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

Liyao Li[a] Ke Miao[b] Zhenlin Yang[c*]

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

[b]*School of Economics, Fudan University, Shanghai, China*

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

April 14, 2024

Abstract

We propose an M-estimation method for estimating dynamic spatial panel data models with interactive fixed effects based on (relatively) short panels. Unbiased estimating functions (EF) are obtained by adjusting the concentrated conditional quasi scores, given the initial values and with the factor loadings being concentrated out, to account for the effects of conditioning and concentration. Solving the estimating equations gives the M-estimators of the common parameters and common factors. Under fixed T, $\sqrt{n}$-consistency and joint asymptotic normality of the two sets of M-estimators are established. Under $T = o(n)$, the M-estimators of the common parameters are shown to be $\sqrt{nT}$-consistent and asymptotically normal. For inference, difficulty lies in the estimation of the variance-covariance (VC) matrix of the EF. We decompose the EF into a sum of n nearly uncorrelated terms. Outer products of these n terms together with a covariance adjustment lead to a consistent estimator of the VC matrix under both fixed T and $T = o(n)$. Important extensions of the methods, allowing for unknown heteroskedasticity, time-varying spatial weight matrices, high-order dynamic and spatial effects, are critically discussed. Monte Carlo results show that the proposed methods perform well in finite sample.

Key Words: Adjusted quasi scores; Dynamic effects; Initial conditions; Incidental parameters; Interactive fixed effects; High-order spatial effects.

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

1. Introduction

Dynamic spatial panel data (DSPD) 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

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

Durbin, and spatial error),[1] and (*iii*) control for unobserved spatiotemporal heterogeneity (individual-specific and time-specific). The bulk of the literature has focused on the DSPD 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 (2015) for a survey on earlier works.

A major recent advancement in the literature of DSPD models is the incorporation of interactive fixed effects (IFE) (Shi and Lee 2017 or SL; Kuersteiner and Prucha 2020 or KP; Bai and Li 2021 or BL; Cui et al. 2023; and Higgins 2023). Besides the existing attractive features, this extended model draws further on the strength of IFE in controlling for the multiple unobserved time-specific effects f_t (the *common factors*) and the corresponding individual-specific responses γ_i (the *factor loadings*). However, this strand of literature is still quite sparse and important asymptotic frameworks with fixed or relatively small T have not been formally considered from likelihood perspective due to technical difficulties caused by IFE.[2]

SL and BL both adopt conditional QML (CQML) approach, given initial observations, to estimate similar first-order DSPD-IFE models. Under a simultaneous passage of n and T to ∞, the CQML estimators are consistent but have non-negligible biases of order $O(\frac{1}{T}) + O(\frac{1}{n})$. A bias-correction removes these biases but it leaves the asymptotic variance unchanged only when $\frac{T}{n} \to c \neq 0$ (see Sec. 2 for further details). KP adopt GMM approach to estimate a high-order DSPD-IFE model (with a different spatial error structure) under a large n and small T setup. Their method allows several (important) additional features (see Sec. 2 for details). The key challenges in the estimation of a DSPD-IFE model are (*i*) the *initial values problem* (IVP) and (*ii*) the *incidental parameters problem* (IPP). The CQML-based methods handle these problems through concentrations and after-estimation bias-corrections. The GMM method handles the IVP by taking use of sequential exogeneity in setting up moments and the IPP by a novel *forward orthogonal deviations* (FOD) transformation that eliminates the factor loadings and at the same time adjusts the degrees of freedom loss. The GMM method does not require further bias-corrections for valid inferences but does require T to be small. Cui et al. (2023) propose instrumental variables (IV) estimation of a simple DSPD-IFE model under Pesaran's (2006) common correlated effects setup, which is valid for large n and large T. Higgins (2023) also propose an IV approach to estimate a slightly more general model

[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.

[2] This is in stark contrast to the large literature on regular panel models with IFE; see but a few Ahn et al. (2001, 2013), Bai (2009), Bai and Ng (2013), Moon and Weidner (2015, 2017). 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.

based on a transformation approach, which is valid for small T. Except KP, all the other four papers discussed above assume first-order spatial effects with time-invariant spatial weights. The later two papers further restrict the model to be free from spatial errors, which may be the key for their IV approaches to work. Clearly, these assumptions are too restrictive, in particular, from perspectives of network effects and social interaction as discussed in KP.

In this paper, we study a general class of DSPD-IFE models similar to that studied by KP, using likelihood-based methods and focusing on the most important asymptotic scenarios of (*i*) T being fixed and (*ii*) T being large but small relative to n. Scenario (*i*) serves as an alternative to KP's GMM approach. We introduce M-estimation methods, which are valid for both of these asymptotic scenarios. We obtain a set of unbiased estimating functions (EF) by **adjusting** the concentrated conditional quasi scores (CCQS) of the common parameters and the factor parameters, given initial observations and with factor loadings being concentrated out, to **directly remove** the effects of conditioning (or IVP) and concentration (or IPP) before estimation. Solving the resulting estimating equations gives **M-estimators** of both sets of parameters that possess usual asymptotic properties. In particular, under fixed T, they are $\sqrt{n}$-consistent and asymptotically normal with zero mean; under $T = o(n)$, the M-estimators of the common parameters are $\sqrt{nT}$-consistent and asymptotically normal with zero mean.[3]

For statistical inference, difficulty lies in the estimation of the variance-covariance (VC) matrix of the EF. We propose to decompose the EF into a sum of n nearly uncorrelated terms. Outer products of these n terms together with a covariance adjustment lead to a consistent estimator of the VC matrix under both cases when T is fixed and $T = o(n)$. The proposed methods are extended to accommodate unknown heteroskedasticity, time-varying spatial weight matrices, and high-order dynamic effects, high-order spatial effects, etc.

Our work complements KP's fixed-T GMM by providing alternative, likelihood-based methods, which are valid when either T is fixed or $T = o(n)$, covering both of the most interesting scenarios in spatial panel data analyses. Furthermore, our methods do not require a transformation but KP's methods depend critically on the FOD transformation; our methods allow cross-sectional heteroskedasticity to be of an unknown form but their methods require it to be a function of a finite number of parameters; their methods allow sequential exogeneity in spatial weight matrices and some regressors but our methods allow only the endogeneity of a 'known form' (time lags of responses, control functions for endogenous spatial weights and endogenous regressors, etc.); and finally, Monte Carlo results suggest that M-estimator is more efficient than GMM estimator under strict exogeneity. Our work also complements those

[3]The proposed method is related to Yang (2018, 2021) and Li and Yang (2020) on a first-order DSPD model with additive fixed effects under small T, where unit-specific effects are eliminated by first-differencing. With the allowance of IFE and large T, the first-differencing or other transformation method is inapplicable. The proposed methods are in line with the *modified equations of maximum likelihood* of Neyman and Scott (1948, Sec. 5), in a search of a systematic method of addressing the incidental parameters problem.

of SL and BL by providing likelihood-based methods for DSPD-IFE models under the 'fixed or relatively small T' asymptotic frameworks. However, our methods differ from theirs in that we adjust the CCQS functions before estimation to correct the IVP due to conditioning and the IPP due to concentration. As a result, our M-estimators have a bias of order $O(\frac{1}{n})$ only, and hence our inferences for common parameters are valid as long as $T/n \to 0$.

The rest of the paper goes as follows. Section 2 discusses model specifications. Section 3 introduces M-estimator, its asymptotic properties, and standard error estimation for a first-order DSPD-IFE model. Section 4 presents M-estimation for extended DSPD-IFE models to allow for heteroskedasticity, time-varying spatial weight matrices, and higher-order spatial and dynamic effects in the model. Section 5 presents Monte Carlo results. Section 6 concludes the paper. All technical proofs are collected in appendix.

2. Model Specifications

The high-order dynamic spatial panel data (DSPD) model with interactive fixed effects (IFE) recently studied by Kuersteiner and Prucha (2020) is by far the most general DSPD-IFE model in the literature. The model can be written in a more explicit form:

$$\begin{aligned} y_t &= \sum_{s=1}^{p} \rho_s y_{t-s} + \sum_{\ell=1}^{q_1} \lambda_{1\ell} W_{1\ell t} y_t + \sum_{s=1}^{p}\sum_{\ell=1}^{q_2} \lambda_{2\ell s} W_{2\ell,t-s} y_{t-s} + x_t\beta \ + u_t, \\ u_t &= \sum_{\ell=1}^{q_3} \lambda_{3\ell} W_{3\ell t} u_t + \Gamma f_t + v_t, \quad t = p, \ldots, T, \end{aligned} \tag{2.1}$$

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; x_t is an $n \times k$ matrix of regressors' values; $W_{\nu\ell t}, \nu = 1, 2, 3, \ell = 1, \ldots, q_\nu, t = 1, \ldots, T$, are $n \times n$ spatial weight matrices; and f_t is a $r \times 1$ vector of common factors and Γ is the corresponding $n \times r$ matrix of factor loadings.

KP propose a GMM for the estimation of the model under the small T setup, assuming $v_{it} \sim (0, \varrho_i(\gamma)\sigma_t^2)$, where $\varrho_i(\gamma)$ are functions with finite number of parameters γ, $W_{\nu\ell t}$ to be sequentially exogenous, and x_t to contain exogenous and sequentially exogenous regressors and their spatial lags. At the core of KP's GMM method are (i) the reduced form: $B_{3t}(\lambda_3)y_t = B_{3t}(\lambda_3)R_t\psi + \Gamma f_t + v_t$, where $B_{3t}(\lambda_3) = I_n - \sum_{\ell=1}^{q_3} \lambda_{3\ell} W_{3\ell t}$, $\lambda_3 = (\lambda_{31}, \ldots, \lambda_{3q_3})'$, R_t collects all the right hand side terms except u_t, and ψ collects the corresponding coefficients; and (ii) the FOD transformation, a $(T-r) \times T$ matrix function of $\{f_t\}$ and $\{\sigma_t^2\}$ that eliminates Γ and maintains zero correlation of the (transformed) $v_{it}'s$ and sequential exogeneity of the variables and the spatial weight matrices so that linear and quadratic moments are formed.

KP's model specifies that spatial interactions among model's disturbances act equally on their components, Γf_t and v_t. An alternative and perhaps more popular specification may be:

$$
\begin{aligned}
y_t &= \sum_{s=1}^{p} \rho_s y_{t-s} + \sum_{\ell=1}^{q_1} \lambda_{1\ell} W_{1\ell t} y_t + \sum_{s=1}^{p}\sum_{\ell=1}^{q_2} \lambda_{2\ell} W_{2\ell,t-s} y_{t-s} + x_t\beta + \Gamma f_t + u_t, \\
u_t &= \sum_{\ell=1}^{q_3} \lambda_{3\ell} W_{3\ell t} u_t + v_t, \quad t = p, \ldots, T,
\end{aligned} \tag{2.2}
$$

which stresses on that spatial interactions occur only in 'remainder' errors, not in unobserved (unshown) individual and time specific effects Γf_t. However, with this model specification, the first equation cannot be written in a simple form in $\Gamma f_t + v_t$ but rather in $B_{3t}(\lambda_3)\Gamma f_t + v_t$ or in $\Gamma f_t + B_{3t}^{-1}(\lambda_3) v_t$. Hence, the FOD-based GMM may not be implementable unless $B_{3t}(\lambda_3)$ is time-invariant so that model's reduced form has disturbance $\Gamma^* f_t + v_t$, where $\Gamma^* = B_3(\lambda_3)\Gamma$ and FOD can be applied to eliminate Γ^*.

Model (2.1) specifies a single spatial autoregressive (SAR) process for the disturbances that is driven by factors and idiosyncratic errors, $\Gamma f_t + v_t$, together, whereas Model (2.2) specifies a single SAR process that is driven by v_t only. A more general model would naturally be that the disturbances contain two SAR processes, driven independently by Γf_t and v_t:

$$
\begin{aligned}
y_t &= \sum_{s=1}^{p} \rho_s y_{t-s} + \sum_{\ell=1}^{q_1} \lambda_{1\ell} W_{1\ell t} y_t + \sum_{s=1}^{p}\sum_{\ell=1}^{q_2} \lambda_{2\ell} W_{2\ell,t-s} y_{t-s} + x_t\beta + \varepsilon_t + u_t, \\
u_t &= \sum_{\ell=1}^{q_3} \lambda_{3\ell} W_{3\ell t} u_t + v_t, \\
\varepsilon_t &= \sum_{\ell=1}^{q_4} \lambda_{4\ell} W_{4\ell t} \varepsilon_t + \Gamma f_t, \quad t = p, \ldots, T.
\end{aligned} \tag{2.3}
$$

Again, the FOD-based GMM may not be implementable, unless $B_{3t}(\lambda_3)$ and $B_{4t}(\lambda_4)$ are both time-invariant, where $B_{4t}(\lambda_4) = I_n - \sum_{\ell=1}^{q_4} \lambda_{4\ell} W_{4\ell t}$ and $\lambda_4 = (\lambda_{41}, \ldots, \lambda_{4q_4})'$. In this case, FOD works on $\Gamma^\diamond f_t + v_t$, where $\Gamma^\diamond = B_3(\lambda_3) B_4^{-1}(\lambda_4)\Gamma$, and GMM proceeds as for (2.1).

Model (2.3) exhibits a great generality and should be highly useful in modeling spatial and network data, in particular in the era of big data. It contains Model (2.1) as a special case with $q_3 = q_4$, $\lambda_{3\ell} = \lambda_{4\ell}$ and $W_{3\ell t} = W_{4\ell t}$, and it reduces to Model (2.2) by setting $\lambda_{4\ell} = 0$. A very interesting special case of Model (2.2) is when $p = q_1 = q_2 = q_3 = 1$, i.e., the first-order DSPD-IFE model that will be rigorously studied in this paper:

$$
\begin{aligned}
y_t &= \rho y_{t-1} + \lambda_1 W_{1t} y_t + \lambda_2 W_{2t} y_{t-1} + x_t\beta + \Gamma f_t + u_t, \\
u_t &= \lambda_3 W_{3t} u_t + v_t, \quad t = 1, 2, \ldots, T.
\end{aligned} \tag{2.4}
$$

In our study, we view $t = 0$ as the initial period of data collection but the process may have started m periods earlier, where m may be finite or infinite. Thus, under Model (2.4), y_0 represents the vector of *initial observations*. SL and BL consider a conditional quasi maximum likelihood (CQML) approach treating y_0 as exogenously given for the estimation of Model (2.4) assuming $W_{1t} = W_{2t} = W$ and $W_{3t} = \tilde{W}$ with W, $\tilde{W}$ and $\{x_t\}$ being exogenously given. The

CQML estimation ignores the information contained in y_0 about the common parameters and therefore will be inconsistent when T is fixed (the IVP, see Nickel 1981). Even when both n and T are large, valid statistical inferences depend on a successful bias correction on the CQML estimators to remove the first-order biases caused by both IVP and the estimation of Γ and $\{f_t\}$ (the IPP of Neyman and Scott, 1948). BL allow for cross-sectional heteroskedasticity explicitly and estimates the individual variances along with the common parameters. SL's assume homoskedasticity and their inference methods depends critically on the perturbation theory that hinders the extension to allow for heteroskedasticity as commented by BL.

The advantages of KP's FOD-based GMM approach are (i) it offers an easy way to avoid the effect of IVP, (ii) it allows for sequential exogeneity (of an unknown form) in spatial weight matrices and regressors, and (iii) it avoids the IPP by eliminating factor loadings through an innovative FOD transformation. Both (i) and (ii) are realized through skillful choices of instrumental variables. KP's GMM is limited to small and fixed T and allows cross-sectional heteroskedasticity to be a function of finite number of parameters.

As discussed in the introduction, likelihood-based methods with T fixed or T large but small relative to n have not been given.[4] We do so in this paper by introducing M-estimation and inference methods. We further extend the methods to allow for time-varying spatial weights, high-order spatial effects, and unknown cross-sectional heteroskedasticity. A distinguishing feature of our approach is that we derive a (minimum) set of unbiased and consistent moment conditions from the conditional concentrated quasi scores. From GMM perspective, likelihood-based approach can be motivated as a way of reducing the number of moments available for estimation, and hence the extent of bias . . . (Alvarez and Arellano, 2022).

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 $\mathtt{vec}(\cdot)$ vectorizes a matrix; $\otimes$ denotes the Kronecker product; $\|\cdot\|$ denotes the Frobenius norm, $\|\cdot\|_{\mathtt{sp}}$ the spectrum norm, $\|\cdot\|_1$ 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.

3. M-Estimation and Inference: Basic DSPD-IFE Model

For ease of exposition and to fix ideas, we start with a basic model, which is Model (2.4) with $W_\nu, \nu = 1, 2, 3$, being time-invariant and exogenously given; $\{v_{it}\}$ being independent and identically distributed (iid) across i and t, i.e. $v_{it} \sim iid(0, \sigma_v^2)$; and $\{x_t\}$ being $n \times k$ matrices of time-varying exogenous variables. The first two assumptions will be relaxed in Sec. 4, where

[4] Alvarez and Arellano (2022) commented: the GMM is routinely employed in the estimation of autoregressive models for short panels, because it provides simple estimates that are fixed-T consistent and optimally enforce the model's restrictions on the data covariance matrix. Yet they are known to frequently exhibit poor properties in finite samples and may be asymptotically biased if T is not treated as fixed.

the model is further extended to allow higher-order spatial and dynamic effects.

In the model, ρy_{t-1} captures the time dynamic effects; 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.

3.1. CQML estimation

Define $B_\nu(\lambda_\nu) = I_n - \lambda_\nu W_\nu, \nu = 1, 3$, and $B_2(\rho, \lambda_2) = \rho I_n + \lambda_2 W_2$. Let $\theta = (\beta', \rho, \lambda_1, \lambda_2)'$, and $\psi = (\theta', \lambda_3, \sigma_v^2)'$, and $F' = (f_1, \ldots, f_T)$. The **quasi** Gaussian loglikelihood function treating y_0 as exogenously given, or the conditional quasi loglikelihood (CQL) function, is:

$$
\begin{aligned}
\ell_{nT}(\psi, \Gamma, F) = &- \tfrac{nT}{2} \log(2\pi\sigma_v^2) - \tfrac{T}{2} \log|\Omega(\lambda_3)| + T \log|B_1(\lambda_1)| \\
&- \tfrac{1}{2\sigma_v^2} \textstyle\sum_{t=1}^{T} [z_t(\theta) - \Gamma f_t]' \Omega^{-1}(\lambda_3) [z_t(\theta) - \Gamma f_t] \qquad (3.1)\\
= &- \tfrac{nT}{2} \log(2\pi\sigma_v^2) + T \log|B_3(\lambda_3)| + T \log|B_1(\lambda_1)| \\
&- \tfrac{1}{2\sigma_v^2} \mathrm{tr}[(\mathbb{Z}(\theta) - \Gamma F')' \Omega^{-1}(\lambda_3)(\mathbb{Z}(\theta) - \Gamma F')], \qquad (3.2)
\end{aligned}
$$

where $z_t(\theta) = B_1(\lambda_1) y_t - B_2(\rho, \lambda_2) y_{t-1} - x_t \beta$, $\mathbb{Z}(\theta) = [z_1(\theta), z_2(\theta), \ldots, z_T(\theta)]$, 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 ψ.[5]

Solving the first order condition, $\frac{\partial}{\partial \Gamma} \ell_{nT}(\psi, \Gamma, F) = 0$, using (3.2),[6] we obtain the constrained CQML estimator of Γ as a matrix function of θ and F:

$$
\tilde{\Gamma}(\theta, F) = \mathbb{Z}(\theta) F (F'F)^{-1}. \qquad (3.3)
$$

With $\mathbb{Z}(\theta) - \tilde{\Gamma}(\theta, F) F' = \mathbb{Z}(\theta) - \mathbb{Z}(\theta) F (F'F)^{-1} F' \equiv \mathbb{Z}(\theta) M_F$, where $M_F = I_T - F(F'F)^{-1} F'$, plugging $\tilde{\Gamma}(\theta, F)$ in $\ell_{nT}(\psi, \Gamma, F)$ gives the concentrated CQL (CCQL) function of ψ and F:

$$
\begin{aligned}
\ell_{nT}^c(\psi, F) = &- \tfrac{nT}{2} \log(2\pi\sigma_v^2) + T \log|B_3(\lambda_3)| + T \log|B_1(\lambda_1)| \\
&- \tfrac{1}{2\sigma_v^2} \mathrm{tr}[M_F \mathbb{Z}'(\theta) \Omega^{-1}(\lambda_3) \mathbb{Z}(\theta)]. \qquad (3.4)
\end{aligned}
$$

Maximizing the CCQL $\ell_{nT}^c(\psi, F)$ gives the CQML estimators of ψ and F subject to the constraints imposed on F (details to be given later), and hence the CQML estimator of Γ.

[5]Under $W_1 = W_2$, SL show that $\hat{\psi}_{\texttt{CQML}}$ is consistent only when $(n, T) \to \infty$, and that $\sqrt{nT}(\hat{\psi}_{\texttt{CQML}} - \psi_0)$ has a non-zero asymptotic mean and a bias correction (BC) has to be made for proper inference. BL propose a BC-CQML estimation of a simpler model but allowing explicitly the cross-sectional heteroskedasticity.

[6]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.

3.2. M-estimation with fixed T

To facilitate the derivation of unbiased and consistent estimating functions, it is convenient to use the $nT \times 1$ vector $\mathbf{Z}(\theta) = [z_1'(\theta), z_2'(\theta), \ldots, z_T'(\theta)]' = \mathtt{vec}(\mathbb{Z}(\theta))$. Working directly with (3.1) and (3.3), or using the identity $\mathrm{tr}[M_F \mathbb{Z}'(\theta)\Omega^{-1}(\lambda_3)\mathbb{Z}(\theta)] = \mathbf{Z}'(\theta)[M_F \otimes \Omega^{-1}(\lambda_3)]\mathbf{Z}(\theta)$ on (3.4),[7] the CCQL function can be written as

$$\begin{aligned} \ell_{nT}^c(\psi, F) = & -\tfrac{nT}{2}\log(2\pi\sigma_v^2) + T\log|B_3(\lambda_3)| + T\log|B_1(\lambda_1)| \\ & -\tfrac{1}{2\sigma_v^2}\mathbf{Z}'(\theta)[M_F \otimes \Omega^{-1}(\lambda_3)]\mathbf{Z}(\theta). \end{aligned} \tag{3.5}$$

The ψ-component of the concentrated conditional quasi score (CCQS) can be derived in a straightforward manner. For the F-component, we note that F enters the CCQL function (3.5) in the form of $P_F = F(F'F)^{-1}F'$. As a result, $\ell_{nT}^c(\psi, F)$ is 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 et al. (2013) and Kuersteiner and Prucha (2020), we normalize F as $(F^{*\prime}, I_r)'$, where F^* is a $(T-r)\times r$ matrix of unrestricted parameters.[9] Let $\phi = \mathtt{vec}(F^*)$ with elements $\phi_s, s = 1, \ldots, k_\phi$, where $k_\phi = \dim(\phi) = (T-r)r$. Denote the CCQL function by $\ell_{nT}^c(\psi, \phi)$. One can then derive the CCQS functions of ψ and ϕ.

Let $\mathbf{Y} = (y_1', y_2', \ldots, y_T')'$ and $\mathbf{Y}_{-1} = (y_0', y_1', \ldots, y_{T-1}')'$, 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. Let $\mathbf{W}_\nu = I_T \otimes W_\nu, \nu = 1, 2, 3$, $\mathbf{B}_\nu(\lambda_\nu) = I_T \otimes B_\nu(\lambda_\nu), \nu = 1, 3$, and $\mathbf{B}_2(\rho, \lambda_2) = I_T \otimes B_2(\rho, \lambda_2)$. Then, $\mathbf{Z}(\theta) = \mathbf{B}_1(\lambda_1)\mathbf{Y} - \mathbf{B}_2(\rho, \lambda_2)\mathbf{Y}_{-1} - \mathbf{X}\beta$. Denote $\mathbf{\Omega}(\lambda_3) = I_T \otimes \Omega(\lambda_3)$ and $\mathbf{M}_F = M_F \otimes I_n$. The CCQS functions of ψ and ϕ, $S_{nT}^c(\psi, \phi) = (\frac{\partial}{\partial \psi'}\ell_{nT}^c(\psi, \phi), \frac{\partial}{\partial \phi'}\ell_{nT}^c(\psi, \phi))'$, take the form:

$$S_{nT}^c(\psi, \phi) = \begin{cases} \frac{1}{\sigma_v^2}\mathbf{X}'\mathbf{M}_F\mathbf{\Omega}^{-1}(\lambda_3)\mathbf{Z}(\theta), \\ \frac{1}{\sigma_v^2}\mathbf{Z}'(\theta)\mathbf{M}_F\mathbf{\Omega}^{-1}(\lambda_3)\mathbf{Y}_{-1}, \\ \frac{1}{\sigma_v^2}\mathbf{Z}'(\theta)\mathbf{M}_F\mathbf{\Omega}^{-1}(\lambda_3)\mathbf{W}_1\mathbf{Y} - \mathrm{tr}[\mathbf{W}_1\mathbf{B}_1^{-1}(\lambda_1)], \\ \frac{1}{\sigma_v^2}\mathbf{Z}'(\theta)\mathbf{M}_F\mathbf{\Omega}^{-1}(\lambda_3)\mathbf{W}_2\mathbf{Y}_{-1}, \\ \frac{1}{\sigma_v^2}\mathbf{Z}'(\theta)\mathbf{M}_F\mathbf{B}_3'(\lambda_3)\mathbf{W}_3\mathbf{Z}(\theta) - \mathrm{tr}[\mathbf{W}_3\mathbf{B}_3^{-1}(\lambda_3)], \\ \frac{1}{2\sigma_v^4}\mathbf{Z}'(\theta)\mathbf{M}_F\mathbf{\Omega}^{-1}(\lambda_3)\mathbf{Z}(\theta) - \frac{nT}{2\sigma_v^2}, \\ \frac{1}{\sigma_v^2}\mathbf{Z}'(\theta)[M_F\dot{F}_s(F'F)^{-1}F' \otimes \Omega^{-1}(\lambda_3)]\mathbf{Z}(\theta), \quad s = 1, \ldots, k_\phi, \end{cases} \tag{3.6}$$

[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, $\mathrm{tr}(ABCD) = \mathtt{vec}(D')'(C' \otimes A)\mathtt{vec}(B) = \mathtt{vec}(D)'(A \otimes C')\mathtt{vec}(B')$.

[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 $\Gamma C C^{-1} F' = \Gamma F'$ for any $r \times r$ non-singular matrix C.

