Spatial Dynamic Panel Data Models with Interactive Fixed Effects: M-Estimation and Inference with Small T

Liyao Li[a,b] and Zhenlin Yang[a*]

[a] *School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903*

[b] *Academy of Statistics and Interdisciplinary Sciences, East China Normal University, Shanghai 200062, China*

September 29, 2021

Abstract

We propose an M-estimation method for estimating the spatial dynamic panel data models with interactive fixed effects based on short panels. Unbiased estimating function (EF) is obtained by adjusting the concentrated conditional quasi score, given initial values and with factor loadings being concentrated out, to account for the effects of conditioning and concentration. For inference, the EF is decomposed into a sum of n nearly uncorrelated terms. Average of outer products of these n terms together with a covariance adjustment gives a consistent estimator of the variance-covariance (VC) matrix of the EF and hence the VC matrix of the M-estimator. Consistency and asymptotic normality of the proposed estimator are established, and consistency of the VC matrix estimator is proved. Monte Carlo results show the proposed methods perform very well in finite sample. Compared with the existing methods, our methods are much simpler, more efficient in point estimation and more reliable in inference when T is small.

Key Words: Adjusted quasi scores; Dynamic panels; Interactive fixed effects; Initial-condition; Martingale difference; Spatial effects; Short panels.

JEL classifications: C10, C13, C21, C23, C15

1. Introduction

Spatial dynamic panel data (SDPD) model has triggered a fast growing literature due to its important features of being able to (i) take into account temporal dynamics (time lag and space-time lag), (ii) capture spatial interaction effects (spatial lag, space-time lag, spatial Durbin, and spatial error),[1] and (iii) control for unobserved spatiotemporal heterogeneity

*Corresponding author. E-mail: zlyang@smu.edu.sg

[1] These have a close connection to Manski's (1993) social interaction framework, where he labeled these effects as endogenous effects, contextual effects and correlated effects.

(individual-specific and time-specific). The bulk of the literature has focused on the SDPD models with additive individual and time effects, being treated as fixed effects (Yu et al. 2008; Lee and Yu 2010, 2014; Su and Yang 2015; Yang 2018, 2021; Li and Yang 2020; Baltagi et al. 2021), or random effects (Yang et al. 2006; Mutl 2006; Su and Yang 2015), or correlated random effects (Li and Yang 2021). See Lee and Yu (2013) for a survey on earlier works.

A major recent advancement in the literature of SDPD model is the incorporation of interactive fixed effects (IFE) in the model (Shi and Lee 2017; Kuersteiner and Prucha 2020; and Bai and Li 2021). Besides the existing attractive features, this extended model specification draws further on the strength of IFE in controlling for the multiple unobserved time-specific effects f_t (common factors) and the corresponding individual-specific responses γ_i (factor loadings). For the large literature on regular panel models with IFE, see, among the others, Ahn et al. (2001, 2013), Bai (2009), Bai and Ng (2013), Moon and Weidner (2015, 2017).[2]

Shi and Lee (2017) and Bai and Li (2021) consider a conditional quasi maximum likelihood (CQML) approach given the initial observations for the estimation of SDPD-IFE models. The former has richer structures in temporal dynamics and spatial interactions, and the latter allows the individual variance to vary. Both methods depend critically on the large n and large T setup, and for valid statistical inferences it is necessary to carry out bias corrections on the CQML estimators for their asymptotic biases. Kuersteiner and Prucha (2020) consider a GMM approach under a large n and small T setup, for the estimation of a larger model with a different spatial error specification. They emphasize on method's generality by allowing additional features in the model: weakly exogenous covariates, multiple time lags, multiple spatial lags, network formation, unknown heteroskedasticity, etc. However, their methods may face issues of inefficiency and computational complexity when a less general model specification holds, which may hinder their practical applications. Therefore, it is very much desirable to have a set of simple and efficient estimation and inference methods for a fairly general SDPD-IFE model, which are valid when T is small.

In this paper, we propose an M-estimation method for estimating an SDPD-IFE model with a similar model specification as in Shi and Lee (2017) but under large n and small T

[2] Panel data models with interactive effects also specify (i) γ_i as fixed but f_t random, (ii) γ_i as random but f_t fixed, and (iii) both as random (see Hsiao 2018 for details). Case (i) is also of interest in connection with spatial econometrics literature as it induces error cross-section dependence (CD) as does the spatial error term. Pesaran and Tosetti (2011) refer to the former as strong CD and the latter as weak CD. They are perhaps the first researchers who join the two strands in literature in dealing with error cross-section dependence.

setup. Unbiased estimating functions are obtained by **adjusting** the concentrated conditional quasi scores, given initial values and with factor loadings being concentrated out, to account for the effects of conditioning and concentration. The resulting estimating functions (or moment conditions) have a close connection to the *modified equations of maximum likelihood* (Neyman and Scott 1948, Sec. 5).[3] The nature of the proposed estimating functions suggests that the M-estimator would be more efficient than the GMM estimator of Kuersteiner and Prucha (2020) if this less general model specification holds. The proposed method extends that of Yang (2018) for an SDPD model with additive fixed effects (AFE).

For statistical inferences, the vector of estimating functions (EF) is decomposed into a sum of n nearly uncorrelated terms. Then, the average of outer products of these n terms together with a covariance adjustment gives a consistent estimator of the variance-covariance (VC) matrix of the EF and hence the VC matrix of the M-estimator. We establish consistency and asymptotic normality of the proposed M-estimator, and prove consistency of the VC matrix estimator. We perform Monte Carlo experiments extensively to investigate the finite sample performance of the proposed estimation and inference methods. The results indeed show that when T is not large the proposed point estimation method is more efficient and the proposed inference method is more reliable, when compared with the existing methods.

The nature of the proposed estimation and inference methods suggests that there is a great potential for the methods to be extended to allow for multiple lags in time and in space as in Kuersteiner and Prucha (2020), cross-sectional heteroskedasticity explicitly in the model as in Bai and Li (2021) for an SDPD-IFE model or implicitly in the model as in Li and Yang (2020) for an SDPD-AFE model, etc. See Section 6 for further discussions.

The rest of the paper is as follows. Section 2 introduces the model and estimation method. Section 3 studies the asymptotic properties of the proposed estimator. Section 4 introduces the method of estimating the VC matrix of the proposed estimator. Section 5 presents Monte Carlo results. Section 6 concludes the paper. Technical proofs are collected in Appendix.

Notation. $|\cdot|$ denotes the determinant and $\mathrm{tr}(\cdot)$ the trace of a square matrix; $\mathtt{bdiag}(\cdot)$ forms a block-diagonal matrix from given matrices and vectors, and $\mathtt{vec}(\cdot)$ vectorizes a matrix by stacking its columns; $\otimes$ denotes the Kronecker product; $\|\cdot\|$ denotes the Frobenius norm,

[3] In a search of a systematic method of addressing the *incidental parameters problem*, Neyman and Scott (1948) suggest to modified the likelihood equations (score functions) to remove the effect of estimating the incidental parameters on the estimation of the common parameters.

$\|\cdot\|_1$ the maximum column sum norm and $\|\cdot\|_\infty$ the maximum row sum norm; and $\gamma_{\min}(\cdot)$ and $\gamma_{\max}(\cdot)$ denote, respectively, the smallest and largest eigenvalues of a real symmetric matrix.

2. Model and Estimation

Consider the spatial dynamic panel data (SDPD) model with interactive effects:

$$\begin{aligned} y_t &= \rho y_{t-1} + \lambda_1 W_1 y_t + \lambda_2 W_2 y_{t-1} + x_t\beta + \Gamma f_t + u_t, \\ u_t &= \lambda_3 W_3 u_t + v_t, \quad t = 1, 2, \ldots, T, \end{aligned} \tag{2.1}$$

where $y_t = (y_{1t}, y_{2t}, \ldots, y_{nt})'$ and $v_t = (v_{1t}, v_{2t}, \ldots, v_{nt})'$ are $n \times 1$ vectors of response values and idiosyncratic errors at time t, and $\{v_{it}\}$ are independent and identically distributed (*iid*) across i and t with mean zero and variance σ_v^2; $\{x_t\}$ are $n \times k$ matrices of time-varying exogenous variables; ρy_{t-1} captures the time dynamic effects; $W_r, r = 1, 2, 3$, are the given $n \times n$ spatial weight matrices that are used to model spatial dependences; the spatial lag (`SL`) term $\lambda_1 W_1 y_t$ captures the contemporaneous spatial interactions among cross-sectional units, the space-time lag (`STL`) term $\lambda_2 W_2 y_{t-1}$ captures the dynamic spatial interactions, and the spatial error (`SE`) term $\lambda_3 W_3 u_t$ captures the pure cross-sectional error dependence; f_t is a $r \times 1$ vector of unobserved time-specific effects (common factors) at time t, and $\Gamma = (\gamma_1, \gamma_2, \ldots, \gamma_n)'$ is an $n \times r$ matrix of unobserved individual-specific effects (factor loadings), whose rows, γ_i', are individuals' heterogeneous (interactive) responses to the common shocks f_t.

We adopt the fixed effect approach, which allows for arbitrary correlations between the regressors x_{it} and the interactive effects (common factor component) $\gamma_i' f_t$, giving an SDPD model with interactive fixed effects (IFE). The flexibility of the IFE approach comes at a cost. Unlike models with additive fixed effects, there is no simple linear transformation to eliminate the IFE. To control for the IFE, one has to treat factors and factor loadings as parameters and estimate them together with model's common parameters. However, estimating the γ_i parameters leads to the *incidental parameters problem* (Neyman and Scott, 1948), and this problem is greatly amplified by the *initial values problem* if a likelihood-type approach is followed. As a result, the CQML estimators of certain common parameters (e.g., ρ, λ and σ^2) are inconsistent when T is fixed and asymptotically biased when T goes large with n.

To solve the issue of small-T estimation and inference for the SDPD-IFE model, we introduce an M-estimation strategy where a set of unbiased and consistent estimating functions

(EF) of the common parameters and the identifiable part of F is obtained by **adjusting** the concentrated conditional quasi score (CCQS) functions treating y_0 as exogenously given and concentrating out Γ. This estimation strategy is naturally also called the *adjusted quasi score* (AQS) method for a better connection to the *modified equations of maximum likelihood* of Neyman and Scott (1948, Sec. 5). As discussed in the introduction, Kuersteiner and Prucha (2020) consider the small-T estimation and inference for an SDPD-IFE model based on GMM approach where they stress on the generality of the method by allowing several additional features in the model. As such, their general GMM estimation may suffer from efficiency loss besides the computational complexity. We stress on the efficiency and simplicity of the method, which roots in a 'likelihood' function. Indeed, the Monte Carlo results given in Sec. 5 show that the proposed M-estimator can be much more efficient than the GMM estimator and is much easier to implement, which greatly benefits applied researchers.

Let $B_r(\lambda_r) = I_n - \lambda_r W_r, r = 1, 3$, and $B_2(\rho, \lambda_2) = \rho I_n + \lambda_2 W_2$. Denote $\psi = (\beta', \sigma_v^2, \rho, \lambda')'$, the set of common parameters, where $\lambda = (\lambda_1, \lambda_2, \lambda_3)'$, and let $\theta = (\beta', \rho, \lambda_1, \lambda_2)'$. The joint **quasi** Gaussian loglikelihood function treating y_0 as exogenously given, referred to as the conditional quasi loglikelihood (CQL) function in this paper, takes the form:

$$\begin{aligned}\ell_{nT}(\psi, \Gamma, F) &= -\tfrac{nT}{2}\log(2\pi\sigma_v^2) - \tfrac{T}{2}\log|\Omega(\lambda_3)| + T\log|B_1(\lambda_1)| \\ &\quad - \tfrac{1}{2\sigma_v^2}\textstyle\sum_{t=1}^T [z_t(\theta) - \Gamma f_t]'\Omega^{-1}(\lambda_3)[z_t(\theta) - \Gamma f_t] \qquad (2.2)\\ &= -\tfrac{nT}{2}\log(2\pi\sigma_v^2) + T\log|B_3(\lambda_3)| + T\log|B_1(\lambda_1)| \\ &\quad - \tfrac{1}{2\sigma_v^2}\mathrm{tr}[(\mathbb{Z}(\theta) - \Gamma F')'\Omega^{-1}(\lambda_3)(\mathbb{Z}(\theta) - \Gamma F')], \qquad (2.3)\end{aligned}$$

where $z_t(\theta) = B_1(\lambda_1)y_t - B_2(\rho, \lambda_2)y_{t-1} - x_t\beta$, $\mathbb{Z}(\theta) = [z_1(\theta), z_2(\theta), \dots, z_T(\theta)]$, and $\Omega(\lambda_3) = \sigma_v^{-2}\mathrm{E}(u_t u_t') = (B_3'(\lambda_3)B_3(\lambda_3))^{-1}$. Maximizing $\ell_{nT}(\psi, \Gamma, F)$ under a set of constraints on $\{\gamma_i\}$ and $\{f_t\}$ gives the conditional quasi maximum likelihood (CQML) estimator $\hat{\psi}_{\texttt{CQML}}$ of the common parameters ψ. Shi and Lee (2017) show that $\hat{\psi}_{\texttt{CQML}}$ is consistent only when n and T are both large, and in this case, the asymptotic distribution of $\sqrt{nT}(\hat{\psi}_{\texttt{CQML}} - \psi_0)$ has a non-zero mean – the asymptotic bias. For proper statistical inference, a bias correction (BC) has to be made on $\hat{\psi}_{\texttt{CQML}}$. Along the similar ideas, Bai and Li (2021) propose a BC-CQML estimation of a smaller model (without `STL` and `SE`) but allowing explicitly the cross-sectional heteroskedasticity. The results of both papers depend critically on the large n and large T setup for a reasonable finite sample performance of their BC-CQML estimators. Shi and Lee's

BC-CQML method depends critically on the perturbation theory that hinders the extension to allow for heteroskedasticity as commented by Bai and Li (2021).

Solving the first order condition, $\frac{\partial}{\partial \Gamma}\ell_{nT}(\psi, \Gamma, F) = 0$,[4] using (2.3), we obtain the constrained CQML estimator of Γ as function of θ and F:

$$\tilde{\Gamma}(\theta, F) = \mathbb{Z}(\theta)F(F'F)^{-1}. \tag{2.4}$$

Plugging $\tilde{\Gamma}(\theta, F)$ in $\ell_{nT}(\psi, \Gamma, F)$ gives the concentrated conditional quasi loglikelihood (CCQL) function of ψ and F, noting that $\mathbb{Z}(\theta) - \tilde{\Gamma}(\theta, F)F' = \mathbb{Z}(\theta) - \mathbb{Z}(\theta)F(F'F)^{-1}F' \equiv \mathbb{Z}(\theta)M_F$:

$$\begin{aligned} \ell^c_{nT}(\psi, F) = & -\tfrac{nT}{2}\log(2\pi\sigma_v^2) + T\log|B_3(\lambda_3)| + T\log|B_1(\lambda_1)| \\ & - \tfrac{1}{2\sigma_v^2}\mathrm{tr}[M_F\mathbb{Z}'(\theta)\Omega^{-1}(\lambda_3)\mathbb{Z}(\theta)]. \end{aligned} \tag{2.5}$$

Maximizing the CCQL $\ell^c_{nT}(\psi, F)$ gives the CQML estimators of ψ, F, and hence Γ. However, there are two major issues that render the estimation based on maximizing $\ell^c_{nT}(\psi, F)$ inconsistent when T is fixed, both inducing the *incidental parameters problem* of Neyman and Scott (1948).[5] The first is the initial values problem. When y_0 is generated in the same way as the other values of $y_t, t = 1, \ldots, T$, it depends on the past values of time-varying regressors, which are not observable, leading to incidental parameters. Formulating the likelihood function (2.3) conditional on y_0 ignores the information y_0 contains about the common parameters, causing the CQML estimators to be inconsistent. The second issue relates to estimating the factors loadings. Concentrating out Γ from (2.3) is the QML estimation of the $n \times r$ matrix of factor loadings, which results in a large number of degrees of freedom loss and renders the QMLE inconsistent when T is small. The root-cause of the inconsistency problem is that $\mathrm{E}[\frac{1}{nT}\frac{\partial}{\partial\psi}\ell^c_{nT}(\psi_0, F_0)] \neq 0$ and $\mathrm{plim}_{n\to\infty}\frac{1}{nT}\frac{\partial}{\partial\psi}\ell^c_{nT}(\psi_0, F_0) \neq 0$ as seen below, i.e., the estimating functions derived directly from the CCQL functions are biased and inconsistent, where (ψ_0, F_0) are the true values of the parameters (ψ, F).[6] To solve this problem, we adopt the fundamental idea of Yang (2018) by adjusting the concentrated conditional quasi score functions to account for the effects of conditioning on the initial values and estimating the factor loadings, so as to give a set of unbiased estimating functions or moment conditions.

[4]This is done using the matrix differential formulas of Magnus and Neudecker (2019, p.200): $\frac{\partial}{\partial X}\mathrm{tr}(AX) = A'$, and $\frac{\partial}{\partial X}\mathrm{tr}(XAX'B) = B'XA' + BXA$, where X is a matrix.

[5]In Neyman and Scott's terminology, parameters are 'incidental' if information about them stops accumulating after a finite number of observations have been taken.

[6]When T grows with n, as in Shi and Lee (2017) and Bai and Li (2021), the incidental parameters problem is alleviated since information accumulates in both space and time dimensions and consistent estimation can be reached. However, asymptotic bias remains and a bias-correction has to be carried out for proper inferences.

To facilitate the derivation of the unbiased and consistent estimating functions, it is convenient to use the long vector $Z(\theta) = [z_1'(\theta), z_2'(\theta), \ldots, z_T'(\theta)]' = \texttt{vec}(\mathbb{Z})$. Working directly with (2.2) and (2.4), or using the identity $\text{tr}[M_F\mathbb{Z}'(\theta)\Omega^{-1}(\lambda_3)\mathbb{Z}(\theta)] = Z'(\theta)[M_F \otimes \Omega^{-1}(\lambda_3)]Z(\theta)$,[7] the CCQL function can be written as

$$\begin{aligned} \ell_{nT}^c(\psi, F) = & -\tfrac{nT}{2}\log(2\pi\sigma_v^2) + T\log|B_3(\lambda_3)| + T\log|B_1(\lambda_1)| \\ & -\tfrac{1}{2\sigma_v^2}Z'(\theta)[M_F \otimes \Omega^{-1}(\lambda_3)]Z(\theta). \end{aligned} \tag{2.6}$$

Let $Y = (y_1', y_2', \ldots, y_T')'$ and $Y_{-1} = (y_0', y_1', \ldots, y_{T-1}')'$, the $(nT \times 1)$ vectors of response and lagged response values, and $\mathbf{X} = (x_1', x_2', \ldots, x_T')'$, the $nT \times k$ matrix of regressors values. Denote $\mathbf{W}_r = I_T \otimes W_r, \mathbf{B}_r = I_T \otimes B_r, r = 1, 2, 3$, $\boldsymbol{\Omega}(\lambda_3) = I_T \otimes \Omega(\lambda_3)$, and $\mathbf{M}_F = M_F \otimes I_n$. We have the ψ-component of the concentrated conditional quasi score (CCQS):

$$\frac{\partial}{\partial\psi}\ell_{nT}^c(\psi, F) = \begin{cases} \frac{1}{\sigma_v^2}\mathbf{X}'\mathbf{M}_F\boldsymbol{\Omega}^{-1}(\lambda_3)Z(\theta), \\ \frac{1}{2\sigma_v^4}Z(\theta)'\mathbf{M}_F\boldsymbol{\Omega}^{-1}(\lambda_3)Z(\theta) - \frac{nT}{2\sigma_v^2}, \\ \frac{1}{\sigma_v^2}Z'(\theta)\mathbf{M}_F\boldsymbol{\Omega}^{-1}(\lambda_3)Y_{-1}, \\ \frac{1}{\sigma_v^2}Z(\theta)'\mathbf{M}_F\boldsymbol{\Omega}^{-1}(\lambda_3)\mathbf{W}_1 Y - \text{tr}(\mathbf{B}_1^{-1}(\lambda_1)\mathbf{W}_1), \\ \frac{1}{\sigma_v^2}Z'(\theta)\mathbf{M}_F\boldsymbol{\Omega}^{-1}(\lambda_3)\mathbf{W}_2 Y_{-1}, \\ \frac{1}{\sigma_v^2}Z'(\theta)\mathbf{M}_F\mathbf{B}_3'(\lambda_3)\mathbf{W}_3 Z(\theta) - \text{tr}(\mathbf{B}_3^{-1}(\lambda_3)\mathbf{W}_3). \end{cases} \tag{2.7}$$

We derive the expectation of the ψ-component of the CCQS function at true parameter values ψ_0 and F_0, and show that the $(\sigma_v^2, \rho, \lambda)$ components of $\frac{1}{nT}\text{E}[\frac{\partial}{\partial\psi}\ell_{nT}^c(\psi_0, F_0)]$ are generally not zero, and more seriously the $(\sigma_v^2, \rho, \lambda)$ components of $\text{plim}_{n\to\infty}\frac{1}{nT}\frac{\partial}{\partial\psi}\ell_{nT}^c(\psi_0, F_0)$ are not zero. Thus, the CQML estimator of ψ cannot be consistent as a necessary condition for consistent estimation is violated. To proceed, the following basic assumptions are required.

Assumption A. *Process started at $t = -m$ ($m \geq 0$) and data collection started at $t = 0$: (i) y_0 is independent of $\{v_t, t \geq 1\}$, and (ii) time-varying regressors $\{x_t,\ t = 0, 1, \ldots, T\}$, factors F and factor loadings Γ are independent of the idiosyncratic errors $\{v_t,\ t = 0, 1, \ldots, T\}$.*

From now on, we view that Model (2.1) holds only at the true parameters, and the usual expectation and variance operators $\text{E}(\cdot)$ and $\text{Var}(\cdot)$ correspond to the true model. Denote a parametric quantity evaluated at the true parameters by dropping its arguments and then adding a subscript "0", e.g., $B_{10} = B_1(\lambda_{10})$, and $\Omega_0 = \Omega(\lambda_{30})$, except $z_t = z_t(\theta_0)$. Define

[7] This follows from, e.g., Magnus and Neudecker (2019, p.36): for conformable matrices A, B, C and D such that $ABCD$ is defined and square, $\text{tr}(ABCD) = \texttt{vec}(D')'(C' \otimes A)\texttt{vec}(B) = \texttt{vec}(D)'(A \otimes C')\texttt{vec}(B')$.

