# 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]*School of Statistics, East China Normal University, Shanghai, China*

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

[c]*School of Economics, Singapore Management University, Singapore*

February 23, 2025

**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. We apply our methods to study the peer effects in firms' innovation decisions.

**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

*Corresponding author. E-mail: `lyli@fem.ecnu.edu.cn`

# 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 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 of earlier work.

A major 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; and Cui et al. 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 $\gamma_i$ (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, in particular from the likelihood perspective due to technical difficulties caused by IFE.[2]

SL and BL both adopt the conditional QML (CQML) approach, given initial observations, to estimate similar first-order DSPD-IFE models. Under a simultaneous passage of $n$ and $T$ to $\infty$, the CQML estimators are consistent but have nonnegligible 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 adopts the 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 for 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 by bias-correcting the CQML estimators. The GMM method handles the IVP by taking advantage of sequential exogeneity in setting moments and the IPP by a novel *forward orthogonal deviations* (FOD) transformation that eliminates factor loadings and

[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*) $\gamma_i$ as fixed but $f_t$ random, (*ii*) $\gamma_i$ as random but $f_t$ fixed, and (*iii*) both as random (see Hsiao 2018 for details). The case (*i*) is also of interest in connection with the 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.

at the same time adjusts the degrees of freedom loss. The GMM method does not require further bias corrections for valid inferences but requires that $T$ be small. Cui et al. (2023) propose instrumental variable (IV) estimation of a simple DSPD-IFE model without spatial errors, under Pesaran's (2006) common correlated effects setup, which is valid for large $n$ and large $T$. Except KP, all the other three papers discussed above assume first-order spatial effects with time-invariant spatial weights. Clearly, these assumptions are too restrictive, in particular, from the 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, but use likelihood-based methods and focus on the two most important asymptotic scenarios: (*i*) $T$ is fixed and (*ii*) $T$ is large but small relative to $n$. The scenario (*i*) is studied by KP based on the GMM approach, but the likelihood-based approach has not been considered. Scenario (*ii*) has not been considered at all. We introduce M-estimation methods that 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, the 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. The outer products of these $n$ terms, together with a covariance adjustment, lead to a consistent estimator of the VC matrix in both cases where $T$ is fixed and $T = o(n)$. The proposed methods are extended to accommodate unknown heteroskedasticity, time-varying spatial weight matrices, high-order dynamic effects, high-order spatial effects, etc.

Our work complements KP's fixed-$T$ GMM by providing alternative, likelihood-based methods, but our methods are also valid when $T = o(n)$, and thereby cover 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

[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 methods are not applicable. 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. See also Arellano and Hanh (2007) for literature on bias correction for nonlinear panel models.

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 the M-estimator is more efficient than the GMM estimator under strict exogeneity. Our work also complements those 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 bias-correct the CCQS functions instead of the CQML estimators. As a result, our M-estimators are free from asymptotic biases, and our inferences for common parameters are valid as long as $T/n \to 0$. Moreover, our method can be extended to suit the scenario with $T/n \to c(> 0)$ by further correcting the bias in CCQS caused by estimating the factor parameters, but it is not considered in this paper as it can be quite involved.

The rest of the paper goes as follows. Section 2 discusses the model specifications. Section 3 introduces the 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 provides an empirical application on the peer effects of firms' innovation decisions. Section 7 concludes the paper. All technical proofs are collected in the 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 $\Gamma$ is the corresponding $n \times r$ matrix of factor loadings.

KP proposes a GMM for the estimation of the model in a small $T$ set-up, assuming $v_{it} \sim (0, \varrho_i(\gamma)\sigma_t^2)$, where $\varrho_i(\gamma)$ are functions with a finite number of parameters $\gamma$, $W_{\nu\ell t}$ that are sequentially exogenous and $x_t$ that 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 $\psi$ 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 $\Gamma$ 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.

The KP model specifies that spatial interactions among model disturbances act equally on their components $\Gamma f_t$ and $v_t$. An alternative specification may be as follows:

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

which stresses that spatial interactions occur only in 'remainder' errors, not in unobserved (unshown) individual and time specific effects $\Gamma f_t$. 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 the 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 $\Gamma^*$.

The model (2.1) specifies a single spatial autoregressive (SAR) process for disturbances driven by factors and idiosyncratic errors, $\Gamma f_t + v_t$, together, whereas Model (2.2) specifies a single SAR process driven only by $v_t$. A more general model would naturally be that the disturbances contain two SAR processes, driven independently by $\Gamma f_t$ 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 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, in 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 $\Gamma$ 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 assume homoskedasticity and their inference methods depend 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 that ($i$) offers an easy way to avoid the effect of IVP, ($ii$) allows for sequential exogeneity (of an unknown form) in spatial weight matrices and regressors, and ($iii$) avoids 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 a 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 a 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

[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. However, they are known to frequently exhibit poor properties in finite samples and may be asymptotically biased if $T$ is not treated as fixed.

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

In the model, $\rho y_{t-1}$ captures the time dynamic effects; the spatial lag term $\lambda_1 W_1 y_t$ captures the contemporaneous spatial interactions among cross-sectional units, the space-time lag term $\lambda_2 W_2 y_{t-1}$ captures the dynamic spatial interactions, and the spatial error 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 $\Gamma = (\gamma_1, \gamma_2, \dots, \gamma_n)'$ is an $n \times r$ matrix of unobserved individual-specific effects (factor loadings), whose rows, $\gamma_i'$, are individuals' heterogeneous (interactive) responses to 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 $F' = (f_1, \dots, f_T)$. 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 $z_t(\theta) = B_1(\lambda_1)y_t - B_2(\rho, \lambda_2)y_{t-1} - x_t\beta$, $\mathbb{Z}(\theta) = [z_1(\theta), z_2(\theta), \dots, z_T(\theta)]$, and $\Omega(\lambda_3) = \sigma_v^{-2}\mathrm{E}(u_t u_t') = (B_3'(\lambda_3)B_3(\lambda_3))^{-1}$. Maximizing $\ell_{nT}(\psi, \Gamma, F)$ under a set of constraints on $\{\gamma_i\}$ and $\{f_t\}$ gives the conditional quasi maximum likelihood (CQML) estimator $\hat\psi_{\mathtt{CQML}}$ of $\psi$.[5]

Solving the first order condition, $\frac{\partial}{\partial\Gamma}\ell_{nT}(\psi, \Gamma, F) = 0$, using (3.2),[6] we obtain the con-

[5]Under $W_1 = W_2$, SL show that $\hat\psi_{\mathtt{CQML}}$ is consistent only when $(n, T) \to \infty$, and that $\sqrt{nT}(\hat\psi_{\mathtt{CQML}} - \psi_0)$ 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 explicitly allowing 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.

strained CQML estimator of $\Gamma$ as a matrix function of $\theta$ and $F$:

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

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 $\psi$ and $F$:

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

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

### 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

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

The $\psi$-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, $\ell^c_{nT}(\psi, F)$ 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 $\ell^c_{nT}(\psi, \phi)$. Then we can derive the CCQS functions of $\psi$ and $\phi$.

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.

[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 $\Gamma$ and $F$ separately without restrictions as $\Gamma CC^{-1}F' = \Gamma F'$ 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 to establish the set of unbiased and consistent estimating functions. See Bai and Ng (2013) for a detailed discussion of alternative normalizations.

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 $\psi$ and $\phi$, $S^c_{nT}(\psi,\phi) = (\frac{\partial}{\partial\psi'}\ell^c_{nT}(\psi,\phi), \frac{\partial}{\partial\phi'}\ell^c_{nT}(\psi,\phi))'$, take the form:

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

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

However, we show that the $(\sigma_v^2, \rho, \lambda)$ components of $\lim_{n\to\infty}\frac{1}{nT}\mathrm{E}[S^c_{nT}(\psi_0,\phi_0)]$ and more seriously those of $\mathrm{plim}_{n\to\infty}\frac{1}{nT}S^c_{nT}(\psi_0,\phi_0)$ are generally not zero at the true $\psi_0$ and $\phi_0$. Thus, the CQML estimator of $(\psi,\phi)$ cannot be consistent as a necessary condition for a consistent estimation is violated. To see this, the following basic assumptions are required.

**Assumption A.** *The 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 $\Gamma$ are independent of 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 the time-invariant $W_\nu$ 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, $\mathbf{Z} = \mathbf{Z}(\theta_0)$, $\boldsymbol{\eta} = \mathbf{D}\mathbf{X}\beta_0$, $\boldsymbol{\eta}_{-1} = \mathbf{D}_{-1}\mathbf{X}\beta_0$,

$\mathbf{Q} = \texttt{bdiag}(\mathcal{B}_0, \mathcal{B}_0^2, \ldots, \mathcal{B}_0^T)$, $\mathbf{Q}_{-1} = \texttt{bdiag}(I_n, \mathcal{B}_0, \ldots, \mathcal{B}_0^{T-1})$,

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

Based on the 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 $\phi$-part are given at the end of this subsection. The result of (3.9) clearly reveals that $\frac{1}{nT}\mathrm{E}[S_{nT}^c(\psi_0, \phi_0)] \neq 0$ and does not even converge to 0 when only $n$ approaches to $\infty$, and therefore $\mathrm{plim}_{n\to\infty}\frac{1}{nT}S_{nT}^c(\psi_0, \phi_0) \neq 0$.

Note that $\mathrm{E}[S_{nT}^c(\psi_0, \phi_0)]$ 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 $(\psi, \phi)$, free from $m$, $\Gamma$ 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 $\psi$ and $\phi$. Our **AQS or M-estimators** $\hat{\psi}_{\mathtt{M}}$ and $\hat{\phi}_{\mathtt{M}}$ of $\psi$ and $\phi$ are therefore defined as the solution of the estimating equations: $S_{nT}^*(\psi, \phi) = 0$.