[9] This is obtained through the rotation. Denote $F = (F_1', F_2')'$ with F_2 being $r \times r$ and invertible, and take $C = F_2^{-1}$. Then, $FC = FF_2^{-1} = (F_2'^{-1}F_1', I_r)'$, and therefore $F^* = F_1F_2^{-1}$. Ahn et al. (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.

where $\dot{F}_s = \frac{\partial}{\partial \phi_s} F$, a $T \times r$ matrix with elements 1 at the ϕ_s-position and 0 elsewhere. Under mild conditions, maximizing (3.4) w.r.t. ψ and ϕ is equivalent to solving $S_{nT}^c(\psi, \phi) = 0$.

However, we show that the $(\sigma_v^2, \rho, \lambda)$ components of $\lim_{n\to} \frac{1}{nT}\mathrm{E}[S_{nT}^c(\psi_0, \phi_0)]$ and more seriously these of $\mathrm{plim}_{n\to\infty} \frac{1}{nT} S_{nT}^c(\psi_0, \phi_0)$ are generally not zero at the true ψ_0 and ϕ_0. Thus, the CQML estimator of (ψ, ϕ) cannot be consistent as a necessary condition for consistent estimation is violated. To see these, 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 (2.4) holds only at the true parameters, and the usual expectation and variance operators $\mathrm{E}(\cdot)$ and $\mathrm{Var}(\cdot)$ 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 $\mathcal{B}_0 = \mathcal{B}(\rho_0, \lambda_{10}, \lambda_{20}) \equiv B_1^{-1}(\lambda_{10}) B_2(\rho_0, \lambda_{20})$. The first equation of (2.4) under time-invariance of W_ν is written as $y_t = \mathcal{B}_0 y_{t-1} + B_{10}^{-1} x_t \beta_0 + B_{10}^{-1} z_t$. Backward substitution gives

$$y_t = \mathcal{B}_0^t y_0 + \sum_{s=0}^{t-1} \mathcal{B}_0^s B_{10}^{-1} x_{t-s} \beta_0 + \sum_{s=0}^{t-1} \mathcal{B}_0^s B_{10}^{-1} z_{t-s}, \quad t = 1, \ldots, T. \tag{3.7}$$

This leads to the following simple but important representations for $\mathbf{Y}$ and $\mathbf{Y}_{-1}$:

$$\mathbf{Y} = \mathbf{Q}\mathbf{y}_0 + \boldsymbol{\eta} + \mathbf{D}\mathbf{Z} \quad \text{and} \quad \mathbf{Y}_{-1} = \mathbf{Q}_{-1}\mathbf{y}_0 + \boldsymbol{\eta}_{-1} + \mathbf{D}_{-1}\mathbf{Z}, \tag{3.8}$$

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

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

Based on the representation (3.8), we obtain under Assumption A with $\{v_{it}\}$ being iid,

$$\mathrm{E}[S_{nT}^c(\psi_0, \phi_0)] = \begin{cases} 0_k, \\ \mathrm{tr}(\mathbf{M}_{F_0}\mathbf{D}_{-1}), \\ \mathrm{tr}(\mathbf{M}_{F_0}\mathbf{W}_1\mathbf{D}) - \mathrm{tr}(\mathbf{W}_1\mathbf{B}_{10}^{-1}), \\ \mathrm{tr}(\mathbf{M}_{F_0}\mathbf{W}_2\mathbf{D}_{-1}), \\ \mathrm{tr}(\mathbf{M}_{F_0}\mathbf{W}_3\mathbf{B}_{30}^{-1}) - \mathrm{tr}(\mathbf{W}_3\mathbf{B}_{30}^{-1}), \\ \frac{n(T-r)}{2\sigma_{v0}^2} - \frac{nT}{2\sigma_{v0}^2}, \\ 0_{k_\phi}, \end{cases} \tag{3.9}$$

where 0_m denotes an $m \times 1$ vector of zeros, and some details on the ϕ-part are given at the end of this subsection. The result of (3.9) clearly reveals that $\frac{1}{nT}\mathrm{E}[S^c_{nT}(\psi_0, \phi_0)] \neq 0$ and does not even converge to 0 when only n approaches to ∞, and therefore $\mathrm{plim}_{n\to\infty}\frac{1}{nT}S^c_{nT}(\psi_0, \phi_0) \neq 0$.

Note that $\mathrm{E}[S^c_{nT}(\psi_0, \phi_0)]$ is a parametric vector **free** from initial conditions, process starting time and factor loadings. Therefore, it can be used to adjust (3.6) to give a set of *adjusted quasi score* (AQS) functions or EFs for (ψ, ϕ), free from m, Γ and the conditions on y_0:

$$S^*_{nT}(\psi, \phi) = \begin{cases} \frac{1}{\sigma_v^2}\mathbf{X}'\mathbf{M}_F\mathbf{\Omega}^{-1}(\lambda_3)\mathbf{Z}(\theta), \\ \frac{1}{\sigma_v^2}\mathbf{Z}'(\theta)\mathbf{M}_F\mathbf{\Omega}^{-1}(\lambda_3)\mathbf{Y}_{-1} - \mathrm{tr}[\mathbf{M}_F\mathbf{D}_{-1}(\rho, \lambda_1, \lambda_2)], \\ \frac{1}{\sigma_v^2}\mathbf{Z}'(\theta)\mathbf{M}_F\mathbf{\Omega}^{-1}(\lambda_3)\mathbf{W}_1\mathbf{Y} - \mathrm{tr}[\mathbf{M}_F\mathbf{W}_1\mathbf{D}(\rho, \lambda_1, \lambda_2)], \\ \frac{1}{\sigma_v^2}\mathbf{Z}'(\theta)\mathbf{M}_F\mathbf{\Omega}^{-1}(\lambda_3)\mathbf{W}_2\mathbf{Y}_{-1} - \mathrm{tr}[\mathbf{M}_F\mathbf{W}_2\mathbf{D}_{-1}(\rho, \lambda_1, \lambda_2)], \\ \frac{1}{\sigma_v^2}\mathbf{Z}'(\theta)\mathbf{M}_F\mathbf{B}_3'(\lambda_3)\mathbf{W}_3\mathbf{Z}(\theta) - \mathrm{tr}[\mathbf{M}_F\mathbf{W}_3\mathbf{B}_3^{-1}(\lambda_3)], \\ \frac{1}{2\sigma_v^4}\mathbf{Z}'(\theta)\mathbf{M}_F\mathbf{\Omega}^{-1}(\lambda_3)\mathbf{Z}(\theta) - \frac{n(T-r)}{2\sigma_v^2}, \\ \frac{1}{\sigma_v^2}\mathbf{Z}'(\theta)[M_F\dot{F}_s(F'F)^{-1}F' \otimes \Omega^{-1}(\lambda_3)]\mathbf{Z}(\theta), \quad s = 1, \ldots, k_\phi. \end{cases} \tag{3.10}$$

Clearly, $\mathrm{E}[S^*_{nT}(\psi_0, \phi_0)] = 0$. One can further show that $\mathrm{plim}_{n\to\infty}\frac{1}{nT}S^*_{nT}(\psi_0, \phi_0) = 0$. Thus, $S^*_{nT}(\psi, \phi)$ gives a set of unbiased and consistent estimating functions, which paves the way for a consistent estimation of ψ and ϕ. Our **AQS or M-estimators** $\hat{\psi}_{\mathtt{M}}$ and $\hat{\phi}_{\mathtt{M}}$ of ψ and ϕ are therefore defined as the solution of the estimating equations: $S^*_{nT}(\psi, \phi) = 0$.

A computational note. Given ψ, Model (2.4) reduces to a pure factor model. The constrained M-estimator of F or ϕ can be obtained by maximizing $\frac{1}{nT}\mathrm{tr}[P_F Z'(\theta)\Omega^{-1}(\lambda_3)Z(\theta)]$,[10] and the solution is the eigenvector matrix of $\frac{1}{nT}Z'(\theta)\Omega^{-1}(\lambda_3)Z(\theta)$ corresponding to the r largest eigenvalues.[11] Denoting the ψ-component of $S^*_{nT}(\psi, \phi)$ by $S^*_{nT,\psi}(\psi, F)$, the computation of the M-estimators can simply be done as follows:

1. *Given F, compute the estimator of ψ: $\hat{\psi}(F) = \arg\{S^*_{nT,\psi}(\psi, F) = 0\}$,*
2. *Given ψ, compute the estimator of F: $\widehat{F}(\psi)$, which is the matrix of eigenvectors corresponding to the r largest eigenvalues of the $T \times T$ matrix $\frac{1}{nT}\mathbb{Z}'(\theta)\Omega^{-1}(\lambda_3)\mathbb{Z}(\theta)$,*[12]
3. *Iterate between 1. and 2. until convergence, to give $\hat{\psi}_{\mathtt{M}}$ and $\hat{\phi}_{\mathtt{M}} = \mathtt{vec}(\widehat{F}_1(\hat{\psi}_{\mathtt{M}})\widehat{F}_2^{-1}(\hat{\psi}_{\mathtt{M}}))$.*

See Footnote 9, Kiefer (1980), Ahn, et al. (2001, 2013), and Bai (2009), for more discussions.

The root-finding process in Step 1 can be further simplified. First solving the first two sets of equations for β and σ^2, we obtain analytical solutions in terms of $\delta = (\rho, \lambda', \phi')'$:

$$\hat{\beta}(\delta) = [\mathbf{X}'\mathbf{M}_F\mathbf{\Omega}^{-1}(\lambda_3)\mathbf{X}]^{-1}\mathbf{X}'\mathbf{M}_F\mathbf{\Omega}^{-1}(\lambda_3)[\mathbf{B}_1(\lambda_1)\mathbf{Y} - \mathbf{B}_2(\lambda_2)\mathbf{Y}_{-1}], \text{ and} \tag{3.11}$$

$$\hat{\sigma}_v^2(\delta) = \frac{1}{n(T-r)}\hat{\mathbf{Z}}'(\delta)\mathbf{M}_F\mathbf{\Omega}^{-1}(\lambda_3)\hat{\mathbf{Z}}(\delta), \tag{3.12}$$

[10]This is equivalent to the objective function of the least square estimation of a pure factor model, $B_3Z = B_3\Gamma F' + \mathbb{V}$, after the factor loadings Γ being concentrated out, where $\mathbb{V} = (v_1, \ldots, v_T)$. See, e.g., Bai (2009).

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

[12]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 $\Gamma' B_{30}' B_{30}\Gamma/n$ and $\mathbb{V}'\mathbb{V}/n$, respectively. If $\Sigma_v = \sigma_{v0}^2 I_T$, 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.

where $\hat{\mathbf{Z}}(\delta) = \mathbf{B}_1(\lambda_1)\mathbf{Y} - \mathbf{B}_2(\lambda_2)\mathbf{Y}_{-1} - \mathbf{X}\hat{\beta}(\delta)$. Substituting $\hat{\beta}(\delta)$ and $\hat{\sigma}_v^2(\delta)$ back into the (ρ, λ)-components of $S^*_{nT,\psi}(\psi, \phi)$ gives the concentrated AQS function (detailed expression is given in Appendix B). Then, solving the concentrated AQS equations gives the constrained (given F) estimators of ρ and λ, and thus the constrained estimators (given F) of β and σ^2.

Before moving to the study of the asymptotic properties of the proposed M-estimator, some important remarks on the proposed M-estimation strategy are as follows.

Remark 3.1. *The proposed method is likelihood-based, and also the method of moments under just identified situation. From a GMM perspective, likelihood-based estimation can be motivated as a way of reducing the number of moments available for the estimation, and hence the extent of bias in second-order or double asymptotics (Alvarez and Arellano, 2022).* Missing in the ref list

Remark 3.2. *The importance of the joint EF, $S^*_{nT}(\psi, \phi)$, also lies in the fact that it leads to a simple way to establish the joint asymptotic distribution of $\hat{\psi}_{\mathtt{M}}$ and $\hat{\phi}_{\mathtt{M}}$, and a simple and reliable way to obtain the VC matrix estimate as seen in the subsequent sections.*

Remark 3.3. *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.*[13] *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 cross-sectional heteroskedasticity. See Section 4 for details.*

Remark 3.4. *When $\Gamma f_t = \gamma + f_t 1_n$ where γ is an $n \times 1$ vector and f_t is a scalar, we have a DSPD model with additive fixed effects. In this case, our method provides an alternative to Yang (2018). The advantage of our method is that it does not require a transformation to eliminate γ and thus can accommodate time-varying spatial weights. See Section 4 for details.*

Remark 3.5. *Setting $\lambda_1 = \lambda_2 = \lambda_3 = 0$ and $F = 1_T$, Model* (2.4) *reduces to a regular dynamic panel data model with individual FE only, and our M-estimator reduces to the bias-corrected conditional score estimator under small-T proposed by Alvarez and Arellano (2022).*

Finally, it is useful to give some details for the ϕ-component of $S^*_{nT}(\psi, \phi)$. With $\dot{F}_s$ defined in (3.6), 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

$$\begin{aligned} \frac{\partial}{\partial \phi_s} \ell^c_{nT}(\psi, \phi) &= \frac{1}{2\sigma_v^2} \mathbf{Z}'(\theta)[\dot{P}_{F,s} \otimes \Omega^{-1}(\lambda_3)]\mathbf{Z}(\theta) \\ &= \frac{1}{\sigma_v^2} \mathbf{Z}'(\theta)[M_F \dot{F}_s (F'F)^{-1} F' \otimes \Omega^{-1}(\lambda_3)]\mathbf{Z}(\theta). \end{aligned} \tag{3.13}$$

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

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

assumptions on the errors, we have, for $s = 1, \ldots, k_\phi$, noting that $F_0' M_{F_0} = 0$,

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

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

3.3. Asymptotic properties of M-estimator with fixed T

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. F_0 has full column rank.*

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 $\nu = 1, 2, 3$, the elements $w_{\nu,ij}$ of W_ν are at most of order h_n^{-1}, uniformly in all i and j, and $w_{\nu,ii} = 0$ for all i; (ii) $h_n/n \to 0$ as $n \to \infty$; (iii) $\{W_\nu, \nu = 1, 2, 3\}$ and $\{B_{\nu 0}^{-1}, \nu = 1, 3\}$ are uniformly bounded in both row and column sum norms; (iv) For $B_\nu = B_\nu(\lambda_\nu)$ with $\nu = 1, 3$, either $\|B_\nu^{-1}\|_\infty$ or $\|B_\nu^{-1}\|_1$ is bounded, uniformly in λ_ν in a compact parameter space $\mathbf{\Lambda}_\nu$, and $0 < \underline{c}_\nu \le \inf_{\lambda_\nu \in \mathbf{\Lambda}_\nu}\gamma_{\min}(B_\nu' B_\nu) \le \sup_{\lambda_\nu \in \mathbf{\Lambda}_\nu}\gamma_{\max}(B_\nu' B_\nu) \le \bar{c}_\nu < \infty$.*

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 $h_n^{-1/2}$, (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 context as they are dealt with by the spatial lag, time lag, space-time lag, spatial error terms. Assumption C(i) is standard for establishing the consistency of the M-estimator $\hat{\delta}_{\mathtt{M}}$ of δ. The consistency of $\hat{\beta}_{\mathtt{M}}$ and $\hat{\sigma}^2_{v,\mathtt{M}}$ follows from that of $\hat{\delta}_{\mathtt{M}}$ 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 to an important spatial layout where the *degree of spatial dependence* increases with n, see Lee (2004) and Yang (2015) for related discussions. Assumption F is of low level, to ensure the initial observations to have a proper stochastic behavior. It is satisfied if the process has evolved according to (2.4) since it started and if $\sum_{i=0}^{\infty} \mathcal{B}_0^i$ exists and is uniformly bounded in both row and column sums, as in Yu et al. (2008) and Lee and Yu (2014).

Given δ, solving the AQS equations for β and σ_v^2 from (3.10), we obtain the constrained M-estimators $\hat{\beta}(\delta)$ and $\hat{\sigma}_v^2(\delta)$ as in (3.11) and (3.12). Now, substituting $\hat{\beta}(\delta)$ and $\hat{\sigma}_v^2(\delta)$ back into the δ-component of $S_{nT}^*(\psi,\phi)$ gives the concentrated AQS function $S_{nT}^{*c}(\delta)$ (detailed expression is given in Appendix B). Similarly, let $\bar{S}_{nT}^{*c}(\delta)$ be the population counterpart of the concentrated AQS function (see Appendix B). It is easy to see that $S_{nT}^{*c}(\hat{\delta}_{\mathtt{M}}) = \mathbf{0}$, and $\bar{S}_{nT}^{*c}(\delta_0) = \mathbf{0}$. By Theorem 5.9 of van der Vaart (1998), $\hat{\delta}_{\mathtt{M}}$ will be consistent for δ_0 if $\sup_{\delta\in\mathbf{\Delta}} \frac{1}{nT}\big\|S_{nT}^{*c}(\delta) - \bar{S}_{nT}^{*c}(\delta)\big\| \xrightarrow{p} 0$, and the following identification condition holds.

Assumption G. $\inf_{\delta\ d(\delta,\delta)\geq\varepsilon} \big\|\bar{S}_{nT}^{*c}(\delta)\big\| > 0$ *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)* $\gamma_{\max}[\mathrm{Var}(\mathbf{Y})]$ *and* $\gamma_{\max}[\mathrm{Var}(\mathbf{Y}_{-1})]$ *are bounded, and (ii)* $\inf_{\delta\in\mathbf{\Delta}} \gamma_{\min}\big[\mathrm{Var}(\mathbf{B}_1\mathbf{Y} - \mathbf{B}_2\mathbf{Y}_{-1})\big] \geq \underline{c}_y > 0$*. 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* $\hat{\sigma}_{v,\mathtt{M}}^2 \xrightarrow{p} \sigma_{v0}^2$*.*

Let $\boldsymbol{\psi} = (\psi', \phi')'$. To establish joint asymptotic normality of $\hat{\boldsymbol{\psi}}_{\mathtt{M}}$, we have by (3.8) at $\boldsymbol{\psi}_0$,

$$
S_{nT}^*(\boldsymbol{\psi}_0) = \begin{cases} \Pi_1'\mathbf{Z} \\ \mathbf{Z}'\Psi_1\mathbf{y}_0 + \mathbf{Z}'\Phi_1\mathbf{Z} + \Pi_2'\mathbf{Z} - \mu_{\rho_0}, \\ \mathbf{Z}'\Psi_2\mathbf{y}_0 + \mathbf{Z}'\Phi_2\mathbf{Z} + \Pi_3'\mathbf{Z} - \mu_{\lambda_{10}}, \\ \mathbf{Z}'\Psi_3\mathbf{y}_0 + \mathbf{Z}'\Phi_3\mathbf{Z} + \Pi_4'\mathbf{Z} - \mu_{\lambda_{20}}, \\ \mathbf{Z}'\Phi_4\mathbf{Z} - \mu_{\lambda_{30}}, \\ \mathbf{Z}'\Phi_5\mathbf{Z} - \mu_{\sigma_{v0}^2}, \\ \mathbf{Z}'\Phi_{5+s}\mathbf{Z},\ s = 1,2,\ldots,k_\phi, \end{cases} \tag{3.14}
$$

where $\Pi_1 = \frac{1}{\sigma_{v0}^2}(M_{F_0}\otimes\Omega_0^{-1})\mathbf{X}$, $\Pi_2 = \frac{1}{\sigma_{v0}^2}(M_{F_0}\otimes\Omega_0^{-1})\boldsymbol{\eta}_{-1}$, $\Pi_3 = \frac{1}{\sigma_{v0}^2}(M_{F_0}\otimes\Omega_0^{-1})\mathbf{W}_1\boldsymbol{\eta}$, and $\Pi_4 = \frac{1}{\sigma_{v0}^2}(M_{F_0}\otimes\Omega_0^{-1})\mathbf{W}_2\boldsymbol{\eta}_{-1}$; $\Phi_1 = \frac{1}{\sigma_{v0}^2}(M_{F_0}\otimes\Omega_0^{-1})\mathbf{D}_{-1}$, $\Phi_2 = \frac{1}{\sigma_{v0}^2}(M_{F_0}\otimes\Omega_0^{-1})\mathbf{W}_1\mathbf{D}$, $\Phi_3 = \frac{1}{\sigma_{v0}^2}(M_{F_0}\otimes\Omega_0^{-1})\mathbf{W}_2\mathbf{D}_{-1}$, $\Phi_4 = \frac{1}{\sigma_{v0}^2}(M_{F_0}\otimes B_{30}'W_3)$, $\Phi_5 = \frac{1}{2\sigma_{v0}^4}(M_{F_0}\otimes\Omega_0^{-1})$, and $\Phi_{5+s} = \frac{1}{\sigma_{v0}^2}[M_{F_0}\dot{F}_{s0}(F_0'F_0)^{-1}F_0'\otimes\Omega_0^{-1}]$, $s = 1,\ldots,k_\phi$; $\Psi_1 = \frac{1}{\sigma_{v0}^2}(M_{F_0}\otimes\Omega_0^{-1})\mathbf{Q}_{-1}$, $\Psi_2 = \frac{1}{\sigma_{v0}^2}(M_{F_0}\otimes\Omega_0^{-1})\mathbf{W}_1\mathbf{Q}$, and $\Psi_3 = \frac{1}{\sigma_{v0}^2}(M_{F_0}\otimes\Omega_0^{-1})\mathbf{W}_2\mathbf{Q}_{-1}$; $\mu_{\sigma_v^2} = \frac{n(T-r)}{2\sigma_v^2}$, $\mu_\rho = \mathrm{tr}(\mathbf{M}_{F_0}\mathbf{D}_{-1})$, $\mu_{\lambda_1} = \mathrm{tr}(\mathbf{M}_{F_0}\mathbf{W}_1\mathbf{D})$, $\mu_{\lambda_2} = \mathrm{tr}(\mathbf{M}_{F_0}\mathbf{W}_2\mathbf{D}_{-1})$, and $\mu_{\lambda_3} = \mathrm{tr}(\mathbf{M}_{F_0}\mathbf{W}_3\mathbf{B}_{30}^{-1})$.

Using the relation $\mathbf{Z} = \mathbf{B}_{30}^{-1}\mathbf{v} + \mathtt{vec}(\Gamma_0F_0')$, the AQS vector at true $\boldsymbol{\psi}_0$, $S_{nT}^*(\boldsymbol{\psi}_0)$, is further expressed as linear combinations of terms linear or quadratic in $\mathbf{v}$ and bilinear in $\mathbf{v}$ and $\mathbf{y}_0$; see (B.7). This leads to a simple way for establishing the asymptotic normality of $S_{nT}^*(\boldsymbol{\psi}_0)$, and the asymptotic normality of $\hat{\psi}_{\mathtt{M}}$ through a first-order expansion of $S^*_{nT}(\hat{\psi}_{\mathtt{M}})$ at ψ_0.

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

$$\sqrt{nT}\left(\hat{\psi}_{\mathtt{M}} - \psi_0\right) \xrightarrow{D} N\left(0, \lim_{n\to\infty} H^{-1}_{nT}(\psi_0)\Sigma_{nT}(\psi_0)H'^{-1}_{nT}(\psi_0)\right),$$

where $H_{nT}(\psi_0) = -\frac{1}{nT}\mathrm{E}[\frac{\partial}{\partial \psi'}S^*_{nT}(\psi_0)]$ and $\Sigma_{nT}(\psi_0) = \frac{1}{nT}\mathrm{Var}[S^*_{nT}(\psi_0)]$, both assumed to exist and $H_{nT}(\psi_0)$ to be positive definite, for sufficiently large n.

3.4. Robust VC matrix estimation with fixed T

While Theorems 3.1 and 3.2 provide theoretical foundations for fixed-T inferences based on the DSPD-IFE model, empirical applications of the results depend on the availability of consistent estimators of the two matrices $H_{nT}(\psi_0)$ and $\Sigma_{nT}(\psi_0)$. The former can be consistently estimated by its observed counterpart, $H_{nT}(\hat{\psi}_{\mathtt{M}}) = -\frac{1}{nT}\frac{\partial}{\partial \psi'}S^*_{nT}(\hat{\psi}_{\mathtt{M}})$. The analytical expression of $\frac{\partial}{\partial \psi'}S^*_{nT}(\psi)$ is given in Appendix B. Unfortunately, the estimation of the latter is not straightforward. From (3.14) we see that the joint AQS function $S^*_{nT}(\psi_0)$ contains three types of elements, $\Pi'\mathbf{Z}$, $\mathbf{Z}'\Psi\mathbf{y}_0$, and $\mathbf{Z}'\Phi\mathbf{Z}$, where Π, Ψ and Φ are non-stochastic vectors or matrices. The traditional plug-in method requires the closed-form expression of $\Sigma_{nT}(\psi_0)$, but the variance of $\mathbf{Z}'\Psi\mathbf{y}_0$ and its covariances with $\Pi'\mathbf{Z}$ and $\mathbf{Z}'\Phi\mathbf{Z}$ 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,[14] and a consistent estimate of the $n \times r$ matrix Γ_0 is impossible to obtain when T is fixed. Thus, the plug-in method based on the analytical expression of $\Sigma_{nT}(\psi_0)$ 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 DSPD-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 DSPD-IFE model due to the fact that the original errors v_t are not estimable,[15] the idea of decomposition prevails!

Inspired by the OPMD method, we decompose the AQS function as $S^*_{nT}(\psi_0) = \sum_{i=1}^n \mathbf{g}_i$, where $\{\mathbf{g}_i\}$ are defined in terms of z_{it} and some nonstochastic quantities that depend on ψ_0 and $W_\nu, \nu = 1, 2, 3$. Based on this, a feasible estimator of $\Sigma_{nT}(\psi_0)$ may be obtained through the following, taking the advantage that $\{\mathbf{g}_i\}$ are nearly an MD sequence and are 'estimable':

$$\Sigma_{nT}(\psi_0) = \frac{1}{nT}\mathrm{E}[S^*_{nT}(\psi_0)S^{*\prime}_{nT}(\psi_0)] = \frac{1}{nT}\sum_{i=1}^n \mathrm{E}(\mathbf{g}_i\mathbf{g}_i') + \frac{1}{nT}\sum_{i=1}^n\sum_{j\neq i}\mathrm{E}(\mathbf{g}_i\mathbf{g}_j'). \quad (3.15)$$

[14] A valid model for y_0, as that in Su and Yang (2015) for an DSPD 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).

[15] This is seen from the relation $z_t = B_3^{-1}v_t + \Gamma f_t$, 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.