$\mathcal{B}_0 = \mathcal{B}(\rho_0, \lambda_{10}, \lambda_{20}) \equiv B_1^{-1}(\lambda_{10})B_2(\rho_0, \lambda_{20})$. Backward substitution on (2.1) gives

$$y_t = \mathcal{B}_0^t y_0 + \sum_{s=0}^{t} \mathcal{B}_0^s B_{10}^{-1} x_{t-s}\beta_0 + \sum_{s=0}^{t} \mathcal{B}_0^s B_{10}^{-1} z_{t-s}, \quad t = 1, \ldots, T. \tag{2.8}$$

This leads to the following simple but important representations for Y and Y_{-1}:

$$Y = \mathbf{Q}\mathbf{y}_0 + \boldsymbol{\eta} + \mathbf{D}Z \quad \text{and} \quad Y_{-1} = \mathbf{Q}_{-1}\mathbf{y}_0 + \boldsymbol{\eta}_{-1} + \mathbf{D}_{-1}Z, \tag{2.9}$$

where $\mathbf{y}_0 = 1_T \otimes y_0$, $Z = Z(\theta_0)$, $\boldsymbol{\eta} = \mathbf{D}\mathbf{X}\beta_0$, $\boldsymbol{\eta}_{-1} = \mathbf{D}_{-1}\mathbf{X}\beta_0$, $\mathbf{Q} = \texttt{bdiag}(\mathcal{B}_0, \mathcal{B}_0^2, \ldots, \mathcal{B}_0^T)$, $\mathbf{Q}_{-1} = \texttt{bdiag}(I_n, \mathcal{B}_0, \ldots, \mathcal{B}_0^{T-1})$,

$$\mathbf{D} = \begin{pmatrix} I_n & 0 & \cdots & 0 & 0 \\ \mathcal{B}_0 & I_n & \cdots & 0 & 0 \\ \mathcal{B}_0^2 & \mathcal{B}_0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathcal{B}_0^{T-1} & \mathcal{B}_0^{T-2} & \cdots & \mathcal{B}_0 & I_n \end{pmatrix} \mathbf{B}_{10}^{-1}, \text{ and } \mathbf{D}_{-1} = \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ I_n & 0 & \cdots & 0 & 0 \\ \mathcal{B}_0 & I_n & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathcal{B}_0^{T-2} & \mathcal{B}_0^{T-3} & \cdots & I_n & 0 \end{pmatrix} \mathbf{B}_{10}^{-1}.$$

Based on the representations given in (2.9), we obtain under Assumption A and the assumption that the errors $\{v_{it}\}$ in Model (2.1) are *iid* $(0, \sigma_{v0}^2)$ across i and t,

$$\mathrm{E}[\tfrac{\partial}{\partial\psi}\ell_{nT}^c(\psi_0, F_0)] = \begin{cases} 0, \\ \frac{n(T-r)}{2\sigma_{v0}^2} - \frac{nT}{2\sigma_{v0}^2}, \\ \mathrm{tr}(\mathbf{M}_{F0}\mathbf{D}_{-1}), \\ \mathrm{tr}(\mathbf{M}_{F0}\mathbf{W}_1\mathbf{D}) - \mathrm{tr}(\mathbf{B}_{10}^{-1}\mathbf{W}_1), \\ \mathrm{tr}(\mathbf{M}_{F0}\mathbf{W}_2\mathbf{D}_{-1}), \\ (T-r)\mathrm{tr}(B_{30}^{-1}W_3) - T\mathrm{tr}(B_{30}^{-1}W_3). \end{cases} \tag{2.10}$$

From (2.10), we see that $\mathrm{E}[\frac{\partial}{\partial\psi}\ell_{nT}^c(\psi_0, F_0)] \neq 0$ and $\mathrm{plim}_{n\to\infty} \frac{1}{nT}\frac{\partial}{\partial\psi}\ell_{nT}^c(\psi_0, F_0) \neq 0$. A necessary condition for consistency is violated, and the CQML estimators cannot be consistent.

Note that $\mathrm{E}[\frac{\partial}{\partial\psi}\ell_{nT}^c(\psi_0, F_0)]$ is a parametric vector free from the initial conditions, the process starting time and the loadings. Therefore, it can be used to adjust/center (2.7) to give a set of unbiased estimating functions for ψ, free from m, Γ and the conditions on y_0:

$$S_{nT,\psi}^*(\psi, F) = \begin{cases} \frac{1}{\sigma_v^2}\mathbf{X}'\mathbf{M}_F\boldsymbol{\Omega}^{-1}(\lambda_3)Z(\theta), \\ \frac{1}{2\sigma_v^4}Z(\theta)'\mathbf{M}_F\boldsymbol{\Omega}^{-1}(\lambda_3)Z(\theta) - \frac{n(T-r)}{2\sigma_v^2}, \\ \frac{1}{\sigma_v^2}Z'(\theta)\mathbf{M}_F\boldsymbol{\Omega}^{-1}(\lambda_3)Y_{-1} - \mathrm{tr}[\mathbf{M}_F\mathbf{D}_{-1}(\rho, \lambda_1, \lambda_2)], \\ \frac{1}{\sigma_v^2}Z(\theta)'\mathbf{M}_F\boldsymbol{\Omega}^{-1}(\lambda_3)\mathbf{W}_1 Y - \mathrm{tr}[\mathbf{M}_F\mathbf{W}_1\mathbf{D}(\rho, \lambda_1, \lambda_2)], \\ \frac{1}{\sigma_v^2}Z'(\theta)\mathbf{M}_F\boldsymbol{\Omega}^{-1}(\lambda_3)\mathbf{W}_2 Y_{-1} - \mathrm{tr}[\mathbf{M}_F\mathbf{W}_2\mathbf{D}_{-1}(\rho, \lambda_1, \lambda_2)], \\ \frac{1}{\sigma_v^2}Z'(\theta)\mathbf{M}_F\mathbf{B}_3'(\lambda_3)\mathbf{W}_3 Z(\theta) - (T-r)\mathrm{tr}[B_3^{-1}(\lambda_3)W_3]. \end{cases} \tag{2.11}$$

We have $\mathrm{E}[S^*_{nT,\psi}(\psi_0, F_0)] = 0$, and one can show that $\mathrm{plim}_{n\to\infty}\frac{1}{nT}S^*_{nT,\psi}(\psi_0, F_0) = 0$. Thus, given F, the common parameters ψ can be consistently estimated by solving $S^*_{nT,\psi}(\psi, F) = 0$.

It remains to derive the F-component of the AQS function. Notice that F enters the CCQL function (2.6) and the AQS function (2.11) in the form of $P_F = F(F'F)^{-1}F'$. As a result, both (2.6) and (2.11) are invariant to the transformation $F^\dagger = FC$ for any $r \times r$ invertible matrix C as $P_{F^\dagger} = P_F$. Thus, we are not able to identify F without restrictions. As an arbitrary $r \times r$ invertible matrix has r^2 free elements, exactly r^2 restrictions are needed.[8] Following Ahn and Schmidt (2013), we normalize F such that $F = (F^{*\prime}, I_r)'$, where F^* is a $(T-r)\times r$ matrix of unrestricted parameters.[9] Collect all the free parameters of F in a vector $\phi = \mathtt{vec}(F^*)$, with its sth elements denoted by ϕ_s, $s = 1, \ldots, k_\phi$, where $k_\phi = \dim(\phi) = (T-r)r$. Then, we derive the CCQS components corresponding to the free parameters F^*, and we show that adjustments are not needed for these components.

Let $\dot{F}_s = \frac{\partial}{\partial \phi_s}F$, we have $\dot{P}_{F,s} = \frac{\partial}{\partial \phi_s}P_F = M_F\dot{F}_s(F'F)^{-1}F' + F(F'F)^{-1}\dot{F}_s'M_F$, $s = 1, \ldots, k_\phi$. Then, the CCQS component corresponding to ϕ_s, $s = 1, \ldots, k_\phi$, is

$$\begin{aligned}\frac{\partial}{\partial \phi_s}\ell^c_{nT}(\psi,\phi) &= \frac{1}{2\sigma_v^2}Z'(\theta)[\dot{P}_{F,s}\otimes\Omega^{-1}(\lambda_3)]Z(\theta)\\ &= \frac{1}{\sigma_v^2}Z'(\theta)[M_F\dot{F}_s(F'F)^{-1}F'\otimes\Omega^{-1}(\lambda_3)]Z(\theta).\end{aligned} \tag{2.12}$$

Let $\mathbf{v} = (v_1', \ldots, v_T')'$, we can write $Z = \mathtt{vec}(\Gamma_0 F_0') + \mathbf{B}_{30}^{-1}\mathbf{v}$. Under Assumption A and the assumptions on the errors, we have, for $s = 1, \ldots, k_\phi$,

$$\begin{aligned}&\mathrm{E}[\tfrac{\partial}{\partial \phi_s}\ell^c_{nT}(\psi_0,\phi_0)]\\ &= \tfrac{1}{\sigma_{v0}^2}\mathrm{E}\{[\mathbf{v}+\mathbf{B}_{30}\mathtt{vec}(\Gamma_0F_0')]'[M_{F0}\dot{F}_{s0}(F_0'F_0)^{-1}F_0'\otimes I_n][\mathbf{v}+\mathbf{B}_{30}\mathtt{vec}(\Gamma_0F_0')]\}\\ &= \tfrac{1}{\sigma_{v0}^2}\mathrm{E}\{\mathbf{v}'[M_F\dot{F}_{s0}(F_0'F_0)^{-1}F_0'\otimes I_n]\mathbf{v}\} + \tfrac{1}{\sigma_{v0}^2}\mathtt{vec}(\Gamma_0F_0')'[M_F\dot{F}_{s0}(F_0'F_0)^{-1}F_0'\otimes\Omega_0^{-1}]\mathtt{vec}(\Gamma_0F_0')\\ &= n\,\mathrm{tr}[M_{F_0}\dot{F}_{s0}(F_0'F_0)^{-1}F_0'] + \tfrac{1}{\sigma_{v0}^2}\mathrm{tr}[M_{F_0}\dot{F}_{s0}\Gamma_0'B_{30}'B_{30}\Gamma_0F_0'] = 0.\end{aligned}$$

This shows that the ϕ-component of the CCQS function is unbiased. Further, one shows that $\mathrm{plim}_{n\to\infty}\frac{1}{nT}\frac{\partial}{\partial\phi_s}\ell^c_{nT}(\psi_0,\phi_0) = 0, s = 1, \ldots, k_\phi$. Therefore, we do not need to adjust these CCQS components. In another word, given ψ, maximizing the CCQL function in (2.6) gives

[8] This is equivalent to the so-called "rotation problem" in factor models, which says that it is impossible to identify Γ and F separately without restrictions as $\Gamma CC^{-1}F' = \Gamma F'$ for any $r\times r$ non-singular matrix C. See Bai (2009) and Bai and Ng (2013) for detailed discussions.

[9] This is obtained, if we denote $F = (F_1, F_2)'$ where F_2 is $r \times r$ and is invertible, through the rotation: $F^\dagger = FC = FF_2^{-1} = (F_2'^{-1}F_1', I_r)' = (F^{*\prime}, I_r)'$. Ahn and Schmidt (2013) use the same normalization in their study of a regular panel data model with IFE under short T. The choice of normalization is not important because we are interested in controlling for the IFE, not interpreting them. However, in our paper, this normalization leads to a simpler way of establishing the set of unbiased and consistent estimating functions. See Bai and Ng (2013) for a detailed discussion of alternative normalizations.

a consistent estimate of ϕ, and therefore gives a consistent estimate of (a rotation of) F.

Combining the ψ components (2.11) and the ϕ components (2.12), we obtain the following important joint estimating function or adjusted quasi score function for (ψ, ϕ):

$$S^*_{nT}(\psi,\phi) = \begin{cases} \frac{1}{\sigma_v^2}\mathbf{X}'\mathbf{M}_F\mathbf{\Omega}^{-1}(\lambda_3)Z(\theta), \\ \frac{1}{2\sigma_v^4}Z(\theta)'\mathbf{M}_F\mathbf{\Omega}^{-1}(\lambda_3)Z(\theta) - \frac{n(T-r)}{2\sigma_v^2}, \\ \frac{1}{\sigma_v^2}Z'(\theta)\mathbf{M}_F\mathbf{\Omega}^{-1}(\lambda_3)Y_{-1} - \mathrm{tr}[\mathbf{M}_F\mathbf{D}_{-1}(\rho,\lambda_1,\lambda_2)], \\ \frac{1}{\sigma_v^2}Z(\theta)'\mathbf{M}_F\mathbf{\Omega}^{-1}(\lambda_3)\mathbf{W}_1Y - \mathrm{tr}[\mathbf{M}_F\mathbf{W}_1\mathbf{D}(\rho,\lambda_1,\lambda_2)], \\ \frac{1}{\sigma_v^2}Z'(\theta)\mathbf{M}_F\mathbf{\Omega}^{-1}(\lambda_3)\mathbf{W}_2Y_{-1} - \mathrm{tr}[\mathbf{M}_F\mathbf{W}_2\mathbf{D}_{-1}(\rho,\lambda_1,\lambda_2)], \\ \frac{1}{\sigma_v^2}Z'(\theta)\mathbf{M}_F\mathbf{B}_3'(\lambda_3)\mathbf{W}_3Z(\theta) - (T-r)\mathrm{tr}[B_3^{-1}(\lambda_3)W_3] \\ \frac{1}{\sigma_v^2}Z'(\theta)[M_F\dot{F}_s(F'F)^{-1}F'\otimes\Omega^{-1}(\lambda_3)]Z(\theta), \quad s=1,\ldots,k_\phi. \end{cases} \tag{2.13}$$

Solving the estimating equations: $S^*_{nT}(\psi,\phi)=0$, leads to the **AQS or M-estimators** $\hat{\psi}_{\mathtt{M}}$ and $\hat{\phi}_{\mathtt{M}}$ of ψ and ϕ. As $\mathrm{E}[S^*_{nT}(\psi_0,\phi_0)]=0$, one can show that $\mathrm{plim}_{n\to\infty}\frac{1}{nT}S^*_{nT}(\psi_0,\phi_0)=0$, i.e., a key condition for the consistency of $\hat{\psi}_{\mathtt{M}}$ and $\hat{\phi}_{\mathtt{M}}$ is satisfied.

Remark 2.1. *The importance of the joint EF, $S^*_{nT}(\psi,\phi)$, also lies in the fact that it leads to a simpler way to establish the joint asymptotic distribution of $\hat{\psi}_{\mathtt{M}}$ and $\hat{\phi}_{\mathtt{M}}$, and a simpler and reliable way to obtain the VC matrix estimate as seen in the subsequent sections.*

Remark 2.2. *It is interesting to note that the $(\beta_0,\sigma_0^2,\phi_0)$-components of $S^*_{nT}(\psi_0,\phi_0)$ remain unbiased and consistent under cross-sectional heteroskedasticity.*[10] *Therefore, if we are able to adjust the (ρ_0,λ_0)-components of $S^*_{nT}(\psi_0,\phi_0)$ so that they possess the same property, we then obtain a set of AQS functions and hence M-estimators that are robust against unknown heteroskedasticity. See Section 6 for more discussions.*

A computational note. Given ψ, Model (2.1) reduces to a pure factor model. The constrained M-estimator of F or ϕ can be obtained by maximizing $\frac{1}{nT}\mathrm{tr}[P_F\mathbb{Z}'(\theta)\Omega^{-1}(\lambda_3)\mathbb{Z}(\theta)]$,[11] and the solution is the eigenvector matrix of $\frac{1}{nT}\mathbb{Z}'(\theta)\Omega^{-1}(\lambda_3)\mathbb{Z}(\theta)$ corresponding to the r largest eigenvalues.[12] Thus, the computation of the M-estimators can simply be done by a

[10]Suppose $\mathrm{Var}(v_{it})=\sigma_v^2h_{n,i}$, such that $h_{n,i}>0$ and $\frac{1}{n}\sum_{i=1}^n h_{n,i}=1$. Let $\mathcal{H}=\mathtt{diag}(h_{n,1},\ldots,h_{n,n})$. Then, $\mathrm{Var}(\mathbf{v})=\sigma_{v0}^2I_T\otimes\mathcal{H}$, and $\mathrm{E}\{\mathbf{v}'[M_F\dot{F}_{s0}(F_0'F_0)^{-1}F_0'\otimes I_n]\mathbf{v}\}=\sigma_{v0}^2\mathrm{tr}\{(I_T\otimes\mathcal{H})[M_F\dot{F}_{s0}(F_0'F_0)^{-1}F_0'\otimes I_n]\}=\sigma_{v0}^2\mathrm{tr}\{[M_F\dot{F}_{s0}(F_0'F_0)^{-1}F_0']\otimes\mathcal{H}\}=\sigma_{v0}^2\mathrm{tr}(\mathcal{H})\mathrm{tr}[M_F\dot{F}_{s0}(F_0'F_0)^{-1}F_0']=0$, for the ϕ-component. It is much easier to verify that the same holds for the (β,σ^2)-components.

[11]This is equivalent to the objective function of the least square estimation of a pure factor model, $B_3\mathbb{Z}=B_3\Gamma F'+\mathbb{V}$, after the factor loadings Γ being concentrated out, where $\mathbb{V}=(v_1,\ldots,v_T)$. See Connor and Korajzcyk (1986), Stock and Watson (2002), and Bai (2003, 2009).

[12]See Magnus and Neudecker (2019, Ch. 17) and Ahn et al. (2013) for more details.

recursive process: iterating between Steps 1. and 2. until results converge:

1. Given F, the constrained M-estimator of ψ is $\hat{\psi}^*_{nT}(F) = \arg\{S^*_{nT,\psi}(\psi, F) = 0\}$,
2. Given ψ, the estimator of F is the matrix of eigenvectors corresponding to the r largest eigenvalues of the $T \times T$ matrix $\frac{1}{nT}\mathbb{Z}'(\theta)\Omega^{-1}(\lambda_3)\mathbb{Z}(\theta)$.[13]

See, for more discussions, Kiefer (1980), Ahn, et al. (2001, 2013), and Bai (2009). The root-finding process in Step 1. can be further simplified by first solving the first two sets of equations for β and σ^2 to obtain analytical solutions, and then solving the equations corresponding to the concentrated AQS functions (see Sec. 3 for details).

3. Asymptotic Properties of the M-Estimator

Rigorous studies on the asymptotic properties of the proposed M-estimator require the following basic regularity conditions. Denote $\delta = (\rho, \lambda', \phi')'$, the set of parameters that appear in the AQS function nonlinearly (i.e., their AQS equations cannot be solved analytically).

Assumption B. *The innovations v_{it} are iid for all i and t with $E(v_{it}) = 0$, $Var(v_{it}) = \sigma^2_{v0}$, and $E|v_{it}|^{4+\epsilon_0} < \infty$ for some $\epsilon_0 > 0$.*

Assumption C. *(i) The parameter space $\mathbf{\Delta}$ of δ is compact, and the true parameter vector δ_0 lies in its interior; (ii) The number of factors r_0 is constant and less than T. The elements of Γ_0 and F_0 are uniformly bounded.*

Assumption D. *The elements of the time-varying regressors $\{x_t, t = 1, \ldots, T\}$ are uniformly bounded, and the limit $\lim_{n\to\infty} \frac{1}{nT}\mathbf{X}'\mathbf{M}_F\mathbf{X}$ exists and is nonsingular.*

Assumption E. *(i) For $r = 1, 2, 3$, the elements $w_{r,ij}$ of W_r are at most of order h_n^{-1}, uniformly in all i and j, and $w_{r,ii} = 0$ for all i; (ii) $h_n/n \to 0$ as $n \to \infty$; (iii) $\{W_r, r = 1, 2, 3\}$ and $\{B_{r0}^{-1}, r = 1, 3\}$ are uniformly bounded in both row and column sums; (iv) For $r = 1, 3$, either $\|B_r^{-1}\|_\infty$ or $\|B_r^{-1}\|_1$ is bounded, uniformly in λ_r in a compact parameter space $\mathbf{\Lambda}_r$, and $0 < \underline{c}_r \leq \inf_{\lambda_r \in \mathbf{\Lambda}_r} \gamma_{\min}(B_r'B_r) \leq \sup_{\lambda_r \in \mathbf{\Lambda}_r} \gamma_{\max}(B_r'B_r) \leq \bar{c}_r < \infty$, where $B_r = B_r(\lambda_r)$.*

Assumption F. *For an $n \times n$ matrix Φ uniformly bounded in either row or column sums, with elements of uniform order h_n^{-1}, and an $n \times 1$ vector b with elements of uniform order*

[13] When T is fixed $\frac{1}{n}\mathbb{Z}'\Omega^{-1}\mathbb{Z} \to \Sigma_Z = F\Sigma_{\Gamma^*}F' + \Sigma_v$, where Σ_{Γ^*} and Σ_v are the limits of $\Gamma^{*\prime}B_{30}'B_{30}\Gamma/n$ and $\mathbb{V}'\mathbb{V}/n$. If $\Sigma_v = \sigma^2_{v0}I_T$, the matrix of the first r eigenvectors of Σ_Z is a rotation of F. See Bai (2009) and Chamberlain and Rothschild (1982) for more detailed discussions.

$h_n^{-1/2}$, *(i)* $\frac{h_n}{n} y_0' \Phi y_0 = O_p(1)$; *(ii)* $\frac{h_n}{n}[y_0 - \mathrm{E}(y_0)]' b = o_p(1)$; *(iii)* $\frac{h_n}{n}[y_0' \Phi y_0 - \mathrm{E}(y_0' \Phi y_0)] = o_p(1)$.

Assumption B assumes that the idiosyncratic error v_{it} to be independent over cross section and time. Cross sectional and time correlations are not a major concern in the present scenario as they are dealt with by the spatial lag term, the spatial error term and the dynamic term. Assumption C(i) is standard for establishing the consistency of $\hat{\delta}$. The consistency of $\hat{\beta}$ and $\hat{\sigma}_v^2$ follows from that of $\hat{\delta}$ and Assumption D. Assumption E imposes standard assumptions on the spatial weight matrices. It parallels Assumption E of Yang (2018) and relates to Lee (2004). Allowing h_n to grow with n but at a slower rate is useful as it corresponds a spatial layout where the *degree of spatial dependence* increases with n. See Lee (2004) and Yang (2015) for related discussions. Assumption F ensures that the initial observations y_0 have proper stochastic behavior. If the process evolves according to (2.1), Assumption F is satisfied by the assumption: $\sum_{i=0}^{\infty} \mathcal{B}_0^i$ exists and is uniformly bounded in both row and column sums, which parallels Assumption 6 in Yu et al. (2008) and Lee and Yu (2014).

