t=2}^T (y_{i,t-1}-\bar{y}_{-1})^2}$, where $\bar{y}_{-1} = \frac{1}{N(T-1)} \sum_{i=1}^N \sum_{t=2}^T y_{i,t-1}$. Because

$$
\begin{aligned}
& \hat{\alpha}_{PLSE} - \alpha^* \\
= & \frac{\sum_{i=1}^N \sum_{t=2}^T (y_{i,t-1} - \bar{y}_{-1}) \left(u_{it} + m_i(1-\alpha^*)\right)}{\sum_{i=1}^N \sum_{t=2}^T (y_{i,t-1} - \bar{y}_{-1})^2} \\
= & \frac{\sum_{i=1}^N \sum_{t=2}^T \left\{(x_{i,t-1} - \bar{x}_{-1} + m_i - \bar{m})u_{it} + (x_{i,t-1} - \bar{x}_{-1})m_i(1-\alpha^*)\right\}}{\sum_{i=1}^N \sum_{t=2}^T (y_{i,t-1} - \bar{y}_{-1})^2} \\
& + \frac{(1-\alpha^*)(T-1) \sum_{i=1}^N (m_i - \bar{m})m_i}{\sum_{i=1}^N \sum_{t=2}^T (y_{i,t-1} - \bar{y}_{-1})^2}, \qquad (11)
\end{aligned}
$$

we have under Assumption 1,

$$\hat{\alpha}_{PLSE} - \alpha^* \xrightarrow{p} \frac{(1-\alpha^*)(T-1)\sigma_{\mathrm{m}}^2}{p\lim \frac{1}{N} \sum_{i=1}^N \sum_{t=2}^T (y_{i,t-1} - \bar{y}_{-1})^2}.$$

This means that $\hat{\alpha}_{PLSE}$ is inconsistent for α^* unless $\alpha^* = 1$. The BCPLSE using the QMLE is defined by

$$\hat{\alpha}_{BCPLSE} = \hat{\alpha}_{PLSE} - \frac{(1-\hat{\alpha})(T-1)\hat{\sigma}_{\mathrm{m}}^2}{\frac{1}{N} \sum_{i=1}^N \sum_{t=2}^T (y_{i,t-1} - \bar{y}_{-1})^2}. \qquad (12)$$

Asymptotic properties of the bias-corrected PLSE are reported in the following theorem.

Theorem 3 *Suppose that Assumption 1 holds. Then,*
(i) $\hat{\alpha}_{BCPLSE} \xrightarrow{p} \alpha^*$;

(ii) $\sqrt{N}(\hat{\alpha}_{BCPLSE} - \alpha^*) = O_p(1)$.

In practice, we may use BCPLSE to construct a two-stage BCPLSE using the formula (12) with $\hat{\alpha}$ being replaced by $\hat{\alpha}_{BCPLSE}$. The two-stage BCPLSE performs better than BCPLSE according to some unreported simulation results. Comparing PLSE and BCPLSE, we found by simulation that the latter has a smaller magnitude of bias and larger variance than the former in finite samples. Thus, the often-observed bias-variance trade-off is also occurring here.

Because the limiting distribution of $\hat{\alpha}_{BCPLSE}$ depends on that of a linear combination of $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(m_i^2 - \sigma_{\mathrm{m}}^{*2}\right)$, $\sqrt{N}\left(\hat{\sigma}_{\mathrm{m}}^2 - \sigma_{\mathrm{m}}^{*2}\right)$ and $\sqrt{N}(\hat{\alpha} - \alpha^*)$, it is difficult to derive the limiting distribution of $\hat{\alpha}_{BCPLSE}$. Under this circumstance, we can use bootstrapping for interval estimation of α^*. The interval estimation can also be used for point hypothesis testing in the usual way. The following steps provide bootstrap confidence intervals of α.

> **Step 1**: Let $\mathbf{y}_i = [y_{i1}, ..., y_{iT}]'$. Choose $\mathbf{y}_1^*, ..., \mathbf{y}_N^*$ randomly from $\{\mathbf{y}_i\}_{i=1,...,N}$ with replacements.
> **Step 2**: Calculate BCPLSE of α^* using $\{\mathbf{y}_i^*\}_{i=1,...,N}$, which is denoted as $\hat{\alpha}_{BCPLSE,b}^*$.
> **Step 3**: Repeat Steps 1 and 2 B times and record the values of $\{\sqrt{N}(\hat{\alpha}_{BCPLSE,b}^* - \hat{\alpha}_{BCPLSE})\}_{b=1,...,B}$.
> **Step 4:** Obtain the $\gamma/2$-th and $(1-\gamma/2)$-th percentiles of $\{\sqrt{N}(\hat{\alpha}_{BCPLSE,b}^* - \hat{\alpha}_{BCPLSE})\}_{b=1,...,B}$, which are denoted as $c_{\gamma/2}$ and $c_{(1-\gamma/2)}$.

The $(1-\gamma)\times 100$ percent bootstrap confidence interval for α^* is defined as $(\hat{\alpha}_{BCPLSE} - c_{\gamma/2}/\sqrt{N},\ \hat{\alpha}_{BCPLSE} - c_{(1-\gamma/2)}/\sqrt{N})$. Finite sample properties of the bootstrap confidence intervals will be studied in Section 7.

Alternatively, one may resample the data using time series residuals for each i as in the time series literature. According to some experimental results which are unreported here, the bootstrap procedure given above provides far better results in finite samples. A plausible reason for this is that the resampled data using the steps above mimic the original data better than those based on resampled residuals.

6 An extension to the dynamic AR(1) model with endogenous regressors

6.1 The model and assumptions

An extended version of Model (1) is

$$y_{it} = \mu_i + \gamma' p_i + x_{it},\ (i = 1, \ldots, N;\ t = 1, \ldots, T), \tag{13}$$

where $\{p_i\}$ is a sequence of observed, time-invariant variables of dimension l_p and $\{x_{it}\}$ is the same as in equation (2). Model (13) yields the conventional dynamic panel data model

$$\begin{aligned} y_{it} &= \mu_i(1-\alpha) + (1-\alpha)\gamma' p_i + \alpha y_{i,t-1} + u_{it} \\ &= \mu(1-\alpha) + (1-\alpha)\gamma' p_i + \alpha y_{i,t-1} + u_{it} + m_i(1-\alpha). \end{aligned} \quad (14)$$

We assume for Model (13)

Assumption 4 *Assume*

$$\begin{pmatrix} p_i \\ m_i \\ u_i \\ x_{i1} \end{pmatrix} \sim \text{i. i. d.} \left(\mathbf{0}, \begin{bmatrix} \phi_{\text{pp}} & \phi_{\text{pm}} & 0 & 0 \\ \phi'_{\text{pm}} & \sigma^2_{\text{m}} & 0 & 0 \\ 0 & 0 & \sigma^2_{\text{u}} I_{T-1} & 0 \\ 0 & 0 & 0 & \sigma^2_{\text{x}_1} \end{bmatrix} \right),$$

where $\sigma^2_{\text{m}} > 0$, $\sigma^2_{\text{u}} > 0$, $\sigma^2_{\text{x}_1} > 0$ *and* $\Psi = \begin{bmatrix} \phi_{\text{pp}} & \phi_{\text{pm}} \\ \phi'_{\text{pm}} & \sigma^2_{\text{m}} \end{bmatrix} > 0.$

This assumption implies that all the regressors of Model (14) are endogenous when $\phi_{\text{pm}} \neq 0$. This feature makes OLS or Within-OLS unusable for Model (14).

6.2 QMLE

Now, we consider QMLE for Model (13) as in Section 3. Assume $\mu = 0$ for simplicity from now on. Model (13) can be written as

$$y_{it} = \gamma' p_i + \alpha^{t-1} x_{i1} + w_{it} + m_i, \ (t = 2, ..., T),$$

where $w_{it} = \sum_{j=0}^{t-2} \alpha^j u_{i,t-j}$. Because $y_{i1} = \mu_i + \gamma' p_i + x_{i1}$, we have for $t = T$

$$\begin{aligned} y_{iT} &= \gamma' p_i + \alpha^{T-1}(y_{i1} - \mu_i - \gamma' p_i) + w_{iT} + m_i \\ &= (1-\alpha^{T-1})\gamma' p_i + \alpha^{T-1} y_{i1} + w_{iT} + (1-\alpha^{T-1}) m_i. \end{aligned} \quad (15)$$

All the regressors $\{p_i\}$ and $\{y_{i1}\}$ are endogenous if $\phi_{\text{pm}} \neq 0$ and $\alpha \neq 1$.