**A computational note.** Given $\psi$, Model (2.4) reduces to a pure factor model. The constrained M-estimator of $F$ or $\phi$ can be obtained by maximizing $\frac{1}{nT}\mathrm{tr}[P_F Z'(\theta)\Omega^{-1}(\lambda_3)Z(\theta)]$,[10]

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

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 $\psi$-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 $\psi$: $\hat{\psi}(F) = \arg\{S^*_{nT,\psi}(\psi,F) = 0\}$,*
2. *Given $\psi$, 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 $\beta$ and $\sigma^2$, we obtain analytical solutions in terms of $\delta = (\rho, \lambda', \phi')'$:

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

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

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 $(\rho,\lambda)$-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 estimators (given $F$) of $\rho$ and $\lambda$, and thus the constrained estimators (given $F$) of $\beta$ and $\sigma^2$.

Before we move on 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 in a 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).*

**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 estimate of the VC matrix 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 $(\rho_0,\lambda_0)$-components of $S^*_{nT}(\psi_0,\phi_0)$ so that they possess the same property, we*

[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 $\Sigma_{\Gamma^*}$ and $\Sigma_v$ 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 $\Sigma_Z$ is a rotation of $F$. See Bai (2009) and Chamberlain and Rothschild (1982) for more detailed discussions.

[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 $\phi$-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 $(\beta,\sigma^2)$-components.

*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* $\gamma$ *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* $\gamma$ *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 $\phi$-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 $\phi_s$, $s = 1, \ldots, k_\phi$, is

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

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

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

This shows that the $\phi$-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 other words, given $\psi$, maximizing the CCQL function in (3.5) gives a consistent estimate of $\phi$, 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 (that is, their AQS equations cannot be solved analytically).

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

**Assumption C.** *(i) The parameter space* $\mathbf{\Delta}$ *of* $\delta$ *is compact and the true parameter vector* $\delta_0$ *lies in its interior; (ii) The number of factors* $r_0$ *is constant and less than* $T$*. The elements of* $\Gamma_0$ *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* $\lim_{n\to\infty} \frac{1}{nT}\mathbf{X}'\mathbf{M}_{F_0}\mathbf{X}$ *exists and is nonsingular.*

**Assumption E.** *(i) For* $\nu = 1, 2, 3$*, the elements* $w_{\nu,ij}$ *of* $W_\nu$ *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* $\lambda_\nu$ *in a compact parameter space* $\mathbf{\Lambda}_\nu$*, and* $0 < \underline{c}_\nu \leq \inf_{\lambda_\nu \in \mathbf{\Lambda}_\nu} \gamma_{\min}(B_\nu' B_\nu) \leq \sup_{\lambda_\nu \in \mathbf{\Lambda}_\nu} \gamma_{\max}(B_\nu' B_\nu) \leq \bar{c}_\nu < \infty$*.*

**Assumption F.** *For an* $n \times n$ *matrix* $\Phi$ *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 $v_{it}$ is independent of the 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, and spatial error terms. Assumption C($i$) is standard for establishing the consistency of the M-estimator $\hat{\delta}_{\mathtt{M}}$ of $\delta$. 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 Fensures that the initial observations 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 $\delta$, solving the AQS equations for $\beta$ and $\sigma_v^2$ 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 $\delta$-component of $S^*_{nT}(\psi, \phi)$ gives the concentrated AQS function $S^{*c}_{nT}(\delta)$ (a detailed expression is given in Appendix B). Similarly, let $\bar{S}^{*c}_{nT}(\delta)$ be the population counterpart of the concentrated AQS function (see Appendix B). It is easy to see that $S^{*c}_{nT}(\hat{\delta}_{\mathtt{M}}) = \mathbf{0}$, and $\bar{S}^{*c}_{nT}(\delta_0) = \mathbf{0}$. By Theorem 5.9 of van der Vaart (1998), $\hat{\delta}_{\mathtt{M}}$ will be consistent for $\delta_0$ if $\sup_{\delta \in \mathbf{\Delta}} \frac{1}{nT} \left\| S^{*c}_{nT}(\delta) - \bar{S}^{*c}_{nT}(\delta) \right\| \xrightarrow{p} 0$, and the following identification condition holds.

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

Similar conditions are required in Lee and Yu (2014) and in KP. As the AQS equations are highly nonlinear, it is extremely difficult to provide more primitive conditions, in general. KP provide a set of more primitive conditions similar to Kelejian and Prucha (1998) and Lee (2007) under which they show separately that the linear and quadratic moments have a unique solution at $\delta_0$. Using representations in (3.14), we see that AQS functions also contain linear combinations of a set of linear and quadratic moment conditions as in KP, and a similar set of primitive conditions may be established following their approach, except that the additional bilinear terms in our AQS functions may further complicate the matter.

**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)* $\inf_{\delta\in\boldsymbol{\Delta}}\gamma_{\min}\big[\mathrm{Var}(\mathbf{B}_1\mathbf{Y}-\mathbf{B}_2\mathbf{Y}_{-1})\big]\geq \underline{c}_y>0$. *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* $\hat{\sigma}^2_{v,\mathtt{M}}\xrightarrow{p}\sigma^2_{v0}$.

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

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

where $\Pi_1=\frac{1}{\sigma^2_{v0}}(M_{F_0}\otimes\Omega_0^{-1})\mathbf{X}$, $\Pi_2=\frac{1}{\sigma^2_{v0}}(M_{F_0}\otimes\Omega_0^{-1})\boldsymbol{\eta}_{-1}$, $\Pi_3=\frac{1}{\sigma^2_{v0}}(M_{F_0}\otimes\Omega_0^{-1})\mathbf{W}_1\boldsymbol{\eta}$, and
$\Pi_4=\frac{1}{\sigma^2_{v0}}(M_{F_0}\otimes\Omega_0^{-1})\mathbf{W}_2\boldsymbol{\eta}_{-1}$; $\Phi_1=\frac{1}{\sigma^2_{v0}}(M_{F_0}\otimes\Omega_0^{-1})\mathbf{D}_{-1}$, $\Phi_2=\frac{1}{\sigma^2_{v0}}(M_{F_0}\otimes\Omega_0^{-1})\mathbf{W}_1\mathbf{D}$,
$\Phi_3=\frac{1}{\sigma^2_{v0}}(M_{F_0}\otimes\Omega_0^{-1})\mathbf{W}_2\mathbf{D}_{-1}$, $\Phi_4=\frac{1}{\sigma^2_{v0}}(M_{F_0}\otimes B_{30}'W_3)$, $\Phi_5=\frac{1}{2\sigma^4_{v0}}(M_{F_0}\otimes\Omega_0^{-1})$, and
$\Phi_{5+s}=\frac{1}{\sigma^2_{v0}}[M_{F_0}\dot{F}_{s0}(F_0'F_0)^{-1}F_0'\otimes\Omega_0^{-1}]$, $s=1,\ldots,k_\phi$; $\Psi_1=\frac{1}{\sigma^2_{v0}}(M_{F_0}\otimes\Omega_0^{-1})\mathbf{Q}_{-1}$,
$\Psi_2=\frac{1}{\sigma^2_{v0}}(M_{F_0}\otimes\Omega_0^{-1})\mathbf{W}_1\mathbf{Q}$, and $\Psi_3=\frac{1}{\sigma^2_{v0}}(M_{F_0}\otimes\Omega_0^{-1})\mathbf{W}_2\mathbf{Q}_{-1}$; $\mu_{\sigma^2_v}=\frac{n(T-r)}{2\sigma^2_v}$,
$\mu_\rho=\mathrm{tr}(\mathbf{M}_{F_0}\mathbf{D}_{-1})$, $\mu_{\lambda_1}=\mathrm{tr}(\mathbf{M}_{F_0}\mathbf{W}_1\mathbf{D})$, $\mu_{\lambda_2}=\mathrm{tr}(\mathbf{M}_{F_0}\mathbf{W}_2\mathbf{D}_{-1})$, 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 to establish the asymptotic normality of $S^*_{nT}(\boldsymbol{\psi}_0)$, and the asymptotic normality of $\hat{\boldsymbol{\psi}}_{\mathtt{M}}$ through a first-order expansion of $S^*_{nT}(\hat{\boldsymbol{\psi}}_{\mathtt{M}})$ at $\boldsymbol{\psi}_0$.

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

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

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

### 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 $H_{nT}(\boldsymbol{\psi}_0)$ and $\Sigma_{nT}(\boldsymbol{\psi}_0)$. 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 $\Pi$, $\Psi$ and $\Phi$ 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 factor loadings $\Gamma_0$. 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 $\Gamma_0$ 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 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 non-stochastic quantities that depend on $\psi_0$ 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 advantage that $\{\mathbf{g}_i\}$ are nearly an MD sequence and are 'estimable':

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

The first term in (3.15) can be estimated by its sample analog $\frac{1}{nT}\sum_{i=1}^n \hat{\mathbf{g}}_i\hat{\mathbf{g}}_i'$, where $\hat{\mathbf{g}}_i$ is a plug-in estimate of $\mathbf{g}_i$. The full analytical expression of $\Upsilon(\psi_0) = \sum_{i=1}^n\sum_{j\neq i}\mathrm{E}(\mathbf{g}_i\mathbf{g}_j')$ is derived. Due to the way $\{\mathbf{g}_i\}$ are constructed, the $(k+5+k_\phi)\times(k+5+k_\phi)$ matrix $\Upsilon(\psi_0)$ does not involve the initial conditions or factor loadings and it depends only on $\psi_0$. Therefore, the covariance term $\Upsilon(\psi_0)$ can be consistently estimated using the plug-in method. The estimator of the VC matrix of the estimating functions is given by the following

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

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

Now, we present the details of the decomposition, $S^*_{nT}(\psi_0) = \sum_{i=1}^n \mathbf{g}_i$, and derive the correction term $\Upsilon(\psi_0)$. Recall that the components of the joint AQS vector $S^*_{nT}(\psi_0)$ are

[14] A valid model for $y_0$, as that in Su and Yang (2015) for an DSPD model with spatial error 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 $\Gamma$ and hence $v_t$ cannot be consistently estimated when $T$ is fixed.

