A Bias-Corrected Method of Moments Approach for Estimation of Dynamic Panels*

Alexander Chudik
Federal Reserve Bank of Dallas

M. Hashem Pesaran
USC Dornsife INET, and Trinity College, Cambridge

15 February 2016
Work in progress. Please do not circulate.

Abstract

This paper develops a novel set of bias-corrected moment conditions for estimation of short-T dynamic panel data models. In contrast with GMM, which searchers for instruments that are uncorrelated with the errors but are highly correlated with the target variables (the included regressors), our approach uses the target variables as instruments but corrects for the non-zero correlations between the errors and the instruments. The main advantage of our approach lies in the fact that, by construction, the instruments have maximum correlation with the target variables. The proposed bias-corrected method of moments (BMM) estimation procedure allows for fixed effects, error heteroskedasticity and it is also applicable to the case of heterogeneous slopes. Allowing for slope heterogeneity is achieved by assuming that the coefficient heterogeneity can be characterized in terms of a probability distribution with a finite number of unknown parameters. The BMM estimator is shown to be consistent when the errors are weakly cross-sectionally dependent, and asymptotically normally distributed when the errors are cross-sectionally independent. A semi-parametric estimator of the asymptotic covariance matrix is also proposed. Monte Carlo evidence shows that the proposed estimator works well even for T as small as 4. The results are reasonably robust to modest misspecification in the assumed distribution of parameter heterogeneity. In contrast, the inference based on estimates that incorrectly assume parameter homogeneity can be grossly misleading.

Keywords: Short Dynamic Panels, Coefficient Heterogeneity, Cross Section Dependence and Heteroskedasticity, Bias-Corrected Method of Moments
JEL Classification: C12, C13, C23

*The views expressed in this paper are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of Dallas.

1 Introduction

Important advances have been made in econometric analysis of dynamic panel data models where the time dimension of the panel, T, is small (typically $2 < T \leq 10$), while the cross section dimension (N) is large (often $N \geq 1000$). A number of estimation methods have been developed for such models, known as short-T dynamic panels. These include the instrumental variable and GMM methods (Anderson and Hsiao, 1981 and 1982, Arellano and Bond, 1991, Ahn and Schmidt, 1995, Arellano and Bover, 1995, Blundell and Bond, 1998, Hayakawa, 2012); maximum likelihood methods (Hsiao et al., 2002, Su and Yang, 2007, Hayakawa et al., 2014, and Hayakawa and Pesaran, 2015), and other methods (see for example X-differencing proposed by Han et al., 2010, and a factor analytical method adopted by Bai, 2013). Closely related to single equation dynamic panels are multivariate dynamic panels, or panel VARs. The estimation of short-T panel VARs with homogeneous slopes has been considered, for example, by Holtz-Eakin et al. (1988) and Binder et al. (2005).[1] Early surveys of literature on short-T dynamic panels are provided in Bond (2002), Arellano (2003), Hsiao (2003), and Pesaran (2015, Ch. 27).

The above estimation methods all assume homogeneous slopes, and only allow for intercept heterogeneity. Such a strategy works well when the regressors are strictly exogenous, since the effects of slope heterogeneity get absorbed into the error terms, and if random has little consequence for consistent estimation of the mean of the slope coefficients. But as shown by Pesaran and Smith (1995), such an estimation strategy does not deal with slope heterogeneity in the case of dynamic panels where the regressors are weakly exogenous. Allowing for dynamic heterogeneity is important since firms, households, or institutions tend to converge to their respective equilibrium at different rates, reflecting their different initial states, risk aversion and constraints. When T is sufficiently large, such dynamic heterogeneity can be explored by estimating separate regressions for different cross units and then pooling the outcomes. But when T is short a different estimation strategy is required.

To our knowledge, the only estimation strategy in the literature that allows for coefficients heterogeneity in the case of dynamic short-T panels, is the likelihood approach by Mavroeidis et al.

[1] Literature on the estimation of short-T dynamic homogeneous panels is vast and this brief survey is not comprehensive.

(2015) developed for panel autoregressive (AR) models of order 1, assuming cross sectional independence and normally distributed idiosyncratic errors. In this paper we develop an alternative approach which is applicable more generally, allows for fixed effects which can have an arbitrary degree of correlation with the error terms and the autoregressive coefficients, as well as error heteroskedasticity. To achieve this we assume that the heterogeneity of the autoregressive coefficients can be characterized in terms of a probability distribution with a finite number of unknown parameters, including the mean and the variance of the autoregressive coefficients. The proposed estimators are based on set of estimating equations, derived from the specification of the way the dynamic process is initialized. We consider stationary AR panels where the unit-specific processes have been initialized from an infinite past, and other, possibly non-stationary AR panels, where the unit specific processes are all started from a finite past. Our approach is quite flexible in that any distributional assumption on the autoregressive coefficients can be easily accommodated. We derive theoretical results for a general distribution and present few specific examples. We develop our approach for the case of autoregressive panels of order 1, and we subsequently consider generalization of our method to the vector case, allowing also for higher order lags.

The main idea behind our framework is to allow weakly exogenous variables to serve as instruments even when they are correlated with errors, so long as the extent of the correlation between the instruments and errors is properly taken into account in the derivation of the system of estimating equations. In contrast with GMM, which searchers for instruments that are uncorrelated with the errors but are highly correlated with the target variables (the included regressors), our approach uses the target variables as instruments but corrects for the non-zero correlations between the errors and the instruments. Both approaches employ method of moments, but differ in the way the moments are derived. The main advantage of our approach lies in the fact that, by construction, the instruments have maximum correlation with the target variables. This idea is applicable to a wide variety of estimation problems, where the extent of the correlation of the chosen instruments with errors can be intrinsically derived from the complete specification of the model, as is the case with panel VARs, or spatial autoregressive models. We refer to this estimation strategy as the "bias-corrected method of moments", or BMM for short. In the special case of dynamic panels with homogeneous slopes, BMM has some advantages over existing GMM estimators. Specifically, BMM does not suffer from large N and T bias of GMM estimators in the literature,[2] and the performance of BMM does not suffer when the AR parameter is close to 1.

This paper also derives the asymptotic distribution of the proposed estimators, but under slightly more stringent assumption of cross-sectionally independent yet heteroskedastic idiosyncratic shocks. Heteroskedasticity of shocks affects the asymptotic distribution and introduces nuisance parameters. We show that it is nevertheless possible to consistently estimate the asymptotic covariance matrix semi-parametrically. This also allows us to easily test for coefficient homogeneity or for other aspects of the coefficients' distribution.

Monte Carlo evidence presented in the paper documents satisfactory small sample performance in terms of bias and root mean square errors. The size and the power performance are also very good when innovations are cross-sectionally independent (yet heteroskedastic). The Monte Carlo evidence also shows that the inference based on estimates that (wrongly) assume slope homogeneity can be grossly misleading when slopes are in fact heterogeneous.

The remainder of the paper is organized as follows. Section 2 describes the model and its assumptions and proposes BMM estimators in the case of a stationary panel autoregressive model of order one, or AR(1) for short. The next Section proposes BMM estimators in the case of panel AR(1) model initialized such that the covariance stationarity does not necessarily hold. Section 4 derives asymptotic distributions and proposes a semi-parametric estimator of the asymptotic covariance matrix in the case of AR(1) panels. Section 5 documents satisfactory small sample performance by means of Monte Carlo simulations. Section 6 extends the theory to AR(p) and VAR(p) models. Section 7 concludes the paper. The Appendix presents supplementary derivations and proofs.

Notation

Generic positive finite constant (that does not depend on N) is denoted by K. It can take different values at different instances. All vectors are column vectors denoted by bold lowercase letters. Matrices are denoted by bold uppercase letters. The notation $\rightarrow_p$ denotes convergence in probability, and $\rightarrow_d$ denotes convergence in distribution.

[2] See Alvarez and Arellano (2003) for results on asymptotic bias of GMM estimators when both N and T tend to infinity jointly.

2 Stationary Panel Autoregressive Model

Consider the following stationary panel AR(1) model with fixed effects:

$$y_{it} = c_i + \phi_i y_{i,t-1} + u_{it}, \tag{1}$$

for $i = 1, 2, ..., N$, and $t = ... - 1, 0, 1, 2, ..., T$; where u_{it} is the idiosyncratic shock for the unit i and time period t, and $c_i = (1 - \phi_i)\mu_i$ are the fixed effects that are allowed to be correlated with ϕ_j and $u_{jt'}$, for all i, j and t. The fixed effects can be non-stochastic and bounded or could be stochastic with a well defined probability distribution. It is assumed that only observations for time periods $t = 1, 2, ..., T$ are available, T is fixed and N is large.

We follow the GMM literature and eliminate the fixed effects by first-differencing of (1),

$$\Delta y_{it} = \phi_i \Delta y_{i,t-1} + \Delta u_{it}, \tag{2}$$

for $i = 1, 2, ..., N$ and $t = ..., -1, 0, 1, ..., T$. This is in contrast to the likelihood approach of Mavroeidis et al. (2015) that assumes c_i to be random draws distributed independently of u_{it} and ϕ_i. The treatment of c_i as fixed effects renders our analysis robust to misspecification errors that could arise from particular distributional assumptions about c_i.

Initially, we consider dynamic panels where the processes, $\{y_{it}\}$, are assumed to have started a long time in the past, and accordingly make the following assumption:

ASSUMPTION 1 *(Initialization with an infinite past) The support of ϕ_i lies strictly within the unit circle, and Δy_{it} is given by (2) for $i = 1, 2, ..., N$, and $t = ..., 0, 1, 2, ..., T$.*

Remark 1 *We discuss the extension of the initialization of the dynamic process to the case of a finite past in the next Section. Assumption 1 is an important case, which allows us to present the main idea in a simple way.*

For the idiosyncratic shocks we adopt the following assumption:

ASSUMPTION 2 *(Idiosyncratic shocks) u_{it}, for $i = 1, 2, ..., N$ and $t = ..., 0, 1, ..., T$ are serially independent and distributed such that:*

(i) $E(u_{it}) = 0$,

(ii) $E\left(u_{it}^2 \middle| \sigma_i^2\right) = \sigma_i^2 < K < \infty$, $p\lim_{N\to\infty} N^{-1}\sum_{i=1}^N u_{it}^2 = \sigma^2 < K$,*and*

(iii) $p\lim_{N\to\infty} \bar{u}_t = 0$, *where* $\bar{u}_t = N^{-1}\sum_{i=1}^N u_{it}$.

Remark 2 *Assumption 2 imposes a minimal set of requirements on the idiosyncratic errors. Condition (i) of Assumption 2 requires zero means, condition (ii) requires finite second moments and the existence of probability limit of the cross section average of squared errors, and by assuming that the cross section average of errors tends to* 0 *in probability, condition (iii) restricts the idiosyncratic errors to be weakly cross correlated. (See, for example, Chudik, Pesaran, and Tosetti, 2011).*

The last assumption needed for the derivation of estimating equations concerns the distribution of autoregressive coefficients.

ASSUMPTION 3 *(Autoregressive coefficients)* ϕ_i, *for* $i = 1, 2, ..., N$, *are random draws from a known distribution,* $G(\boldsymbol{\theta})$, *and support* S_ϕ, *where* $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_k)' \in \mathcal{C}_k$ *is a* $k \times 1$ *vector of unknown parameters that do not depend on* N, *and* $\mathcal{C}_k$ *is a finite* k*-dimensional compact set. The autoregressive coefficients,* ϕ_i, *are distributed independently of innovations* $u_{i't}$ *for any* $i, i' = 1, 2, .., N$, *and any* $t = ..., -1, 0, 1, ..., T$.

Remark 3 *Assumption 3 postulates that the distribution of the autoregressive parameters is known up to a* k *unknown parameters collected in the vector* $\boldsymbol{\theta}$. k *is assumed to be finite. Note that under Assumption 3, the support* S_ϕ *lies within the unit circle, which ensures that all moments of* ϕ_i *exist.*

The main object of interest is the vector of unknown parameters $\boldsymbol{\theta}$. We provide some examples below.

Example 1 *The simplest example to consider is a degenerate distribution, that is homogeneous autoregressive coefficients* $\phi_i = \theta$, *in which case the object of interest is the homogeneous AR parameter* θ.

Example 2 *Another possibility is to consider uniformly distributed autoregressive parameters,* $\phi_i \sim IIDU[\phi_{\min}, \phi_{\max}]$, *with* $-1 < \phi_{\min} \leq \phi_{\max} < 1$. *In this case, it is more convenient to use the mean* *and standard deviation of the distribution as unknown parameters of interest, namely* $\boldsymbol{\theta} = (\theta_\mu, \theta_\sigma)$*, where* $\theta_\mu = E(\phi_i)$ *and* $\theta_\sigma^2 = Var(\phi_i)$*.* $\phi_{\min}$ *and* $\phi_{\max}$ *can be recovered from* $\boldsymbol{\theta}$ *as follows:* $\phi_{\min} = \theta_\mu - \sqrt{12}\theta_\sigma/2$, $\phi_{\max} = 2\theta_\mu - \phi_{\min}$.

Example 3 *Alternative example is to consider beta distributed autoregressive parameters,* $\phi_i \sim IIDBeta\left[\phi_\alpha, \phi_\beta\right]$*. As in the previous example, it is more convenient to use the mean and standard deviation of the distribution as unknown parameters of interest, namely* $\boldsymbol{\theta} = (\theta_\mu, \theta_\sigma)$*, where* $\theta_\mu = E(\phi_i)$ *and* $\theta_\sigma^2 = Var(\phi_i)$*.* ϕ_α *and* ϕ_β *can be recovered from* $\boldsymbol{\theta}$ *as follows:* $\phi_\alpha = \theta_\mu\left[\theta_\mu(1-\theta_\mu)/\theta_\sigma^2 - 1\right]$ *and* $\phi_\beta = (1-\theta_\mu)\left[\theta_\mu(1-\theta_\mu)/\theta_\sigma^2 - 1\right]$*.*

Assumptions 1-3 are not sufficient for the parameters in $\boldsymbol{\theta}$ to be identified and we shall postulate additional requirements that would ensure the identification of $\boldsymbol{\theta}$ below.

By focussing on the first-differenced formulation (2), our approach is closely related to the IV or GMM procedure used in the literature, with this important difference that we also use $\Delta y_{i,t-1}$ as an "instrument", which is regarded as being "invalid" in the GMM literature since it is correlated with Δu_{it}. By using $\Delta y_{i,t-1}$ as an instrument we then adjust the resultant moment condition for the non-zero correlation between $\Delta y_{i,t-1}$ and Δu_{it}. The advantage of using $\Delta y_{i,t-1}$ as an instrument lies in the fact that by construction it has a maximum correlation with the target variable (itself), so long as we are able to correct for the bias that arises due to $Cov(\Delta y_{i,t-1}, \Delta u_{it}) \neq 0$. To summarize, GMM searchers for instruments that are uncorrelated with the errors but are highly correlated with the target variables, the included regressors. In contrast, our approach uses the target variables as instruments but corrects for the non-zero correlations between the errors and the instruments. Both approaches employ method of moments, but differ in the way the moments are derived. However, it is clear that our proposed method can only be applied if the specification of estimation problem is sufficiently rich such that an expression for $Cov(\Delta y_{i,t-1}, \Delta u_{it})$ can be derived. The standard IV or GMM do not have this problem, but can lead to the weak instrumental variable problem as they can not ensure that the correlations between the instruments and the targets are sufficiently large.

To illustrate the basic idea of the paper, we consider the homogeneous case where $\phi_i = \phi$, and using $\Delta y_{i,t-1}$ as an instrument, we note that

$$E\left[\Delta y_{it}\Delta y_{i,t-1} - \phi\left(\Delta y_{i,t-1}\right)^2\right] = E\left(\Delta u_{it}\Delta y_{i,t-1}\right), \text{ for } t = 2, 3, ..., T \tag{3}$$

To derive $E(\Delta u_{it}\Delta y_{i,t-1})$, using (2) to solve for Δy_{it}, we have

$$\begin{aligned}\Delta y_{it} &= \sum_{\ell=0}^{\infty}\phi_i^{\ell}\Delta u_{i,t-\ell} \\ &= u_{it}+\sum_{\ell=1}^{\infty}\phi_i^{\ell-1}(\phi_i-1)\,u_{i,t-\ell}. \end{aligned} \tag{4}$$

Then it is readily seen that

$$E(\Delta u_{it}\Delta y_{i,t-1}) = -E\left(\sigma_i^2\right),$$

and hence

$$E\left[\Delta y_{it}\Delta y_{i,t-1}-\phi\left(\Delta y_{i,t-1}\right)^2\right] = -E\left(\sigma_i^2\right) \tag{5}$$

Furthermore

$$E\left[\left(\Delta y_{it}\right)^2\right] = \left(\frac{2}{1+\phi}\right)E\left(\sigma_i^2\right),$$

which if used in (5) yields the following bias-corrected moment condition (BMM)

$$E\left[\Delta y_{it}\Delta y_{i,t-1}-\phi\left(\Delta y_{i,t-1}\right)^2\right] = -\left(\frac{1+\phi}{2}\right)E\left[\left(\Delta y_{it}\right)^2\right].$$

Suppose that observations Δy_{it}, $t=2,3,...,T$ and $i=1,2,...,N$ are available to estimate ϕ. Then the sample moment associated to the above BMM condition is given by

$$\sum_{i=1}^{N}\sum_{t=2}^{T}\Delta y_{it}\Delta y_{i,t-1}-\hat{\phi}_{BMM}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(\Delta y_{i,t-1}\right)^2 = -\left(\frac{1+\hat{\phi}_{BMM}}{2}\right)\left(\frac{T-1}{T}\right)\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\Delta y_{it}\right)^2,$$

where $\hat{\phi}_{BMM}$ denotes the bias-corrected MM estimator of ϕ. An explicit expression for $\hat{\phi}_{BMM}$ is given by

$$\hat{\phi}_{BMM} = \frac{\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T-2}\sum_{t=3}^{T}\Delta y_{it}\Delta y_{i,t-1}+\frac{1}{2}\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T-1}\sum_{t=2}^{T}\left(\Delta y_{it}\right)^2}{\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T-2}\sum_{t=3}^{T}\left(\Delta y_{i,t-1}\right)^2-\frac{1}{2}\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T-1}\sum_{t=2}^{T}\left(\Delta y_{it}\right)^2}. \tag{6}$$

This estimator is robust to heteroskedastic errors of an arbitrary form, so long as $E\left(\sigma_i^2\right)$ exists, and does not suffer from the weak instrument problem even if ϕ is close to unity. Also, it does not suffer from the large N and T bias of GMM estimators of ϕ discussed in the literature. We shall consider the homogenous slope case in further detail below.

Having illustrated the basic idea we now return to the heterogeneous case which is more complicated since we also need to take account of the correlation of ϕ_i and $\Delta y_{i,t-1}$, considering the dependence of $\Delta y_{i,t-1}$ on ϕ_i. We also need to make the further assumption that $\sigma^2 = E\left(\sigma_i^2\right)$. Note that this is not the same as assuming that the errors are cross sectionally homoskedastic, namely $\sigma_i^2 = \sigma^2$, for all i, but that σ_i^2 whose realizations differ across i are random draws from a distribution with the common mean $E(\sigma_i^2)$. Then, using (4), we obtain

$$\begin{aligned} E\left(\Delta y_{it}\Delta y_{i,t-s}\right) &= E\left[\left(u_{it} + \sum_{\ell=1}^{\infty}\phi_i^{\ell-1}\left(\phi_i - 1\right)u_{i,t-\ell}\right)\left(u_{i,t-s} + \sum_{\ell=1}^{\infty}\phi_i^{\ell-1}\left(\phi_i - 1\right)u_{i,t-\ell-s}\right)\right] \\ &= 2\sigma^2 E\left(\frac{\phi_i^s}{1+\phi_i}\right), \end{aligned} \tag{7}$$

for $t = 2, 3, ..., T$ and $s = 0, 1, ..., t-2$. Let us define the function

$$\omega_s = 2\sigma^2 E\left(\frac{\phi_i^s}{1+\phi_i}\right),$$

for $s = 0, 1, ..., t-2$, which, in the stationary case, does not depend on t. Note that, under Assumption 3, $E\left[\phi_i^s/\left(1+\phi_i\right)\right]$ is a function of the unknown parameters in $\boldsymbol{\theta}$. We denote this function as $h_s\left(\boldsymbol{\theta}\right)$, and note that it can be expressed in terms of the moments of ϕ_i as

$$h_s\left(\boldsymbol{\theta}\right) = E\left(\frac{\phi_i^s}{1+\phi_i}\right) = m_s\left(\boldsymbol{\theta}\right) + \sum_{\ell=1}^{\infty}\left(-1\right)^{\ell} m_{s+\ell}\left(\boldsymbol{\theta}\right), \tag{8}$$

for $s = 0, 1, ..., t-2$, where we define

$$m_{\ell}\left(\boldsymbol{\theta}\right) = \begin{cases} 1, & \text{for } \ell = 0 \\ E\left(\phi_i^{\ell}\right), & \text{for } \ell = 1, 2, ... \end{cases}. \tag{9}$$

The series expression (8) is useful for the empirical implementation of the proposed approach, where it is sufficient to specify the moment functions $m_{\ell}\left(\boldsymbol{\theta}\right)$ only. But more compact expressions for $h_s\left(\boldsymbol{\theta}\right)$ can be obtained when ϕ_i is specified to follow some known distributions such as uniform or beta distributions. See Example 4 below.

Suppose that we have the observations Δy_{it}, $t = 2, ..., T$ and $i = 1, 2, ..., N$, and consider the sample moments defined by

$$\hat{\omega}_{s,NT}=\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T-s-1}\sum_{t=s+2}^{T}\Delta y_{it}\Delta y_{i,t-s}\right),\ \text{for } s=0,1,..T-2 \tag{10}$$

The following lemma establishes convergence of the sample moments for a given T and as $N\to\infty$.