Let $c_{iT} = w_{iT} + (1-\alpha^{T-1}) m_i$. Then,

$$\begin{pmatrix} c_{iT} \\ p_i \\ y_{i1} \end{pmatrix} \sim \text{i. i. d.} \left(0, \ \Delta\right),$$

where $\Delta = \begin{bmatrix} \delta_{11} & \delta_{12} & \delta_{13} \\ \delta_{12}' & \delta_{22} & \delta_{23} \\ \delta_{13} & \delta_{23}' & \delta_{33} \end{bmatrix}$ and

$$\begin{aligned}
\delta_{11} &= \sigma_{\mathrm{u}}^2 \sum_{j=0}^{T-2} \alpha^{2j} + (1-\alpha^{T-1})^2 \sigma_{\mathrm{m}}^2;\ \delta_{12} = (1-\alpha^{T-1})\phi_{\mathrm{pm}}';\\
\delta_{13} &= (1-\alpha^{T-1})\sigma_{\mathrm{m}}^2 + (1-\alpha^{T-1})\gamma'\phi_{\mathrm{pm}};\ \delta_{22} = \phi_{\mathrm{pp}};\\
\delta_{23} &= \phi_{\mathrm{pm}} + \phi_{\mathrm{pp}}\gamma;\ \delta_{33} = \sigma_{\mathrm{m}}^2 + \gamma'\phi_{\mathrm{pp}}\gamma + \sigma_{\mathrm{x}_1}^2 + 2\gamma'\phi_{\mathrm{pm}}.
\end{aligned}$$

Let $\Delta = \begin{bmatrix} \delta_{11} & \delta_{\bullet 1}' \\ \delta_{\bullet 1} & \delta_{\bullet\bullet} \end{bmatrix}$, $q_i = \begin{pmatrix} p_i' & y_{i1} \end{pmatrix}'$ and $\delta_{11.\bullet} = \delta_{11} - \delta_{\bullet 1}'\delta_{\bullet\bullet}^{-1}\delta_{\bullet 1}$. As in Section 3, write

$$c_{iT} = \delta_{\bullet 1}'\delta_{\bullet\bullet}^{-1} q_i + r_{iT}.$$

Then,

$$\begin{pmatrix} r_{iT} \\ q_i \end{pmatrix} \sim \text{i. i. d.} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \delta_{11.\bullet} & 0 \\ 0 & \delta_{\bullet\bullet} \end{bmatrix} \right).$$

Let $\varrho = (\alpha,\ \gamma,\ \sigma_{\mathrm{u}}^2,\ \phi_{\mathrm{pp}},\ \phi_{\mathrm{pm}},\ \sigma_{\mathrm{m}}^2,\ \sigma_{\mathrm{x}_1}^2)'$, $z_i = (y_{iT}, q_i')'$, $\Sigma_{zz} = E(z_1 z_1')$ and $M_{zz} = \frac{1}{n}\sum_{i=1}^n z_i z_i'$. The objective function we employ is

$$\begin{aligned}
L(\varrho) &= -\ln\det(\Sigma_{zz}) - tr(\Sigma_{zz}^{-1} M_{zz}) \qquad (16)\\
&= -\ln(\delta_{11.\bullet}) - \ln(\det(\delta_{\bullet\bullet}))\\
&\quad -\delta_{11.\bullet}^{-1}\left[M_{TT} - 2\tau' M_{T1} + \tau' M_{11}\tau\right] - tr(\delta_{\bullet\bullet}^{-1} M_{11}),
\end{aligned}$$

where $\tau = \begin{pmatrix} \left(1-\alpha^{T-1}\right)\gamma \\ \alpha^{T-1} \end{pmatrix} + \delta_{\bullet\bullet}^{-1}\delta_{\bullet 1}$, $M_{TT} = \frac{1}{N}\sum_{i=1}^N y_{iT}^2$, $M_{T1} = \frac{1}{N}\sum_{i=1}^N y_{iT} q_i$ and $M_{11} = \frac{1}{N}\sum_{i=1}^N q_i q_i'$. Let Ω be any compact subset of $\{v = (v_\alpha, v_\gamma, v_u, \nu_{pp}, v_{pm}, v_m, v_x)' \mid v \in \mathbb{R}\times\mathbb{R}^{l_p}\times\mathbb{R}^+\times\mathbb{R}^{l_p(l_p+1)/2}\times\mathbb{R}^{l_p}\times\mathbb{R}^+\times\mathbb{R}^+\}$, where the partition of v is conformable to that of ϱ. The QMLE of ϱ is defined as

$$\hat{\varrho} = \arg\max_{\varrho\in\Omega} l(\varrho).$$

The following theorem reports the consistency properties of the QMLE.

Theorem 4 *Suppose that Assumption 4 holds.*
(i) If $\alpha^* \neq 1$, $\hat{\varrho} \xrightarrow{p} \varrho^*$.
(ii) If $\alpha^* = 1$, $\hat{\alpha} \xrightarrow{p} 1$, $\hat{\sigma}_{\mathrm{u}}^2 \xrightarrow{p} \sigma_{\mathrm{u}}^{*2}$, $\hat{\phi}_{\mathrm{pp}} \xrightarrow{p} \phi_{\mathrm{pp}}^*$, $\hat{\sigma}_{\mathrm{x}_1}^2 + \hat{\sigma}_{\mathrm{m}}^2 - \hat{\phi}_{\mathrm{pm}}'\hat{\phi}_{\mathrm{pp}}^{-1}\hat{\phi}_{\mathrm{pm}} \xrightarrow{p} \sigma_{\mathrm{x}_1}^{*2} + \sigma_{\mathrm{m}}^{*2} - \phi_{\mathrm{pm}}^{*\prime}\phi_{\mathrm{pp}}^{*-1}\phi_{\mathrm{pm}}^*$ *and* $\hat{\phi}_{\mathrm{pm}} + \hat{\phi}_{\mathrm{pp}}\hat{\gamma} \xrightarrow{p} \phi_{\mathrm{pm}}^* + \phi_{\mathrm{pp}}^*\gamma^*$.

If $\alpha^* \neq 1$, $\hat{\varrho}$ is consistent for ϱ^*. But if $\alpha^* = 1$, only α^*, σ_{u}^{*2}, ϕ_{pp}^*, $\sigma_{\mathrm{x}_1}^{*2} + \sigma_{\mathrm{m}}^{*2} - \phi_{\mathrm{pm}}^{*\prime}\phi_{\mathrm{pp}}^{*-1}\phi_{\mathrm{pm}}^*$ and $\phi_{\mathrm{pm}}^* + \phi_{\mathrm{pp}}^*\gamma^*$ can be estimated consistently. The parameters γ^*, $\sigma_{\mathrm{x}_1}^{*2}$, σ_{m}^{*2} and ϕ_{pm}^* cannot be identified separately when $\alpha^* = 1$. But this does not affect asymptotic properties of BCPLSE of Section 5.

To reduce the number of parameters to estimate as in Section 3, assume ϕ_{pp}, δ_{23},

δ_{33} and $\lambda = Var(y_{i2} - y_{i1}) = \sigma^2_{\mathrm{u}} + (1-\alpha)^2\sigma^2_{\mathrm{x}_1}$ are known. Then, because $\sigma^2_{\mathrm{u}} = \lambda - (1-\alpha)^2\sigma^2_{\mathrm{x}_1}$, $\phi_{\mathrm{pm}} = \delta_{23} - \phi_{\mathrm{pp}}\gamma$ and $\sigma^2_{\mathrm{x}_1} = \delta_{33} - \sigma^2_{\mathrm{m}} - \gamma'\phi_{\mathrm{pp}}\gamma - 2\gamma'\delta_{23}$, only α, γ and σ^2_{m} need to be estimated by maximizing the objective function. Lastly, σ^2_{m} can be estimated using the objective function (16). Then, the objective function contains only α and γ, which are the parameters of our main interest.

Adding time-variant endogenous variables $\{p_{it}\}$ to Model (14) and deriving similar results are also possible. We discuss this issue briefly here. For simplicity, assume that there are no time-invariant variables. Now, it is required to change Assumption 4 to

$$\begin{pmatrix} p_{i1} \\ p_{iT} \\ m_i \\ u_i \\ x_{i1} \end{pmatrix} \sim \mathrm{i.i.d.} \left(\mathbf{0}, \begin{bmatrix} \phi^{11}_{\mathrm{pp}} & \phi^{1T}_{\mathrm{pp}} & \phi^{1}_{\mathrm{pm}} & 0 & 0 \\ \phi^{1T\prime}_{\mathrm{pp}} & \phi^{TT}_{\mathrm{pp}} & \phi^{T}_{\mathrm{pm}} & 0 & 0 \\ \phi^{1\prime}_{\mathrm{pm}} & \phi^{T\prime}_{\mathrm{pm}} & \sigma^2_{\mathrm{m}} & 0 & 0 \\ 0 & 0 & 0 & \sigma^2_{\mathrm{u}} I_{T-1} & 0 \\ 0 & 0 & 0 & 0 & \sigma^2_{\mathrm{x}_1} \end{bmatrix} \right).$$

We need an assumption involving only p_{i1} and p_{iT} because only these are used for our regression. The regression model with time-variant regressors is

$$y_{iT} = \gamma' p^{\dagger}_i + \alpha^{T-1} y_{i1} + w_{iT} + (1-\alpha^{T-1}) m_i,$$

where $p^{\dagger}_i = p_{iT} - \alpha^{T-1} p_{i1}$. We can show that $\begin{pmatrix} c_{iT} \\ p^{\dagger}_i \\ y_{i1} \end{pmatrix} \sim \mathrm{i.i.d.}\left(0,\ \Delta^{\#}\right)$, where $\Delta^{\#} = \begin{bmatrix} \delta^{\#}_{11} & \delta^{\#\prime}_{21} & \delta^{\#}_{13} \\ \delta^{\#}_{21} & \delta^{\#}_{22} & \delta^{\#}_{23} \\ \delta^{\#}_{13} & \delta^{\#\prime}_{23} & \delta^{\#}_{33} \end{bmatrix}$ and

$$\begin{aligned}
\delta^{\#}_{11} &= \sigma^2_{\mathrm{u}} \sum_{j=0}^{T-2} \alpha^{2j} + (1-\alpha^{T-1})^2 \sigma^2_{\mathrm{m}}; \\
\delta^{\#}_{21} &= (1-\alpha^{T-1})\phi^{T}_{\mathrm{pm}} - \alpha^{T-1}(1-\alpha^{T-1})\phi^{1}_{\mathrm{pm}};\ \delta^{\#}_{22} = \phi^{TT}_{\mathrm{pp}} + \alpha^{2(T-1)}\phi^{11}_{\mathrm{pp}} - 2\alpha^{(T-1)}\phi^{1T}_{\mathrm{pp}}; \\
\delta^{\#}_{13} &= (1-\alpha^{T-1})\sigma^2_{\mathrm{m}} + (1-\alpha^{T-1})\gamma'\phi^{1}_{\mathrm{pm}}; \\
\delta^{\#}_{23} &= \phi^{T}_{\mathrm{pm}} - \alpha^{T-1}\phi^{1}_{\mathrm{pm}} + \phi^{1T\prime}_{\mathrm{pp}}\gamma - \alpha^{T-1}\phi^{11}_{\mathrm{pp}}\gamma; \\
\delta^{\#}_{33} &= \sigma^2_{\mathrm{m}} + \gamma'\phi^{11}_{\mathrm{pp}}\gamma + \sigma^2_{\mathrm{x}_1} + 2\gamma'\phi^{1}_{\mathrm{pm}}.
\end{aligned}$$

The objective function is formulated as above using $\Delta^{\#}$. Assuming that ϕ^{11}_{pp}, ϕ^{1T}_{pp}, ϕ^{TT}_{pp}, $\delta^{\#}_{33}$, and $Var(y_{iT} - y_{i1})$ are known, the objective function is rewritten as a function of α, γ, ϕ^{1}_{pm}, ϕ^{T}_{pm} and σ^2_{m}. Finally, we estimate ϕ^{1}_{pm}, ϕ^{T}_{pm} and σ^2_{m} using the original objective function and plug the estimators into the rewritten objective function from which we obtain the QMLEs of α and γ.