The first term in (3.15) 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 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

$$\hat{\Sigma}_{nT}=\tfrac{1}{nT}\textstyle\sum_{i=1}^{n}\hat{\mathbf{g}}_i\hat{\mathbf{g}}_i'+\tfrac{1}{nT}\Upsilon(\hat{\boldsymbol{\psi}}_{\mathtt{M}}). \tag{3.16}$$

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 $\Pi'\mathbf{Z}$, $\mathbf{Z}'\Psi\mathbf{y}_0$, and $\mathbf{Z}'\Phi\mathbf{Z}$, 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 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.

First, consider a linear term $\Pi'\mathbf{Z}$.[16] Write $\Pi'\mathbf{Z}=\Pi^{*\prime}\mathbf{v}+\Pi'\mathtt{vec}(\Gamma_0F_0')$, where $\Pi^*=\mathbf{B}_3^{-1\prime}\Pi$. From (3.14), we see that Π takes the form $\mathbf{M}_{F_0}K$ for a suitably defined nonstochastic vector K involving $\boldsymbol{\psi}_0$, $\mathbf{X}$, or $W_r, r=1,2,3$. Therefore, the second term of $\Pi'\mathbf{Z}$ equals 0. This is seen as follows. 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, $\Pi'\mathtt{vec}(\Gamma_0F_0')=K'(M_{F_0}\otimes I_n)\mathtt{vec}(\Gamma_0F_0')=\mathrm{tr}(\Gamma_0F_0'M_{F_0}\mathbb{K}')=0$. Thus, $\Pi'\mathbf{Z}=\Pi^{*\prime}\mathbf{v}$. This leads to the following decomposition for any Π term defined in (3.14):

$$\Pi'\mathbf{Z}=\Pi^{*\prime}\mathbf{v}=\textstyle\sum_{i=1}^{n}(\sum_{t=1}^{T}\Pi_{it}^{*}v_{it})\equiv\sum_{i=1}^{n}g_{\Pi,i}, \tag{3.17}$$

where Π_{it}^{*} is the ith element of Π_t^{*}. Clearly, $\{g_{\Pi,i}\}$ are uncorrelated under this decomposition and it is easy to see that they constitute an MD sequence.

Next, consider a bilinear term $\mathbf{Z}'\Psi\mathbf{y}_0$, which can be separated into $\mathbf{Z}'\Psi\mathbf{y}_0=\mathbf{v}'\Psi^{*}\mathbf{y}_0+\mathtt{vec}(\Gamma_0F_0')'\Psi\mathbf{y}_0$, where $\Psi^{*}=\mathbf{B}_3^{-1\prime}\Psi$. Similarly, the second term equals zero,[17] and thus $\mathbf{Z}'\Psi\mathbf{y}_0=\mathbf{v}'\Psi^{*}\mathbf{y}_0$. With $\mathrm{E}(\mathbf{v}'\Psi^{*}\mathbf{y}_0)=0$ due to the independence between y_0 and $\{v_t, t\geq 1\}$,

[16] Without loss of generality, assume Π is a vector $(nT\times 1)$, as if not we can work on each column of it.

[17] By the expressions of Ψ given in (3.14), 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 $W_r, r=1,2,3$.

we have the following MD decomposition of a bilinear term for any Ψ defined in (3.14):

$$Z'\Psi \mathbf{y}_0 = \mathbf{v}'\Psi^* \mathbf{y}_0 = \textstyle\sum_{i=1}^n \sum_{t=1}^T v_{it}\xi_{it} \equiv \sum_{i=1}^n g_{\Psi,i}, \tag{3.18}$$

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 $\mathbf{Z}'\Phi\mathbf{Z}$, we separate the first $\mathbf{Z}$ into two parts to give $\mathbf{Z}'\Phi\mathbf{Z} = \mathbf{v}'\Phi^*\mathbf{Z} + \mathtt{vec}(\Gamma_0 F_0')'\Phi\mathbf{Z}$, where $\Phi^* = \mathbf{B}_3^{-1\prime}\Phi\mathbf{B}_3^{-1}$. Again, the second term equals zero.[18] Therefore, $\mathbf{Z}'\Phi\mathbf{Z} = \mathbf{v}'\Phi^*\mathbf{Z}$ and the latter can be decomposed for any Φ defined in (3.14) as,

$$\begin{aligned} \mathbf{v}'\Phi^*\mathbf{Z} = &\ \textstyle\sum_{t=1}^T \sum_{s=1}^T v_t'\Phi_{ts}^* z_s \\ = &\ \textstyle\sum_{t=1}^T \sum_{s=1}^T v_t'\Phi_{ts}^{*u} z_s + \sum_{t=1}^T \sum_{s=1}^T v_t'\Phi_{ts}^{*\ell} z_s + \sum_{t=1}^T \sum_{s=1}^T v_t'\Phi_{ts}^{*d} z_s \\ = &\ \textstyle\sum_{i=1}^n (\sum_{t=1}^T v_{it}\varphi_{it} + \sum_{t=1}^T v_{it} z_{it}^d), \end{aligned} \tag{3.19}$$

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:

$$\mathbf{v}'\Phi^*\mathbf{Z} - \mathrm{E}(\mathbf{v}'\Phi^*\mathbf{Z}) = \textstyle\sum_{i=1}^n [\sum_{t=1}^T v_{it}\varphi_{it} + \sum_{t=1}^T (v_{it} z_{it}^d - d_{it})] \equiv \sum_{i=1}^n g_{\Phi,i}. \tag{3.20}$$

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 (3.17)-(3.20) lead immediately to the decomposition $S_{nT}^*(\boldsymbol{\psi}_0) = \sum_{i=1}^n \mathbf{g}_i$, where

$$\mathbf{g}_i = \begin{cases} g_{\Pi_1,i} \\ g_{\Pi_2,i} + g_{\Phi_1,i} + g_{\Psi_1,i} \\ g_{\Pi_3,i} + g_{\Phi_2,i} + g_{\Psi_2,i} \\ g_{\Pi_4,i} + g_{\Phi_3,i} + g_{\Psi_3,i} \\ g_{\Phi_4,i} \\ g_{\Phi_5,i} \\ g_{\Phi_{5+s},i}, \ s = 1, 2, \ldots, k_\phi \end{cases} \tag{3.21}$$

$g_{\Pi_r,i}$ is defined according to (3.17) for each $\Pi_r, r = 1, 2, 3, 4$; $g_{\Psi_r,i}$ according to (3.18) for each $\Psi_r, r = 1, 2, 3$; and $g_{\Phi_r,i}$ according to (3.20) for each $\Phi_r, r = 1, 2, \ldots, 5 + k_\phi$, defined in (3.14).

This decomposition result in $\mathbf{g}_i$ that contains many uncorrelated terms. Moreover, this particular decomposition allow us to utilise the M_F structure embedded in Π, Ψ, and Φ, so that when we replace v_{it} by z_{it}, the factor component is cancelled out in sum of $\mathbf{g}_i'\mathbf{g}_i$. These features open up a simple way to consistently estimate the VC matrix of the AQS function. From its general form given in (3.15), 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 (3.21) by $\hat{z}_{it}$, and replacing $\boldsymbol{\psi}_0$ by $\hat{\boldsymbol{\psi}}_{\mathtt{M}}$. This is justified in the proof of Theorem 3.3, where we show that

[18] From (3.14), 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 $W_r, r = 1, 2, 3$.

$\sum_{i=1}^n \mathbf{g}_i^* \mathbf{g}_i^{*\prime} = \sum_{i=1}^n \mathbf{g}_i \mathbf{g}_i'$, where $\mathbf{g}_i^*$ is obtained by replacing v_{it} by z_{it} in (3.21).

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 (3.21) are all zero, for all r. First, by Assumptions A and B and the expressions (3.17) and (3.18), we show that $(g_{\Pi_r,i}, g_{\Psi_\nu,i})$ are uncorrelated, i.e., $\mathrm{E}(g_{\Pi_r,i} g_{\Pi_\nu,j})$, $\mathrm{E}(g_{\Psi_r,i} g_{\Psi_\nu,j})$ and $\mathrm{E}(g_{\Pi_r,i} g_{\Psi_\nu,j})$ are all zero, for $i \neq j$, $r = 1,2,3,4$, and $\nu = 1,2,3$. Next, by (3.17)-(3.20) and Assumptions A and B, we have, for $i \neq j$ $(= 1, \ldots, n)$,

$$\begin{aligned} &\mathrm{E}(g_{\Phi_r,i} g_{\Pi_\nu,j}) = \mathrm{E}\{[\textstyle\sum_{t=1}^T v_{it}\varphi_{r,it} + \sum_{t=1}^T (v_{it} z_{r,it}^d - d_{r,it})](\sum_{t=1}^T \Pi_{\nu,jt} v_{jt})\} \\ &= \mathrm{E}[(\textstyle\sum_{t=1}^T v_{it}\varphi_{r,it})(\sum_{t=1}^T \Pi_{\nu,jt} v_{jt})] + \mathrm{E}[\sum_{t=1}^T (v_{it} z_{r,it}^d - d_{r,it})(\sum_{t=1}^T \Pi_{\nu,jt} v_{jt})] = 0; \end{aligned} \tag{3.22}$$

$$\begin{aligned} &\mathrm{E}(g_{\Phi_r,i} g_{\Psi_\nu,j}) = \mathrm{E}\{[\textstyle\sum_{t=1}^T v_{it}\varphi_{r,it} + \sum_{t=1}^T (v_{it} z_{r,it}^d - d_{r,it})](\sum_{t=1}^T v_{jt}\xi_{\nu,jt})\} \\ &= \mathrm{E}[(\textstyle\sum_{t=1}^T v_{it}\varphi_{r,it})(\sum_{t=1}^T v_{jt}\xi_{\nu,jt})] + \mathrm{E}[\sum_{t=1}^T (v_{it} z_{r,it}^d - d_{r,it})(\sum_{t=1}^T v_{jt}\xi_{\nu,jt})] = 0. \end{aligned} \tag{3.23}$$

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 $i \neq j$,

$$\begin{aligned} \mathrm{E}(g_{\Phi_r,i} g_{\Phi_\nu,j}) &= \mathrm{E}[(\textstyle\sum_{t=1}^T v_{it}\varphi_{r,it})(\sum_{s=1}^T v_{js}\varphi_{\nu,jt})] \\ &= \textstyle\sum_{t=1}^T \sum_{s=1}^T \mathrm{E}[v_{it} v_{js} (\sum_{p=1}^T a_{r,itp}' z_p^*)(\sum_{p=1}^T a_{\nu,jsp}' z_p^*)] \\ &= \textstyle\sum_{t=1}^T \sum_{s=1}^T \mathrm{E}[v_{it}(\sum_{p=1}^T a_{\nu,jsp}' z_p^*)]\mathrm{E}[v_{js}(\sum_{p=1}^T a_{r,itp}' z_p^*)] \\ &= \textstyle\sum_{t=1}^T \sum_{s=1}^T \mathrm{E}(v_{it} a_{\nu,jst}' z_t^*)\mathrm{E}(v_{js} a_{r,its}' z_s^*) \\ &= \textstyle\sum_{t=1}^T \sum_{s=1}^T \mathrm{E}(a_{\nu,jist} v_{it} z_{it}^*)\mathrm{E}(a_{r,ijts} v_{js} z_{js}^*) \\ &= \sigma_{v0}^4 \textstyle\sum_{t=1}^T \sum_{s=1}^T a_{\nu,jist} a_{r,ijts}. \end{aligned} \tag{3.24}$$

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

$$\begin{aligned} \Upsilon_{k+r,k+\nu}(\boldsymbol{\psi}_0) &= \textstyle\sum_{i=1}^n \sum_{j\neq i}^n \mathrm{E}(g_{\Phi_r,i} g_{\Phi_\nu,j}) \\ &= \textstyle\sum_{i=1}^n \sum_{j\neq i}^n \sigma_{v0}^4 \sum_{t=1}^T \sum_{s=1}^T a_{\nu,jist} a_{r,ijts} \\ &= \sigma_{v0}^4 \mathrm{tr}(\Phi_r \Phi_\nu) - \sigma_{v0}^4 \textstyle\sum_{i=1}^n \sum_{t,s=1}^T \Phi_{\nu ii,st} \Phi_{rii,ts}, \end{aligned} \tag{3.25}$$

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

3.5. Number of factors under fixed T

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 (3.10), we see that, when $r < r_0$, $\text{rank}(M_F(\phi)) < r_0$ 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 $\mathrm{E}[S_{nT}^*(\psi, \phi)] = 0$ holds only at $\psi = \psi_0$.

To see this, write $\mathbf{Z}(\theta) = \mathbf{Z}(\theta_0) - \sum_{p=1}^{k+3} \mathbf{X}_p(\beta_p - \beta_{p0})$, where $\mathbf{X}_p$ is the pth column of $\mathbf{X}, p = 1, \cdots, k$, $\mathbf{X}_{k+1} = \mathbf{Y}_{-1}, \mathbf{X}_{k+2} = \mathbf{W}_1\mathbf{Y}$, and $\mathbf{X}_{k+3} = \mathbf{W}_2\mathbf{Y}_{-1}$, with $\beta_{k+1} = \rho$, $\beta_{k+2} = \lambda_1$, and $\beta_{k+3} = \lambda_2$. Then, for example, the β_1-component of the AQS function can be written as

$$\begin{aligned}\tfrac{1}{\sigma_v^2}\mathbf{X}_1'\mathbf{M}_F(\phi)\mathbf{\Omega}^{-1}(\lambda_3)\mathbf{Z}(\theta) = \tfrac{1}{\sigma_v^2}\mathbf{X}_1'\mathbf{M}_F(\phi)\mathbf{\Omega}^{-1}(\lambda_3)\mathtt{vec}(\Gamma_0 F_0') \\ + \tfrac{1}{\sigma_v^2}\mathbf{X}_1'\mathbf{M}_F(\phi)\mathbf{B}_{30}'(\lambda_3)\mathbf{v} - \tfrac{1}{\sigma_v^2}\textstyle\sum_{p=1}^{k+3}\mathbf{X}_1'\mathbf{M}_F(\phi)\mathbf{\Omega}^{-1}(\lambda_3)\mathbf{X}_k(\beta_p - \beta_{p0}).\end{aligned} \tag{3.26}$$

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 $\mathbf{M}_F(\phi)\mathtt{vec}(\Gamma_0 F_0') = 0$. 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 et al. (2013). This feature is also discussed in Moon and Weidner (2015) for regular panel models, and in Shi and Lee (2017) for DSPD models. Kuersteiner and Prucha (2020), on the other hand, requires r to be correctly specified for their estimator to be consistent. A formal study on this issue is beyond the scope of the paper. We instead provide simulation results for the misspecified case $r > r_0$ in Sec. 5.

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 et al. (2013) propose to estimate r_0 for (non-spatial) short panels with IFE by the following Bayesian information criteria (BIC):

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

where $g(r) = ar$, $f(n) = \frac{\ln n}{n}$, a is an arbitrarily chosen positive number, and $\hat{\sigma}_v^2(r)$ is the estimated error variance based a chosen r. Under BIC, we have $nf(n) \to \infty$, and $f(n) \to 0$ as $n \to \infty$, where the first condition ensures that $\text{plim}_{n\to\infty} \Pr(\widehat{r} > r_0) = 0$, and the second condition is to ensure $\text{plim}_{n\to\infty} \Pr(\hat{r} < r_0) = 0$.

The above BIC may also be used in our case and a similar study would be interesting for our short DSPD-IFE models but is beyond scope of the paper. Finally, it is very interesting to note that our AQS functions may provide a potential framework for the construction of a formal M-test for the identification of a subset of factors that are significant, and thus the identification of the true r_0. This would be an interesting topic for future research.

3.6. M-estimation with relatively small T

As mentioned in the introduction, the asymptotic framework with $T = o(n)$, i.e., T increases with n but at a slower rate, is of great practical interest but has not been formally studied due to technical difficulty. As T increases the dimension of the factor parameters ϕ (and of the corresponding AQS vector) increases. Therefore, it may not be possible to study the asymptotic behavior of $\hat{\boldsymbol{\psi}}_{\mathtt{M}} = (\hat{\psi}_{\mathtt{M}}, \hat{\phi}_{\mathtt{M}})'$ jointly. We focus on the M-estimator $\hat{\psi}_{\mathtt{M}}$ of the common parameters ψ. We show that, under $T = o(n)$, the M-estimation process remains largely unchanged, $\hat{\psi}_{\mathtt{M}}$ is now $\sqrt{nT}$-consistent, and the variance of $\hat{\psi}_{\mathtt{M}}$ can be estimated by the ψ-ψ block of $H_{nT}^{-1}(\hat{\boldsymbol{\psi}}_{\mathtt{M}})\widehat{\Sigma}_{nT}H_{nT}'^{-1}(\hat{\boldsymbol{\psi}}_{\mathtt{M}})$ given in Theorem 3.3. These show that for inference on ψ the finite-T asymptotics still apply when T is large but smaller than n. We further show that the elements of $\hat{\phi}_{\mathtt{M}}$ only have $\sqrt{n}$ consistency rate, and inference for a finite number of linear contrasts of ϕ can be made based on the ϕ-ϕ block of $H_{nT}^{-1}(\hat{\boldsymbol{\psi}}_{\mathtt{M}})\widehat{\Sigma}_{nT}H_{nT}'^{-1}(\hat{\boldsymbol{\psi}}_{\mathtt{M}})$.

Which these results seem to be quite intuitive, the asymptotic arguments leading them are much more complicated than the case of fixed T. Now $\boldsymbol{\psi} = (\psi', \phi')'$ is a high-dimensional parameter, so is the AQS function (3.10) and the asymptotic orders of entries corresponds to ψ and ϕ are distinct. Directly studying the joint AQS function can be troublesome. Instead, we consider the concentrated AQS function with ϕ being 'concentrated' out:

$$\tilde{S}_{nT}^{*}(\psi) = S_{nT,\psi}^{*}\big(\psi, \hat{\phi}(\psi)\big), \tag{3.27}$$

where $\hat{\phi}(\psi) = \mathtt{vec}[\widehat{F}_1(\psi)\widehat{F}_2^{-1}(\psi)]$ and $(\widehat{F}_1'(\psi), \widehat{F}_2'(\psi))' = \widehat{F}(\psi) = \mathtt{eigv}_r\big(\frac{1}{nT}Z'(\theta)\Omega^{-1}(\lambda_3)Z(\theta)\big)$, the eigenvectors corresponding to the r largest eigenvalues. Note that $\hat{\psi}_{\mathtt{M}}$ is the solution to $\tilde{S}_{nT}^{*}(\psi) = 0$. It suffices to derive the asymptotic properties of the concentrated AQS and its derivatives at true parameters. To concentrate out ϕ, one needs an analytical expression of $\widehat{F}(\psi)$ (or $M_{\hat{F}(\psi)}$), which is not possible. To overcome this difficulty, we employ the perturbation theory for linear operators (Kato 2013) to obtain an asymptotic expansion of $M_{\hat{F}(\psi)}$ around ψ_0, so as to give an approximation to $M_{\hat{F}(\psi)}$ using the leading term(s). This method greatly helps in analyzing the asymptotic behavior of the concentrated AQS function, but it still faces a great challenge. If we allow T to go to infinity at a very slow rate compared to n, the leading terms of $M_{\hat{F}(\psi)}$ can be too many, which complicate the analysis greatly. To simplify the analysis, we follow Bai (2003) to restrict $n/T^2 \to 0$ and further assume the following.

To proceed, we impose an assumption on the $n \times T$ matrix of idiosyncratic errors .

Assumption H. *Arrange the idiosyncratic errors v_{it} into an $n \times T$ matrix $\mathbb{V}$. Assume $||\mathbb{V}||_{\mathtt{sp}} = O_p(\sqrt{n+T})$, $||\mathbb{V}\mathbb{V}' - T\sigma_{v0}^2 I_n||_{\mathtt{sp}} = O_p(\frac{n}{T})$, and $||\mathbb{V}'\mathbb{V} - n\sigma_{v0}^2 I_T||_{\mathtt{sp}} = O_p(\frac{T}{n})$.*

Assumption H bounds spectrum norm of matrices related to the idiosyncratic errors. This is a standard assumption in factor analysis literature. Similar assumptions also appear in Moon and Weidner (2015) and Miao et al. (2020). If v_{it}'s are sub-Gaussian, we can prove these properties under Assumption A.[19] The corollary below summarizes the desired results.

Corollary 3.1. *Suppose Assumptions A-H hold, $(n,T)\to\infty$ and $\frac{T}{n}+\frac{\sqrt{n}}{T}\to 0$. Then,*

$$(i)\ \hat{\psi}_{\mathtt{M}}-\psi_0=O_p\big(\tfrac{1}{\sqrt{nT}}\big)\ \text{ and }\ \hat{\phi}_{s,\mathtt{M}}-\phi_{s0}=O_p(\tfrac{1}{\sqrt{n}}),\ \text{for each } s=1,\ldots,k_\phi;$$
$$(ii)\ \sqrt{nT}\big(\hat{\psi}_{\mathtt{M}}-\psi_0\big)\xrightarrow{D}N\Big(0,\ \lim_{(n,T)\to\infty}\tilde{H}_{nT}^{-1}(\psi_0)\tilde{\Sigma}_{nT}(\psi_0)\tilde{H}_{nT}^{\prime-1}(\psi_0)\Big),$$

where $\tilde{H}_{nT}(\psi_0)=-\frac{1}{nT}\mathrm{E}\big[\frac{\partial}{\partial\psi'}\tilde{S}_{nT}^*(\psi_0)\big]$ and $\tilde{\Sigma}_{nT}(\psi_0)=\frac{1}{nT}\mathrm{Var}\big[\tilde{S}_{nT}^*(\psi_0)\big]$, both assumed to exist and $\tilde{H}_{nT}(\psi_0)$ to be positive definite for large enough (n,T).

Corollary 3.1 shows that $\hat{\psi}_{\mathtt{M}}$ is consistent with $\sqrt{nT}$ rate. On top of that, there are no asymptotic bias terms that may affect the inference. In contrast, the QML estimator of Bai and Li (2021) has three asymptotic bias terms and bias correction has to be made after estimation for valid inference. For inference, Corollary 3.1 indicates that we can further find the limit of the covariance matrix and proposes a new consistent estimator. However, this will lead to a different inference procedure compared to the fixed T framework. In practice, one has to determine which inference procedure to use which will largely complicate the inference procedure. To address this issue, we have carefully studied the asymptotic property of $\hat{\Sigma}_{nT}$ as in (3.16) and found a delightful fact: the covariance estimator proposed in Sec. 3.4 continues to valid. More specifically, the ψ-ψ block of the covariance matrix given in Theorem 3.3 still consistently estimates the covariance matrix of $\hat{\psi}_{\mathtt{M}}$ under the large T framework.

Corollary 3.2. *Suppose Assumptions A-H hold. We have as $(n,T)\to\infty$ and $\frac{T}{n}+\frac{\sqrt{n}}{T}\to 0$,*

$$\big[H_{nT}^{-1}(\hat{\boldsymbol{\psi}}_{\mathtt{M}})\widehat{\Sigma}_{nT}H_{nT}^{\prime-1}(\hat{\boldsymbol{\psi}}_{\mathtt{M}})\big]_{\psi\psi}-\tilde{H}_{nT}^{-1}(\psi_0)\tilde{\Sigma}_{nT}(\psi_0)\tilde{H}_{nT}^{\prime-1}(\psi_0)\longrightarrow 0,$$

where $[\,\cdot\,]_{\psi\psi}$ takes the ψ-ψ block of the given matrix.

Corollary 3.2 confirms the inference procedure on the finite-dimensional parameter vector ψ is valid for both the case of fixed T and the case of $T=o(n)$.[20] Establishing the Corollary is challenging as we have to deal with high-dimensional matrices. While the target matrix is of fixed dimension, two inverses of high-dimensional matrices are involved in the analysis. In addition, the entries corresponding to ψ and ϕ are of different asymptotic orders. We need to handle these two issues carefully in the proofs of the results.

Statistical inference for ϕ_s or a finite number of linear contrasts $C\phi$ of ϕ might be of interest. With $\sqrt{nT}$ consistent rate of $\hat{\psi}_{\mathtt{M}}$, the estimate of factor is asymptotically equivalent to that obtained from a pure factor model (Bai 2003). We can also show that, inference on $C\phi_0$ can be carried out based on the result: $\sqrt{n}C(\hat{\phi}_{s,\mathtt{M}}-\phi_{s0})\xrightarrow{D}N(0,\ \Omega)$, where Ω can be consistently estimated by $C\big[H_{nT}^{-1}(\hat{\boldsymbol{\psi}}_{\mathtt{M}})\hat{\Sigma}_{nT}H_{nT}^{\prime-1}(\hat{\boldsymbol{\psi}}_{\mathtt{M}})\big]_{\phi\phi}C'$.

[19]For a detailed discussion, one can look up the book Vershynin (2018).

[20]The restrictive assumption $\frac{\sqrt{n}}{T}\to 0$ is only of technical reasons and the cost of removing it may be just many extra lines in the proofs. The Monte Carlo simulation results show that this assumption is not necessary.

4. M-Estimation of Extended DSPD-IFE Models

In this section, we present some critical details on the following extensions: (*i*) DSPD-IFE model with time-varying spatial weight matrices, (*ii*) DSPD-IFE model with unknown cross-sectional heteroskedasticity, and (*iii*) High-order DSPD-IFE models. We also give some discussions on the potential applications of our methods to estimate DSPD-IFE models with endogenous spatial weights and additional endogenous regressors.