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 $\Pi_t, \Phi_{ts}$ and $\Psi_{ts}$ the submatrices of $\Pi, \Phi$ and $\Psi$ partitioned according to $t, s = 1, \ldots, T$. Similarly, for a vector $K$, let $K_t$ denote its subvectors partitioned according to $t = 1, \ldots, T$. Denote the partial sum of the 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 $\Phi_{ts}, \Pi_t$ and other time-indexed quantities.

First, consider a linear term $\Pi'\mathbf{Z}$.[16] Write $\Pi'\mathbf{Z} = \Pi^{*\prime}\mathbf{v} + \Pi'\mathtt{vec}(\Gamma_0 F_0')$, where $\Pi^* = \mathbf{B}_3^{-1\prime}\Pi$. From (3.14), we see that $\Pi$ takes the form $\mathbf{M}_{F_0}K$ for a suitably defined non-stochastic 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, $\Pi'\mathtt{vec}(\Gamma_0 F_0') = K'(M_{F_0} \otimes I_n)\mathtt{vec}(\Gamma_0 F_0') = \mathrm{tr}(\Gamma_0 F_0' M_{F_0} \mathbb{K}') = 0$. Thus, $\Pi'\mathbf{Z} = \Pi^{*\prime}\mathbf{v}$. This leads to the following decomposition for any $\Pi$ 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 $\Pi_{it}^*$ is the $i$th element of $\Pi_t^*$. Clearly, $\{g_{\Pi,i}\}$ are not correlated 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 $\mathbf{Z}'\Psi\mathbf{y}_0 = \mathbf{v}'\Psi^*\mathbf{y}_0 + \mathtt{vec}(\Gamma_0 F_0')'\Psi\mathbf{y}_0$, 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 $\{v_t, t \geq 1\}$, we have the following MD decomposition of a bilinear term for any $\Psi$ 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 $\Phi$ defined in (3.14) as,

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

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

[17]By the expressions of $\Psi$ 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 $W_r, r = 1, 2, 3$.

[18]From (3.14), we see that $\Phi Z^*$ can also be written in the form $\mathbf{M}_F K$ for a suitably defined vector $K$ involving $\boldsymbol{\psi}_0$, $Z$, and $W_r, r = 1, 2, 3$.

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 $i$th diagonal element of $\Phi_{tt}$. 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 $\mathbf{g}_i$ which contains many uncorrelated terms. Moreover, this particular decomposition allows us to use the $M_F$ structure embedded in $\Pi$, $\Psi$, and $\Phi$, so that when we replace $v_{it}$ by $z_{it}$, the factor component is canceled in the 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 $\sum_{i=1}^n \mathbf{g}_i^*\mathbf{g}_i^{*\prime} = \sum_{i=1}^n \mathbf{g}_i\mathbf{g}_i'$, 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 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 $i$th row of the $n \times n$ matrix $\Phi^u_{ts} + \Phi^\ell_{ts}$, and $a_{ijts}$ be the $j$th 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, $\psi$ 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 $r < r_0$, $\mathrm{rank}(M_F(\phi)) < r_0$ and thus $M_F(\phi)$ cannot completely remove $\Gamma_0 F'_0$ from $Z(\theta)$. Therefore, no $\phi$ can satisfy $\mathrm{E}[S^*_{nT}(\psi, \phi)] = 0$. On the other hand, when $\mathrm{rank}(M_F(\phi)) > r_0$, there are infinitely many $\phi$ such that $M_F(\phi)$ can completely remove $\Gamma F'$. While $\phi$ is not identified when $r > r_0$, $\psi$ 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 $p$th 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 $\beta_1$-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 $\phi$ such that $\mathbf{M}_F(\phi)\mathtt{vec}(\Gamma_0 F_0') = 0$. When $r > r_0$, there are infinitely many $\phi$'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 by Moon and Weidner (2015) for regular panel models and by Shi and Lee (2017) for DSPD models. Kuersteiner and Prucha (2020), on the other hand, require $r$ 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 the 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 $\text{plim}_{n\to\infty} \Pr(\hat{r} > r_0) = 0$, 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 the 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. Under Assumptions A-G, the components of $\hat{\boldsymbol{\psi}}_{\mathtt{M}}$ remain consistent, but with different convergence rates. We show that $\hat{\psi}_{\mathtt{M}}$ is now $\sqrt{nT}$-consistent, but the elements of $\hat{\phi}_{\mathtt{M}}$ are $\sqrt{n}$-consistent. We derive the asymptotic distribution of $\hat{\psi}_{\mathtt{M}}$ and prove that the inference procedure in Theorem 3.3 remains valid for $\psi$. That is, one can use one inference procedure in both fixed $T$ and large $T$ scenarios.

To derive its asymptotic behavior, we can no longer study $\hat{\boldsymbol{\psi}}_{\texttt{M}} = (\hat{\psi}_{\texttt{M}}, \hat{\phi}_{\texttt{M}})'$ jointly, as $\phi$ is now high-dimensional. In addition, the asymptotic orders of the entries corresponding to $\psi$ and $\phi$ are distinct. Now we focus on $\psi$, the common parameter, and consider the concentrated AQS function with $\phi$ being 'concentrated' out:

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

where $\hat{\phi}(\psi) = \texttt{vec}[\widehat{F}_1(\psi)\widehat{F}_2^{-1}(\psi)]$ and $(\widehat{F}_1'(\psi), \widehat{F}_2'(\psi))' = \widehat{F}(\psi) = \texttt{eigv}_r\big(\frac{1}{nT}Z'(\theta)\Omega^{-1}(\lambda_3)Z(\theta)\big)$, the eigenvectors corresponding to the $r$ largest eigenvalues. Note that $\hat{\psi}_{\texttt{M}}$ 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 $\phi$, 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 $\psi_0$, so as to give an approximation to $M_{\hat{F}(\psi)}$ using the leading term(s). To proceed, let $\|\cdot\|_{\texttt{sp}}$ denote the spectrum norm, and assume the following:

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

*(ii) The $r$th eigenvalue of $(\Gamma' B_{30}' B_{30}\Gamma)(F_0'F_0)/(nT)$ converges to a constant $C_{min} > 0$.*

*(iii) Assume $H_{nT}(\psi_0)$ converges to a positive definite matrix, where $H_{nT}(\psi_0)$ is a $(k+4) \times (k+4)$ matrix given in Lemma C.3.*

Assumption H(i) bounds the spectrum norm of matrices related to the idiosyncratic errors. This is a standard assumption in the 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. Detailed discussions can be found in Vershynin (2018). Assumption H(ii) is a standard strong factor assumption. Assumption H(iii) assumes that the Hessian matrix is asymptotically nonsingular. The desired results are summarized as follows.

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

$$(i)\ \hat{\psi}_{\texttt{M}} - \psi_0 = O_p\big(\tfrac{1}{\sqrt{nT}}\big) \ \text{ and } \ \hat{\phi}_{s,\texttt{M}} - \phi_{s0} = O_p(\tfrac{1}{\sqrt{n}}), \text{ for each } s = 1, \ldots, k_\phi;$$

$$(ii)\ \sqrt{nT}\big(\hat{\psi}_{\texttt{M}} - \psi_0\big) \xrightarrow{D} N\Big(0, \lim_{(n,T)\to\infty} \tilde{H}^{-1}_{nT}(\psi_0)\tilde{\Sigma}_{nT}(\psi_0)\tilde{H}'^{-1}_{nT}(\psi_0)\Big),$$

*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 $\hat{\psi}_{\texttt{M}}$ is consistent with $\sqrt{nT}$ rate. On top of that, there are no asymptotic bias terms that may affect the inference as the bias arising from the initial values and factor loadings is eliminated through the recentering of the concentrated conditional quasi-scores. In contrast, the QML estimator proposed by Bai and Li (2021) exhibits three asymptotic bias terms, requiring bias correction for valid inference after estimation. For inference, Corollary 3.1 indicates that we can further find the limit of the covariance matrix and propose a new consistent estimator. However, this will lead to a different inference procedure compared to the fixed $T$ framework, which will greatly 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 VC martrix estimator proposed in Sec. 3.4 continues to be valid. More specifically, the $\psi$-$\psi$ block of the VC matrix given in Theorem 3.3 still consistently estimates the VC matrix of $\hat{\psi}_{\mathtt{M}}$ given in Corollary 3.1 under the large $T$ framework.

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

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

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

Corollary 3.2 confirms the inference procedure on the finite-dimensional parameter vector $\psi$ is valid for both the case of fixed $T$ and the case of $T = o(n)$. 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 $\psi$ and $\phi$ are of different asymptotic orders. We need to handle these two issues carefully in the proofs.