Solving the AQS equations in (2.13) for β and σ_v^2 given δ, we obtain the constrained M-estimators $\hat{\beta}(\delta) = [\mathbf{X}'\mathbf{M}_F \mathbf{\Omega}^{-1}(\lambda_3)\mathbf{X}]^{-1}\mathbf{X}'\mathbf{M}_F\mathbf{\Omega}^{-1}(\lambda_3)[\mathbf{B}_1(\lambda_1)Y - \mathbf{B}_2(\lambda_2)Y_{-1}]$, and $\hat{\sigma}_v^2(\delta) = \frac{1}{n(T-r)}\hat{Z}'(\delta)\mathbf{M}_F\mathbf{\Omega}^{-1}(\lambda_3)\hat{Z}(\delta)$, where $\hat{Z}(\delta) = \mathbf{B}_1(\lambda_1)Y - \mathbf{B}_2(\lambda_2)Y_{-1} - \mathbf{X}\hat{\beta}(\delta)$. Substituting $\hat{\beta}(\delta)$ and $\hat{\sigma}_v^2(\delta)$ back into the δ-components of $S_{nT}^*(\psi, \phi)$ gives the concentrated AQS functions $S_{nT}^{*c}(\delta)$ (details are presented in Appendix B). Similarly, let $\bar{S}_{nT}^{*c}(\delta)$ be the population counterpart of the concentrated AQS functions (see Appendix B). It is easy to see that $S_{nT}^{*c}(\hat{\delta}) = \mathbf{0}$, and $\bar{S}_{nT}^{*c}(\delta_0) = \mathbf{0}$. By Theorem 5.9 of van der Vaart (1998), $\hat{\delta}$ will be consistent for δ_0 if $\sup_{\delta\in\mathbf{\Delta}} \frac{1}{nT}\left\|S_{nT}^{*c}(\delta) - \bar{S}_{nT}^{*c}(\delta)\right\| \xrightarrow{p} 0$, and the following identification condition holds.

Assumption G. $\inf_{\delta\ d(\delta,\delta)\geq\varepsilon}\left\|\bar{S}_{nT}^{*c}(\delta)\right\| > 0$ *for every* $\varepsilon > 0$*, where* $d(\delta, \delta_0)$ *is a measure of distance between* δ *and* δ_0*.*

Theorem 3.1. *Suppose Assumptions A-G hold. Assume further that* (i) $\gamma_{\max}[\mathrm{Var}(Y)]$ *and* $\gamma_{\max}[\mathrm{Var}(Y_{-1})]$ *are bounded, and* (ii) $\inf_{\delta\in\mathbf{\Delta}}\gamma_{\min}\left[\mathrm{Var}(\mathbf{B}_1 Y - \mathbf{B}_2 Y_{-1})\right] \geq \underline{c}_y > 0$*. We have as* $n \to \infty$, $\hat{\delta}_{\mathtt{M}} \xrightarrow{p} \delta_0$*. It follows that* $\hat{\beta}_{\mathtt{M}} \xrightarrow{p} \beta_0$*, and* $\hat{\sigma}_{v,\mathtt{M}}^2 \xrightarrow{p} \sigma_{v0}^2$*.*

So far we have assumed that the true number of factors r_0 is known. In fact, ψ could be consistently estimated with a choice of r not less than r_0. From the AQS function in (2.13), we see that, when $r < r_0$, $\mathrm{rank}(M_F(\phi)) < r_0$ and thus $M_F(\phi)$ cannot completely remove $\Gamma_0 F_0'$ from $Z(\theta)$. Therefore, no ϕ can satisfy $\mathrm{E}[S_{nT}^*(\psi, \phi)] = 0$. On the other hand, when $\text{rank}(M_F(\phi)) > r_0$, there are infinitely many ϕ such that $M_F(\phi)$ can completely remove $\Gamma F'$. While ϕ is not identified when $r > r_0$, ψ is, because $\text{E}[S^*_{nT}(\psi, \phi)] = 0$ holds only at $\psi = \psi_0$. To see this, write $Z(\theta) = Z(\theta_0) - \sum_{p=1}^{k+3} X_k(\beta_p - \beta_{p0})$, where X_p is the p th column of $\mathbf{X}, p = 1, \cdots, k$, $X_{k+1} = Y_{-1}, X_{k+2} = \mathbf{W}_1 Y$, and $X_{k+3} = \mathbf{W}_2 Y_{-1}$, with $\beta_{k+1} = \rho$, $\beta_{k+2} = \lambda_1$, and $\beta_{k+3} = \lambda_2$. Take the β_1-component of the AQS function as an example, we have

$$
\begin{aligned}
& \tfrac{1}{\sigma_v^2} X_1' \mathbf{M}_F(\phi) \boldsymbol{\Omega}^{-1}(\lambda_3) Z(\theta) \\
= \; & \tfrac{1}{\sigma_v^2} X_1' \mathbf{M}_F(\phi) \boldsymbol{\Omega}^{-1}(\lambda_3) Z(\theta_0) - \tfrac{1}{\sigma_v^2} \textstyle\sum_{p=1}^{k+3} X_1' \mathbf{M}_F(\phi) \boldsymbol{\Omega}^{-1}(\lambda_3) X_k(\beta_p - \beta_{p0}) \\
= \; & \tfrac{1}{\sigma_v^2} X_1' \mathbf{M}_F(\phi) \boldsymbol{\Omega}^{-1}(\lambda_3) \mathtt{vec}(\Gamma F') + \tfrac{1}{\sigma_v^2} X_1' \mathbf{M}_F(\phi) \boldsymbol{\Omega}^{-1}(\lambda_3) \mathbf{v} \\
& - \tfrac{1}{\sigma_v^2} \textstyle\sum_{p=1}^{k+3} X_1' \mathbf{M}_F(\phi) \boldsymbol{\Omega}^{-1}(\lambda_3) X_k(\beta_p - \beta_{p0}).
\end{aligned}
\tag{3.1}
$$

The expectation of the second term is always zero by Assumption A. When $r < r_0$, the first term cannot be zero as there is no ϕ such that $\mathbf{M}_F(\phi)\mathtt{vec}(\Gamma F') = 0$. When $r > r_0$, there are infinitely many ϕ's such that the first term is zero. The third term is zero only when $\beta_p = \beta_{p0}$. Similar arguments are made in Ahn and Schmidt (2013). This feature is also discussed in Moon and Weidner (2015) for regular panel models, and in Shi and Lee (2017) for SDPD models. We provide simulation results for the misspecified case $r > r_0$ in Sec. 5.[14]

Next, we establish the asymptotic normality of the proposed M-estimator $\hat{\boldsymbol{\psi}}_{\mathtt{M}}$ of $\boldsymbol{\psi} = (\psi', \phi')'$. We expand $S^*_{nT}(\hat{\boldsymbol{\psi}}_{\mathtt{M}})$ at $\boldsymbol{\psi}_0$ and study the asymptotic behaviour of $S^*_{nT}(\boldsymbol{\psi}_0)$ and $\frac{\partial}{\partial \boldsymbol{\psi}'} S^*_{nT}(\bar{\boldsymbol{\psi}})$, for some $\bar{\boldsymbol{\psi}}$ lying between $\hat{\boldsymbol{\psi}}_{\mathtt{M}}$ and $\boldsymbol{\psi}_0$ elementwise. Using the representations given in (2.9) and letting $Z^* = \mathbf{B}_{30} Z$, the AQS vector at $\boldsymbol{\psi}_0$ can be written as follows

$$
S^*_{nT}(\boldsymbol{\psi}_0) = \begin{cases}
\Pi_1' Z^* \\
Z^{*\prime} \Phi_1 Z^* - \mu_{\sigma^2_{v0}}, \\
Z^{*\prime} \Psi_1 \mathbf{y}_0 + Z^{*\prime} \Phi_2 Z^* + \Pi_2' Z^* - \mu_{\rho_0}, \\
Z^{*\prime} \Psi_2 \mathbf{y}_0 + Z^{*\prime} \Phi_3 Z^* + \Pi_3' Z^* - \mu_{\lambda_{10}}, \\
Z^{*\prime} \Psi_3 \mathbf{y}_0 + Z^{*\prime} \Phi_4 Z^* + \Pi_4' Z^* - \mu_{\lambda_{20}}, \\
Z^{*\prime} \Phi_5 Z^* - \mu_{\lambda_{30}}, \\
Z^{*\prime} \Phi_{5+s} Z^*, \; s = 1, 2, \ldots, k_\phi,
\end{cases}
\tag{3.2}
$$

[14]Although the proposed M-estimator remains consistent when $r > r_0$, its limiting distribution is derived under the premise that number of factors is correctly specified. Ahn and Schmidt (2013) propose to estimate r_0 for short panels with IFE by the following information criteria which can also be used in our case:

$$
\hat{r} = \underset{0 \le r \le T-1}{\text{argmin}} \; \ln(\hat{\sigma}_v^2(r)) + g(r) f(n)
$$

where $g(r) = ar$, $f(n) = \frac{\ln n}{n}$, and a is an arbitrarily chosen positive number. Under BIC, we have $nf(n) \to \infty$, and $f(n) \to 0$ as $n \to \infty$, where the first condition ensures that $\text{plim}_{n\to\infty} \Pr(\hat{r} > r_o) = 0$, and the second condition is to ensure $\text{plim}_{N\to\infty} \Pr(\hat{r} < r_0) = 0$. See Ahn and Schmidt (2013) for detailed discussions.

where $\Pi_1 = \frac{1}{\sigma_{v0}^2}(M_{F0} \otimes B_{30})\mathbf{X}$, $\Pi_2 = \frac{1}{\sigma_{v0}^2}(M_{F0} \otimes B_{30})\boldsymbol{\eta}_{-1}$, $\Pi_3 = \frac{1}{\sigma_{v0}^2}(M_{F0} \otimes B_{30})\mathbf{W}_1\boldsymbol{\eta}$, and $\Pi_4 = \frac{1}{\sigma_{v0}^2}(M_{F0} \otimes B_{30})\mathbf{W}_2\boldsymbol{\eta}_{-1}$; $\Phi_1 = \frac{1}{2\sigma_{v0}^4}\mathbf{M}_{F0}$, $\Phi_2 = \frac{1}{\sigma_{v0}^2}(M_{F0} \otimes B_{30})\mathbf{D}_{-1}\mathbf{B}_{30}^{-1}$, $\Phi_3 = \frac{1}{\sigma_{v0}^2}(M_{F0} \otimes B_{30})\mathbf{W}_1\mathbf{D}\mathbf{B}_{30}^{-1}$, $\Phi_4 = \frac{1}{\sigma_{v0}^2}(M_{F0} \otimes B_{30})\mathbf{W}_2\mathbf{D}_{-1}\mathbf{B}_{30}^{-1}$, $\Phi_5 = \frac{1}{\sigma_{v0}^2}(M_{F0} \otimes W_3 B_{30}^{-1})$, and $\Phi_{5+s} = \frac{1}{\sigma_{v0}^2}[M_{F0}\dot{F}_{s0}(F_0'F_0)^{-1}F_0' \otimes I_n], s = 1, \ldots, k_\phi$; $\Psi_1 = \frac{1}{\sigma_{v0}^2}(M_{F0} \otimes B_{30})\mathbf{Q}_{-1}$, $\Psi_2 = \frac{1}{\sigma_{v0}^2}(M_{F0} \otimes B_{30})\mathbf{W}_1\mathbf{Q}$, and $\Psi_3 = \frac{1}{\sigma_{v0}^2}(M_{F0} \otimes B_{30})\mathbf{W}_2\mathbf{Q}_{-1}$; $\mu_{\sigma_v^2} = \frac{n(T-r)}{2\sigma_v^2}$, $\mu_\rho = \mathrm{tr}(\mathbf{M}_F\mathbf{D}_{-1})$, $\mu_{\lambda_1} = \mathrm{tr}(\mathbf{M}_F\mathbf{W}_1\mathbf{D})$, $\mu_{\lambda_2} = \mathrm{tr}(\mathbf{M}_F\mathbf{W}_2\mathbf{D}_{-1})$, and $\mu_{\lambda_3} = (T-r)\mathrm{tr}(B_3^{-1}W_3)$.

Using the relation $Z^* = \mathbf{v} + \mathtt{vec}(B_{30}\Gamma_0 F_0')$, the components of $S_{nT,\psi}^*(\boldsymbol{\psi}_0)$ can be further expressed as linear combinations of terms linear or quadratic in $\mathbf{v}$ and bilinear in $\mathbf{v}$ and $\mathbf{y}_0$ (see Appendix B). These lead to a simple way for establishing the asymptotic normality of the AQS vector and thus the asymptotic normality of the proposed estimator.

Theorem 3.2. *Under the assumptions of Theorem 3.1, we have, as $n \to \infty$*

$$\sqrt{nT}\left(\hat{\boldsymbol{\psi}}_{\mathtt{M}} - \boldsymbol{\psi}_0\right) \xrightarrow{D} N\left(0, \lim_{n\to\infty} H_{nT}^{-1}(\boldsymbol{\psi}_0)\Sigma_{nT}(\boldsymbol{\psi}_0)H_{nT}'^{-1}(\boldsymbol{\psi}_0)\right),$$

where $H_{nT}(\boldsymbol{\psi}_0) = -\frac{1}{nT}\mathrm{E}[\frac{\partial}{\partial\boldsymbol{\psi}'}S_{nT}^*(\boldsymbol{\psi}_0)]$ and $\Sigma_{nT}(\boldsymbol{\psi}_0) = \frac{1}{nT}\mathrm{Var}[S_{nT}^*(\boldsymbol{\psi}_0)]$, both assumed to exist and $H_{nT}(\boldsymbol{\psi}_0)$ to be positive definite, for sufficiently large n.

4. Robust VC Matrix Estimation

While Theorems 3.1 and 3.2 provide theoretical foundations for small-T inferences based on the SDPD-IFE model, empirical applications of the results depend on the availability of consistent estimators of the two matrices $H_{nT}(\boldsymbol{\psi}_0)$ and $\Sigma_{nT}(\boldsymbol{\psi}_0)$. The former can be consistently estimated by its observed counterpart, $H_{nT}(\hat{\boldsymbol{\psi}}_{\mathtt{M}}) = -\frac{1}{nT}\frac{\partial}{\partial\boldsymbol{\psi}'}S_{nT}^*(\hat{\boldsymbol{\psi}}_{\mathtt{M}})$. The analytical expression of $\frac{\partial}{\partial\boldsymbol{\psi}'}S_{nT}^*(\boldsymbol{\psi})$ is given in Appendix B. Unfortunately, the estimation of the latter is not straightforward. From (3.2) we see that the joint AQS function $S_{nT}^*(\boldsymbol{\psi}_0)$ contains three types of elements, $\Pi' Z^*$, $Z^{*\prime}\Psi\mathbf{y}_0$, and $Z^{*\prime}\Phi Z^*$, where Π, Ψ and Φ are non-stochastic vectors or matrices. The traditional plug-in method requires the closed-form expression of $\Sigma_{nT}(\boldsymbol{\psi}_0)$, but the variance of $Z^{*\prime}\Psi\mathbf{y}_0$ and its covariances with $\Pi' Z^*$ and $Z^{*\prime}\Phi Z^*$ involve the unconditional distribution of y_0 and the factor loadings Γ_0. The distribution of y_0 depends on the past values of the regressors and the process starting positions, which are unobserved,[15] and a consistent estimate of the $n \times r$ matrix Γ_0 is impossible to obtain when T is fixed. Thus,

[15] A valid model for y_0, as that in Su and Yang (2015) for an SDPD model with SE only, is very difficult (if not impossible) to formulate due to the existence of spatial lag terms, as commented by Yang (2018).

the plug-in method based on the analytical expression of $\Sigma_{nT}(\boldsymbol{\psi}_0)$ does not work in this case.

To overcome the difficulties induced by the initial conditions, Yang (2018) proposed an *outer-product-of-martingale-difference* (OPMD) method for estimating the VC matrix of an SDPD-AFE model. The central idea behind this method is to decompose the AQS functions into a sum of n terms, which form a martingale difference (MD) sequence so that the average of the outer products of the MDs gives a consistent estimate of the VC matrix of that AQS function. While this OPMD method does not directly apply to our SDPD-IFE model due to the fact that the original errors v_t are not estimable,[16] the idea of decomposition prevails!

Inspired by the OPMD method, we decompose the AQS function as $S^*_{nT}(\boldsymbol{\psi}_0) = \sum_{i=1}^n \mathbf{g}_i$, where $\{\mathbf{g}_i\}$ are defined in terms of z^*_{it} and some nonstochastic quantities that depend on $\boldsymbol{\psi}_0$ and $W_r, r = 1, 2, 3$, taking full use of the independence of z^*_{it} across i and the fact that T is small and fixed. Based on our decomposition, $\{\mathbf{g}_i\}$ are nearly an MD sequence, which are 'estimable' and thus lead to a feasible estimator for $\Sigma_{nT}(\boldsymbol{\psi}_0)$ through the average of the outer products of $\mathbf{g}_i$ and their analytical covariances:

$$\Sigma_{nT}(\boldsymbol{\psi}_0) = \tfrac{1}{nT}\mathrm{E}[S^*_{nT}(\boldsymbol{\psi}_0)S^{*\prime}_{nT}(\boldsymbol{\psi}_0)] = \tfrac{1}{nT}\textstyle\sum_{i=1}^n \mathrm{E}(\mathbf{g}_i\mathbf{g}_i') + \tfrac{1}{nT}\sum_{i=1}^n\sum_{j\neq i}\mathrm{E}(\mathbf{g}_i\mathbf{g}_j'). \qquad (4.1)$$

The first term in (4.1) can be estimated by its sample analogue $\frac{1}{nT}\sum_{i=1}^n \hat{\mathbf{g}}_i\hat{\mathbf{g}}_i'$, where $\hat{\mathbf{g}}_i$ a plug-in estimate of $\mathbf{g}_i$. The full analytical expression of $\Upsilon(\boldsymbol{\psi}_0) = \sum_{i=1}^n\sum_{j\neq i}\mathrm{E}(\mathbf{g}_i\mathbf{g}_j')$ is derived. Due to the way $\{\mathbf{g}_i\}$ are constructed, the $(k+5+k_\phi)\times(k+5+k_\phi)$ matrix $\Upsilon(\boldsymbol{\psi}_0)$ does not involve the initial conditions or factor loadings and it depends only on $\boldsymbol{\psi}_0$. Therefore, the covariance term $\Upsilon(\boldsymbol{\psi}_0)$ can be consistently estimated using the plug-in method. The estimator of the VC matrix of the estimating functions is given by the following

$$\hat{\Sigma}_{nT} = \tfrac{1}{nT}\textstyle\sum_{i=1}^n \hat{\mathbf{g}}_i\hat{\mathbf{g}}_i' + \tfrac{1}{nT}\Upsilon(\hat{\boldsymbol{\psi}}_{\mathtt{M}}). \qquad (4.2)$$

For this we term our method of VC matrix estimation as the *extended* OPMD method.

Now, we present the details of the decomposition, $S^*_{nT}(\boldsymbol{\psi}_0) = \sum_{i=1}^n \mathbf{g}_i$, and derive the correction term $\Upsilon(\boldsymbol{\psi}_0)$. Recall that components of the joint AQS vector $S^*_{nT}(\boldsymbol{\psi}_0)$ are linear combinations of three types of terms $\Pi' Z^*$, $Z^{*\prime}\Psi\mathbf{y}_0$, and $Z^{*\prime}\Phi Z^*$, we decompose each type separately into $\sum_{i=1}^n g_{\Pi i}$, $\sum_{i=1}^n g_{\Psi i}$ and $\sum_{i=1}^n g_{\Phi i}$. Then, we can use the linear combinations of $g_{ri}, r = \Pi, \Psi, \Phi$ to construct the vector $\mathbf{g}_i$. And naturally, elements of $\mathrm{E}(\mathbf{g}_i\mathbf{g}_j')$ are linear

[16] This is seen from the relation $z^*_t = v_t + B_3\Gamma f_t$, where z^*_t can be consistently estimated by $\hat{z}^*_t$, but the factor loadings Γ and hence v_t cannot be consistently estimated when T is fixed.

combinations of $\mathrm{E}(g_{ri}g_{\nu i}), r, \nu = \Pi, \Psi, \Phi$. To proceed, for a square matrix A, let A^u, A^l and A^d be, respectively, its upper-triangular, lower-triangular, and diagonal matrix such that $A = A^u + A^l + A^d$. Denote by Π_t, Φ_{ts} and Ψ_{ts} the submatrices of Π, Φ and Ψ partitioned according to $t, s = 1, \dots, T$. Similarly, for a vector K, let K_t denote its subvectors partitioned according to $t = 1, \dots, T$. Denote the partial sum of time-indexed quantities using the '$_+$' notation: e.g., $\Psi_{t+} = \sum_{s=1}^T \Psi_{ts}, \Psi_{+s} = \sum_{t=1}^T \Psi_{ts}$ $\Psi_{++} = \sum_{t=1}^T \sum_{s=1}^T \Psi_{ts}$, and similarly for Φ_{ts}, Π_t and other time-indexed quantities. **Recall** $Z^* = \mathbf{v} + \mathtt{vec}(B_{30}\Gamma_0 F_0')$.

First, consider a linear term $\Pi' Z^*$. We have $\Pi' Z^* = \Pi' \mathbf{v} + \Pi' \mathtt{vec}(B_{30}\Gamma_0 F_0')$. From (3.2), we see that Π takes the form $\mathbf{M}_{F_0} K$ for a suitably defined nonstochastic vector or matrix K involving $\boldsymbol{\psi}_0$, $\mathbf{X}$, and $W_r, r = 1, 2, 3$. Without loss of generality, assume Π is a vector ($nT \times 1$) and so is K, as if not we can work on each column of it. Using $\Pi = \mathbf{M}_{F_0} K$ and letting $\mathbb{K}$ be such that $K = \mathtt{vec}(\mathbb{K})$, we have by the matrix result in Footnote 7, $\Pi' \mathtt{vec}(B_{30}\Gamma_0 F_0') = K'(M_{F_0} \otimes I_n)\mathtt{vec}(B_{30}\Gamma_0 F_0') = \mathrm{tr}(B_{30}\Gamma_0 F_0' M_{F_0} \mathbb{K}') = 0$. Therefore, $\Pi' Z^* = \Pi' \mathbf{v}$, and we have the following decomposition for any Π defined in (3.2), noting that $\mathrm{E}(\Pi' \mathbf{v}) = 0$:

$$\Pi' Z^* = \Pi' \mathbf{v} = \textstyle\sum_{i=1}^n (\sum_{t=1}^T \Pi_{it}' v_{it}) \equiv \sum_{i=1}^n g_{\Pi,i}, \tag{4.3}$$

where Π_{it}' be the ith row of Π_t. Clearly, $\{g_{\Pi,i}\}$ are uncorrelated.

Next, consider a bilinear term $Z^{*\prime} \Psi \mathbf{y}_0$. Again we separate Z^* into two parts and write $Z^{*\prime} \Psi \mathbf{y}_0 = \mathbf{v}' \Psi \mathbf{y}_0 + \mathtt{vec}(B_{30}\Gamma_0 F_0')' \Psi \mathbf{y}_0$. By the expressions of Ψ given in (3.2), each $nT \times 1$ vector $\Psi \mathbf{y}_0$ can be written in the form $\Psi \mathbf{y}_0 = \mathbf{M}_{F_0} K$ for a suitably defined vector K involving $\mathbf{y_0}$, $\boldsymbol{\psi}_0$, and $W_r, r = 1, 2, 3$. Again, by the matrix result in Footnote 7, we show that $\mathtt{vec}(B_{30}\Gamma_0 F_0')' \Psi \mathbf{y}_0 = \mathtt{vec}(B_{30}\Gamma_0 F_0')' \mathbf{M}_{F_0} K = 0$. Therefore, $Z^{*\prime} \Psi \mathbf{y}_0 = \mathbf{v}' \Psi \mathbf{y}_0$. With $\mathrm{E}(\mathbf{v}' \Psi \mathbf{y}_0) = 0$ due to the independence between y_0 and $\{v_t, t \geq 1\}$, we have the following decomposition of a bilinear term for any Ψ defined in (3.2):,

$$Z^{*\prime} \Psi \mathbf{y}_0 = \mathbf{v}' \Psi \mathbf{y}_0 = \textstyle\sum_{i=1}^n \sum_{t=1}^T v_{it} \xi_{it} \equiv \sum_{i=1}^n g_{\Psi,i}, \tag{4.4}$$

where $\{\xi_{it}\} = \xi_t = \Psi_{t+} y_0$, $\{g_{\Psi,i}\}$ are uncorrelated, and $g_{\Psi,i}$ is uncorrelated with $g_{\Pi,j}$, $i \neq j$.