(i) Time-varying spatial weight matrices. First, consider Model (2.4) but with $W_{3t} = W_3$. The model has the reduced form: $y_t = \mathcal{B}_t y_{t-1} + B_{1t0}^{-1} x_t \beta_0 + B_{1t0}^{-1} z_t$, $t = 1, \ldots, T$, where $\mathcal{B}_t = B_{1t0}^{-1} B_{2t0}$, $B_{1t0} = I_n - \lambda_{10} W_{1t}$ and $B_{2t0} = \rho_0 I_n + \lambda_{20} W_{2t}$. Define $\mathbf{W}_\nu = \mathtt{bdiag}(W_{\nu t}, t = 1, \ldots, T)$, $\nu = 1, 2$, $\mathbf{B}_1(\lambda_1) = I_{nT} - \lambda_1 \mathbf{W}_1$, and $\mathbf{B}_2(\rho, \lambda_2) = \rho I_{nT} + \lambda_2 \mathbf{W}_2$. The representations for $\mathbf{Y}$ and $\mathbf{Y}_{-1}$ given in (3.8) still hold with redefined $\mathbf{Q}$, $\mathbf{Q}_{-1}$, $\mathbf{D}$, and $\mathbf{D}_{-1}$:

$\mathbf{Q} = \mathtt{bdiag}(\mathcal{B}_1,\ \mathcal{B}_1\mathcal{B}_2,\ \ldots,\ \mathcal{B}_1 \cdots \mathcal{B}_T)$, $\mathbf{Q}_{-1} = \mathtt{bdiag}(I_n,\ \mathcal{B}_1,\ \ldots,\ \mathcal{B}_1 \cdots \mathcal{B}_{T-1})$,

$$\mathbf{D} = \begin{pmatrix} I_n & 0 & \ldots & 0 & 0 \\ \mathcal{B}_2 & I_n & \ldots & 0 & 0 \\ \mathcal{B}_2\mathcal{B}_3 & \mathcal{B}_2 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathcal{B}_2 \cdots \mathcal{B}_T & \mathcal{B}_2 \cdots \mathcal{B}_{(T-1)} & \ldots & \mathcal{B}_2 & I_n \end{pmatrix} \mathbf{B}_{10}^{-1}, \quad \text{and} \tag{4.1}$$

$$\mathbf{D}_{-1} = \begin{pmatrix} 0 & 0 & \ldots & 0 & 0 \\ I_n & 0 & \ldots & 0 & 0 \\ \mathcal{B}_2 & I_n & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathcal{B}_2 \cdots \mathcal{B}_{(T-1)} & \mathcal{B}_2 \cdots \mathcal{B}_{(T-2)} & \ldots & I_n & 0 \end{pmatrix} \mathbf{B}_{10}^{-1}. \tag{4.2}$$

Then, with $\mathbf{Z}(\theta) = \mathbf{B}_1(\lambda_1)\mathbf{Y} - \mathbf{B}_2(\rho, \lambda_2)\mathbf{Y}_{-1} - \mathbf{X}\beta$, we see that the AQS function takes the form identical to (3.10), and M-estimation proceeds.

Next, we further allow W_3 to be time varying and define $B_{3t0} = I_n - \lambda_{30} W_{3t}$, $\mathbf{W}_3 = \mathtt{bdiag}(W_{3t}, t = 1, \ldots, T)$ and $\mathbf{B}_3(\lambda_3) = I_{nT} - \lambda_3 \mathbf{W}_3$. With time varying W_{3t}, concentrating out the factor loadings Γ from the CQL function in (3.1) is no longer straightforward. Let $\Gamma_v = \mathtt{vec}(\Gamma)$, we can rewrite the CQL function as

$$\begin{aligned} \ell_{nT}(\psi, \Gamma_v, F) &= -\tfrac{nT}{2} \log(2\pi\sigma_v^2) - \log|\mathbf{B}_3(\lambda_3)| + \log|\mathbf{B}_1(\lambda_1)| \\ &- \tfrac{1}{2\sigma_v^2} \textstyle\sum_{t=1}^T [z_t'(\theta)\Omega_t^{-1}(\lambda_3) z_t(\theta) - 2\Gamma_v'(f_t \otimes \Omega_t^{-1}(\lambda_3)) z_t(\theta) + \Gamma_v'(f_t \otimes \Omega_t^{-1}(\lambda_3))\Gamma_v], \end{aligned} \tag{4.3}$$

where $\Omega_t(\lambda_3) = (B_{3t}'(\lambda_3) B_{3t}(\lambda_3))^{-1}$. Solving the first order condition, $\frac{\partial}{\partial \Gamma_v} \ell_{nT}(\psi, \Gamma_v, F) = 0$ gives the constrained CQML estimator of Γ_v

$$\tilde{\Gamma}_v(\theta, \lambda_3, F) = [\textstyle\sum_{t=1}^T (f_t f_t' \otimes \Omega_t^{-1})]^{-1} [\sum_{t=1}^T (f_t \otimes \Omega_t^{-1}) z_t(\theta)]. \tag{4.4}$$

Then we obtain the CCQL function by plugging $\tilde{\Gamma}_v(\theta, \lambda_3, F)$ into $\ell_{nT}(\psi, \Gamma_v, F)$ as

$$\begin{aligned}\ell_{nT}^{c}(\psi,F) = &-\tfrac{nT}{2}\log(2\pi\sigma_v^2) - \log|\mathbf{B}_3(\lambda_3)| + \log|\mathbf{B_1}(\boldsymbol{\lambda_1})| \\ &-\tfrac{1}{2\sigma_v^2}\mathbf{Z}'(\theta)\mathbf{B}_3'(\lambda_3)\mathbf{M}_{F^\dagger}(F,\lambda_3)\mathbf{B}_3(\lambda_3)\mathbf{Z}(\theta),\end{aligned} \tag{4.5}$$

where $\mathbf{M}_{F^\dagger}(F,\lambda_3) = I_{nT} - \mathbf{F}^\dagger(\mathbf{F}^{\dagger\prime}\mathbf{F}^\dagger)^{-1}\mathbf{F}^{\dagger\prime}$ and $\mathbf{F}^\dagger = \mathbf{B}_3(F \otimes I_n)$. It can been verified easily that (4.5) reduces to (3.5) when W_3 is time-invariant. The CCQS functions of ψ and ϕ defined in (3.6) now becomes

$$S_{nT}^{c}(\psi,\phi) = \begin{cases} \frac{1}{\sigma_v^2}\mathbf{X}'\mathbf{B}_3'(\lambda_3)\mathbf{M}_{F^\dagger}\mathbf{B}_3(\lambda_3)\mathbf{Z}(\theta), \\ \frac{1}{\sigma_v^2}\mathbf{Z}'(\theta)\mathbf{B}_3'(\lambda_3)\mathbf{M}_{F^\dagger}\mathbf{B}_3(\lambda_3)\mathbf{Y}_{-1}, \\ \frac{1}{\sigma_v^2}\mathbf{Z}'(\theta)\mathbf{B}_3'(\lambda_3)\mathbf{M}_{F^\dagger}\mathbf{B}_3(\lambda_3)\mathbf{W}_1\mathbf{Y} - \mathrm{tr}[\mathbf{W}_1\mathbf{B}_1^{-1}(\lambda_1)], \\ \frac{1}{\sigma_v^2}\mathbf{Z}'(\theta)\mathbf{B}_3'(\lambda_3)\mathbf{M}_{F^\dagger}\mathbf{B}_3(\lambda_3)\mathbf{W}_2\mathbf{Y}_{-1}, \\ \frac{1}{\sigma_v^2}\mathbf{Z}'(\theta)\mathbf{B}_3'(\lambda_3)\mathbf{M}_{F^\dagger}\mathbf{W}_3\mathbf{Z}(\theta) - \mathrm{tr}[\mathbf{W}_3\mathbf{B}_3^{-1}(\lambda_3)], \\ \frac{1}{2\sigma_v^4}\mathbf{Z}'(\theta)\mathbf{B}_3'(\lambda_3)\mathbf{M}_{F^\dagger}\mathbf{B}_3(\lambda_3)\mathbf{Z}(\theta) - \frac{nT}{2\sigma_v^2}, \\ \frac{1}{\sigma_v^2}\mathbf{Z}'(\theta)\mathbf{B}_3'(\lambda_3)\mathbf{A}_s(\phi,\lambda_3)\mathbf{B}_3(\lambda_3)\mathbf{Z}(\theta), \quad s = 1,\ldots,k_\phi, \end{cases} \tag{4.6}$$

where $\mathbf{A}_s = \mathbf{M}_{F^\dagger}[\dot{\mathbf{F}}_s^\dagger(\mathbf{F}^{\dagger\prime}\mathbf{F}^\dagger)^{-1}\mathbf{F}^{\dagger\prime}]$, and $\dot{\mathbf{F}}_s^\dagger = \frac{\partial}{\partial\phi_s}\mathbf{F}^\dagger = \mathbf{B}_3(\dot{F}_s \otimes I_n)$.

With $\mathbf{D}$ and $\mathbf{D}_{-1}$ defined in (4.1) and (4.2), defining $\mathbf{M}_{F^\dagger}^{*} = \mathbf{B}_3\mathbf{M}_{F^\dagger}\mathbf{B}_3^{-1}$, the AQS function can be written as

$$S_{nT}^{*}(\psi,\phi) = \begin{cases} \frac{1}{\sigma_v^2}\mathbf{X}'\mathbf{B}_3'(\lambda_3)\mathbf{M}_{F^\dagger}\mathbf{B}_3(\lambda_3)\mathbf{Z}(\theta), \\ \frac{1}{\sigma_v^2}\mathbf{Z}'(\theta)\mathbf{B}_3'(\lambda_3)\mathbf{M}_{F^\dagger}\mathbf{B}_3(\lambda_3)\mathbf{Y}_{-1} - \mathrm{tr}[\mathbf{M}_{F^\dagger}^{*}\mathbf{D}_{-1}(\rho,\lambda_1,\lambda_2)], \\ \frac{1}{\sigma_v^2}\mathbf{Z}'(\theta)\mathbf{B}_3'(\lambda_3)\mathbf{M}_{F^\dagger}\mathbf{B}_3(\lambda_3)\mathbf{W}_1\mathbf{Y} - \mathrm{tr}[\mathbf{M}_{F^\dagger}^{*}\mathbf{W}_1\mathbf{D}(\rho,\lambda_1,\lambda_2)], \\ \frac{1}{\sigma_v^2}\mathbf{Z}'(\theta)\mathbf{B}_3'(\lambda_3)\mathbf{M}_{F^\dagger}\mathbf{B}_3(\lambda_3)\mathbf{W}_2\mathbf{Y}_{-1} - \mathrm{tr}[\mathbf{M}_{F^\dagger}^{*}\mathbf{W}_2\mathbf{D}_{-1}(\rho,\lambda_1,\lambda_2)], \\ \frac{1}{\sigma_v^2}\mathbf{Z}'(\theta)\mathbf{B}_3'(\lambda_3)\mathbf{M}_{F^\dagger}\mathbf{W}_3\mathbf{Z}(\theta) - \mathrm{tr}[\mathbf{M}_{F^\dagger}\mathbf{W}_3\mathbf{B}_3^{-1}(\lambda_3)], \\ \frac{1}{2\sigma_v^4}\mathbf{Z}'(\theta)\mathbf{B}_3'(\lambda_3)\mathbf{M}_{F^\dagger}\mathbf{B}_3(\lambda_3)\mathbf{Z}(\theta) - \frac{n(T-r)}{2\sigma_v^2}, \\ \frac{1}{\sigma_v^2}\mathbf{Z}'(\theta)\mathbf{B}_3'(\lambda_3)\mathbf{A}_s(\phi,\lambda_3)\mathbf{B}_3(\lambda_3)\mathbf{Z}(\theta), \quad s = 1,\ldots,k_\phi. \end{cases} \tag{4.7}$$

The above AQS functions, allowing all three weight matrices being time varying, take similar form as these in (3.10). Our proposed M-estimation and inference methods proceed as before.

(ii) Cross-sectional heteroskedasticity. An interesting extension to consider is to allow for cross-sectional heteroskedasticity in the error vector $\mathbf{v}$. For ease of exposition, we extend the model considered in Sec. 3 by allowing $\mathbf{v} \sim (0, \sigma_{v0}^2\mathbf{H})$ where $\mathbf{H} = (I_T \otimes \mathcal{H})$ (see Remark 3.3 and Footnote 13). It is easy to verify the following results:

$$\sigma_{v0}^{-2}\mathrm{E}(\mathbf{Z}'\mathbf{M}_{F_0}\boldsymbol{\Omega}_0^{-1}\mathbf{Y}_{-1}) = \mathrm{tr}(\mathbf{D}_{-1}\mathbf{M}_{F_0}\mathbf{B}_{30}^{-1}\mathbf{H}\mathbf{B}_{30}), \tag{4.8}$$

$$\sigma_{v0}^{-2}\mathrm{E}(\mathbf{Z}'\mathbf{M}_{F_0}\boldsymbol{\Omega}_0^{-1}\mathbf{W}_1\mathbf{Y}) = \mathrm{tr}(\mathbf{D}\mathbf{M}_{F_0}\mathbf{B}_{30}^{-1}\mathbf{H}\mathbf{B}_{30}\mathbf{W}_1), \tag{4.9}$$

$$\sigma_{v0}^{-2}\mathrm{E}(\mathbf{Z}'\mathbf{M}_{F_0}\boldsymbol{\Omega}_0^{-1}\mathbf{W}_2\mathbf{Y}_{-1}) = \mathrm{tr}(\mathbf{D}_{-1}\mathbf{M}_{F_0}\mathbf{B}_{30}^{-1}\mathbf{H}\mathbf{B}_{30}\mathbf{W}_2), \tag{4.10}$$

$$\sigma_{v0}^{-2}\mathrm{E}(\mathbf{Z}'\mathbf{M}_{F_0}\mathbf{B}_{30}'\mathbf{W}_3\mathbf{Z}) = \mathrm{tr}(\mathbf{M}_{F_0}\mathbf{B}_{30}^{-1}\mathbf{H}\mathbf{W}_3). \tag{4.11}$$

Therefore, the ρ and λ components $\mathrm{E}[\frac{\partial}{\partial\psi}\ell^c_{nT}(\psi_0,\phi_0)]$ 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 (2020) for an DSPD model with additive FE, instead of directly subtracting the expectation, we can find a set of quadratic terms in $\mathbf{Z}$ with expectations being identical to (4.8)-(4.11):

$$\sigma_{v0}^{-2}\mathrm{E}(\mathbf{Z}'\mathbf{\Omega}_0^{-1}\mathbf{D}_{-1}\mathbf{M}_{F_0}\mathbf{Z}) = \mathrm{tr}(\mathbf{D}_{-1}\mathbf{M}_{F_0}\mathbf{B}_{30}^{-1}\mathbf{H}\mathbf{B}_{30}), \tag{4.12}$$

$$\sigma_{v0}^{-2}\mathrm{E}(\mathbf{Z}'\mathbf{\Omega}_0^{-1}\mathbf{W}_1\mathbf{D}\mathbf{M}_{F_0}\mathbf{Z}) = \mathrm{tr}(\mathbf{D}\mathbf{M}_{F_0}\mathbf{B}_{30}^{-1}\mathbf{H}\mathbf{B}_{30}\mathbf{W}_1), \tag{4.13}$$

$$\sigma_{v0}^{-2}\mathrm{E}(\mathbf{Z}'\mathbf{\Omega}_0^{-1}\mathbf{W}_2\mathbf{D}_{-1}\mathbf{M}_{F_0}\mathbf{Z}) = \mathrm{tr}(\mathbf{D}_{-1}\mathbf{M}_{F_0}\mathbf{B}_{30}^{-1}\mathbf{H}\mathbf{B}_{30}\mathbf{W}_2), \tag{4.14}$$

$$\sigma_{v0}^{-2}\mathrm{E}\{\mathbf{Z}'\mathbf{B}_{30}'[I_T\otimes \mathtt{diag}(W_3B_{30}^{-1})]\mathbf{B}_{30}\mathbf{M}_{F_0}\mathbf{Z}\} = \mathrm{tr}(\mathbf{M}_{F_0}\mathbf{B}_{30}^{-1}\mathbf{H}\mathbf{W}_3). \tag{4.15}$$

Taking the differences between these two sets and drop the expectations lead to a set of unbiased estimating functions for ρ and λ, robust against unknown $\mathcal{H}$. The ϕ-component of the EF vector given in (3.10) is naturally robust against unknown $\mathcal{H}$ as shown in Footnote 13. Moreover the β' and σ_v^2 components also do not need further adjustment under heteroskedasticity. Therefore, a full set of EFs robust against unknown $\mathcal{H}$ is given as follows.

$$S^r_{nT}(\psi,\phi) = \begin{cases} \mathbf{X}'\mathbf{M}_F\mathbf{\Omega}^{-1}(\lambda_3)\mathbf{Z}(\theta), \\ \mathbf{Z}'(\theta)\mathbf{M}_F\mathbf{\Omega}^{-1}(\lambda_3)\mathbf{Y}_{-1} - \mathbf{Z}'(\theta)\mathbf{\Omega}^{-1}(\lambda_3)\mathbf{D}_{-1}(\rho,\lambda_1,\lambda_2)\mathbf{M}_F\mathbf{Z}(\theta), \\ \mathbf{Z}'(\theta)\mathbf{M}_F\mathbf{\Omega}^{-1}(\lambda_3)\mathbf{W}_1\mathbf{Y} - \mathbf{Z}'(\theta)\mathbf{\Omega}^{-1}(\lambda_3)\mathbf{W}_1\mathbf{D}(\rho,\lambda_1,\lambda_2)\mathbf{M}_F\mathbf{Z}(\theta), \\ \mathbf{Z}'(\theta)\mathbf{M}_F\mathbf{\Omega}^{-1}(\lambda_3)\mathbf{W}_2\mathbf{Y}_{-1} - \mathbf{Z}'(\theta)\mathbf{\Omega}^{-1}(\lambda_3)\mathbf{W}_2\mathbf{D}_{-1}(\rho,\lambda_1,\lambda_2)\mathbf{M}_F\mathbf{Z}(\theta), \\ \mathbf{Z}'(\theta)\mathbf{M}_F\mathbf{B}_3'(\lambda_3)\big\{\mathbf{W}_3 - [I_T\otimes \mathtt{diag}(W_3B_3^{-1}(\lambda_3))]\mathbf{B}_3(\lambda_3)\big\}\mathbf{Z}(\theta), \\ \frac{1}{2\sigma_v^2}\mathbf{Z}'(\theta)\mathbf{M}_F\mathbf{\Omega}^{-1}(\lambda_3)\mathbf{Z}(\theta) - \frac{n(T-r)}{2}, \\ \mathbf{Z}'(\theta)[M_F\dot{F}_s(F'F)^{-1}F'\otimes\Omega^{-1}(\lambda_3)]\mathbf{Z}(\theta),\ \ s=1,\ldots,k_\phi. \end{cases} \tag{4.16}$$

We have $\mathrm{E}[S^r_{nT}(\psi_0,\phi_0)] = 0$. We further show that $\mathrm{plim}_{n\to\infty}\frac{1}{nT}S^r_{nT}(\psi_0,\phi_0) = 0$. Therefore, solving $S^r_{nT}(\psi,\phi) = 0$ would give consistent M-estimators of ψ and ϕ robust against unknown $\mathcal{H}$. The two-step computation approach still works under heteroskedasticity (see footnote 11 for details). With the EF vector (4.16), our M-estimation method will go through as before and remain valid. Our inference method will go through as well **provided** that either the Γ term or the λ_3 term 'exists'. When both terms are absent, the $\mathcal{H}$-robust inference for σ_0^2 faces difficulty. This suggests one should work with $S^r_{nT}(\psi,\phi)$ without the σ^2-component for $\mathcal{H}$-robust inference. This is particularly meaningful as the subvector is free from σ^2. While the fundamental ideas are clear, these extensions require additional complicated algebra and rigorous proofs, and can only be handled by a separate research.

The same set of results can also be worked out for KP's type of model given in (2.1) with exogenous spatial weights and regressors, where all spatial weight matrices are allowed to change with time in light of the remarks given at the end of **(i)**. Extensions to high-order DSPD-IFE models are possible. See the discussions below.

(iii) High-order DSPD-IFE models. Our methods can also be extended to allow for multiple space and time lags as in Models (2.1), (2.2), and (2.3). First, for Model (2.1) with $p = 1$ and spatial weights and regressors being exogenous. Let $\lambda_\nu = (\lambda_{\nu 1}, \ldots, \lambda_{\nu, q_\nu})', \nu = 1, 2, 3$. Define $B_{\nu t}(\lambda_\nu) = I_n - \sum_{\ell=1}^{q_\nu} \lambda_{\nu\ell} W_{\nu\ell t}$, $\nu = 1, 3$, and $B_{2t}(\rho, \lambda_2) = \rho I_n + \sum_{\ell=1}^{q_2} \lambda_{2\ell} W_{2\ell, t-1}$. Then, Model (2.1) can be written in the following compact form:

$$B_{1t}(\lambda_1) y_t = B_{2t}(\rho, \lambda_2) y_{t-1} + x_t \beta + B_{3t}^{-1}(\lambda_3)(\Gamma f_t + v_t), \quad t = 1, \ldots, T.$$

Redefine $z_t(\theta) = B_{3t}(\lambda_1)[B_{1t}(\lambda_1) y_t - B_{2t}(\rho, \lambda_2) y_{t-1} - x_t \beta]$, and let $\mathbb{Z}(\theta) = [z_1(\theta), \ldots, z_T(\theta)]$. Referring to (3.1) and (3.2), the only component in the quasi Gaussian loglikelihood that involves $\Gamma F'$ has the form: $-\frac{1}{2\sigma^2} \text{tr}[(\mathbb{Z}(\theta) - \Gamma F')'(\mathbb{Z}(\theta) - \Gamma F')]$. With this new $\mathbb{Z}(\theta)$, the CQML estimate of Γ, $\tilde{\Gamma}(\theta, F)$, has an identical form as (3.3). The rest of the derivations for the M-estimation can be done in a manner similar to Sec. 3. For Model (2.2), if further $B_{3t}(\lambda_3) = B_3(\lambda_3)$, then, the quasi Gaussian loglikelihood remain the same as (3.1) and (3.2). The rest of derivations is similar to those in Sec. 3, although much more tedious due to the existence of multiple spatial lag effects of three different forms. For Model (2.3), combining the above ideas, if both $B_{3t}(\lambda_3)$ and $B_{4t}(\lambda_4)$ are time-invariant, and the spatial weight matrices and regressors are exogenous, our M-estimation method will go through.

We end this section by offering some comments on the DSPD-IFE models with endogenous spatial weights and regressors. Our methods have potential to be extended to cover the cases where the spatial weights and some regressors are generated by some endogenous economic variables through some functional relationship as in Qu et al. (2017). In this case, we are able to derive the CQL function, and thereby the adjustments, and so on.

5. Monte Carlo Study

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

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

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, and (ii) the GMM estimator in Kuersteiner and Prucha (2020) using DGP2. 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.[21]

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 $H_{nT}^{-1}(\hat{\boldsymbol{\psi}}_{\mathtt{M}})\hat{\Sigma}_{nT} H_{nT}^{-1}(\hat{\boldsymbol{\psi}}_{\mathtt{M}})$, which should be compared with the corresponding empirical sd. Similar types of Monte Carlo results are also reported for BC-QMLE for direct comparisons. Due to the issues of numerical stability and code availability of the GMM estimator, a smaller scale of comparison is made.

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 dominate, in terms of bias and efficiency, the bias-corrected CQML method of Shi and Lee (2017),[22] and the GMM method of Kuersteiner and Prucha (2020).[23]

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-CQMLE 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-CQMLE of λ_1 performs quite well, but these of λ_2 and λ_3 are slightly biased. While the biases of the BC-CQMLEs of λ_2 and λ_3 are not severe, the

[21] 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.

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

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

standard error estimate 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, leading to reliable inferences.

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-CQMLEs 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 perform quite well and are generally very close to the corresponding Monte Carlo sds, whereas the ses of BC-CQMLE 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-CQMLEs, on the other hand, are severely biased under this setting, especially for ρ and λ_1. The associated standard error estimates of the BC-CQMLEs perform even worse.

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-CQMLE 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) when spatial weights and covariates are strictly exogenous. 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 proposed M-estimator is 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 suggest that when extra conditions (strict exogeneity) are met, the proposed likelihood-type estimator can be more efficient than the more general GMM estimator of Kuersteiner and Prucha (2020) which are valid when spatial weights and (some) regressors are sequentially exogenous. The Monte Carlo results for GMM estimator do not include $\widehat{\text{rse}}$, based on the code we received from the authors. It would be interesting to study the efficiency of the proposed estimator from theoretical perspective, but it is clearly beyond the scope of current paper. We plan to carry out such a study in a future research.

6. Conclusion

This paper proposes a set of new estimation and inference methods for spatial dynamic panel data models with interactive fixed effect based on fixed and relatively large T set up, 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 before the start of estimation process, rendering the estimators possessing the usual asymptotic properties, i.e., consistency and asymptotic normality with zero mean. 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.