6.3 Indirect inference estimation

The indirect inference estimation for Model (15) proceeds in the same manner as in Section 4. We should assume $\alpha^* \neq 1$ in this subsection because σ_{m}^{2*} is not identified when $\alpha^* = 1$. Suppose that we can generate simulated data by using Model (15) once the value of α and γ are given, and denote the h-th simulated data as $\{\mathbf{y}_T^h(\alpha,\gamma),\ \mathbf{y}_1^h(\alpha,\gamma)\}$, where $\mathbf{y}_T^h(\alpha,\gamma)$ and $\mathbf{y}_1^h(\alpha,\gamma)$ are vectors of the regressand and regressor, respectively. Let $\theta^h(\alpha,\gamma) = \begin{pmatrix} \hat{\alpha}^h \\ \hat{\gamma}^h \end{pmatrix}$ be the QMLE of $\theta^* = \begin{pmatrix} \alpha^* \\ \gamma^* \end{pmatrix}$ using the h-th simulated data $\{\mathbf{y}_1^h(\alpha,\gamma), ...,\ \mathbf{y}_T^h(\alpha,\gamma)\}$. The indirect inference estimator (IIE) of θ^* is defined by

$$\hat{\theta}_{IIE} = \arg\min_{\theta\in\Lambda} \left\| \hat{\theta} - \frac{1}{H}\sum_{h=1}^{H} \theta^h(\alpha,\gamma) \right\|,$$

where Λ is a compact subset of $\mathbb{R}^{l_p+1}$ and $\|\cdot\| : \mathbb{R}^{l_p+1} \to [0,\infty)$ is the Euclidian norm. Letting $b_N^H(\alpha,\gamma) = \frac{1}{H}\sum_{h=1}^{H}\theta^h(\alpha,\gamma)$, assume as in Section 4

Assumption 5 $b_N^H(\alpha,\gamma) : \Lambda \to \Lambda$ *is an one-to-one, continuous function of* α *and* γ *for all* N *and* H.

Assumption 6 $b_N^{\infty}(\alpha,\gamma) = \theta + N^{-1}c(\theta)$, *where* $c(\theta) : \Lambda \to \mathbb{R}^{l_p+1}$ *is a continuously differentiable function of* θ.

Now, we have

Theorem 5 *Strengthen Assumption 4 such that the random vector there is normally distributed, and suppose that the strengthened Assumption 4, Assumption 5 and Assumption 6 hold. Then, if* $\alpha^* \neq 1$,
(i) $\hat{\theta}_{IIE}^{\infty} \xrightarrow{p} \theta^*$.
(ii) $\sqrt{N}(\hat{\theta}_{IIE}^{\infty} - \theta^*)$ *has the same asymptotic distribution as* $\sqrt{N}(\hat{\theta} - \theta^*)$.
(iii) $E(\hat{\theta}_{IIE}^{\infty}) = \theta^* + E(\zeta_N)$, *where* $\zeta_N = O_p(N^{-3/2})$.

6.4 Bias-corrected pooled least squares estimator

Let $\psi = [\gamma_\alpha',\ \alpha]'$ with $\gamma_\alpha = (1-\alpha)\gamma$. Running OLS using Model (14), we obtain the PLSE of ψ as

$$\begin{aligned} \hat{\psi}_{PLSE} &= \begin{bmatrix} \sum_{i=1}^{N}\sum_{t=2}^{T}(p_i-\bar{p})^2 & \sum_{i=1}^{N}\sum_{t=2}^{T}(p_i-\bar{p})(y_{i,t-1}-\bar{y}_{-1}) \\ \sum_{i=1}^{N}\sum_{t=2}^{T}(p_i-\bar{p})(y_{i,t-1}-\bar{y}_{-1}) & \sum_{i=1}^{N}\sum_{t=2}^{T}(y_{i,t-1}-\bar{y}_{-1})^2 \end{bmatrix}^{-1} \\ &\quad \times \begin{bmatrix} \sum_{i=1}^{N}\sum_{t=2}^{T}(p_i-\bar{p})\,y_{it} \\ \sum_{i=1}^{N}\sum_{t=2}^{T}(y_{i,t-1}-\bar{y}_{-1})y_{it} \end{bmatrix}. \end{aligned}$$

Under Assumption 4, we obtain

$$\hat{\psi}_{PLSE} - \psi^* \xrightarrow{p} \begin{bmatrix} (T-1)\phi^*_{\mathrm{pp}} & (T-1)(\phi^*_{\mathrm{pp}}\gamma^* + \phi^*_{\mathrm{pm}}) \\ (T-1)(\gamma^{*\prime}\phi^*_{\mathrm{pp}} + \phi^{*\prime}_{\mathrm{pm}}) & (T-1)(\sigma^{*2}_{\mathrm{m}} + \gamma^{*\prime}\phi^*_{\mathrm{pp}}\gamma^* + 2\gamma^{*\prime}\phi^*_{\mathrm{pm}}) + \kappa \end{bmatrix}^{-1} \times \begin{bmatrix} (1-\alpha^*)(T-1)\phi^*_{\mathrm{pm}} \\ (1-\alpha^*)(T-1)(\sigma^{*2}_{\mathrm{m}} + \gamma^{*\prime}\phi^*_{\mathrm{pm}}) \end{bmatrix},$$

where $\kappa = E\left(\sum_{t=2}^{T}(x_{i,t-1} - \bar{x}_{-1})^2\right)$. This shows that $\hat{\psi}_{PLSE}$ is inconsistent unless $\alpha^* = 1$. Thus, the BCPLSE of ψ is defined as

$$\begin{aligned} \hat{\psi}_{BCPLSE} = {} & \hat{\psi}_{PLSE} - \\ & \left[\frac{1}{N}\begin{pmatrix} \sum_{i=1}^{N}\sum_{t=2}^{T}(p_i - \bar{p})^2 & \sum_{i=1}^{N}\sum_{t=2}^{T}(p_i - \bar{p})(y_{i,t-1} - \bar{y}_{-1}) \\ \sum_{i=1}^{N}\sum_{t=2}^{T}(p_i - \bar{p})(y_{i,t-1} - \bar{y}_{-1}) & \sum_{i=1}^{N}\sum_{t=2}^{T}(y_{i,t-1} - \bar{y}_{-1})^2 \end{pmatrix}\right]^{-1} \\ & \times \begin{bmatrix} (1-\hat{\alpha})(T-1)\hat{\phi}_{\mathrm{pm}} \\ (1-\hat{\alpha})(T-1)(\hat{\sigma}^2_{\mathrm{m}} + \hat{\gamma}'\hat{\phi}_{\mathrm{pm}}) \end{bmatrix}, \end{aligned}$$

where $\hat{\alpha}$, $\hat{\gamma}$, $\hat{\phi}_{\mathrm{pm}}$ and $\hat{\sigma}^2_{\mathrm{m}}$ are the QMLEs of the previous subsection.

Let $\hat{\psi}_{BCPLSE}$ be partitioned as $[\hat{\psi}^{\gamma\prime}_{BCPLSE}, \ \hat{\psi}^{\alpha}_{BCPLSE}]'$, where $\hat{\psi}^{\gamma}_{BCPLSE}$ is of dimension l_p, and let $\hat{\gamma}_{BCPLSE} = \hat{\psi}^{\gamma}_{BCPLSE}/(1 - \hat{\psi}^{\alpha}_{BCPLSE})$. Then, we obtain:

Theorem 6 *Suppose that Assumption 4 holds. Then,*

(i) $\hat{\psi}^{\alpha}_{BCPLSE} \xrightarrow{p} \alpha^*$ *and* $\sqrt{N}\left(\hat{\psi}^{\alpha}_{BCPLSE} - \alpha^*\right) = O_p(1)$;

(ii) Assume $\alpha^* \neq 1$. *Then,* $\hat{\gamma}_{BCPLSE} \xrightarrow{p} \gamma^*$ *and* $\sqrt{N}(\hat{\gamma}_{BCPLSE} - \gamma^*) = O_p(1)$.

Part (i) of this theorem shows that $\hat{\psi}^{\alpha}_{BCPLSE}$ estimates α^* consistently and that $\sqrt{N}$-asymptotics applies to it. The same holds for $\hat{\gamma}_{BCPLSE}$ unless $\alpha^* = 1$. When $\alpha^* = 1$, $\hat{\gamma}_{BCPLSE}$ diverges in probability. The bootstrap method of Section 5 can also be used for $\hat{\psi}_{BCPLSE}$.