Lemma 1 *Let y_{it} be given by (2) for $t=...0,1,2,...T$, and $i=1,2,...,N$, and suppose Assumptions 1-3 hold. Then*

$$\hat{\omega}_{s,NT}\xrightarrow{p}\omega_s,\ \text{for } t=2,3,...,T \text{ and } s=0,1,...,t-2, \tag{11}$$

as $N\to\infty$ and T is fixed, where

$$\omega_s=2\sigma^2E\left(\frac{\phi_i^s}{1+\phi_i}\right)=2\sigma^2h_s\left(\boldsymbol{\theta}\right),$$

in which $\sigma^2=p\lim_{N\to\infty}N^{-1}\sum_{i=1}^{N}u_{it}^2$, and $h_s\left(\boldsymbol{\theta}\right)=E\left[\phi_i^s/\left(1+\phi_i\right)\right]$ can be written as (8).

Proof. Under Assumptions 1-2, we have $N^{-1}\sum_{i=1}^{N}u_{it}u_{i,t-s}\xrightarrow{p}0$ as $N\to\infty$, for any $t=...,0,1,...,T$ and $s>0$. Since also ϕ_i is independently distributed of u_{jt} for any i,j,t, it follows that $N^{-1}\sum_{i=1}^{N}\Delta y_{it}\Delta y_{i,t-s}\xrightarrow{p}\omega_s$. This in turn establishes (11). ■

(11) can be explored for a consistent estimation of the k unknown parameters in the vector $\boldsymbol{\theta}$ and the unknown parameter σ^2. Let $\boldsymbol{\psi}=\left(\boldsymbol{\theta}',\sigma^2\right)'$ collect the $k+1$ unknown parameters, and define

$$\begin{aligned} g_{NT}\left(\boldsymbol{\psi},s\right) &= \hat{\omega}_{s,NT}-\omega_s \\ &= \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T-s-1}\sum_{t=s+2}^{T}\Delta y_{it}\Delta y_{i,t-s}\right)-2\sigma^2h_s\left(\boldsymbol{\theta}\right), \end{aligned}$$

where we opt for a simplified notation $g_{NT}\left(\boldsymbol{\psi},s\right)\equiv g\left(\boldsymbol{\psi},s,\{y_{it}\}_{i=1,t=2}^{N,T}\right)$. Moreover, let us define function $\mathbf{g}_{NT}\left(\boldsymbol{\psi}\right)$ by stacking the individual $T-1$ functions $g_{NT}\left(\boldsymbol{\psi},s\right)$, for $s=0,1,...,T-2$, namely

$$\underset{T-1\times1}{\mathbf{g}_{NT}\left(\boldsymbol{\psi}\right)}=\left[g_{NT}\left(\boldsymbol{\psi},0\right),g_{NT}\left(\boldsymbol{\psi},1\right),...,g_{NT}\left(\boldsymbol{\psi},T-2\right)\right]', \tag{12}$$

and similarly let function $\boldsymbol{\zeta}\left(\boldsymbol{\psi}\right)$ by defined by stacking the $T-1$ individual functions $\omega_s=2\sigma^2h_s\left(\boldsymbol{\theta}\right)$, for $s = 0, 1, ..., T-2$, namely

$$\underset{T-1\times 1}{\boldsymbol{\xi}_T}(\boldsymbol{\psi}) = 2\sigma^2 \left[h_0(\boldsymbol{\theta}), h_1(\boldsymbol{\theta}), ..., h_{T-2}(\boldsymbol{\theta})\right]'. \tag{13}$$

Lemma 1 implies

$$p \lim_{N\to\infty} \mathbf{g}_{NT}(\boldsymbol{\psi}) = \mathbf{0}, \tag{14}$$

which can be used for estimation, provided that $\boldsymbol{\psi}$ is identified. A sufficient condition for the identification of $\boldsymbol{\psi}$ from $\boldsymbol{\xi}_T(\boldsymbol{\theta})$ is provided below.

ASSUMPTION 4 *(Identification) Let $\boldsymbol{\xi}_T(\boldsymbol{\psi})$ defined in (13) be continuous in $\boldsymbol{\psi} = \left(\boldsymbol{\theta}', \sigma^2\right)'$, and for any $\boldsymbol{\psi}_1, \boldsymbol{\psi}_2 \in \Omega$ such that $\boldsymbol{\psi}_1 \neq \boldsymbol{\psi}_2$ we have*

$$\boldsymbol{\xi}_T(\boldsymbol{\psi}_1) \neq \boldsymbol{\xi}_T(\boldsymbol{\psi}_2),$$

where $\Omega = \mathcal{C}_k \times \mathbb{R}_0^+$ is the space of unknown parameters.

Remark 4 *Assumption 4 states that, for a given T, function $\boldsymbol{\xi}_T(\boldsymbol{\psi})$ is invertible, that is the unknown parameters could be identified from the function $\boldsymbol{\xi}_T(\boldsymbol{\psi})$. Note that at most $T-1$ parameters could be identified since there are at most $T-1$ ($s = 0, 1, 2, ..., T-2$) independent functions in $\boldsymbol{\xi}_T(\boldsymbol{\psi})$.*

Example 4 *As an illustrative example consider the case of uniformly distributed coefficients, $\phi_i \sim IIDU[0, \theta]$, $0 < \theta \leq 1-\epsilon$, for some $0 < \epsilon < 1$, and $T = 3$. In this case we have two unknown parameters, $\boldsymbol{\psi} = \left(\theta, \sigma^2\right)' \in \mathcal{C}_k \times \mathbb{R}_0^+$, $\mathcal{C}_k \equiv [0, 1-\epsilon]$, and $m_\ell(\theta)$, for $\ell > 0$, is given by*

$$m_\ell(\theta) = \frac{1}{\ell+1}\theta^\ell.$$

In this example, function $\boldsymbol{\xi}_3(\boldsymbol{\psi})$ is 2×1 and is given by

$$\boldsymbol{\xi}_3(\boldsymbol{\psi}) = 2\sigma^2 \begin{pmatrix} h_0(\theta) \\ h_1(\theta) \end{pmatrix}, \tag{15}$$

and $h_0(\theta)$ *and* $h_1(\theta)$ *can be derived directly as (recall that* $\phi_i \sim IIDU[0,\theta]$*)*

$$h_0(\theta) = E\left(\frac{1}{1+\phi}\right) = \frac{1}{\theta}\int_0^{\theta} \frac{1}{1+\phi} d\phi = \theta^{-1}\ln(1+\theta),$$

and

$$h_1(\theta) = \frac{1}{\theta}\int_0^{\theta} \frac{\phi}{1+\phi} d\phi = 1 - \theta^{-1}\ln(1+\theta).$$

To see that θ *and* σ^2 *are identified, consider the ratio of the first and second term in* $\boldsymbol{\xi}_3(\boldsymbol{\psi})$*, namely*

$$f(\theta) = 2\sigma^2 h_0(\theta)/2\sigma^2 h_1(\theta) = \frac{\ln(1+\theta)}{\theta - \ln(1+\theta)}. \tag{16}$$

It is also easily seen that

$$f'(\theta) = \frac{\theta - (1+\theta)\ln(1+\theta)}{(1+\theta)\left[\theta - \ln(1+\theta)\right]^2} < 0, \text{ for all } \theta > -1.$$

Therefore, $f(\theta)$ *as a function of* θ *is strictly monotonic on* $[0,1]$*, and hence* θ *is identified.* σ^2 *can now be identified using either of the two elements of* $\boldsymbol{\xi}_3(\boldsymbol{\psi})$*.*

We propose the following consistent estimation of unknown parameters in the vector $\boldsymbol{\psi}$.

Theorem 1 *Let* y_{it} *be given by (2) for* $t = ...0,1,2,...T$*, and* $i = 1,2,...,N$*, suppose Assumptions 1-4 hold, and define* $\hat{\boldsymbol{\psi}}_{BMM}$ *as*

$$\hat{\boldsymbol{\psi}}_{BMM} = \arg\min_{\boldsymbol{\psi}\in\Omega} \left\|\mathbf{g}_{NT}(\boldsymbol{\psi})\right\|, \tag{17}$$

where $\mathbf{g}_{NT}(\boldsymbol{\psi})$ *is defined in (12), and* $\|.\|$ *denotes the Euclidean norm. Then* $\hat{\boldsymbol{\psi}}_{BMM}$ *satisfies*

$$p\lim_{N\to\infty}\left\|\hat{\boldsymbol{\psi}}_{BMM} - \boldsymbol{\psi}\right\| = 0. \tag{18}$$

Proof. The dimension of the function $\mathbf{g}_{TM}(\boldsymbol{\psi})$ in (12) is finite since T is finite. Since also $\boldsymbol{\xi}_{TM}(\boldsymbol{\psi})$ is continuous and invertible, it follows that $p\lim_{N\to\infty}\left\|\hat{\boldsymbol{\psi}}_{BMM} - \boldsymbol{\psi}\right\| = 0$ if and only if $p\lim_{N\to\infty}\mathbf{g}_{TM}(\boldsymbol{\psi}) = 0$ for all $s = 0,1,2,...,T-2$. Noting that (14) holds completes the proof. ■

Remark 5 *In the case of Example 4 with* $T = 3$*, the vector of bias-corrected moment conditions* *are given by*

$$\mathbf{g}_{N,3}(\boldsymbol{\psi}) = \begin{bmatrix} 2\sigma^2\theta^{-1}\ln(1+\theta) - N^{-1}\sum_{i=1}^{N}\Delta y_{i2}^2 \\ 2\sigma^2\left[1-\theta^{-1}\ln(1+\theta)\right] - N^{-1}\sum_{i=1}^{N}\Delta y_{i2}\Delta y_{i3} \end{bmatrix}, \tag{19}$$

The bias-corrected MM estimator of $\boldsymbol{\psi} = \left(\theta, \sigma^2\right)'$ *is thus given by* $\hat{\boldsymbol{\psi}}_{BMM} = \arg\min_{\boldsymbol{\psi}\in\Omega}\left\|\mathbf{g}_{N,3}(\boldsymbol{\psi})\right\|$. *Given that there are no excess moment conditions the bias-corrected estimators of* θ *and* σ^2 *are given*

$$\widehat{\theta}_{BMM} = f^{-1}\left(\frac{N^{-1}\sum_{i=1}^{N}\Delta y_{i2}^2}{N^{-1}\sum_{i=1}^{N}\Delta y_{i2}\Delta y_{i3}}\right),$$

where $f^{-1}(\theta)$ *is the inverse of* $f(\theta)$ *defined by (16). The BMM estimator of* σ^2 *can now be obtained using either of the two equations given by*

$$2\hat{\sigma}_{BMM}^2\left[\hat{\theta}_{BMM}^{-1}\ln\left(1+\hat{\theta}_{BMM}\right), 1-\hat{\theta}_{BMM}^{-1}\ln\left(1+\hat{\theta}_{BMM}\right)\right] = \left(\hat{\omega}_{0,NT}, \hat{\omega}_{1,NT}\right),$$

where $\hat{\omega}_{s,NT}$, $s = 0, 1$, *are defined by (10).*

There are different ways how one could combine the individual consistent estimating equations in $\mathbf{g}_{TM}(\boldsymbol{\psi})$ to construct an estimator of $\boldsymbol{\psi}$, when the number of estimating equations is larger than the number of unknown parameters. Ideally, this would be done in view of efficiency of the resulting estimator. Theorem 1 established that a consistent estimation is possible (under rather general assumptions) and without specifying any requirements on $G(\boldsymbol{\theta})$ besides the identification (and constrained support to ensure moments exist). Minimizing Euclidean norm of $\mathbf{g}_{TM}(\boldsymbol{\psi})$ seems to be a good starting point, but note that the same consistency result holds for any other norm. For example, one could also choose to minimize $\left\|\mathbf{g}_{TM}(\boldsymbol{\psi})\right\|_W = \left\|\mathbf{W}'\mathbf{g}_{TM}(\boldsymbol{\psi})\right\|$, for some suitable $T-1\times(k+1)$ weights matrix $\mathbf{W}$. The asymptotic distribution of the resulting estimator denoted as $\hat{\boldsymbol{\psi}}_W$ and the asymptotic distribution of $\hat{\boldsymbol{\psi}}_{BMM}$ is studied in Section 4. We note that utilizing all equations need not necessarily be the best for the finite sample performance and choices of $\mathbf{W}$ where some of the rows are zero vectors could also be considered.

Finally, it is worth noting that the bias-corrected MM estimator of ϕ for the homogeneous case, defined by (6), can also be written in terms of the moment estimators, $\hat{\omega}_{s,NT}$, as

$$\hat{\phi}_{BMM} = \frac{\hat{\omega}_{2,NT} + \frac{1}{2}\hat{\omega}_{0,NT}}{\hat{\omega}_{1,NT} - \frac{1}{2}\hat{\omega}_{0,NT}}.$$

3 Nonstationary Panel Autoregressive Models

The assumption of stationarity in the previous Section is an interesting and important case, which allowed us to motivate the main idea in a simple way. But this assumption might not necessarily hold in empirical applications. Assumption 1 together with Assumption 2 are responsible for $E(\Delta y_{it}\Delta y_{i,t-s})$ to depend only on s and not on t. Such covariance stationarity can be easily tested using the results derived in the next Section. When the null of covariance stationarity is rejected, it is well possible that the process $\{y_{it}\}$ started from a finite past. This section considers nonstationary panel autoregressive models initialized from a finite past.

We continue to consider the panel AR(1) model given by (1) for $i = 1, 2, ..., N$, but we assume (1) holds for $t = -M, -M + 1..., 1, 2, ..., T$ and the starting values are $y_{i,-M-1}$; where $M \geq -1$. As before, only observations for time periods $t = 1, 2, ..., T$ are available, T is fixed and N is large. We postulate the starting values given by

$$y_{i,-M-1} - \mu_i = \eta_i. \tag{20}$$

This ensures the starting values of the first-difference representation of data in (2) satisfy

$$\begin{aligned} \Delta y_{i,-M} &= y_{i,-M} - y_{i,-M-1} \\ &= (1 - \phi_i)\mu_i - (1 - \phi_i)y_{i,-M-1} + u_{i,-M} \\ &= \eta_i(\phi_i - 1) + u_{i,-Mi}, \end{aligned} \tag{21}$$

for $i = 1, 2, ..., N$, where we have first substituted (1) for $y_{i,-M}$, and then (20) for $y_{i,-M-1}$. It is clear that (2) with starting values (21) do not depend on μ_i and consequently the fixed effects $c_i = (1 - \phi_i)\mu_i$ are still arbitrary.

(20) allows starting values to deviate from μ_i and these deviations are denoted as η_i. Un-

fortunately, in short-T dynamic panels with slope heterogeneity, it does not seem possible to treat the starting values together with fixed effects as arbitrary (without assuming that the distribution of errors is known); and some assumption on η_i must be imposed. As it is always the case in the literature, we assume that η_i is independently distributed from shocks u_{it} for $t = -M, -M+1..., 1, 2, ..., T$, and we assume the following assumption instead of Assumption 1.

ASSUMPTION 5 *(Initialization with finite past) The process Δy_{it}, for $i = 1, 2, ..., N$, and $t = -M, -M+1, ..., 1, 2, ..., T$, is given by (2) with a starting values $\Delta y_{i,-M}$ given by (21), for $i = 1, 2, ..., N$, where $-1 \leq M < \infty$ is fixed. η_i is independently distributed of ϕ_j and u_{jt} for any i, j and t. In addition, $p\lim_{N\to\infty} N^{-1}\sum_{i=1}^{N}\eta_i^2 = E\left(\eta_i^2\right) = \delta < \infty$ and M is known.*

Remark 6 *Assumption 5 postulates that the process y_{it} started from a finite past ($M < \infty$). M is assumed to be known for simplicity of exposition. This assumption can be extended to the case when M is finite and fixed but otherwise unknown. Another possible extension is to allow the starting period to differ across i, say M_i, and to assume that M_i is stochastic.*

Remark 7 *The independence of η_i from ϕ_i stated in Assumption 5 could be relaxed at the expense of introducing additional unknown parameters that need to be estimated. One possibility to allow for η_i to depend on ϕ_i is to assume (linear dependence)*

$$\eta_i = b_1\phi_i + \upsilon_i,$$

where b_1 is an unknown parameter to be estimated, and υ_i is independent of ϕ_i and satisfies $p\lim_{N\to\infty} N^{-1}\sum_{i=1}^{N}\upsilon_i^2 = \delta_\upsilon < \infty$. Under these assumptions, we have

$$p\lim_{N\to\infty} N^{-1}\sum_{i=1}^{N}\eta_i^2\phi_i^s = \delta_\upsilon\left(b_1 \cdot p\lim_{N\to\infty} N^{-1}\sum_{i=1}^{N}\phi_i^{s+1}\right),$$

for $s = 1, 2, ..., 2T + 2M + 2$.

Remark 8 *Non-linear dependence of η_i and ϕ_i could also be considered, such as*

$$\eta_i = \sum_{\ell=1}^{p_\eta} b_\ell\phi_i^\ell + \upsilon_i,$$

for some finite $p_\ell > 1$, in which case we obtain (assuming υ_i satisfies the same assumptions as in the previous remark)

$$p \lim_{N\to\infty} N^{-1} \sum_{i=1}^{N} \eta_i^2 \phi_i^s = \delta_\upsilon \sum_{\ell=1}^{p_\eta} \left(b_\ell \cdot p \lim_{N\to\infty} N^{-1} \sum_{i=1}^{N} \phi_i^{s+\ell} \right),$$

for $s = 1, 2, ..., 2T + 2M + 2$.

Remark 9 *Assumption 5 does not restrict the support of ϕ_i.*

Since Assumption 5 no longer restricts the support of ϕ_i, we need an additional assumption that complements Assumption 3 to ensure existence of higher moments of ϕ_i.

ASSUMPTION 6 *(Moments of autoregressive coefficients) The first $2M+2T$ moments of ϕ_i are finite.*

Under Assumption 5, it is useful to write Δy_{it} as

$$\begin{aligned}
\Delta y_{it} &= \sum_{\ell=0}^{t+M-1} \phi_i^\ell \Delta u_{i,t-\ell} + \phi_i^{t+M} \left[\delta_i (\phi_i - 1) + u_{i,-M}\right] \\
&= u_{it} + \sum_{\ell=1}^{t+M} \phi_i^{\ell-1} (\phi_i - 1) u_{i,t-\ell} + \phi_i^{t+M} (\phi_i - 1) \eta_i.
\end{aligned}$$

Consider the limiting behavior of cross section averages of the product $\Delta y_{it} \Delta y_{i,t-s}$ for $t = 2, 3, ..., T$ and $s = 0, 1, 2, ..., t-2$. For $s = 0$, we have

$$N^{-1} \sum_{i=1}^{N} \Delta y_{it}^2 = N^{-1} \sum_{i=1}^{N} \left[u_{it} + \sum_{\ell=1}^{t+M} \phi_i^{\ell-1} (\phi_i - 1) u_{i,t-\ell} + \phi_i^{t+M} (\phi_i - 1) \eta_i \right]^2,$$

and taking probability limit (as $N \to \infty$) under Assumptions 2-3, 5 and 6, we obtain

$$p \lim_{N\to\infty} N^{-1} \sum_{i=1}^{N} \Delta y_{it}^2 = E \left\{ \frac{2\sigma_i^2}{1+\phi_i} + \frac{1-\phi_i}{1+\phi_i} \phi_i^{2(M+t)} \left[\left(1-\phi_i^2\right) \eta_i^2 - \sigma_i^2 \right] \right\} \tag{22}$$

for $t = 2, 3, ..., T$. For easy empirical implementation of the proposed approach, it is useful to express the right side of (22), under Assumptions 2-3, 5 and 6, as

$$E\left\{\frac{2\sigma_i^2}{1+\phi_i}+\frac{1-\phi_i}{1+\phi_i}\phi_i^{2(M+t)}\left[\left(1-\phi_i^2\right)\eta_i^2-\sigma_i^2\right]\right\}=$$
$$\sigma^2\left(1+\sum_{\ell=1}^{t+M}\left[m_{2\ell}\left(\boldsymbol{\theta}\right)-2m_{2\ell-1}\left(\boldsymbol{\theta}\right)+m_{2\ell-2}\left(\boldsymbol{\theta}\right)\right]\right) \tag{23}$$
$$+\delta\left[m_{2t+2M+2}\left(\boldsymbol{\theta}\right)-2m_{2t+2M+1}\left(\boldsymbol{\theta}\right)+m_{2t+2M}\left(\boldsymbol{\theta}\right)\right] \tag{24}$$

where $m_\ell\left(\boldsymbol{\theta}\right)$ for $\ell=0,1,2,...,2T+2M+2$ is defined in (9).