Statistical inference for $\phi_s$ or a finite number of linear contrasts $C\phi$ of $\phi$ might be of interest. With $\sqrt{nT}$ consistent rate of $\hat{\psi}_{\mathtt{M}}$, the estimate of the 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 $\Omega$ can be consistently estimated by $C\big[H_{nT}^{-1}(\hat{\boldsymbol{\psi}}_{\mathtt{M}})\hat{\Sigma}_{nT}H_{nT}^{\prime -1}(\hat{\boldsymbol{\psi}}_{\mathtt{M}})\big]_{\phi\phi}C'$.

## 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 $\mathbf{D}_{-1}$:

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

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

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

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 $B_{3t0} = I_n - \lambda_{30}W_{3t}$, $\mathbf{W}_3 = \texttt{bdiag}(W_{3t}, t = 1, \ldots, T)$ and $\mathbf{B}_3(\lambda_3) = I_{nT} - \lambda_3\mathbf{W}_3$. With time varying $W_{3t}$, concentrating out the factor loadings $\Gamma$ from the CQL function in (3.1) is no longer straightforward. Let $\Gamma_v = \texttt{vec}(\Gamma)$, 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 $\Gamma_v$

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

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

$$\begin{aligned} \ell_{nT}^c(\psi, F) = &- \tfrac{nT}{2}\log(2\pi\sigma_v^2) - \log|\mathbf{B}_3(\lambda_3)| + \log|\mathbf{B}_1(\lambda_1)| \\ &- \tfrac{1}{2\sigma_v^2}\mathbf{Z}'(\theta)\mathbf{B}_3'(\lambda_3)\mathbf{M}_{F^\dagger}(F, \lambda_3)\mathbf{B}_3(\lambda_3)\mathbf{Z}(\theta), \end{aligned} \tag{4.5}$$

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 be easily verified that (4.5) reduces to (3.5) when $W_3$ is time-invariant. The CCQS functions of $\psi$ and $\phi$ defined in (3.6) now becomes

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

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 to be 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 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:

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

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

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

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

Therefore, the $\rho$ and $\lambda$ components $\mathrm{E}[\frac{\partial}{\partial\psi}\ell^c_{nT}(\psi_0,\phi_0)]$ are no longer functions of only $(\psi_0,\phi_0)$; they contain the unknown heteroskedasticity matrix $\mathcal{H}$.

Although 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 a DSPD model with additive FE, instead of directly subtracting the expectation, we can find a set of quadratic terms in $\mathbf{Z}$ with expectations identical to (4.8)-(4.11). Define $\mathbf{R}_{\rho0}=\mathtt{diag}(\mathbf{B}_{30}\mathbf{D}_{-1}\mathbf{B}_{30}^{-1}\mathbf{M}_{F_0})\mathtt{diag}(\mathbf{M}_{F_0})^{-1}$, $\mathbf{R}_{\lambda_10}=\mathtt{diag}(\mathbf{B}_{30}\mathbf{W}_1\mathbf{D}\mathbf{B}_{30}^{-1}\mathbf{M}_{F_0})\mathtt{diag}(\mathbf{M}_{F_0})^{-1}$, and $\mathbf{R}_{\lambda_2 0} = \texttt{diag}(\mathbf{B}_{30}\mathbf{W}_2\mathbf{D}_{-1}\mathbf{B}_{30}^{-1}\mathbf{M}_{F_0})\texttt{diag}(\mathbf{M}_{F_0})^{-1}$, we have the following quadratic terms:

$$\sigma_{v0}^{-2}\mathrm{E}(\mathbf{Z}'\mathbf{B}_{30}'\mathbf{R}_{\rho 0}\mathbf{M}_{F_0}\mathbf{B}_{30}\mathbf{Z}) = \mathrm{tr}(\mathbf{D}_{-1}\mathbf{M}_{F_0}\mathbf{B}_{30}^{-1}\mathbf{H}\mathbf{B}_{30}), \tag{4.12}$$

$$\sigma_{v0}^{-2}\mathrm{E}(\mathbf{Z}'\mathbf{B}_{30}'\mathbf{R}_{\lambda_1 0}\mathbf{M}_{F_0}\mathbf{B}_{30}\mathbf{Z}) = \mathrm{tr}(\mathbf{D}\mathbf{M}_{F_0}\mathbf{B}_{30}^{-1}\mathbf{H}\mathbf{B}_{30}\mathbf{W}_1), \tag{4.13}$$

$$\sigma_{v0}^{-2}\mathrm{E}(\mathbf{Z}'\mathbf{B}_{30}'\mathbf{R}_{\lambda_2 0}\mathbf{M}_{F_0}\mathbf{B}_{30}\mathbf{Z}) = \mathrm{tr}(\mathbf{D}_{-1}\mathbf{M}_{F_0}\mathbf{B}_{30}^{-1}\mathbf{H}\mathbf{B}_{30}\mathbf{W}_2), \tag{4.14}$$

$$\sigma_{v0}^{-2}\mathrm{E}\{\mathbf{Z}'\mathbf{B}_{30}'[I_T \otimes \texttt{diag}(W_3 B_{30}^{-1})]\mathbf{B}_{30}\mathbf{M}_{F_0}\mathbf{Z}\} = \mathrm{tr}(\mathbf{M}_{F_0}\mathbf{B}_{30}^{-1}\mathbf{H}\mathbf{W}_3). \tag{4.15}$$

Taking the differences between these two sets and dropping the expectations leads to a set of unbiased estimating functions for $\rho$ and $\lambda$, robust against unknown $\mathcal{H}$. The $\phi$-component of the EF vector given in (3.10) is naturally robust against unknown $\mathcal{H}$ as shown in Footnote 13. Moreover, the $\beta'$ and $\sigma_v^2$ components also do not need further adjustment under heteroskedasticity. Therefore, a full set of EFs robust against unknown $\mathcal{H}$ is given below.

$$S_{nT}^r(\psi,\phi) = \begin{cases} \mathbf{X}'\mathbf{M}_F\mathbf{\Omega}^{-1}(\lambda_3)\mathbf{Z}(\theta), \\ \mathbf{Z}'(\theta)\mathbf{M}_F\mathbf{\Omega}_3(\lambda_3)\mathbf{Y}_{-1} - \mathbf{Z}'(\theta)\mathbf{B}_3'(\lambda_3)\mathbf{R}_\rho(\delta)\mathbf{M}_F\mathbf{B}_3(\lambda_3)\mathbf{Z}(\theta), \\ \mathbf{Z}'(\theta)\mathbf{M}_F\mathbf{\Omega}^{-1}(\lambda_3)\mathbf{W}_1\mathbf{Y} - \mathbf{Z}'(\theta)\mathbf{B}_3'(\lambda_3)\mathbf{R}_{\lambda_1}(\delta)\mathbf{M}_F\mathbf{B}_3(\lambda_3)\mathbf{Z}(\theta), \\ \mathbf{Z}'(\theta)\mathbf{M}_F\mathbf{\Omega}^{-1}(\lambda_3)\mathbf{W}_2\mathbf{Y}_{-1} - \mathbf{Z}'(\theta)\mathbf{B}_3'(\lambda_3)\mathbf{R}_{\lambda_2}(\delta)\mathbf{M}_F\mathbf{B}_3(\lambda_3)\mathbf{Z}(\theta), \\ \mathbf{Z}'(\theta)\mathbf{M}_F\mathbf{B}_3'(\lambda_3)\big\{\mathbf{W}_3 - [I_T \otimes \texttt{diag}(W_3 B_3^{-1}(\lambda_3))]\mathbf{B}_3(\lambda_3)\big\}\mathbf{Z}(\theta), \\ \frac{1}{2\sigma_v^2}\mathbf{Z}'(\theta)\mathbf{M}_F\mathbf{\Omega}^{-1}(\lambda_3)\mathbf{Z}(\theta) - \frac{n(T-r)}{2}, \\ \mathbf{Z}'(\theta)[M_F\dot{F}_s(F'F)^{-1}F' \otimes \Omega^{-1}(\lambda_3)]\mathbf{Z}(\theta), \;\; s = 1,\ldots,k_\phi. \end{cases} \tag{4.16}$$

We have $\mathrm{E}[S_{nT}^r(\psi_0,\phi_0)] = 0$. We further show that $\mathrm{plim}_{n\to\infty}\frac{1}{nT}S_{nT}^r(\psi_0,\phi_0) = 0$. Therefore, solving $S_{nT}^r(\psi,\phi) = 0$ would give consistent M-estimators of $\psi$ and $\phi$ 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 as before and remain valid. Our inference method will also go through **provided** that either the $\Gamma$ term or the $\lambda_3$ term exists. When both terms are absent, the $\mathcal{H}$-robust inference for $\sigma_0^2$ faces difficulty. This suggests that one should work with $S_{nT}^r(\psi,\phi)$ without the $\sigma^2$-component for $\mathcal{H}$-robust inference. This is particularly meaningful as the subvector is free from $\sigma^2$. 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 discussion 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 $\Gamma$, $\tilde{\Gamma}(\theta, F)$, 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 remains the same as (3.1) and (3.2). The rest of the 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 the 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 $\Gamma$ 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 the standard normal distribution, and $c$ is a constant. We use $c = 1$ for DGP1 and $c = 2$ for DGP2 as the numerical stability of the GMM method requires a significantly larger signal-to-noise ratio. The spatial weight matrices are generated according to the following schemes: Rook contiguity, Queen contiguity, or group interaction.[19]

The error ($v_t$) distribution can be ($i$) normal, ($ii$) normal mixture $(10\%N(0,4), 90\%N(0,1))$, or ($iii$) chi-squared with degrees of freedom 3. In both ($ii$) and ($iii$), the errors generated are standardized to have mean zero and variance $\sigma_v^2$. 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 robust standard errors (ses) based on the VC matrix estimate $H_{nT}^{-1}(\hat{\psi}_{\texttt{M}})\hat{\Sigma}_{nT}H_{nT}^{-1}(\hat{\psi}_{\texttt{M}})$, 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 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),[20] and the GMM method of Kuersteiner and Prucha (2020).[21]