The proposed set of estimation and inferences methods constitute an important set of econometric tools relevant to applied researchers dealing with a broad class of pertinent issues, such as spatial spillovers, endogenous social effects, social interactions, network effects, time persistence, spatial diffusion, common factors, cross-sectional dependence, cross-sectional heteroskedasticity, etc. In addition, the nature of the proposed estimation and inference methods suggests that there is a great potential for further extensions to allow for even more features in the model. A rigorous comparison of various estimators for the DSPD models with IFE, in particular between M-estimator and GMM estimator, would be an interesting topic of future research. Specification tests for identifying the number of factors based on our AQS functions would be another interesting topic of future research. These (proposed methods and planned research) are of a great relevance in the era of machine learning with big data,

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 (2.4), 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 neighborhood 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 of Theorems

To simplify notation, a parametric quantity (scalar, vector or matrix) evaluated at parameters' general values 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 $\Omega(\lambda_3) \equiv \Omega$. The following matrix results are repeatedly 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, $\gamma_{\min}(A)\text{tr}(B) \leq \text{tr}(AB) \leq \gamma_{\max}(A)\text{tr}(B)$; (iv) for symmetric matrices A and B, $\gamma_{\max}(A+B) \leq \gamma_{\max}(A) + \gamma_{\max}(B)$; and (v) for p.s.d. matrices A and B, $\gamma_{\max}(AB) \leq \gamma_{\max}(A)\gamma_{\max}(B)$. 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 $\sup_{\delta\in\mathbf{\Delta}} \frac{1}{nT}\|S^{*c}_{nT}(\delta) - \bar{S}^{*c}_{nT}(\delta)\| \xrightarrow{p} 0$ as $n \to \infty$, where $S^{*c}_{nT}(\delta)$ is the concentrated AQS function for δ and $\bar{S}^{*c}_{nT}(\delta)$ is its population counterpart. Both quantities are defined above Theorem 3.1 and their exact expressions are given below:

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

where recall $\hat{\mathbf{Z}}(\delta)(=\mathbf{B}_1\mathbf{Y} - \mathbf{B}_2\mathbf{Y}_{-1} - \mathbf{X}\hat{\beta}(\delta))$, $\hat{\sigma}^2_v(\delta)$ and $\hat{\beta}(\delta)$ from (3.11) and (3.12);

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

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

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

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

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

(d) $\sup_{\delta\in\mathbf{\Delta}} \frac{1}{nT}\left|\hat{\mathbf{Z}}'(\delta)\mathbf{M}_F\mathbf{\Omega}^{-1}\mathbf{W}_1\mathbf{Y} - \text{E}[\bar{\mathbf{Z}}'(\delta)\mathbf{M}_F\mathbf{\Omega}^{-1}\mathbf{W}_1\mathbf{Y}]\right| = o_p(1)$,

(e) $\sup_{\delta\in\mathbf{\Delta}} \frac{1}{nT}\left|\hat{\mathbf{Z}}'(\delta)\mathbf{M}_F\mathbf{\Omega}^{-1}\mathbf{W}_2\mathbf{Y}_{-1} - \text{E}[\bar{\mathbf{Z}}'(\delta)\mathbf{M}_F\mathbf{\Omega}^{-1}\mathbf{W}_2\mathbf{Y}_{-1}]\right| = o_p(1)$,

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

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

Denote $\mathbf{A} = \mathbf{M}_F\boldsymbol{\Omega}^{-1} = M_F\otimes(B_3'B_3)$, and let $\mathbf{A}^{\frac{1}{2}}$ be a square-root matrix of $\mathbf{A}$. Define $\bar{\mathbf{Z}}^\dagger(\delta) = \mathbf{A}^{\frac{1}{2}}\bar{\mathbf{Z}}(\delta)$, $\hat{\mathbf{Z}}^\dagger(\delta) = \mathbf{A}^{\frac{1}{2}}\hat{\mathbf{Z}}(\delta)$, and $\mathbf{B}_\nu^\dagger = \mathbf{A}^{\frac{1}{2}}\mathbf{B}_\nu, r = 1,2$. Let $\mathbf{Y}^\circ = \mathbf{Y} - \mathrm{E}(\mathbf{Y})$ and $\mathbf{Y}_{-1}^\circ = \mathbf{Y}_{-1} - \mathrm{E}(\mathbf{Y}_{-1})$. 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:

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

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

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

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

By Assumption E(iv) and the assumptions given in the theorem, we have for the first term, $\inf_{\delta\in\boldsymbol{\Delta}} \frac{1}{n(T-r)}\mathrm{tr}[\mathbf{A}\mathrm{Var}(\mathbf{B}_1\mathbf{Y} - \mathbf{B}_2\mathbf{Y}_{-1})] \geq \frac{1}{n(T-r)}\inf_{\delta\in\boldsymbol{\Delta}}\gamma_{\min}[\mathrm{Var}(\mathbf{B}_1\mathbf{Y} - \mathbf{B}_2\mathbf{Y}_{-1})]\mathrm{tr}(M_F\otimes B_3'B_3) \geq \frac{1}{n}\underline{c}_y\inf_{\lambda_3\in\Lambda_3}\mathrm{tr}(B_3'B_3) \geq \frac{1}{n}\underline{c}_y n[\inf_{\lambda_3\in\Lambda_3}\gamma_{\min}(B_3'B_3)] \geq \underline{c}_y\underline{c}_3 > 0$. The second term is non-negative uniformly in $\delta\in\boldsymbol{\Delta}$ as $\mathbf{M}$ is positive semi-definite (p.s.d). It follows that $\inf_{\delta\in\boldsymbol{\Delta}}\bar{\sigma}_v^2(\delta) > c > 0$, 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 $Q_1 = \frac{1}{n(T-r)}\mathbf{Y}'\mathbf{B}_1^{\dagger\prime}\mathbf{M}\mathbf{B}_1^\dagger\mathbf{Y}$, $Q_2 = \frac{1}{n(T-r)}\mathbf{Y}_{-1}'\mathbf{B}_2^{\dagger\prime}\mathbf{M}\mathbf{B}_2^\dagger\mathbf{Y}_{-1}$, $Q_3 = \frac{2}{n(T-r)}\mathbf{Y}'\mathbf{B}_1^{\dagger\prime}\mathbf{M}\mathbf{B}_2^\dagger\mathbf{Y}_{-1}$ and $Q_4 = \frac{1}{n(T-r)}(\mathbf{B}_1^\dagger\mathbf{Y}^\circ - \mathbf{B}_2^\dagger\mathbf{Y}_{-1}^\circ)'\mathbf{P}(\mathbf{B}_1^\dagger Y^\circ - \mathbf{B}_2^\dagger\mathbf{Y}_{-1}^\circ)$. 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 $\delta\in\boldsymbol{\Delta}$.

Recall from (3.8): $\mathbf{Y} = \mathbf{Q}\mathbf{y}_0 + \boldsymbol{\eta} + \mathbf{D}\mathbf{Z}$ and $\mathbf{Y}_{-1} = \mathbf{Q}_{-1}\mathbf{y}_0 + \boldsymbol{\eta}_{-1} + \mathbf{D}_{-1}\mathbf{Z}$. By $\mathbf{B}_{30}\mathbf{Z} = \mathbf{v} + \mathtt{vec}(B_{30}\Gamma_0F_0')$, we can further write $\mathbf{Y} = \mathbf{Q}\mathbf{y}_0 + \boldsymbol{\eta}^* + \mathbf{D}\mathbf{B}_{30}^{-1}\mathbf{v}$, and $\mathbf{Y}_{-1} = \mathbf{Q}_{-1}\mathbf{y}_0 + \boldsymbol{\eta}_{-1}^* + \mathbf{D}_{-1}\mathbf{B}_{30}^{-1}\mathbf{v}$, where $\boldsymbol{\eta}^* = \boldsymbol{\eta} + \mathbf{D}\mathtt{vec}(\Gamma_0F_0')$ and $\boldsymbol{\eta}_{-1}^* = \boldsymbol{\eta}_{-1} + \mathbf{D}_{-1}\mathtt{vec}(\Gamma_0F_0')$. Using these expressions and letting $\mathbf{M}^\dagger = \mathbf{A}^{\frac{1}{2}}\mathbf{M}\mathbf{A}^{\frac{1}{2}}$, we can write

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

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 $\mathbf{M}^\dagger$, 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, A.2, and A.3, $\mathbf{B}_1$, $\mathbf{B}_2$ and $\mathbf{M}^\dagger$ 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 $\mathbf{R}_5$ are uniformly bounded. Hence, by Assumption F, we immediately have the results that $\frac{1}{n(T-r)}[\mathbf{y}_0'\mathbf{R}_1\mathbf{y}_0 - \mathrm{E}(\mathbf{y}_0'\mathbf{R}_1\mathbf{y}_0)] = o_p(1)$, and $\frac{1}{n(T-r)}[\mathbf{y}_0'\mathbf{R}_4 - \mathrm{E}(\mathbf{y}_0')\mathbf{R}_4] = o_p(1)$. The pointwise convergence of the quadratic terms $\frac{1}{n(T-r)}\mathbf{v}'\mathbf{R}_2\mathbf{v}$, and the bilinear term $\frac{1}{n(T-r)}\mathbf{y}_0'\mathbf{R}_3\mathbf{v}$, can be established by Assumptions B, E and results (v) and (vi) in Lemma A.4. The pointwise convergence of the linear terms $\frac{1}{n(T-r)}\mathbf{v}'\mathbf{R}_5$ can be proved using Chebyshev's inequality. Therefore, for $k = 1, 2, 3$, and all ℓ,

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

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 $\sup_{\delta\in\boldsymbol{\Delta}} |\frac{\partial}{\partial\omega}Q_{k,\ell}(\delta)| = O_p(1)$, 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 $\sup_{\delta\in\boldsymbol{\Delta}} \gamma_{\max}(\dot{\mathbf{A}}_{\lambda_3}) = \sup_{\delta\in\boldsymbol{\Delta}} \gamma_{\max}(B_3'W_3 + W_3'B_3) < c$. Moreover, $\sup_{\delta\in\boldsymbol{\Delta}} \gamma_{\max}(\mathbf{G}) = \sup_{\delta\in\boldsymbol{\Delta}} \gamma_{\max}(\mathbf{X}(\mathbf{X}'\mathbf{A}\mathbf{X})^{-1}\mathbf{X}'\mathbf{A}) = \sup_{\delta\in\boldsymbol{\Delta}} \gamma_{\max}(\mathbf{A}^{\frac{1}{2}}\mathbf{X}(\mathbf{X}'\mathbf{A}\mathbf{X})^{-1}\mathbf{X}'\mathbf{A}^{\frac{1}{2}}) = 1$. By applying Lemmas A.1, A.4, and Assumption F repeatedly, we can show that, for $k = 1, 2, 3$, and all ℓ, $\sup_{\delta\in\boldsymbol{\Delta}} |\frac{\partial}{\partial\lambda_3}Q_{k,\ell}(\delta)| = O_p(1)$. For example, for $|\frac{\partial}{\partial\lambda_3}Q_{1,1}(\delta)|$,

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

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, $\sup_{\delta\in\boldsymbol{\Delta}} |\frac{\partial}{\partial\phi_s}Q_{k,\ell}(\delta)| = O_p(1), s = 1, 2, \ldots, k_\phi$. 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,

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

It left to show that $\mathrm{E}Q_4(\delta) = \frac{1}{n(T-r)}\mathrm{E}[(\mathbf{B}_1^*\mathbf{Y}^\circ - \mathbf{B}_2^*\mathbf{Y}^\circ_{-1})'\mathbf{P}(\mathbf{B}_1^*\mathbf{Y}^\circ - \mathbf{B}_2^*\mathbf{Y}^\circ_{-1})] \to 0$, uniformly in $\delta \in \boldsymbol{\Delta}$. By Assumption D, $\gamma_{\min}\big(\frac{\mathbf{X}'\mathbf{A}\mathbf{X}}{nT}\big) > \underline{c}_x$. We have by the assumptions in Theorem 3.1 and Assumption D, $\mathrm{E}Q_4 = \frac{1}{n(T-r)}\mathrm{tr}[\mathbf{A}\mathbf{X}(\mathbf{X}'\mathbf{A}\mathbf{X})^{-1}\mathbf{X}'\mathbf{A}\mathrm{Var}(\mathbf{B}_1\mathbf{Y} - \mathbf{B}_2\mathbf{Y}_{-1})] \leq \frac{1}{n(T-r)}\gamma^2_{\max}(\mathbf{A})\gamma^{-1}_{\min}\big(\frac{\mathbf{X}'\mathbf{A}\mathbf{X}}{nT}\big)\bar{c}_y\ \frac{1}{nT}\mathrm{tr}(\mathbf{X}'\mathbf{X}) = O(n^{-1})$. Hence, $\hat{\sigma}^2_v(\delta) - \bar{\sigma}^2_v(\delta) \xrightarrow{p} 0$, 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 $\mathbf{Y}$ and $\mathbf{Y}_{-1}$ in (3.8), 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 (3.14), we write the AQS vector as linear combinations of terms linear or quadratic in $\mathbf{Z}$ and bilinear in $\mathbf{Z}$ and $\mathbf{y}_0$. Using $\mathbf{Z} = \mathbf{B}_{30}^{-1}\mathbf{v} + \mathtt{vec}(\Gamma_0 F_0')$, and the matrix multiplication result $\mathtt{vec}(\Gamma_0 F_0')'\mathbf{M}_{F0}K = 0$ for any $nT\times 1$ vector K, the AQS vector at the true parameters can be written as follows:

$$S^*_{nT}(\boldsymbol{\psi}_0) = \begin{cases} \Pi_1^{\dagger\prime}\mathbf{v} \\ \mathbf{v}'\Psi_1^\dagger\mathbf{y}_0 + \mathbf{v}'\Phi_1^\dagger\mathbf{v} + \Pi_2^{\dagger\prime}\mathbf{v} - \mu_{\rho_0} \\ \mathbf{v}'\Psi_2^\dagger\mathbf{y}_0 + \mathbf{v}'\Phi_2^\dagger\mathbf{v} + \Pi_3^{\dagger\prime}\mathbf{v} - \mu_{\lambda_{10}} \\ \mathbf{v}'\Psi_3^\dagger\mathbf{y}_0 + \mathbf{v}'\Phi_3^\dagger\mathbf{v} + \Pi_4^{\dagger\prime}\mathbf{v} - \mu_{\lambda_{20}} \\ \mathbf{v}'\Phi_4^\dagger\mathbf{v} - \mu_{\lambda_{30}} \\ \mathbf{v}'\Phi_5^\dagger\mathbf{v} - \mu_{\sigma^2_{v0}} \\ \mathbf{v}'\Phi_{5+s}^\dagger\mathbf{v} + \Pi_{4+s}^{\dagger\prime}\mathbf{v},\ s = 1,\ldots,k_\phi \end{cases} \tag{B.7}$$

where $\Pi_1^\dagger = \frac{1}{\sigma^2_{v0}}(M_{F_0}\otimes B_{30})\mathbf{X}$, $\Pi_2^\dagger = \frac{1}{\sigma^2_{v0}}(M_{F_0}\otimes B_{30})\boldsymbol{\eta}^*_{-1}$, $\Pi_3^\dagger = \frac{1}{\sigma^2_{v0}}(M_{F_0}\otimes B_{30})\mathbf{W}_1\boldsymbol{\eta}^*$,
$\Pi_4^\dagger = \frac{1}{\sigma^2_{v0}}(M_{F_0}\otimes B_{30})\mathbf{W}_2\boldsymbol{\eta}^*_{-1}$, $\Pi_{4+s}^\dagger = \frac{1}{\sigma^2_{v0}}(M_{F_0}\dot{F}_{s0}(F_0'F_0)^{-1}F_0\otimes B_{30})\mathtt{vec}(\Gamma_0F_0')$, $s=1,\ldots,k_\phi$;
$\Phi_1^\dagger = \frac{1}{\sigma^2_{v0}}(M_{F_0}\otimes B_{30})\mathbf{D}_{-1}\mathbf{B}_{30}^{-1}$, $\Phi_2^\dagger = \frac{1}{\sigma^2_{v0}}(M_{F_0}\otimes B_{30})\mathbf{W}_1\mathbf{D}\mathbf{B}_{30}^{-1}$, $\Phi_3^\dagger = \frac{1}{\sigma^2_{v0}}(M_{F_0}\otimes B_{30})\mathbf{W}_2\mathbf{D}_{-1}\mathbf{B}_{30}^{-1}$,
$\Phi_4^\dagger = \frac{1}{\sigma^2_{v0}}(M_{F_0}\otimes W_3B_{30}^{-1})$, $\Phi_5^\dagger = \frac{1}{2\sigma^4_{v0}}\mathbf{M}_{F_0}$, $\Phi_{5+s}^\dagger = \frac{1}{\sigma^2_{v0}}[M_{F_0}\dot{F}_{s0}(F_0'F_0)^{-1}F_0'\otimes I_n]$, $s=1,\ldots,k_\phi$;
$\Psi_1^\dagger = \frac{1}{\sigma^2_{v0}}(M_{F_0}\otimes B_{30})\mathbf{Q}_{-1}$, $\Psi_2^\dagger = \frac{1}{\sigma^2_{v0}}(M_{F_0}\otimes B_{30})\mathbf{W}_1\mathbf{Q}$, and $\Psi_3^\dagger = \frac{1}{\sigma^2_{v0}}(M_{F_0}\otimes B_{30})\mathbf{W}_2\mathbf{Q}_{-1}$.

By Assumptions C, E, and Lemma A.1, the $nT\times nT$ matrices $\Phi^\dagger$ and $\Psi^\dagger$ are uniformly bounded in both row and column sums, and elements of vectors $\Pi^\dagger$ 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 $\mu = (0_k', \mu_{\sigma_v^2}, \mu_\rho, \mu_{\lambda_1}, \mu_{\lambda_2}, \mu_{\lambda_3}, 0_{k_\gamma}')'$. 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 $\mathbf{X}X_p$, $p = 1, \cdots, k$, be the pth column of $\mathbf{X}$. Denote $nT \times 1$ vectors, $\mathbf{X}_{k+1} = \mathbf{Y}_{-1}$, $\mathbf{X}_{k+2} = \mathbf{W}_1\mathbf{Y}$, $\mathbf{X}_{k+3} = \mathbf{W}_2\mathbf{Y}_{-1}$. 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:

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

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 (3.14) 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,

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

where $(\tilde{\lambda}_3, \tilde{\phi}')$ is between $(\bar{\lambda}_3, \bar{\phi}')$ and (λ_{30}, ϕ_0'). By (3.8), Assumptions C, E, F, Lemmas A.1, A.4, and the consistency of $\hat{\boldsymbol{\psi}}$, $\frac{1}{nT}\mathbf{X}_p'[M_F(\bar{\phi}) \otimes \Omega^{-1}(\bar{\lambda}_3)]\mathbf{X}_q = \frac{1}{nT}\mathbf{X}_p'(M_{F0} \otimes \Omega_0^{-1})\mathbf{X}_q + o_p(1)$.

For the convergence of $\dot{\mu}_{\beta_p,\beta_q}$, consider $\mu_{\rho,\rho}(\bar{\boldsymbol{\psi}}) = \mathrm{tr}[(\frac{\partial}{\partial \rho}\mathbf{D}_{-1}(\bar{\rho}, \bar{\lambda}))M_F(\bar{\phi})]$ for example. By the expression of $\mathbf{D}_{-1}$ in (3.8) 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 $\bar{\mathbf{Z}} = \mathbf{Z} - \sum_{p=1}^{k+3}\mathbf{X}_p(\bar{\beta}_p - \beta_{p0})$ and representations for $\mathbf{Y}$ and $\mathbf{Y}_{-1}$ given in (3.8), the convergence of other terms in $H(\boldsymbol{\psi})$ that involve $\mathbf{Z}(\theta)$ 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 (3.8), 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 3.3: First, the result $H_{nT}(\hat{\boldsymbol{\psi}}_{\mathtt{M}}) - H_{nT}(\boldsymbol{\psi}_0) \xrightarrow{p} 0$ is implied by result **(b)** in the proof of Theorem 3.2. Next, the result $\hat{\Sigma}_{nT} - \Sigma_{nT}(\boldsymbol{\psi}_0) \xrightarrow{p} 0$ follows from

(a) $\frac{1}{nT}\sum_{i=1}^{n}[\hat{\mathbf{g}}_i\hat{\mathbf{g}}_i' - \mathrm{E}(\mathbf{g}_i\mathbf{g}_i')] = o_p(1)$, and **(b)** $\frac{1}{nT}[\Upsilon(\hat{\boldsymbol{\psi}}) - \Upsilon(\boldsymbol{\psi}_0)] = o_p(1)$.

By the expression of Υ presented in Section 4, the proof of (**b**) is straightforward by the MVT and consistency of $\hat{\boldsymbol{\psi}}_{\mathtt{M}}$. We focus on the proof of (**a**), which follows if

(*i*) $\frac{1}{nT}\sum_{i=1}^{n}(\hat{\mathbf{g}}_i\hat{\mathbf{g}}_i' - \mathbf{g}_i^*\mathbf{g}_i^{*\prime}) \xrightarrow{p} 0$,

(*ii*) $\sum_{i=1}^{n}\mathbf{g}_i^*\mathbf{g}_i^{*\prime} = \sum_{i=1}^{n}\mathbf{g}_i\mathbf{g}_i'$, and

(*iii*) $\frac{1}{nT}\sum_{i=1}^{n}[\mathbf{g}_i\mathbf{g}_i' - \mathrm{E}(\mathbf{g}_i\mathbf{g}_i')] \xrightarrow{p} 0$.

The proof of (*i*) is straightforward by MVT. We focus on the proof of (*ii*) and (*iii*).

Proof of (*ii*): Recall that $g_{ri}^* = g_{\Pi i}^*, g_{\Psi i}^*, g_{\Phi i}^*$ is obtained by replacing v_{it} by z_{it}^* in $g_{ri} = g_{\Pi i}, g_{\Psi i}, g_{\Phi i}$ presented in (3.17), (3.18) and (3.20). It suffices to show that, for $r = 1, \ldots, 4$, $\nu = 1, 2, 3$, and $\iota = 1, \ldots, 5 + k_\phi$,

$$\textstyle\sum_{i=1}^{n} g_{\kappa,i}^* g_{\varpi,i}^{*\prime} = \sum_{i=1}^{n} g_{\kappa,i} g_{\varpi,i}', \quad \text{for } \kappa, \varpi = \Pi_r, \Psi_\nu, \Phi_\iota.$$

First, we show that $\sum_{i=1}^{n} g_{\Pi_r,i}^* g_{\Pi_\nu,i}^* = \sum_{i=1}^{n} g_{\Pi_r,i} g_{\Pi_\nu,i}$ for $r, \nu = 1, \ldots, 4$. Assuming without loss of generality Π_{it} are scalars and letting b_i' is the ith row of B_{30}, we have by (3.17),

$$g_{\Pi_r,i}^* = \textstyle\sum_{t=1}^{T}\Pi_{r,it}v_{it} + \sum_{t=1}^{T}\Pi_{r,it}b_i'(\Gamma_0 f_{t0}) = g_{\Pi_r,i} + \sum_{t=1}^{T}\Pi_{r,it}b_i'(\Gamma_0 f_{t0}),$$

Let $\mathtt{diag}(A)$ be the diagonal matrix formed by the diagonal elements of A. We can write

$$\begin{aligned}\textstyle\sum_{i=1}^{n} g_{\Pi_r,i}^* g_{\Pi_\nu,i}^* &= \textstyle\sum_{i=1}^{n} g_{\Pi_r,i} g_{\Pi_\nu,i} + \sum_{i=1}^{n} g_{\Pi_r,i}\big[\sum_{t=1}^{T}\Pi_{\nu,it}b_i'(\Gamma_0 f_{t0})\big] \\ &\quad + \textstyle\sum_{i=1}^{n} g_{\Pi_\nu,i}\big[\sum_{t=1}^{T}\Pi_{r,it}b_i'(\Gamma_0 f_{t0})\big] + \sum_{i=1}^{n}\big[\sum_{t=1}^{T}\Pi_{r,it}b_i'(\Gamma_0 f_{t0})\big]\big[\sum_{s=1}^{T}\Pi_{\nu,it}b_i'(\Gamma_0 f_{s0})\big] \\ &= \textstyle\sum_{i=1}^{n}(g_{\Pi_r,i}g_{\Pi_\nu,i}) + g_{\Pi_r}'\mathtt{diag}(\mathbb{D}_{\Pi,\nu}F_0\Gamma_0'B_{30}') + g_{\Pi_\nu}'\mathtt{diag}(\mathbb{D}_{\Pi,r}F_0\Gamma_0'B_{30}') \\ &\quad + \mathtt{diag}(\mathbb{D}_{\Pi,r}F_0\Gamma_0'B_{30}')'\mathtt{diag}(\mathbb{D}_{\Pi,\nu}F_0\Gamma_0'B_{30}'),\end{aligned}$$

where $\mathbb{D}_{\Pi,r} = (\Pi_{r,1}, \Pi_{r,2}, \ldots, \Pi_{r,T})$ is a $n \times T$ matrix whose tth column corresponds to $\Pi_{r,t}$, the subvectors of Π_r corresponding to $t = 1, \ldots, T$. According to the expressions of Π_r in (3.2), $\mathbb{D}_{\Pi,r}$ can be written as $\mathbb{D}_{\Pi,r} = \mathbb{K}_r M_{F_0}$, where $\mathbb{K}_r$ are some $n \times T$ matrices constructed from $\mathbf{X}, W_\ell, \ell = 1, 2, 3$ and $\boldsymbol{\psi}_0$. Therefore we have $\mathbb{D}_{\Pi,r}F_0\Gamma_0'B_{30}' = \mathbb{K}_r M_{F_0}F_0\Gamma_0'B_{30}' = \mathbf{0}_{n\times n}$. Hence the result $\sum_{i=1}^{n} g_{\Pi_r,i}^* g_{\Pi_\nu,i}^* = \sum_{i=1}^{n} g_{\Pi_r,i} g_{\Pi_\nu,i}$ follows.

Second, we show that $\sum_{i=1}^n g^*_{\Psi_r,i} g^*_{\Psi_\nu,i} = \sum_{i=1}^n g_{\Psi_r,i} g_{\Psi_\nu,i}$, for $r, \nu = 1, 2, 3$. By (3.18), the bilinear term $g^*_{\Psi_r,i}$ can be written as,

$$g^*_{\Psi_r,i} = \sum_{t=1}^T \xi_{r,it} v_{it} + \sum_{t=1}^T \xi_{r,it} b_i'(\Gamma_0 f_{t0}) = g_{\Psi_r,i} + \sum_{t=1}^T \xi_{r,it} b_i'(\Gamma_0 f_{t0}).$$

So, we can write $\sum_{i=1}^n g^*_{\Psi_r,i} g^*_{\Psi_\nu,i}$ as

$$\begin{aligned}
&\sum_{i=1}^n g^*_{\Psi_r,i} g^*_{\Psi_\nu,i} = \sum_{i=1}^n g_{\Psi_r,i} g_{\Psi_\nu,i} + \sum_{i=1}^n g_{\Psi_r,i} \big[\sum_{t=1}^T \xi_{\nu,it} b_i'(\Gamma_0 f_{t0})\big] \\
&\quad + \sum_{i=1}^n g_{\Psi_\nu,i} \big[\sum_{t=1}^T \xi_{r,it} b_i'(\Gamma_0 f_{t0})\big] + \sum_{i=1}^n \big[\sum_{t=1}^T \xi_{r,it} b_i'(\Gamma_0 f_{t0})\big]\big[\sum_{s=1}^T \xi_{\nu,it} b_i'(\Gamma_0 f_{s0})\big] \\
&= \sum_{i=1}^n g_{\Psi_r,i} g_{\Psi_\nu,i} + g'_{\Psi_r} \mathrm{diag}(\mathbb{D}_{\xi,\nu} F_0 \Gamma_0' B_{30}') + g'_{\Psi_\nu} \mathrm{diag}(\mathbb{D}_{\xi,r} F_0 \Gamma_0' B_{30}') \\
&\quad + \mathrm{diag}(\mathbb{D}_{\xi,r} F_0 \Gamma_0' B_{30}')' \mathrm{diag}(\mathbb{D}_{\xi,\nu} F_0 \Gamma_0' B_{30}'),
\end{aligned}$$

where $\mathbb{D}_{\xi,r}$ is a $n \times T$ matrix whose t-th column is $\xi_{r,t} = \Psi_{r,t+} y_0$. According to the expressions of Ψ_r given in (3.2), $\mathbb{D}_{\xi,r}$ can also be written as $\mathbb{K}_r M_{F_0}$, where $\mathbb{K}_r$ are some $n \times T$ matrices constructed from $y_0, \mathbf{X}, W_\ell, \ell = 1, 2, 3$ and $\boldsymbol{\psi}_0$. Therefore we have $\mathbb{D}_{\Psi,r} F_0 \Gamma_0' B_{30}' = \mathbf{0}_{n\times n}$, and the result $\sum_{i=1}^n g^*_{\Psi_r,i} g^*_{\Psi_\nu,i} = \sum_{i=1}^n g_{\Psi_r,i} g_{\Psi_\nu,i}$ follows.