Similarly, for $1\leq s<t-1$, we obtain

$$\begin{aligned} p\lim_{N\to\infty}N^{-1}\sum_{i=1}^N\Delta y_{it}\Delta y_{i,t-s} &= \sigma^2\left(m_s\left(\boldsymbol{\theta}\right)-m_{s-1}\left(\boldsymbol{\theta}\right)+\sum_{\ell=1}^{t-s+M}\left[m_{2\ell+s}\left(\boldsymbol{\theta}\right)-2m_{2\ell+s-1}\left(\boldsymbol{\theta}\right)+m_{2\ell+s-2}\left(\boldsymbol{\theta}\right)\right]\right) \\ &+\delta\left[m_{2t+2M-s+2}\left(\boldsymbol{\theta}\right)-m_{t+M+1}\left(\boldsymbol{\theta}\right)-m_{t-s+M+1}\left(\boldsymbol{\theta}\right)+m_{2t+2M-s}\left(\boldsymbol{\theta}\right)\right], \end{aligned} \tag{25}$$

where $m_\ell\left(\boldsymbol{\theta}\right)=0$ for $\ell=-1$, and $m_\ell\left(\boldsymbol{\theta}\right)$ for $\ell=1,2,...$ is defined in (9). It is easily seen that (25) reduces to (24) for $s=0$. Finally, let us define the following moment function

$$\begin{aligned} \xi\left(\boldsymbol{\theta},\sigma^2,\delta,t,s,M\right) &= \sigma^2\left[m_s\left(\boldsymbol{\theta}\right)-m_{s-1}\left(\boldsymbol{\theta}\right)+\sum_{\ell=1}^{t+M-s}\left[m_{s+2\ell}\left(\boldsymbol{\theta}\right)-2m_{s+2\ell-1}\left(\boldsymbol{\theta}\right)+m_{s+2\ell-2}\left(\boldsymbol{\theta}\right)\right]\right] \\ &+\delta\left[m_{2t+2M-s+2}\left(\boldsymbol{\theta}\right)-m_{t+M+1}\left(\boldsymbol{\theta}\right)-m_{t-s+M+1}\left(\boldsymbol{\theta}\right)+m_{2t+2M-s}\left(\boldsymbol{\theta}\right)\right], \end{aligned} \tag{26}$$

for $t=2,3,...,T$ and $s=0,1,2,...,t-2$. Consider the following estimating equations,

$$g\left(\boldsymbol{\theta},\sigma^2,\delta,t,s,M,\{y_{it}\}_{i=1,t=2}^{N,T}\right)=\xi\left(\boldsymbol{\theta},\sigma^2,\delta,t,s,M\right)-N^{-1}\sum_{i=1}^N\Delta y_{it}\Delta y_{i,t-s}, \tag{27}$$

for $t=2,3,...,T$ and $s=0,1,2,...,t-2$. Note that by construction we have,

$$p\lim_{N\to\infty}g\left(\boldsymbol{\theta},\sigma^2,\delta,t,s,M,\{y_{it}\}_{i=1,t=2}^{N,T}\right)=0. \tag{28}$$

We stack $\sum_{t=2}^{T} t-1 = T(T-1)/2$ estimating equations into one vector function,

$$\underset{T(T-1)/2\times 1}{\mathbf{g}}\left(\boldsymbol{\theta},\sigma^2,\delta,T,M,\{y_{it}\}_{i=1,t=2}^{N,T}\right) = \begin{pmatrix} g\left(\boldsymbol{\theta},\sigma^2,\delta,t=2,s=0,M,\{y_{it}\}_{i=1,t=2}^{N,T}\right) \\ g\left(\boldsymbol{\theta},\sigma^2,\delta,t=3,s=0,M,\{y_{it}\}_{i=1,t=2}^{N,T}\right) \\ g\left(\boldsymbol{\theta},\sigma^2,\delta,t=3,s=1,M,\{y_{it}\}_{i=1,t=2}^{N,T}\right) \\ g\left(\boldsymbol{\theta},\sigma^2,\delta,t=4,s=0,M,\{y_{it}\}_{i=1,t=2}^{N,T}\right) \\ \vdots \\ g\left(\boldsymbol{\theta},\sigma^2,\delta,t=T,s=T-2,M,\{y_{it}\}_{i=1,t=2}^{N,T}\right) \end{pmatrix} = \mathbf{0}. \quad (29)$$

Similarly, define $T(T-1)/2$ vector $\boldsymbol{\xi}\left(\boldsymbol{\theta},\sigma^2,\delta,T,M\right)$ based on stacking individual functions $\xi\left(\boldsymbol{\theta},\sigma^2,\delta,t,s,M\right)$ in a similar way as $\mathbf{g}\left(\boldsymbol{\theta},\sigma^2,\delta,T,M,\{y_{it}\}_{i=1,t=2}^{N,T}\right)$, namely

$$\boldsymbol{\xi}\left(\boldsymbol{\theta},\sigma^2,\delta,T,M\right) = \begin{pmatrix} \xi\left(\boldsymbol{\theta},\sigma^2,\delta,t=2,s=0,M\right) \\ \xi\left(\boldsymbol{\theta},\sigma^2,\delta,t=3,s=0,M\right) \\ \xi\left(\boldsymbol{\theta},\sigma^2,\delta,t=3,s=1,M\right) \\ \xi\left(\boldsymbol{\theta},\sigma^2,\delta,t=4,s=0,M\right) \\ \vdots \\ \xi\left(\boldsymbol{\theta},\sigma^2,\delta,t=T,s=T-2,M\right) \end{pmatrix}. \quad (30)$$

There are $k+2$ unknown parameters collected in the vector $\boldsymbol{\psi} = \left(\boldsymbol{\theta}',\sigma^2,\delta\right)'$, when M is known and finite. On the other hand, there are $T(T-1)/2$ consistent estimating equations. As before, denote the space of the unknown parameters as Ω. Also, let us write $\boldsymbol{\xi}\left(\boldsymbol{\theta},\sigma^2,\delta,T,M\right)$ and $\mathbf{g}\left(\boldsymbol{\theta},\sigma^2,\delta,T,M,\{y_{it}\}_{i=1,t=2}^{N,T}\right)$ simply as $\boldsymbol{\xi}_{TM}(\boldsymbol{\psi})$ and $\mathbf{g}_{TM}(\boldsymbol{\psi})$, respectively from now on to economize on notations. We assume the following sufficient assumption for the identification of the unknown parameters in the vector $\boldsymbol{\psi}$, which replaces Assumption 4.

ASSUMPTION 7 *(Identification)* $\boldsymbol{\xi}_{TM}(\boldsymbol{\psi}) \equiv \boldsymbol{\xi}\left(\boldsymbol{\theta},\sigma^2,\delta,T,M\right)$ *defined in (30) is continuous in* $\boldsymbol{\psi} = \left(\boldsymbol{\theta}',\sigma^2,\delta\right)'$*, and for any* $\boldsymbol{\psi}_1,\boldsymbol{\psi}_2 \in \Omega$*, such that* $\boldsymbol{\psi}_1 \neq \boldsymbol{\psi}_2$*, we have*

$$\boldsymbol{\xi}_{TM}(\boldsymbol{\psi}_1) \neq \boldsymbol{\xi}_{TM}(\boldsymbol{\psi}_2).$$

Assumption 7 states that, for a given M and T, function $\boldsymbol{\xi}_{TM}(\boldsymbol{\psi})$ is invertible, that is the unknown parameters could be identified from the function $\boldsymbol{\xi}_{TM}(\boldsymbol{\psi})$.

Similarly to Theorem 1, we propose the following consistent estimation of unknown parameters in the vector $\boldsymbol{\psi}$.

Theorem 2 *Let y_{it} be given by (2) for $-M, -M+1, ..., 1, 2, ..., T$, and $i = 1, 2, ..., N$, suppose Assumptions 2-3 and 5-7 hold, and define $\hat{\boldsymbol{\psi}}_{BMM}$ as*

$$\hat{\boldsymbol{\psi}}_{BMM} = \arg\min_{\boldsymbol{\psi}\in\Omega} \left\|\mathbf{g}_{TM}(\boldsymbol{\psi})\right\|, \tag{31}$$

where $\mathbf{g}_{TM}(\boldsymbol{\psi}) = \mathbf{g}\left(\boldsymbol{\theta},\sigma^2, \delta, T, M, \{y_{it}\}_{i=1,t=2}^{N,T}\right)$ is defined in (29), and $\|.\|$ denotes the Euclidean norm. Then $\hat{\boldsymbol{\psi}}_{BMM}$ satisfies

$$p\lim_{N\to\infty} \left\|\hat{\boldsymbol{\psi}}_{BMM} - \boldsymbol{\psi}\right\| = 0. \tag{32}$$

Proof. The dimension of the function $\mathbf{g}_{TM}(\boldsymbol{\psi})$ in (29) is finite since T is finite. Since also $\boldsymbol{\xi}_{TM}(\boldsymbol{\psi})$ is continuous and invertible, it follows that $p\lim_{N\to\infty}\left\|\hat{\boldsymbol{\psi}}_{BMM} - \boldsymbol{\psi}\right\| = 0$ if and only if $p\lim_{N\to\infty}\mathbf{g}_{TM}(\boldsymbol{\psi}) = 0$ for all $t = 2, ..., T$ and $s = 0, 1, 2, ..., t-2$. Noting that (28) indeed holds for all $t = 2, 3, ..., T$ and $s = 0, 1, 2, ..., t-2$ completes the proof. ■

As in the previous Section, there are different ways how one could combine the individual consistent estimating equations in $\mathbf{g}_{TM}(\boldsymbol{\psi})$ to construct an estimator of $\boldsymbol{\psi}$.

Let us consider the following example to illustrate the previous exposition in the context of Beta distributed autoregressive coefficients and $M = 0$.

Example 5 *Suppose $T = 3$, $M = 0$, $\phi_i \sim Beta[\theta_\alpha, \theta_\beta]$, $(\theta_\alpha, \theta_\beta) \in \Omega_\theta$, where $\Omega_\theta \equiv \mathbb{R}_0^+ \times \mathbb{R}_0^+$. The starting deviations η_i in Assumption 5 are for simplicity and illustrative purposes set to 0. Thus the vector of unknown parameters is $\boldsymbol{\psi} = \left(\theta_\alpha, \theta_\beta, \sigma^2\right) \in \Omega_\theta \times \mathbb{R}_0^+$ and the moments of $\phi_i \sim Beta[\theta_\alpha, \theta_\beta]$ can be expressed recursively as*

$$m_\ell(\boldsymbol{\theta}) = \frac{\theta_\alpha + \ell - 1}{\theta_\alpha + \theta_\beta + \ell - 1} m_{\ell-1}(\boldsymbol{\theta}), \text{ for } \ell = 1, 2, ...$$

In this example, function $\boldsymbol{\xi}_{3,0}(\boldsymbol{\psi})$ *is* 3×1 *and given by*

$$\boldsymbol{\xi}_{3,0}(\boldsymbol{\psi}) = \begin{pmatrix} \xi\left(\boldsymbol{\theta},\sigma^2,\delta=0,t=2,s=0,M=0\right) \\ \xi\left(\boldsymbol{\theta},\sigma^2,\delta=0,t=3,s=0,M=0\right) \\ \xi\left(\boldsymbol{\theta},\sigma^2,\delta=0,t=3,s=1,M=0\right) \end{pmatrix}, \tag{33}$$

where

$$\begin{aligned} \xi\left(\boldsymbol{\theta},\sigma^2,\delta=0,t=2,s=0,M=0\right) &= \sigma^2\left[2+\sum_{\ell=1}^{2}\left[m_{2\ell}(\boldsymbol{\theta})-2m_{2\ell-1}(\boldsymbol{\theta})+m_{2\ell-2}(\boldsymbol{\theta})\right]\right], \\ &= \sigma^2\left[3+m_4(\boldsymbol{\theta})+2\sum_{h=1}^{3}(-2)^h m_h(\boldsymbol{\theta})\right] \qquad (34) \\ \xi\left(\boldsymbol{\theta},\sigma^2,\delta=0,t=3,s=0,M=0\right) &= \sigma^2\left[2+\sum_{\ell=1}^{3}\left[m_{2\ell}(\boldsymbol{\theta})-2m_{2\ell-1}(\boldsymbol{\theta})+m_{2\ell-2}(\boldsymbol{\theta})\right]\right] \\ &= \sigma^2\left[3+m_6(\boldsymbol{\theta})+2\sum_{h=1}^{5}(-2)^h m_h(\boldsymbol{\theta})\right] \qquad (35) \end{aligned}$$

and

$$\begin{aligned} \xi\left(\boldsymbol{\theta},\sigma^2,\delta=0,t=3,s=1,M=0\right) &= \sigma^2\left[m_1(\boldsymbol{\theta})-1+\sum_{\ell=1}^{2}\left[m_{1+2\ell}(\boldsymbol{\theta})-2m_{1+2\ell-1}(\boldsymbol{\theta})+m_{1+2\ell-2}(\boldsymbol{\theta})\right]\right] \\ &= \sigma^2\left[m_5(\boldsymbol{\theta})-1+2\sum_{\ell=1}^{4}(-2)^{h+1} m_h(\boldsymbol{\theta})\right]. \qquad (36) \end{aligned}$$

There are 3 unknown parameters and 3 functions in $\boldsymbol{\xi}_{3,0}(\boldsymbol{\psi})$ *and the function* $\boldsymbol{\xi}_{3,0}(\boldsymbol{\psi})$ *is differentiable. Parameters* $\boldsymbol{\psi} = \left(\theta_\alpha, \theta_\beta, \sigma^2\right)$ *are identified if and only if the Jacobian*

$$\mathbf{J}_{\boldsymbol{\xi}}(\boldsymbol{\psi}) = \frac{\partial \boldsymbol{\xi}_{3,0}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}'}$$

has nonzero determinant for all $\boldsymbol{\psi} \in \Omega_\theta$. *It is not straightforward to verify this condition analytically. Therefore, one could examine numerically whether the invertibility requirement in Assumption 7 is satisfied. We show that the parameters are identified numerically in Monte Carlo section for* $T \geq 4$ *and for a range of values for* $E(\phi_i)$ *and* $Var(\phi_i)$.

4 Inference

Inference requires stricter assumptions on errors and their cross section dependence in order to achieve the typical $\sqrt{N}$ convergence. However, even when the cross section dependence of innovations is restricted sufficiently to achieve the $\sqrt{N}$ convergence, the asymptotic distribution would still depend on a nuisance parameters that characterize the degree of cross section dependence. It remains an open research question whether it would be possible to estimate the asymptotic variance in the presence of unknown (but sufficiently weak) form of cross section dependence without restoring to a particular type of spatial model.

We therefore derive the asymptotic distribution of $\hat{\boldsymbol{\psi}}_{BMM}$ in the case of cross sectionally independent innovations below, while still allowing for cross sectional heteroskedasticity. Heteroskedasticity also introduces nuisance parameters in the asymptotic distribution, but we show that it is possible to deal with these parameters and propose a semi-parametric approach that is robust to unknown form of heteroskedasticity. The robustness of our approach in the case of cross sectionally dependent innovations is investigated in Monte Carlo experiments in Section 5.

In addition to Assumptions above, we also assume that the conditions in the following assumption hold.

ASSUMPTION 8 *(Cross sectionally independent idiosyncratic shocks and higher moments of ϕ_i) u_{it} is independently distributed from u_{jt}, for any t and any $i \neq j$, $i, j = 1, 2, ..., N$, and the fourth moments of u_{it} are bounded. In addition,*

$$\lim_{N\to\infty} N^{-1} \sum_{i=1}^{N} \left[E\left(u_{it}^2\right)\right]^2 = \varkappa,$$

$$\lim_{N\to\infty} N^{-1} \sum_{i=1}^{N} E\left(u_{it}^3\right) = \upsilon_3,$$

and

$$\lim_{N\to\infty} N^{-1} \sum_{i=1}^{N} E\left(u_{it}^4\right) = \upsilon_4,$$

where $0 < \varkappa, \upsilon_3, \upsilon_4 < \infty$. Furthermore, η_i is identically and independently distributed across i with $E\left(\eta_i\right) = b_{\delta 1}$, $E\left(\eta_i^2\right) = \delta$, $E\left(\eta_i^4\right) = b_{\delta 4} < \infty$, first $4\left(T + M + 1\right)$ moments of ϕ_i exist and $\boldsymbol{\xi}_{TM}\left(\boldsymbol{\psi}\right)$

is differentiable (with respect to ψ) in a neighborhood of ψ.

We derive the asymptotic distribution for $\hat{\psi}_{BMM}$ defined in (31). We proceed in two steps.

First, we derive the asymptotic distribution of $\mathbf{g}_{TM}(\psi)$ in the proposition below. These derivations are rather tedious and the proof is therefore relegated to the Appendix.

Proposition 1 *Let y_{it} be given by (2) for $-M, -M+1, ..., 1, 2, ..., T,$ and $i = 1, 2, ..., N,$ suppose Assumptions 2-3 and 5-8 hold, T is fixed and $N \to \infty$. Then $\mathbf{g}_{TM}(\psi) = \mathbf{g}\left(\boldsymbol{\theta}, \sigma^2, \delta, T, M, \{y_{it}\}_{i=1,t=2}^{N,T}\right)$ defined in (29) satisfies*

$$\sqrt{N}\mathbf{g}_{TM}(\psi) \xrightarrow{d} N(0, \boldsymbol{\Theta}), \tag{37}$$

where $\boldsymbol{\Theta}$ is $T(T-1)/2$ dimensional asymptotic covariance matrix with its elements given by

$$\phi_{\ell h} = \phi^*_{t(\ell),s(\ell),r(\ell,h),p(\ell,h)}, \text{ for } \ell, h = 1, 2, ..., T(T-1)/2, \tag{38}$$

in which the index functions $t(\ell), s(\ell), r(\ell,h), p(\ell,h)$ are defined in (B.3)-(B.5), and

$$\phi^*_{tsrh} = \varrho_1 + \varrho_2 + \varrho_3 + \varrho_4 - \varkappa \sum_{\ell=s}^{t+M} \sum_{\ell'=p}^{t+M} \gamma_{\ell,\ell-s}(\boldsymbol{\theta}) \gamma_{\ell'-r,\ell'-p}(\boldsymbol{\theta}) - \gamma_{t+M+1,t-s+M+1}(\boldsymbol{\theta}) \gamma_{t-p+M+1,t-r+M+1}(\boldsymbol{\theta}) \delta,$$

for $t = 2, 3, ..., T,$ $s = 0, 1, 2, ..., t-2,$ $r = 0, 1, ..., t-2,$ and $p = r, r+1, ..., t-2;$ where $\gamma_{\ell_1}(\boldsymbol{\theta}) = E(\varphi_{i\ell_1})$, $\gamma_{\ell_1\ell_2}(\boldsymbol{\theta}) = E(\varphi_{i\ell_1}\varphi_{i\ell_2})$, $\gamma_{\ell_1\ell_2\ell_3}(\boldsymbol{\theta}) = E(\varphi_{i\ell_1}\varphi_{i\ell_2}\varphi_{i\ell_3})$, $\gamma_{\ell_1\ell_2\ell_3\ell_4}(\boldsymbol{\theta}) = E(\varphi_{i\ell_1}\varphi_{i\ell_2}\varphi_{i\ell_3}\varphi_{i\ell_4})$,

$$\begin{aligned}
\varrho_1 &= \upsilon_4 \sum_{\ell=p}^{t+M} \gamma_{\ell,\ell-s,\ell-r,\ell-p}(\boldsymbol{\theta}) + \varkappa \sum_{\ell_1=s}^{t+M} \sum_{\ell_3=p;\ \ell_3\neq\ell_1}^{t+M} \gamma_{\ell_1,\ell_1-s,\ell_3-r,\ell_3-p}(\boldsymbol{\theta}) \\
&\quad + \varkappa \sum_{\ell_1=r}^{t+M+1} \sum_{\ell_2=p;\ \ell_2\neq\ell_1}^{t+M+1} \gamma_{\ell_1,\ell_2-s,\ell_1-r,\ell_2-p}(\boldsymbol{\theta}) + \varkappa \sum_{\ell_1=p}^{t+M+1} \sum_{\ell_2=r;\ \ell_2\neq\ell_1}^{t+M+1} \gamma_{\ell_1,\ell_2-s,\ell_2-r,\ell_1-p}(\boldsymbol{\theta}),
\end{aligned}$$

$$\begin{aligned}
\varrho_2 &= b_{\delta 1}\upsilon_3 \left(\gamma_{t-r+M+1}(\boldsymbol{\theta}) \sum_{\ell=\max\{s,p\}}^{t+M} \gamma_{\ell,\ell-s,\ell-p}(\boldsymbol{\theta}) + \gamma_{t-p+M+1}(\boldsymbol{\theta}) \sum_{\ell=\max\{s,r\}}^{t+M} \gamma_{\ell,\ell-s,\ell-r}(\boldsymbol{\theta}) \right. \\
&\quad \left. + \gamma_{t-s+M+1}(\boldsymbol{\theta}) \sum_{\ell=p}^{t+M} \gamma_{\ell,\ell-p,\ell-r}(\boldsymbol{\theta}) + \gamma_{t+M+1}(\boldsymbol{\theta}) \sum_{\ell=\max\{s,p\}}^{t+M} \gamma_{\ell-s,\ell-p,\ell-r}(\boldsymbol{\theta}) \right),
\end{aligned}$$

$$\varrho_3 = \delta\sigma^2 \left(\gamma_{t-p+M+1,t-r+M+1}(\boldsymbol{\theta}) \sum_{\ell=s}^{t+M} \gamma_{\ell,\ell-s}(\boldsymbol{\theta}) + \gamma_{t-s+M+1,t-r+M+1}(\boldsymbol{\theta}) \sum_{\ell=p}^{t+M} \gamma_{\ell,\ell-p}(\boldsymbol{\theta}) \right.$$
$$+\gamma_{t-s+M+1,t-p+M+1}(\boldsymbol{\theta}) \sum_{\ell=r}^{t+M} \gamma_{\ell,\ell-r}(\boldsymbol{\theta}) + \gamma_{t+M+1,t-r+M+1}(\boldsymbol{\theta}) \sum_{\ell=\max\{s,p\}}^{t+M} \gamma_{\ell-s,\ell-p}(\boldsymbol{\theta})$$
$$\left. +\gamma_{t+M+1,t-p+M+1}(\boldsymbol{\theta}) \sum_{\ell=\max\{s,r\}}^{t+M} \gamma_{\ell-s,\ell-r}(\boldsymbol{\theta}) + \gamma_{t+M+1,t-s+M+1}(\boldsymbol{\theta}) \sum_{\ell=p}^{t+M} \gamma_{\ell-p,\ell-r}(\boldsymbol{\theta}) \right),$$

$$\varrho_4 = \gamma_{t+M+1,t-s+M+1,t-p+M+1,t-r+M+1}(\boldsymbol{\theta}) \cdot \upsilon_4,$$

and

$$\varphi_{i\ell} = \begin{cases} 1, & \text{for } \ell = 0 \\ \phi_i^{\ell} - \phi_i^{\ell-1}, & \text{for } \ell = 1, 2, ..., t+M \end{cases}. \tag{39}$$

Proposition 1 establishes that $\mathbf{g}_{TM}(\boldsymbol{\psi})$ converges to a multivariate normal distribution at the rate $\sqrt{N}$ with the asymptotic variance-covariance matrix being a complicated function of parameters in θ (recall θ belongs to $\boldsymbol{\psi}$). The precise functional form depends on the distribution of the autoregressive coefficients, in particular through the moment functions $\gamma_{\ell_1}(\boldsymbol{\theta})$, $\gamma_{\ell_1\ell_2}(\boldsymbol{\theta})$, $\gamma_{\ell_1\ell_2\ell_3}(\boldsymbol{\theta}) = E\left(\varphi_{i\ell_1}\varphi_{i\ell_2}\varphi_{i\ell_3}\right)$ and $\gamma_{\ell_1\ell_2\ell_3\ell_4}(\boldsymbol{\theta}) = E\left(\varphi_{i\ell_1}\varphi_{i\ell_2}\varphi_{i\ell_3}\varphi_{i\ell_4}\right)$, where $\varphi_{i\ell}$ is defined in (39). These functions are straightforward to derive by substituting (39) for $\varphi_{i\ell}$ and noting that $m_\ell(\boldsymbol{\theta}) = E\left(\phi_i^\ell\right)$. Function $\gamma_{\ell_1}(\boldsymbol{\theta})$ thus can be written as

$$\gamma_{\ell_1}(\boldsymbol{\theta}) = \begin{cases} 1 & \text{for } \ell_1 = 0, \\ m_{\ell_1}(\boldsymbol{\theta}) - m_{\ell_1-1}(\boldsymbol{\theta}) & \text{for } \ell_1 = 1, 2, ..., t+M, \end{cases},$$

and function $\gamma_{\ell_1\ell_2}(\boldsymbol{\theta})$ can be written as

$$\gamma_{\ell_1\ell_2}(\boldsymbol{\theta}) = \begin{cases} 1 & \text{for } \ell_1 = \ell_2 = 0, \\ m_{\ell_2}(\boldsymbol{\theta}) - m_{\ell_2-1}(\boldsymbol{\theta}) & \text{for } \ell_1 = 0, \ell_2 = 1, 2, ..., t+M, \\ m_{\ell_1}(\boldsymbol{\theta}) - m_{\ell_1-1}(\boldsymbol{\theta}) & \text{for } \ell_1 = 1, 2, ..., t+M, \ell_2 = 0, \\ m_{\ell_1+\ell_2}(\boldsymbol{\theta}) - 2m_{\ell_1+\ell_2-1}(\boldsymbol{\theta}) + m_{\ell_1+\ell_2-2}(\boldsymbol{\theta}) & \text{for } \ell_1, \ell_2 = 1, 2, ..., t+M, \end{cases}$$

Similarly, $\gamma_{\ell_1\ell_2\ell_3}(\boldsymbol{\theta})$ and $\gamma_{\ell_1\ell_2\ell_3\ell_4}(\boldsymbol{\theta})$ can be written as functions involving indices $\ell_1, \ell_2, \ell_3, \ell_4$ and the moment function $m_{.}(.)$, but we do not do so here since these expressions are long and not very insightful.