Finally, for a quadratic term $Z^{*\prime} \Phi Z^*$, we separate the first Z^* into two parts and write $Z^{*\prime} \Phi Z^* = \mathbf{v}' \Phi Z^* + \mathtt{vec}(B_{30}\Gamma_0 F_0')' \Phi Z^*$. From (3.2), we see that ΦZ^* can also be written in the form $\mathbf{M}_F K$ for a suitably defined vector K involving $\boldsymbol{\psi}_0$, Z, and $W_r, r = 1, 2, 3$. Thus, $\mathtt{vec}(B_{30}\Gamma_0 F_0')' \Phi Z^* = \mathtt{vec}(B_{30}\Gamma_0 F_0')' \mathbf{M}_{F_0} K = 0$, by the matrix result in Footnote 7.

Therefore, $Z^{*\prime}\Phi Z^* = \mathbf{v}'\Phi Z^*$, and the latter can be decomposed for any Φ defined in (3.2) as,

$$\begin{aligned}\mathbf{v}'\Phi Z^* &= \textstyle\sum_{t=1}^T \sum_{s=1}^T v_t'\Phi_{ts} z_s^* \\ &= \textstyle\sum_{t=1}^T \sum_{s=1}^T v_t'\Phi_{ts}^u z_s^* + \sum_{t=1}^T \sum_{s=1}^T v_t'\Phi_{ts}^l z_s^* + \sum_{t=1}^T \sum_{s=1}^T v_t'\Phi_{ts}^d z_s^* \\ &= \textstyle\sum_{i=1}^n (\sum_{t=1}^T v_{it}\varphi_{it} + \sum_{t=1}^T v_{it} z_{it}^d),\end{aligned} \tag{4.5}$$

where $\{\varphi_{it}\} = \varphi_t = \sum_{s=1}^T (\Phi_{ts}^u + \Phi_{ts}^\ell) z_s^*$, and $\{z_{it}^d\} = z_t^d = \sum_{s=1}^T \Phi_{ts}^d z_s^*$. By Assumptions A and B, $\mathrm{E}(v_{it}\varphi_{it}) = 0$ and $\mathrm{E}(v_{it} z_{it}^d) = \sigma_{v0}^2 \Phi_{ii,tt} \equiv d_{it}$, where $\Phi_{ii,tt}$ is the ith diagonal element of Φ_{tt}. These lead to the following decomposition for a quadratic term:

$$\mathbf{v}'\Phi Z^* - \mathrm{E}(\mathbf{v}'\Phi Z^*) = \textstyle\sum_{i=1}^n [\sum_{t=1}^T v_{it}\varphi_{it} + \sum_{t=1}^T (v_{it} z_{it}^d - d_{it})] \equiv \sum_{i=1}^n g_{\Phi,i}. \tag{4.6}$$

While $\{g_{\Phi,i}\}$ are correlated, $g_{\Phi,i}$ is uncorrelated with $g_{\Pi,j}$ and $g_{\Psi,j}$, $i \neq j$, as shown below.

The decompositions of the three types of quantities given by (4.3)-(4.6) lead immediately to a decomposition of $S_{nT}^*(\boldsymbol{\psi}_0)$. For for each $\Pi_r, r = 1,2,3,4$ defined in (3.2), define $g_{\Pi_r,i}$ according to (4.3); for each $\Psi_r, r = 1,2,3$ defined in (3.2), define $g_{\Psi_r,i}$ according to (4.4); and for each $\Phi_r, r = 1,2,\ldots,5+k_\phi$ defined in (3.2), define $g_{\Phi_r,i}$ according to (4.6). Define,

$$\mathbf{g}_i = \begin{cases} g_{\Pi_1,i} \\ g_{\Phi_1,i} \\ g_{\Pi_2,i} + g_{\Phi_2,i} + g_{\Psi_1,i} \\ g_{\Pi_3,i} + g_{\Phi_3,i} + g_{\Psi_2,i} \\ g_{\Pi_4,i} + g_{\Phi_4,i} + g_{\Psi_3,i} \\ g_{\Phi_5,i} \\ g_{\Phi_{5+s},i}, \ s = 1,2,\ldots,k_\phi \end{cases} \tag{4.7}$$

Then, the AQS vector at the true parameter value is $S_{nT}^*(\boldsymbol{\psi}_0) = \sum_{i=1}^n \mathbf{g}_i$. The $\{\mathbf{g}_i\}$ are nearly uncorrelated as seen from (4.3)-(4.6), and the details given below.

The nature of such decompositions (many terms are uncorrelated) opens up a simple way for a consistent estimate of the VC matrix of the AQS function. From its general form given in (4.1), the first term $\frac{1}{nT}\sum_{i=1}^n \mathrm{E}(\mathbf{g}_i\mathbf{g}_i')$ can be estimated by its sample analogue $\frac{1}{nT}\sum_{i=1}^n \hat{\mathbf{g}}_i\hat{\mathbf{g}}_i'$, where $\hat{\mathbf{g}}_i$ is obtained by replacing both v_{it} and z_{it}^* in (4.7) by $\hat{z}_{it}^*$, and replacing $\boldsymbol{\psi}_0$ by $\hat{\boldsymbol{\psi}}$. This is justified by the results $\Pi' Z^* = \Pi'\mathbf{v}$, $Z^{*\prime}\Psi\mathbf{y}_0 = \mathbf{v}'\Psi\mathbf{y}_0$, and $Z^{*\prime}\Phi Z^* = \mathbf{v}'\Phi Z^*$ given above, and the consistency of the M-estimator $\hat{\boldsymbol{\psi}}_{\mathtt{M}}$. See the proof of Theorem 4.1 for details.

To derive the analytical form of $\Upsilon(\boldsymbol{\psi}_0) = \sum_{i=1}^n \sum_{j\neq i} \mathrm{E}(\mathbf{g}_i\mathbf{g}_j')$. Note that the expectations of $g_{\Pi_r,i}$, $g_{\Psi_r,i}$ and $g_{\Phi_r,i}$ in (4.7) are all zero, for all r. First, by Assumptions A and B and the expressions (4.3) and (4.4), we show that $(g_{\Pi_r,i}, g_{\Psi_\nu,i})$ are uncorrelated, i.e., $\mathrm{E}(g_{\boldsymbol{\pi}_r,i} g_{\boldsymbol{\pi}_\nu,j})$, $\mathrm{E}(g_{\Psi_r,i} g_{\Psi_\nu,j})$ and $\mathrm{E}(g_{\boldsymbol{\pi}_r,i} g_{\Psi_\nu,j})$ are all zero, for $i \neq j$, $r = 1,2,3,4$, and $\nu = 1,2,3$. Next, by

(4.3)-(4.6) and Assumptions A and B, we have, for $i \neq j$ $(= 1, \ldots, n)$,

$$
\begin{aligned}
&\mathrm{E}(g_{\Phi_r,i}g_{\Pi_\nu,j}) = \mathrm{E}\{[\textstyle\sum_{t=1}^T v_{it}\varphi_{r,it} + \sum_{t=1}^T(v_{it}z_{r,it}^d - d_{r,it})](\sum_{t=1}^T \Pi_{\nu,jt}v_{jt})\} \\
= \; &\mathrm{E}[(\textstyle\sum_{t=1}^T v_{it}\varphi_{r,it})(\sum_{t=1}^T \Pi_{\nu,jt}v_{jt})] + \mathrm{E}[\sum_{t=1}^T(v_{it}z_{r,it}^d - d_{r,it})(\sum_{t=1}^T \Pi_{\nu,jt}v_{jt})] = 0;
\end{aligned} \tag{4.8}
$$

$$
\begin{aligned}
&\mathrm{E}(g_{\Phi_r,i}g_{\Psi_\nu,j}) = \mathrm{E}\{[\textstyle\sum_{t=1}^T v_{it}\varphi_{r,it} + \sum_{t=1}^T(v_{it}z_{r,it}^d - d_{r,it})](\sum_{t=1}^T v_{jt}\xi_{\nu,jt})\} \\
= \; &\mathrm{E}[(\textstyle\sum_{t=1}^T v_{it}\varphi_{r,it})(\sum_{t=1}^T v_{jt}\xi_{\nu,jt})] + \mathrm{E}[\sum_{t=1}^T(v_{it}z_{r,it}^d - d_{r,it})(\sum_{t=1}^T v_{jt}\xi_{\nu,jt})] = 0.
\end{aligned} \tag{4.9}
$$

Therefore, $g_{\Phi_r,i}$ is uncorrelated with $g_{\Pi_\nu,j}$ and $g_{\Psi_\nu,j}$, $i \neq j$. These results show that the covariance between $\mathbf{g}_i$ and $\mathbf{g}_j$ comes only from the covariance between $g_{\Phi_r,i}$ and $g_{\Phi_\nu,j}, i \neq j$, and $r, \nu = 1, 2, \ldots, 5 + k_\phi$. Let a'_{its} be the ith row of the $n \times n$ matrix $\Phi_{ts}^u + \Phi_{ts}^\ell$, and a_{ijts} be the jth element of a'_{its}. Under Assumptions A and B, we have for $i \neq j$,

$$
\begin{aligned}
\mathrm{E}(g_{\Phi_r,i}g_{\Phi_\nu,j}) \; &= \mathrm{E}[(\textstyle\sum_{t=1}^T v_{it}\varphi_{r,it})(\sum_{s=1}^T v_{js}\varphi_{\nu,jt})] \\
&= \textstyle\sum_{t=1}^T \sum_{s=1}^T \mathrm{E}[v_{it}v_{js}(\sum_{p=1}^T a'_{r,itp}z_p^*)(\sum_{p=1}^T a'_{\nu,jsp}z_p^*)] \\
&= \textstyle\sum_{t=1}^T \sum_{s=1}^T \mathrm{E}[v_{it}(\sum_{p=1}^T a'_{\nu,jsp}z_p^*)]\mathrm{E}[v_{js}(\sum_{p=1}^T a'_{r,itp}z_p^*)] \\
&= \textstyle\sum_{t=1}^T \sum_{s=1}^T \mathrm{E}[v_{it}a'_{\nu,jst}z_t^*)]\mathrm{E}[v_{js}a'_{r,its}z_s^*)] \\
&= \textstyle\sum_{t=1}^T \sum_{s=1}^T \mathrm{E}[a_{\nu,jist}v_{it}z_{it}^*)]\mathrm{E}[a_{r,ijts}v_{js}z_{js}^*)] \\
&= \sigma_{v0}^4 \textstyle\sum_{t=1}^T \sum_{s=1}^T a_{\nu,jist}a_{r,ijts}.
\end{aligned} \tag{4.10}
$$

Collecting all the results above, we have the non-zero elements of $\Upsilon(\boldsymbol{\psi}_0)$ as follows,

$$
\begin{aligned}
\Upsilon_{k+r,k+\nu}(\boldsymbol{\psi}_0) \; &= \textstyle\sum_{i=1}^n \sum_{j\neq i}^n \mathrm{E}(g_{\Phi_r,i}g_{\Phi_\nu,j}) \\
&= \textstyle\sum_{i=1}^n \sum_{j\neq i}^n \sigma_{v0}^4 \sum_{t=1}^T \sum_{s=1}^T a_{\nu,jist}a_{r,ijts} \\
&= \sigma_{v0}^4 \mathrm{tr}(\Phi_r\Phi_\nu) - \sigma_{v0}^4 \textstyle\sum_{i=1}^n \sum_{t,s=1}^T a_{\nu,iist}a_{r,iits},
\end{aligned} \tag{4.11}
$$

for $r, \nu = 1, 2, \ldots 5 + k_\phi$. These show that the covariance matrix $\Upsilon(\boldsymbol{\psi}_0)$ has a simple form and depends only on $\boldsymbol{\psi}_0$. Thus, it can be consistently estimated by plugging in a consistent estimate of $\boldsymbol{\psi}_0$. Finally, the consistency of the proposed estimator of the variance of the estimating functions, $\hat{\Sigma}_{nT} = \frac{1}{nT}\sum_{i=1}^n \hat{\mathbf{g}}_i\hat{\mathbf{g}}'_i + \frac{1}{nT}\Upsilon(\hat{\boldsymbol{\psi}}_{\mathtt{M}})$, is proved in the following theorem.

Theorem 4.1. *Under the assumptions of Theorem 3.1, we have, as $n \to \infty$*

$$
\widehat{\Sigma}_{nT} - \Sigma(\boldsymbol{\psi}_0) = \frac{1}{nT}\sum_{i=1}^n [\hat{\mathbf{g}}_i\hat{\mathbf{g}}'_i - \mathrm{E}(\mathbf{g}_i\mathbf{g}'_i)] + \frac{1}{nT}[\Upsilon(\hat{\boldsymbol{\psi}}_{\mathtt{M}}) - \Upsilon(\boldsymbol{\psi}_0)] \xrightarrow{p} 0,
$$

and hence $H_{nT}^{-1}(\hat{\boldsymbol{\psi}}_{\mathtt{M}})\widehat{\Sigma}_{nT}H_{nT}^{-1\prime}(\hat{\boldsymbol{\psi}}_{\mathtt{M}}) - H_{nT}^{-1}(\boldsymbol{\psi}_0)\Sigma_{nT}(\boldsymbol{\psi}_0)H_{nT}^{-1\prime}(\boldsymbol{\psi}_0) \xrightarrow{p} 0.$

5. Monte Carlo Study

Extensive Monte Carlo experiments are run to investigate the finite sample performance of the proposed M-estimator of the SDPD-IFE model and the extended OPMD estimator of its VC matrix. We use the following two data generating processes (DGPs):

$$\text{DGP1: } y_t = \rho y_{t-1} + \lambda_1 W_1 y_t + \lambda_2 W_2 y_{t-1} + x_t\beta + \Gamma f_t + u_t, \quad u_t = \lambda_3 W_3 u_t + v_t;$$

$$\text{DGP2: } y_t = \rho y_{t-1} + \lambda_1 W_1 y_t + \lambda_2 W_2 y_{t-1} + x_t\beta + \Gamma f_t + v_t.$$

To substantiate our claim that the proposed methods are superior when T is small, comparisons are made with (i) the bias corrected CQML estimator (BC-CQMLE) of Shi and Lee (2017) using DGP1,[17] and (ii) the GMM estimator in Kuersteiner and Prucha (2020) using DGP2.[18] The former is designed for large T and the latter is valid for small T.

The exogenous time varying regressors x_t, the $T \times r$ matrix of unobserved factors F and their $n \times r$ loadings matrix Γ are generated in a similar fashion as Shi and Lee (2017). $x_t = (x_{1,t}, x_{2,t})$ is an $n \times 2$ matrix of regressors, whose elements are generated according to $x_{1,it} = 0.25(\gamma_i' f_t + (\gamma_i' f_t)^2 + 1'\gamma_i + 1' f_t) + \eta_{1,it}$, and $x_{2,it} = c\eta_{2,it}$. The elements of $\gamma_i, f_t, \eta_{1it}$, and $\eta_{2,it}$ are generated independently from standard normal distribution, and c is a constant. We use $c = 1$ for DGP1 and $c = 2$ for DGP2 as the numerical stability of the GMM method requires a significantly larger signal-to-noise ratio. The spatial weight matrices are generated according to the following schemes: Rook contiguity, Queen contiguity, or group interaction.[19]

The error (v_t) distribution can be (i) normal, (ii) normal mixture $(10\%N(0,4), 90\%N(0,1))$, or (iii) chi-squared with degrees of freedom 3. In both (ii) and (iii), the generated errors are standardized to have mean zero and variance σ_v^2. We choose $\beta_1 = \beta_2 = \sigma_v^2 = 1$, $\rho = 0.3$, and $\lambda_1 = \lambda_2 = \lambda_3 = 0.2$. The number of factors $r = 1$ or 2. We set the processes starting time at $t = -10$ $(m = 10)$, $n = 50, 100, 200, 400$ for $T = 3$, and $n = 25, 50, 100, 200$ for $T = 10$. Each set of Monte Carlo results, under a set of values of $(n, T, \rho, \lambda' s)$, is based on 2000 samples.

Monte Carlo (empirical) mean and standard deviation (sd) are reported for the proposed M-estimator, along with $\widehat{\text{rse}}$, the empirical average of the robust standard errors (ses) based on the VC matrix estimate $H_{nT}^{-1}(\hat{\psi}_{\mathtt{M}})\hat{\Sigma}_{nT}H_{nT}^{-1}(\hat{\psi}_{\mathtt{M}})$. $\widehat{\text{rse}}$ should be compared with the corre-

[17] We thank the authors for making their codes available at `https://www.w-shi.net/research.html`.

[18] We thank the authors for codes at `http://econweb.umd.edu/%7Ekuersteiner/research_UMD.html`.

[19] The Rook and Queen schemes are standard. For group interaction, we first generate $k = n^\alpha$ groups of sizes $n_g \sim U(.5\bar{n}, 1.5\bar{n})$, $g = 1, \cdots, k$, where $0 < \alpha < 1$ and $\bar{n} = n/k$, and then adjust n_g so that $\sum_{g=1}^k n_g = n$. The reported results correspond to $\alpha = 0.5$. See Yang (2015) for details in generating these spatial layouts.

sponding empirical sd. The ses of the M-estimator, $\widehat{\text{se}}$ and $\widetilde{\text{se}}$, based on $\widehat{\Sigma}_{nT}$ only and on $\widehat{H}_{nT}$ only are also computed, and the results (unreported to conserve space) show that they are not robust. A subset of results are reported in Tables 1-7. Monte Carlo results involved in the discussions but unreported due to space constraint can be found online through Appendix C.

The results show an excellent finite sample performance of the proposed M-estimator and the OPMD-type estimator of the VC matrix of the M-estimator, irrespective of the spatial layouts, the error distributions, the number of factors, etc. The proposed estimation and inference methods clearly dominant, in terms of bias and efficiency, the bias-corrected CQML method of Shi and Lee (2017) valid under large T (Tables 1-6), and the GMM method of Kuersteiner and Prucha (2020) under a more general setup (Table 7).

Table 1 presents the results with $T = 3$, $r = r_0 = 1$ and Rook contiguity spatial layout. The M-estimator of the dynamic parameter is nearly unbiased, whereas the corresponding BC-QMLE can be quite biased and as n increases it does not show a sign of convergence. This shows that their bias correction does not address the initial values problem when T is small. The M-estimators of the spatial parameters λ_1 and λ_2 also show an excellent finite sample performance, whereas that of λ_3 shows some small bias when errors are drawn from the chi-squared distribution. The BC-QMLE of λ_1 performs quite well, but these of λ_2 and λ_3 are slightly biased. While the biases of the BC-QMLEs of λ_2 and λ_3 are not severe, the standard error estimate (reported in Appendix C) performs poorly. In contrast, the robust ses (rses) of M-estimator are on average very close to the corresponding Monte Carlo sds, showing the robustness and good finite sample performance of the proposed VC matrix estimate.

Table 2 presents the results with $T = 3$, $r = r_0 = 1$, group interaction for W_1 and W_2, and Queen contiguity W_3. Under these much denser spatial layouts, the proposed robust M-estimators continue to perform very well, whereas the BC-QMLEs for ρ and $\lambda's$ deteriorate significantly, which can be severely biased and show a clear pattern of inconsistency. Moreover, the rses of our M-estimator still performs quite well and are generally very close to the corresponding Monte Carlo sds, whereas the ses of BC-QMLE again show large biases.

Table 3 presents the results with $T = 3$, $r = r_0 = 2$, and Rook contiguity spatial weight matrices. Compared with Table 1, the M-estimators have slightly larger bias and sds when the number of factors increases as expected, but their performance is still satisfactory and more importantly the sign of convergence is clear. Moreover, the rses are also generally close to the corresponding Monte Carlo sds. The BC-QMLEs, on the other hand, are severely biased under this setting, especially for ρ and λ_1. The associated standard error estimates of the BC-QMLEs perform even worse (see Appendix C).

Tables 4 and 5 present the results with $T = 10$, $r = r_0 = 1$, under Rook contiguity spatial layouts and a combination of group interaction and Queen spatial layouts, respectively. Results show that increasing T further improves performance of the M-estimators and their robust standard error estimates. Increasing T significantly improves the performance of the BC-CQML estimators so that they become comparable with the M-estimators except the BC-CQMLE of the error variance. Further, the standard errors estimates of the BC-CQMLEs are still noticeably biased, whereas the proposed rses of the M-estimators are very accurate.

Table 6 presents the results when number of factors is misspecified. The true number of factor is $r_0 = 1$ but number of factor assumed in the estimation is $r = 2$. The proposed M-estimators perform reasonably well under misspecification. The M-estimator of σ_v^2 show slightly larger bias than that in the correctly specified case while the M-estimators of the other parameters show similar performance in terms of bias as in Table 1. The sds are slightly larger than that in the correctly specified cases. As expected, the rses show some bias as the asymptotic distribution of the AQS estimator is established based on true number of factor. The BC-QMLE performs poorly with much larger bias as compared to the M-estimators.

Table 7 presents the estimation results under DGP2, for the purpose of comparing our M-estimator with the GMM estimator of Kuersteiner and Prucha (2020). From the results we see that (i) both estimators show clear patterns of convergence, (ii) both perform well in terms of bias with M-estimator being slightly better, and (iii) the M-estimator is much more efficient than the GMM estimator as shown by the empirical sds, for all sample sizes and all error distributions considered. Furthermore, our Monte Carlo experiments show that the GMM estimator requires a larger signal-to-noise ratio for numerical stability. These confirm the general statements made in the introduction: the proposed strategy focuses on efficiency and simplicity, whereas that of Kuersteiner and Prucha (2020) stresses on generality.

6. Conclusion and Discussion

This paper proposes a set of new estimation and inference methods for spatial dynamic panel data models with interactive fixed effect based on short panels, the adjusted quasi score (AQS) or M-estimation method and the extended *outer-product-of-martingale-difference* method. The advantage of the proposed AQS estimation methodology is that it **adjusts** the conditional concentrated quasi score functions to **remove** the effects of conditioning and concentration. Thus, it is free from the initial conditions, the process starting time and the factor loadings. It is simple and reliable, preserving the efficiency properties of the likelihood-type of estimation, and leading naturally to a simple method for standard error estimation. In contrast, the existing methods are either invalid under short panels or inefficient and computationally complicated due to the use of GMM method under a more general model setup. In addition, the nature of the proposed estimation and inference methods suggests that there is a great potential for extensions to allow for additional features in the model.