7 Simulation

This section reports simulation results for the following estimators: QMLE, IIE[7] and BCPLSE of this paper, GMM estimators of Arellano and Bond (1991) (GMM1) , Ahn and Schmidt (1995) (GMM2[8]) , Blundell and Bond (1998) (GMM3) and Blundell and Bond with homoskedasticity restrictions (GMM4), Han, Phillips and Sul's (2014) panel fully aggregated estimator (PFAE)[9], Anderson and Hsiao's (1982) MLE (AH)and Bun and Carree's (2005) bias-corrected Within-OLS estimator (BCWOLS).

[7] We consider the case $\alpha^* = 1$ for IIE, although IIE's properties under that case are not studied in this paper.

[8] We assumed $Var(u_{it}) = \sigma^2_{\mathrm{u}}$ for every i and t for this estimator.

[9] The previous versions of this paper also considered Han and Phillips' (2010) first difference least squares estimator. But we found that it performs worse than PFAE in finite samples.

All the estimators we employ for the purpose of comparison are explained briefly in Supplement 1 to this paper.

7.1 Efficiency comparison

Data were generated by Model (1) under Assumption 1 with normality. For simulation, values of the parameters α, μ, σ_{u}^2, σ_{m}^2 and $\sigma_{\mathrm{x}_1}^2$, and sample sizes N and T should be selected. First of all, we set $\mu = 0$ and $\sigma_{\mathrm{u}}^2 = 1$. For the initial variable x_{i1}, we set

$$\sigma_{\mathrm{x}_1}^2 = \begin{cases} \sigma_{\mathrm{u}}^2/(1-\alpha^2), & \text{if } \alpha < 1 \\ 5, & \text{if } \alpha \geq 1 \end{cases} \tag{17}$$

and let $\sigma_{\mathrm{m}}^2/\sigma_{\mathrm{u}}^2 = k$. The set-up for the initial variable follows previous studies (e.g., Blundell and Bond, 1998) when $\alpha < 1$. But when $\alpha \geq 1$, the conventional set-up cannot be used. Thus, we chose $\sigma_{\mathrm{x}_1}^2 = 5$[10], which is larger than the variance for the stationary case. It is reasonable to do so, because the variance of observations $\{x_{it}\}$ becomes larger as the value of α increases. The parameter values considered are $\alpha = 0.5.0.8, 1.0, 1.1$; $k = 1,\ 2$, and the sample sizes are $N = 100,\ 500$ and $T = 4,\ 10$. The number of iterations for simulation is 1,000, except the case of IIE for which it is 500. For IIE, we set $H = 10$. Bias-corrected Within-OLS is not considered in the case of $\alpha = 1.1$ because it impossible to identify α when it does not belong in $[-1, 1)$. All the computation was done by Matlab, and QMLE was calculated by a Matlab procedure fmincon.

Table 1 reports empirical biases, variances and RMSEs of aforementioned estimators[11] under the assumption of normality. The results reported in Table 1 can be summarized as follows.

(i) IIE tends to be less biased than QMLE as is expected from Theorem 2. But it has slightly higher variance than QMLE in some cases.

(ii) Comparing the 10 estimators in terms of RMSEs, BCPLSE performs best: its RMSE is the lowest in 26 cases out of 32. IIE and QMLE also show good performance and their RMSEs are quite close to those of BCPLSE in most cases. PFAE performs best in 6 cases. When PFAE performs best, the differences between BCPLSE and PFAE are marginal. In contrast, when BCPLSE performs best, its RMSE is sometimes less than 5% of that of PFAE (see the case of $(T, \alpha, k) = (4, 1.1, , 1.0)$ at $N = 500$).

(iii) At fixed N, RMSEs of most of the estimators tend to decrease as T increases.

(iv) At fixed T, RMSEs of most of the estimators decrease as N increases. Exceptions are sometimes observed for GMM1 and GMM2 at $\alpha = 1$. This occurs due to the problem of weak instruments. At the other values of α, their RMSEs also decrease as N increases.

[10] When $\sigma_{\mathrm{x}_1}^2 = 10$, it is found that the RMSEs of QMLE and BCPLSE decrease and thate of PFAE increases. QMLE and BCPLSE continue to show better performance than the rest in this case.

[11] BCWOLS of Bun and Carree (2005) cannot be computed for the case $|\alpha| > 1$. So there are no simulation results for BCWOLS in that case.

(v) At $\alpha = 1$ and $\alpha = 1.1$, GMM1 and GMM2 perform poorly compared with IIE, QMLE, BCPLSE, GMM3 and GMM4. This stems from the problem of weak instruments that these estimators share at these values of α.

(vi) When $T = 4$, GMM4 tends to perform better than GMM3. GMM3 shows better performance than GMM4 at $T = 10$.

(vii) At $\alpha = 1.1$, PFAE is strongly biased, causing high RMSEs.

(viii) As the value of k increases, most of the estimators perform poorer. PFAE and BCWOLS are invariant to the value of k.

(ix) As the value of α grows, the RMSEs of QMLE, BCPLSE and AH decrease, while those of PFAE show the opposite behavior.

(x) When $T = 4$, QMLE tends to perform better than AH. But QMLE has higher RMSEs than AH at $T = 10$.

(xi) QMLE's and BCPLSE's biases tend to move in the same direction because they are related by formula (12).

To summarize the overall simulation results, Table 2 reports frequencies of the 10 estimators performing as best and second-best players in terms of RMSEs. BCPLSE performs best in 26 cases out of 32 and PFAE in 6 cases. Those 6 cases are from $\alpha = 0.5,\ 0.8$ when $T = 10$. In the spots for the second-best players, QMLE comes 13 times and IIE 2. One last comment to add is that IIE, QMLE and BCPLSE show quite similar performance, although BCPLSE performs best in a strict sense. For BCPLSE, we can use bootstrapping for the construction of confidence intervals as will be in the next subsection.

7.2 Bootstrap confidence intervals

This subsection reports finite-sample performance of the bootstrap confidence interval of the PAR(1) coefficient explained in Section 5. Data were generated as for Table 1. We considered only the case $N = 100$ to save space and computation time. The number of bootstrap iterations (B) is set at 1,000. Empirical coverage ratios of the 95% and 90% confidence intervals based on 300 iterations are reported in Table 3. The results in Table 3 show the coverage ratios are reasonably close to the nominal coverage ratios. It is expected that they can be improved further by increasing the number of bootstrap iterations, although this incurs longer computation time.

8 Summary and further remarks

We have proposed three new estimators for the PAR(1) models with short T and large N. These estimators are based on the cross-sectional regression model using the first time series observations as a regressor and the last as a dependent variable. The first estimator is the cross-sectional QMLE that is consistent in the presence of the regressor-error dependency of the cross-sectional regression model. Using the cross-sectional QMLE, we constructed the indirect inference estimator and the bias-corrected pooled least squares estimator. These three estimators were also extended to the PAR model with endogenous regressors. The estimators of this paper are consistent for the PAR coefficients of stationary, unit root and explosive PAR models, estimate the coefficients of endogenous regressors consistently and can be computed as long as $T \geq 2$. The estimators were shown to perform quite well in finite samples relative to existing estimators.

The QMLE of this paper was employed to investigate the effects of agglomeration economy on cities' production and employment by Jung (2020). The QMLE and GMM3 provide similar results, but GMM1's results are sometimes quite different from the rest.

Appendix: Proofs

Proof of Theorem 1: Add a constant $\ln\det(\Sigma_{zz}^*) + 2 = \ln(\omega_{11.2}^*) + \ln(\omega_{22}^*) + 2$ to the objective function $L(\xi)$ and rewrite it as

$$L(\xi) = \bar{L}(\xi) + r(\xi),$$

where

$$\begin{aligned}
\bar{L}(\xi) &= -\ln\det(\Sigma_{zz}) - tr(\Sigma_{zz}^{-1}\Sigma_{zz}^*) + 2 + \ln\det(\Sigma_{zz}^*) \\
&= -\ln(\omega_{11.2}) - \ln(\omega_{22}) + \ln(\omega_{11.2}^*) + \ln(\omega_{22}^*) \\
&\quad - \left[\omega_{11.2}^{-1}\left(\omega_{11.2}^* + \chi^{*2}\omega_{22}^*\right) - 2\chi\chi^*\omega_{11.2}^{-1}\omega_{22}^* + \chi^2\omega_{11.2}^{-1}\omega_{22}^* + \omega_{22}^{-1}\omega_{22}^*\right] + 2,
\end{aligned}$$

and

$$\begin{aligned}
r(\xi) &= -tr(\Sigma_{zz}^{-1}(M_{zz} - \Sigma_{zz}^*)) \\
&= -\omega_{11.2}^{-1}\left(M_{TT} - 2\chi M_{T1} + \chi^2 M_{11}\right) - \omega_{22}^{-1}M_{11} \\
&\quad + \omega_{11.2}^{-1}\left(\omega_{11.2}^* + \chi^{*2}\omega_{22}^* - 2\chi\chi^*\omega_{22}^* + \chi^2\omega_{22}^*\right) + \omega_{22}^{-1}\omega_{22}^*.
\end{aligned}$$