Third, we show that $\sum_{i=1}^n g^*_{\Phi_r,i} g^*_{\Phi_\nu,i} = \sum_{i=1}^n g_{\Phi_r,i} g_{\Phi_\nu,i}$ for $r = 1, \ldots, 5 + k_\gamma$. By (3.20), the quadratic term $g^*_{\Phi_r,i}$ can be written as

$$\begin{aligned}
g^*_{\Phi_r,i} &= \sum_{t=1}^T z^*_{it} \varphi_{r,it} + \sum_{t=1}^T (z^*_{it} z^d_{r,it} - d_{it}) \\
&= \sum_{t=1}^T v_{it} \varphi_{r,it} + \sum_{t=1}^T (v_{it} z^d_{r,it} - d_{it}) + \sum_{t=1}^T b_i'(\Gamma_0 f_{t0})(\varphi_{r,it} + z^d_{r,it}) \\
&= g_{\Phi_r,i} + \sum_{t=1}^T b_i'(\Gamma_0 f_{t0})(\varphi_{r,it} + z^d_{r,it}) \\
&= g_{\Phi_r,i} + \sum_{t=1}^T b_i'(\Gamma_0 f_{t0}) \varphi^*_{r,it}
\end{aligned}$$

where $\varphi^*_{r,it} = \varphi_{r,it} + z^d_{r,it}$. Then, we can write

$$\begin{aligned}
&\sum_{i=1}^n g^*_{\Phi_r,i} g^*_{\Phi_\nu,i} \\
= &\sum_{i=1}^n g_{\Phi_r,i} g_{\Phi_\nu,i} + \sum_{i=1}^n [g_{\Phi_r,i} \sum_{s=1}^T b_i'(\Gamma_0 f_{s0}) \varphi^*_{\nu,is}] + \sum_{i=1}^n [g_{\Phi_\nu,i} \sum_{t=1}^T b_i'(\Gamma_0 f_{t0}) \varphi^*_{r,it}] \\
&+ \sum_{i=1}^n [\sum_{t=1}^T b_i'(\Gamma_0 f_{t0}) \varphi^*_{r,it}][\sum_{s=1}^T b_i'(\Gamma_0 f_{s0}) \varphi^*_{\nu,is}] \\
= &\sum_{i=1}^n g_{\Phi_r,i} g_{\Phi_\nu,i} + g'_{\Phi_r} \mathrm{diag}(\mathbb{D}_{\varphi,\nu} F_0 \Gamma_0' B_{30}') + g'_{\Phi_\nu} \mathrm{diag}(\mathbb{D}_{\varphi,r} F_0 \Gamma_0' B_{30}') \\
&+ \mathrm{diag}(\mathbb{D}_{\varphi,r} F_0 \Gamma_0' B_{30}')' \mathrm{diag}(\mathbb{D}_{\varphi,\nu} F_0 \Gamma_0' B_{30}') = \sum_{i=1}^n g_{\Phi_r,i} g_{\Phi_\nu,i}
\end{aligned}$$

where $\mathbb{D}_{\varphi,r}$ is a $n \times T$ matrix whose tth column is $\varphi_{r,t} = \sum_{s=1}^T \Phi_{r,ts} z^*_s$. Similarly, by the expressions of Φ_r in (3.2), we have $\mathbb{D}_{\varphi,r} F_0 \Gamma_0' B_{30}' = \mathbf{0}_{n\times n}$. Hence, $\sum_{i=1}^n g^*_{\Phi_r,i} g^*_{\Phi_\nu,i} = \sum_{i=1}^n g_{\Phi_r,i} g_{\Phi_\nu,i}$.

Fourth, we examine the cross-product terms. Similarly to the early cases, we have

$$\begin{aligned}
&\sum_{i=1}^n g^*_{\Pi_r,i} g^*_{\Psi_\nu,i} = \sum_{i=1}^n g_{\Pi_r,i} g_{\Psi_\nu,i} + \sum_{i=1}^n g_{\Pi_r,i} [\sum_{t=1}^T \xi_{\nu,it} b_i'(\Gamma_0 f_{t0})] \\
&\quad + \sum_{i=1}^n g_{\Psi_\nu,i} [\sum_{t=1}^T \Pi_{r,it} b_i'(\Gamma_0 f_{t0})] + \sum_{i=1}^n [\sum_{t=1}^T \Pi_{r,it} b_i'(\Gamma_0 f_{t0})][\sum_{s=1}^T \xi_{\nu,it} b_i'(\Gamma_0 f_{s0})] \\
&= \sum_{i=1}^n g_{\Pi_r,i} g_{\Psi_\nu,i} + g'_{\Pi_r} \mathrm{diag}(\mathbb{D}_{\xi,\nu} F_0 \Gamma_0' B_{30}') + g'_{\Psi_\nu} \mathrm{diag}(\mathbb{D}_{\Pi,r} F_0 \Gamma_0' B_{30}') \\
&\quad + \mathrm{diag}(\mathbb{D}_{\Pi,r} F_0 \Gamma_0' B_{30}')' \mathrm{diag}(\mathbb{D}_{\xi,\nu} F_0 \Gamma_0' B_{30}') = \sum_{i=1}^n g_{\Pi_r,i} g_{\Psi_\nu,i}.
\end{aligned}$$

$$\sum_{i=1}^n g^*_{\Pi_r,i} g^*_{\Phi_\nu,i} = \sum_{i=1}^n g_{\Pi_r,i} g_{\Phi_\nu,i} + \sum_{i=1}^n g_{\Pi_r,i} [\sum_{t=1}^T b_i'(\Gamma_0 f_{t0}) \varphi^*_{\nu,it}]$$
$$+ \sum_{i=1}^n g_{\Phi_\nu,i} [\sum_{t=1}^T b_i'(\Gamma_0 f_{t0}) \Pi_{r,it}] + \sum_{i=1}^n [\sum_{t=1}^T b_i'(\Gamma_0 f_{t0}) \Pi_{r,it}][\sum_{t=1}^T b_i'(\Gamma_0 f_{t0}) \varphi^*_{\nu,it}]$$
$$= \sum_{i=1}^n g_{\Pi_r,i} g_{\Phi_\nu,i} + g'_{\Pi_r} \mathrm{diag}(\mathbb{D}_{\varphi,\nu} F_0 \Gamma_0' B_{30}') + g'_{\Phi_\nu} \mathrm{diag}(\mathbb{D}_{\Pi,r} F_0 \Gamma_0' B_{30}')$$
$$+ \mathrm{diag}(\mathbb{D}_{\Pi,r} F_0 \Gamma_0' B_{30}') \mathrm{diag}(\mathbb{D}_{\varphi,\nu} F_0 \Gamma_0' B_{30}') = \sum_{i=1}^n g_{\Pi_r,i} g_{\Phi_\nu,i},$$

$$\sum_{i=1}^n g^*_{\Psi_r,i} g^*_{\Phi_\nu,i} = \sum_{i=1}^n g_{\Psi_r,i} g_{\Phi_\nu,i} + \sum_{i=1}^n g_{\Psi_r,i} [\sum_{t=1}^T b_i'(\Gamma_0 f_{t0}) \varphi^*_{\nu,it}]$$
$$+ \sum_{i=1}^n g_{\Phi_\nu,i} [\sum_{t=1}^T b_i'(\Gamma_0 f_{t0}) \xi_{r,it}] + \sum_{i=1}^n [\sum_{t=1}^T b_i'(\Gamma_0 f_{t0}) \xi_{r,it}][\sum_{t=1}^T b_i'(\Gamma_0 f_{t0}) \varphi^*_{\nu,it}]$$
$$= \sum_{i=1}^n g_{\Psi_r,i} g_{\Phi_\nu,i} + g'_{\Psi_r} \mathrm{diag}(\mathbb{D}_{\varphi,\nu} F_0 \Gamma_0' B_{30}') + g'_{\Phi_\nu} \mathrm{diag}(\mathbb{D}_{\xi,r} F_0 \Gamma_0' B_{30}')$$
$$+ \mathrm{diag}(\mathbb{D}_{\xi,r} F_0 \Gamma_0' B_{30}')' \mathrm{diag}(\mathbb{D}_{\varphi,\nu} F_0 \Gamma_0' B_{30}') = \sum_{i=1}^n g_{\Psi_r,i} g_{\Phi_\nu,i}.$$

Summarizing all the results above, we have $\sum_{i=1}^n \mathbf{g}_i^* \mathbf{g}_i^{*\prime} = \sum_{i=1}^n \mathbf{g}_i \mathbf{g}_i'$.

Proof of (*iii*). To show $\frac{1}{nT}\sum_{i=1}^n [\mathbf{g}_i \mathbf{g}_i' - \mathrm{E}(\mathbf{g}_i \mathbf{g}_i')] \xrightarrow{p} 0$, it suffices to show that

$$\frac{1}{nT}\sum_{i=1}^n [g_{\kappa,i} g'_{\varpi,i} - \mathrm{E}(g_{\kappa,i} g'_{\varpi,i})] \xrightarrow{p} 0, \quad \text{for } \kappa, \varpi = \Pi_r, \Psi_\nu, \Phi_\iota,$$

where $r = 1, \dots, 4$, $\nu = 1, 2, 3$, and $\iota = 1, \dots, 5 + k_\phi$.

First, we show $\frac{1}{nT}\sum_{i=1}^n [g_{\Pi_r i} g_{\Pi_\nu i} - \mathrm{E}(g_{\Pi_r i} g_{\Pi_\nu i})] \xrightarrow{p} 0$. Letting $v_{i\cdot} = (v_{i1}, v_{i2}, \dots, v_{iT})'$ and $\Pi_{r,i\cdot}$ be similarly defined, we can write

$$\frac{1}{nT}\sum_{i=1}^n [g_{\Pi_r i} g_{\Pi_\nu i} - \mathrm{E}(g_{\Pi_r i} g_{\Pi_\nu i})] = \frac{1}{nT}\sum_{i=1}^n \Pi'_{r,i\cdot}(v_{i\cdot} v'_{i\cdot} - \sigma^2_{v0} I_T)\Pi_{\nu,i\cdot} \equiv \frac{1}{nT}\sum_{i=1}^n U_{n,i}.$$

By Assumptions A and B, $U_{n,i}$ are independent across i. Elements of $\Pi_r, r = 1, \dots, 4$ are uniformly bounded by Assumptions C, D, E, and Lemma A.1. Then, it is straightforward to show that $\frac{1}{nT}\sum_{i=1}^n U_{n,i} = o_p(1)$ by Chebyshev's inequality.

Second, we show $\frac{1}{nT}\sum_{i=1}^n [g_{\Psi_r i} g_{\Psi_\nu i} - \mathrm{E}(g_{\Psi_r i} g_{\Psi_\nu i})] \xrightarrow{p} 0$, $r, \nu = 1, 2, 3$. By (3.18), we have

$$\begin{aligned}
&\frac{1}{nT}\sum_{i=1}^n [g_{\Psi_r i} g_{\Psi_\nu i} - \mathrm{E}(g_{\Psi_r i} g_{\Psi_\nu i})] \\
&= \frac{1}{nT}\sum_{i=1}^n \xi'_{r,i\cdot}(v_{i\cdot} v'_{i\cdot} - \sigma^2_{v0} I_T)\xi_{\nu,i\cdot} + \frac{\sigma^2_{v0}}{nT}\sum_{i=1}^n [\xi'_{r,i\cdot}\xi_{\nu,i\cdot} - \mathrm{E}(\xi'_{r,i\cdot}\xi_{\nu,i\cdot})] \\
&= \frac{1}{nT}\sum_{i=1}^n U_{1n,i} + \frac{1}{nT}\sum_{i=1}^n U_{2n,i}.
\end{aligned}$$

Let $\{\mathcal{G}_{n,i}\}$ be the increasing sequence of σ -fields generated by $(v_{j1}, \dots, v_{jT}, j = 1, \dots, i)$, $i = 1, \dots, n, n \geq 1$. Let $\mathcal{F}_{n,0}$ be the σ -field generated by (v_0, y_0), and define $\mathcal{F}_{n,i} = \mathcal{F}_{n,0} \otimes \mathcal{G}_{n,i}$. Clearly, $\mathcal{F}_{n,i-1} \subseteq \mathcal{F}_{n,i}$ for each $n \geq 1$, i.e., $\{\mathcal{F}_{n,i}\}_{i=1}^n$ is an increasing sequence of σ-fields. As $\xi'_{r,i\cdot}$ is $\mathcal{F}_{n,i-1}$ -measurable, $\mathrm{E}(\mathrm{U_{1n,i}} \mid \mathcal{F}_{\mathrm{n,i-1}}) = 0$. Thus, $\{U_{1n,i}, \mathcal{F}_{n,i}\}$ form a M.D. array. Using Assumptions A, B, E, and F, it is easy to see that $\mathrm{E}\left|\mathrm{U}^{1+\epsilon}_{\mathrm{1n,i}}\right| \leq \mathrm{K_v} < \infty$, for some $\epsilon > 0$. Thus, $\{U_{1n,i}\}$ is uniformly integrable. With constant coefficients $\frac{1}{nT}$, the other two conditions of weak law of large numbers (WLLN) for MD array of Theorem 19.7 of Davidson (1994, p .299) are satisfied. Thus, $\frac{1}{nT}\sum_{i=1}^n U_{1n,i} \xrightarrow{p} 0$. The convergence of the second term $\frac{1}{nT}\sum_{i=1}^n U_{2n,i} \xrightarrow{p} 0$ follows from Assumption F.

Third, we show $\frac{1}{nT}\sum_{i=1}^n[g_{\Phi_r i}g_{\Phi_\nu i}-\mathrm{E}(g_{\Phi_r i}g_{\Phi_\nu i})]\xrightarrow{p}0$, $r,\nu=1,\ldots,5+k_\phi$, without loss of generality we show $\frac{1}{nT}\sum_{i=1}^n[g_{\Phi_r i}^2-\mathrm{E}(g_{\Phi_r i}^2)]\xrightarrow{p}0$, for $r=1,\ldots,5+k_\phi$. Recall expression (4.7), $g_{\Phi,i}=\sum_{t=1}^T v_{it}\varphi_{it}+\sum_{t=1}^T(v_{it}z_{it}^d-d_{it})$, where $\{\varphi_{it}\}=\varphi_t=\sum_{s=1}^T\left(\Phi_{ts}^u+\Phi_{ts}^\ell\right)z_s^*$, and $\left\{z_{it}^d\right\}=z_t^d=\sum_{s=1}^T\Phi_{ts}^d z_s^*$, further recall that $z_t^*=v_t+B_{30}\Gamma_0 f_{t0}$, we can write,

$$
\begin{aligned}
g_{\Phi_r i}=&\ \sum_{t=1}^T v_{it}\varphi_{r,it}+\sum_{t=1}^T(v_{it}z_{r,it}^d-d_{r,it})\\
=&\ \sum_{t=1}^T v_{it}\varphi_{r,it}^v+\sum_{t=1}^T(v_{it}v_{r,it}^*-d_{r,it})+\sum_{t=1}^T v_{it}c_{r,it}\\
=&\ v_{i\cdot}'\varphi_{r,i\cdot}^v+v_{i\cdot}'v_{r,i\cdot}^*-1_T'd_{r,i\cdot}+v_{i\cdot}'c_{r,i\cdot}
\end{aligned}
$$

where $\{\varphi_{r,it}^v\}=\varphi_t^v=\sum_{s=1}^T\left(\Phi_{r,ts}^u+\Phi_{r,ts}^\ell\right)v_s$, $\{v_{r,it}^*\}=v_t^*=\sum_{s=1}^T\Phi_{ts}^d v_s$, and $\{c_{r,it}\}=c_{r,t}=\sum_{s=1}^T\Phi_{r,ts}B_{30}\Gamma_0 f_{s0}$. It follows that for $r=1,\ldots,5+k_\phi$,

$$
\frac{1}{nT}\sum_{i=1}^n[g_{\Phi_r i}^2-\mathrm{E}(g_{\Phi_r i}^2)]=\sum_{k=1}^9 U_k,
$$

where $U_9=\frac{2}{nT}\sum_{i=1}^n\{(v_{i\cdot}'v_{r,i\cdot}^*)(v_{i\cdot}'c_{r,i\cdot})-\mathrm{E}[(v_{i\cdot}'v_{r,i\cdot}^*)(v_{i\cdot}'c_{r,i\cdot})]\}$,

$$
\begin{aligned}
&U_1=\frac{1}{nT}\sum_{i=1}^n\{(v_{i\cdot}'\varphi_{r,i\cdot}^v)^2-\mathrm{E}[(v_{i\cdot}'\varphi_{r,i\cdot}^v)^2]\}, && U_2=\frac{1}{nT}\sum_{i=1}^n\{(v_{i\cdot}'v_{r,i\cdot}^*)^2-\mathrm{E}[(v_{i\cdot}'v_{r,i\cdot}^*)^2]\},\\
&U_3=\frac{1}{nT}\sum_{i=1}^n\{(v_{i\cdot}'c_{r,i\cdot})^2-\mathrm{E}[(v_{i\cdot}'c_{r,i\cdot})^2]\}, && U_4=\frac{2}{nT}\sum_{i=1}^n(v_{i\cdot}'\varphi_{r,i\cdot}^v)(v_{i\cdot}'v_{r,i\cdot}^*),\\
&U_5=-\frac{2}{nT}\sum_{i=1}^n(v_{i\cdot}'\varphi_{r,i\cdot}^v)(1_T'd_{r,i\cdot}), && U_6=\frac{2}{nT}\sum_{i=1}^n(v_{i\cdot}'\varphi_{r,i\cdot}^v)(v_{i\cdot}'c_{r,i\cdot})\\
&U_7=-\frac{2}{nT}\sum_{i=1}^n(v_{i\cdot}'v_{r,i\cdot}^*)(1_T'd_{r,i\cdot}), && U_8=-\frac{2}{nT}\sum_{i=1}^n(v_{i\cdot}'c_{r,i\cdot})(1_T'd_{r,i\cdot}).
\end{aligned}
$$

For U_1, we can write $(v_{i\cdot}'\varphi_{i\cdot}^v)^2=(\sum_{t=1}^T v_{it}\varphi_{it}^v)^2=\sum_{t=1}^T(v_{it}\varphi_{it}^v)^2+\sum_{t=1}^T\sum_{s\neq t}v_{it}\varphi_{it}^v v_{is}\varphi_{is}^v$. The second term can be written as $\sum_{t=1}^T v_{it}\kappa_{it}$, where $\kappa_{it}=\sum_{s\neq t}\varphi_{it}^v v_{is}\varphi_{is}^v$. By Assumptions A and B, κ_{it} is independent of v_{it}. Recall that a_{its}' is the ith row of the $n\times n$ matrix $\Phi_{ts}^u+\Phi_{ts}^\ell$, we have $\mathrm{E}(\kappa_{it}^2)=\sigma_{v0}^6\sum_t\sum_s a_{its}'a_{its}$, which equals the (i,i) element of matrix $A=(\Phi^u+\Phi^\ell)(\Phi^u+\Phi^\ell)'$. By Assumption E and Lemma A.1, A is uniformly bounded in both row and column sums with elements of uniform order $O\left(h_n^{-1}\right)$. So, by Lemma A.4, we have $\frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T v_{it}\kappa_{it}=o_p(1)$. For the first term, as v_{it} is independent of φ_{it}^v, we have,

$$
\begin{aligned}
&\sum_{t=1}^T\{(v_{it}\varphi_{it}^v)^2-\mathrm{E}[(v_{it}\varphi_{it}^v)^2]\}=\sum_{t=1}^T\{v_{it}^2(\phi_{it}^u+\phi_{it}^\ell)^2-\mathrm{E}[(v_{it}\varphi_{it}^v)^2]\}\\
=&\ \sum_{t=1}^T(v_{it}^2-\sigma_{v0}^2)\phi_{it}^{u^2}+\sum_{t=1}^T(v_{it}^2-\sigma_{v0}^2)\phi_{it}^{\ell 2}+2\sum_{t=1}^T v_{it}^2\phi_{it}^u\phi_{it}^\ell+\sigma_{v0}^2\sum_{t=1}^T[\phi_{it}^{u^2}-\mathrm{E}(\phi_{it}^{u^2})]\\
&+\sigma_{v0}^2\sum_{t=1}^T[\phi_{it}^{\ell 2}-\mathrm{E}(\phi_{it}^{\ell 2})]\ \equiv\ \sum_{r=1}^5 H_{rn,i},
\end{aligned}
$$

where $\phi_{it}^u=\sum_{s=1}^T a_{its}^{u\prime}v_s$, $\phi_{it}^\ell=\sum_{s=1}^T a_{its}^{\ell\prime}v_s$, and $a_{its}^{u\prime}$, and $a_{its}^{\ell\prime}$ are the ith rows of Φ_{ts}^u, and Φ_{ts}^ℓ.

First, we consider $H_{1n,i}$. By Assumptions A and B, we have $\mathrm{E}(H_{1n,i})=0$, and for $i\neq j$,

$$
\mathrm{E}(H_{1n,i}H_{1n,j})=\mathrm{E}[\phi_{i\cdot}^{u\prime}(v_{i\cdot}v_{i\cdot}'-\sigma_{v0}^2I_T)\phi_{i\cdot}^u][\phi_{j\cdot}^{u\prime}(v_{j\cdot}v_{j\cdot}'-\sigma_{v0}^2I_T)\phi_{j\cdot}^u]=0.
$$

Therefore, $\{H_{1n,i}\}$ are uncorrelated across i with 0 mean. By Assumptions A and B, we have,

$$
\begin{aligned}
\mathrm{E}(H_{1n,i}^2)=&\ \sum_{t=1}^T\mathrm{E}[(v_{it}^2-\sigma_{v0}^2)^2\phi_{it}^{u^4}]=\sum_{t=1}^T\{\mathrm{E}[(v_{it}^2-\sigma_{v0}^2)^2]\mathrm{E}(\phi_{it}^{u^4})\}\\
=&\ (\mu_{v0}^{(4)}-\sigma_{v0}^4)\sum_{t=1}^T\mathrm{E}[(\sum_{s=1}^T a_{its}^{u\prime}v_s)^4],\ \text{and}
\end{aligned}
$$

$$
\begin{aligned}
&\sum_{t=1}^T \mathrm{E}[(\sum_{s=1}^T a_{its}^{u\prime} v_s)^4] = \sum_{t=1}^T \mathrm{E}[\sum_{p=1}^T \sum_{q=1}^T (a_{itp}^{u\prime} v_p)^2 (a_{itq}^{u\prime} v_q)^2 + \sum_{s=1}^T (a_{its}^{u\prime} v_s)^4] \\
&= \sum_{t=1}^T \{\sigma_{v0}^4 (\sum_{p=1}^T a_{itp}^{u\prime} a_{itp}^u)(\sum_{q=1}^T a_{itq}^{u\prime} a_{itq}^u) + \sum_{s=1}^T \mathrm{E}[(a_{its}^{u\prime} v_s)^4]\} \\
&= \sum_{t=1}^T \{\sigma_{v0}^4 (\sum_{p=1}^T a_{itp}^{u\prime} a_{itp}^u)^2 + \sum_{s=1}^T [\sigma_{v0}^4 (a_{its}^{u\prime} a_{its}^u)^2 + \mu_{v0}^{(4)} (\sum_{j=1}^n a_{its,j}^{u^4})]\} \\
&= \sigma_{v0}^4 \sum_{t=1}^T (\sum_{s=1}^T a_{its}^{u\prime} a_{its}^u)^2 + \sigma_{v0}^4 \sum_{t=1}^T \sum_{s=1}^T (a_{its}^{u\prime} a_{its}^u)^2 + \mu_{v0}^{(4)} \sum_{t=1}^T \sum_{s=1}^T \sum_{j=1}^n a_{its,j}^{u^4},
\end{aligned}
$$

where $a_{its,j}^u$ is the jth element of a_{its}^u, which is, by Assumption E and Lemma A.1 uniformly bounded. $a_{its}^{u\prime} a_{its}^u$ is the (i,i) element of $\Phi_{ts}^u \Phi_{ts}^{u\prime}$, which is, by Assumption E and Lemma A.1, uniformly bounded. So, as T is fixed and small, we have $\sum_{t=1}^T (\sum_{s=1}^T a_{its}^{u\prime} a_{its}^u)^2 \le C < \infty$, $\sum_{t=1}^T \sum_{s=1}^T (a_{its}^{u\prime} a_{its}^u)^2 \le C < \infty$, and $\sum_{j=1}^n a_{its,j}^{u^4} \le \max_j |a_{its,j}^2| \sum_{j=1}^n a_{its,j}^{u^2} = \max_j |a_{its,j}^2| (a_{its}^{u\prime} a_{its}^u) \le C < \infty$. Thus, we have $\mathrm{E}(H_{1n,i}^2) \le C < \infty$. Therefore, by the WLLN we have $\frac{1}{nT} \sum_{i=1}^n H_{1n,i} = o_p(1)$.

Next, consider $H_{2n,i} = \sum_{t=1}^T (v_{it}^2 - \sigma_{v0}^2) \phi_{it}^{\ell 2}$. As $\phi_{it}^\ell = \sum_s a_{its}^{\ell\prime} v_s$ is $\mathcal{G}_{n,i-1}$-measurable, we have $\mathrm{E}(H_{2n,i} | \mathcal{G}_{n,i-1}) = 0$. Thus $\{H_{2n,i}, \mathcal{G}_{n,i}\}$ form a M.D. array. Similar to $H_{1n,i}$, we show $\mathrm{E}(H_{2n,i}^2) \le C < \infty$. With constant coefficients $\frac{1}{nT}$, the other two conditions of WLLN for MD array of Theorem 19.7 of Davidson (1994, p .299) are satisfied. Thus, $\frac{1}{nT} \sum_{i=1}^n H_{2n,i} = o_p(1)$.

For $H_{3n,i}$, we can write $H_{3n,i} = \sum_{t=1}^T v_{it}^2 (\sum_{p=1}^T a_{itp}^{u\prime} v_p)(\sum_{s=1}^T a_{its}^{\ell\prime} v_s) = \sum_{s=1}^T v_s' \kappa_{is}$, where $\kappa_{is} = \sum_{t=1}^T \sum_{p=1}^T a_{its}^\ell a_{itp}^{u\prime} v_p v_{it}^2$. So we can write $\frac{1}{nT} \sum_{i=1}^n H_{3n,i} = \frac{1}{nT} \sum_{t=1}^T v_t' (\sum_{i=1}^n \kappa_{it})$, which is a bilinear form. By Assumptions A, B, E and Lemma A.1, we can verify the conditions of Lemma A.4 (vi) holds. Therefore we have $\frac{1}{nT} \sum_{i=1}^n H_{3n,i} = o_p(1)$.