Extensions. An interesting extension to consider is to allow for cross-sectional heteroskedasticity in the error terms, as discussed in Remark 2.2 and specified in Footnote 10. Letting $\mathbf{H} = (I_T \otimes \mathcal{H})$, it is easy to verify the following results:

$$\mathrm{E}(Z'\mathbf{M}_{F0}\mathbf{\Omega}_0^{-1}Y_{-1}) = \sigma_{v0}^2\mathrm{tr}(\mathbf{D}_{-1}\mathbf{M}_{F0}\mathbf{B}_{30}^{-1}\mathbf{H}\mathbf{B}_{30}), \tag{6.1}$$

$$\mathrm{E}(Z'\mathbf{M}_{F0}\mathbf{\Omega}_0^{-1}\mathbf{W}_1 Y) = \sigma_{v0}^2\mathrm{tr}(\mathbf{D}\mathbf{M}_{F0}\mathbf{B}_{30}^{-1}\mathbf{H}\mathbf{B}_{30}\mathbf{W}_1), \tag{6.2}$$

$$\mathrm{E}(Z'\mathbf{M}_{F0}\mathbf{\Omega}_0^{-1}\mathbf{W}_2 Y_{-1}) = \sigma_{v0}^2\mathrm{tr}(\mathbf{D}_{-1}\mathbf{M}_{F0}\mathbf{B}_{30}^{-1}\mathbf{H}\mathbf{B}_{30}\mathbf{W}_2), \tag{6.3}$$

$$\mathrm{E}(Z'\mathbf{M}_{F0}\mathbf{B}_{30}'\mathbf{W}_3 Z) = (T-r)\sigma_{v0}^2\mathrm{tr}(B_{30}^{-1}\mathcal{H}W_3). \tag{6.4}$$

Therefore, the ρ and λ components $\mathrm{E}[\frac{\partial}{\partial\psi}\ell_{nT}^c(\psi_0,\phi_0)]$ are no longer functions of only (ψ_0,ϕ_0); they contain the unknown heteroskedasticity matrix $\mathcal{H}$.

While this makes the direct adjustment method as in the paper infeasible, the idea of AQS prevails, showing the generality and flexibility of the AQS method. As in Li and Yang (2019) for an SDPD model with additive FE, instead of directly subtracting the expectation, we can find a set of quadratic terms in Z with expectations being identical to (6.1)-(6.4):

$$\mathrm{E}(Z'\Omega_0^{-1}\mathbf{D}_{-1}\mathbf{M}_{F0}Z) = \sigma_{v0}^2\mathrm{tr}(\mathbf{D}_{-1}\mathbf{M}_{F0}\mathbf{B}_{30}^{-1}\mathbf{H}\mathbf{B}_{30}), \tag{6.5}$$

$$\mathrm{E}(Z'\Omega_0^{-1}\mathbf{W}_1\mathbf{D}\mathbf{M}_{F0}Z) = \sigma_{v0}^2\mathrm{tr}(\mathbf{D}\mathbf{M}_{F0}\mathbf{B}_{30}^{-1}\mathbf{H}\mathbf{B}_{30}\mathbf{W}_1), \tag{6.6}$$

$$\mathrm{E}(Z'\Omega_0^{-1}\mathbf{W}_2\mathbf{D}_{-1}\mathbf{M}_{F0}Z) = \sigma_{v0}^2\mathrm{tr}(\mathbf{D}_{-1}\mathbf{M}_{F0}\mathbf{B}_{30}^{-1}\mathbf{H}\mathbf{B}_{30}\mathbf{W}_2), \tag{6.7}$$

$$\mathrm{E}[Z'\mathbf{B}_{30}'[I_T\otimes \mathtt{diag}(W_3B_{30}^{-1})]\mathbf{B}_{30}\mathbf{M}_{F0}Z] = (T-r)\sigma_{v0}^2\mathrm{tr}(B_{30}^{-1}\mathcal{H}W_3). \tag{6.8}$$

Modifying the CCQS with the set of quadratic terms above will lead to a set of unbiased estimating equations robust against unknown $\mathcal{H}$. Note that the ϕ-component of the AQS is naturally robust against unknown $\mathcal{H}$ as shown in Footnote 10, and the two step computation approach still works under heteroskedasticity (see footnote 11 for details). Moreover the β' and σ_v^2 components also do not need further adjustment under heteroskedasticity. After the adjustment, our estimation and inference method will go through as before and remain valid. While the fundamental ideas are clear, these extensions require additional complicated algebra and proofs, and can only be handled by a separate research.

Our methods can also be extended to allow for multiple time lags and multiple spatial lags as in Kuersteiner and Prucha (2020), to give the following high-order SDPD-IFE model:

$$\begin{aligned} y_t &= \rho_1 y_{t-1} + \rho_2 y_{t-2} + \textstyle\sum_{\ell=1}^{p_1} \lambda_{1\ell} W_{1\ell} y_t + \sum_{\ell=1}^{p_2} \lambda_{2\ell} W_{2\ell} y_{t-1} + x_t\beta + \Gamma f_t + u_t, \\ u_t &= \textstyle\sum_{\ell=1}^{p_3} \lambda_{3\ell} W_{3\ell} u_t + v_t, \quad t = 1, 2, \ldots, T. \end{aligned} \tag{6.9}$$

Our AQS estimation and OPMD-type inference processes can be extended by letting $\lambda_r = (\lambda_{r1}, \ldots, \lambda_{r,p_r})', r = 1, 2, 3,\ B_1(\lambda_1) = I_n - \sum_{\ell=1}^{p_1} \lambda_{1\ell} W_{1\ell},\ B_2(\rho_1, \lambda_2) = \rho_1 I_n + \sum_{\ell=1}^{p_2} \lambda_{2\ell} W_{2\ell}$, and $B_3(\lambda_3) = I_n - \sum_{\ell=1}^{p_3} \lambda_{3\ell} W_{3\ell}$. It would also be interesting to extend our methods to allow for endogenous spatial weights as in Kuersteiner and Prucha (2020) in future works.

A very interesting and important extension to consider in future research is to 'unify' the small-T and large-T estimation and inference for the SDPF-IFE models. It is clear that under large T, our joint EF or AQS function for ψ and ϕ given in (2.13) remain valid, except that the ϕ parameters become incidental. Estimation and inference concern ψ. Thus, if the impact (if any) of estimating ϕ on the estimation of ψ can be 'removed' to give an asymptotically consistent marginal AQS function of ψ, the resulting estimation method would be valid whether T is small or grows with n. To extend our inference method to cater the case of large T, the idea of decomposition can still be followed, but it would be necessary to decompose the marginal AQS function of ψ into nT terms instead of n terms. The averaged outer products of these terms together with a covariance calculation may still provide a simple and consistent estimator of the VC matrix of the AQS estimator of ψ.

Appendix A: Some Basic Lemmas

Lemma A.1. *(Kelejian and Prucha, 1999; Lee, 2002): Let $\{A_n\}$ and $\{B_n\}$ be two sequences of $n \times n$ matrices that are uniformly bounded in both row and column sums. Let C_n be a sequence of conformable matrices whose elements are uniformly $O(h_n^{-1})$. Then*

(i) the sequence $\{A_n B_n\}$ are uniformly bounded in both row and column sums,

(ii) the elements of A_n are uniformly bounded and $tr(A_n) = O(n)$, and

(iii) the elements of $A_n C_n$ and $C_n A_n$ are uniformly $O(h_n^{-1})$.

Lemma A.2. *(Lee, 2004, p.1918): For W_1 and B_1 defined in Model (2.1), if $\|W_1\|$ and $\|B_{10}^{-1}\|$ are uniformly bounded, where $\|\cdot\|$ is a matrix norm, then $\|B_1^{-1}\|$ is uniformly bounded in a neighbourhood of λ_{10}.*

Lemma A.3. *(Lee, 2004, p.1918): Let X_n be an $n \times p$ matrix. If the elements X_n are uniformly bounded and $\lim_{n\to\infty} \frac{1}{n} X_n' X_n$ exists and is nonsingular, then $P_n = X_n(X_n'X_n)^{-1}X_n'$ and $M_n = I_n - P_n$ are uniformly bounded in both row and column sums.*

Lemma A.4. *(Lemma A.4, Yang, 2018): Let $\{A_n\}$ be a sequence of $n \times n$ matrices that are uniformly bounded in either row or column sums. Suppose that the elements $a_{n,ij}$ of A_n are $O(h_n^{-1})$ uniformly in all i and j. Let v_n be a random n-vector of iid elements with mean zero, variance σ^2 and finite 4th moment, and b_n a constant n-vector of elements of uniform order $O(h_n^{-1/2})$. Then*

(i) $\mathrm{E}(v_n' A_n v_n) = O(\frac{n}{h_n})$,	*(ii)* $\mathrm{Var}(v_n' A_n v_n) = O(\frac{n}{h_n})$,
(iii) $\mathrm{Var}(v_n' A_n v_n + b_n' v_n) = O(\frac{n}{h_n})$,	*(iv)* $v_n' A_n v_n = O_p(\frac{n}{h_n})$,
(v) $v_n' A_n v_n - \mathrm{E}(v_n' A_n v_n) = O_p((\frac{n}{h_n})^{\frac{1}{2}})$,	*(vi)* $v_n' A_n b_n = O_p((\frac{n}{h_n})^{\frac{1}{2}})$,

and (vii), the results (iii) and (vi) remain valid if b_n is a random n-vector independent of v_n such that $\{\mathrm{E}(b_{ni}^2)\}$ are of uniform order $O(h_n^{-1})$.

Lemma A.5. *(Lemma A.5, Yang, 2018): Let $\{\Phi_n\}$ be a sequence of $n \times n$ matrices with row and column sums uniformly bounded, and elements of uniform order $O(h_n^{-1})$. Let $v_n = (v_1, \cdots, v_n)'$ be a random vector of iid elements with mean zero, variance σ_v^2, and finite $(4 + 2\epsilon_0)$th moment for some $\epsilon_0 > 0$. Let $b_n = \{b_{ni}\}$ be an $n \times 1$ random vector, independent of v_n, such that (i) $\{\mathrm{E}(b_{ni}^2)\}$ are of uniform order $O(h_n^{-1})$, (ii) $sup_i E|b_{ni}|^{2+\epsilon_0} < \infty$, (iii) $\frac{h_n}{n}\sum_{i=1}^n [\phi_{n,ii}(b_{ni} - \mathrm{E}b_{ni})] = o_p(1)$ where $\{\phi_{n,ii}\}$ are the diagonal elements of Φ_n, and (iv)*

$\frac{h_n}{n}\sum_{i=1}^n[b_{ni}^2 - \mathrm{E}(b_{ni}^2)] = o_p(1)$. *Define the bilinear-quadratic form:* $Q_n = b_n'v_n + v_n'\Phi_n v_n - \sigma_v^2 tr(\Phi_n)$, *and let* $\sigma_{Q_n}^2$ *be the variance of* Q_n. *If* $lim_{n\to\infty}h_n^{1+2/\epsilon_0}/n = 0$ *and* $\{\frac{h_n}{n}\sigma_{Q_n}^2\}$ *are bounded away from zero, then* $Q_n/\sigma_{Q_n} \xrightarrow{d} N(0,1)$.

Appendix B: Proofs for Section 3 and 4

To simplify the notation, a parametric quantity (scalar, vector or matrix) evaluated at the general values of the parameters is denoted by dropping its arguments, e.g., $B_1 \equiv B_1(\lambda_1), \mathbf{B}_1 \equiv \mathbf{B}_1(\lambda_1)$, and $\Omega(\lambda_3) \equiv \Omega$. In proving the theorems, the following matrix results are used: (i) eigenvalues of a projection matrix are either 0 or 1; (ii) eigenvalues of a positive definite matrix are strictly positive; (iii) for symmetric matrix A and positive semidefinite (p.s.d.) matrix B, $\gamma_{\min}(A)\mathrm{tr}(B) \le \mathrm{tr}(AB) \le \gamma_{\max}(A)\mathrm{tr}(B)$; ($iv$) for symmetric matrices A and B, $\gamma_{\max}(A+B) \le \gamma_{\max}(A) + \gamma_{\max}(B)$; and ($v$) for p.s.d. matrices A and B, $\gamma_{\max}(AB) \le \gamma_{\max}(A)\gamma_{\max}(B)$. See, e.g, Bernstein (2009).

Proof of Theorem 3.1: Under Assumption G, by Theorem 5.9 of van der Vaart (1998) the consistency of $\hat{\delta}$ follows if $\sup_{\delta\in\boldsymbol{\Delta}}\frac{1}{nT}\|S_{nT}^{*c}(\delta) - \bar{S}_{nT}^{*c}(\delta)\| \xrightarrow{p} 0$ as $n \to \infty$, where $S_{nT}^{*c}(\delta)$ is the concentrated AQS function for δ and $\bar{S}_{nT}^{*c}(\delta)$ is its population counterpart. Both quantities are defined above Theorem 3.1 and their exact expressions are given below:

$$S_{nT}^{*c}(\delta) = \begin{cases} \frac{1}{\hat{\sigma}_v^2(\delta)}\hat{Z}'(\delta)\mathbf{M}_F\boldsymbol{\Omega}^{-1}Y_{-1} - \mathrm{tr}(\mathbf{M}_F\mathbf{D}_{-1}), \\ \frac{1}{\hat{\sigma}_v^2(\delta)}\hat{Z}(\delta)'\mathbf{M}_F\boldsymbol{\Omega}^{-1}\mathbf{W}_1Y - \mathrm{tr}(\mathbf{M}_F\mathbf{W}_1\mathbf{D}), \\ \frac{1}{\hat{\sigma}_v^2(\delta)}\hat{Z}'(\delta)\mathbf{M}_F\boldsymbol{\Omega}^{-1}\mathbf{W}_2Y_{-1} - \mathrm{tr}(\mathbf{M}_F\mathbf{W}_2\mathbf{D}_{-1}), \\ \frac{1}{\hat{\sigma}_v^2(\delta)}\hat{Z}'(\delta)\mathbf{M}_F\mathbf{B}_3'\mathbf{W}_3\hat{Z}(\delta) - (T-r)\mathrm{tr}(B_3^{-1}W_3), \\ \frac{1}{\hat{\sigma}_v^2(\delta)}\hat{Z}'(\delta)[M_F\dot{F}_s(F'F)^{-1}F' \otimes \Omega^{-1}]\hat{Z}(\delta),\ s = 1,\dots,k_\phi. \end{cases} \tag{B.1}$$

$$\bar{S}_{nT}^{*c}(\delta) = \begin{cases} \frac{1}{\bar{\sigma}_v^2(\delta)}\mathrm{E}[\bar{Z}'(\delta)\mathbf{M}_F\boldsymbol{\Omega}^{-1}Y_{-1}] - \mathrm{tr}(\mathbf{M}_F\mathbf{D}_{-1}), \\ \frac{1}{\bar{\sigma}_v^2(\delta)}\mathrm{E}[\bar{Z}'(\delta)\mathbf{M}_F\boldsymbol{\Omega}^{-1}\mathbf{W}_1Y] - \mathrm{tr}(\mathbf{M}_F\mathbf{W}_1\mathbf{D}), \\ \frac{1}{\bar{\sigma}_v^2(\delta)}\mathrm{E}[\bar{Z}'(\delta)\mathbf{M}_F\boldsymbol{\Omega}^{-1}\mathbf{W}_2Y_{-1}] - \mathrm{tr}(\mathbf{M}_F\mathbf{W}_2\mathbf{D}_{-1}), \\ \frac{1}{\bar{\sigma}_v^2(\delta)}\mathrm{E}[\bar{Z}'(\delta)\mathbf{M}_F\mathbf{B}_3'\mathbf{W}_3\bar{Z}(\delta)] - (T-r)\mathrm{tr}(B_3^{-1}W_3), \\ \frac{1}{\bar{\sigma}_v^2(\delta)}\mathrm{E}\{\bar{Z}'(\delta)[M_F\dot{F}_s(F'F)^{-1}F' \otimes \Omega^{-1}]\bar{Z}(\delta)\},\ s = 1,\dots,k_\phi, \end{cases} \tag{B.2}$$

where $\bar{\sigma}_v^2(\delta) = \frac{1}{n(T-r)}\mathrm{E}\left[\bar{Z}(\delta)'\mathbf{M}_F\mathbf{\Omega}^{-1}\bar{Z}(\delta)\right]$, $\bar{Z}(\delta) = Z(\theta)|_{\beta=\bar{\beta}(\delta)} = \mathbf{B}_1Y - \mathbf{B}_2Y_{-1} - \mathbf{X}\bar{\beta}(\delta)$, and $\bar{\beta}(\delta) = (\mathbf{X}'\mathbf{M}_F\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{M}_F\mathbf{\Omega}^{-1}(\mathbf{B}_1\mathrm{EY} - \mathbf{B}_2\mathrm{EY}_{-1})$. With (B.1) and (B.2), the proof of consistency of $\hat{\delta}$ boils down to the proofs of the following:

(a) $\inf_{\delta\in\mathbf{\Delta}} \bar{\sigma}_v^2(\delta)$ is bounded away from zero,

(b) $\sup_{\delta\in\mathbf{\Delta}} \left|\hat{\sigma}_v^2(\delta) - \bar{\sigma}_v^2(\delta)\right| = o_p(1)$,

(c) $\sup_{\delta\in\mathbf{\Delta}} \frac{1}{nT}\left|\hat{Z}'(\delta)\mathbf{M}_F\mathbf{\Omega}^{-1}Y_{-1} - \mathrm{E}[\bar{Z}'(\delta)\mathbf{M}_F\mathbf{\Omega}^{-1}Y_{-1}]\right| = o_p(1)$,

(d) $\sup_{\delta\in\mathbf{\Delta}} \frac{1}{nT}\left|\hat{Z}'(\delta)\mathbf{M}_F\mathbf{\Omega}^{-1}\mathbf{W}_1Y - \mathrm{E}[\bar{Z}'(\delta)\mathbf{M}_F\mathbf{\Omega}^{-1}\mathbf{W}_1Y]\right| = o_p(1)$,

(e) $\sup_{\delta\in\mathbf{\Delta}} \frac{1}{nT}\left|\hat{Z}'(\delta)\mathbf{M}_F\mathbf{\Omega}^{-1}\mathbf{W}_2Y_{-1} - \mathrm{E}[\bar{Z}'(\delta)\mathbf{M}_F\mathbf{\Omega}^{-1}\mathbf{W}_2Y_{-1}]\right| = o_p(1)$,

(f) $\sup_{\delta\in\mathbf{\Delta}} \frac{1}{nT}\left|\hat{Z}'(\delta)\mathbf{M}_F\mathbf{B}_3'\mathbf{W}_3\hat{Z}(\delta) - \mathrm{E}[\bar{Z}'(\delta)\mathbf{M}_F\mathbf{B}_3'\mathbf{W}_3\bar{Z}(\delta)]\right| = o_p(1)$,

(g) $\sup_{\delta\in\mathbf{\Delta}} \frac{1}{nT}\left|\hat{Z}'(\delta)[M_F\dot{F}_s(F'F)^{-1}F'\otimes\Omega^{-1}]\hat{Z}(\delta) - \mathrm{E}\{\bar{Z}'(\delta)[M_F\dot{F}_s(F'F)^{-1}F'\otimes\Omega^{-1}]\bar{Z}(\delta)\}\right| = o_p(1)$, $s = 1, \ldots, k_\phi$.

Denote $\mathbf{A} = \mathbf{M}_F\mathbf{\Omega}^{-1} = M_F \otimes (B_3'B_3)$, and let $\mathbf{A}^{\frac{1}{2}}$ be a square-root matrix of $\mathbf{A}$. Define $\bar{Z}^\dagger(\delta) = \mathbf{A}^{\frac{1}{2}}\bar{Z}(\delta)$, $\hat{Z}^\dagger(\delta) = \mathbf{A}^{\frac{1}{2}}\hat{Z}(\delta)$, and $\mathbf{B}_r^\dagger = \mathbf{A}^{\frac{1}{2}}\mathbf{B}_r, r = 1, 2$. Let $Y^\circ = Y - \mathrm{E}(Y)$ and $Y_{-1}^\circ = Y_{-1} - \mathrm{E}(Y_{-1})$. Further define the projection matrices: $\mathbf{M} = I_{nT} - \mathbf{A}^{\frac{1}{2}}\mathbf{X}(\mathbf{X}'\mathbf{A}\mathbf{X})^{-1}\mathbf{X}'\mathbf{A}^{\frac{1}{2}}$ and $\mathbf{P} = I_{nT} - \mathbf{M}$. Then, we can write:

$$\bar{Z}^\dagger(\delta) = \mathbf{M}(\mathbf{B}_1^\dagger Y - \mathbf{B}_2^* Y_{-1}) + \mathbf{P}(\mathbf{B}_1^* Y^\circ - \mathbf{B}_2^\dagger Y_{-1}^\circ), \tag{B.3}$$

$$\hat{Z}^\dagger(\delta) = \mathbf{M}(\mathbf{B}_1^\dagger Y - \mathbf{B}_2^\dagger Y_{-1}). \tag{B.4}$$

Proof of (a). Using the expression (B.3) and by the orthogonality between $\mathbf{M}$ and $\mathbf{P}$, we can write $\bar{\sigma}_v^2(\delta) = \frac{1}{n(T-r)}\mathrm{E}[\bar{Z}^{\dagger\prime}(\delta)\bar{Z}^\dagger(\delta)]$ as follows:

$$\bar{\sigma}_v^2(\delta) = \tfrac{1}{n(T-r)}\mathrm{tr}[\mathrm{Var}(\mathbf{B}_1^\dagger Y - \mathbf{B}_2^\dagger Y_{-1})] + \tfrac{1}{n(T-r)}(\mathbf{B}_1^\dagger\mathrm{E}Y - \mathbf{B}_2^\dagger\mathrm{E}Y_{-1})'\mathbf{M}(\mathbf{B}_1^\dagger\mathrm{E}Y - \mathbf{B}_2^\dagger\mathrm{E}Y_{-1}).$$

By Assumption E(iv) and the assumptions given in the theorem, we have for the first term, $\inf_{\delta\in\mathbf{\Delta}} \frac{1}{n(T-r)}\mathrm{tr}[\mathbf{A}\mathrm{Var}(\mathbf{B}_1Y - \mathbf{B}_2Y_{-1})] \geq \frac{1}{n(T-r)}\inf_{\delta\in\mathbf{\Delta}}\gamma_{\min}[\mathrm{Var}(\mathbf{B}_1Y - \mathbf{B}_2Y_{-1})]\mathrm{tr}(M_F \otimes B_3'B_3) \geq \frac{1}{n}\underline{c}_y\inf_{\lambda_3\in\Lambda_3}\mathrm{tr}(B_3'B_3) \geq \frac{1}{n}\underline{c}_y n[\inf_{\lambda_3\in\Lambda_3}\gamma_{\min}(B_3'B_3)] \geq \underline{c}_y\underline{c}_3 > 0$. The second term is non-negative uniformly in $\delta \in \mathbf{\Delta}$ as $\mathbf{M}$ is positive semi-definite (p.s.d). It follows that $\inf_{\delta\in\mathbf{\Delta}}\bar{\sigma}_v^2(\delta) > c > 0$, and result (a) is proved.