Obviously, $\bar{L}(\xi^*) = 0$. Moreover, relation (8) derived from Assumption 1 yields

$$\begin{aligned}
\sup_{\xi\in\Xi}|r(\theta)| &\leq \sup_{\xi\in\Xi}\left|\omega_{11.2}^{-1}\right|\left|\left(M_{TT} - 2\chi M_{T1} + \chi^2 M_{11}\right) - \left(\omega_{11.2}^* + \chi^{*2}\omega_{22}^* - 2\chi\chi^*\omega_{22}^* + \chi^2\omega_{22}^*\right)\right| \\
&\quad + \sup_{\xi\in\Xi}\left|\omega_{22}^{-1}\right|\left|M_{11} - \omega_{22}^*\right| \\
&= o_p(1),
\end{aligned}$$

since ξ lies in a compact set. The QMLE $\hat{\xi}$ maximizes $L(\xi)$ so that

$$L(\hat{\xi}) = \bar{L}(\hat{\xi}) + r(\hat{\xi}) \geq \bar{L}(\xi^*) + r(\xi^*) = r(\xi^*),$$

from which we obtain $\bar{L}(\hat{\xi}) \geq r(\xi^*) - r(\hat{\xi}) \geq -\left|r(\xi^*) - r(\hat{\xi})\right|$. But $\left|r(\xi^*) - r(\hat{\xi})\right| \leq 2\sup_{\xi\in\Xi}|r(\theta)| = o_p(1)$. Therefore,

$$\bar{L}(\hat{\xi}) \geq -\left|o_p(1)\right|. \tag{A.1}$$

We may write

$$\begin{aligned}\bar{L}(\xi) &= -\ln(\omega_{11.2}) + \ln(\omega^*_{11.2}) - \omega^{-1}_{11.2}\omega^*_{11.2} + 1 \\ &\quad - \ln(\omega_{22}) + \ln(\omega^*_{22}) - \omega^{-1}_{22}\omega^*_{22} + 1 - \omega^{-1}_{11.2}\omega^*_{22}(\chi^* - \chi)^2.\end{aligned}$$

Here $-\ln(\omega_{11.2}) + \ln(\omega^*_{11.2}) - \omega^{-1}_{11.2}\omega^*_{11.2} + 1 \leq 0$ for $\omega_{11.2} \in \mathbb{R}^+$ and becomes zero at $\omega_{11.2} = \omega^*_{11.2}$. In addition, $-\ln(\omega_{22}) + \ln(\omega^*_{22}) - \omega^{-1}_{22}\omega^*_{22} + 1$ is also maximized at $\omega_{22} = \omega^*_{22}$. Thus, $\bar{L}(\xi)$ achieve its maximum at $\xi = \xi^*$, which yields

$$\bar{L}(\xi) \leq \bar{L}(\xi^*) = 0. \tag{A.2}$$

Relations (A.1) and (A.2) indicate $\bar{L}(\hat{\xi}) = o_p(1)$.

We have

$$\begin{aligned}\bar{L}(\hat{\xi}) &= -\ln\left(\hat{\sigma}^2_{\mathrm{u}}\sum_{j=0}^{T-2}\hat{\alpha}^{2j} + (1-\hat{\alpha}^{T-1})^2\hat{\sigma}^2_{\mathrm{m}} - \frac{(1-\hat{\alpha}^{T-1})^2\hat{\sigma}^4_{\mathrm{m}}}{\hat{\sigma}^2_{\mathrm{x}_1} + \hat{\sigma}^2_{\mathrm{m}}}\right) \\ &\quad + \ln\left(\sigma^{*2}_{\mathrm{u}}\sum_{j=0}^{T-2}\alpha^{*2j} + (1-\alpha^{*(T-1)})^2\sigma^{*2}_{\mathrm{m}} - \frac{(1-\alpha^{*(T-1)})^2\sigma^{*4}_{\mathrm{m}}}{\sigma^{*2}_{\mathrm{x}_1} + \sigma^{*2}_{\mathrm{m}}}\right) \\ &\quad - \left(\hat{\sigma}^2_{\mathrm{u}}\sum_{j=0}^{T-2}\hat{\alpha}^{2j} + (1-\hat{\alpha}^{T-1})^2\hat{\sigma}^2_{\mathrm{m}} - \frac{(1-\hat{\alpha}^{T-1})^2\hat{\sigma}^4_{\mathrm{m}}}{\hat{\sigma}^2_{\mathrm{x}_1} + \hat{\sigma}^2_{\mathrm{m}}}\right)^{-1} \\ &\quad \times \left(\sigma^{*2}_{\mathrm{u}}\sum_{j=0}^{T-2}\alpha^{*2j} + (1-\alpha^{*(T-1)})^2\sigma^{*2}_{\mathrm{m}} - \frac{(1-\alpha^{*(T-1)})^2\sigma^{*4}_{\mathrm{m}}}{\sigma^{*2}_{\mathrm{x}_1} + \sigma^{*2}_{\mathrm{m}}}\right) + 1 \\ &\quad - \ln(\hat{\sigma}^2_{\mathrm{x}_1} + \hat{\sigma}^2_{\mathrm{m}}) + \ln(\sigma^{*2}_{\mathrm{x}_1} + \sigma^{*2}_{\mathrm{m}}) - \left(\hat{\sigma}^2_{\mathrm{x}_1} + \hat{\sigma}^2_{\mathrm{m}}\right)^{-1}\left(\sigma^{*2}_{\mathrm{x}_1} + \sigma^{*2}_{\mathrm{m}}\right) + 1 \\ &\quad -\hat{\omega}^{-1}_{11.2}\omega^*_{22}\left(\hat{\alpha}^{T-1} - \alpha^{*(T-1)} + \frac{\hat{\sigma}^2_{\mathrm{m}}}{\hat{\sigma}^2_{\mathrm{x}_1} + \hat{\sigma}^2_{\mathrm{m}}} - \frac{\sigma^{*2}_{\mathrm{m}}}{\sigma^{*2}_{\mathrm{x}_1} + \sigma^{*2}_{\mathrm{m}}} - \frac{\hat{\alpha}^{T-1}\hat{\sigma}^2_{\mathrm{m}}}{\hat{\sigma}^2_{\mathrm{x}_1} + \hat{\sigma}^2_{\mathrm{m}}} + \frac{\alpha^{*T-1}\sigma^{*2}_{\mathrm{m}}}{\sigma^{*2}_{\mathrm{x}_1} + \sigma^{*2}_{\mathrm{m}}}\right)^2.\end{aligned}$$

Assume $\alpha^* \neq 1$ and suppose that all or any of the QMLEs are inconsistent. Then, it is easy to find that $\bar{L}(\hat{\xi}) \xrightarrow{p} c$, where c is a non-zero constant. This is a contradiction to the fact $L(\hat{\xi}) = o_p(1)$. For example, suppose that $\hat{\sigma}^2_{\mathrm{m}} \xrightarrow{p} \sigma^{*2}_{\mathrm{m}} + 1$ and the rest of the QMLEs are consistent. Assume that $T = 2$, $\alpha^* = 0.5$ and $\sigma^{*2}_{\mathrm{u}} = \sigma^{*2}_{\mathrm{x}_1} = \sigma^{*2}_{\mathrm{m}} = 1$. Then,

$L(\hat{\xi}) \xrightarrow{p} \ln(\frac{7}{8}) - \frac{4}{7} \simeq -0.70$. If $\alpha^* = 1$, $\bar{L}(\hat{\xi})$ is simplified as

$$\begin{aligned}
\bar{L}(\hat{\xi}) &= -\ln\left(\hat{\sigma}_{\mathrm{u}}^2 \sum_{j=0}^{T-2} \hat{\alpha}^{2j} + (1-\hat{\alpha}^{T-1})^2 \hat{\sigma}_{\mathrm{m}}^2 - \frac{(1-\hat{\alpha}^{T-1})^2 \hat{\sigma}_{\mathrm{m}}^4}{\hat{\sigma}_{\mathrm{x}_1}^2 + \hat{\sigma}_{\mathrm{m}}^2}\right) + \ln\left(\sigma_{\mathrm{u}}^{*2}(T-1)\right) \\
&\quad - \left(\hat{\sigma}_{\mathrm{u}}^2 \sum_{j=0}^{T-2} \hat{\alpha}^{2j} + (1-\hat{\alpha}^{T-1})^2 \hat{\sigma}_{\mathrm{m}}^2 - \frac{(1-\hat{\alpha}^{T-1})^2 \hat{\sigma}_{\mathrm{m}}^4}{\hat{\sigma}_{\mathrm{x}_1}^2 + \hat{\sigma}_{\mathrm{m}}^2}\right)^{-1} \left(\sigma_{\mathrm{u}}^{*2}(T-1)\right) + 1 \\
&\quad - \ln(\hat{\sigma}_{\mathrm{x}_1}^2 + \hat{\sigma}_{\mathrm{m}}^2) + \ln(\sigma_{\mathrm{x}_1}^{*2} + \sigma_{\mathrm{m}}^{*2}) - \left(\hat{\sigma}_{\mathrm{x}_1}^2 + \hat{\sigma}_{\mathrm{m}}^2\right)^{-1} \left(\sigma_{\mathrm{x}_1}^{*2} + \sigma_{\mathrm{m}}^{*2}\right) + 1 \\
&\quad - \hat{\omega}_{11.2}^{-1} \omega_{22}^* \left(\hat{\alpha}^{T-1} - 1 + \frac{\hat{\sigma}_{\mathrm{m}}^2}{\hat{\sigma}_{\mathrm{x}_1}^2 + \hat{\sigma}_{\mathrm{m}}^2} - \frac{\hat{\alpha}^{T-1} \hat{\sigma}_{\mathrm{m}}^2}{\hat{\sigma}_{\mathrm{x}_1}^2 + \hat{\sigma}_{\mathrm{m}}^2}\right)^2.
\end{aligned}$$

This shows that $\bar{L}(\hat{\xi}) = o_p(1)$ if and only if $\hat{\alpha} \xrightarrow{p} 1$, $\hat{\sigma}_{\mathrm{u}}^2 \xrightarrow{p} \sigma_{\mathrm{u}}^{*2}$ and $\hat{\sigma}_{\mathrm{x}_1}^2 + \hat{\sigma}_{\mathrm{m}}^2 \xrightarrow{p} \sigma_{\mathrm{x}_1}^{*2} + \sigma_{\mathrm{m}}^{*2}$.

Proof of Theorem 2: (i) Under the given assumptions, $\hat{\alpha} = \hat{\alpha}_{IIE}^{\infty} + N^{-1} c(\hat{\alpha}_{IIE}^{\infty})$. Since $\hat{\alpha} \xrightarrow{p} \alpha^*$ and $\hat{\alpha}_{IIE}^{\infty} = O_p(1)$, $N^{-1} c(\hat{\alpha}_{IIE}^{\infty}) = o_p(1)$ and $\hat{\alpha}_{IIE}^{\infty} \xrightarrow{p} \alpha^*$.