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 $\lambda_1$ and $\lambda_2$ also show excellent finite sample performance, whereas that of $\lambda_3$ shows some small bias when errors are drawn from the chi-squared distribution. The BC-CQMLE of $\lambda_1$ performs quite well, but those of $\lambda_2$ and $\lambda_3$ are slightly biased. While the biases of the BC-CQMLEs of $\lambda_2$ and $\lambda_3$ are not severe, the 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-

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

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

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

estimators continue to perform very well, whereas the BC-CQMLEs for $\rho$ 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 $\rho$ and $\lambda_1$. 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 for the BC-CQMLE of the error variance. Further, the standard error 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 the number of factors is misspecified. The true number of factors is $r_0 = 1$ but the number of factors assumed in the estimation is $r = 2$. The proposed M-estimators perform reasonably well under misspecification. The M-estimator of $\sigma_v^2$ shows 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 those in the correctly specified cases. As expected, the rses show some bias as the asymptotic distribution of the AQS estimator is established based on the true number of factors. 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 the 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 is valid when spatial weights and (some) regressors are sequentially exogenous. The Monte Carlo results for the 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 a theoretical perspective, but it is clearly beyond the scope of this paper. We plan to carry out such a study in future research.

## 6. Empirical Illustration: Peer Effects in Innovation

In this section, we apply the estimation and inference methods for the DSPD-IFE model proposed in this paper to investigate peer effects in firms' innovation decisions. Theoretical frameworks recognize that substantial inter-firm interactions among peers play a significant role in various corporate actions. In the decision-making process, firms typically consider the choices made by their peers. For instance, a firm's investment decisions are substantially influenced by the investment decisions of its peer firms (Dougal et al., 2015; Bustamante and Frésard,2021; Grieser et al., 2022a). Indeed, many financial and corporate policy decisions made by firms can be viewed as equilibrium outcomes that incorporate the influence of peer effects among firms. Leary and Roberts (2014) demonstrated that the average capital structure of peer firms significantly influences a firm's own capital structure decisions. Utilizing a static spatial panel data model, Grieser et al. (2022a) similarly identified a moderate yet substantial degree of strategic interactions among firms in their capital structure decisions. Building on this econometric framework, Grieser et al. (2022b) further highlighted the critical role of peer effects across various outcomes, including corporate investment behavior, financial policies, and firm performance.

Building upon the work of Grieser et al. (2022a, 2022b), we employ a more comprehensive model, as specified in equation (2.4), to analyze the peer effects of R&D spending among 1,179 publicly listed firms on Chinese stock exchanges from 2015 to 2018. The time-lagged dependent variable $y_{t-1}$ measures the stability of firms' innovation decisions. The spatial lag term $W_1 y_t$ reflects how the investment decisions of peer firms affect the own R&D investment level of a firm. The interaction may also be dynamic in the sense that the own R&D investment decision of a firm can depend on the investment level of its neighbors in the past, reflected by the space-time lag term $W_2 y_{t-1}$. The unobserved shocks that affect the investment level are likely to be correlated across peers, reflected by the spatial error term $W_3 u_t$. Time-specific effects $f_t$ capture macroeconomic conditions general to all firms in each year. We select a comparable set of time-varying regressors $x_t$ that could influence R&D investment, following Grieser et al. (2022a, 2022b). This set includes variables such as firm size, leverage, return on assets (ROA), asset tangibility (PPE), and growth. Detailed descriptions of these variables and their construction are provided in the Appendix E. All data are obtained from the CSMAR database. We considered two types of spatial weight matrices: $W_{\text{ind}}$ that treats firms as neighbors if they are in the same industry, and $W_{\text{geo}}$ that treats firms as neighbors if they are in the same city. We specify $W_1 = W_2 = W_{\text{ind}}$, and $W_3 = W_{\text{geo}}$. Both weight matrices are row-normalized with zero on the diagonals. Table 8 summarizes the main empirical findings.

The point estimate of the spatial lag parameter $\lambda_1$ is 0.391 and significant at 1% level, suggesting a positive and strong spatial interaction in innovation investment among firms within the same industry. Conversely, dynamic interaction does not seem to influence firms' innovation decision as $\lambda_2$ is estimated to be 0.158 with standard error 0.021. Additionally, the spatial error parameter $\lambda_3$ is estimated to be 0.088 with standard error 0.0016,which indicates moderate but statistically significant correlations or spillovers in unobserved shocks among firms located in the same area.The small and insignificant coefficient estimate for time lag investment suggests that prior innovation decisions have a minimal impact on firms' current investment actions.

**Table 5.** Peer Effects in Innovation Decisions

| Parameters | $\rho$ | $\lambda_1$ | $\lambda_2$ | $\lambda_3$ | $\beta_{Size}$ | $\beta_{Leverage}$ | $\beta_{ROA}$ | $\beta_{PPE}$ | $\beta_{Growth}$ |
|---|---|---|---|---|---|---|---|---|---|
| M-estimates | 0.053 | 0.391 | 0.158 | 0.088 | -0.078 | -0.074 | -0.036 | 0.035 | 0.012 |
| Standard error | 0.0028 | 0.0053 | 0.021 | 0.0016 | 0.0022 | 0.0005 | 0.0002 | 0.0005 | 0.00003 |

**Note**: Model (2.4) is estimated with $W_{1t} = W_{2t} = W_{\text{ind}}$, and $W_{3t} = W_{\text{geo}}$

## 7. Conclusion

This paper proposes a set of new estimation and inference methods for spatial dynamic panel data models with interactive fixed effects 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 the 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 estimation, and leading to a simple method for standard error estimation.

The proposed set of estimation and inference methods constitutes 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 for future research. These (proposed methods and planned research) are of 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 $\lambda_{10}$.*

**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 $\sigma^2$ 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 $\sigma_v^2$, 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 $\Phi_n$, 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 $B$, $\gamma_{\min}(A)\text{tr}(B) \le \text{tr}(AB) \le \gamma_{\max}(A)\text{tr}(B)$; ($iv$) for symmetric matrices $A$ and $B$, $\gamma_{\max}(A+B) \le \gamma_{\max}(A) + \gamma_{\max}(B)$; and ($v$) for p.s.d. matrices $A$ and $B$, $\gamma_{\max}(AB) \le \gamma_{\max}(A)\gamma_{\max}(B)$. See, e.g, Bernstein (2009).

**Proof of Theorem 3.1:** Under Assumption G, by Theorem 5.9 of van der Vaart (1998) the consistency of $\hat{\delta}$ follows if $\sup_{\delta\in\mathbf{\Delta}} \frac{1}{nT}\|S^{*c}_{nT}(\delta) - \bar{S}^{*c}_{nT}(\delta)\| \xrightarrow{p} 0$ as $n \to \infty$, where $S^{*c}_{nT}(\delta)$ is the concentrated AQS function for $\delta$ 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:

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

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

**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

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

where $Q_{k\ell}$ takes one of the forms: $\frac{1}{n(T-r)}\mathbf{y}_0'\mathbf{R}_1\mathbf{y}_0$, $\frac{1}{n(T-r)}\mathbf{v}'\mathbf{R}_2\mathbf{v}$, $\frac{1}{n(T-r)}\mathbf{y}_0'\mathbf{R}_3\mathbf{v}$, $\frac{1}{n(T-r)}\mathbf{y}_0'\mathbf{R}_4$, and $\frac{1}{n(T-r)}\mathbf{v}'\mathbf{R}_5$. $\mathbf{R}_1, \mathbf{R}_2$, and $\mathbf{R}_3$ are $nT\times nT$ matrices while $\mathbf{R}_4$ and $\mathbf{R}_5$ are $nT\times 1$ vectors. These parametric quantities $\mathbf{R}_s, s = 1,\ldots,5$ depend on $\delta$ 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 $\ell$,

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

Now, all the $Q_{k,\ell}(\delta)$ terms are linear or quadratic in $\rho, \lambda_1$ and $\lambda_2$, 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 $\lambda_3$ and $\phi$, 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 $\ell$, $\sup_{\delta\in\boldsymbol{\Delta}} |\frac{\partial}{\partial\lambda_3}Q_{k,\ell}(\delta)| = O_p(1)$. 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 $\ell$, $\sup_{\delta\in\boldsymbol{\Delta}} |\frac{\partial}{\partial\phi_s}Q_{k,\ell}(\delta)| = O_p(1), s = 1, 2, \ldots, k_\phi$. 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,