Proof of (b). Using (B.3) and (B.4), we can decompose $\hat{\sigma}_v^2(\delta) - \bar{\sigma}_v^2(\delta)$ into four terms

$$\hat{\sigma}_v^2(\delta) - \bar{\sigma}_v^2(\delta) = (Q_1 - \mathrm{E}Q_1) + (Q_2 - \mathrm{E}Q_2) - 2(Q_3 - \mathrm{E}Q_3) - \mathrm{E}Q_4. \tag{B.5}$$

where $Q_1 = \frac{1}{n(T-r)}Y'\mathbf{B}_1^{\dagger\prime}\mathbf{M}\mathbf{B}_1^\dagger Y$, $Q_2 = \frac{1}{n(T-r)}Y_{-1}'\mathbf{B}_2^{\dagger\prime}\mathbf{M}\mathbf{B}_2^\dagger Y_{-1}$, $Q_3 = \frac{2}{n(T-r)}Y'\mathbf{B}_1^{\dagger\prime}\mathbf{M}\mathbf{B}_2^\dagger Y_{-1}$

and $Q_4 = \frac{1}{n(T-r)}(\mathbf{B}_1^{\dagger} Y^{\circ} - \mathbf{B}_2^{\dagger} Y_{-1}^{\circ})'\mathbf{P}(\mathbf{B}_1^{\dagger} Y^{\circ} - \mathbf{B}_2^{\dagger} Y_{-1}^{\circ})$. The result in **(b)** follows if $Q_j - \mathrm{E}Q_j \xrightarrow{p} 0,\ j = 1, 2, 3$, and $\mathrm{E}Q_4 \to 0$, uniformly in $\delta \in \mathbf{\Delta}$.

Recall from (2.9): $Y = \mathbf{Q}\mathbf{y}_0 + \boldsymbol{\eta} + \mathbf{D}Z$ and $Y_{-1} = \mathbf{Q}_{-1}\mathbf{y}_0 + \boldsymbol{\eta}_{-1} + \mathbf{D}_{-1}Z$. By $\mathbf{B}_{30}Z = \mathbf{v} + \mathtt{vec}(B_{30}\Gamma_0 F_0')$, we can further write $Y = \mathbf{Q}\mathbf{y}_0 + \boldsymbol{\eta}^* + \mathbf{D}\mathbf{B}_{30}^{-1}\mathbf{v}$, and $Y_{-1} = \mathbf{Q}_{-1}\mathbf{y}_0 + \boldsymbol{\eta}_{-1}^* + \mathbf{D}_{-1}\mathbf{B}_{30}^{-1}\mathbf{v}$, where $\boldsymbol{\eta}^* = \boldsymbol{\eta} + \mathbf{D}\mathtt{vec}(\Gamma_0 F_0')$ and $\boldsymbol{\eta}_{-1}^* = \boldsymbol{\eta}_{-1} + \mathbf{D}_{-1}\mathtt{vec}(\Gamma_0 F_0')$. Using these expressions and letting $\mathbf{M}^{\dagger} = \mathbf{A}^{\frac{1}{2}}\mathbf{M}\mathbf{A}^{\frac{1}{2}}$, we can write

$$
\begin{aligned}
Q_1 &= \textstyle\sum_{\ell=1}^{5} Q_{1,\ell} + \frac{1}{n(T-r)}\boldsymbol{\eta}^{*\prime}\mathbf{B}_1'\mathbf{M}^{\dagger}\mathbf{B}_1\boldsymbol{\eta}^*, \\
Q_2 &= \textstyle\sum_{\ell=1}^{5} Q_{2,\ell} + \frac{1}{n(T-r)}\boldsymbol{\eta}_{-1}^{*\prime}\mathbf{B}_2'\mathbf{M}^{\dagger}\mathbf{B}_2\boldsymbol{\eta}_{-1}^*, \\
Q_3 &= \textstyle\sum_{\ell=1}^{8} Q_{3,\ell} + \frac{2}{n(T-r)}\boldsymbol{\eta}^{*\prime}\mathbf{B}_1'\mathbf{M}^{\dagger}\mathbf{B}_2\boldsymbol{\eta}_{-1}^*,
\end{aligned}
$$

where $Q_{k\ell}$ takes one of the forms: $\frac{1}{n(T-r)}\mathbf{y}_0'\mathbf{R}_1\mathbf{y}_0$, $\frac{1}{n(T-r)}\mathbf{v}'\mathbf{R}_2\mathbf{v}$, $\frac{1}{n(T-r)}\mathbf{y}_0'\mathbf{R}_3\mathbf{v}$, $\frac{1}{n(T-r)}\mathbf{y}_0'\mathbf{R}_4$, and $\frac{1}{n(T-r)}\mathbf{v}'\mathbf{R}_5$. $\mathbf{R}_1, \mathbf{R}_2$, and $\mathbf{R}_3$ are $nT \times nT$ matrices while $\mathbf{R}_4$ and $\mathbf{R}_5$ are $nT \times 1$ vectors. These parametric quantities $\mathbf{R}_s, s = 1, \ldots, 5$ depend on δ through $\mathbf{B}_1$, $\mathbf{B}_2$ and $\mathbf{M}^*$, and involve $\mathbf{Q}$, $\mathbf{Q}_{-1}$, $\mathbf{D}$, $\mathbf{D}_{-1}$, $\boldsymbol{\eta}^*$ and $\boldsymbol{\eta}_{-1}^*$, which are all matrix or vector functions of true parameters.

By Assumptions D, E and Lemma A.1, the $nT \times nT$ matrices $\mathbf{Q}$, $\mathbf{Q}_{-1}$, $\mathbf{D}$, and $\mathbf{D}_{-1}$ are uniformly bounded in both row and column sums, and the elements of the $nT \times 1$ vectors $\boldsymbol{\eta}^*$ and $\boldsymbol{\eta}_{-1}^*$ are uniformly bounded. By Assumptions D, E(iii) and Lemmas A.1 and A.3, $\mathbf{B}_1$, $\mathbf{B}_2$ and $\mathbf{M}^*$ are uniformly bounded in both row and column sums. Therefore, by Lemma A.1(i) matrices $\mathbf{R}_{\ell}, \ell = 1, 2, 3$ are uniformly bounded in both row and column sums and by Lemma A.1(iii) elements of vectors $\mathbf{R}_4$ and $\mathbf{R}_4$ are uniformly bounded. Hence, by Assumption F, we immediately have the results that $\frac{1}{nT}[\mathbf{y}_0'\mathbf{R}_1\mathbf{y}_0 - \mathrm{E}(\mathbf{y}_0'\mathbf{R}_1\mathbf{y}_0)] = o_p(1)$, and $\frac{1}{nT}[\mathbf{y}_0'\mathbf{R}_4 - \mathrm{E}(\mathbf{y}_0')\mathbf{R}_4] = o_p(1)$. The point wise convergence of the quadratic terms $\frac{1}{nT}\mathbf{v}'\mathbf{R}_2\mathbf{v}$, and the bilinear term $\frac{1}{nT}\mathbf{y}_0'\mathbf{R}_3\mathbf{v}$, can be established by Assumptions B, E and results (v) and (vi) in Lemma A.4. The point wise convergence of the linear terms $\frac{1}{nT}\mathbf{v}'\mathbf{R}_5$ can be easily proved using Chebyshev's inequality. Therefore, for $k = 1, 2, 3$, and all ℓ,

$$
Q_{k,\ell}(\delta) - \mathrm{E}Q_{k,\ell}(\delta) \xrightarrow{p} 0, \quad \text{for each } \ \delta \in \mathbf{\Delta}.
$$

Now, all the $Q_{k,\ell}(\delta)$ terms are linear or quadratic in ρ, λ_1 and λ_2, and it is easy to show that $\sup_{\delta \in \mathbf{\Delta}} |\frac{\partial}{\partial \omega} Q_{k,\ell}(\delta)| = O_p(1)$, for $\omega = \rho, \lambda_1, \lambda_2$. For λ_3 and ϕ, they only enter $Q_{k,\ell}(\delta)$ through $\mathbf{A}$ in matrix $\mathbf{M}^{\dagger}$. For $\omega = \lambda_3, \phi_s, s = 1, \ldots, k_{\phi}$, some algebra leads to the following expression $\frac{d}{d\omega}\mathbf{M}^{\dagger} = \mathbf{G}'\dot{\mathbf{A}}_{\omega}\mathbf{G}$, where $\mathbf{G} = I_{nT} - \mathbf{X}(\mathbf{X}'\mathbf{A}\mathbf{X})^{-1}\mathbf{X}'\mathbf{A}$, $\dot{\mathbf{A}}_{\lambda_3} = \frac{\partial}{\partial \lambda_3}\mathbf{A} = M_F \otimes (B_3'W_3 + W_3'B_3)$, and $\dot{\mathbf{A}}_{\phi_s} = \frac{\partial}{\partial \phi_s}\mathbf{A} = -\dot{P}_{F,s} \otimes (B_3'B_3)$. By Assumption E$(iv)$, we have $\sup_{\delta\in\mathbf{\Delta}}\gamma_{\max}(\dot{\mathbf{A}}_{\lambda_3}) = \sup_{\delta\in\mathbf{\Delta}}\gamma_{\max}(B_3'W_3 + W_3'B_3) < c$. Moreover, $\sup_{\delta\in\mathbf{\Delta}}\gamma_{\max}(\mathbf{G}) = \sup_{\delta\in\mathbf{\Delta}}\gamma_{\max}(\mathbf{X}(\mathbf{X}'\mathbf{A}\mathbf{X})^{-1}\mathbf{X}'\mathbf{A}) = \sup_{\delta\in\mathbf{\Delta}}\gamma_{\max}(\mathbf{A}^{\frac{1}{2}}\mathbf{X}(\mathbf{X}'\mathbf{A}\mathbf{X})^{-1}\mathbf{X}'\mathbf{A}^{\frac{1}{2}}) = 1$. By applying Lemmas A.1, A.4, and Assumption F repeatedly, we can show that, for $k = 1, 2, 3$, and all ℓ, $\sup_{\delta\in\mathbf{\Delta}}|\frac{\partial}{\partial\lambda_3}Q_{k,\ell}(\delta)| = O_p(1)$. For example, for $|\frac{\partial}{\partial\lambda_3}Q_{1,1}(\delta)|$,

$$\begin{aligned}\sup_{\delta\in\mathbf{\Delta}}\big|\tfrac{\partial}{\partial\lambda_3}Q_{1,1}(\delta)\big| &= \sup_{\delta\in\mathbf{\Delta}}\big|\tfrac{1}{n(T-r)}\tfrac{\partial}{\partial\lambda_3}\mathbf{y}_0'\mathbf{Q}'\mathbf{B}_1'\mathbf{M}^{\dagger}\mathbf{B}_1\mathbf{Q}\mathbf{y}_0'\big| \\ &\leqslant \sup_{\delta\in\mathbf{\Delta}}\gamma_{\max}(\dot{\mathbf{A}}_{\lambda_3})\gamma_{\max}(\mathbf{G}'\mathbf{G})\gamma_{\max}(\mathbf{B}_1'\mathbf{B}_1)\tfrac{1}{n(T-r)}\big|\mathbf{y}_0'\mathbf{Q}'\mathbf{Q}\mathbf{y}_0'\big| = O_p(1).\end{aligned}$$

Recall $\dot{P}_{F,s} = M_F\dot{F}_s(F'F)^{-1}F' + F(F'F)^{-1}\dot{F}_s'M_F$, by Assumptions C and E(iv), it is easy to see that $\gamma_{\max}(\dot{\mathbf{A}}_{\phi_s})$ is uniformly bounded. Therefore by Lemmas A.1, A.4, and Assumption F, we have for $k = 1, 2, 3$, and all ℓ, $\sup_{\delta\in\mathbf{\Delta}}|\frac{\partial}{\partial\phi_s}Q_{k,\ell}(\delta)| = O_p(1)$, $s = 1, 2, \ldots, k_\phi$. It follows that $Q_{k,\ell}(\delta)$ are stochastically equicontinuous. By Theorem 2.1 of Newey (1991), the pointwise convergence and stochastic equicontinuity therefore lead to,

$$Q_{k,\ell}(\delta) - \mathrm{E}Q_{k,\ell}(\delta) \xrightarrow{p} 0, \quad \text{uniformly in } \delta \in \mathbf{\Delta}.$$

It left to show that $\mathrm{E}Q_4(\delta) = \frac{1}{n(T-r)}\mathrm{E}[(\mathbf{B}_1^*Y^\circ - \mathbf{B}_2^*Y_{-1}^\circ)'\mathbf{P}(\mathbf{B}_1^*Y^\circ - \mathbf{B}_2^*Y_{-1}^\circ)] \to 0$, uniformly in $\delta \in \mathbf{\Delta}$. By Assumption D, $\gamma_{\min}\big(\frac{\mathbf{X}'\mathbf{A}\mathbf{X}}{nT}\big) > \underline{c}_x$. Therefore, we have by the assumptions in Theorem 3.1 and Assumption D,

$$\begin{aligned}\mathrm{E}Q_4 &= \tfrac{1}{n(T-r)}\mathrm{tr}[\mathbf{A}\mathbf{X}(\mathbf{X}'\mathbf{A}\mathbf{X})^{-1}\mathbf{X}'\mathbf{A}\mathrm{Var}(\mathbf{B}_1Y - \mathbf{B}_2Y_{-1})] \\ &\leq \tfrac{1}{n(T-r)}\gamma_{\max}^2(\mathbf{A})\gamma_{\min}^{-1}\big(\tfrac{\mathbf{X}'\mathbf{A}\mathbf{X}}{nT}\big)\bar{c}_y\,\tfrac{1}{nT}\mathrm{tr}(\mathbf{X}'\mathbf{X}) = O(n^{-1}).\end{aligned}$$

Hence, $\hat{\sigma}_v^2(\delta) - \bar{\sigma}_v^2(\delta) \xrightarrow{p} 0$, uniformly in $\delta \in \mathbf{\Delta}$, completing the proof of **(b)**.

Proofs of (c)-(g). Using the expressions (B.3) and (B.4) and the representations of Y and Y_{-1} in (2.9), all the quantities inside $|\cdot|$ in **(c)-(g)** can all be expressed in the forms similar to (B.5). Thus, the proofs of **(c)-(g)** follow the proof of **(b)**. ■

Proof of Theorem 3.2: By applying the mean value theorem (henceforth MVT) to each element of $S_{nT}^*(\hat{\boldsymbol{\psi}})$, we have,

$$\tfrac{1}{nT}S_{nT}^*(\hat{\boldsymbol{\psi}}) = \tfrac{1}{nT}S_{nT}^*(\boldsymbol{\psi}_0) + \big[\tfrac{1}{nT}\tfrac{\partial}{\partial\boldsymbol{\psi}'}S_{nT}^*(\boldsymbol{\psi})|_{\boldsymbol{\psi}=\bar{\boldsymbol{\psi}}_r \text{ in } r\text{th row}}\big](\hat{\boldsymbol{\psi}}_{\mathtt{M}} - \boldsymbol{\psi}_0) = 0, \tag{B.6}$$

where $\{\bar{\boldsymbol{\psi}}_r\}$ are between $\hat{\boldsymbol{\psi}}$ and $\boldsymbol{\psi}_0$ elementwise. The result of the theorem follows if

(a) $\frac{1}{\sqrt{nT}}S_{nT}^*(\boldsymbol{\psi}_0) \xrightarrow{D} N\big[0, \lim_{n\to\infty}\Sigma_{nT}(\boldsymbol{\psi}_0)\big]$,

(b) $\frac{1}{nT}\big[\frac{\partial}{\partial\boldsymbol{\psi}'}S_{nT}^*(\boldsymbol{\psi}|_{\boldsymbol{\psi}=\bar{\boldsymbol{\psi}}_r \text{ in } r\text{th row}}) - \frac{\partial}{\partial\boldsymbol{\psi}'}S_{nT}^*(\boldsymbol{\psi}_0)\big] \xrightarrow{p} 0$, and

(c) $\frac{1}{nT}\big[\frac{\partial}{\partial\boldsymbol{\psi}'}S_{nT}^*(\boldsymbol{\psi}_0) - \mathrm{E}\big(\frac{\partial}{\partial\boldsymbol{\psi}'}S_{nT}^*(\boldsymbol{\psi}_0)\big)\big] \xrightarrow{p} 0$.

Proof of (a). In (3.2), we write the AQS vector as linear combinations of terms linear or quadratic in Z^* and bilinear in Z^* and $\mathbf{y}_0$. Using $Z^* = \mathbf{v} + \mathtt{vec}(B_{30}\Gamma_0 F_0')$, and the matrix multiplication result $\mathtt{vec}(B_{30}\Gamma_0 F_0')'\mathbf{M}_{F0}K = 0$ for any $nT \times 1$ vector K, the AQS vector at the true parameters can be written as follows:

$$S^*_{nT}(\boldsymbol{\psi}_0) = \begin{cases} \Pi_1'\mathbf{v} \\ \mathbf{v}'\Phi_1\mathbf{v} - \mu_{\sigma^2_{v0}} \\ \mathbf{v}'\Psi_1\mathbf{y}_0 + \mathbf{v}'\Phi_2\mathbf{v} + \Pi_2^{*\prime}\mathbf{v} - \mu_{\rho_0} \\ \mathbf{v}'\Psi_2\mathbf{y}_0 + \mathbf{v}'\Phi_3\mathbf{v} + \Pi_3^{*\prime}\mathbf{v} - \mu_{\lambda_{10}} \\ \mathbf{v}'\Psi_3\mathbf{y}_0 + \mathbf{v}'\Phi_4\mathbf{v} + \Pi_4^{*\prime}\mathbf{v} - \mu_{\lambda_{20}} \\ \mathbf{v}'\Phi_5\mathbf{v} - \mu_{\lambda_{30}} \\ \mathbf{v}'\Phi_{5+s}\mathbf{v},\ s = 1, \ldots, k_\phi \end{cases} \tag{B.7}$$

where $\Pi_2^* = \frac{1}{\sigma^2_{v0}}(M_{F_0} \otimes B_{30})\boldsymbol{\eta}^*_{-1}$, $\Pi_3^* = \frac{1}{\sigma^2_{v0}}(M_{F_0} \otimes B_{30})\mathbf{W}_1\boldsymbol{\eta}^*$, $\Pi_4^* = \frac{1}{\sigma^2_{v0}}(M_{F_0} \otimes B_{30})\mathbf{W}_2\boldsymbol{\eta}^*_{-1}$.

By Assumptions C, E, and Lemma A.1, the $nT \times nT$ matrices Φ and Ψ are uniformly bounded in both row and column sums, and elements of vectors Π or Π^* are uniformly bounded. For every non-zero $(k + 5 + k_\phi) \times 1$ vector of constants ℓ, we can express,

$$\ell' S^*_{nT}(\boldsymbol{\psi}_0) = \textstyle\sum_{t=1}^T \sum_{s=1}^T v_t' A_{ts} v_s + \sum_{t=1}^T v_t' g(y_0) - \ell'\mu,$$

for suitably defined non-stochastic matrices A_{ts}, vector μ, and functions $g(y_0)$ that are linear in y_0, where $\mu = (0_k', \mu_{\sigma^2_v}, \mu_\rho, \mu_{\lambda_1}, \mu_{\lambda_2}, \mu_{\lambda_3}, 0_{k_\gamma}')'$. As $\{y_0, v_1, \ldots, v_T\}$ are independent, the asymptotic normality of $\frac{1}{\sqrt{nT}}\ell' S^*_{nT}(\boldsymbol{\psi}_0)$ follows from Lemma A.5. The Cramér-Wold devise leads to the joint asymptotic normality of $\frac{1}{\sqrt{nT}} S^*_{nT}(\boldsymbol{\psi}_0)$.

Proof of (b). Let the $nT \times 1$ vector X_p, $p = 1, \cdots, k$, be the pth column of $\mathbf{X}$. Denote $nT \times 1$ vectors, $X_{k+1} = Y_{-1}$, $X_{k+2} = \mathbf{W}_1 Y$, $X_{k+3} = \mathbf{W}_2 Y_{-1}$. Further, denote $\beta_{k+1} = \rho$, $\beta_{k+2} = \lambda_1$, and $\beta_{k+3} = \lambda_2$. The Hessian matrix, $H(\boldsymbol{\psi}) = \frac{\partial}{\partial \boldsymbol{\psi}'} S^*_{nT}(\boldsymbol{\psi})$, has the elements:

$$\begin{aligned}
&H_{\beta_p\beta_q} = -\tfrac{1}{\sigma_v^2} X_p'(M_F \otimes \Omega^{-1})X_q - \dot{\mu}_{\beta_p,\beta_q}, && H_{\beta_p\lambda_3} = -\tfrac{1}{\sigma_v^2} X_p'[M_F \otimes (W_3'B_3 + B_3'W_3)]Z(\theta) \\
&H_{\beta_p\sigma_v^2} = -\tfrac{1}{\sigma_v^4} X_p'(M_F \otimes \Omega^{-1})Z(\theta), && H_{\sigma_v^2\sigma_v^2} = -\tfrac{1}{\sigma_v^6} Z'(\theta)(M_F \otimes \Omega^{-1})Z(\theta) + \tfrac{n(T-r)}{2\sigma_v^4} \\
&H_{\sigma_v^2\lambda_3} = -\tfrac{1}{\sigma_v^4} Z'(\theta)(M_F \otimes W_3'B_3)Z(\theta), && H_{\sigma_v^2\beta_p} = H_{\beta_p\sigma_v^2}, H_{\lambda_3\beta_p} = H_{\beta_p\lambda_3}, H_{\lambda_3\sigma_v^2} = H_{\sigma_v^2\lambda_3} \\
&H_{\lambda_3\lambda_3} = -\tfrac{1}{\sigma_v^2} Z'(\theta)(M_F \otimes W_3'W_3)Z(\theta) - (T-r)\mathrm{tr}(B_3^{-1}W_3B_3^{-1}W_3) && \\
&H_{\beta_p\phi_s} = -\tfrac{1}{\sigma_v^2} X_p'(\dot{P}_{F,s} \otimes \Omega^{-1})Z'(\theta) - \dot{\mu}_{\beta_p,\phi_s} && H_{\sigma_v^2\phi_s} = -\tfrac{1}{2\sigma_v^4} Z'(\theta)(\dot{P}_{F,s} \otimes \Omega^{-1})Z(\theta) \\
&H_{\lambda_3\phi_s} = -\tfrac{1}{\sigma_v^2} Z'(\theta)(\dot{P}_{F,s} \otimes B_3'W_3)Z(\theta), && H_{\phi_s\beta_p} = -\tfrac{1}{\sigma_v^2} X_p'(\dot{P}_{F,s} \otimes \Omega^{-1})Z'(\theta) \\
&H_{\phi_s\sigma_v^2} = H_{\sigma_v^2\phi_s},\ H_{\phi_s\lambda_3} = H_{\lambda_3\phi_s}, && H_{\phi_s\phi_\ell} = -\tfrac{1}{\sigma_v^2} Z'(\theta)(\dot{A}_{s,\ell} \otimes \Omega^{-1})Z(\theta).
\end{aligned}$$

where $p, q = 1, \ldots, k+3$, $s, \ell = 1, \ldots, k_\phi$, $A_s = M_F \dot{F}_s (F'F)^{-1}F'$, $\dot{A}_{s,\ell} = \frac{\partial}{\partial \phi_\ell} A_s$, $\dot{\mu}_{\beta_p,\beta_q} = \frac{\partial}{\partial \beta_q}\mu_{\beta_p}$, and $\dot{\mu}_{\beta_p,\phi_s} = \frac{\partial}{\partial \phi_s}\mu_{\beta_p}$, where $\mu_{\beta_p} = 0$ for $p \le k$, and defined under (3.2) for $p > k$.