(ii) $\sqrt{N}(\hat{\alpha} - \alpha^*) - \sqrt{N}(\hat{\alpha}_{IIE}^{\infty} - \alpha^*) = N^{-1/2} c(\hat{\alpha}_{IIE}^{\infty}) = O_p(N^{-1/2})$. Thus, the stated result follows.

(iii) We obtain by the mean-value theorem

$$\begin{aligned}
E(\hat{\alpha}_{IIE}^{\infty}) &= E(\hat{\alpha}) - E c_N(\hat{\alpha}_{IIE}^{\infty}) \\
&= \alpha^* + N^{-1} c(\alpha^*) - N^{-1} E\left[c(\hat{\alpha}_{IIE}^{\infty})\right] \\
&= \alpha^* - E\left[N^{-3/2} c'(\alpha^{\dagger}) \sqrt{N}(\hat{\alpha}_{IIE}^{\infty} - \alpha^*)\right],
\end{aligned}$$

where $\alpha^{\dagger}$ lies in between $\hat{\alpha}_{IIE}^{\infty}$ and α^*. The result follows because $N^{-3/2} c'(\alpha^{\dagger}) \sqrt{N}(\hat{\alpha}_{IIE}^{\infty} - \alpha^*) = O_p(N^{-3/2})$.

Proof of Theorem 3: (i) This follows from Assumption 1 and Theorem 1.
(ii) We have

$$\begin{aligned}
&\sqrt{N}\left(\hat{\alpha}_{BCPLSE} - \alpha^*\right) \\
&= \sqrt{N}\left(\hat{\alpha}_{PLSE} - \alpha^* - \frac{(1-\alpha^*)(T-1)\sum_{i=1}^N \sum_{t=2}^T (m_i - \bar{m}) m_i}{\sum_{i=1}^N \sum_{t=2}^T (y_{i,t-1} - \bar{y}_{-1})^2}\right) \\
&\quad + \sqrt{N}\left(\frac{(1-\alpha^*)(T-1)\sum_{i=1}^N \sum_{t=2}^T (m_i - \bar{m}) m_i / N}{\sum_{i=1}^N \sum_{t=2}^T (y_{i,t-1} - \bar{y}_{-1})^2 / N} - \frac{(1-\hat{\alpha})(T-1)\hat{\sigma}_{\mathrm{m}}^2}{\sum_{i=1}^N \sum_{t=2}^T (y_{i,t-1} - \bar{y}_{-1})^2 / N}\right) \\
&= A_N + B_N, \text{ say.}
\end{aligned}$$

Equation (11) indicates $A_N = O_p(1)$ under Assumption 1. In addition,

$$\begin{aligned}
&\sqrt{N}\left[(1-\alpha^*)(T-1)\sum_{i=1}^{N}\sum_{t=2}^{T}(m_i-\bar{m})m_i/N-(1-\hat{\alpha})(T-1)\hat{\sigma}_{\mathrm{m}}^2\right] \\
= \;& (T-1)\sqrt{N}\left[\sum_{i=1}^{N}(m_i-\bar{m})m_i/N-\hat{\sigma}_{\mathrm{m}}^2\right] \\
&+(T-1)\sqrt{N}\left[\hat{\alpha}\hat{\sigma}_{\mathrm{m}}^2-\alpha^*(m_i-\bar{m})m_i/N\right] \\
= \;& (T-1)\sqrt{N}\left[\frac{1}{N}\sum_{i=1}^{N}\left((m_i-\bar{m})m_i-\sigma_{\mathrm{m}}^{*2}\right)-\left(\hat{\sigma}_{\mathrm{m}}^2-\sigma_{\mathrm{m}}^{*2}\right)\right] \\
&+(T-1)\sqrt{N}\left[(\hat{\alpha}-\alpha^*)\hat{\sigma}_{\mathrm{m}}^2-\alpha^*\left(\sum_{i=1}^{N}(m_i-\bar{m})m_i/N-\hat{\sigma}_{\mathrm{m}}^2\right)\right] \\
= \;& (T-1)\sqrt{N}\left[\frac{1}{N}\sum_{i=1}^{N}\left((m_i-\bar{m})m_i-\sigma_{\mathrm{m}}^{*2}\right)-\left(\hat{\sigma}_{\mathrm{m}}^2-\sigma_{\mathrm{m}}^{*2}\right)\right] \\
&+\hat{\sigma}_{\mathrm{m}}^2(T-1)\sqrt{N}\left[(\hat{\alpha}-\alpha^*)\right] \\
&-\alpha^*(T-1)\sqrt{N}\left[\frac{1}{N}\sum_{i=1}^{N}\left((m_i-\bar{m})m_i-\sigma_{\mathrm{m}}^{*2}\right)-\left(\hat{\sigma}_{\mathrm{m}}^2-\sigma_{\mathrm{m}}^{*2}\right)\right] \\
= \;& (1-\alpha^*)(T-1)\sqrt{N}\left[\frac{1}{N}\sum_{i=1}^{N}\left((m_i-\bar{m})m_i-\sigma_{\mathrm{m}}^{*2}\right)-\left(\hat{\sigma}_{\mathrm{m}}^2-\sigma_{\mathrm{m}}^{*2}\right)\right] \\
&+\hat{\sigma}_{\mathrm{m}}^2(T-1)\sqrt{N}(\hat{\alpha}-\alpha^*) \\
= \;& O_p(1),
\end{aligned}$$

which implies $B_N = O_p(1)$. Thus, the result follows.

Proof of Theorem 4: We can show easily that $\bar{L}(\hat{\varrho}) = o_p(1)$ as in the proof of Theorem 1. The rest of the argument depends on the functional form of $\bar{L}(\varrho)$. But it is highly complicated, so we put it in Supplement 3. Inspection of $\bar{L}(\varrho)$ reveals that $\bar{L}(\hat{\varrho}) = o_p(1)$ if the stated results hold.

Proof of Theorem 5: The result can shown in the same way as for Theorem 2**.**

Proof of Theorem 6: As in Section 5, it is straightforward to show under Assumption 4 (i) $\hat{\psi}_{BCPLSE} \xrightarrow{p} \psi$ and (ii) $\sqrt{N}(\hat{\psi}_{BCPLSE} - \psi) = O_p(1)$. The results follow from these.

Compliance with Ethical Standards:

Funding: This study was funded by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2015S1A5A2A01009333).

Conflict of Interest: Author In Choi declares that he/she has no conflict of interest. Author Sanghyun Jung declares that he/she has no conflict of interest.

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.

References

Ahn, S. C., & Schmidt, P. (1995). Efficient estimation of models for dynamic panel data. Journal of Econometrics, 68(1), 5-27.

Ahn, S. C., & Thomas, G. C. (2006). Likelihood based inference for dynamic panel data models. Unpublished manuscript, Arizona State University.

Alvarez, J., & Arellano, M. (2004). Robust likelihood estimation of dynamic panel data models. CEMFI working paper.

Anderson, T. W., & Hsiao, C. (1981). Estimation of dynamic models with error components. Journal of the American Statistical Association, 76(375), 598-606.

Anderson, T. W., & Hsiao, C. (1982). Formulation and estimation of dynamic models using panel data. Journal of Econometrics, 18(1), 47-82.

Andrews, D. W. K. (1993). Exactly median-unbiased estimation of first order autoregressive/unit root models. Econometrica, 61(1), 139–165.

Arellano, M., & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. The Review of Economic Studies, 58(2), 277-297.

Ashley, R. A., & Sun, X. (2016). Subset-Continuous-Updating GMM Estimators for Dynamic Panel Data Models. Econometrics, 4(4), 47.

Bai, J. (2013). Fixed-effects dynamic panel models, a factor analytical method. Econometrica, 81(1), 285-314.

Bai, J., & Li, K. (2012). Statistical analysis of factor models of high dimension. The Annals of Statistics, 40(1), 436-465.

Baltagi, B. (2008). Econometric analysis of panel data. John Wiley & Sons.

Blundell, R., & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. Journal of Econometrics, 87(1), 115-143.

Bun, M. J. & Carree, M. A. (2005) Bias-corrected estimation in dynamic panel data models. Journal of Business & Economic Statistics, 23(2), 200–210.

Bun, M. J. & Carree, M. A. (2006). Bias-corrected estimation in dynamic panel data models with heteroscedasticity. Economics Letters, 92(2), 220-227.

Bun, M. J., & Sarafidis, V. (2015). Dynamic panel data models. Oxford Handbook of Panel Data Econometrics, 76-110, Oxford University Press.

Bun, M. J., & Windmeijer, F. (2010). The weak instrument problem of the system GMM estimator in dynamic panel data models. The Econometrics Journal, 13(1), 95-126.

Choi, I. (2019) "Unit root tests for dependent and heterogeneous micro-panels," Japanese Economic Review 70 (2), 145-167.

Cox, D. R. and Snell, E. J. (1968). A general definition of residuals. Journal of the Royal Statistical Society, Series B, 30(2), 248–275.

Gourieroux, C., Monfort, A., & Renault, E. (1993). Indirect inference. Journal of Applied Econometrics, 8(S1), S85-S118.

Gourieroux, C., Monfort, A., & Trognon, A. (1984). Pseudo maximum likelihood methods: Theory. Econometrica, 52(3), 681-700.

Gourieroux, C., Phillips, P. C., & Yu, J. (2010). Indirect inference for dynamic panel models. Journal of Econometrics, 157(1), 68-77.