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

It left to show that $\mathrm{E}Q_4(\delta) = \frac{1}{n(T-r)}\mathrm{E}[(\mathbf{B}_1^*\mathbf{Y}^\circ - \mathbf{B}_2^*\mathbf{Y}_{-1}^\circ)'\mathbf{P}(\mathbf{B}_1^*\mathbf{Y}^\circ - \mathbf{B}_2^*\mathbf{Y}_{-1}^\circ)] \to 0$, uniformly in $\delta \in \boldsymbol{\Delta}$. By Assumption D, $\gamma_{\min}(\frac{\mathbf{X}'\mathbf{A}\mathbf{X}}{nT}) > \underline{c}_x$. We have by the assumptions in Theorem 3.1 and Assumption D, $\mathrm{E}Q_4 = \frac{1}{n(T-r)}\mathrm{tr}[\mathbf{A}\mathbf{X}(\mathbf{X}'\mathbf{A}\mathbf{X})^{-1}\mathbf{X}'\mathbf{A}\mathrm{Var}(\mathbf{B}_1\mathbf{Y} - \mathbf{B}_2\mathbf{Y}_{-1})] \leq \frac{1}{n(T-r)}\gamma^2_{\max}(\mathbf{A})\gamma^{-1}_{\min}(\frac{\mathbf{X}'\mathbf{A}\mathbf{X}}{nT})\bar{c}_y\ \frac{1}{nT}\mathrm{tr}(\mathbf{X}'\mathbf{X}) = O(n^{-1})$. 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 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 $\Pi_1^\dagger = \frac{1}{\sigma^2_{v0}}(M_{F_0} \otimes B_{30})\mathbf{X}$, $\Pi_2^\dagger = \frac{1}{\sigma^2_{v0}}(M_{F_0} \otimes B_{30})\boldsymbol{\eta}^*_{-1}$, $\Pi_3^\dagger = \frac{1}{\sigma^2_{v0}}(M_{F_0} \otimes B_{30})\mathbf{W}_1\boldsymbol{\eta}^*$,
$\Pi_4^\dagger = \frac{1}{\sigma^2_{v0}}(M_{F_0}\otimes B_{30})\mathbf{W}_2\boldsymbol{\eta}^*_{-1}$, $\Pi_{4+s}^\dagger = \frac{1}{\sigma^2_{v0}}(M_{F_0}\dot{F}_{s0}(F_0'F_0)^{-1}F_0\otimes B_{30})\mathtt{vec}(\Gamma_0 F_0')$, $s = 1, \ldots, k_\phi$;
$\Phi_1^\dagger = \frac{1}{\sigma^2_{v0}}(M_{F_0}\otimes B_{30})\mathbf{D}_{-1}\mathbf{B}_{30}^{-1}$, $\Phi_2^\dagger = \frac{1}{\sigma^2_{v0}}(M_{F_0}\otimes B_{30})\mathbf{W}_1\mathbf{D}\mathbf{B}_{30}^{-1}$, $\Phi_3^\dagger = \frac{1}{\sigma^2_{v0}}(M_{F_0}\otimes B_{30})\mathbf{W}_2\mathbf{D}_{-1}\mathbf{B}_{30}^{-1}$,
$\Phi_4^\dagger = \frac{1}{\sigma^2_{v0}}(M_{F_0}\otimes W_3 B_{30}^{-1})$, $\Phi_5^\dagger = \frac{1}{2\sigma^4_{v0}}\mathbf{M}_{F_0}$, $\Phi_{5+s}^\dagger = \frac{1}{\sigma^2_{v0}}[M_{F_0}\dot{F}_{s0}(F_0'F_0)^{-1}F_0'\otimes I_n]$, $s = 1, \ldots, k_\phi$;
$\Psi_1^\dagger = \frac{1}{\sigma^2_{v0}}(M_{F_0} \otimes B_{30})\mathbf{Q}_{-1}$, $\Psi_2^\dagger = \frac{1}{\sigma^2_{v0}}(M_{F_0} \otimes B_{30})\mathbf{W}_1\mathbf{Q}$, 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 $\ell$, 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 $\mu$, 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 device 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 $p$th 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:

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

where $p, q = 1, \ldots, k+3$, $s, \ell = 1, \ldots, k_\phi$, $A_s = M_F\dot{F}_s(F'F)^{-1}F'$, $\dot{A}_{s,\ell} = \frac{\partial}{\partial \phi_\ell}A_s$, $\dot{\mu}_{\beta_p,\beta_q} = \frac{\partial}{\partial \beta_q}\mu_{\beta_p}$, and $\dot{\mu}_{\beta_p,\phi_s} = \frac{\partial}{\partial \phi_s}\mu_{\beta_p}$, where $\mu_{\beta_p} = 0$ for $p \le k$, and defined under (3.14) 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 $(\lambda_{30}, \phi_0')$. 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 $(\rho, \lambda)$ in a neighborhood of $(\rho_0, \lambda_0)$ by Lemma A.2 and Assumptions C and E. So, the derivatives of $\mu_{\rho,\rho}(\bar{\boldsymbol{\psi}})$ with respect to $\rho$ , $\lambda$ and $\phi$ 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 $\hat{\psi}_{\mathtt{M}}$.

**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 $H_{nT}(\hat{\psi}_{\mathtt{M}}) - H_{nT}(\psi_0) \xrightarrow{p} 0$ is implied by result **(b)** in the proof of Theorem 3.2. Next, the result $\hat{\Sigma}_{nT} - \Sigma_{nT}(\psi_0) \xrightarrow{p} 0$ 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)** $\frac{1}{nT}[\Upsilon(\hat{\psi}) - \Upsilon(\psi_0)] = o_p(1)$.

By the expression of $\Upsilon$ presented in Section 4, the proof of (**b**) is straightforward by the MVT and consistency of $\hat{\psi}_{\mathtt{M}}$. 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 $\iota = 1, \ldots, 5 + k_\phi$,

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

**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 $\Pi_{it}$ are scalars and letting $b_i'$ is the $i$th row of $B_{30}$, we have by (3.17),

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

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

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

where $\mathbb{D}_{\Pi,r} = (\Pi_{r,1}, \Pi_{r,2}, \ldots, \Pi_{r,T})$ is a $n \times T$ matrix whose $t$th column corresponds to $\Pi_{r,t}$, the subvectors of $\Pi_r$ corresponding to $t = 1, \ldots, T$. According to the expressions of $\Pi_r$ 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 $\psi_0$. 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 $\Psi_r$ 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 $t$th column is $\varphi_{r,t} = \sum_{s=1}^T \Phi_{r,ts} z^*_s$. Similarly, by the expressions of $\Phi_r$ 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 $r = 1, \ldots, 4$, $\nu = 1, 2, 3$, and $\iota = 1, \ldots, 5 + k_\phi$.

**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 $v_{i\cdot} = (v_{i1}, v_{i2}, \ldots, v_{iT})'$ and $\Pi_{r,i\cdot}$ be similarly defined, we can write

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

By Assumptions A and B, $U_{n,i}$ are independent across $i$. Elements of $\Pi_r, r = 1, \ldots, 4$ 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

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

Let $\{\mathcal{G}_{n,i}\}$ be the increasing sequence of $\sigma$ -fields generated by $(v_{j1}, \ldots, v_{jT}, j = 1, \ldots, i)$, $i = 1, \ldots, n, n \geq 1$. Let $\mathcal{F}_{n,0}$ be the $\sigma$ -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 $\sigma$-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, $\{U_{1n,i}, \mathcal{F}_{n,i}\}$ forms a M.D. array. Using Assumptions A, B, E, and F, it is easy to see that $\mathrm{E}\left|\mathrm{U_{1n,i}^{1+\epsilon}}\right| \leq \mathrm{K_v} < \infty$, for some $\epsilon > 0$. Thus, $\{U_{1n,i}\}$ is uniformly integrable. With constant coefficients $\frac{1}{nT}$,, the other two conditions of the weak law of large numbers (WLLN) for the 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 $U_9=\frac{2}{nT}\sum_{i=1}^{n}\{(v_{i\cdot}'v_{r,i\cdot}^*)(v_{i\cdot}'c_{r,i\cdot})-\mathrm{E}[(v_{i\cdot}'v_{r,i\cdot}^*)(v_{i\cdot}'c_{r,i\cdot})]\}$,

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

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, $\kappa_{it}$ is independent of $v_{it}$. Recall that $a_{its}'$ is the $i$th 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 $\varphi_{it}^v$, 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 $i$th rows of $\Phi_{ts}^u$, and $\Phi_{ts}^\ell$.

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,

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

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

where $a_{its,j}^u$ is the $j$th 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 $\sum_{t=1}^T (\sum_{s=1}^T a_{its}^{u\prime} a_{its}^u)^2 \leq C < \infty$, $\sum_{t=1}^T \sum_{s=1}^T (a_{its}^{u\prime} a_{its}^u)^2 \leq C < \infty$, and $\sum_{j=1}^n a_{its,j}^{u^4} \leq \max_j |a_{its,j}^2| \sum_{j=1}^n a_{its,j}^{u^2} = \max_j |a_{its,j}^2| (a_{its}^{u\prime} a_{its}^u) \leq C < \infty$. Thus, we have $\mathrm{E}(H_{1n,i}^2) \leq C < \infty$. 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 $\mathrm{E}(H_{2n,i}^2) \leq C < \infty$. 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,

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

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

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

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 (i):** Under Assumptions A-G, the proof of consistency as in 3.1 continues to hold. We establish the desired results based on the fact that $\hat{\psi} - \psi_0 = o_p(1)$. Then we conduct the analysis of $\hat{\psi}$ local to $\psi_0$. The following analysis determines the convergence rate of $\hat{\psi}$ and derives its asymptotic distribution.

We rely on Lemmas C.1-C.3. 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 $M_{\hat{F}(\psi_0)}$. Lemma D.6 provides a detailed expansion of $M_{\hat{F}(\psi)}$ local to $\psi_0$. Lemma C.1 provides a simplified version of the expansion of $M_{\hat{F}(\psi)}$ at true value. Then we can readily study the asymptotic orders of the score in Lemma C.2. And Lemma C.3 studies the Hessian matrix at $\psi_0$.

As the factors enter the AQS functions in the form of $M_{\hat{F}(\psi_0)}$. 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 $\psi_0$. Then we plug in the expression and find the asymptotic order of each term. Finally, analysis of the leading terms helps establish the desired results.