First, it is easy to show that $\frac{1}{nT}H(\bar{\boldsymbol{\psi}}) = O_p(1)$ by Lemmas A.1, A.4 and the model assumptions, where we use $H(\bar{\boldsymbol{\psi}})$ to denote $\frac{\partial}{\partial \boldsymbol{\psi}'} S^*_{nT}(\boldsymbol{\psi}|_{\boldsymbol{\psi}=\bar{\boldsymbol{\psi}}_r \text{ in } r_{th} \text{ row}})$ for notation simplicity. As $\sigma_v^{-r}, r = 2, 4, 6$, appear in $H(\boldsymbol{\psi})$ multiplicatively, we have $\frac{1}{nT}H(\bar{\boldsymbol{\psi}}) = \frac{1}{nT}H(\bar{\lambda}, \bar{\beta}, \bar{\rho}, \sigma_{v0}^2) + o_p(1)$ as $\bar{\sigma}_v^{-r} = \sigma_{v0}^{-r} + o_p(1)$. Consider the term $H_{\beta_p\beta_q}(\bar{\lambda}, \bar{\beta}, \bar{\rho}, \bar{\gamma}, \sigma_{v0}^2)$. By MVT we have,

$$\begin{aligned}
& X_p'[M_F(\bar{\phi}) \otimes \Omega^{-1}(\bar{\lambda}_3)]X_q \\
= \;& X_p'(M_{F0} \otimes \Omega_0^{-1})X_q + X_p'[M_F(\tilde{\phi}) \otimes (B_3'(\tilde{\lambda}_3)W_3 + W_3'B_3(\tilde{\lambda}_3))]X_q(\bar{\lambda}_3 - \lambda_{30}) \\
& - \textstyle\sum_{s=1}^{k_\phi} X_p'[\dot{P}_{F,s}(\tilde{\phi}) \otimes \Omega^{-1}(\tilde{\lambda}_3)](\bar{\phi}_s - \phi_{s0}),
\end{aligned}$$

where $(\tilde{\lambda}_3, \tilde{\phi}')$ is between $(\bar{\lambda}_3, \bar{\phi}')$ and (λ_{30}, ϕ_0'). By (2.9), Assumptions C, E, F, Lemmas A.1, A.4, and the consistency of $\hat{\boldsymbol{\psi}}$, $\frac{1}{nT}X_p'[M_F(\bar{\phi}) \otimes \Omega^{-1}(\bar{\lambda}_3)]X_q = \frac{1}{nT}X_p'(M_{F0} \otimes \Omega_0^{-1})X_q + o_p(1)$.

For the convergence of $\dot{\mu}_{\beta_p,\beta_q}$, consider $\mu_{\rho,\rho}(\bar{\boldsymbol{\psi}}) = \text{tr}[(\frac{\partial}{\partial\rho}\mathbf{D}_{-1}(\bar{\rho}, \bar{\lambda}))M_F(\bar{\phi})]$ for example. By the expression of $\mathbf{D}_{-1}$ in (2.9) it is easy to see that blocks of $\frac{\partial}{\partial\rho}\mathbf{D}_{-1}$ are products of matrices B_1^{-1}, B_2, and W_2, which are bounded in both row and column sums for (ρ, λ) in a neighborhood of (ρ_0, λ_0) by Lemma A.2 and Assumptions C and E. So, the derivatives of $\mu_{\rho,\rho}(\bar{\boldsymbol{\psi}})$ with respect to ρ , λ and ϕ are the traces of matrices that are products of M_F, B_1^{-1}, B_2, W_1, and W_2, and are bounded in both row and column sums by Lemma A.1, A.2 and Assumption C. Hence, by the MVT and consistency of $\hat{\boldsymbol{\psi}}_{\mathtt{M}}$ we have $\frac{1}{nT}\mu_{\rho,\rho}(\bar{\boldsymbol{\psi}}) = \frac{1}{nT}\mu_{\rho,\rho}(\boldsymbol{\psi}_0) + o_p(1)$. For $p, q = 1, \cdots k+3$, the convergence of $\dot{\mu}_{\beta_p,\beta_q}(\bar{\boldsymbol{\psi}})$ can be shown similarly. So we have established that $\frac{1}{nT}H_{\beta_p\beta_q}(\bar{\boldsymbol{\psi}}) = \frac{1}{nT}H_{\beta_p\beta_q}(\boldsymbol{\psi}) + o_p(1)$. Using $\bar{Z} = Z - \sum_{p=1}^{k+3} X_p(\bar{\beta}_p - \beta_{p0})$ and representations for Y and Y_{-1} given in (2.9), the convergence of other terms in $H(\boldsymbol{\psi})$ that involve $Z(\theta)$ can be shown similarly by repeatedly applying the MVT and Assumptions C, E, F, Lemmas A.1 and A.4, and the consistency of $\hat{\boldsymbol{\psi}}_{\mathtt{M}}$.

Proof of (c). By the representations given in (2.9), the elements of Hessian matrix can be written as linear combinations of quadratic and linear terms of $\mathbf{v}$, quadratic and linear terms of $\mathbf{y}_0$, bilinear terms of $\mathbf{v}$ and $\mathbf{y}_0$. Thus, the results follow by repeatedly applying Assumption F, Lemma A.1, and Lemma A.4. ∎

Proof of Theorem 4.1: The proof is given online with information in Appendix C. ∎

Appendix C: Supplementary Data

Supplementary material containing the proof of Theorem 4.1 and additional Monte Carlo results are given online at `http://www.mysmu.edu.sg/faculty/zlyang/SubPages/research.htm`

Acknowledgments

The bulk of the paper was written when Liyao Li was with Singapore Management University. Early versions of this paper were presented at Shanghai Workshop of Econometrics 2019, Shanghai, and XIII Conference of Spatial Econometrics Association 2019, Pittsburgh. We thank Ingmar Prucha, Lung-Fei Lee, James LeSage, Yichong Zhang, and the participants of the conferences for their helpful comments. Zhenlin Yang gratefully acknowledges the financial support from Singapore Management University under Lee Kong Chian Fellowship.

References

[1] Ahn, S.C., Lee, Y.H., Schmidt, P., 2001. GMM estimation of linear panel data models with time-varying individual effects. *Journal of Econometrics* 101, 219-255.

[2] Ahn, S.C., Lee, Y.H., Schmidt, P., 2013. Panel data models with multiple time-varying individual effects. *Journal of Econometrics* 174, 1-14.

[3] Bai, J., 2003. Inferential theory for factor models of large dimensions. *Econometrica* 71, 135-171.

[4] Bai, J., 2009. Panel data models with interactive fixed effects. *Econometrica* 77, 1229-1279.

[5] Bai, J., Li, K., 2021. Dynamic spatial panel data models with common shocks. *Journal of Econometrics*, 224, 134-160.

[6] Bai, J., Ng, S., 2013. Principal components estimation and identification of static factors. *Journal of Econometrics* 176, 18-29.

[7] Baltagi, B. H., Pirotte, A. and Yang, Z. L., 2021. Diagnostic tests for homoscedasticity in spatial cross-sectional or panel models. *Journal of Econometrics* 224, 245-270.

[8] Bernstein, D. S., 2009. *Matrix Mathematics: Theory, Facts, and Formulas.* Princeton University Press, Princeton.

[9] Connor, G., Korajczyk, R.A., 1986. Performance measurement with the arbitrage pricing theory: A new framework for analysis. *Journal of Financial Economics* 15, 373-394.

[10] Chamberlain, G., Rothschild, M., 1982. Arbitrage, factor structure, and mean-variance analysis on large asset markets. NBER Working Paper 996.

[11] Davidson, J., 1994. *Stochastic Limit Theory.* Oxford University Press, Oxford.

[12] Hsiao, C., 2018. Panel models with interactive effects. *Journal of Econometrics* 206, 645-673.

[13] Kelejian, H. H. and Prucha, I. R., 1999. A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review* 40, 509-533.

[14] Kiefer, N. M., 1980. A time series-cross section model with fixed effects with an intertemporal factor structure. *Working Paper*, Cornell University.

[15] Kuersteiner, G. M., Prucha, I. R., 2020. Dynamic panel data models: networks, common shocks, and sequential exogeneity. *Econometrica* 88, 2109-2146.

[16] Lee, L.-F., 2002. Consistency and efficiency of least squares estimation for mixed regressive spatial autoregressive models. *Econometric Theory* 18, 252-277.

[17] Lee, L.-F., 2004. Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica* 72, 1899-1925.

[18] Lee, L.-F., Yu, J., 2010. A spatial dynamic panel data model with both time and individual fixed effects. *Econometric Theory* 26, 564-597.

[19] Lee, L.-F., Yu, J., 2014. Efficient GMM estimation of spatial dynamic panel data models with fixed effects. *Journal of Econometrics* 180, 174-197.

[20] Lee, L.-F., Yu, J., 2015. Spatial panel data models. In: Baltagi, B. H. (Ed.), The Oxford Handbook of Panel Data. Oxford University Press, Oxford, pp.363-401.

[21] Li, L., Yang, Z. L., 2020. Estimation of fixed effects spatial dynamic panel data models with small T and unknown heteroskedasticity. *Regional Science and Urban Economics* 81, p.103520.

[22] Li, L., Yang, Z. L., 2021. Spatial dynamic panel data models with correlated random effects. *Journal of Econometrics* 221, 424-454.

[23] Manski, C. F., 1993. Identification of endogenous social effects: the reflection problem. *The Review of Economic Studies* 60, 531-542.

[24] Magnus, J. R., Neudecker, H., 2019. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley & Sons.

[25] Moon, H. R., Weidner, M., 2015. Linear regression for panel with unknown number of factors as interactive fixed effects. *Econometrica* 83, 1543-1579.

[26] Moon, H. R., Weidner, M., 2017. Dynamic linear regression models with interactive fixed effects. *Econometric Theory* 33, 158-195.

[27] Mutl, J., 2006. Dynamic panel data models with spatially correlated disturbances. *PhD Thesis*, University of Maryland, College Park.

[28] Newey, W. K., 1991. Uniform convergence in probability and stochastic equicontinuity. *Econometrica* 59, 1161-1167.

[29] Neyman, J., Scott, E.L., 1948. Consistent estimates based on partially consistent observations. *Econometrica* 16, 1-32.

[30] Pesaran, M. H., Tosetti, E., 2011. Large panels with common factors and spatial correlation. *Journal of Econometrics* 161, 182-202.

[31] Shi, W., Lee, L.-F., 2017. Spatial dynamic panel data models with interactive fixed effects. *Journal of Econometrics* 197, 323-347.

[32] Stock, J. H., Watson, M. W., 2002. Forecasting using principal components from a large number of predictors. *Journal of the American statistical association* 97, 1167-1179.

[33] Su, L., Yang, Z. L., 2015. QML estimation of dynamic panel data models with spatial errors. *Journal of Econometrics* 185, 230-258.

[34] van der Vaart, A. W., 1998. *Asymptotic Statistics.* Cambridge University Press.

[35] Yang, Z. L., 2015. A general method for third-order bias and variance correction on a nonlinear estimator. *Journal of Econometrics* 186, 178-200.

[36] Yang, Z. L., 2018. Unified M-estimation of fixed-effects spatial dynamic models with short panels. *Journal of Econometrics* 205, 423-447.

[37] Yang, Z. L., 2021. Joint tests for dynamic and spatial effects in short dynamic panel data models with fixed effects and heteroskedasticity. *Empirical Economics* 60, 51-92.

[38] Yang, Z. L., Li, C., Tse, Y. K., 2006. Functional form and spatial dependence in dynamic panels. *Economics Letters* 91, 138-145.

[39] Yu, J., de Jong, R., Lee, L.-F., 2008. Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large. *Journal of Econometrics* 146, 118-134.

Table 1. Empirical Mean(sd)[$\widehat{\text{rse}}$] of BC-CQMLE and M-Estimator: DGP1, $T = 3$, $m = 10$
$W_1 = W_2 = W_3$: `Rook Contiguity`, $r_0 = 1$, $r = 1$

	Normal Error		Normal Mixture		Chi-Square	
ψ	BC-CQMLE	M-Est	BC-CQMLE	M-Est	BC-CQMLE	M-Est
	$n = 50$					
1	.9746(.103)	.9982(.100)[.100]	.9770(.103)	.9998(.100)[.098]	.9736(.104)	.9955(.104)[.098]
1	.9744(.106)	.9925(.103)[.099]	.9691(.110)	.9890(.107)[.099]	.9782(.112)	.9965(.109)[.099]
1	.5801(.088)	.9007(.141)[.132]	.5674(.167)	.8832(.212)[.202]	.5752(.123)	.8930(.200)[.177]
.3	.2427(.072)	.2959(.062)[.062]	.2407(.083)	.2930(.068)[.065]	.2435(.073)	.2953(.060)[.061]
.2	.1766(.141)	.1929(.129)[.124]	.1741(.136)	.1936(.121)[.127]	.1720(.133)	.1904(.125)[.120]
.2	.2219(.076)	.2028(.079)[.077]	.2224(.075)	.2062(.077)[.078]	.2210(.078)	.2032(.078)[.080]
.2	.1881(.200)	.1931(.195)[.181]	.1835(.195)	.1869(.187)[.187]	.1896(.191)	.1892(.190)[.180]
	$n = 100$					
1	.9994(.076)	.9988(.075)[.073]	1.0005(.078)	1.0004(.078)[.072]	1.0009(.080)	1.0012(.078)[.074]
1	.9929(.073)	.9984(.072)[.072]	.9892(.075)	.9960(.073)[.072]	.9934(.076)	.9993(.075)[.073]
1	.6306(.065)	.9497(.099)[.095]	.6196(.134)	.9341(.195)[.185]	.6300(.098)	.9493(.150)[.141]
.3	.3117(.058)	.2996(.046)[.047]	.3146(.064)	.2990(.050)[.051]	.3137(.062)	.3016(.049)[.050]
.2	.1956(.092)	.1936(.091)[.091]	.2062(.085)	.2032(.084)[.087]	.1960(.093)	.1957(.092)[.089]
.2	.1869(.079)	.1989(.073)[.073]	.1829(.081)	.1959(.075)[.076]	.1877(.079)	.1994(.074)[.074]
.2	.1921(.133)	.1971(.133)[.132]	.1799(.127)	.1899(.127)[.130]	.1939(.134)	.1983(.133)[.130]
	$n = 200$					
1	.9851(.051)	1.0003(.053)[.052]	.9852(.052)	1.0002(.054)[.052]	.9811(.051)	.9963(.053)[.051]
1	.9792(.051)	.9997(.052)[.051]	.9798(.053)	.9995(.054)[.051]	.9812(.054)	1.0014(.054)[.051]
1	.6252(.046)	.9756(.075)[.072]	.6210(.092)	.9688(.143)[.140]	.6262(.072)	.9773(.119)[.107]
.3	.2571(.031)	.3003(.034)[.033]	.2583(.034)	.3002(.036)[.036]	.2577(.033)	.3009(.037)[.035]
.2	.1874(.065)	.1974(.065)[.064]	.1903(.067)	.2000(.064)[.064]	.1937(.065)	.2012(.064)[.064]
.2	.2007(.048)	.1996(.052)[.051]	.1995(.047)	.1983(.050)[.050]	.1990(.049)	.1997(.051)[.050]
.2	.1993(.091)	.1980(.090)[.089]	.1980(.091)	.1976(.090)[.089]	.1960(.088)	.1960(.087)[.087]
	$n = 400$					
1	.9951(.036)	.9985(.036)[.036]	.9964(.036)	.9997(.036)[.035]	.9971(.036)	.9999(.036)[.036]
1	.9861(.037)	1.0005(.037)[.036]	.9858(.038)	1.0000(.037)[.036]	.9837(.037)	.9980(.036)[.037]
1	.6425(.032)	.9899(.050)[.051]	.6367(.068)	.9891(.105)[.105]	.6424(.051)	.9980(.081)[.078]
.3	.2593(.027)	.2999(.023)[.023]	.2595(.031)	.3001(.027)[.027]	.2595(.029)	.3000(.025)[.024]
.2	.1993(.048)	.1994(.048)[.048]	.1985(.049)	.1999(.047)[.047]	.2008(.049)	.2002(.048)[.048]
.2	.2057(.030)	.2001(.031)[.031]	.2046(.030)	.1998(.032)[.032]	.2041(.030)	.1999(.031)[.032]
.2	.1954(.065)	.1995(.066)[.066]	.1994(.068)	.1994(.066)[.066]	.1915(.066)	.1982(.066)[.066]

Note: 1. $\psi = (\beta', \sigma_v^2, \rho, \lambda')'$; 2. r_0 = true number of factor, r = assumed number of factor.

Table 2. Empirical Mean(sd)[$\widehat{\text{rse}}$] of BC-CQMLE and M-Estimator: DGP1, $T = 3$, $m = 10$
$W_1 = W_2$: `Group Interaction`; W_3: `Queen Contiguity`, $r_0 = 1$, $r = 1$

	Normal Error		Normal Mixture		Chi-Square	
ψ	BC-CQMLE	M-Est	BC-CQMLE	M-Est	BC-CQMLE	M-Est
	$n = 50$					
1	.9637(.109)	.9988(.116)[.116]	.9624(.108)	.9971(.116)[.112]	.9599(.112)	.9934(.119)[.112]
1	.9811(.109)	1.0012(.109)[.107]	.9753(.108)	.9973(.109)[.102]	.9745(.111)	.9964(.111)[.104]
1	.6111(.096)	.9072(.144)[.137]	.6096(.179)	.9064(.237)[.232]	.6114(.135)	.9134(.201)[.185]
.3	.2601(.064)	.3011(.070)[.069]	.2590(.067)	.3005(.072)[.069]	.2555(.065)	.2977(.071)[.068]
.2	.1403(.135)	.1716(.126)[.132]	.1366(.137)	.1768(.130)[.119]	.1360(.136)	.1685(.121)[.118]
.2	.2141(.099)	.2062(.092)[.093]	.2093(.099)	.2004(.091)[.086]	.2100(.101)	.2028(.091)[.087]
.2	.1477(.162)	.1908(.178)[.188]	.1547(.155)	.1797(.187)[.180]	.1451(.154)	.1849(.174)[.179]
	$n = 100$					
1	.9691(.079)	.9987(.081)[.079]	.9729(.080)	1.0017(.085)[.081]	.9674(.079)	.9956(.083)[.080]
1	.9577(.083)	.9973(.080)[.079]	.9594(.087)	.9966(.082)[.079]	.9601(.083)	.9995(.080)[.079]
1	.6444(.068)	.9554(.104)[.099]	.6406(.135)	.9497(.200)[.181]	.6447(.100)	.9557(.145)[.141]
.3	.2638(.050)	.3005(.053)[.052]	.2637(.059)	.2993(.061)[.060]	.2648(.055)	.3010(.059)[.058]
.2	.0689(.084)	.1877(.089)[.086]	.0686(.085)	.1867(.090)[.085]	.0687(.089)	.1816(.085)[.086]
.2	.3473(.086)	.2174(.083)[.080]	.3447(.085)	.2190(.089)[.082]	.3403(.086)	.2110(.085)[.082]
.2	.2132(.117)	.1917(.127)[.125]	.2093(.122)	.1845(.125)[.124]	.2212(.110)	.1887(.123)[.124]
	$n = 200$					
1	.9918(.042)	.9988(.041)[.042]	.9909(.044)	.9980(.043)[.043]	.9933(.043)	1.0002(.043)[.042]
1	.9935(.052)	.9979(.050)[.049]	.9957(.050)	.9999(.049)[.049]	.9939(.050)	.9989(.049)[.049]
1	.6683(.048)	.9744(.071)[.069]	.6708(.097)	.9779(.136)[.134]	.6694(.075)	.9780(.103)[.100]
.3	.3105(.031)	.3001(.028)[.028]	.3118(.044)	.3004(.039)[.037]	.3096(.036)	.2992(.032)[.032]
.2	.0408(.037)	.1894(.063)[.062]	.0392(.039)	.1894(.065)[.062]	.0414(.035)	.1894(.061)[.062]
.2	.3381(.031)	.2095(.056)[.056]	.3378(.032)	.2082(.059)[.057]	.3367(.032)	.2115(.057)[.057]
.2	.2178(.096)	.1948(.085)[.085]	.2194(.096)	.1898(.087)[.085]	.2161(.092)	.1927(.086)[.085]
	$n = 400$					
1	.9561(.036)	1.0005(.036)[.036]	.9607(.036)	.9997(.036)[.036]	.9558(.036)	.9989(.036)[.036]
1	.9481(.041)	1.0001(.036)[.036]	.9508(.046)	.9991(.037)[.036]	.9467(.043)	.9997(.037)[.036]
1	.6382(.033)	.9866(.051)[.050]	.6315(.066)	.9897(.110)[.109]	.6375(.049)	.9898(.083)[.082]
.3	.1532(.047)	.2999(.023)[.023]	.1618(.064)	.3001(.028)[.027]	.1527(.054)	.2995(.025)[.024]
.2	.1161(.048)	.2004(.057)[.056]	.1181(.048)	.1981(.054)[.055]	.1126(.047)	.1979(.057)[.056]
.2	.2611(.063)	.2001(.043)[.043]	.2664(.060)	.2008(.045)[.045]	.2551(.060)	.2007(.046)[.046]
.2	.1880(.083)	.1991(.059)[.059]	.1968(.078)	.1989(.058)[.059]	.1843(.083)	.1977(.060)[.059]

Note: 1. $\psi = (\beta', \sigma_v^2, \rho, \lambda')'$; 2. r_0 = true number of factor, r = assumed number of factor.