Finally, the proof for convergence of $H_{4n,i}$ and $H_{5n,i}$ are the same. So, we only show the proof for $H_{4n,i}$. Write,

$$
\begin{aligned}
\frac{1}{nT} \sum_{i=1}^n H_{4n,i} &= \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T (\sum_{p=1}^T a_{itp}^{u\prime} v_p)(\sum_{q=1}^T a_{itq}^{u\prime} v_q) \\
&= \frac{1}{nT} \sum_{t=1}^T \sum_{p=1}^T \sum_{q=1}^T v_p' \sum_{i=1}^n (a_{itp}^u a_{itq}^{u\prime}) v_q \\
&= \frac{1}{nT} \sum_{t=1}^T \sum_{p=1}^T \sum_{q=1}^T v_p' \Phi_{tp}^{u\prime} \Phi_{tq}^u v_q = \frac{1}{nT} \mathbf{v}' \Phi^{u\prime} \Phi^u \mathbf{v}
\end{aligned}
$$

By Lemma A.1 and Assumption E, $\Phi^{u\prime}\Phi^u$ is uniformly bounded in either row or column sums. Thus, the result $\frac{1}{nT} \sum_{i=1}^n H_{4n,i} = o_p(1)$, and $\frac{1}{nT} \sum_{i=1}^n H_{5n,i} = o_p(1)$ follow from Lemma A.4. Combining the results above, we have $U_1 = o_p(1)$.

The $U_r, r = 2, 3, 7, 8, 9$, are the means of n independent terms, therefore their convergence can be shown using WLLN similarly as in the proof of $\frac{1}{nT} \sum_{i=1}^n H_{1n,i} = o_p(1)$ in U_1.

The proof of $U_r, r = 4, 5, 6$, are similar, and hence only the proof for U_4 is given. Write

$$
\begin{aligned}
U_4 &= \frac{2}{nT} \sum_{i=1}^n (v_{i\cdot}' \varphi_{i\cdot}^v)(v_{i\cdot}' v_{i\cdot}^*) \\
&= \frac{2}{nT} \sum_{i=1}^n [v_{i\cdot}' (\phi_{i\cdot}^u + \phi_{i\cdot}^\ell)](v_{i\cdot}' v_{i\cdot}^*) \\
&= \frac{2}{nT} \sum_{i=1}^n (v_{i\cdot}' \phi_{i\cdot}^u)(v_{i\cdot}' v_{i\cdot}^*) + \frac{2}{nT} \sum_{i=1}^n (v_{i\cdot}' \phi_{i\cdot}^\ell)(v_{i\cdot}' v_{i\cdot}^*) \\
&= \frac{2}{nT} \sum_{i=1}^n \phi_{i\cdot}^{u\prime} (v_{i\cdot} v_{i\cdot}' v_{i\cdot}^* - \mu_{v0}^{(3)} d_{i\cdot}) + \frac{2}{nT} \sum_{i=1}^n \phi_{i\cdot}^{\ell\prime} (v_{i\cdot} v_{i\cdot}' v_{i\cdot}^* - \mu_{v0}^{(3)} d_{i\cdot}) + \frac{2\mu_{v0}^{(3)}}{nT} \sum_{i=1}^n \varphi_{i\cdot}^{v\prime} d_{i\cdot}.
\end{aligned}
$$

The first term is the mean of n uncorrelated terms, its convergence can be shown using WLLN similarly as in the proof of $\frac{1}{nT}\sum_{i=1}^{n} H_{1n,i} = o_p(1)$ in U_1. The second term is the mean of a M.D. array , its convergence can be shown using WLLN for MD array similarly as in the proof of $\frac{1}{nT}\sum_{i=1}^{n} H_{2n,i} = o_p(1)$ in U_1. The convergence of the third term can be shown similarly as in the proof of $\frac{1}{nT}\sum_{i=1}^{n} H_{4n,i} = o_p(1)$ in U_1.

Subsequently, the cross-product terms: $\frac{1}{nT}\sum_{i=1}^{n}[g_{\Pi i}g_{\Phi i} - \mathrm{E}(g_{\Pi i}g_{\Phi i})]$, $\frac{1}{nT}\sum_{i=1}^{n}[g_{\Pi i}g_{\Psi i} - \mathrm{E}(g_{\Pi i}g_{\Psi i})]$ and $\frac{1}{nT}\sum_{i=1}^{n}[g_{\Psi i}g_{\Phi i} - \mathrm{E}(g_{\Psi i}g_{\Phi i})]$ can all be decomposed in a similar manner, and the convergence of each of the decomposed terms can be proved in a similar way. These complete the proof of Theorem 3.3. ∎

Proof of Corollary 3.1: The proof of part (i) relies on Lemmas C.1-C.3 in Appendix C in `Supplementary Material`. The main task of our analysis is to find the leading terms of the concentrated score function and its derivative with respect to coefficients at true parameters. Note that $\hat{F}(\psi)$ is not the true factor matrix F_0. It contains eigenvectors of $\mathbb{Z}'(\theta)\Omega^{-1}(\lambda_3)\mathbb{Z}(\theta)$, which is of high dimension as $T \to \infty$. And it enters the AQS functions in $M_{\hat{F}(\psi_0)}$. We need an explicit expression of $M_{\hat{F}(\psi_0)}$. To this end, we employ Lemma C.1, which gives an expansion of $M_{\hat{F}(\psi)}$ around ψ_0. Then we plug in the expression and find the asymptotic order of each term. Finally, analysis of the leading terms helps establishing the desired results.

(i) We first write out the AQS functions explicitly. Let $\tilde{S}^*_{nT,j}(\psi)$ denote the jth entry of the AQS function. We have the following expressions:

$$
\begin{aligned}
\tilde{S}^*_{nT,j}(\psi) &= \tfrac{1}{\sigma_v^2}\,\mathrm{tr}[X_j' B_3(\lambda_3)' B_3(\lambda_3)\mathbb{Z}(\theta)M_{\hat{F}(\psi)}], \quad \text{for } j = 1, \cdots, k,\\
\tilde{S}^*_{nT,k+1}(\psi) &= \tfrac{1}{\sigma_v^2}\,\mathrm{tr}[Y_{-1}' B_3(\lambda_3)' B_3(\lambda_3)\mathbb{Z}(\theta)M_{\hat{F}(\psi)}] - \mathrm{tr}[\mathbf{M}_{\hat{F}(\psi)}\mathbf{D}_{-1}(\rho,\lambda_1,\lambda_2)],\\
\tilde{S}^*_{nT,k+2}(\psi) &= \tfrac{1}{\sigma_v^2}\,\mathrm{tr}[(W_1Y)' B_3(\lambda_3)' B_3(\lambda_3)\mathbb{Z}(\theta)M_{\hat{F}(\psi)}] - \mathrm{tr}[\mathbf{M}_{\hat{F}(\psi)}\mathbf{W}_1\mathbf{D}(\rho,\lambda_1,\lambda_2)],\\
\tilde{S}^*_{nT,k+3}(\psi) &= \tfrac{1}{\sigma_v^2}\,\mathrm{tr}[(W_2Y_{-1})' B_3(\lambda_3)' B_3(\lambda_3)\mathbb{Z}(\theta)M_{\hat{F}(\psi)}] - \mathrm{tr}[\mathbf{M}_{\hat{F}(\psi)}\mathbf{W}_2\mathbf{D}_{-1}(\rho,\lambda_1,\lambda_2)],\\
\tilde{S}^*_{nT,k+4}(\psi) &= \tfrac{1}{\sigma_v^2}\,\mathrm{tr}[\mathbb{Z}(\theta)' B_3(\lambda_3)' W_3\mathbb{Z}(\theta)M_{\hat{F}(\psi)}] - (T-r)\,\mathrm{tr}[W_3 B_3^{-1}(\lambda_3)],\\
\tilde{S}^*_{nT,k+5}(\psi) &= \tfrac{1}{2\sigma_v^4}\,\mathrm{tr}[\mathbb{Z}(\theta)' B_3(\lambda_3)' B_3(\lambda_3)\mathbb{Z}(\theta)M_{\hat{F}(\psi)}] - \tfrac{n(T-r)}{2\sigma_v^2}.
\end{aligned}
$$

The estimate $\hat{F}(\psi)$ are eigenvectors of $\mathbb{Z}'(\theta)\Omega^{-1}(\lambda_3)\mathbb{Z}(\theta)$, which is of high dimension as $T \to \infty$. Lemma C.1 gives an asymptotic expansion of the projection matrix $M_{\hat{F}(\psi_0)}$. Lemma C.2 expands the concentrated AQS function at true parameter ψ_0 as

$$
\tilde{S}^*_{nT}(\psi_0) = \tilde{S}_{nT} + o_P(\sqrt{nT}),
$$

with the detailed expression of $\tilde{S}_{nT}$ being given therein, and further verifies that $\mathrm{E}(\tilde{S}^*_{nT}(\psi_0)) = o(\sqrt{nT})$ and $\tilde{S}^*_{nT}(\psi_0) = O_p(\sqrt{nT})$. Lemma C.3 studies the Hessian matrix and shows it is invertible at true parameters. These lead to the result $\hat{\psi} - \psi_0 = O_p(1/\sqrt{nT})$.

As $\hat{F}$ are eigenvectors of $\mathbb{Z}'(\hat{\theta})\Omega^{-1}(\hat{\lambda}_3)\mathbb{Z}(\hat{\theta})$, we can follow a similar analysis in Bai (2003) to show that $\hat{\phi}_s$ has $\sqrt{n}$ convergence rate.

(ii) In Appendix C, we show that $\tilde{S}_{nT}$ has the following representation:

$$
\tilde{S}_{nT}=\begin{cases}
\tilde{\Pi}_1'\mathbf{v}\\
\mathbf{v}'\tilde{\Psi}_1\mathbf{y}_0+\mathbf{v}'\tilde{\Phi}_1\mathbf{v}+\tilde{\Pi}_2'\mathbf{v}-\sigma_{v0}^2\operatorname{tr}(\tilde{\Phi}_1)\\
\mathbf{v}'\tilde{\Psi}_2\mathbf{y}_0+\mathbf{v}'\tilde{\Phi}_2\mathbf{v}+\tilde{\Pi}_3'\mathbf{v}-\sigma_{v0}^2\operatorname{tr}(\tilde{\Phi}_2)\\
\mathbf{v}'\tilde{\Psi}_3\mathbf{y}_0+\mathbf{v}'\tilde{\Phi}_3\mathbf{v}+\tilde{\Pi}_4'\mathbf{v}-\sigma_{v0}^2\operatorname{tr}(\tilde{\Phi}_3)\\
\mathbf{v}'\tilde{\Phi}_4\mathbf{v}-\sigma_{v0}^2\operatorname{tr}(\tilde{\Phi}_4)\\
\mathbf{v}'\tilde{\Phi}_5\mathbf{v}-\sigma_{v0}^2\operatorname{tr}(\tilde{\Phi}_5),
\end{cases}
$$

where the detailed expressions of the $\tilde{\Pi}$, $\tilde{\Phi}$ and $\tilde{\Psi}$ quantities being given therein. Based on this representation, We show that $\lim_{(n,T)\to\infty}\operatorname{Var}(\tilde{S}_{nT}/\sqrt{nT})=\lim_{(n,T)\to\infty}\tilde{\Sigma}_{nT}$ in Lemma C.4. Similar to the proof of Theorem 3.2, we can show that $\tilde{S}_{nT}/\sqrt{nT}\xrightarrow{D}N(0,\lim_{(n,T)\to\infty}\tilde{\Sigma}_{nT})$. Lemma C.3 finds the leading term of the Hessian matrix, which is denoted as $\tilde{H}_{nT}$. It follows that

$$\sqrt{nT}(\hat{\psi}-\psi_0)\xrightarrow{D}N(0,\lim_{(n,T)\to\infty}\tilde{H}_{nT}^{-1}\tilde{\Sigma}_{nT}\tilde{H}_{nT}^{-1}),$$

by Slutsky's Theorem. ∎

Proof of Corollary 3.2: The Hessian matrix $\tilde{H}_{nT}$ is studied in Lemma C.3 and the VC matrix $\tilde{\Sigma}_{nT}$ is studied in Lemma C.4. These together give the leading term of $\tilde{H}_{nT}^{-1}\tilde{\Sigma}_{nT}\tilde{H}_{nT}^{-1}$.

Next, we show that the VC matrix estimator $H_{nT}^{-1}(\hat{\boldsymbol{\psi}}_{\mathtt{M}})\widehat{\Sigma}_{nT}H_{nT}^{-1\prime}(\hat{\boldsymbol{\psi}}_{\mathtt{M}})$ is still valid for inference on ψ. That is, its ψ-ψ sub-matrix $[H_{nT}^{-1}(\hat{\boldsymbol{\psi}}_{\mathtt{M}})\widehat{\Sigma}_{nT}H_{nT}^{-1\prime}(\hat{\boldsymbol{\psi}}_{\mathtt{M}})]_{\psi\psi}$ does not differ asymptotically from $\tilde{H}_{nT}^{-1}\tilde{\Sigma}_{nT}\tilde{H}_{nT}^{-1}$ in probability. It is chanllaenging to prove this result as with large T, the matrices $H_{nT}(\hat{\boldsymbol{\psi}}_{\mathtt{M}})$ and $\widehat{\Sigma}_{nT}$ are both of high dimension and their entries are not all $O_p(1)$. Partition the two high-dimensional matrices according to ψ and ϕ, we can write

$$
\widehat{\Sigma}_{nT}=\begin{bmatrix}\widehat{\Sigma}_{nT,\psi\psi} & \widehat{\Sigma}_{nT,\psi\phi}\\ \widehat{\Sigma}_{nT,\phi\psi} & \widehat{\Sigma}_{nT,\phi\phi}\end{bmatrix}\quad\text{and } H_{nT}^{-1}(\hat{\boldsymbol{\psi}}_{\mathtt{M}})=\begin{bmatrix}[H_{nT}^{-1}(\hat{\boldsymbol{\psi}}_{\mathtt{M}})]_{\psi\psi} & [H_{nT}^{-1}(\hat{\boldsymbol{\psi}}_{\mathtt{M}})]_{\psi\phi}\\ [H_{nT}^{-1}(\hat{\boldsymbol{\psi}}_{\mathtt{M}})]_{\phi\psi} & [H_{nT}^{-1}(\hat{\boldsymbol{\psi}}_{\mathtt{M}})]_{\phi\phi}\end{bmatrix},
$$

and we have that $[H_{nT}^{-1}(\hat{\boldsymbol{\psi}}_{\mathtt{M}})\widehat{\Sigma}_{nT}H_{nT}^{-1\prime}(\hat{\boldsymbol{\psi}}_{\mathtt{M}})]_{\psi\psi}$ is the summation of the following four terms:
(i) $[H_{nT}^{-1}(\hat{\boldsymbol{\psi}}_{\mathtt{M}})]_{\psi\psi}\widehat{\Sigma}_{nT,\psi\psi}[H_{nT}^{-1}(\hat{\boldsymbol{\psi}}_{\mathtt{M}})]'_{\psi\psi}$; (ii) $[H_{nT}^{-1}(\hat{\boldsymbol{\psi}}_{\mathtt{M}})]_{\psi\psi}\widehat{\Sigma}_{nT,\psi\phi}[H_{nT}^{-1}(\hat{\boldsymbol{\psi}}_{\mathtt{M}})]'_{\psi\phi}$;
(iii) $[H_{nT}^{-1\prime}(\hat{\boldsymbol{\psi}}_{\mathtt{M}})]_{\psi\phi}\widehat{\Sigma}_{nT,\phi\psi}[H_{nT}^{-1}(\hat{\boldsymbol{\psi}}_{\mathtt{M}})]'_{\psi\psi}$; (iv) $[H_{nT}^{-1}(\hat{\boldsymbol{\psi}}_{\mathtt{M}})]_{\psi\phi}\widehat{\Sigma}_{nT,\phi\phi}[H_{nT}^{-1}(\hat{\boldsymbol{\psi}}_{\mathtt{M}})]'_{\psi\phi}$.
Our task is to find the asymptotic leading order of the above four terms. We present the key steps here with detailed arguments given in Appendix C:

1. Study the asymptotic properties of $\widehat{\Sigma}_{nT}$, with $(n,T)\to\infty$. Under large T, only entries of $\widehat{\Sigma}_{nT,\psi\psi}$ are of order $O_p(1)$ and entries of the other three sub-matrices of $\widehat{\Sigma}_{nT}$ are of order $o_p(1)$. However, as $\widehat{\Sigma}_{nT,\psi\phi}$ has $O(T)$ terms and $\widehat{\Sigma}_{nT,\phi\phi}$ has $O(T^2)$ terms, their influence on terms (ii-iv) of $[H_{nT}^{-1}(\hat{\boldsymbol{\psi}}_{\mathtt{M}})\widehat{\Sigma}_{nT}H_{nT}^{-1\prime}(\hat{\boldsymbol{\psi}}_{\mathtt{M}})]_{\psi\psi}$ are not negligible. Therefore, their leading

terms must be studied even if they are elementwise $o_p(1)$. The detailed analysis is given in Lemma C.5 of Appendix C.

2. Study the asymptotic properties of $H_{nT}(\hat{\boldsymbol{\psi}}_{\mathtt{M}})$ and its inverse, with $(n,T)\to\infty$. Similar to the last step, we find the leading terms of the inverse of $H_{nT}(\hat{\boldsymbol{\psi}}_{\mathtt{M}})$. To study $H_{nT}^{-1}(\hat{\boldsymbol{\psi}}_{\mathtt{M}})$, we have an additional challenge, we need to find a closed-form expression of $H_{nT}^{-1}(\hat{\boldsymbol{\psi}}_{\mathtt{M}})$. We first partition H_{nT} according to ψ and ϕ:

$$H_{nT}=\begin{bmatrix} H_{nT,\psi\psi} & H_{nT,\psi\phi} \\ H_{nT,\psi\phi} & H_{nT,\phi\phi} \end{bmatrix}.$$

where $H_{nT,\psi\psi}$ is $(k+5)\times(k+5)$ and $H_{nT,\phi\phi}$ is $(rT-r^2)\times(rT-r^2)$. Then we use the inverse formula for partition matrices to obtain:

$$H_{nT}^{-1}=\begin{bmatrix} H_{nT,*}^{-1} & -H_{nT,*}^{-1}H_{nT,\psi\phi}H_{nT,\psi\psi}^{-1} \\ -H_{nT,\phi\phi}^{-1}H_{nT,\phi\psi}H_{nT,*}^{-1} & H_{nT,\phi\phi}^{-1}+H_{nT,\phi\phi}^{-1}H_{nT,\phi\psi}H_{nT,*}^{-1}H_{nT,\psi\phi}H_{nT,\phi\phi}^{-1} \end{bmatrix},$$

where $H_{nT,*}=H_{nT,\psi\psi}-H_{nT,\psi\phi}(H_{nT,\phi\phi})^{-1}H_{nT,\phi\psi}$. We show that $H_{nT,\phi\phi}$ and $H_{nT,*}$ are invertible and find their leading terms. In Lemma C.6 (i-iv), we give the leading term of H_{nT}'s entries. And in Lemma C.6 (v) we show that $H_{nT,\phi\phi}$ has a nice closed form, which assists in finding its inverse. Then we find the leading term of $H_{nT,*}$ and show it is asymptotically equal to $\tilde{H}_{nT}$. Details of the analysis can be found in Lemma C.6 of Appendix C.

3. Use the leading terms of $H_{nT}^{-1}(\hat{\boldsymbol{\psi}}_{\mathtt{M}})$ and $\widehat{\Sigma}_{nT}$, studied in the first two steps to find the asymptotic leading terms of $[H_{nT}^{-1}(\hat{\boldsymbol{\psi}}_{\mathtt{M}})\widehat{\Sigma}_{nT}H_{nT}^{-1\prime}(\hat{\boldsymbol{\psi}}_{\mathtt{M}})]_{\psi\psi}$. And show it is asymptotically equal to $\tilde{H}_{nT}^{-1}\tilde{\Sigma}_{nT}\tilde{H}_{nT}^{-1}$. The proof of this result is given in Lemma C.7.

Thus, we have shown that the inference method given in Section 3.4 continues to be valid when T is large but small relatively to n. ■

Supplementary Material

The `Supplementary Material` contains additional lemmas for the proofs of Corollaries 3.1 and 3.2, and can be found online at `http://www.mysmu.edu/faculty/zlyang/`.

Acknowledgments

Early versions of this paper were presented at Shanghai Workshop of Econometrics 2019, Shanghai; XIII Conference of Spatial Econometrics Association 2019, Pittsburgh; 2022 Asian Meeting of the Econometric Society, Shenzhen; 2023 Asian Meeting of the Econometric Society,Beijing; 2023 Asian Meeting of the Econometric Society, Singapore. We thank Ingmar Prucha, Lung-Fei Lee, James LeSage, Yichong Zhang, and the participants of the conferences for their helpful comments. Liyao Li gratefully acknowledges the financial support provided by the National Natural Science Foundation of China under grant number 72203062. 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] Chamberlain, G., Rothschild, M., 1982. Arbitrage, factor structure, and mean-variance analysis on large asset markets. NBER Working Paper 996.

[10] Cui, G., Sarafidis, V., Yamagata, T., 2023. IV estimation of spatial dynamic panels with interactive effects: large sample theory and an application on bank attitude towards risk. The Econometrics Journal 26, 124–146.

[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] Higgins, A., 2023. Instrumental Variables for Dynamic Spatial Models with Interactive Effects. *Working Paper*, University of Oxford.

[14] Kato, T., 2013. Perturbation theory for linear operators. Springer Science & Business Media.

[15] 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.

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

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

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

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

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

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

[22] 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.

[23] 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.

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

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

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

[27] Miao, K., Li, K. and Su, L., 2020. Panel threshold models with interactive fixed effects. *Journal of Econometrics*, 219, 137-170.

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

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

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

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

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

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

[34] Qu, X., Lee, L.-F., YU, J., 2017. Estimation of spatial dynamic panel data models with endogenous time varying spatial weights matrices. *Journal of Econometrics* 197, 173-201.

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

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

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

[38] Vershynin, R., 2018. *High-dimensional probability: An introduction with applications in data science.* Cambridge university press.

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

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

[41] 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.