Besides the unknown parameters in the vector θ, the asymptotic variance of $\mathbf{g}_{TM}(\boldsymbol{\psi})$ depends also on the character of heteroskedasticity and fourth moments. In particular, $\boldsymbol{\Theta}$ depends on the limits of arithmetic cross section averages of $\left[E\left(u_{it}^2\right)\right]^2$ and $E\left(u_{it}^4\right)$ (parameters $\varkappa$ and υ_4 in Assumption 8). Moreover, the asymptotic distribution also depends on the moments $b_{\delta 1} = E(\eta_i)$, $\delta = E\left(\eta_i^2\right)$ and $b_{\delta 4} = E\left(\eta_i^4\right)$; and in the case when $b_{\delta 1} \neq 0$, then the asymptotic distribution depends also on the skewness of the distribution of innovations, parameter υ_3.

In the second step of deriving the asymptotic distribution of $\hat{\boldsymbol{\psi}}_{BMM}$, let us use the following decomposition $\hat{\boldsymbol{\psi}}_{BMM} = \boldsymbol{\psi} + \boldsymbol{\varepsilon}$ and note that (37) implies $\boldsymbol{\varepsilon} = O_p\left(N^{-1/2}\right)$. Denote $\sqrt{N}\mathbf{g}_{TM}(\boldsymbol{\psi})$ simply as $\boldsymbol{\vartheta}_\Theta$ and consider the Taylor expansion of $\boldsymbol{\xi}_{TM}(\boldsymbol{\psi}+\boldsymbol{\varepsilon}) = \boldsymbol{\xi}_{TM}(\boldsymbol{\psi}) + \mathbf{J}_{\boldsymbol{\xi}}(\boldsymbol{\psi})\boldsymbol{\varepsilon} + O_p\left(N^{-1}\right)$, where $\mathbf{J}_{\boldsymbol{\xi}}(\boldsymbol{\psi})$ is (Jacobian) matrix of partial derivatives of $\boldsymbol{\xi}_{TM}(\boldsymbol{\psi})$ with respect to $\boldsymbol{\psi}$. We shall denote $\mathbf{J}_{\boldsymbol{\xi}}(\boldsymbol{\psi})$ simply as $\mathbf{J}_{\boldsymbol{\xi}}$ below (unless we need to emphasize the dependence on $\boldsymbol{\psi}$). Using this expansion in (31) we obtain (multiplying the objective function by $\sqrt{N}$),

$$\boldsymbol{\varepsilon} = \arg\min_{\boldsymbol{\varepsilon}} \left\| \mathbf{J}_{\boldsymbol{\xi}} \cdot \left(\sqrt{N}\boldsymbol{\varepsilon}\right) + O_p\left(N^{-1/2}\right) - \boldsymbol{\vartheta}_\Theta \right\|.$$

Noting that $\boldsymbol{\vartheta}_\Theta \sim N(0, \boldsymbol{\Theta})$, the asymptotic distribution of $\boldsymbol{\varepsilon}$ is the same as the distribution of $\tilde{\boldsymbol{\varepsilon}}$ defined by

$$\tilde{\boldsymbol{\varepsilon}} = \arg\min_{\tilde{\boldsymbol{\delta}}} \left\| \mathbf{J}_{\boldsymbol{\xi}} \cdot \left(\sqrt{N}\tilde{\boldsymbol{\varepsilon}}\right) - \boldsymbol{\vartheta}_{\boldsymbol{\Theta}} \right\|. \tag{40}$$

It follows that

$$\sqrt{N}\left(\hat{\boldsymbol{\psi}}_{BMM} - \boldsymbol{\psi}\right) - \sqrt{N}\tilde{\boldsymbol{\varepsilon}} \xrightarrow{d} 0.$$

If $\mathbf{J}_{\boldsymbol{\xi}}$ is invertible, then $\sqrt{N}\tilde{\boldsymbol{\varepsilon}} = \mathbf{J}_{\boldsymbol{\xi}}^{-1}\boldsymbol{\vartheta}_{\boldsymbol{\Theta}}$ and (40) implies

$$\sqrt{N}\left(\hat{\boldsymbol{\psi}}_{BMM} - \boldsymbol{\psi}\right) \xrightarrow{d} N\left(0, \mathbf{J}_{\boldsymbol{\xi}}^{-1}\boldsymbol{\Theta}\mathbf{J}_{\boldsymbol{\xi}}^{-1\prime}\right).$$

Regardless whether $\mathbf{J}_{\boldsymbol{\xi}}$ is invertible or not, we have $\tilde{\boldsymbol{\delta}} = \left(\mathbf{J}_{\boldsymbol{\xi}}'\mathbf{J}_{\boldsymbol{\xi}}\right)^{-1}\mathbf{J}_{\boldsymbol{\xi}}'\boldsymbol{\vartheta}_{\boldsymbol{\Theta}}$, and the asymptotic variance of $\hat{\boldsymbol{\psi}}_{BMM}$ can be always written as $\left(\mathbf{J}_{\boldsymbol{\xi}}'\mathbf{J}_{\boldsymbol{\xi}}\right)^{-1}\mathbf{J}_{\boldsymbol{\xi}}'\boldsymbol{\Theta}\mathbf{J}_{\boldsymbol{\xi}}\left(\mathbf{J}_{\boldsymbol{\xi}}'\mathbf{J}_{\boldsymbol{\xi}}\right)^{-1}$. Our main findings are summarized in the following theorem.

Theorem 3 *Let y_{it} be given by (2) for $-M,-M+1,...,1,2,...,T,$ and $i=1,2,...,N,$ suppose Assumptions 2-3 and 5-8 hold, T is fixed and $N\to\infty$. Then $\hat{\boldsymbol{\psi}}_{BMM}$ defined in (37) has the following asymptotic distribution,*

$$\sqrt{N}\left(\hat{\boldsymbol{\psi}}_{BMM}-\boldsymbol{\psi}\right)\xrightarrow{d}N\left(0,\boldsymbol{\Sigma}_{\psi}\right) \tag{41}$$

where

$$\boldsymbol{\Sigma}_{\psi}=\left(\mathbf{J}_{\boldsymbol{\xi}}'\mathbf{J}_{\boldsymbol{\xi}}\right)^{-1}\mathbf{J}_{\boldsymbol{\xi}}'\boldsymbol{\Theta}\mathbf{J}_{\boldsymbol{\xi}}\left(\mathbf{J}_{\boldsymbol{\xi}}'\mathbf{J}_{\boldsymbol{\xi}}\right)^{-1}, \tag{42}$$

$\mathbf{J}_{\boldsymbol{\xi}}$ is the Jacobian of $\boldsymbol{\xi}_{TM}(\boldsymbol{\psi})$ and $\boldsymbol{\Theta}$ is the asymptotic variance of $\mathbf{g}_{TM}(\boldsymbol{\psi})$ defined in Proposition 1.

Theorem 3 can be easily extended to estimators derived based on (31) but with different objective function than minimizing the euclidean norm of $\mathbf{g}_{TM}(\boldsymbol{\psi})$. Let $\widehat{\boldsymbol{\psi}}_w$ be defined as

$$\hat{\boldsymbol{\psi}}_w=\arg\min_{\boldsymbol{\psi}\in\Omega}\left\|\mathbf{g}\left(\boldsymbol{\psi},T,M\right)\right\|_{\mathbf{W}}, \tag{43}$$

where $\mathbf{W}$ is $T(T-1)/2\times(k+1)$ matrix of weights such that $\mathbf{W}'\mathbf{W}$ is positive definite, and $\|.\|_{\mathbf{W}}$ is a linear operator defined by $\|\mathbf{x}\|_{\mathbf{W}}=\mathbf{x}'\mathbf{W}'\mathbf{W}\mathbf{x}$ for any $T(T-1)/2$ dimensional vector $\mathbf{x}$. Using the same steps as before, it can be shown that

$$\sqrt{N}\left(\widehat{\boldsymbol{\psi}}_w-\boldsymbol{\psi}\right)\xrightarrow{d}N\left[0,\left(\mathbf{J}_{\boldsymbol{\xi}}'\mathbf{W}\mathbf{W}'\mathbf{J}_{\boldsymbol{\xi}}\right)^{-1}\mathbf{J}_{\boldsymbol{\xi}}'\mathbf{W}\boldsymbol{\Theta}\mathbf{W}'\mathbf{J}_{\boldsymbol{\xi}}\left(\mathbf{J}_{\boldsymbol{\xi}}'\mathbf{W}\mathbf{W}'\mathbf{J}_{\boldsymbol{\xi}}\right)^{-1}\right].$$

Theorems 1 and 3 can be easily extended to the case when $M\to\infty$.

4.0.1 Estimation of $\boldsymbol{\Sigma}_{\psi}$

In empirical applications, the second moments of the asymptotic variance of $\hat{\boldsymbol{\psi}}_{BMM}$ need to be estimated. This can be done semi-parametrically, based on

$$\phi_{tsrp}^{*}=p\lim_{N\to\infty}E\left\{\frac{1}{N}\sum_{i=1}^{N}\left[\Delta y_{it}\Delta y_{i,t-s}-E\left(\Delta y_{it}\Delta y_{i,t-s}\right)\right]\cdot\left[\Delta y_{i,t-r}\Delta y_{i,t-p}-E\left(\Delta y_{i,t-r}\Delta y_{i,t-p}\right)\right]\right\},$$

Fully parametric estimation necessitates estimation of all nuisance parameters of the asymptotic distribution. However, as we shall see below, the performance of the semi-parametric approach is rather good and this approach is also relatively easy to implement, and therefore fully parametric approach is not necessary and is likely to be less robust than the semi-parametric approach below.

Semi–parametric estimator of $\mathbf{\Sigma}_\psi$ is given by[3]

$$\hat{\mathbf{\Sigma}}_\psi = \left(\hat{\mathbf{J}}'_{\boldsymbol{\xi}}\hat{\mathbf{J}}_{\boldsymbol{\xi}}\right)^{-1}\hat{\mathbf{J}}'_{\boldsymbol{\xi}}\hat{\mathbf{\Theta}}\hat{\mathbf{J}}_{\boldsymbol{\xi}}\left(\hat{\mathbf{J}}'_{\boldsymbol{\xi}}\hat{\mathbf{J}}_{\boldsymbol{\xi}}\right)^{-1},$$

where $\hat{\mathbf{J}}_{\boldsymbol{\xi}} = \mathbf{J}_{\boldsymbol{\xi}}\left(\hat{\boldsymbol{\psi}}_{BMM}\right)$, the individual elements of $\hat{\mathbf{\Theta}}$ are given by (38) with ϕ^*_{tsrp} replaced by its non-parametric estimate,

$$\hat{\phi}^*_{tsrp} = \frac{1}{N}\sum_{i=1}^{N}\left(\Delta y_{it}\Delta y_{i,t-s} - \hat{\omega}_{ts}\right)\cdot\left(\Delta y_{i,t-r}\Delta y_{i,t-p} - \hat{\omega}_{t-r,p-r}\right),$$

in which

$$\hat{\omega}_{ts} = \frac{1}{N}\sum_{i=1}^{N}\Delta y_{it}\Delta y_{i,t-s}.$$

5 Monte Carlo Evidence

5.1 Data Generating Process (DGP)

We generate the dependent variable based on the panel AR(1) model,

$$y_{it} = c_i + \phi_i y_{i,t-1} + u_{it}, \tag{44}$$

for $i = 1, 2, ..., N$ and $t = 0, 1, 2, ..., T$, where the fixed effects, c_i, and starting values, $y_{i,-1}$, are generated as

$$\begin{aligned} c_i &= (1-\phi_i)\,\mu_i, \\ y_{i,-1} &= \mu_i, \end{aligned}$$

[3]Note that we have shown that $\boldsymbol{\psi}$ can be consistently estimated. Hence $J_{\boldsymbol{\xi}}(\boldsymbol{\psi})$ and γ's can be consistently estimated.

for $i = 1, 2, ..., N$, with $\mu_i \sim IIDN(1, 1)$. We consider four possibilities for autoregressive parameters:

1. homogeneous autoregressive parameters, $\phi_i = \theta_\mu = 0.6$,
2. uniformly distributed autoregressive parameters, $\phi_i \sim IIDU[\phi_{\min}, \phi_{\max}]$,
3. normally distributed autoregressive parameters, $\phi_i \sim IIDN(\theta_\mu, \theta_\sigma^2)$, and
4. beta distributed autoregressive parameters, $\phi_i \sim IIDBeta[\phi_\alpha, \phi_\beta]$.

In all heterogeneous cases we choose parameters of the distribution of autoregressive coefficients so that $E(\phi_i) = \theta_\mu = 0.6$ and $Var(\phi_i) = \theta_\sigma^2 = 0.1^2$ and in all cases the key parameters of interest are the mean and the standard error of the coefficients, $\theta = (\theta_\mu, \theta_\sigma)'$. [4]

Idiosyncratic errors are generated to be serially independent, cross sectionally dependent and heteroskedastic, based on the following spatial model.

$$u_{1t} \sim IIDN\left(0, \sigma_1^2\right), \text{ for } t = 0, 1, 2, ..., T, \tag{45}$$

and

$$u_{it} = b_u \frac{\sigma_i}{\sigma_{i-1}} u_{i-1,t} + \varepsilon_{it}, \; \varepsilon_{it} \sim IIDN\left[0, \left(1 - b_u^2\right)\sigma_i^2 / \left(1 - \phi_i^2\right)\right], \tag{46}$$

for $i = 2, 3, ..., N$ and $t = 0, 1, 2, ...T$. Cross-sectional dependence of $\{u_{it}\}$ is given by the spatial autoregressive parameter b_u and we set $b_u = 0.6$. Note that for $|b_u| < 1$ the cross-sectional dependence of $\{u_{it}\}$, as defined in Chudik, Pesaran, and Tosetti (2011), is weak. Variance of u_{it} is generated as $\sigma_i^2 = 1/2 + \kappa_i/4$, where κ_i is an IID draw from $\chi^2(2)$.

It is not straightforward to verify Assumption 1.i analytically when coefficients are stochastic. We assume $\delta = 0$ is known and the unknown parameters that we estimate are $\boldsymbol{\psi} = \left[\theta_\mu, \theta_\sigma, E\left(\sigma_i^2\right)\right]'$. We investigate validity of Assumption 4 by means of numerical simulations in Appendix and find out that it is satisfied for a reasonable range of parameter values considered.

[4] In particular, we set $\phi_{\min} = \theta_\mu - \sqrt{12}\theta_\sigma/2$, $\phi_{\max} = 2\theta_\mu - \phi_{\min}$, $\phi_\alpha = \theta_\mu\left[\theta_\mu(1-\theta_\mu)/\theta_\sigma^2 - 1\right]$ and $\phi_\beta = (1-\theta_\mu)\left[\theta_\mu(1-\theta_\mu)/\theta_\sigma^2 - 1\right]$.

5.2 Monte Carlo findings

5.2.1 Root mean squared errors (RMSE) and bias

Table 1 presents findings with homogeneous autoregressive coefficients, and Tables 2, 3, and 4, respectively, present findings with uniformly, normally and beta distributed autoregressive coefficients.

In each of these experiments, we consider four estimators, depending on the functional form assumed for the distribution of autoregressive coefficients. In particular, we assume degenerate ($\theta_\sigma = 0$), uniform, normal and Beta distributions.

Findings in Tables 1-4 are very encouraging and show that the approach proposed in this paper performs well even in the heterogeneous cases when the distribution is not correctly specified so long as it is not wrongly assumed that the coefficients are homogeneous. The reported bias and RMSE are largest when it is wrongly assumed that coefficients are homogeneous. The results are surprisingly similar across different non-degenerate distributional forms assumed for the estimation of the mean of ϕ_i.

5.2.2 Size and power

Tables 5-8 report on the size and power for the estimation of $E(\phi_i)$ in the four sets of experiments (homogeneous slopes, uniformly, normally and beta distributed slopes, respectively). We present findings for experiments with cross-sectionally independent errors ($b_u = 0$), since this assumption is required for correct inference. It can be seen that the inference based on the homogeneous slopes assumption can be grossly misleading in cases when the slopes are in fact heterogeneous (Tables 6-8), with the size approaching 100% as $N \to \infty$. Misspecification of the non-degenerate distributional form (uniform, normal or beta), on the other hand does not seem to have any severe consequences.

6 Extension to AR(p) and VAR(p) models

Since AR(p) as well as VAR(p) models can be represented by a VAR(1) companion model, we first present the derivations of a general panel VAR(1) model and then we discuss special cases of panel

AR(p) and panel VAR(p) models.

6.1 Panel VAR(1) model

We extend the model in Section 2 into VAR(1) model below. Let $y_{it} = (y_{i1t}, y_{i2t}, ..., y_{int})'$ be $n \times 1$ vector of variables specific to the unit i and consider the following panel VAR(1) model with fixed effects

$$\mathbf{y}_{it} = \mathbf{c}_i + \mathbf{\Phi}_i \mathbf{y}_{i,t-1} + \mathbf{u}_{it}, \tag{47}$$

for $i = 1, 2, ..., N$, and $t = -M, -M+1, ..., 1, 2, ..., T$; where $M \geq 0$, $T > 2$, $\mathbf{c}_i$ is $n \times 1$ vectors of fixed effects for the unit i, $\mathbf{\Phi}_i$ is $n \times n$ matrix of unknown coefficients for the unit i, and $\mathbf{u}_{it}$ is $n \times 1$ vector of the idiosyncratic shocks. As before, it is assumed that only observations for time periods $t = 1, 2, ..., T$ are available, T is fixed and N is large. To deal with fixed effects, we again take the first differences of (47),

$$\Delta \mathbf{y}_{it} = \mathbf{\Phi}_i \Delta \mathbf{y}_{i,t-1} + \Delta \mathbf{u}_{it}, \tag{48}$$

for $i = 1, 2, ..., N$ and $t = -M+1, -M+2, ..., 1, 2, ..., T$, and postulate the following assumptions, which replace previous assumptions on the initialization of the dynamic processes, coefficients, and idiosyncratic shocks.

ASSUMPTION 9 *(Initialization)*

(i) Finite past: The process $\Delta \mathbf{y}_{it}$, *for* $i = 1, 2, ..., N$, *and* $t = -M+1, -M+2, ..., 1, 2, ..., T$, *is given by (48) with starting values*

$$\Delta \mathbf{y}_{i,-M} = (\mathbf{\Phi}_i - \mathbf{I}_n) \boldsymbol{\eta}_i + \mathbf{u}_{i,-M}, \tag{49}$$

where $\eta_i = (\eta_{i1}, \eta_{i2}, ..., \eta_{in})'$ *is* $n \times 1$ *vector independently distributed of* $\mathbf{\Phi}_j$ *and* $\mathbf{u}_{jt}$ *for any* i, j *and* t, $p\lim_{N\to\infty} N^{-1} \sum_{i=1}^{N} \eta_i \eta_i' = \mathbf{B}_\delta < \infty$, *and* $M \geq 0$. *Furthermore,* M *is known.*

(ii) Infinite past: The support of $|\lambda_1 (\mathbf{\Phi}_i)|$, *where* $\lambda_1 (\mathbf{\Phi}_i)$ *is the largest eigenvalue of* $\mathbf{\Phi}_i$, *lies strictly within the unit circle,* $M \to \infty$, *and* $\Delta \mathbf{y}_{it}$ *is given by (48) for* $i = 1, 2, ..., N$, *and* $t = ..., 0, 1, 2, ..., T$.

ASSUMPTION 10 *(Idiosyncratic shocks) Shocks* $\mathbf{u}_{it}$*, for* $i = 1, 2, ..., N$ *and* $t = -M, -M + 1, ..., 0, 1, ..., T$ *are serially independent and distributed such that:*

(i) $E(\mathbf{u}_{it}) = 0$,

(ii) $E(\mathbf{u}_{it}\mathbf{u}_{it}') = \mathbf{\Sigma}_i < K$ *for some constant* $0 < K < \infty$ *that does not depend on* N, $p\lim_{N\to\infty} N^{-1}\sum_{i=1}^{N}\mathbf{u}_{it}\mathbf{u}_{it}' = \mathbf{\Sigma}$*, and*

(iii) $p\lim_{N\to\infty}\bar{\mathbf{u}}_t = 0$*, where* $\bar{\mathbf{u}}_t = N^{-1}\sum_{i=1}^{N}\mathbf{u}_{it}$.