Table 3. Empirical Mean(sd)[$\widehat{\text{rse}}$] of BC-QMLE and M-Estimator: DGP1, $T = 3$, $m = 10$
$W_1 = W_2 = W_3$: `Rook Contiguity`, $r_0 = 2$, $r = 2$

	Normal Error		Normal Mixture		Chi-Square	
ψ	BC-CQMLE	M-Est	BC-CQMLE	M-Est	BC-CQMLE	M-Est
	$n = 50$					
1	.7616(.106)	1.0201(.200)[.165]	.7876(.132)	1.0404(.208)[.167]	.7760(.122)	1.0350(.205)[.165]
1	.6464(.141)	.9876(.172)[.159]	.6812(.173)	.9923(.178)[.160]	.6705(.145)	.9965(.177)[.165]
1	.2120(.046)	.7848(.176)[.170]	.2017(.061)	.7828(.184)[.177]	.2104(.052)	.7481(.189)[.179]
.3	-.1793(.108)	.2698(.110)[.098]	-.1395(.170)	.2616(.115)[.110]	-.1591(.132)	.2520(.119)[.113]
.2	.2638(.177)	.1941(.190)[.188]	.2488(.168)	.1931(.198)[.190]	.2487(.164)	.1898(.197)[.190]
.2	.2348(.151)	.2133(.143)[.142]	.2202(.147)	.2266(.143)[.140]	.2174(.154)	.2166(.147)[.141]
.2	.0139(.272)	.1674(.309)[.303]	.0476(.263)	.1621(.310)[.298]	.0374(.266)	.1706(.311)[.297]
	$n = 100$					
1	.7475(.135)	.9699(.141)[.144]	.7750(.149)	.9817(.142)[.142]	.7556(.143)	.9789(.144)[.147]
1	.7796(.104)	.9863(.105)[.109]	.7989(.119)	.9729(.106)[.109]	.7881(.109)	.9724(.110)[.114]
1	.2053(.031)	.8981(.115)[.121]	.1968(.041)	.9023(.149)[.146]	.2024(.036)	.8468(.133)[.136]
.3	-.0547(.123)	.2906(.097)[.093]	-.0094(.169)	.2991(.101)[.092]	-.0454(.137)	.2897(.100)[.102]
.2	.1294(.254)	.1964(.160)[.163]	.1234(.241)	.1950(.163)[.166]	.1123(.237)	.1933(.164)[.167]
.2	.1771(.208)	.2011(.098)[.095]	.1797(.194)	.2024(.114)[.110]	.1675(.199)	.1987(.114)[.117]
.2	.1992(.302)	.1902(.202)[.201]	.2117(.288)	.1845(.204)[.207]	.2263(.287)	.1951(.212)[.215]
	$n = 200$					
1	.9759(.176)	1.0102(.087)[.087]	1.0022(.167)	1.0021(.088)[.087]	.9866(.168)	1.0014(.089)[.088]
1	.9668(.137)	1.0071(.071)[.072]	.9769(.123)	1.0055(.070)[.072]	.9739(.131)	1.0087(.074)[.075]
1	.2973(.029)	.9489(.083)[.087]	.2837(.046)	.9640(.096)[.099]	.2920(.036)	.9348(.103)[.104]
.3	.2091(.192)	.3011(.050)[.049]	.2346(.181)	.3061(.051)[.049]	.2251(.184)	.3051(.051)[.049]
.2	.1786(.103)	.1982(.083)[.084]	.1808(.094)	.1983(.084)[.084]	.1754(.106)	.1981(.090)[.091]
.2	.1900(.063)	.1993(.060)[.059]	.1858(.063)	.1994(.061)[.062]	.1843(.068)	.1982(.067)[.069]
.2	.1933(.139)	.1994(.125)[.123]	.1839(.126)	.1980(.127)[.129]	.1926(.138)	.1978(.131)[.131]
	$n = 400$					
1	.9289(.047)	.9996(.028)[.028]	.9290(.048)	.9989(.031)[.031]	.9301(.048)	.9984(.029)[.030]
1	.8905(.091)	.9963(.049)[.050]	.8978(.089)	.9983(.051)[.051]	.8925(.089)	.9865(.048)[.049]
1	.3138(.022)	.9893(.071)[.071]	.3073(.034)	.9888(.084)[.085]	.3095(.027)	.9833(.083)[.083]
.3	.1682(.180)	.2996(.030)[.030]	.1970(.185)	.2988(.031)[.032]	.1807(.182)	.2983(.034)[.034]
.2	.1662(.043)	.1994(.031)[.031]	.1680(.046)	.1960(.032)[.033]	.1662(.044)	.1973(.032)[.033]
.2	.2073(.032)	.2003(.026)[.026]	.1999(.035)	.1970(.028)[.029]	.2041(.032)	.2000(.027)[.028]
.2	.1910(.078)	.1996(.074)[.075]	.1982(.076)	.1961(.075)[.076]	.1930(.078)	.1962(.076)[.076]

Note: 1. $\psi = (\beta', \sigma_v^2, \rho, \lambda')'$; 2. r_0 = true number of factor, r = assumed number of factor.

Table 4. Empirical Mean(sd)[$\widehat{\text{rse}}$] of BC-CQMLE and M-Estimator: DGP1, $T = 10$, $m = 10$
$W_1 = W_2 = W_3$: Rook Contiguity, $r_0 = 1$, $r = 1$

	Normal Error		Normal Mixture		Chi-Square	
ψ	BC-CQMLE	M-Est	BC-CQMLE	M-Est	BC-CQMLE	M-Est
	$n = 25$					
1	.9958(.069)	.9957(.069)[.065]	.9995(.070)	.9994(.070)[.065]	.9991(.069)	.9990(.069)[.066]
1	.9966(.070)	.9967(.070)[.066]	.9926(.072)	.9927(.072)[.066]	.9995(.069)	.9996(.069)[.065]
1	.8256(.078)	.9176(.087)[.085]	.8199(.186)	.9112(.193)[.170]	.8265(.133)	.9186(.147)[.136]
.3	.2979(.038)	.2987(.038)[.035]	.3018(.037)	.3015(.037)[.035]	.2986(.038)	.2994(.038)[.035]
.2	.1941(.076)	.1949(.076)[.073]	.1971(.074)	.1971(.074)[.071]	.1978(.072)	.1976(.072)[.071]
.2	.2020(.064)	.2011(.064)[.061]	.1982(.063)	.1974(.063)[.061]	.1976(.063)	.1968(.063)[.061]
.2	.2064(.120)	.2027(.121)[.117]	.1983(.113)	.2033(.115)[.113]	.1975(.110)	.2028(.112)[.113]
	$n = 50$					
1	.9978(.044)	.9979(.044)[.042]	.9992(.045)	.9993(.045)[.043]	.9996(.046)	.9997(.046)[.043]
1	.9985(.046)	.9985(.046)[.047]	.9997(.048)	.9997(.048)[.046]	1.0007(.049)	1.0007(.049)[.047]
1	.8610(.059)	.9568(.066)[.064]	.8686(.136)	.9653(.141)[.139]	.8649(.098)	.9611(.109)[.103]
.3	.2985(.026)	.2994(.026)[.026]	.2991(.027)	.3000(.027)[.027]	.2979(.026)	.2998(.026)[.026]
.2	.1973(.060)	.1974(.060)[.059]	.1980(.060)	.1981(.060)[.059]	.1952(.059)	.1983(.059)[.059]
.2	.2000(.042)	.1997(.042)[.041]	.1986(.044)	.1988(.044)[.042]	.2017(.042)	.2013(.042)[.042]
.2	.1984(.087)	.2012(.088)[.086]	.1963(.087)	.2011(.087)[.085]	.1987(.084)	.2013(.084)[.084]
	$n = 100$					
1	.9995(.029)	.9995(.029)[.030]	1.0000(.032)	1.0000(.032)[.032]	1.0003(.031)	1.0009(.031)[.030]
1	1.0013(.033)	1.0004(.033)[.033]	1.0016(.034)	1.0006(.034)[.033]	.9971(.033)	.9981(.033)[.033]
1	.8837(.041)	.9841(.046)[.046]	.8843(.098)	.9848(.107)[.105]	.8821(.071)	.9882(.078)[.075]
.3	.2997(.018)	.3002(.018)[.018]	.2985(.019)	.2998(.019)[.018]	.2999(.018)	.3001(.018)[.018]
.2	.1961(.038)	.1986(.038)[.037]	.1990(.038)	.1989(.038)[.037]	.1998(.038)	.1997(.038)[.037]
.2	.2014(.029)	.2001(.029)[.029]	.2006(.029)	.2001(.029)[.029]	.1983(.029)	.1989(.029)[.029]
.2	.2006(.056)	.2006(.056)[.057]	.1982(.058)	.2002(.058)[.057]	.1982(.058)	.2003(.058)[.057]
	$n = 200$					
1	.9990(.023)	.9998(.023)[.023]	1.0005(.024)	1.0003(.024)[.024]	.9997(.023)	1.0002(.023)[.023]
1	.9990(.022)	.9997(.022)[.023]	.9996(.023)	.9998(.023)[.023]	1.0009(.023)	1.0001(.023)[.023]
1	.8901(.030)	.9989(.033)[.033]	.8905(.070)	.9981(.076)[.076]	.8886(.051)	.9980(.057)[.056]
.3	.2978(.013)	.2999(.013)[.013]	.2971(.014)	.2999(.014)[.014]	.2975(.014)	.2998(.014)[.014]
.2	.2006(.028)	.2001(.028)[.028]	.1991(.029)	.1988(.029)[.029]	.1985(.029)	.1982(.029)[.029]
.2	.2003(.021)	.1999(.021)[.021]	.2015(.021)	.2001(.021)[.021]	.2001(.021)	.1996(.021)[.021]
.2	.1974(.042)	.1997(.042)[.042]	.2006(.043)	.2001(.043)[.043]	.2011(.043)	.2002(.043)[.043]

Note: 1. $\psi = (\beta', \sigma_v^2, \rho, \lambda')'$; 2. r_0 = true number of factor, r = assumed number of factor.

Table 5. Empirical Mean(sd)[$\widehat{\text{rse}}$] of BC-CQMLE and M-Estimator: DGP1, $T = 10$, $m=10$
$W_1 = W_3$: `Queen Contiguity`, W_2: `Group Interaction`, $r_0 = 1$, $r = 1$

	Normal Error		Normal Mixture		Chi-Square	
ψ	BC-CQMLE	M-Est	BC-CQMLE	M-Est	BC-CQMLE	M-Est
	$n = 25$					
1	1.0018(.071)	1.0015(.071)[.065]	1.0015(.070)	1.0013(.070)[.065]	1.0000(.070)	.9999(.070)[.066]
1	.9966(.069)	.9965(.069)[.064]	1.0007(.067)	1.0005(.067)[.063]	1.0017(.066)	1.0015(.066)[.065]
1	.8284(.078)	.9208(.087)[.085]	.8161(.180)	.9071(.200)[.169]	.8347(.130)	.9278(.145)[.131]
.3	.2970(.037)	.2982(.037)[.035]	.2970(.037)	.2981(.037)[.036]	.2937(.039)	.2949(.039)[.036]
.2	.1950(.081)	.1949(.082)[.075]	.1956(.079)	.1952(.078)[.074]	.1975(.078)	.1971(.078)[.075]
.2	.1995(.051)	.1992(.051)[.049]	.1960(.053)	.1957(.053)[.048]	.1951(.052)	.1947(.052)[.048]
.2	.1888(.140)	.1888(.145)[.146]	.1798(.141)	.1795(.145)[.150]	.1851(.136)	.1864(.141)[.146]
	$n = 50$					
1	.9956(.046)	.9960(.046)[.043]	1.0012(.046)	1.0010(.046)[.044]	1.0014(.045)	1.0009(.045)[.042]
1	.9999(.046)	1.0001(.046)[.047]	.9995(.047)	.9997(.047)[.047]	1.0012(.047)	1.0022(.047)[.047]
1	.8663(.059)	.9628(.066)[.063]	.8676(.137)	.9643(.152)[.140]	.8673(.100)	.9640(.111)[.102]
.3	.3017(.023)	.3004(.023)[.022]	.3001(.025)	.2988(.025)[.023]	.2990(.022)	.2977(.023)[.022]
.2	.1999(.043)	.2003(.043)[.044]	.1989(.045)	.1992(.045)[.043]	.1956(.046)	.1959(.046)[.044]
.2	.1986(.028)	.1993(.028)[.026]	.1986(.028)	.1987(.028)[.027]	.2004(.027)	.2005(.027)[.026]
.2	.1869(.090)	.1942(.092)[.091]	.1896(.090)	.1890(.092)[.090]	.1945(.091)	.1948(.092)[.090]
	$n = 100$					
1	1.0006(.031)	1.0005(.031)[.030]	1.0004(.031)	1.0004(.031)[.030]	1.0001(.030)	1.0001(.030)[.030]
1	1.0002(.034)	1.0002(.034)[.033]	1.0000(.033)	1.0000(.033)[.033]	.9996(.032)	.9986(.032)[.033]
1	.8836(.042)	.9928(.047)[.046]	.8795(.095)	.9773(.105)[.102]	.8807(.070)	.9786(.078)[.075]
.3	.2993(.018)	.2998(.018)[.018]	.3012(.018)	.3006(.018)[.018]	.2992(.018)	.2987(.018)[.018]
.2	.1964(.041)	.1986(.041)[.040]	.1976(.040)	.1997(.040)[.040]	.1957(.040)	.1959(.040)[.040]
.2	.1997(.033)	.1999(.033)[.033]	.1976(.033)	.1997(.033)[.033]	.1984(.033)	.2003(.033)[.033]
.2	.2005(.069)	.2001(.070)[.069]	.1962(.067)	.1978(.068)[.068]	.1992(.067)	.2007(.067)[.068]
	$n = 200$					
1	.9994(.021)	.9997(.021)[.021]	.9990(.021)	.9993(.021)[.021]	.9993(.021)	.9995(.021)[.021]
1	.9999(.023)	.9999(.023)[.023]	.9986(.023)	.9996(.023)[.023]	.9994(.024)	.9995(.024)[.024]
1	.8909(.029)	.9902(.032)[.033]	.8927(.071)	.9923(.075)[.075]	.8944(.049)	.9941(.055)[.055]
.3	.3002(.012)	.2999(.012)[.012]	.3008(.013)	.2999(.013)[.013]	.3001(.013)	.2999(.013)[.013]
.2	.1986(.028)	.1996(.028)[.028]	.1994(.026)	.1996(.026)[.026]	.1998(.029)	.2000(.029)[.029]
.2	.1992(.027)	.1998(.027)[.027]	.1984(.026)	.1998(.026)[.026]	.1979(.028)	.1983(.028)[.026]
.2	.1986(.050)	.1998(.050)[.049]	.1963(.047)	.1995(.048)[.048]	.1953(.048)	.2001(.048)[.048]

Note: 1. $\psi = (\beta', \sigma_v^2, \rho, \lambda')'$; 2. r_0 = true number of factor, r = assumed number of factor.

Table 6. Empirical Mean(sd)[$\widehat{\text{rse}}$] of BC-CQMLE and M-Estimator: DPG1, $T = 3$, $m = 10$
$W_1 = W_2 = W_3$: `Rook Contiguity`, $r_0 = 1$, $r = 2$

	Normal Error		Normal Mixture		Chi-Square	
ψ	BC-CQMLE	M-Est	BC-CQMLE	M-Est	BC-CQMLE	M-Est
	$n = 50$					
1	.7243(.174)	.9988(.154)[.151]	.7857(.196)	.9895(.142)[.137]	.7507(.185)	.9899(.155)[.151]
1	.7370(.181)	.9838(.172)[.160]	.8110(.199)	.9965(.154)[.148]	.7728(.190)	.9877(.162)[.157]
1	.1701(.039)	.6797(.144)[.147]	.1607(.048)	.6254(.195)[.155]	.1653(.043)	.6560(.180)[.155]
.3	-.1715(.210)	.2939(.107)[.102]	-.0564(.262)	.2932(.096)[.090]	-.1249(.234)	.2885(.100)[.102]
.2	.0957(.282)	.1870(.190)[.191]	.1202(.246)	.1806(.171)[.167]	.1032(.264)	.1622(.183)[.191]
.2	.1705(.246)	.2053(.143)[.150]	.1852(.209)	.2024(.121)[.126]	.1716(.234)	.1899(.135)[.147]
.2	.1402(.356)	.1876(.303)[.301]	.1377(.329)	.1767(.290)[.284]	.1523(.345)	.1980(.307)[.315]
	$n = 100$					
1	.8124(.195)	.9979(.111)[.119]	.8778(.179)	.9943(.109)[.110]	.8396(.192)	.9967(.106)[.118]
1	.8458(.149)	.9950(.107)[.115]	.8929(.150)	.9986(.105)[.109]	.8674(.161)	.9924(.113)[.122]
1	.2444(.033)	.7933(.103)[.119]	.2243(.052)	.7402(.178)[.158]	.2360(.042)	.7570(.144)[.139]
.3	.1514(.258)	.2972(.076)[.077]	.2215(.228)	.2999(.073)[.073]	.1780(.253)	.2965(.073)[.083]
.2	.1662(.177)	.1997(.147)[.149]	.1676(.162)	.1961(.135)[.140]	.1666(.176)	.2028(.136)[.160]
.2	.1957(.149)	.1976(.124)[.135]	.1806(.142)	.1973(.120)[.123]	.1877(.142)	.1985(.119)[.140]
.2	.1591(.243)	.1921(.206)[.224]	.1900(.214)	.1930(.197)[.212]	.1819(.248)	.1960(.207)[.233]
	$n = 200$					
1	.8223(.069)	.9987(.081)[.087]	.8431(.084)	.9986(.076)[.087]	.8371(.078)	1.0006(.079)[.085]
1	.7641(.076)	.9978(.078)[.086]	.7966(.102)	.9985(.073)[.085]	.7788(.088)	.9971(.078)[.085]
1	.2247(.025)	.8775(.088)[.100]	.2199(.035)	.8154(.140)[.145]	.2229(.028)	.8644(.115)[.123]
.3	-.0425(.074)	.2982(.060)[.060]	-.0077(.120)	.2998(.057)[.063]	-.0255(.098)	.2987(.059)[.065]
.2	.1351(.143)	.1993(.101)[.112]	.1421(.133)	.1976(.099)[.109]	.1467(.133)	.2011(.101)[.111]
.2	.1101(.101)	.1999(.094)[.095]	.1249(.104)	.1996(.082)[.093]	.1195(.100)	.1996(.089)[.096]
.2	.1984(.185)	.2003(.131)[.152]	.1959(.170)	.1968(.130)[.151]	.1922(.168)	.1963(.138)[.151]
	$n = 400$					
1	.9381(.056)	.9991(.055)[.060]	.9462(.057)	.9994(.053)[.060]	.9397(.058)	.9995(.053)[.060]
1	.9412(.058)	.9980(.055)[.059]	.9548(.056)	.9986(.052)[.059]	.9449(.058)	.9981(.053)[.064]
1	.2859(.022)	.9591(.059)[.070]	.2745(.035)	.9230(.108)[.118]	.2811(.028)	.9025(.085)[.088]
.3	.2165(.082)	.2993(.036)[.039]	.2236(.078)	.2993(.038)[.046]	.2217(.083)	.2973(.039)[.036]
.2	.2168(.072)	.2002(.069)[.078]	.2034(.071)	.1992(.070)[.074]	.2047(.076)	.1977(.072)[.084]
.2	.2156(.049)	.2001(.048)[.054]	.2166(.046)	.2005(.047)[.053]	.2150(.047)	.2008(.049)[.058]
.2	.1888(.096)	.1998(.097)[.108]	.1989(.099)	.1998(.098)[.106]	.1960(.105)	.2008(.102)[.118]

Note: 1. $\psi = (\beta', \sigma_v^2, \rho, \lambda')'$; 2. r_0 = true number of factor, r = assumed number of factor.

Table 7. Empirical Mean(sd)[$\widehat{\text{rse}}$] of GMM and M Estimators: DGP2, $T = 3$, $m = 10$
$W_1 = W_2$: `Rook Contiguity`, $r_0 = 1$, $r = 1$

	Normal Error		Normal Mixture		Chi-Square	
ψ	KP-GMM	M-Est	KP-GMM	M-Est	KP-GMM	M-Est
	$n = 50$					
1	.9907(.084)	.9992(.050)[.049]	.9922(.082)	.9992(.053)[.048]	.9880(.083)	.9991(.052)[.048]
1	.9651(.106)	.9984(.050)[.048]	.9656(.098)	.9998(.051)[.047]	.9724(.097)	1.0011(.049)[.048]
.2	.1951(.073)	.1995(.034)[.034]	.1990(.070)	.1992(.035)[.034]	.1951(.070)	.2010(.035)[.033]
.2	.1890(.104)	.1960(.056)[.054]	.1985(.104)	.1985(.055)[.053]	.1903(.103)	.1958(.055)[.053]
.2	.1993(.094)	.2006(.051)[.048]	.1973(.091)	.2020(.049)[.047]	.1966(.089)	.1979(.050)[.047]
	$n = 100$					
1	.9694(.063)	.9986(.037)[.037]	.9722(.061)	1.0012(.038)[.037]	.9728(.064)	1.0007(.038)[.036]
1	.9772(.059)	.9999(.037)[.036]	.9813(.057)	1.0010(.037)[.036]	.9836(.060)	1.0007(.038)[.036]
.2	.1855(.064)	.1998(.026)[.026]	.1886(.063)	.2024(.027)[.027]	.1856(.062)	.2007(.026)[.026]
.2	.2048(.074)	.1999(.041)[.041]	.2054(.067)	.1989(.039)[.040]	.2031(.067)	.1980(.042)[.041]
.2	.2148(.082)	.2022(.044)[.044]	.2073(.078)	.2002(.045)[.043]	.2086(.075)	.1996(.045)[.043]
	$n = 200$					
1	.9968(.040)	1.0001(.027)[.026]	.9976(.038)	1.0003(.025)[.026]	.9978(.040)	1.0008(.027)[.026]
1	.9935(.042)	.9975(.027)[.025]	.9949(.041)	.9991(.026)[.025]	.9937(.042)	.9997(.026)[.026]
.2	.1962(.033)	.1999(.019)[.019]	.1968(.032)	.2003(.020)[.019]	.1966(.033)	.2006(.020)[.019]
.2	.1996(.048)	.2008(.031)[.030]	.2005(.049)	.1991(.031)[.030]	.2016(.049)	.2006(.030)[.030]
.2	.1974(.053)	.1984(.031)[.030]	.1985(.054)	.1992(.030)[.029]	.2013(.053)	.2000(.030)[.030]
	$n = 400$					
1	.9986(.029)	.9990(.019)[.019]	.9892(.029)	.9988(.019)[.019]	.9921(.029)	.9999(.018)[.018]
1	1.0063(.028)	1.0002(.017)[.018]	1.0062(.027)	.9999(.018)[.018]	1.0076(.028)	1.0000(.017)[.018]
.2	.2104(.020)	.2000(.013)[.013]	.2092(.020)	.1991(.013)[.013]	.2092(.020)	.1990(.014)[.013]
.2	.1982(.035)	.1995(.021)[.021]	.1920(.037)	.2004(.022)[.022]	.1892(.036)	.2001(.021)[.021]
.2	.2063(.037)	.2004(.021)[.021]	.2067(.036)	.1997(.023)[.023]	.2071(.036)	.1997(.022)[.023]

Note: 1. $\psi = (\beta', \rho, \lambda_1, \lambda_2)'$; 2. r_0 = true number of factor, r = assumed number of factor.