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

[43] 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)[.075]	.9982(.100)[.100]	.9770(.103)[.073]	.9998(.100)[.098]	.9736(.104)[.075]	.9955(.104)[.098]
1	.9744(.106)[.078]	.9925(.103)[.099]	.9691(.110)[.077]	.9890(.107)[.099]	.9782(.112)[.077]	.9965(.109)[.099]
1	.5801(.088)[−]	.9007(.141)[.132]	.5674(.167)[[−]	.8832(.212)[.202]	.5752(.123)[−]	.8930(.200)[.177]
.3	.2427(.072)[.047]	.2959(.062)[.062]	.2407(.083)[.045]	.2930(.068)[.065]	.2435(.073)[.046]	.2953(.060)[.061]
.2	.1766(.141)[.099]	.1929(.129)[.124]	.1741(.136)[.097]	.1936(.121)[.127]	.1720(.133)[.098]	.1904(.125)[.120]
.2	.2219(.076)[.057]	.2028(.079)[.077]	.2224(.075)[.057]	.2062(.077)[.078]	.2210(.078)[.057]	.2032(.078)[.080]
.2	.1881(.200)[.135]	.1931(.195)[.181]	.1835(.195)[.135]	.1869(.187)[.187]	.1896(.191)[.135]	.1892(.190)[.180]
	$n = 100$					
1	.9994(.076)[.057]	.9988(.075)[.073]	1.0005(.078)[.056]	1.0006(.078)[.072]	1.0009(.080)[.057]	1.0012(.078)[.074]
1	.9929(.073)[.058]	.9984(.072)[.072]	.9892(.075)[.057]	.9960(.073)[.072]	.9934(.076)[.058]	.9993(.075)[.073]
1	.6306(.065)[−]	.9497(.099)[.095]	.6196(.134)[−]	.9341(.204)[.185]	.6300(.098)[−]	.9493(.150)[.141]
.3	.3117(.058)[.030]	.2996(.046)[.047]	.3146(.064)[.030]	.2990(.050)[.051]	.3137(.062)[.030]	.3016(.049)[.050]
.2	.1956(.092)[.072]	.1936(.091)[.091]	.2062(.085)[.070]	.2023(.084)[.087]	.1960(.093)[.072]	.1947(.092)[.089]
.2	.1869(.079)[.056]	.1989(.073)[.073]	.1829(.081)[.055]	.1959(.075)[.076]	.1877(.079)[.055]	.1994(.074)[.074]
.2	.1921(.133)[.101]	.1971(.133)[.132]	.1799(.127)[.101]	.1899(.127)[.130]	.1939(.134)[.101]	.1983(.133)[.130]
	$n = 200$					
1	.9851(.051)[.041]	1.0003(.053)[.052]	.9852(.052)[.040]	1.0002(.054)[.052]	.9811(.051)[.041]	.9963(.053)[.051]
1	.9792(.051)[.040]	.9997(.052)[.051]	.9798(.053)[.040]	.9995(.054)[.051]	.9812(.054)[.040]	1.0014(.054)[.051]
1	.6252(.046)[−]	.9756(.075)[.072]	.6210(.092)[−]	.9688(.143)[.140]	.6262(.072)[−]	.9773(.119)[.107]
.3	.2571(.031)[.024]	.3003(.034)[.033]	.2583(.034)[.024]	.3002(.036)[.036]	.2577(.033)[.024]	.3009(.037)[.035]
.2	.1874(.065)[.054]	.1974(.065)[.064]	.1903(.067)[.053]	.2000(.064)[.064]	.1937(.065)[.054]	.2012(.064)[.064]
.2	.2007(.048)[.037]	.1996(.052)[.050]	.1995(.047)[.037]	.1983(.050)[.050]	.1990(.049)[.037]	.1997(.052)[.050]
.2	.1993(.091)[.074]	.1980(.090)[.089]	.1980(.091)[.073]	.1976(.090)[.089]	.1960(.088)[.074]	.1960(.087)[.089]
	$n = 400$					
1	.9951(.036)[.029]	.9985(.036)[.036]	.9964(.036)[.029]	.9997(.036)[.035]	.9971(.036)[.029]	.9999(.036)[.036]
1	.9861(.037)[.029]	1.0005(.037)[.036]	.9858(.038)[.029]	1.0000(.037)[.036]	.9837(.037)[.029]	.9980(.036)[.037]
1	.6425(.032)[−]	.9899(.050)[.051]	.6367(.068)[−]	.9891(.105)[.105]	.6424(.051)[−]	.9880(.081)[.078]
.3	.2593(.027)[.018]	.2999(.023)[.023]	.2595(.031)[.018]	.3001(.027)[.027]	.2595(.029)[.018]	.3000(.025)[.024]
.2	.1993(.048)[.040]	.1994(.048)[.048]	.1985(.049)[.040]	.1999(.048)[.047]	.2008(.049)[.040]	.2002(.048)[.048]
.2	.2057(.030)[.024]	.2001(.031)[.031]	.2046(.030)[.023]	.1998(.032)[.032]	.2041(.030)[.024]	.1999(.031)[.032]
.2	.1954(.065)[.054]	.1995(.066)[.066]	.1994(.068)[.054]	.2035(.066)[.066]	.1915(.066)[.054]	.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)[.088]	.9988(.116)[.116]	.9624(.108)[.087]	.9971(.116)[.112]	.9599(.112)[.088]	.9934(.119)[.112]
1	.9811(.109)[.084]	1.0012(.109)[.107]	.9753(.108)[.083]	.9973(.109)[.102]	.9745(.111)[.084]	.9964(.111)[.104]
1	.6111(.096)[−]	.9072(.144)[.137]	.6096(.179)[−]	.9064(.237)[.232]	.6114(.135)[−]	.9134(.201)[.185]
.3	.2601(.064)[.049]	.3011(.070)[.069]	.2590(.067)[.049]	.3005(.072)[.069]	.2555(.065)[.049]	.2977(.071)[.068]
.2	.1403(.135)[.082]	.1716(.126)[.132]	.1366(.137)[.082]	.1768(.130)[.119]	.1360(.136)[.082]	.1685(.121)[.118]
.2	.2141(.099)[.069]	.2062(.092)[.093]	.2093(.099)[.069]	.2004(.091)[.086]	.2100(.101)[.070]	.2028(.091)[.087]
.2	.1477(.162)[.130]	.1908(.178)[.188]	.1547(.155)[.130]	.1797(.187)[.180]	.1451(.154)[.130]	.1849(.174)[.179]
	$n = 100$					
1	.9691(.079)[.061]	.9987(.081)[.079]	.9729(.080)[.060]	1.0017(.085)[.081]	.9674(.079)[.060]	.9956(.083)[.080]
1	.9577(.083)[.063]	.9973(.080)[.079]	.9594(.087)[.063]	.9966(.082)[.079]	.9601(.083)[.063]	.9995(.080)[.079]
1	.6444(.068)[−]	.9554(.104)[.099]	.6406(.135)[−]	.9497(.209)[.181]	.6447(.100)[−]	.9557(.1545)[.141]
.3	.2638(.050)[.035]	.3005(.053)[.052]	.2637(.059)[.034]	.2993(.061)[.060]	.2648(.055)[.035]	.3010(.059)[.058]
.2	.0689(.084)[.075]	.1881(.089)[.086]	.0686(.085)[.075]	.1867(.090)[.085]	.0687(.089)[.075]	.1816(.085)[.086]
.2	.3473(.086)[.066]	.2174(.083)[.080]	.3447(.085)[.066]	.2190(.089)[.082]	.3403(.086)[.066]	.2110(.085)[.082]
.2	.2132(.117)[.091]	.1917(.127)[.125]	.2093(.122)[.091]	.1845(.125)[.124]	.2212(.110)[.091]	.1887(.123)[.124]
	$n = 200$					
1	.9918(.042)[.035]	.9988(.041)[.042]	.9909(.044)[.035]	.9980(.043)[.043]	.9933(.043)[.035]	1.0002(.043)[.042]
1	.9935(.052)[.041]	.9979(.050)[.049]	.9957(.050)[.041]	.9999(.049)[.049]	.9939(.050)[.041]	.9989(.049)[.049]
1	.6683(.048)[−]	.9744(.071)[.069]	.6708(.097)[−]	.9779(.136)[.134]	.6694(.075)[−]	.9780(.103)[.100]
.3	.3105(.031)[.020]	.3001(.028)[.028]	.3118(.044)[.020]	.3004(.039)[.037]	.3096(.036)[.020]	.2992(.032)[.032]
.2	.0408(.037)[.059]	.1894(.063)[.062]	.0392(.039)[.059]	.1894(.065)[.062]	.0414(.035)[.059]	.1894(.061)[.062]
.2	.3381(.031)[.051]	.2095(.056)[.056]	.3378(.032)[.051]	.2082(.059)[.057]	.3367(.032)[.051]	.2115(.057)[.057]
.2	.2178(.096)[.065]	.1948(.085)[.085]	.2194(.096)[.065]	.1898(.087)[.085]	.2161(.092)[.065]	.1927(.086)[.085]
	$n = 400$					
1	.9561(.036)[.027]	1.0005(.036)[.036]	.9607(.036)[.027]	.9997(.036)[.036]	.9558(.036)[.027]	.9989(.036)[.036]
1	.9481(.041)[.029]	1.0001(.037)[.036]	.9508(.046)[.029]	.9991(.037)[.036]	.9467(.043)[.029]	.9997(.037)[.036]
1	.6382(.033)[−]	.9866(.051)[.050]	.6315(.066)[−]	.9797(.110)[.109]	.6375(.049)[−]	.9898(.083)[.082]
.3	.1532(.047)[.015]	.2999(.023)[.023]	.1618(.064)[.015]	.3001(.028)[.027]	.1527(.054)[.014]	.2995(.025)[.024]
.2	.1161(.048)[.048]	.2004(.057)[.056]	.1181(.048)[.047]	.1981(.054)[.055]	.1126(.047)[.048]	.1979(.057)[.056]
.2	.2611(.063)[.040]	.2001(.043)[.043]	.2664(.060)[.039]	.2008(.045)[.045]	.2551(.060)[.039]	.2007(.046)[.046]
.2	.1880(.083)[.047]	.1997(.060)[.059]	.1968(.078)[.047]	.1989(.058)[.059]	.1843(.083)[.048]	.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-CQMLE 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)[.064]	1.0212(.200)[.165]	.7876(.132)[.062]	1.0404(.208)[.167]	.7760(.122)[.063]	1.0350(.205)[.155]
1	.6464(.141)[.072]	.9876(.172)[.159]	.6812(.173)[.070]	.9923(.178)[.160]	.6705(.145)[.072]	.9965(.177)[.155]
1	.2120(.046)[−]	.7848(.176)[.170]	.2017(.061)[−]	.7028(.184)[.177]	.2104(.052)[−]	.7481(.189)[.179]
.3	-.1793(.108)[.045]	.2698(.110)[.098]	-.1395(.170)[.043]	.2616(.115)[.110]	-.1591(.132)[.045]	.2520(.119)[.113]
.2	.2638(.177)[.090]	.1941(.190)[.188]	.2488(.168)[.088]	.1931(.198)[.190]	.2487(.164)[.091]	.1898(.197)[.190]
.2	.2348(.151)[.085]	.2133(.143)[.142]	.2282(.147)[.079]	.2266(.143)[.140]	.2174(.154)[.083]	.2166(.147)[.141]
.2	.0139(.272)[.133]	.1574(.329)[.303]	.0476(.263)[.131]	.1521(.310)[.294]	.0374(.266)[.134]	.1706(.311)[.297]
	$n = 100$					
1	.7475(.135)[.066]	.9699(.141)[.144]	.7750(.149)[.063]	.9817(.142)[.142]	.7556(.143)[.065]	.9759(.144)[.147]
1	.7796(.104)[.051]	.9863(.105)[.109]	.7989(.119)[.050]	.9729(.106)[.109]	.7881(.109)[.051]	.9724(.110)[.114]
1	.2053(.031)[−]	.8981(.115)[.121]	.1968(.041)[−]	.9023(.149)[.146]	.2024(.036)[−]	.7435(.133)[.136]
.3	-.0547(.123)[.043]	.2906(.097)[.093]	-.0094(.169)[.040]	.2991(.101)[.092]	-.0454(.137)[.042]	.2837(.100)[.102]
.2	.1294(.254)[.095]	.1964(.160)[.163]	.1234(.241)[.089]	.1950(.163)[.166]	.1123(.237)[.094]	.1833(.164)[.167]
.2	.1771(.208)[.075]	.2011(.098)[.095]	.1797(.194)[.069]	.2024(.114)[.110]	.1675(.199)[.073]	.1987(.114)[.117]
.2	.1992(.302)[.109]	.1902(.202)[.201]	.2117(.288)[.105]	.1845(.204)[.207]	.2263(.287)[.108]	.1951(.212)[.215]
	$n = 200$					
1	.9759(.176)[.037]	1.0102(.087)[.087]	1.0022(.167)[.036]	1.0021(.088)[.087]	.9866(.168)[.037]	1.0014(.089)[.088]
1	.9668(.137)[.039]	1.0071(.071)[.072]	.9769(.123)[.038]	1.0055(.070)[.072]	.9739(.131)[.038]	1.0087(.074)[.075]
1	.2973(.029)[−]	.9489(.083)[.087]	.2837(.046)[−]	.9640(.096)[.099]	.2920(.036)[−]	.9348(.103)[.104]
.3	.2091(.192)[.022]	.3011(.050)[.048]	.2346(.181)[.021]	.3061(.051)[.049]	.2251(.184)[.022]	.3175051(.051)[.049]
.2	.1786(.103)[.052]	.1982(.083)[.084]	.1808(.094)[.050]	.1983(.084)[.084]	.1754(.106)[.051]	.1981(.090)[.091]
.2	.1900(.063)[.034]	.1993(.060)[.059]	.1858(.063)[.033]	.1994(.061)[.062]	.1843(.068)[.033]	.1982(.067)[.069]
.2	.1933(.139)[.073]	.1994(.125)[.123]	.1839(.126)[.072]	.1980(.127)[.129]	.1926(.138)[.073]	.1978(.131)[.147]
	$n = 400$					
1	.9289(.047)[.019]	.9996(.028)[.028]	.9290(.048)[.018]	.9989(.031)[.031]	.9301(.048)[.018]	.9984(.029)[.030]
1	.8905(.091)[.029]	.9963(.049)[.050]	.8978(.089)[.029]	.9983(.051)[.051]	.8925(.089)[.029]	.9865(.048)[.049]
1	.3138(.022)[−]	.9893(.071)[.071]	.3073(.034)[−]	.9888(.084)[.085]	.3095(.027)[−]	.9833(.083)[.083]
.3	.1682(.180)[.020]	.2996(.030)[.030]	.1970(.185)[.019]	.2988(.031)[.032]	.1807(.182)[.019]	.2983(.034)[.034]
.2	.1662(.043)[.026]	.1994(.031)[.031]	.1680(.046)[.025]	.1960(.032)[.033]	.1662(.044)[.026]	.1973(.032)[.033]
.2	.2073(.032)[.018]	.2003(.026)[.028]	.1999(.035)[.018]	.1970(.028)[.029]	.2041(.032)[.018]	.2000(.027)[.028]
.2	.1910(.078)[.045]	.1996(.074)[.075]	.1982(.076)[.045]	.1961(.075)[.076]	.1930(.078)[.045]	.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)[.062]	.9957(.069)[.065]	.9995(.070)[.062]	.9994(.070)[.065]	.9991(.069)[.062]	.9990(.069)[.066]
1	.9966(.070)[.063]	.9967(.070)[.066]	.9926(.072)[.062]	.9927(.072)[.065]	.9995(.069)[.062]	.9996(.069)[.065]
1	.8256(.078)[−]	.9176(.087)[.085]	.8199(.186)[−]	.9112(.207)[.170]	.8265(.133)[−]	.9186(.147)[.136]
.3	.2979(.038)[.034]	.2987(.038)[.035]	.3018(.037)[.034]	.3015(.037)[.035]	.2986(.038)[.034]	.2994(.038)[.035]
.2	.1941(.076)[.069]	.1940(.076)[.073]	.1971(.074)[.069]	.1971(.074)[.071]	.1978(.072)[.069]	.1976(.072)[.071]
.2	.2020(.064)[.058]	.2011(.064)[.061]	.1982(.063)[.057]	.1974(.063)[.061]	.1976(.063)[.057]	.1968(.063)[.061]
.2	.2064(.120)[.101]	.2017(.121)[.117]	.1983(.113)[.101]	.2033(.115)[.113]	.1975(.110)[.101]	.2028(.112)[.113]
	$n = 50$					
1	.9978(.044)[.041]	.9979(.044)[.042]	.9992(.045)[.041]	.9993(.045)[.043]	.9996(.046)[.041]	.9997(.046)[.043]
1	.9985(.046)[.045]	.9985(.046)[.047]	.9997(.048)[.045]	.9997(.048)[.046]	1.0007(.049)[.045]	1.0007(.049)[.047]
1	.8610(.059)[−]	.9568(.066)[.064]	.8686(.136)[−]	.9653(.141)[.139]	.8649(.098)[−]	.9611(.103)[.101]
.3	.2985(.026)[.024]	.2994(.026)[.026]	.2991(.027)[.024]	.3000(.027)[.027]	.2979(.026)[.024]	.2998(.026)[.026]
.2	.1973(.060)[.055]	.1974(.060)[.059]	.1980(.060)[.055]	.1981(.060)[.059]	.1952(.059)[.055]	.1983(.059)[.059]
.2	.2000(.042)[.038]	.1997(.042)[.041]	.1986(.044)[.038]	.1988(.044)[.042]	.2017(.042)[.038]	.2013(.042)[.042]
.2	.1984(.087)[.078]	.2012(.088)[.086]	.1963(.087)[.078]	.2011(.087)[.085]	.1987(.084)[.078]	.2013(.084)[.084]
	$n = 100$					
1	.9995(.029)[.028]	.9995(.029)[.030]	1.0000(.032)[.028]	1.0000(.032)[.032]	1.0009(.031)[.028]	1.0009(.031)[.030]
1	1.0013(.033)[.031]	1.0004(.033)[.033]	1.0016(.034)[.031]	1.0016(.034)[.033]	.9981(.033)[.031]	.9971(.033)[.033]
1	.8837(.041)[−]	.9841(.046)[.046]	.8843(.098)[−]	.9848(.107)[.105]	.8821(.071)[−]	.9882(.078)[.075]
.3	.2997(.018)[.017]	.3002(.018)[.018]	.2985(.019)[.017]	.2992(.019)[.018]	.2999(.018)[.017]	.3001(.018)[.018]
.2	.1961(.038)[.035]	.1986(.038)[.037]	.1990(.038)[.035]	.1989(.038)[.037]	.1998(.038)[.035]	.1997(.038)[.037]
.2	.2014(.029)[.027]	.2001(.029)[.029]	.2006(.029)[.027]	.2001(.029)[.029]	.1983(.029)[.027]	.1989(.029)[.029]
.2	.2006(.056)[.053]	.2006(.056)[.057]	.1982(.058)[.053]	.2002(.058)[.057]	.1982(.058)[.053]	.2003(.058)[.057]
	$n = 200$					
1	.9990(.023)[.022]	.9998(.023)[.023]	1.0005(.024)[.022]	1.0003(.024)[.024]	.9997(.023)[.022]	1.0002(.023)[.023]
1	.9990(.022)[.022]	.9997(.022)[.023]	.9996(.023)[.022]	.9998(.023)[.023]	1.0009(.023)[.022]	1.0001(.023)[.023]
1	.8901(.030)[−]	.9989(.033)[.033]	.8905(.070)[−]	.9981(.076)[.076]	.8886(.051)[−]	.9880(.057)[.054]
.3	.2978(.013)[.012]	.2999(.013)[.013]	.2971(.014)[.012]	.2999(.014)[.014]	.2975(.014)[.012]	.2998(.014)[.014]
.2	.2006(.028)[.027]	.2001(.028)[.028]	.1991(.029)[.027]	.1988(.029)[.029]	.1985(.029)[.027]	.1982(.029)[.029]
.2	.2003(.021)[.020]	.1999(.021)[.021]	.2015(.021)[.020]	.2001(.021)[.021]	.2001(.021)[.020]	.1996(.021)[.021]
.2	.1974(.042)[.040]	.1997(.042)[.042]	.2006(.043)[.040]	.2003(.043)[.043]	.2011(.043)[.040]	.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, $T = 10$, $m = 10$
$W_1 = W_3$: `Rook 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)[.062]	1.0017(.071)[.065]	1.0015(.070)[.062]	1.0013(.070)[.065]	1.0000(.070)[.062]	.9999(.070)[.066]
1	.9966(.069)[.062]	.9965(.069)[.064]	1.0007(.067)[.061]	1.0005(.067)[.063]	1.0017(.066)[.062]	1.0015(.066)[.065]
1	.8284(.078)[−]	.9208(.087)[.085]	.8161(.180)[−]	.9071(.200)[.169]	.8347(.130)[−]	.9278(.145)[.129]
.3	.2970(.037)[.033]	.2982(.037)[.035]	.2970(.037)[.032]	.2981(.037)[.036]	.2937(.039)[.033]	.2949(.039)[.036]
.2	.1950(.081)[.071]	.1949(.082)[.075]	.1956(.079)[.070]	.1952(.078)[.074]	.1975(.078)[.071]	.1971(.078)[.075]
.2	.1995(.051)[.046]	.1992(.051)[.049]	.1960(.053)[.046]	.1957(.053)[.048]	.1951(.052)[.046]	.1947(.052)[.048]
.2	.1888(.140)[.117]	.1888(.145)[.146]	.1798(.141)[.117]	.1795(.145)[.150]	.1851(.136)[.118]	.1844(.141)[.146]
	$n = 50$					
1	.9956(.046)[.041]	.9960(.046)[.043]	1.0012(.046)[.041]	1.0010(.046)[.044]	1.0014(.045)[.041]	1.0009(.045)[.042]
1	.9999(.046)[.045]	1.0001(.046)[.047]	.9995(.047)[.045]	.9997(.047)[.047]	1.0020(.047)[.045]	1.0022(.047)[.047]
1	.8663(.059)[−]	.9628(.066)[.063]	.8676(.137)[−]	.9643(.152)[.140]	.8673(.100)[−]	.9640(.111)[.102]
.3	.3017(.023)[.021]	.3004(.023)[.022]	.3001(.025)[.021]	.2988(.025)[.023]	.2990(.022)[.021]	.2977(.023)[.022]
.2	.1999(.043)[.042]	.2003(.043)[.044]	.1989(.045)[.042]	.1992(.045)[.043]	.1956(.046)[.042]	.1959(.046)[.044]
.2	.1986(.028)[.026]	.1993(.028)[.026]	.1986(.028)[.026]	.1987(.028)[.027]	.2004(.027)[.026]	.2005(.027)[.026]
.2	.1869(.090)[.081]	.1942(.092)[.091]	.1896(.090)[.081]	.1890(.092)[.090]	.1945(.091)[.081]	.1948(.092)[.090]
	$n = 100$					
1	1.0006(.031)[.028]	1.0005(.031)[.030]	1.0004(.031)[.028]	1.0004(.031)[.030]	1.0001(.030)[.028]	1.0001(.030)[.030]
1	1.0002(.034)[.031]	1.0002(.034)[.033]	1.0000(.033)[.031]	1.0000(.033)[.033]	.9986(.032)[.031]	.9986(.032)[.033]
1	.8836(.042)[−]	.9928(.047)[.046]	.8795(.095)[−]	.9773(.105)[.102]	.8807(.070)[−]	.9786(.078)[.075]
.3	.2993(.018)[.016]	.2998(.018)[.018]	.3012(.018)[.016]	.3006(.018)[.018]	.2992(.018)[.016]	.2987(.018)[.018]
.2	.1964(.041)[.038]	.1986(.041)[.040]	.1976(.040)[.038]	.1997(.040)[.040]	.1957(.040)[.038]	.1959(.040)[.040]
.2	.1997(.033)[.031]	.1999(.033)[.033]	.1976(.033)[.031]	.1997(.033)[.033]	.1984(.033)[.031]	.2003(.033)[.033]
.2	.2005(.069)[.063]	.2001(.070)[.069]	.1962(.067)[.063]	.1978(.068)[.068]	.1992(.067)[.063]	.2007(.067)[.068]
	$n = 200$					
1	.9994(.021)[.020]	.9996(.021)[.021]	.9990(.021)[.020]	.9993(.021)[.021]	.9993(.021)[.020]	.9995(.021)[.021]
1	.9999(.023)[.022]	.9999(.023)[.023]	.9986(.023)[.022]	.9996(.023)[.023]	.9994(.024)[.022]	.9995(.024)[.023]
1	.8909(.029)[−]	.9902(.032)[.033]	.8927(.071)[−]	.9923(.079)[.075]	.8944(.049)[−]	.9941(.055)[.055]
.3	.3002(.012)[.012]	.2999(.012)[.012]	.3008(.013)[.012]	.2999(.013)[.013]	.3001(.013)[.012]	.2999(.013)[.012]
.2	.1986(.028)[.026]	.1996(.028)[.028]	.1994(.026)[.026]	.1996(.026)[.026]	.1998(.029)[.026]	.2000(.029)[.028]
.2	.1992(.027)[.025]	.1998(.027)[.026]	.1984(.026)[.025]	.1988(.026)[.026]	.1979(.028)[.025]	.1983(.028)[.026]
.2	.1986(.050)[.045]	.1998(.050)[.049]	.1963(.047)[.045]	.1995(.048)[.048]	.1953(.048)[.045]	.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: DGP1, $T = 3$, $m = 10$
$W_1 = W_2 = W_3$: `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)[.063]	.9988(.154)[.151]	.7857(.196)[.061]	.9895(.142)[.137]	.7507(.185)[.062]	.9899(.155)[.151]
1	.7370(.181)[.076]	.9838(.172)[.160]	.8110(.199)[.071]	.9965(.154)[.148]	.7728(.190)[.074]	.9877(.162)[.157]
1	.1701(.039)[−]	.6797(.144)[.147]	.1607(.048)[−]	.6254(.195)[.155]	.1653(.043)[−]	.6560(.180)[.155]
.3	-.1715(.210)[.054]	.2939(.107)[.102]	-.0564(.262)[.049]	.2932(.096)[.090]	-.1249(.234)[.052]	.2885(.100)[.102]
.2	.0957(.282)[.130]	.1870(.190)[.191]	.1202(.246)[.114]	.1806(.171)[.167]	.1032(.264)[.124]	.1622(.183)[.191]
.2	.1705(.246)[.111]	.2053(.143)[.150]	.1852(.209)[.093]	.2024(.121)[.126]	.1716(.234)[.103]	.1899(.135)[.147]
.2	.1402(.356)[.151]	.1876(.303)[.301]	.1377(.329)[.142]	.1767(.290)[.284]	.1523(.345)[.147]	.1980(.307)[.315]
	$n = 100$					
1	.8124(.195)[.055]	.9979(.111)[.119]	.8778(.179)[.052]	.9943(.109)[.110]	.8396(.192)[.053]	.9967(.106)[.118]
1	.8458(.149)[.055]	.9950(.107)[.115]	.8929(.150)[.052]	.9886(.105)[.109]	.8674(.161)[.053]	.9924(.113)[.122]
1	.2444(.033)[−]	.7933(.103)[.119]	.2243(.052)[−]	.7402(.178)[.158]	.2360(.042)[−]	.7570(.144)[.139]
.3	.1514(.258)[.040]	.2972(.076)[.077]	.2215(.228)[.035]	.2999(.073)[.073]	.1780(.253)[.038]	.2965(.073)[.083]
.2	.1662(.177)[.091]	.1997(.147)[.149]	.1676(.162)[.079]	.1961(.135)[.140]	.1666(.176)[.085]	.2028(.136)[.160]
.2	.1957(.149)[.067]	.1976(.124)[.135]	.1806(.142)[.060]	.1973(.120)[.123]	.1877(.142)[.065]	.1985(.119)[.140]
.2	.1591(.243)[.111]	.1921(.206)[.224]	.1900(.214)[.104]	.1930(.197)[.212]	.1819(.248)[.106]	.1960(.207)[.233]
	$n = 200$					
1	.8223(.069)[.037]	.9987(.081)[.087]	.8431(.084)[.036]	.9986(.076)[.087]	.8371(.078)[.037]	1.0006(.079)[.085]
1	.7641(.076)[.038]	.9978(.078)[.086]	.7966(.102)[.037]	.9985(.073)[.085]	.7788(.088)[.038]	.9971(.078)[.085]
1	.2247(.025)[−]	.8775(.088)[.100]	.2199(.035)[−]	.8154(.140)[.145]	.2229(.028)[−]	.8644(.115)[.123]
.3	-.0425(.074)[.029]	.2982(.060)[.060]	-.0077(.120)[.028]	.2998(.057)[.063]	-.0255(.098)[.029]	.2987(.059)[.065]
.2	.1351(.143)[.068]	.1993(.101)[.112]	.1421(.133)[.065]	.1976(.099)[.109]	.1467(.133)[.066]	.2017(.101)[.111]
.2	.1101(.101)[.055]	.1999(.094)[.095]	.1249(.104)[.051]	.1996(.082)[.093]	.1195(.100)[.053]	.1996(.089)[.096]
.2	.1984(.185)[.082]	.2003(.131)[.152]	.1959(.170)[.080]	.1966(.130)[.151]	.1922(.168)[.081]	.1963(.138)[.151]
	$n = 400$					
1	.9381(.056)[.029]	.9991(.055)[.060]	.9462(.057)[.028]	.9994(.053)[.060]	.9397(.058)[.028]	.9995(.053)[.065]
1	.9412(.058)[.028]	.9980(.055)[.059]	.9548(.056)[.028]	.9986(.052)[.059]	.9449(.058)[.028]	.98941(.053)[.063]
1	.2859(.022)[−]	.9591(.059)[.070]	.2745(.035)[−]	.9230(.108)[.118]	.2811(.028)[−]	.9025(.085)[.088]
.3	.2165(.082)[.017]	.2993(.036)[.039]	.2236(.078)[.017]	.2990(.038)[.046]	.2217(.083)[.017]	.2973(.039)[.036]
.2	.2168(.072)[.040]	.2002(.069)[.078]	.2034(.071)[.039]	.1992(.070)[.074]	.2047(.076)[.039]	.1977(.071)[.084]
.2	.2156(.049)[.025]	.2001(.048)[.054]	.2166(.046)[.024]	.2005(.047)[.053]	.2150(.047)[.025]	.2008(.047)[.058]
.2	.1888(.096)[.054]	.1998(.097)[.108]	.1989(.099)[.053]	.1998(.098)[.106]	.1960(.105)[.054]	.2008(.101)[.117]

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.