Assumptions 9-10 are straightforward extensions of earlier assumptions and they play a similar role. In particular, in the case of finite past, the starting value $\Delta\mathbf{y}_{i,-M}$ in (49) together with the VAR equation (47) for $t = -M$ imply

$$(\mathbf{I}_n - \mathbf{\Phi}_i)\, y_{i,-M-1} = \mathbf{c}_i - (\mathbf{\Phi}_i - \mathbf{I}_n)\,\boldsymbol{\eta}_i,$$

and if in addition $|\lambda_1(\mathbf{\Phi}_i)| < 1$, we have

$$\mathbf{y}_{i,-M-1} - \boldsymbol{\mu}_i = \boldsymbol{\eta}_i,$$

where $\boldsymbol{\mu}_i = (\mathbf{I}_n - \mathbf{\Phi}_i)^{-1}\mathbf{c}_i$; which shows that Assumption 9.i postulates that the vector η_i represents the deviations of the initial values in $\mathbf{y}_{i,-M-1}$ from the long-run mean $\boldsymbol{\mu}_i$, provided that $\boldsymbol{\mu}_i$ exists. Note that Assumption 9.i does not necessarily require that $|\lambda_1(\mathbf{\Phi}_i)| < 1$. Assumption 10 continues to assume a rather general requirements on the idiosyncratic errors, which are allowed to be heteroskedaastic and weakly cross sectionally dependent with the eigenvalues of $E(\mathbf{u}_t\mathbf{u}_t')$, where $\mathbf{u}_t = (\mathbf{u}_{1t}', \mathbf{u}_{2t}', ..., \mathbf{u}_{Nt}')'$, not necessarily bounded in N.

The next assumption replaces Assumption 3. As before, we assume the distribution of the VAR coefficients matrices is known up to a k unknown parameters collected in the vector θ.

ASSUMPTION 11 *(VAR coefficients) Coefficient matrices* $\mathbf{\Phi}_i$*, for* $i = 1, 2, ..., N$*, are independently and identically distributed from a known multivariate distribution* $\mathbf{G}(\boldsymbol{\theta})$ *with finite* $2M+2T+2$ *moments and* $k\times 1$ *dimensional vector of unknown parameters,* $\theta = (\theta_1, \theta_2, ..., \theta_k)' \in \mathcal{C}_k$*, where* $k <$

∞ *does not depend on* N*, and* $\mathcal{C}_k$ *is a* k*-dimensional compact set. Coefficient in* $\mathbf{\Phi}_i$ *are independently distributed of innovations in* $\mathbf{u}_{i't}$ *for any* $i, i' = 1, 2, .., N$*, and any* $t = -M, -M+1, ..., 0, 1, ..., T$*.*

Using (48) to solve for $\Delta \mathbf{y}_{it}$, we obtain

$$\Delta \mathbf{y}_{it} = \mathbf{u}_{it} + \sum_{\ell=1}^{t+M} \mathbf{\Phi}_i^{\ell-1} \left(\mathbf{\Phi}_i - \mathbf{I}_n\right) \mathbf{u}_{i,t-\ell} + \mathbf{\Phi}_i^{t+M} \left(\mathbf{\Phi}_i - \mathbf{I}_n\right) \boldsymbol{\eta}_i.$$

Define the following function

$$\mathbf{M}_{\mathbf{\Sigma},\ell_1,\ell_2}\left(\boldsymbol{\theta}\right) = \begin{cases} E\left[\left(\mathbf{\Phi}_i^{\ell_1} - \mathbf{\Phi}_i^{\ell_1-1}\right) \mathbf{\Sigma} \left(\mathbf{\Phi}_i^{\ell_2} - \mathbf{\Phi}_i^{\ell_2-1}\right)'\right], & \text{for } \ell_1, \ell_2 = 1, 2, ...T+M+1, \\ E\left[\left(\mathbf{\Phi}_i^{\ell_1} - \mathbf{\Phi}_i^{\ell_1-1}\right) \mathbf{\Sigma}\right], & \text{for } \ell_1 = 1, 2, ...T+M+1,\ \ell_2 = 0, \\ E\left[\mathbf{\Sigma} \left(\mathbf{\Phi}_i^{\ell_2} - \mathbf{\Phi}_i^{\ell_2-1}\right)'\right], & \text{for } \ell_1 = 0,\ \ell_2 = 1, 2, ...T+M+1, \text{ and} \\ \mathbf{\Sigma}, & \text{for } \ell_1 = \ell_2 = 0. \end{cases}$$

Consider the limiting behavior of cross section averages of the product $\Delta \mathbf{y}_{it} \Delta \mathbf{y}'_{i,t-s}$ for $t = 2, 3, ..., T$ and $s = 0, 1, 2, ..., t-2$. Using similar steps as in the derivation of (22), we obtain under Assumptions 9-11,

$$p \lim_{N\to\infty} N^{-1} \sum_{i=1}^{N} \Delta \mathbf{y}_{it} \Delta \mathbf{y}'_{i,t-s} = \mathbf{M}_{\mathbf{\Sigma},s,0}\left(\boldsymbol{\theta}\right) + \sum_{\ell=1}^{t+M-s} \mathbf{M}_{\mathbf{\Sigma},\ell+s,\ell}\left(\boldsymbol{\theta}\right) + \mathbf{M}_{\mathbf{B}_\delta,T+M+1,T+M+1-s}\left(\boldsymbol{\theta}\right), \quad (50)$$

for $t = 2, 3, ..., T$, and $s = 0, 1, 2, ..., t-2$. Let us define the following function

$$\Xi\left(\boldsymbol{\theta}, \mathbf{\Sigma}, \mathbf{B}_\delta, t, s, M\right) = \mathbf{M}_{\mathbf{\Sigma},s,0}\left(\boldsymbol{\theta}\right) + \sum_{\ell=1}^{t+M-s} \mathbf{M}_{\mathbf{\Sigma},\ell+s,\ell}\left(\boldsymbol{\theta}\right) + \mathbf{M}_{\mathbf{B}_\delta,T+M+1,T+M+1-s}\left(\boldsymbol{\theta}\right), \quad (51)$$

for $t = 2, 3, ..., T$ and $s = 0, 1, 2, ..., t-2$. Note that when the dynamic process started from infinite past ($M \to \infty$ and Assumption 9.ii holds), then $\mathbf{B}_\delta$ no longer enters the function $\Xi(.)$ defined in (51). Consider the following consistent estimating equations,

$$\mathbf{G}\left(\boldsymbol{\theta}, \mathbf{\Sigma}, \mathbf{B}_\delta, t, s, M, \{y_{it}\}_{i=1,t=2}^{N,T}\right) = \Xi\left(\boldsymbol{\theta}, \mathbf{\Sigma}, \mathbf{B}_\delta, t, s, M\right) - N^{-1} \sum_{i=1}^{N} \Delta \mathbf{y}_{it} \Delta \mathbf{y}'_{i,t-s}, \quad (52)$$

for $t = 2, 3, ..., T$ and $s = 0, 1, 2, ..., t-2$. Note that by construction we have,

$$p \lim_{N\to\infty} \mathbf{G}\left(\boldsymbol{\theta}, \boldsymbol{\Sigma}, \mathbf{B}_\delta, t, s, M, \{y_{it}\}_{i=1,t=2}^{N,T}\right) = \underset{n\times n}{\mathbf{0}}. \tag{53}$$

We stack $\sum_{t=2}^{T} n(t-1) = nT(T-1)/2$ consistent estimating equations into one vector function,

$$\underset{nT(T-1)/2\times 1}{\mathbf{g}\left(\boldsymbol{\theta}, \boldsymbol{\Sigma}, \mathbf{B}_\delta, T, M, \{y_{it}\}_{i=1,t=2}^{N,T}\right)} = \begin{pmatrix} vec\left[\mathbf{G}\left(\boldsymbol{\theta}, \boldsymbol{\Sigma}, \mathbf{B}_\delta, t=2, s=0, M, \{y_{it}\}_{i=1,t=2}^{N,T}\right)\right] \\ vec\left[\mathbf{G}\left(\boldsymbol{\theta}, \boldsymbol{\Sigma}, \mathbf{B}_\delta, t=3, s=0, M, \{y_{it}\}_{i=1,t=2}^{N,T}\right)\right] \\ vec\left[\mathbf{G}\left(\boldsymbol{\theta}, \boldsymbol{\Sigma}, \mathbf{B}_\delta, t=3, s=1, M, \{y_{it}\}_{i=1,t=2}^{N,T}\right)\right] \\ vec\left[\mathbf{G}\left(\boldsymbol{\theta}, \boldsymbol{\Sigma}, \mathbf{B}_\delta, t=4, s=0, M, \{y_{it}\}_{i=1,t=2}^{N,T}\right)\right] \\ \vdots \\ vec\left[\mathbf{G}\left(\boldsymbol{\theta}, \boldsymbol{\Sigma}, \mathbf{B}_\delta, t=T, s=T-2, M, \{y_{it}\}_{i=1,t=2}^{N,T}\right)\right] \end{pmatrix}, \tag{54}$$

where $vec(\mathbf{X}) = (\mathbf{x}_1', \mathbf{x}_2', ..., \mathbf{x}_n')'$ for any matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n]$. Similarly, define $nT(T-1)/2$ vector $\boldsymbol{\xi}\left(\boldsymbol{\theta}, \sigma^2, \delta, T, M\right)$ based on stacking individual functions in $\Xi(\boldsymbol{\theta}, \boldsymbol{\Sigma}, \mathbf{B}_\delta, t, s, M)$ in a similar way, namely

$$\boldsymbol{\xi}(\boldsymbol{\theta}, \boldsymbol{\Sigma}, \mathbf{B}_\delta, T, M) = \begin{pmatrix} vec\left[\Xi(\boldsymbol{\theta}, \boldsymbol{\Sigma}, \mathbf{B}_\delta, t=2, s=0, M)\right] \\ vec\left[\Xi(\boldsymbol{\theta}, \boldsymbol{\Sigma}, \mathbf{B}_\delta, t=3, s=0, M)\right] \\ vec\left[\Xi(\boldsymbol{\theta}, \boldsymbol{\Sigma}, \mathbf{B}_\delta, t=3, s=1, M)\right] \\ vec\left[\Xi(\boldsymbol{\theta}, \boldsymbol{\Sigma}, \mathbf{B}_\delta, t=4, s=0, M)\right] \\ \vdots \\ vec\left[\Xi(\boldsymbol{\theta}, \boldsymbol{\Sigma}, \mathbf{B}_\delta, t=T, s=T-2, M)\right] \end{pmatrix}. \tag{55}$$

There are $k + (n+1)n$ unknown parameters collected in the vector $\boldsymbol{\psi} = \left[\boldsymbol{\theta}', vec(\boldsymbol{\Sigma})', vec(\mathbf{B}_\delta)'\right]'$, when M is known and finite. When $M \to \infty$, then there are $k + (n+1)n/2$ unknown parameters $\boldsymbol{\psi} = \left[\boldsymbol{\theta}', vec(\boldsymbol{\Sigma})'\right]'$. Let us denote the number of unknown parameters as n_ψ and denote the space of the unknown parameters as Ω. Also, same as earlier let us write $\boldsymbol{\xi}(\boldsymbol{\theta}, \boldsymbol{\Sigma}, \mathbf{B}_\delta, T, M)$ and $\mathbf{g}\left(\boldsymbol{\theta}, \boldsymbol{\Sigma}, \mathbf{B}_\delta, T, M, \{y_{it}\}_{i=1,t=2}^{N,T}\right)$ simply as $\boldsymbol{\xi}_{TM}(\boldsymbol{\psi})$ and $\mathbf{g}_{TM}(\boldsymbol{\psi})$, respectively to economize on notations. We assume Assumption 4 (when $M \to \infty$) or Assumption 7 (when M is finite) holds, which is sufficient for identification of the unknown parameters in the vector $\boldsymbol{\psi}$.

Using the same arguments as before, it can be established that $\hat{\boldsymbol{\psi}}_{BMM}$ defined as

$$\hat{\boldsymbol{\psi}}_{BMM} = \arg\min_{\boldsymbol{\psi}\in\Omega} \left\| \mathbf{g}_{TM}\left(\boldsymbol{\psi}\right) \right\|, \tag{56}$$

where $\mathbf{g}_{TM}\left(\boldsymbol{\psi}\right)$ is defined in (54), is consistent as $N \to \infty$ and T is fixed.

Inference can also be conducted in a similar way as in the simpler AR(1) case in Section 4. In particular, inference can be conducted using the semi-parametric estimator of $\boldsymbol{\Sigma}_{\boldsymbol{\psi}}$,

$$\hat{\boldsymbol{\Sigma}}_{\psi} = \left(\hat{\mathbf{J}}_{\boldsymbol{\xi}}'\hat{\mathbf{J}}_{\boldsymbol{\xi}}\right)^{-1}\hat{\mathbf{J}}_{\boldsymbol{\xi}}'\hat{\boldsymbol{\Theta}}\hat{\mathbf{J}}_{\boldsymbol{\xi}}\left(\hat{\mathbf{J}}_{\boldsymbol{\xi}}'\hat{\mathbf{J}}_{\boldsymbol{\xi}}\right)^{-1},$$

where $\hat{\mathbf{J}}_{\boldsymbol{\xi}} = \mathbf{J}_{\boldsymbol{\xi}}\left(\hat{\boldsymbol{\psi}}_{BMM}\right)$, $\mathbf{J}_{\boldsymbol{\xi}}\left(\boldsymbol{\psi}\right)$ is the Jacobian of $\boldsymbol{\xi}_{TM}\left(\boldsymbol{\psi}\right)$, the function $\boldsymbol{\xi}_{TM}\left(\boldsymbol{\psi}\right)$ is given by (55) in the case of panel VAR(1) model, $\hat{\boldsymbol{\Theta}}$ is partitioned into $n \times n$ submatrices $\hat{\boldsymbol{\Phi}}_{\ell h}$, for $\ell, h = 1, 2, ..., T\left(T-1\right)/2$, and $\hat{\boldsymbol{\Phi}}_{\ell h}$ is given by

$$\hat{\boldsymbol{\Phi}}_{\ell h} = \hat{\boldsymbol{\Phi}}^{*}_{t(\ell),s(\ell),r(\ell,h),p(\ell,h)}, \tag{57}$$

in which the index functions $t\left(\ell\right), s\left(\ell\right), r\left(\ell, h\right), p\left(\ell, h\right)$ are defined in (B.3)-(B.5), and $\hat{\boldsymbol{\Phi}}^{*}_{tsrp}$ is non-parametric estimate defined by

$$\hat{\boldsymbol{\Phi}}^{*}_{tsrp} = \frac{1}{N}\sum_{i=1}^{N}\left(\Delta\mathbf{y}_{it}\Delta\mathbf{y}_{i,t-s}' - \hat{\boldsymbol{\Pi}}_{t,t-s}\right)\cdot\left(\Delta\mathbf{y}_{i,t-r}\Delta\mathbf{y}_{i,t-p}' - \hat{\boldsymbol{\Pi}}_{t-r,t-p}\right),$$

where

$$\hat{\boldsymbol{\Pi}}_{t,t-s} = \frac{1}{N}\sum_{i=1}^{N}\Delta\mathbf{y}_{it}\Delta\mathbf{y}_{i,t-s}'.$$

Exact expression for the asymptotic covariance matrix $\boldsymbol{\Theta} = Avar\left[\mathbf{g}_{TM}\left(\boldsymbol{\psi}\right)\right]$, with $\mathbf{g}_{TM}\left(\boldsymbol{\psi}\right)$ defined by (54), can also be derived, as in (37), but this expression is not very insightful and quite complicated and therefore we shall not pursue the derivation of $\boldsymbol{\Theta}$ here.

6.2 AR(p) and VAR(p) models

Consider now the following panel AR(p) model for the variable z_{it},

$$z_{it} = c_i + \sum_{\ell=1}^{p} \rho_{i\ell} z_{i,t-\ell} + e_{it}, \tag{58}$$

where $\rho_{i\ell}$ unknown coefficients and e_{it} is the error term. This model can be equivalently written as a panel VAR(1) model (47) with

$$\underset{p\times p}{\mathbf{\Phi}_i} = \begin{pmatrix} \rho_{i1} & \rho_{i2} & \rho_{i3} & \cdots & \rho_{ip} \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix}, \mathbf{y}_{it} = \begin{pmatrix} z_{it} \\ z_{i,t-1} \\ z_{i,t-2} \\ \vdots \\ z_{i,t-p+1} \end{pmatrix} \text{ and } \mathbf{u}_{it} = \begin{pmatrix} e_{it} \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Note that $E\left(\mathbf{u}_{it}\mathbf{u}_{it}'\right)$ is singular, which is not ruled out in the previous analysis. The estimation of AR(p) models therefore can be conducted by using the companion VAR(1) representation. Similarly, VAR(p) model in $n_z \times 1$ vector $\mathbf{z}_{it}$ can be represented by VAR(1) model in $n \times 1$ vector $\mathbf{y}_{it} = \left(\mathbf{z}_{it}', \mathbf{z}_{i,t-1}', ..., \mathbf{z}_{i,t-p+1}'\right)'$.

7 Conclusion

This paper extends the literature on estimation of short-T dynamic panel VARs in a number of directions. We derive novel bias-corrected moment conditions, and allow for randomly distributed slope coefficients from a distribution with unknown parameters but known functional form. Our approach works for T as small as 4, provided that all parameters are properly identified.

Short time dimension is no excuse for the assumption of homogeneous coefficients. This paper is a step towards understanding the problem of estimation in short-T dynamic heterogeneous panels and we hope that this paper will stimulate future research on the subject.

Table 1: Bias and RMSE findings for the estimation of $\theta_\mu = E(\phi_i) = 0.6$ in experiments with homogeneous autoregressive coefficients

($\phi_i = 0.6$ for all i)

	Bias (×100)					RMSE (×100)				
(N,T)	**4**	**5**	**6**	**7**	**10**	**4**	**5**	**6**	**7**	**10**
	Experiments with cross-sectionally independent errors ($b_u = 0$)									
	Assuming correctly homogeneous coefficients									
1000	-0.02	-0.01	0.05	-0.01	-0.06	5.20	3.99	3.38	2.87	2.16
5000	-0.03	0.08	-0.07	-0.03	0.01	2.35	1.81	1.50	1.30	0.92
10 000	-0.05	0.04	0.02	-0.03	-0.01	1.65	1.32	1.03	0.91	0.67
50 000	-0.01	-0.02	-0.01	0.01	0.00	0.75	0.57	0.48	0.42	0.30
100 000	0.01	-0.01	0.00	-0.01	0.00	0.53	0.42	0.33	0.29	0.21
	Assuming incorrectly uniformly distributed coefficients									
1000	-0.02	-0.01	0.05	-0.01	-0.06	5.20	3.99	3.38	2.87	2.16
5000	-0.03	0.08	-0.07	-0.03	0.01	2.35	1.81	1.50	1.30	0.92
10 000	-0.05	0.04	0.02	-0.03	-0.01	1.65	1.32	1.03	0.91	0.67
50 000	-0.01	-0.02	-0.01	0.01	0.00	0.75	0.57	0.48	0.42	0.30
100 000	0.01	-0.01	0.00	-0.01	0.00	0.53	0.42	0.33	0.29	0.21
	Assuming incorrectly normally distributed coefficients									
1000	0.90	0.91	0.85	0.68	0.40	4.62	3.61	3.12	2.65	1.96
5000	0.78	0.71	0.48	0.41	0.27	2.24	1.79	1.42	1.23	0.87
10 000	0.64	0.55	0.43	0.29	0.16	1.68	1.32	1.03	0.90	0.63
50 000	0.43	0.26	0.19	0.18	0.08	0.89	0.60	0.50	0.43	0.28
100 000	0.33	0.20	0.16	0.10	0.06	0.64	0.45	0.36	0.30	0.19
	Assuming incorrectly Beta distributed coefficients									
1000	1.91	1.41	1.17	0.91	0.52	4.99	3.72	3.22	2.71	1.98
5000	1.08	0.82	0.55	0.46	0.28	2.46	1.87	1.47	1.26	0.88
10 000	0.80	0.60	0.46	0.31	0.17	1.83	1.37	1.06	0.92	0.63
50 000	0.46	0.27	0.20	0.18	0.08	0.94	0.61	0.50	0.44	0.28
100 000	0.34	0.20	0.16	0.10	0.06	0.66	0.46	0.37	0.30	0.19
	Experiments with spatially correlated errors ($b_u = 0.6$)									
	Assuming correctly homogeneous coefficients									
1000	-0.07	0.00	0.00	-0.11	-0.06	7.00	5.70	4.69	3.95	2.95
5000	-0.03	-0.03	0.08	-0.12	-0.02	3.18	2.53	2.05	1.78	1.26
10 000	0.06	-0.04	-0.02	-0.04	-0.03	2.32	1.80	1.43	1.25	0.92
50 000	-0.03	-0.02	-0.02	-0.01	-0.01	1.00	0.80	0.66	0.55	0.42
100 000	-0.02	0.00	0.01	-0.01	0.01	0.74	0.57	0.45	0.39	0.29
	Assuming incorrectly uniformly distributed coefficients									
1000	1.57	1.52	1.20	0.99	0.68	5.89	5.00	4.22	3.60	2.64
5000	1.16	0.94	0.83	0.54	0.36	3.09	2.43	2.01	1.68	1.21
10 000	1.07	0.74	0.59	0.44	0.22	2.39	1.79	1.43	1.21	0.86
50 000	0.58	0.41	0.28	0.22	0.10	1.20	0.89	0.69	0.56	0.39
100 000	0.43	0.29	0.23	0.16	0.09	0.92	0.64	0.50	0.41	0.28
	Assuming incorrectly normally distributed coefficients									
1000	0.85	0.99	0.82	0.68	0.49	6.05	5.08	4.26	3.61	2.67
5000	0.86	0.76	0.71	0.45	0.34	2.96	2.35	1.95	1.64	1.19
10 000	0.87	0.63	0.51	0.39	0.21	2.26	1.72	1.38	1.18	0.85
50 000	0.51	0.37	0.26	0.21	0.10	1.12	0.85	0.67	0.55	0.39
100 000	0.40	0.28	0.22	0.16	0.09	0.87	0.62	0.49	0.41	0.28
	Assuming incorrectly Beta distributed coefficients									
1000	2.53	1.86	1.39	1.11	0.73	6.42	5.13	4.27	3.63	2.65
5000	1.34	0.99	0.84	0.56	0.37	3.27	2.47	2.03	1.69	1.21
10 000	1.14	0.75	0.59	0.44	0.22	2.48	1.80	1.43	1.21	0.86
50 000	0.58	0.40	0.27	0.21	0.10	1.20	0.88	0.68	0.56	0.39
100 000	0.43	0.29	0.23	0.16	0.09	0.91	0.63	0.50	0.41	0.28

Notes: Data is generated based on $y_{it} = c_i + \phi_i y_{i,t-1} + u_{it}$, for $i = 1, 2, ..., N$ and $t = 0, 1, 2, ..., T$, with starting values $y_{i,-1} = \mu_i \sim IIDN(1, 1)$ and fixed effects $c_i = (1 - \phi_i)\mu_i$, for all i. Errors u_{it} are generated to be heteroskedastic and weakly cross sectionally dependent, see (45)-(46). See Section 5 for complete description of the Monte Carlo experiments.