Hahn, J. (1999). How informative is the initial condition in the dynamic panel model with fixed effects?. Journal of Econometrics, 93(2), 309-326.
Hahn, J., Hausman, J., & Kuersteiner, G. (2007). Long difference instrumental variables estimation for dynamic panel models with fixed effects. Journal of Econometrics, 140(2), 574-617.
Hahn, J., & Kuersteiner, G. (2002). Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large. Econometrica, 70(4), 1639-1657.
Han, C., & Phillips, P.C.B. (2010). GMM estimation for dynamic panels with fixed effects and strong instruments at unity. Econometric Theory, 26(01), 119-151
Han, C., Phillips, P.C.B., & Sul, D. (2014). X-differencing and dynamic panel model estimation. Econometric Theory, 30(01), 201-251.
Hansen, L. P., Heaton, J., & Yaron, A. (1996). Finite-sample properties of some alternative GMM estimators. Journal of Business & Economic Statistics, 14(3), 262-280.
Hayakawa, K. (2007). Small sample bias properties of the system GMM estimator in dynamic panel data models. Economics Letters, 95(1), 32-38.
Hayakawa, K. (2012). GMM estimation of short dynamic panel data models with interactive fixed effects. Journal of the Japan Statistical Society, 42(2), 109-123.
Hsiao, C. (2003). Analysis of panel data. Second Edition, Cambridge University Press.
Hsiao, C., Pesaran, M. H., & Tahmiscioglu, A. K. (2002). Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. Journal of Econometrics, 109(1), 107-150.
Hsiao, C. and Q. Zhou (2018). Incidental parameters, initial conditions and sample size in statistical inference for dynamic panel data models. Journal of Econometrics, 207(1), 114-128.
Jung, S.H. (2020). Dynamic externalities and local economic development: evidence from South Korea manufacturing sectors, MA thesis, Sogang University.
Kiviet, J. (1995). On bias, inconsistency, and efficiency of various estimators in dynamic panel data models, Journal of Econometrics, 68(1), 53-78.
Kruiniger, H. (2008). Maximum likelihood estimation and inference methods for the covariance stationary panel AR (1)/unit root model. Journal of Econometrics, 144(2), 447-464.
Kruiniger, H. (2013). Quasi ML estimation of the panel AR (1) model with arbitrary initial conditions. Journal of Econometrics, 173(2), 175-188.
Phillips, P.C.B., Wu, Y., & Yu, J. (2011). Explosive behavior in the 1990s Nasdaq: when did exuberance escalate asset values?. International Economic Review, 52(1), 201-226.
Smith, A. A. (1993). Estimating nonlinear time-series models using simulated vector autoregressions, Journal of Applied Econometrics, 8(S1), S63–S84.

Table 1: Efficiency comparison under normal distributions

Note: Data were generated by Models (1) and (2) under Assumption 1 with $\sigma_{\mathrm{u}}^2 = 1$, assuming that the data are normally distributed. The number of iterations is 1,000 for all estimators except IIE. The number of iterations is 500 for IIE.

Part A: $N = 100$

(T, α, k)		IIE	QMLE	BCPLSE	GMM1	GMM2	GMM3	GMM4	PFAE	AH	BCWOLS
(4,0.5,1.0)	Bias	0.0053	-0.0042	-0.0010	-0.0124	0.0155	-0.0062	0.0141	0.0032	0.0018	0.0030
	Var	0.0115	0.0135	0.0065	0.0355	0.0326	0.0142	0.0180	0.0148	0.0124	0.0117
	RMSE	0.1071	0.1163	0.0806	0.1888	0.1813	0.1192	0.1348	0.1218	0.1112	0.1083
(4,0.5,2.0)	Bias	0.0055	-0.0033	0.0008	-0.0183	0.0234	0.0011	0.0336	0.0032	0.0035	0.0030
	Var	0.0136	0.0175	0.0097	0.0498	0.0443	0.0165	0.0230	0.0148	0.0141	0.0117
	RMSE	0.1167	0.1322	0.0985	0.2239	0.2117	0.1283	0.1553	0.1218	0.1186	0.1083
(4,0.8,1.0)	Bias	-0.0020	-0.0050	-0.0054	-0.0329	-0.0236	-0.0274	-0.0128	0.0006	-0.0285	-0.0031
	Var	0.0034	0.0030	0.0023	0.0774	0.0573	0.0201	0.0194	0.0171	0.0092	0.0141
	RMSE	0.0585	0.0550	0.0487	0.2801	0.2405	0.1443	0.1398	0.1308	0.1001	0.1187
(4,0.8,2.0)	Bias	-0.0040	-0.0074	-0.0071	-0.0462	-0.0293	-0.0271	-0.0062	0.0006	-0.0240	-0.0031
	Var	0.0044	0.0041	0.0032	0.1038	0.0691	0.0224	0.0220	0.0171	0.0117	0.0141
	RMSE	0.0664	0.0642	0.0572	0.3255	0.2644	0.1520	0.1485	0.1308	0.1107	0.1187
(4,1.0,1.0)	Bias	-0.0016	-0.0025	-0.0028	-0.9061	-0.5763	-0.0193	-0.0184	-0.0044	0.0006	-0.0548
	Var	0.0009	0.0008	0.0007	0.8369	0.4535	0.0151	0.0098	0.0197	0.0030	0.0059
	RMSE	0.0296	0.0278	0.0266	1.2876	0.8863	0.1244	0.1005	0.1405	0.0547	0.0942
(4,1.0,2.0)	Bias	-0.0015	-0.0026	-0.0030	-0.9106	-0.5733	-0.0207	-0.0193	-0.0044	-0.0083	-0.0548
	Var	0.0010	0.0009	0.0008	0.7852	0.4636	0.0170	0.0110	0.0197	0.0030	0.0059
	RMSE	0.0320	0.0298	0.0284	1.2706	0.8901	0.1321	0.1065	0.1405	0.0556	0.0942
(4,1.1,1.0)	Bias	-0.0016	-0.0027	-0.0029	-0.1018	-0.2159	-0.0046	-0.0072	0.2039	0.0045	-
	Var	0.0007	0.0006	0.0006	0.1587	0.1753	0.0061	0.0040	0.0231	0.0009	-
	RMSE	0.0270	0.0251	0.0240	0.4112	0.4711	0.0780	0.0639	0.2544	0.0311	-
(4,1.1,2.0)	Bias	-0.0015	-0.0027	-0.0030	-0.1223	-0.2300	-0.0039	-0.0071	0.2039	0.0020	-
	Var	0.0008	0.0007	0.0006	0.1867	0.2065	0.0071	0.0045	0.0231	0.0011	-
	RMSE	0.0279	0.0264	0.0252	0.4491	0.5093	0.0843	0.0674	0.2544	0.0328	-

Part A: $N = 100$ (continued)

(T, α, k)		IIE	QMLE	BCPLSE	GMM1	GMM2	GMM3	GMM4	PFAE	AH	BCWOLS
(10,0.5,1.0)	Bias	-0.0004	-0.0053	-0.0069	-0.0343	0.0855	0.0394	0.1547	-0.0004	0.0002	0.0002
	Var	0.0107	0.0117	0.0066	0.0040	0.0190	0.0195	0.0308	0.0015	0.0016	0.0016
	RMSE	0.1035	0.1083	0.0813	0.0721	0.1623	0.1450	0.2339	0.0386	0.0405	0.0399
(10,0.5,2.0)	Bias	0.0046	-0.0049	-0.0066	-0.0380	0.1248	0.0680	0.2045	-0.0004	0.0004	0.0002
	Var	0.0117	0.0126	0.0082	0.0046	0.0277	0.0260	0.0347	0.0015	0.0017	0.0016
	RMSE	0.1083	0.1122	0.0910	0.0779	0.2080	0.1750	0.2766	0.0386	0.0412	0.0399
(10,0.8,1.0)	Bias	-0.0031	-0.0052	-0.0025	-0.0723	0.0177	-0.0205	0.0682	-0.0019	-0.0031	-0.0007
	Var	0.0028	0.0026	0.0011	0.0065	0.0112	0.0117	0.0120	0.0014	0.0014	0.0018
	RMSE	0.0531	0.0512	0.0328	0.1083	0.1072	0.1100	0.1291	0.0374	0.0375	0.0421
(10,0.8,2.0)	Bias	-0.0022	-0.0074	-0.0041	-0.0838	0.0285	-0.0115	0.0830	-0.0019	-0.0024	-0.0007
	Var	0.0035	0.0034	0.0016	0.0077	0.0136	0.0132	0.0125	0.0014	0.0016	0.0018
	RMSE	0.0591	0.0584	0.0400	0.1213	0.1199	0.1153	0.1393	0.0374	0.0401	0.0421
(10,1.0,1.0)	Bias	-0.0002	-0.0007	-0.0009	-0.5302	-0.1362	-0.0373	0.0161	-0.0032	0.0042	-0.0194
	Var	0.0002	0.0002	0.0001	0.0482	0.0730	0.0075	0.0041	0.0014	0.0002	0.0007
	RMSE	0.0148	0.0144	0.0119	0.5739	0.3026	0.0943	0.0659	0.0376	0.0143	0.0331
(10,1.0,2.0)	Bias	-0.0005	-0.0008	-0.0010	-0.5303	-0.1256	-0.0400	0.0147	-0.0032	0.0025	-0.0194
	Var	0.0002	0.0002	0.0002	0.0487	0.0682	0.0076	0.0041	0.0014	0.0002	0.0007
	RMSE	0.0158	0.0153	0.0125	0.5744	0.2898	0.0960	0.0657	0.0376	0.0154	0.0331
(10,1.1,1.0)	Bias	-0.0004	-0.0006	-0.0006	-0.0384	-0.0059	-0.0029	0.0088	0.1793	-0.0004	-
	Var	0.0001	0.0001	0.0001	0.0022	0.0056	0.0020	0.0017	0.0008	0.0001	-
	RMSE	0.0103	0.0102	0.0082	0.0603	0.0750	0.0447	0.0416	0.1815	0.0088	-
(10,1.1,2.0)	Bias	-0.0001	-0.0005	-0.0006	-0.0392	-0.0088	-0.0044	0.0076	0.1793	-0.0006	-
	Var	0.0001	0.0001	0.0001	0.0022	0.0060	0.0020	0.0019	0.0008	0.0001	-
	RMSE	0.0104	0.0104	0.0083	0.0612	0.0780	0.0453	0.0442	0.1815	0.0093	-