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

$$
\begin{aligned}
\tilde{S}^*_{nT,j}(\psi) &= \tfrac{1}{\sigma_v^2}\operatorname{tr}[X_j' B_3(\lambda_3)' B_3(\lambda_3)\mathbb{Z}(\theta) M_{\hat{F}(\psi)}], \quad \text{for } j = 1, \cdots, k,\\
\tilde{S}^*_{nT,k+1}(\psi) &= \tfrac{1}{\sigma_v^2}\operatorname{tr}[Y_{-1}' B_3(\lambda_3)' B_3(\lambda_3)\mathbb{Z}(\theta) M_{\hat{F}(\psi)}] - \operatorname{tr}[\mathbf{M}_{\hat{F}(\psi)}\mathbf{D}_{-1}(\rho, \lambda_1, \lambda_2)],\\
\tilde{S}^*_{nT,k+2}(\psi) &= \tfrac{1}{\sigma_v^2}\operatorname{tr}[(W_1 Y)' B_3(\lambda_3)' B_3(\lambda_3)\mathbb{Z}(\theta) M_{\hat{F}(\psi)}] - \operatorname{tr}[\mathbf{M}_{\hat{F}(\psi)}\mathbf{W}_1\mathbf{D}(\rho, \lambda_1, \lambda_2)],\\
\tilde{S}^*_{nT,k+3}(\psi) &= \tfrac{1}{\sigma_v^2}\operatorname{tr}[(W_2 Y_{-1})' B_3(\lambda_3)' B_3(\lambda_3)\mathbb{Z}(\theta) M_{\hat{F}(\psi)}] - \operatorname{tr}[\mathbf{M}_{\hat{F}(\psi)}\mathbf{W}_2\mathbf{D}_{-1}(\rho, \lambda_1, \lambda_2)],\\
\tilde{S}^*_{nT,k+4}(\psi) &= \tfrac{1}{\sigma_v^2}\operatorname{tr}[\mathbb{Z}(\theta)' B_3(\lambda_3)' W_3 \mathbb{Z}(\theta) M_{\hat{F}(\psi)}] - (T-r)\operatorname{tr}[W_3 B_3^{-1}(\lambda_3)],\\
\tilde{S}^*_{nT,k+5}(\psi) &= \tfrac{1}{2\sigma_v^4}\operatorname{tr}[\mathbb{Z}(\theta)' B_3(\lambda_3)' B_3(\lambda_3)\mathbb{Z}(\theta) M_{\hat{F}(\psi)}] - \tfrac{n(T-r)}{2\sigma_v^2}.
\end{aligned}
$$

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 $\psi_0$ as

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

with the detailed expression of $\tilde{S}_{nT}$ being given therein. Lemma C.3 studies the Hessian matrix. And it is invertible at true parameters under Assumption H. 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. ■

**Proof of Corollary 3.1 (ii):** Before Lemma C.4, 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 are 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. Similarly to the proof of Theorem 3.1 $(ii)$, 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 $\psi$. That is, its $\psi$-$\psi$ 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 challenging 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 $O_p(1)$. Partitioning the two high-dimensional matrices according to $\psi$ and $\phi$, 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 $\psi$ and $\phi$:

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

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 relative to $n$. ■

## Supplementary Material

The `Supplementary Material` contains additional lemmas for the proofs of Corollaries 3.1 and 3.2, details of variable constructions in the empirical study, 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] Alvarez, J., Arellano, M. 2022. Robust likelihood estimation of dynamic panel data models. *Journal of econometrics*, 226(1), 21-61.

[4] Arellano, M. and Hahn, J. (2007). "Understanding bias in nonlinear panel models: Some recent developments." Econometric Society Monographs, 43, 381–409.

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

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

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

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

[9] 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.

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

[11] Bustamante, M.C., Frésard, L., 2021. Does firm investment respond to peers' investment? *Management Science*. 67 (8), 4703–4724.

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

[13] 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.

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

[15] Dougal, C., Parsons, C.A., Titman, S., 2015. Urban vibrancy and corporate growth. *The Journal of Finance* 70 (1), 163–210.

[16] Grieser, W., Hadlock, C., LeSage, J., Zekhnini, M., 2022a. Network effects in corporate financial policies. *Journal of Financial Economics* 144 (1), 247–272.

[17] Grieser, W., LeSage, J., Zekhnini, M., 2022b. Industry networks and the geography of firm behavior. *Management Science* 68 (8), 6163–6183.

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

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

[20] Kelejian, H. H. and Prucha, I. R., 1998. A Generalized Spatial Two Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model With Autoregressive Disturbances. *Journal of Real Estate Finance and Economics* 17, 99–121.

[21] 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.

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

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

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

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

[26] Lee, L.-F., 2007. GMM and 2SLS Estimation of Mixed Regressive, Spatial Autoregressive Models, *Journal of Econometrics* 137, 489–514.

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

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

[29] 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.

[30] 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.

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

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

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

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

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

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

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

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

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

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

[41] 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.

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

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

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

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

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

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

[48] 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.

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

[50] 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 | |
|---|---|---|---|---|---|---|
| $\psi$ | 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 | |
|---|---|---|---|---|---|---|
| $\psi$ | 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 | |
|---|---|---|---|---|---|---|
| $\psi$ | 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 | |
|---|---|---|---|---|---|---|
| $\psi$ | 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 | |
|---|---|---|---|---|---|---|
| $\psi$ | 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 | |
|---|---|---|---|---|---|---|
| $\psi$ | 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 | |
|---|---|---|---|---|---|---|
| $\psi$ | 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.

# Supplementary Material

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

The `Supplementary Material` contains three additional appendices, C and D, where Appendix C contains Lemmas C1-C7 for the proofs of Corollaries 3.1 and 3.2, Appendix D contains additional technical Lemmas D1-D5 for the results in Appendix C, and Appendix E contains details of variable constructions in the empirical study.

## Appendix C: Asymptotic Analysis with Relatively Large $T$

Recall $\|\cdot\|_{\mathtt{sp}}$, the spectrum norm of a matrix. Let $\dot{X}_j = B_{30}X_j$, for $j = 1, \ldots, k$, $\dot{X}_{k+1} = B_{30}Y_{-1}$, $\dot{X}_{k+2} = B_{30}W_1Y$, $\dot{X}_{k+3} = B_{30}W_2Y_{-1}$, and $\dot{\Gamma} = B_{30}\Gamma_0$. Denote $\tilde{F} = \hat{F}(\psi_0)$.

**Lemma C.1.** *Suppose Assumptions A-H hold and $T/n \to 0$. Define the pseudo-inverses $(\dot{\Gamma}F_0')^\dagger = F_0(F_0'F_0)^{-1}(\dot{\Gamma}'\dot{\Gamma})^{-1}\dot{\Gamma}'$ and $(F_0\dot{\Gamma}')^\dagger = \dot{\Gamma}(\dot{\Gamma}'\dot{\Gamma})^{-1}(F_0'F_0)^{-1}F_0'$. The following expansion holds for the projection matrix $M_{\tilde{F}}$:*

$$M_{\tilde{F}} = M_{F_0} + M_{\tilde{F},v}^{(1)} + M_{\tilde{F},v}^{(2)} + M_{\tilde{F}}^{(rem)},$$

*where the last three $T \times T$ matrices are $M_{\tilde{F},v}^{(1)} = -M_{F_0}\mathbb{V}'(F_0\dot{\Gamma}')^\dagger - (\dot{\Gamma}F_0')^\dagger\mathbb{V}M_{F_0}$,*

$$M_{\tilde{F},v}^{(2)} = -\, M_{F_0}(\mathbb{V}'\mathbb{V} - n\sigma_{v0}^2 I_T)(\dot{\Gamma}F_0')^\dagger(F_0\dot{\Gamma}')^\dagger - (F_0\dot{\Gamma}')^\dagger(\dot{\Gamma}F_0')^\dagger(\mathbb{V}'\mathbb{V} - n\sigma_{v0}^2 I_T)M_{F_0}.$$

*and $\|M_{\tilde{F}}^{(rem)}\|_{\mathtt{sp}} = o_p((nT)^{-1/2})$, which is the remainder term.*

**Lemma C.2.** *Suppose Assumptions A-H hold and $T/n \to 0$. The concentrated AQS vector (3.27) has the expansion $\tilde{S}_{nT}^*(\psi_0) = \tilde{S}_{nT} + o_p(\sqrt{nT})$. The leading term $\tilde{S}_{nT}$ has elements:*

$$\begin{aligned}
\tilde{S}_{j,nT} =&\ \tfrac{1}{\sigma_{v0}^2}\operatorname{tr}(\dot{X}_j M_{F_0}\mathbb{V}' M_{\dot{\Gamma}}), \quad \text{for } j = 1, \ldots, k,\\
\tilde{S}_{k+1,nT} =&\ \tfrac{1}{\sigma_{v0}^2}\operatorname{tr}(\dot{X}_{k+1} M_{F_0}\mathbb{V}' M_{\dot{\Gamma}}) - \operatorname{tr}[(M_{F_0} \otimes M_{\dot{\Gamma}})\mathbf{D}_{-1}],\\
\tilde{S}_{k+2,nT} =&\ \tfrac{1}{\sigma_{v0}^2}\operatorname{tr}(\dot{X}_{k+2} M_{F_0}\mathbb{V}' M_{\dot{\Gamma}}) - \operatorname{tr}[(M_{F_0} \otimes M_{\dot{\Gamma}}W_1)\mathbf{D}],\\
\tilde{S}_{k+3,nT} =&\ \tfrac{1}{\sigma_{v0}^2}\operatorname{tr}(\dot{X}_{k+3} M_{F_0}\mathbb{V}' M_{\dot{\Gamma}}) - \operatorname{tr}[(M_{F_0} \otimes M_{\dot{\Gamma}}W_2)\mathbf{D}_{-1}],\\
\tilde{S}_{k+4,nT} =&\ \tfrac{1}{\sigma_{v0}^2}\operatorname{tr}(W_3 B_{30}^{-1}\mathbb{V} M_{F_0}\mathbb{V}') - (T - r)\operatorname{tr}(W_3 B_{30}^{-1}),\\
\tilde{S}_{k+5,nT} =&\ \tfrac{1}{2\sigma_{v0}^4}\operatorname{tr}(\mathbb{V} M_{F_0}\mathbb{V}') - \tfrac{n(T-r)}{2\sigma_{v0}^2}.
\end{aligned}$$

*In addition, we have $E(\tilde{S}_{j,nT}) = 0$ for all $j = 1, \ldots, k + 5$.*

The next lemma studies the properties of the derivative of the concentrated AQS function.