Table 2: Bias and RMSE findings for the estimation of $\theta_\mu = E(\phi_i) = 0.6$ in experiments with uniformly distributed autoregressive coefficients

($\phi_i \sim IIDU[\phi_{\min}, \phi_{\max}]$ for all i, where $\phi_{\min} = \theta_\mu - \frac{\sqrt{12\theta_{\sigma^2}}}{2} = 0.4268$ and $\phi_{\max} = 2\theta_\mu - \phi_{\min} = 0.7732$)

	Bias (×100)					**RMSE (×100)**				
(N,T)	**4**	**5**	**6**	**7**	**10**	**4**	**5**	**6**	**7**	**10**
	Experiments with cross-sectionally independent errors ($b_u = 0$)									
	Assuming incorrectly homogeneous coefficients									
1000	-1.11	-1.09	-1.00	-0.69	-0.76	5.27	4.25	3.64	3.06	2.34
5000	-0.95	-0.93	-0.92	-0.83	-0.75	2.52	2.09	1.78	1.57	1.24
10 000	-0.94	-0.98	-0.94	-0.90	-0.72	1.96	1.64	1.44	1.30	1.00
50 000	-0.95	-0.98	-0.90	-0.86	-0.73	1.21	1.14	1.03	0.96	0.79
100 000	-0.96	-0.96	-0.93	-0.85	-0.72	1.10	1.05	0.99	0.90	0.76
	Assuming correctly uniformly distributed coefficients									
1000	0.64	0.51	0.41	0.49	0.23	4.37	3.53	2.99	2.62	1.90
5000	0.46	0.25	0.18	0.13	0.03	2.15	1.67	1.39	1.20	0.87
10 000	0.31	0.12	0.03	0.02	0.01	1.66	1.24	1.04	0.88	0.61
50 000	0.04	-0.02	-0.02	-0.01	0.00	0.87	0.62	0.53	0.42	0.28
100 000	0.01	-0.01	-0.02	0.00	0.00	0.69	0.48	0.37	0.31	0.20
	Assuming incorrectly normally distributed coefficients									
1000	0.01	0.07	0.07	0.23	0.02	4.43	3.53	3.02	2.61	1.91
5000	0.17	0.06	0.04	0.02	-0.02	2.04	1.60	1.32	1.15	0.85
10 000	0.10	-0.02	-0.07	-0.06	-0.01	1.55	1.17	0.99	0.84	0.60
50 000	-0.06	-0.09	-0.07	-0.04	0.00	0.80	0.58	0.50	0.40	0.28
100 000	-0.08	-0.07	-0.07	-0.03	0.01	0.64	0.45	0.35	0.29	0.20
	Assuming incorrectly Beta distributed coefficients									
1000	1.27	0.69	0.52	0.56	0.27	4.77	3.63	3.03	2.65	1.92
5000	0.58	0.27	0.18	0.13	0.03	2.28	1.69	1.40	1.20	0.87
10 000	0.36	0.12	0.02	0.01	0.01	1.72	1.24	1.04	0.88	0.61
50 000	0.02	-0.04	-0.04	-0.02	0.00	0.87	0.61	0.52	0.42	0.28
100 000	-0.01	-0.04	-0.04	-0.02	0.00	0.69	0.47	0.37	0.30	0.20
	Experiments with spatially correlated errors ($b_u = 0.6$)									
	Assuming incorrectly homogeneous coefficients									
1000	-0.91	-0.91	-1.22	-0.93	-0.73	7.33	5.67	4.92	4.13	3.13
5000	-1.07	-1.05	-0.91	-0.87	-0.74	3.47	2.78	2.29	2.01	1.57
10 000	-1.01	-0.91	-0.94	-0.87	-0.73	2.55	1.98	1.77	1.56	1.20
50 000	-0.96	-0.97	-0.89	-0.84	-0.74	1.41	1.26	1.12	1.01	0.86
100 000	-0.96	-0.96	-0.92	-0.84	-0.72	1.21	1.12	1.04	0.93	0.78
	Assuming correctly uniformly distributed coefficients									
1000	0.84	0.84	0.44	0.42	0.34	5.99	4.71	3.94	3.50	2.61
5000	0.48	0.37	0.30	0.20	0.07	2.96	2.28	1.86	1.57	1.19
10 000	0.37	0.26	0.13	0.05	0.05	2.16	1.60	1.35	1.15	0.81
50 000	0.13	0.00	-0.01	-0.01	0.00	1.13	0.82	0.68	0.56	0.37
100 000	0.03	0.01	-0.03	0.01	0.00	0.85	0.62	0.49	0.42	0.26
	Assuming incorrectly normally distributed coefficients									
1000	0.07	0.21	-0.10	0.01	0.04	6.16	4.84	4.10	3.58	2.69
5000	0.07	0.08	0.10	0.03	-0.01	2.88	2.21	1.80	1.53	1.18
10 000	0.08	0.09	-0.01	-0.05	0.00	2.06	1.53	1.29	1.11	0.80
50 000	-0.01	-0.09	-0.07	-0.05	-0.01	1.04	0.77	0.64	0.53	0.36
100 000	-0.07	-0.06	-0.08	-0.03	0.00	0.78	0.58	0.46	0.40	0.26
	Assuming incorrectly Beta distributed coefficients									
1000	1.88	1.25	0.68	0.58	0.43	6.56	4.86	3.99	3.53	2.63
5000	0.73	0.44	0.33	0.22	0.08	3.18	2.33	1.88	1.58	1.19
10 000	0.47	0.27	0.13	0.05	0.05	2.28	1.62	1.35	1.15	0.81
50 000	0.12	-0.02	-0.02	-0.02	0.00	1.14	0.81	0.67	0.55	0.37
100 000	0.02	-0.01	-0.05	-0.01	0.00	0.84	0.61	0.48	0.41	0.26

Notes: Data is generated based on $y_{it} = c_i + \phi_i y_{i,t-1} + u_{it}$, for $i = 1, 2, ..., N$ and $t = 0, 1, 2, ..., T$, with starting values $y_{i,-1} = \mu_i \sim IIDN(1, 1)$ and fixed effects $c_i = (1 - \phi_i)\mu_i$, for all i. Errors u_{it} are generated to be heteroskedastic and weakly cross sectionally dependent, see (45)-(46). See Section 5 for complete description of the Monte Carlo experiments.

Table 3: Bias and RMSE findings for the estimation of $\theta_\mu = E(\phi_i) = 0.6$ in experiments with normally distributed autoregressive coefficients

$(\phi_i \sim IIDN(\theta_\mu, \theta_\sigma^2)$ for all $i)$

(N,T)	Bias (×100)					RMSE (×100)				
	4	**5**	**6**	**7**	**10**	**4**	**5**	**6**	**7**	**10**
	Experiments with cross-sectionally independent errors ($b_u = 0$)									
	Assuming incorrectly homogeneous coefficients									
1000	-1.01	-0.95	-0.82	-0.93	-0.78	5.49	4.34	3.58	3.12	2.38
5000	-0.89	-0.98	-0.88	-0.87	-0.73	2.63	2.08	1.76	1.62	1.22
10 000	-1.02	-0.94	-0.88	-0.88	-0.77	1.98	1.62	1.40	1.28	1.03
50 000	-0.92	-0.96	-0.92	-0.86	-0.73	1.19	1.12	1.04	0.96	0.79
100 000	-0.95	-0.96	-0.91	-0.87	-0.73	1.09	1.05	0.97	0.91	0.77
	Assuming incorrectly uniformly distributed coefficients									
1000	0.76	0.71	0.56	0.40	0.22	4.56	3.59	3.02	2.54	1.93
5000	0.51	0.29	0.23	0.10	0.03	2.31	1.69	1.40	1.22	0.85
10 000	0.26	0.18	0.10	0.02	-0.03	1.67	1.25	1.02	0.90	0.62
50 000	0.08	0.03	0.03	0.00	0.00	0.87	0.65	0.51	0.42	0.28
100 000	0.03	0.03	0.02	0.00	-0.01	0.69	0.48	0.38	0.31	0.19
	Assuming correctly normally distributed coefficients									
1000	0.12	0.26	0.24	0.08	0.02	4.60	3.61	3.01	2.56	1.95
5000	0.20	0.08	0.08	-0.01	-0.01	2.19	1.60	1.34	1.18	0.83
10 000	0.05	0.04	0.00	-0.06	-0.05	1.56	1.17	0.97	0.86	0.61
50 000	-0.02	-0.05	-0.03	-0.03	0.00	0.80	0.61	0.47	0.40	0.27
100 000	-0.05	-0.04	-0.03	-0.03	0.00	0.64	0.44	0.35	0.29	0.19
	Assuming incorrectly Beta distributed coefficients									
1000	1.40	0.90	0.67	0.49	0.26	4.97	3.69	3.06	2.59	1.95
5000	0.64	0.31	0.23	0.10	0.04	2.45	1.73	1.41	1.23	0.85
10 000	0.31	0.17	0.09	0.01	-0.03	1.73	1.25	1.02	0.90	0.62
50 000	0.07	0.00	0.01	-0.01	0.00	0.86	0.64	0.50	0.42	0.28
100 000	0.01	0.00	0.00	-0.02	-0.01	0.68	0.47	0.37	0.31	0.19
	Experiments with spatially correlated errors ($b_u = 0.6$)									
	Assuming incorrectly homogeneous coefficients									
1000	-0.68	-0.92	-1.03	-0.94	-0.82	7.36	5.69	4.83	4.28	3.14
5000	-0.98	-0.96	-0.94	-0.84	-0.73	3.47	2.77	2.37	2.02	1.52
10 000	-0.96	-0.92	-0.94	-0.83	-0.77	2.54	2.06	1.76	1.55	1.21
50 000	-0.96	-0.95	-0.92	-0.84	-0.73	1.42	1.23	1.13	1.03	0.85
100 000	-0.94	-0.95	-0.92	-0.85	-0.73	1.21	1.11	1.03	0.95	0.79
	Assuming incorrectly uniformly distributed coefficients									
1000	1.14	0.85	0.52	0.53	0.27	6.01	4.72	4.05	3.51	2.58
5000	0.58	0.43	0.27	0.21	0.11	2.96	2.27	1.88	1.63	1.13
10 000	0.49	0.31	0.13	0.09	0.01	2.20	1.66	1.37	1.19	0.81
50 000	0.15	0.07	0.03	0.02	-0.01	1.13	0.80	0.68	0.58	0.39
100 000	0.10	0.05	0.02	0.00	-0.01	0.86	0.63	0.49	0.43	0.26
	Assuming correctly normally distributed coefficients									
1000	0.34	0.21	0.03	0.08	-0.04	6.18	4.86	4.16	3.63	2.66
5000	0.17	0.15	0.06	0.05	0.03	2.87	2.21	1.84	1.59	1.11
10 000	0.19	0.12	-0.02	-0.01	-0.03	2.08	1.59	1.31	1.15	0.79
50 000	0.01	-0.03	-0.04	-0.03	-0.02	1.03	0.74	0.63	0.54	0.38
100 000	0.00	-0.03	-0.03	-0.03	-0.01	0.79	0.58	0.46	0.40	0.26
	Assuming incorrectly Beta distributed coefficients									
1000	2.24	1.24	0.74	0.70	0.36	6.61	4.84	4.09	3.53	2.60
5000	0.83	0.49	0.29	0.23	0.12	3.19	2.32	1.91	1.64	1.14
10 000	0.61	0.33	0.13	0.09	0.02	2.34	1.69	1.38	1.20	0.81
50 000	0.15	0.05	0.01	0.00	-0.01	1.14	0.80	0.67	0.57	0.39
100 000	0.09	0.02	0.00	-0.02	-0.01	0.86	0.62	0.48	0.42	0.26

Notes: Data is generated based on $y_{it} = c_i + \phi_i y_{i,t-1} + u_{it}$, for $i = 1, 2, ..., N$ and $t = 0, 1, 2, ..., T$, with starting values $y_{i,-1} = \mu_i \sim IIDN(1, 1)$ and fixed effects $c_i = (1 - \phi_i)\mu_i$, for all i. Errors u_{it} are generated to be heteroskedastic and weakly cross sectionally dependent, see (45)-(46). See Section 5 for complete description of the Monte Carlo experiments.

Table 4: Bias and RMSE findings for the estimation of $\theta_\mu = E(\phi_i) = 0.6$ in experiments with beta distributed autoregressive coefficients

$\left(\phi_i \sim Beta\left(\phi_\alpha, \phi_\beta\right)\right.$ for all i, where $\phi_\alpha = \theta_\mu v = 13.8$, $\phi_\beta = (1-\theta_\mu)v = 9.2$, $\left. v = \left[\theta_\mu (1-\theta_\mu)/\theta_\sigma^2 - 1\right]\right)$

	Bias (×100)					**RMSE (×100)**				
(N,T)	**4**	**5**	**6**	**7**	**10**	**4**	**5**	**6**	**7**	**10**
	Experiments with cross-sectionally independent errors ($b_u = 0$)									
	Assuming incorrectly homogeneous coefficients									
1000	-0.92	-1.11	-0.92	-0.91	-0.69	5.47	4.28	3.58	3.15	2.35
5000	-0.90	-0.96	-0.94	-0.86	-0.76	2.60	2.07	1.76	1.58	1.24
10 000	-1.07	-0.95	-0.90	-0.89	-0.73	2.01	1.64	1.40	1.29	1.00
50 000	-0.96	-0.98	-0.91	-0.87	-0.73	1.22	1.15	1.03	0.96	0.80
100 000	-0.96	-0.97	-0.91	-0.86	-0.73	1.10	1.06	0.98	0.91	0.77
	Assuming incorrectly uniformly distributed coefficients									
1000	0.86	0.54	0.50	0.31	0.25	4.53	3.46	2.99	2.62	1.90
5000	0.49	0.26	0.14	0.16	0.01	2.30	1.63	1.40	1.21	0.84
10 000	0.29	0.17	0.05	0.03	0.00	1.65	1.24	1.02	0.89	0.60
50 000	0.07	-0.02	0.00	0.00	0.00	0.89	0.64	0.51	0.42	0.28
100 000	0.01	0.01	0.02	0.00	0.00	0.67	0.46	0.37	0.31	0.20
	Assuming incorrectly normally distributed coefficients									
1000	0.23	0.09	0.16	0.03	0.07	4.60	3.51	2.99	2.62	1.92
5000	0.20	0.07	0.00	0.03	-0.04	2.18	1.57	1.34	1.16	0.83
10 000	0.05	0.03	-0.04	-0.05	-0.02	1.53	1.17	0.97	0.85	0.59
50 000	-0.03	-0.09	-0.06	-0.04	0.00	0.81	0.60	0.48	0.40	0.27
100 000	-0.07	-0.06	-0.03	-0.03	0.00	0.62	0.43	0.35	0.29	0.19
	Assuming correctly Beta distributed coefficients									
1000	1.50	0.75	0.61	0.39	0.29	4.96	3.56	3.03	2.66	1.92
5000	0.62	0.28	0.14	0.16	0.02	2.44	1.65	1.41	1.22	0.84
10 000	0.34	0.17	0.05	0.02	0.00	1.72	1.26	1.02	0.89	0.60
50 000	0.06	-0.04	-0.02	-0.01	0.00	0.88	0.63	0.51	0.42	0.28
100 000	0.00	-0.02	0.00	-0.01	0.00	0.66	0.45	0.37	0.30	0.20
	Experiments with spatially correlated errors ($b_u = 0.6$)									
	Assuming incorrectly homogeneous coefficients									
1000	-1.22	-1.10	-0.91	-0.95	-0.85	7.28	5.85	4.82	4.16	3.15
5000	-0.92	-1.00	-0.97	-0.88	-0.80	3.42	2.77	2.37	2.05	1.58
10 000	-0.96	-0.94	-0.91	-0.86	-0.75	2.53	1.99	1.71	1.53	1.22
50 000	-0.96	-0.96	-0.92	-0.87	-0.75	1.42	1.26	1.13	1.04	0.86
100 000	-0.97	-0.97	-0.91	-0.88	-0.74	1.22	1.13	1.03	0.97	0.80
	Assuming incorrectly uniformly distributed coefficients									
1000	0.69	0.66	0.70	0.42	0.22	5.85	4.85	4.05	3.42	2.55
5000	0.61	0.38	0.25	0.21	0.03	2.98	2.25	1.89	1.63	1.16
10 000	0.42	0.26	0.12	0.12	0.03	2.21	1.59	1.35	1.15	0.82
50 000	0.13	0.01	0.00	0.01	-0.02	1.13	0.81	0.67	0.58	0.37
100 000	0.04	0.02	0.00	0.00	-0.01	0.86	0.61	0.51	0.42	0.27
	Assuming incorrectly normally distributed coefficients									
1000	-0.09	0.06	0.19	0.01	-0.09	6.09	4.97	4.15	3.52	2.65
5000	0.21	0.11	0.05	0.03	-0.05	2.89	2.21	1.85	1.58	1.15
10 000	0.14	0.08	-0.01	0.01	-0.01	2.09	1.51	1.29	1.10	0.80
50 000	-0.01	-0.08	-0.07	-0.04	-0.02	1.03	0.76	0.63	0.54	0.37
100 000	-0.06	-0.06	-0.05	-0.04	-0.01	0.79	0.57	0.47	0.40	0.26
	Assuming correctly Beta distributed coefficients									
1000	1.68	1.05	0.92	0.57	0.31	6.33	4.98	4.09	3.46	2.56
5000	0.84	0.43	0.29	0.23	0.04	3.17	2.30	1.92	1.64	1.17
10 000	0.54	0.28	0.13	0.13	0.04	2.32	1.62	1.36	1.16	0.82
50 000	0.13	-0.01	-0.02	-0.01	-0.02	1.14	0.80	0.66	0.57	0.38
100 000	0.03	-0.01	-0.02	-0.01	-0.01	0.86	0.60	0.50	0.41	0.27

Notes: Data is generated based on $y_{it} = c_i + \phi_i y_{i,t-1} + u_{it}$, for $i = 1, 2, ..., N$ and $t = 0, 1, 2, ..., T$, with starting values $y_{i,-1} = \mu_i \sim IIDN(1,1)$ and fixed effects $c_i = (1-\phi_i)\mu_i$, for all i. Errors u_{it} are generated to be heteroskedastic and weakly cross sectionally dependent, see (45)-(46). See Section 5 for complete description of the Monte Carlo experiments.

Table 5: Size and power (5% level) findings for the estimation of $\theta_\mu = E(\phi_i) = 0.6$ in experiments with homogeneous autoregressive coefficients

($\phi_i = 0.6$ for all i)

	Size ($\times100$, $H_0 : \theta_\mu = 0.6$)					**Power** ($\times100$, $H_1 : \theta_\mu = 0.7$)				
(N,T)	**4**	**5**	**6**	**7**	**10**	**4**	**5**	**6**	**7**	**10**
	Experiments with cross-sectionally independent errors ($b_u = 0$)									
	Assuming correctly homogeneous coefficients									
1000	5.70	4.10	5.20	5.50	5.80	48.00	69.20	84.45	94.15	99.75
5000	4.55	4.50	4.70	5.40	3.60	98.80	99.95	100.00	100.00	100.00
10 000	5.15	5.75	4.50	4.65	5.80	100.00	100.00	100.00	100.00	100.00
50 000	5.35	4.45	5.20	5.60	4.50	100.00	100.00	100.00	100.00	100.00
100 000	4.90	5.85	4.55	5.05	4.55	100.00	100.00	100.00	100.00	100.00
	Assuming incorrectly uniformly distributed coefficients									
1000	3.90	4.00	4.90	3.40	6.85	21.20	47.65	74.70	91.15	100.00
5000	4.60	4.70	3.75	3.95	5.40	89.45	99.95	100.00	100.00	100.00
10 000	6.00	4.75	3.75	4.25	6.00	100.00	100.00	100.00	100.00	100.00
50 000	5.55	2.95	3.30	5.10	5.80	100.00	100.00	100.00	100.00	100.00
100 000	3.85	3.10	3.80	4.00	5.45	100.00	100.00	100.00	100.00	100.00
	Assuming incorrectly normally distributed coefficients									
1000	3.45	3.60	4.45	2.95	6.85	24.35	50.45	75.35	91.60	100.00
5000	3.00	3.95	3.20	3.70	5.35	90.00	99.95	100.00	100.00	100.00
10 000	4.15	4.15	3.35	4.20	6.00	100.00	100.00	100.00	100.00	100.00
50 000	5.60	3.30	3.40	5.20	5.75	100.00	100.00	100.00	100.00	100.00
100 000	4.25	3.35	4.00	4.00	5.50	100.00	100.00	100.00	100.00	100.00
	Assuming incorrectly Beta distributed coefficients									
1000	6.00	4.40	5.00	3.60	6.90	11.30	44.85	73.85	90.90	100.00
5000	4.45	4.55	3.60	3.80	5.50	89.10	99.95	100.00	100.00	100.00
10 000	5.35	4.05	3.55	4.30	6.10	100.00	100.00	100.00	100.00	100.00
50 000	4.75	3.20	3.40	5.20	5.75	100.00	100.00	100.00	100.00	100.00
100 000	3.60	3.00	3.75	3.95	5.50	100.00	100.00	100.00	100.00	100.00

Notes: Data is generated based on $y_{it} = c_i + \phi_i y_{i,t-1} + u_{it}$, for $i = 1, 2, ..., N$ and $t = 0, 1, 2, ..., T$, with starting values $y_{i,-1} = \mu_i \sim IIDN(1, 1)$ and fixed effects $c_i = (1 - \phi_i)\mu_i$, for all i. Errors u_{it} are generated to be heteroskedastic and weakly cross sectionally dependent, see (45)-(46). See Section 5 for complete description of the Monte Carlo experiments.