Part B: $N = 500$

(T, α, k)		IIE	QMLE	BCPLSE	GMM1	GMM2	GMM3	GMM4	PFAE	AH	BCWOLS
(4,0.5,1.0)	Bias	-0.0023	-0.0035	-0.0008	0.0019	0.0074	-0.0020	-0.0008	0.0005	0.0008	0.0004
	Var	0.0031	0.0028	0.0013	0.0069	0.0067	0.0027	0.0027	0.0031	0.0026	0.0028
	RMSE	0.0555	0.0526	0.0366	0.0831	0.0822	0.0524	0.0521	0.0560	0.0509	0.0531
(4,0.5,2.0)	Bias	-0.0030	-0.0038	-0.0014	0.0014	0.0094	-0.0023	0.0009	0.0005	0.0006	0.0004
	Var	0.0041	0.0037	0.0021	0.0095	0.0091	0.0032	0.0033	0.0031	0.0030	0.0028
	RMSE	0.0638	0.0612	0.0462	0.0973	0.0958	0.0563	0.0574	0.0560	0.0543	0.0531
(4,0.8,1.0)	Bias	-0.0009	-0.0015	-0.0009	0.0010	0.0018	-0.0042	-0.0052	0.0004	-0.0039	-0.0002
	Var	0.0006	0.0006	0.0004	0.0142	0.0134	0.0031	0.0030	0.0038	0.0022	0.0033
	RMSE	0.0251	0.0238	0.0208	0.1190	0.1157	0.0562	0.0554	0.0613	0.0474	0.0571
(4,0.8,2.0)	Bias	-0.0014	-0.0021	-0.0014	0.0002	0.0011	-0.0061	-0.0069	0.0004	-0.0022	-0.0002
	Var	0.0008	0.0007	0.0006	0.0179	0.0166	0.0035	0.0034	0.0038	0.0028	0.0033
	RMSE	0.0288	0.0273	0.0243	0.1340	0.1287	0.0596	0.0587	0.0613	0.0529	0.0571
(4,1.0,1.0)	Bias	-0.0001	-0.0004	-0.0003	-0.8893	-0.7471	0.0006	-0.0011	0.0006	0.0081	-0.0234
	Var	0.0002	0.0001	0.0001	0.7894	0.4243	0.0009	0.0009	0.0041	0.0003	0.0012
	RMSE	0.0127	0.0122	0.0115	1.2571	0.9912	0.0300	0.0293	0.0641	0.0202	0.0416
(4,1.0,2.0)	Bias	-0.0001	-0.0005	-0.0004	-0.8863	-0.7308	0.0007	-0.0011	0.0006	-0.0010	-0.0234
	Var	0.0002	0.0002	0.0002	0.7228	0.4241	0.0009	0.0009	0.0041	0.0006	0.0012
	RMSE	0.0140	0.0132	0.0124	1.2281	0.9789	0.0306	0.0292	0.0641	0.0240	0.0416
(4,1.1,1.0)	Bias	0.0004	-0.0003	-0.0003	-0.0201	-0.0451	0.0008	0.0000	0.2115	0.0013	-
	Var	0.0001	0.0001	0.0001	0.0217	0.0259	0.0004	0.0004	0.0045	0.0002	-
	RMSE	0.0118	0.0110	0.0104	0.1485	0.1672	0.0190	0.0191	0.2218	0.0149	-
(4,1.1,2.0)	Bias	0.0001	-0.0003	-0.0004	-0.0233	-0.0502	0.0011	0.0005	0.2115	0.0002	-
	Var	0.0002	0.0001	0.0001	0.0250	0.0301	0.0004	0.0004	0.0045	0.0003	-
	RMSE	0.0123	0.0116	0.0110	0.1598	0.1805	0.0194	0.0191	0.2218	0.0161	-

Part B: $N = 500$ (continued)

(T, α, k)		IIE	QMLE	BCPLSE	GMM1	GMM2	GMM3	GMM4	PFAE	AH	BCWOLS
(10,0.5,1.0)	Bias	-0.0015	-0.0025	-0.0021	-0.0065	0.0224	0.0034	0.0336	0.0000	-0.0001	0.0001
	Var	0.0025	0.0023	0.0013	0.0007	0.0022	0.0008	0.0070	0.0003	0.0003	0.0003
	RMSE	0.0502	0.0481	0.0355	0.0280	0.0518	0.0277	0.0902	0.0175	0.0182	0.0180
(10,0.5,2.0)	Bias	-0.0004	-0.0024	-0.0023	-0.0073	0.0353	0.0088	0.0479	0.0000	-0.0001	0.0001
	Var	0.0028	0.0025	0.0016	0.0009	0.0040	0.0014	0.0094	0.0003	0.0003	0.0003
	RMSE	0.0532	0.0500	0.0396	0.0302	0.0722	0.0382	0.1083	0.0175	0.0185	0.0180
(10,0.8,1.0)	Bias	-0.0008	-0.0020	-0.0009	-0.0129	0.0097	-0.0043	0.0114	-0.0003	-0.0006	-0.0002
	Var	0.0005	0.0005	0.0002	0.0010	0.0016	0.0006	0.0019	0.0003	0.0003	0.0004
	RMSE	0.0228	0.0226	0.0145	0.0335	0.0415	0.0254	0.0452	0.0168	0.0175	0.0188
(10,0.8,2.0)	Bias	-0.0019	-0.0026	-0.0014	-0.0155	0.0143	-0.0019	0.0196	-0.0003	-0.0006	-0.0002
	Var	0.0007	0.0007	0.0003	0.0012	0.0022	0.0009	0.0033	0.0003	0.0003	0.0004
	RMSE	0.0264	0.0257	0.0174	0.0376	0.0495	0.0306	0.0605	0.0168	0.0183	0.0188
(10,1.0,1.0)	Bias	0.0002	-0.0003	-0.0002	-0.5804	-0.2452	-0.0072	-0.0011	-0.0005	0.0043	-0.0085
	Var	0.0000	0.0000	0.0000	0.0629	0.0962	0.0005	0.0006	0.0003	0.0000	0.0002
	RMSE	0.0063	0.0064	0.0053	0.6323	0.3954	0.0245	0.0244	0.0167	0.0057	0.0150
(10,1.0,2.0)	Bias	0.0001	-0.0003	-0.0002	-0.5792	-0.2393	-0.0074	-0.0013	-0.0005	0.0024	-0.0085
	Var	0.0000	0.0000	0.0000	0.0631	0.0953	0.0006	0.0005	0.0003	0.0000	0.0002
	RMSE	0.0068	0.0068	0.0055	0.6313	0.3906	0.0259	0.0226	0.0167	0.0070	0.0150
(10,1.1,1.0)	Bias	0.0003	-0.0002	-0.0001	-0.0061	-0.0027	-0.0015	-0.0005	0.1817	-0.0001	-
	Var	0.0000	0.0000	0.0000	0.0004	0.0008	0.0001	0.0001	0.0002	0.0000	-
	RMSE	0.0046	0.0045	0.0035	0.0201	0.0288	0.0101	0.0117	0.1821	0.0040	-
(10,1.1,2.0)	Bias	0.0003	-0.0002	-0.0001	-0.0064	-0.0038	-0.0011	-0.0005	0.1817	-0.0001	-
	Var	0.0000	0.0000	0.0000	0.0004	0.0008	0.0001	0.0001	0.0002	0.0000	-
	RMSE	0.0046	0.0046	0.0036	0.0207	0.0293	0.0105	0.0105	0.1821	0.0041	-

Table 2: Frequencies of best and second-best performers in terms of RMSE

(N,T)		IIE	QMLE	BCPLSE	GMM1	GMM2	GMM3	GMM4	PFAE	AH	BCWOLS	Total
(100,4)	Best	0	0	8	0	0	0	0	0	0	0	8
	2nd-best	1	6	0	0	0	0	0	0	0	1	8
(100,10)	Best	0	0	5	0	0	0	0	3	0	0	8
	2nd-best	0	1	1	0	0	0	0	1	3	2	8
(500,4)	Best	0	0	8	0	0	0	0	0	0	0	8
	2nd-best	0	6	0	0	0	0	0	0	1	1	8
(500,10)	Best	0	0	5	0	0	0	0	3	0	0	8
	2nd-best	1	0	1	0	0	0	0	1	3	2	8
All	Best	0	0	26	0	0	0	0	6	0	0	32
	2nd-best	2	13	2	0	0	0	0	2	7	6	32

Table 3: Empirical coverage ratios of bootstrap confidence intervals of the PAR(1) coefficient at $N = 100$

Note: Data were generated as for Table 1. Empirical coverage ratios are based on 300 iterations. The number of bootstrap iterations is set at 1,000.

(T,α,k)	95%	90%
(4,0.5,1.0)	0.94	0.89
(4,0.5,2.0)	0.94	0.89
(4,0.8,1.0)	0.95	0.90
(4,0.8,2.0)	0.95	0.90
(4,1.0,1.0)	0.93	0.87
(4,1.0,2.0)	0.94	0.87
(4,1.1,1.0)	0.93	0.87
(4,1.1,2.0)	0.93	0.86
(10,0.5,1.0)	0.94	0.88
(10,0.5,2.0)	0.95	0.89
(10,0.8,1.0)	0.95	0.91
(10,0.8,2.0)	0.95	0.91
(10,1.0,1.0)	0.93	0.88
(10,1.0,2.0)	0.93	0.88
(10,1.1,1.0)	0.94	0.88
(10,1.1,2.0)	0.94	0.89