Table 6: Size and power (5% level) findings for the estimation of $\theta_\mu = E(\phi_i) = 0.6$ in experiments with uniformly distributed autoregressive coefficients

($\phi_i \sim IIDU[\phi_{\min}, \phi_{\max}]$ for all i, where $\phi_{\min} = \theta_\mu - \frac{\sqrt{12\theta_{\sigma^2}}}{2} = 0.4268$ and $\phi_{\max} = 2\theta_\mu - \phi_{\min} = 0.7732$)

	Size ($\times 100$, $H_0 : \theta_\mu = 0.6$)					**Power** ($\times 100$, $H_1 : \theta_\mu = 0.7$)				
(N,T)	**4**	**5**	**6**	**7**	**10**	**4**	**5**	**6**	**7**	**10**
	Experiments with cross-sectionally independent errors ($b_u = 0$)									
	Assuming incorrectly homogeneous coefficients									
1000	5.40	6.85	7.15	5.50	7.05	56.15	77.50	88.95	94.90	99.95
5000	6.50	8.70	8.85	9.85	12.60	99.65	100.00	100.00	100.00	100.00
10 000	10.00	12.25	14.15	17.00	17.65	100.00	100.00	100.00	100.00	100.00
50 000	23.80	40.15	47.95	53.85	65.35	100.00	100.00	100.00	100.00	100.00
100 000	43.95	64.35	77.80	82.90	90.65	100.00	100.00	100.00	100.00	100.00
	Assuming correctly uniformly distributed coefficients									
1000	2.45	2.85	3.40	3.85	4.90	27.95	56.05	82.55	92.40	99.95
5000	3.25	2.90	3.30	3.10	5.25	95.10	100.00	100.00	100.00	100.00
10 000	4.20	3.00	3.95	3.85	5.15	100.00	100.00	100.00	100.00	100.00
50 000	4.00	3.25	5.00	4.85	6.20	100.00	100.00	100.00	100.00	100.00
100 000	4.70	5.00	5.65	5.05	5.30	100.00	100.00	100.00	100.00	100.00
	Assuming incorrectly normally distributed coefficients									
1000	2.05	2.95	3.50	3.65	4.95	31.45	59.35	83.75	92.90	100.00
5000	1.70	2.10	2.40	2.70	4.90	95.25	100.00	100.00	100.00	100.00
10 000	2.25	2.30	3.40	3.75	5.00	100.00	100.00	100.00	100.00	100.00
50 000	2.90	2.70	5.05	4.50	6.05	100.00	100.00	100.00	100.00	100.00
100 000	3.60	4.25	5.30	5.25	5.25	100.00	100.00	100.00	100.00	100.00
	Assuming incorrectly Beta distributed coefficients									
1000	4.45	3.90	3.40	3.65	5.15	15.25	53.20	81.50	92.15	99.90
5000	3.30	2.95	3.20	3.05	5.40	94.80	100.00	100.00	100.00	100.00
10 000	3.90	2.35	3.85	3.50	5.10	100.00	100.00	100.00	100.00	100.00
50 000	3.30	2.90	4.65	4.60	6.25	100.00	100.00	100.00	100.00	100.00
100 000	3.90	4.60	5.30	5.35	5.35	100.00	100.00	100.00	100.00	100.00

Notes: Data is generated based on $y_{it} = c_i + \phi_i y_{i,t-1} + u_{it}$, for $i = 1, 2, ..., N$ and $t = 0, 1, 2, ..., T$, with starting values $y_{i,-1} = \mu_i \sim IIDN(1, 1)$ and fixed effects $c_i = (1 - \phi_i)\mu_i$, for all i. Errors u_{it} are generated to be heteroskedastic and weakly cross sectionally dependent, see (45)-(46). See Section 5 for complete description of the Monte Carlo experiments.

Table 7: Size and power (5% level) findings for the estimation of $\theta_\mu = E(\phi_i) = 0.6$ in experiments with normally distributed autoregressive coefficients

$(\phi_i \sim IIDN(\theta_\mu, \theta_\sigma^2)$ for all $i)$

	Size ($\times 100$, $H_0 : \theta_\mu = 0.6$)					Power ($\times 100$, $H_1 : \theta_\mu = 0.7$)				
(N,T)	**4**	**5**	**6**	**7**	**10**	**4**	**5**	**6**	**7**	**10**
	Experiments with cross-sectionally independent errors ($b_u = 0$)									
	Assuming incorrectly homogeneous coefficients									
1000	6.45	6.85	6.45	6.90	7.70	55.30	74.75	89.45	96.40	99.80
5000	7.30	8.90	9.15	10.85	11.90	99.45	100.00	100.00	100.00	100.00
10 000	9.70	11.45	13.35	15.75	20.20	100.00	100.00	100.00	100.00	100.00
50 000	24.25	37.30	47.35	53.95	65.90	100.00	100.00	100.00	100.00	100.00
100 000	44.10	64.75	75.85	84.65	92.55	100.00	100.00	100.00	100.00	100.00
	Assuming incorrectly uniformly distributed coefficients									
1000	3.30	3.70	3.45	3.50	5.60	26.65	56.70	81.00	94.30	99.90
5000	5.35	4.05	3.20	3.95	4.20	94.45	100.00	100.00	100.00	100.00
10 000	4.10	3.45	3.10	4.30	5.40	100.00	100.00	100.00	100.00	100.00
50 000	3.80	4.55	4.85	4.30	4.95	100.00	100.00	100.00	100.00	100.00
100 000	4.65	5.05	5.10	5.05	4.80	100.00	100.00	100.00	100.00	100.00
	Assuming correctly normally distributed coefficients									
1000	3.25	3.30	3.50	3.75	5.55	31.25	59.25	82.40	94.65	99.90
5000	3.70	2.60	2.65	3.55	4.30	94.90	100.00	100.00	100.00	100.00
10 000	2.50	2.80	2.85	4.00	5.00	100.00	100.00	100.00	100.00	100.00
50 000	3.15	3.95	4.50	4.10	5.05	100.00	100.00	100.00	100.00	100.00
100 000	3.85	4.45	4.65	4.75	4.95	100.00	100.00	100.00	100.00	100.00
	Assuming incorrectly Beta distributed coefficients									
1000	5.55	3.85	3.55	3.75	5.70	16.35	53.05	79.85	93.80	99.90
5000	5.80	3.95	3.20	4.10	4.45	93.45	100.00	100.00	100.00	100.00
10 000	3.45	3.25	3.00	4.55	5.40	100.00	100.00	100.00	100.00	100.00
50 000	3.30	4.15	4.65	4.10	5.00	100.00	100.00	100.00	100.00	100.00
100 000	3.90	4.95	4.95	5.00	4.85	100.00	100.00	100.00	100.00	100.00

Notes: Data is generated based on $y_{it} = c_i + \phi_i y_{i,t-1} + u_{it}$, for $i = 1, 2, ..., N$ and $t = 0, 1, 2, ..., T$, with starting values $y_{i,-1} = \mu_i \sim IIDN(1, 1)$ and fixed effects $c_i = (1 - \phi_i)\mu_i$, for all i. Errors u_{it} are generated to be heteroskedastic and weakly cross sectionally dependent, see (45)-(46). See Section 5 for complete description of the Monte Carlo experiments.

Table 8: Size and power (5% level) findings for the estimation of $\theta_\mu = E(\phi_i) = 0.6$ in experiments with beta distributed autoregressive coefficients

$(\phi_i \sim Beta(\phi_\alpha, \phi_\beta)$ for all i, where $\phi_\alpha = \theta_\mu v = 13.8$, $\phi_\beta = (1-\theta_\mu)v = 9.2$, $v = [\theta_\mu(1-\theta_\mu)/\theta_\sigma^2 - 1])$

	Size ($\times 100$, $H_0 : \theta_\mu = 0.6$)					**Power** ($\times 100$, $H_1 : \theta_\mu = 0.7$)				
(N,T)	**4**	**5**	**6**	**7**	**10**	**4**	**5**	**6**	**7**	**10**
	Experiments with cross-sectionally independent errors ($b_u = 0$)									
	Assuming incorrectly homogeneous coefficients									
1000	6.05	6.30	6.70	7.15	7.30	54.10	76.65	89.15	95.45	99.95
5000	7.55	8.40	9.25	10.00	12.10	99.50	100.00	100.00	100.00	100.00
10 000	10.75	11.90	13.30	16.65	19.60	100.00	100.00	100.00	100.00	100.00
50 000	24.05	39.35	47.00	55.50	66.15	100.00	100.00	100.00	100.00	100.00
100 000	42.65	65.35	75.30	83.05	91.65	100.00	100.00	100.00	100.00	100.00
	Assuming incorrectly uniformly distributed coefficients									
1000	2.65	3.05	3.55	3.90	4.75	27.40	57.85	81.50	93.90	100.00
5000	5.10	3.10	3.20	4.05	4.15	94.25	100.00	100.00	100.00	100.00
10 000	4.25	3.80	3.00	3.65	4.25	100.00	100.00	100.00	100.00	100.00
50 000	4.50	4.05	4.50	5.55	5.50	100.00	100.00	100.00	100.00	100.00
100 000	3.65	4.60	5.25	5.00	5.50	100.00	100.00	100.00	100.00	100.00
	Assuming incorrectly normally distributed coefficients									
1000	2.50	3.00	3.40	3.90	4.95	32.10	60.20	82.75	94.35	100.00
5000	2.50	1.90	2.90	3.10	4.00	94.35	100.00	100.00	100.00	100.00
10 000	2.80	2.65	2.75	3.20	4.25	100.00	100.00	100.00	100.00	100.00
50 000	3.05	3.45	4.50	5.05	5.15	100.00	100.00	100.00	100.00	100.00
100 000	3.30	4.05	4.95	4.95	5.20	100.00	100.00	100.00	100.00	100.00
	Assuming correctly Beta distributed coefficients									
1000	5.45	3.70	3.65	4.00	4.70	16.10	54.90	80.60	93.75	100.00
5000	4.95	2.60	3.35	3.70	4.40	93.40	100.00	100.00	100.00	100.00
10 000	3.80	3.85	3.15	3.50	4.25	100.00	100.00	100.00	100.00	100.00
50 000	3.40	3.55	4.40	5.35	5.40	100.00	100.00	100.00	100.00	100.00
100 000	3.30	4.30	5.15	4.90	5.40	100.00	100.00	100.00	100.00	100.00

Notes: Data is generated based on $y_{it} = c_i + \phi_i y_{i,t-1} + u_{it}$, for $i = 1, 2, ..., N$ and $t = 0, 1, 2, ..., T$, with starting values $y_{i,-1} = \mu_i \sim IIDN(1,1)$ and fixed effects $c_i = (1-\phi_i)\mu_i$, for all i. Errors u_{it} are generated to be heteroskedastic and weakly cross sectionally dependent, see (45)-(46). See Section 5 for complete description of the Monte Carlo experiments.

A Identification in the case of Monte Carlo experiments

There are 3 unknown parameters in $\boldsymbol{\psi} = \left(\theta_\alpha, \theta_\beta, \sigma^2\right)$. These parameters are identified if and only if the Jacobian

$$\mathbf{J}_{\boldsymbol{\xi}_{TM}}\left(\boldsymbol{\psi}\right) = \frac{\partial \boldsymbol{\xi}_{TM}\left(\boldsymbol{\psi}\right)}{\partial \boldsymbol{\psi}'}$$

has nonzero determinant for all values of $\boldsymbol{\psi} \in \Omega_\theta$. It is not straightforward to verify this condition analytically in general. Therefore, one could examine numerically whether the invertibility requirement in Assumption 4 is satisfied. Note that in the Monte Carlo experiments the ratios of the individual elements of $\boldsymbol{\xi}_{TM}\left(\boldsymbol{\psi}\right)$ do not depend on σ^2 and this parameter is identified so long as $\theta_\alpha, \theta_\beta$ are identified. Note also that if θ_α and θ_β are identified for $T = 4$ then they are identified also for any $T > 4$. Suppose that $T = 4$ and consider

$$\mathbf{g}\left(\theta_\alpha, \theta_\beta\right) = \begin{pmatrix} \xi\left(\boldsymbol{\theta},\sigma^2, \delta = 0, t = 2, s = 0, M = 0\right)/\xi\left(\boldsymbol{\theta},\sigma^2, \delta = 0, t = 3, s = 1, M = 0\right) \\ \xi\left(\boldsymbol{\theta},\sigma^2, \delta = 0, t = 3, s = 0, M = 0\right)/\xi\left(\boldsymbol{\theta},\sigma^2, \delta = 0, t = 4, s = 2, M = 0\right) \end{pmatrix}.$$

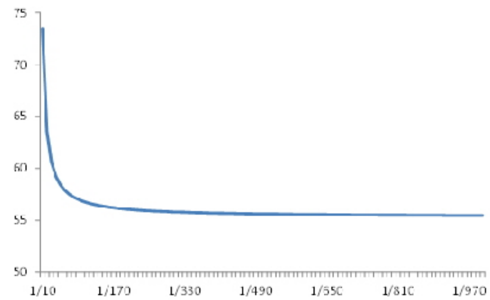
Sufficient condition for θ_α and θ_β to be identified when $T = 4$ is invertibility of $\mathbf{g}\left(\theta_\alpha, \theta_\beta\right)$. This would be the case if the Jacobian

$$\mathbf{J}_g\left(\theta_\alpha, \theta_\beta\right) = \frac{\partial \mathbf{g}\left(\theta_\alpha, \theta_\beta\right)}{\partial\left(\theta_\alpha, \theta_\beta\right)}$$

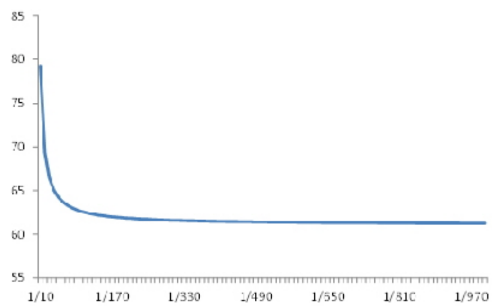
has nonzero determinant. This is again not straightforward to verify analytically, but it is possible to investigate whether $\left|\det\left[\mathbf{J}_g\left(\theta_\alpha, \theta_\beta\right)\right]\right|$ is bounded away from zero numerically. We consider a sequence of grids G_ε of parameter values $0 < \theta_\mu < 1$ and $0 < \theta_{\sigma^2} < 0.16$ with the size of the cells in the grid equal to ε. For a given value of the grid cell size, we compute $f_\varepsilon = \min\left|\det\left[\mathbf{J}_g\left(\theta_\alpha, \theta_\beta\right)\right]\right|$ and ascertain the behavior of f_ε as $\varepsilon \to 0$ is numerically. Figure 1 shows f_ε for $\varepsilon = 1/S$, $S = 10, 20, 30, 40, ...1000$.

Figure 1: Values of $f_\varepsilon = \min_{\theta_\alpha, \theta_\beta \in G_\varepsilon} |\det [\mathbf{J}_g (\theta_\alpha, \theta_\beta)]|$,

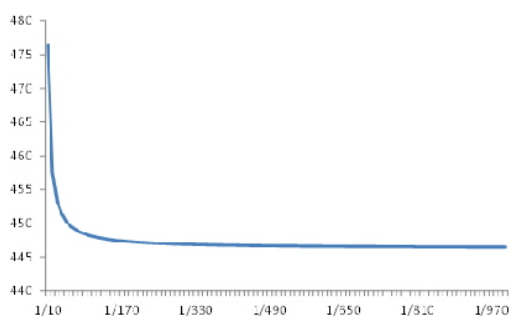
A. Uniform distribution



B. Normal distribution



C. Beta distribution



Note: $(\theta_\alpha, \theta_\beta) \in G_\varepsilon = (0+\varepsilon, 0+2\varepsilon, ..., 0+1) \times \left(0+\sigma^2_{\max}\varepsilon, 0+2\sigma^2_{\max}\varepsilon, ..., \sigma^2_{\max}\right)$, with $\sigma^2_{\max} = 0.16$ and $\varepsilon = 1/S$, for $S = 10, 20, 30, 40, ...1000$.

B Proofs

Proof of Proposition 1. Consider

$$\phi^*_{tsrp,N} = E\left\{\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[\Delta y_{it}\Delta y_{i,t-s} - E\left(\Delta y_{it}\Delta y_{i,t-s}\right)\right]\cdot\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[\Delta y_{i,t-r}\Delta y_{i,t-p} - E\left(\Delta y_{i,t-r}\Delta y_{i,t-p}\right)\right]\right\}, \tag{B.1}$$

for $t = 2, 3, ..., T$, $s = 0, 1, 2, ..., t-2$, $r = 0, 1, ..., t-2$, and $p = r, r+1, ..., t-2$, where Δy_{it} can be conveniently written as

$$\Delta y_{it} = \sum_{\ell=0}^{t+M}\varphi_{i\ell}u_{i,t-\ell} + \varphi_{i,t+M+1}\eta_i, \tag{B.2}$$

in which $\varphi_{i\ell}$ is defined in (39). Note the relationship between $\phi_{\ell h}$ and $\phi^*_{tsrh} = \lim_{N\to\infty}\phi^*_{tsrp,N}$:

$$\phi_{\ell h} = \phi^*_{t(\ell),s(\ell),r(\ell,h),p(\ell,h)}, \text{ for } \ell, h = 1, 2, ..., T\left(T-1\right)/2,$$

where $t\left(\ell\right)$ and $s\left(\ell\right)$ is the following mapping from $\ell = 1, 2, ..., T\left(T - 1/2\right)$ into $(t, s) \in \mathbb{N}\times\mathbb{N}_0$:

$$t\left(\ell\right) = t \in \mathbb{N} : \frac{\left(t-2\right)\left(t-1\right)}{2} < \ell \leq \frac{\left(t-1\right)t}{2}, \tag{B.3}$$

$$s\left(\ell\right) = \ell - \frac{\left[t\left(\ell\right)-2\right]\left[t\left(\ell\right)-1\right]}{2} - 1; \tag{B.4}$$

$$r\left(\ell, h\right) = t\left(\ell\right) - t\left(h\right), \text{ and } p\left(\ell, h\right) = s\left(h\right) + t\left(\ell\right) - t\left(h\right) \tag{B.5}$$

Using (B.2), we obtain

$$\begin{aligned} E\left(\Delta y_{it}\Delta y_{i,t-s}\right) \quad = \quad & E\left\{\sum_{\ell_1=0}^{t+M}\sum_{\ell_2=s}^{t+M}\varphi_{i\ell_1}\varphi_{i,\ell_2-s}u_{i,t-\ell_1}u_{i,t-\ell_2} + \varphi_{i,t-s+M+1}\eta_i\sum_{\ell_1=0}^{t+M}\varphi_{i\ell_1}u_{i,t-\ell_1}\right. \\ & \left. +\varphi_{i,t+M+1}\eta_i\sum_{\ell_2=s}^{t+M}\varphi_{i,\ell_2-s}u_{i,t-\ell_2} + \varphi_{i,t-s+M+1}\varphi_{i,t+M+1}\eta_i^2\right\} \end{aligned}$$

Using $E\left(u_{it}\right) = 0$, the independence of u_{it} across t, the independence of u_{it} and ϕ_i (for all i and t), the independence of u_{it} and η_i (for all i and t) and the independence of ϕ_i and η_i, we obtain

$$\begin{aligned} E\left(\Delta y_{it}\Delta y_{i,t-s}\right) \quad &= \quad \sum_{\ell=s}^{t+M}E\left(\varphi_{i\ell}\varphi_{i,\ell-s}\right)E\left(u^2_{i,t-\ell}\right) + E\left(\varphi_{i,t-s+M+1}\varphi_{i,t+M+1}\right)E\left(\eta_i^2\right) \\ &= \quad \sum_{\ell=s}^{t+M}E\left(\varphi_{i\ell}\varphi_{i,\ell-s}\right)\sigma_i^2 + E\left(\varphi_{i,t-s+M+1}\varphi_{i,t+M+1}\right)\delta, \end{aligned} \tag{B.6}$$

where $E\left(u_{i,t-\ell}^2\right) = \sigma_i^2$ and $E\left(\eta_i^2\right) = \delta$. Similarly, we have (recall that $p \geq r$)

$$E\left(\Delta y_{i,t-r}\Delta y_{i,t-p}\right) = \sum_{\ell=p}^{t+M} E\left(\varphi_{i,\ell-r}\varphi_{i,\ell-p}\right)\sigma_i^2 + E\left(\varphi_{i,t-p+M+1}\varphi_{i,t-r+M+1}\right)\delta. \tag{B.7}$$

Using the independence of u_{it} and η_i across i we have $E\left(\Delta y_{it}\Delta y_{i,t-s}\Delta y_{j,t-r}\Delta y_{j,t-p}\right) = E\left(\Delta y_{it}\Delta y_{i,t-s}\right)\cdot E\left(\Delta y_{j,t-r}\Delta y_{j,t-p}\right)$ for any $i \neq j$, and using also (B.2) and (B.6)-(B.7) we can write (B.1) as

$$\phi_{tsrp,N}^* = \frac{1}{N}\sum_{i=1}^{N} E\left(\Delta y_{it}\Delta y_{i,t-s}\Delta y_{i,t-r}\Delta y_{i,t-p}\right) - \frac{1}{N}\sum_{i=1}^{N}\left[E\left(\Delta y_{it}\Delta y_{i,t-s}\right)\cdot E\left(\Delta y_{i,t-r}\Delta y_{i,t-p}\right)\right] \tag{B.8}$$

Using (B.2), the first term on the right side of (B.8) can be written as

$$\frac{1}{N}\sum_{i=1}^{N} E\left\{\left(\sum_{\ell_1=0}^{t+M}\varphi_{i\ell_1}u_{i,t-\ell_1} + \varphi_{i,t+M+1}\eta_i\right)\cdot\left(\sum_{\ell_2=s}^{t+M}\varphi_{i,\ell_2-s}u_{i,t-\ell_2} + \varphi_{i,t-s+M+1}\eta_i\right)\cdot\right.$$
$$\left.\cdot\left(\sum_{\ell_3=p}^{t+M}\varphi_{i,\ell_3-p}u_{i,t-\ell_3} + \varphi_{i,t-p+M+1}\eta_i\right)\left(\sum_{\ell_4=r}^{t+M}\varphi_{i,\ell_4-r}u_{i,t-\ell_4} + \varphi_{i,t-r+M+1}\eta_i\right)\right\}.$$

Using again the independence of $E(u_{it}) = 0$, the independence of u_{it} across t, the independence of u_{it} and ϕ_i (for all i and t), the independence of u_{it} and η_i (for all i and t) and the independence of ϕ_i and η_i, the expression above can be written as the sum of four terms $\varrho_{1N} + \varrho_{2N} + \varrho_{3N} + \varrho_{4N}$, where

$$\varrho_{1N} = \frac{1}{N}\sum_{i=1}^{N}\sum_{\ell_1=0}^{t+M}\sum_{\ell_2=s}^{t+M}\sum_{\ell_3=p}^{t+M}\sum_{\ell_4=r}^{t+M} E\left(\varphi_{i\ell_1}\varphi_{i,\ell_2-s}\varphi_{i,\ell_3-p}\varphi_{i,\ell_4-r}\right)E\left(u_{i,t-\ell_1}u_{i,t-\ell_2}u_{i,t-\ell_3}u_{i,t-\ell_4}\right), \tag{B.9}$$

$$\begin{aligned}
\varrho_{2N} = \frac{1}{N}\sum_{i=1}^{N}\Bigg[& E\left(\varphi_{i,t-r+M+1}\right)E\left(\eta_i\right)\left(\sum_{\ell_1=0}^{t+M}\sum_{\ell_2=s}^{t+M}\sum_{\ell_3=p}^{t+M} E\left(\varphi_{i\ell_1}\varphi_{i,\ell_2-s}\varphi_{i,\ell_3-p}\right)E\left(u_{i,t-\ell_1}u_{i,t-\ell_2}u_{i,t-\ell_3}\right)\right) \\
& + E\left(\varphi_{i,t-p+M+1}\right)E\left(\eta_i\right)\left(\sum_{\ell_1=0}^{t+M}\sum_{\ell_2=s}^{t+M}\sum_{\ell_4=r}^{t+M} E\left(\varphi_{i\ell_1}\varphi_{i\ell_2-s}\varphi_{i,\ell_4-r}\right)E\left(u_{i,t-\ell_1}u_{i,t-\ell_2}u_{i,t-\ell_4}\right)\right) \\
& + E\left(\varphi_{i,t-s+M+1}\right)E\left(\eta_i\right)\left(\sum_{\ell_1=0}^{t+M}\sum_{\ell_3=p}^{t+M}\sum_{\ell_4=r}^{t+M} E\left(\varphi_{i\ell_1}\varphi_{i,\ell_3-p}\varphi_{i,\ell_4-r}\right)E\left(u_{i,t-\ell_1}u_{i,t-\ell_3}u_{i,t-\ell_4}\right)\right) \\
& + E\left(\varphi_{i,t+M+1}\right)E\left(\eta_i\right)\left(\sum_{\ell_2=s}^{t+M}\sum_{\ell_3=p}^{t+M}\sum_{\ell_4=r}^{t+M} E\left(\varphi_{i,\ell_2-s}\varphi_{i,\ell_3-p}\varphi_{i,\ell_4-r}\right)E\left(u_{i,t-\ell_2}u_{i,t-\ell_3}u_{i,t-\ell_4}\right)\right)\Bigg]
\end{aligned} \tag{B.10}$$

$$\begin{aligned}\varrho_{3N} &= \frac{1}{N}\sum_{i=1}^{N}\Bigg[E\left(\varphi_{i,t-p+M+1}\varphi_{i,t-r+M+1}\right)E\left(\eta_i^2\right)\sum_{\ell_1=0}^{t+M}\sum_{\ell_2=s}^{t+M}E\left(\varphi_{i\ell_1}\varphi_{i,\ell_2-s}\right)E\left(u_{i,t-\ell_1}u_{i,t-\ell_2}\right)\\
&+E\left(\varphi_{i,t-s+M+1}\varphi_{i,t-r+M+1}\right)E\left(\eta_i^2\right)\sum_{\ell_1=0}^{t+M}\sum_{\ell_3=p}^{t+M}E\left(\varphi_{i\ell_1}\varphi_{i,\ell_3-p}\right)E\left(u_{i,t-\ell_1}u_{i,t-\ell_3}\right)\\
&+E\left(\varphi_{i,t-s+M+1}\varphi_{i,t-p+M+1}\right)E\left(\eta_i^2\right)\sum_{\ell_1=0}^{t+M}\sum_{\ell_4=r}^{t+M}E\left(\varphi_{i\ell_1}\varphi_{i,\ell_4-r}\right)E\left(u_{i,t-\ell_1}u_{i,t-\ell_4}\right)\\
&+E\left(\varphi_{i,t+M+1}\varphi_{i,t-r+M+1}\right)E\left(\eta_i^2\right)\sum_{\ell_2=s}^{t+M}\sum_{\ell_3=p}^{t+M}E\left(\varphi_{i,\ell_2-s}\varphi_{i,\ell_3-p}\right)E\left(u_{i,t-\ell_2}u_{i,t-\ell_3}\right)\\
&+E\left(\varphi_{i,t+M+1}\varphi_{i,t-p+M+1}\right)E\left(\eta_i^2\right)\sum_{\ell_2=s}^{t+M}\sum_{\ell_4=r}^{t+M}E\left(\varphi_{i,\ell_2-s}\varphi_{i,\ell_4-r}\right)E\left(u_{i,t-\ell_2}u_{i,t-\ell_4}\right)\\
&+E\left(\varphi_{i,t+M+1}\varphi_{i,t-s+M+1}\right)E\left(\eta_i^2\right)\sum_{\ell_3=p}^{t+M}\sum_{\ell_4=r}^{t+M}E\left(\varphi_{i,\ell_3-p}\varphi_{i,\ell_4-r}\right)E\left(u_{i,t-\ell_3}u_{i,t-\ell_4}\right)\Bigg], \quad \text{(B.11)}\end{aligned}$$

and

$$\varrho_{4N} = \frac{1}{N}\sum_{i=1}^{N}E\left(\varphi_{i,t+M+1}\varphi_{i,t-s+M+1}\varphi_{i,t-p+M+1}\varphi_{i,t-r+M+1}\right)E\left(\eta_i^4\right). \quad \text{(B.12)}$$

But under Assumption 8, we have $E\left(\eta_i\right) = b_{\delta 1}$, $E\left(\eta_i^2\right) = \delta$, $E\left(\eta_i^4\right) = b_{\delta 4}$. Let $\gamma_{\ell_1} = E\left(\varphi_{i\ell_1}\right)$, $\gamma_{\ell_1\ell_2} = E\left(\varphi_{i\ell_1}\varphi_{i\ell_2}\right)$, $\gamma_{\ell_1\ell_2\ell_3} = E\left(\varphi_{i\ell_1}\varphi_{i\ell_2}\varphi_{i\ell_3}\right)$ and $\gamma_{\ell_1\ell_2\ell_3\ell_4} = E\left(\varphi_{i\ell_1}\varphi_{i\ell_2}\varphi_{i\ell_3}\varphi_{i\ell_4}\right)$ (we omit θ in brackets in the definition of these functions throughout this proof to economize on notations). Let also $E\left(u_{it}^2\right) = \sigma_i^2$, $E\left(u_{it}^3\right) = \upsilon_{3i}$ and $E\left(u_{it}^4\right) = \upsilon_{4i}$. We will use these results and notations to derive limiting expressions for ϱ_{1N}, ϱ_{2N}, ϱ_{3N} and ϱ_{4N}.

Consider first the term ϱ_{1N} and in particular the quintuple summation in (B.9). Notice that $E\left(u_{i,t-\ell_1}u_{i,t-\ell_2}u_{i,t-\ell_3}u_{i,t-\ell_4}\right)$ is nonzero only in the following four instances:

$$\begin{aligned}(i) &: \quad \ell_1 = \ell_2 = \ell_3 = \ell_4,\\
(ii) &: \quad \ell_1 = \ell_2 \wedge \ell_3 = \ell_4 \wedge \ell_1 \neq \ell_3,\\
(iii) &: \quad \ell_1 = \ell_3 \wedge \ell_2 = \ell_4 \wedge \ell_1 \neq \ell_2, \text{ and}\\
(iv) &: \quad \ell_1 = \ell_4 \wedge \ell_2 = \ell_3 \wedge \ell_1 \neq \ell_2.\end{aligned}$$

In the first case, we obtain

$$(i): \frac{1}{N}\sum_{i=1}^{N}\sum_{\ell=p}^{t+M}E\left(\varphi_{\ell}\varphi_{\ell-s}\varphi_{\ell-r}\varphi_{\ell-p}\right)E\left(u_{i,t-\ell}^4\right). \quad \text{(B.13)}$$

In the second case we have,

$$(ii) : \frac{1}{N}\sum_{i=1}^{N}\sum_{\ell_1=s}^{t+M}\sum_{\ell_3=p;\ \ell_3\neq\ell_1}^{t+M} E\left(\varphi_{\ell_1}\varphi_{\ell_1-s}\varphi_{\ell_3-r}\varphi_{\ell_3-p}\right)E\left(u_{i,t-\ell_1}^2\right)E\left(u_{i,t-\ell_3}^2\right). \tag{B.14}$$

Similarly, in the last two cases we obtain,

$$(iii) \quad : \quad \frac{1}{N}\sum_{i=1}^{N}\sum_{\ell_1=r}^{t+M}\sum_{\ell_2=p;\ \ell_2\neq\ell_1}^{t+M} E\left(\varphi_{\ell_1}\varphi_{\ell_2-s}\varphi_{\ell_1-r}\varphi_{\ell_2-p}\right)E\left(u_{i,t-\ell_1}^2\right)E\left(u_{i,t-\ell_2}^2\right), \tag{B.15}$$

$$(iv) \quad : \quad \frac{1}{N}\sum_{i=1}^{N}\sum_{\ell_1=p}^{t+M}\sum_{\ell_2=r;\ \ell_2\neq\ell_1}^{t+M} E\left(\varphi_{\ell_1}\varphi_{\ell_2-s}\varphi_{\ell_2-r}\varphi_{\ell_1-p}\right)E\left(u_{i,t-\ell_1}^2\right)E\left(u_{i,t-\ell_2}^2\right). \tag{B.16}$$

Using (B.13)-(B.16) in (B.9), and the notations above we obtain that

$$\begin{aligned}
\varrho_{1N} &= \bar{\upsilon}_4\sum_{\ell=p}^{t+M}\gamma_{\ell,\ell-s,\ell-r,\ell-p} + \bar{\varkappa}\sum_{\ell_1=s}^{t+M}\sum_{\ell_3=p;\ \ell_3\neq\ell_1}^{t+M}\gamma_{\ell_1,\ell_1-s,\ell_3-r,\ell_3-p} \\
&+\bar{\varkappa}\sum_{\ell_1=r}^{t+M}\sum_{\ell_2=p;\ \ell_2\neq\ell_1}^{t+M}\gamma_{\ell_1,\ell_2-s,\ell_1-r,\ell_2-p} + \bar{\varkappa}\sum_{\ell_1=p}^{t+M}\sum_{\ell_2=r;\ \ell_2\neq\ell_1}^{t+M}\gamma_{\ell_1,\ell_2-s,\ell_2-r,\ell_1-p},
\end{aligned}$$

where $\bar{\upsilon}_4 = N^{-1}\sum_{i=1}^{N}\upsilon_{4i}$ and $\bar{\varkappa} = N^{-1}\sum_{i=1}^{N}\left(\sigma_i^2\right)^2$. Similarly, we obtain the following expression for ϱ_{2N}, ϱ_{3N} and ϱ_{4N}:

$$\begin{aligned}
\varrho_{2N} &= b_{\delta 1}\bar{\upsilon}_3\left(\gamma_{t-r+M+1}\sum_{\ell=\max\{s,p\}}^{t+M}\gamma_{\ell,\ell-s,\ell-p} + \gamma_{t-p+M+1}\sum_{\ell=\max\{s,r\}}^{t+M}\gamma_{\ell,\ell-s,\ell-r}\right. \\
&\left. +\gamma_{t-s+M+1}\sum_{\ell=p}^{t+M}\gamma_{\ell,\ell-p,\ell-r} + \gamma_{t+M+1}\sum_{\ell=\max\{s,p\}}^{t+M}\gamma_{\ell-s,\ell-p,\ell-r}\right),
\end{aligned}$$

$$\begin{aligned}
\varrho_{3N} &= \delta\overline{\sigma^2}\left(\gamma_{t-p+M+1,t-r+M+1}\sum_{\ell=s}^{t+M}\gamma_{\ell,\ell-s} + \gamma_{t-s+M+1,t-r+M+1}\sum_{\ell=p}^{t+M}\gamma_{\ell,\ell-p}\right. \\
&+\gamma_{t-s+M+1,t-p+M+1}\sum_{\ell=r}^{t+M}\gamma_{\ell,\ell-r} + \gamma_{t+M+1,t-r+M+1}\sum_{\ell=\max\{s,p\}}^{t+M}\gamma_{\ell-s,\ell-p} \\
&\left.+\gamma_{t+M+1,t-p+M+1}\sum_{\ell=\max\{s,r\}}^{t+M}\gamma_{\ell-s,\ell-r} + \gamma_{t+M+1,t-s+M+1}\sum_{\ell=p}^{t+M}\gamma_{\ell-p,\ell-r}\right),
\end{aligned} \tag{B.17}$$

and

$$\varrho_{4N} = \gamma_{t+M+1,t-s+M+1,t-p+M+1,t-r+M+1}\cdot b_{\delta 4},$$

But $\lim_{N\to\infty}\bar{\varkappa}=\varkappa$, and $\lim_{N\to\infty}\overline{\sigma^2}=\sigma^2$, $\lim_{N\to\infty}\bar{\upsilon}_3=\upsilon_3$, and $\lim_{N\to\infty}\bar{\upsilon}_4=\upsilon_4$. Hence,

$$\begin{aligned}\varrho_1 &= p\lim_{N\to\infty}\varrho_{1N}=\upsilon_4\sum_{\ell=p}^{t+M}\gamma_{\ell,\ell-s,\ell-r,\ell-p}+\varkappa\sum_{\ell_1=s}^{t+M}\sum_{\ell_3=p;\ \ell_3\neq\ell_1}^{t+M}\gamma_{\ell_1,\ell_1-s,\ell_3-r,\ell_3-p}\\ &\quad+\varkappa\sum_{\ell_1=r}^{t+M+1}\sum_{\ell_2=p;\ \ell_2\neq\ell_1}^{t+M+1}\gamma_{\ell_1,\ell_2-s,\ell_1-r,\ell_2-p}+\varkappa\sum_{\ell_1=p}^{t+M+1}\sum_{\ell_2=r;\ \ell_2\neq\ell_1}^{t+M+1}\gamma_{\ell_1,\ell_2-s,\ell_2-r,\ell_1-p},\end{aligned}$$

and similar expressions can be obtained for $\varrho_2=p\lim_{N\to\infty}\varrho_{2N}$, $\varrho_3=p\lim_{N\to\infty}\varrho_{3N}$ and $\varrho_4=p\lim_{N\to\infty}\varrho_{4N}$. Now consider the second term on the right side of (B.8). Using similar arguments as in the derivations above, we obtain

$$\begin{aligned}p\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left[E\left(\Delta y_{it}\Delta y_{i,t-s}\right)\cdot E\left(\Delta y_{i,t-r}\Delta y_{i,t-p}\right)\right] &= \varkappa\sum_{\ell=s}^{t+M}\sum_{\ell'=p}^{t+M}\gamma_{\ell,\ell-s}\gamma_{\ell'-r,\ell'-p}\\ &\quad+\gamma_{t+M+1,t-s+M+1}\gamma_{t-p+M+1,t-r+M+1}\delta^2.\end{aligned}$$

It now follows that $\phi^*_{tsrh}=\lim_{N\to\infty}\phi^*_{tsrp,N}$ is given by

$$\phi^*_{tsrh}=\varrho_1+\varrho_2+\varrho_3+\varrho_4-\varkappa\sum_{\ell=s}^{t+M}\sum_{\ell'=p}^{t+M}\gamma_{\ell,\ell-s}\gamma_{\ell'-r,\ell'-p}-\gamma_{t+M+1,t-s+M+1}\gamma_{t-p+M+1,t-r+M+1}\delta^2,$$

where

$$\begin{aligned}\varrho_1 &= \upsilon_4\sum_{\ell=p}^{t+M}\gamma_{\ell,\ell-s,\ell-r,\ell-p}+\varkappa\sum_{\ell_1=s}^{t+M}\sum_{\ell_3=p;\ \ell_3\neq\ell_1}^{t+M}\gamma_{\ell_1,\ell_1-s,\ell_3-r,\ell_3-p}\\ &\quad+\varkappa\sum_{\ell_1=r}^{t+M+1}\sum_{\ell_2=p;\ \ell_2\neq\ell_1}^{t+M+1}\gamma_{\ell_1,\ell_2-s,\ell_1-r,\ell_2-p}+\varkappa\sum_{\ell_1=p}^{t+M+1}\sum_{\ell_2=r;\ \ell_2\neq\ell_1}^{t+M+1}\gamma_{\ell_1,\ell_2-s,\ell_2-r,\ell_1-p},\end{aligned}$$

$$\begin{aligned}\varrho_2 &= b_{\delta1}\upsilon_3\left(\gamma_{t-r+M+1}\sum_{\ell=\max\{s,p\}}^{t+M}\gamma_{\ell,\ell-s,\ell-p}+\gamma_{t-p+M+1}\sum_{\ell=\max\{s,r\}}^{t+M}\gamma_{\ell,\ell-s,\ell-r}\right.\\ &\quad\left.+\gamma_{t-s+M+1}\sum_{\ell=p}^{t+M}\gamma_{\ell,\ell-p,\ell-r}+\gamma_{t+M+1}\sum_{\ell=\max\{s,p\}}^{t+M}\gamma_{\ell-s,\ell-p,\ell-r}\right),\end{aligned}$$

$$\begin{aligned}
\varrho_3 &= \delta\sigma^2 \Bigg(\gamma_{t-p+M+1,t-r+M+1} \sum_{\ell=s}^{t+M} \gamma_{\ell,\ell-s} + \gamma_{t-s+M+1,t-r+M+1} \sum_{\ell=p}^{t+M} \gamma_{\ell,\ell-p} \\
&\quad + \gamma_{t-s+M+1,t-p+M+1} \sum_{\ell=r}^{t+M} \gamma_{\ell,\ell-r} + \gamma_{t+M+1,t-r+M+1} \sum_{\ell=\max\{s,p\}}^{t+M} \gamma_{\ell-s,\ell-p} \\
&\quad + \gamma_{t+M+1,t-p+M+1} \sum_{\ell=\max\{s,r\}}^{t+M} \gamma_{\ell-s,\ell-r} + \gamma_{t+M+1,t-s+M+1} \sum_{\ell=p}^{t+M} \gamma_{\ell-p,\ell-r} \Bigg),
\end{aligned}$$

and

$$\varrho_4 = \gamma_{t+M+1,t-s+M+1,t-p+M+1,t-r+M+1} \cdot b_{\delta 4}$$

This completes the proof. ■

References

Ahn, S. C. and P. Schmidt (1995). Efficient estimation of models for dynamic panel data. *Journal of Econometrics 68*, 5–27.

Alvarez, J. and M. Arellano (2003). The time series and cross-section asymptotics of dynamic panel data estimators. *Econometrica 71*, 1121–1159.

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

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

Arellano, M. (2003). *Panel data econometrics.* Oxford, UK: Oxford University Press.

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

Arellano, M. and O. Bover (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics 68*, 29–51.

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

Binder, M., C. Hsiao, and M. H. Pesaran (2005). Estimation and inference in short panel vector autoregressions with unit roots and cointegration. *Econometric Theory 21*, 795–837.

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

Bond, S. (2002). Dynamic panel data models: A guide to microdata methods and practice. CeMMAP Working Paper No. CWP09/02.

Chudik, A., M. H. Pesaran, and E. Tosetti (2011). Weak and strong cross section dependence and estimation of large panels. *Econometrics Journal 14*, C45–C90.

Han, C., P. C. B. Phillips, and D. Sul (2010). X-Differencing and dynamic panel model estimation. Cowles Foundation Discussion Paper No. 147.

Hayakawa, K. (2012). GMM estimation of short dynamic panel data models with interactive fixed effects. Mimeo, Hiroshima University, 17 February 2012.

Hayakawa, K. and M. H. Pesaran (2015). Robust standard errors in transformed likelihood estimation of dynamic panel data models. *Journal of Econometrics 188*(1), 111–134.

Hayakawa, K., M. H. Pesaran, and L. V. Smith (2014). Transformed maximum likelihood estimation of short dynamic panel data models with interactive effects. University of Cambridge, Cambridge Working Papers in Economics, No. 1412.

Holtz-Eakin, D., W. Newey, and H. S. Rosen (1988). Estimating vector autoregressions with panel data. *Econometrica 56*, 1371–1395.

Hsiao, C. (2003). *Analysis of Panel Data*. Econometric Society Monographs, Cambridge University Press.

Hsiao, C., M. H. Pesaran, and A. K. Tahmiscioglu (2002). Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. *Journal of Econometrics 109*(1), 107–150.

Mavroeidis, S., Y. Sasaki, and I. Welch (2015). Estimation of heterogeneous autoregressive parameters with short panel data. *Journal of Econometrics 188*(1), 219–235.

Pesaran, M. H. (2015). *Time Series and Panel Data Econometrics*. Oxford: Oxford University Press. ISBN-978-0-19-873691-2.

Pesaran, M. H. and R. Smith (1995). Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics 68*, 79–113.

Su, L. and Z. Yang (2007). QML estimation of dynamic panel data models with spatial errors. Mimeo, February 2007.