**Lemma C.3.** *Suppose Assumptions A-H hold and $T/n \to 0$. We have,*

$$-\frac{1}{nT}\frac{\partial \tilde{S}^*_{nT}(\psi)}{\partial \psi'} \mid_{\psi_0} = \tilde{H}_{nT} + o_p(1).$$

*The leading term $\tilde{H}_{nT}$ has elements $\tilde{H}_{nT,(j,q)}$, $j, q = 1, 2, \ldots, \dim(\psi)$:*

*(i)* $\tilde{H}_{nT,(j,q)} = \frac{1}{nT\sigma^2_{v0}} \operatorname{tr}(\dot{X}'_j M_{\dot{\Gamma}} \dot{X}_q M_{F_0}) - \mathbf{1}\{j = q = k+2\} \cdot \frac{T-r}{nT} \operatorname{tr}(W_1 B_{10}^{-1} W_1 B_{10}^{-1})$, $j \leq k+3$, $q \leq k+3$; $\tilde{H}_{nT,(j,q)} = o_p(1)$, $q = k+4, k+5$;

*(ii)* $\tilde{H}_{nT,(j,q)} = \frac{1}{n} \operatorname{tr}(B_{30}^{-1\prime} W_3' W_3 B_{30}^{-1} - W_3 B_{30}^{-1} W_3 B_{30}^{-1})$, $j = k+4$, $q = k+4$;

*(iii)* $\tilde{H}_{nT,(j,q)} = \frac{1}{2\sigma^4_{v0}} + o_p(1)$, $j = k+5$, $q = k+5$; $\tilde{H}_{nT,(j,q)} = o_p(1)$, $j = k+5$, $q \neq k+5$.

Next lemma studies the variance matrix of $\tilde{S}_{nT}$, for which we write

$$\tilde{S}_{nT} = \begin{cases} \tilde{\Pi}_1' \mathbf{v} \\ \mathbf{v}'\tilde{\Psi}_1 \mathbf{y}_0 + \mathbf{v}'\tilde{\Phi}_1 \mathbf{v} + \tilde{\Pi}_2' \mathbf{v} - \sigma^2_{v0} \operatorname{tr}(\tilde{\Phi}_1) \\ \mathbf{v}'\tilde{\Psi}_2 \mathbf{y}_0 + \mathbf{v}'\tilde{\Phi}_2 \mathbf{v} + \tilde{\Pi}_3' \mathbf{v} - \sigma^2_{v0} \operatorname{tr}(\tilde{\Phi}_2) \\ \mathbf{v}'\tilde{\Psi}_3 \mathbf{y}_0 + \mathbf{v}'\tilde{\Phi}_3 \mathbf{v} + \tilde{\Pi}_4' \mathbf{v} - \sigma^2_{v0} \operatorname{tr}(\tilde{\Phi}_3) \\ \mathbf{v}'\Phi_4 \mathbf{v} - \sigma^2_{v0} \operatorname{tr}(\Phi_4) \\ \mathbf{v}'\Phi_5 \mathbf{v} - \sigma^2_{v0} \operatorname{tr}(\Phi_5) \end{cases}$$

where $\tilde{\Psi}_1 = \frac{1}{\sigma^2_{v0}}(M_{F_0} \otimes M_{\dot{\Gamma}} B_{30})\mathbf{Q}_{-1}$, $\tilde{\Psi}_2 = \frac{1}{\sigma^2_{v0}}(M_{F_0} \otimes M_{\dot{\Gamma}} B_{30} W_1)\mathbf{Q}$,
$\tilde{\Psi}_3 = \frac{1}{\sigma^2_{v0}}(M_{F_0} \otimes M_{\dot{\Gamma}} B_{30} W_2)\mathbf{Q}_{-1}$, $\tilde{\Pi}_1 = \frac{1}{\sigma^2_{v0}}(M_{F_0} \otimes M_{\dot{\Gamma}})\mathbf{X}$,
$\tilde{\Pi}_2 = \frac{1}{\sigma^2_{v0}}(M_{F_0} \otimes M_{\dot{\Gamma}} B_{30})[\boldsymbol{\eta}_{-1} + \mathbf{D}_{-1}\mathtt{vec}(\Gamma_0 F_0')]$,
$\tilde{\Pi}_3 = \frac{1}{\sigma^2_{v0}}(M_{F_0} \otimes M_{\dot{\Gamma}} B_{30} W_1)[\boldsymbol{\eta} + \mathbf{D}\mathtt{vec}(\Gamma_0 F_0')]$,
$\tilde{\Pi}_4 = \frac{1}{\sigma^2_{v0}}(M_{F_0} \otimes M_{\dot{\Gamma}} B_{30} W_2)[\boldsymbol{\eta}_{-1} + \mathbf{D}_{-1}\mathtt{vec}(\Gamma_0 F_0')]$,
$\tilde{\Phi}_1 = \frac{1}{\sigma^2_{v0}}(M_{F_0} \otimes M_{\dot{\Gamma}} B_{30})\mathbf{D}_{-1}\mathbf{B}_{30}^{-1}$, $\tilde{\Phi}_2 = \frac{1}{\sigma^2_{v0}}(M_{F_0} \otimes M_{\dot{\Gamma}} B_{30} W_1)\mathbf{D}\mathbf{B}_{30}^{-1}$,
$\tilde{\Phi}_3 = \frac{1}{\sigma^2_{v0}}(M_{F_0} \otimes M_{\dot{\Gamma}} B_{30} W_2)\mathbf{D}_{-1}\mathbf{B}_{30}^{-1}$, $\Phi_4 = \frac{1}{\sigma^2_{v0}}(M_{F_0} \otimes W_3 B_{30}^{-1})$, $\Phi_5 = \frac{1}{2\sigma^4_{v0}}(M_{F_0} \otimes I_n)$.

**Lemma C.4.** *Suppose Assumptions A-H hold and $T/n \to 0$. Denote the variance of $\tilde{S}_{nT}/\sqrt{nT}$ conditional on $\mathcal{D}$ as $\tilde{\Sigma}_{nT}$. We have $\tilde{\Sigma}_{nT} = \tilde{H}_{nT} + \tilde{\Xi}_{nT} + o_p(1)$. All three matrices are symmetric. The upper triangular part of $\tilde{\Xi}_{nT}$ has $(j,q)$-entries:*

$\tilde{\Xi}_{nT,(1:k,1:k)} = 0$; $\tilde{\Xi}_{nT,(1:k,q)} = \frac{\mu_3}{nT}\tilde{\Pi}_{1,j}' \mathtt{diag}(\tilde{\Phi}_a)$, $q = k+a$ *and* $a = 1, \ldots, 5$;

$$\tilde{\Xi}_{nT,(k+1,k+1)} = \frac{2\mu_3}{nT}(\tilde{\Pi}_2 + \tilde{\Psi}_1 \mathbf{y}_0)'\mathtt{diag}(\tilde{\Phi}_1) + \frac{\sigma^4_{v0}}{nT}\operatorname{tr}(\tilde{\Phi}_1\tilde{\Phi}_1) + \frac{(\mu_4 - 3\sigma^4_{v0})}{nT}\|\mathtt{diag}(\tilde{\Phi}_1)\|^2;$$

$$\tilde{\Xi}_{nT,(k+1,k+2)} = \frac{\mu_3}{nT}(\tilde{\Pi}_2 + \tilde{\Psi}_1 \mathbf{y}_0)'\mathtt{diag}(\tilde{\Phi}_2) + \frac{\mu_3}{nT}(\tilde{\Pi}_3 + \tilde{\Psi}_2 \mathbf{y}_0)'\mathtt{diag}(\tilde{\Phi}_1) + \frac{\sigma^4_{v0}}{nT}\operatorname{tr}(\tilde{\Phi}_1\tilde{\Phi}_2) + \frac{(\mu_4 - 3\sigma^4_{v0})}{nT}\mathtt{diag}(\tilde{\Phi}_1)'\mathtt{diag}(\tilde{\Phi}_2);$$

$$\tilde{\Xi}_{nT,(k+1,k+3)} = \frac{\mu_3}{nT}(\tilde{\Pi}_3 + \tilde{\Psi}_2 \mathbf{y}_0)'\mathtt{diag}(\tilde{\Phi}_3) + \frac{\mu_3}{nT}(\tilde{\Pi}_4 + \tilde{\Psi}_3 \mathbf{y}_0)'\mathtt{diag}(\tilde{\Phi}_2) + \frac{\sigma^4_{v0}}{nT}\operatorname{tr}(\tilde{\Phi}_2\tilde{\Phi}_3) + \frac{(\mu_4 - 3\sigma^4_{v0})}{nT}\mathtt{diag}(\tilde{\Phi}_2)'\mathtt{diag}(\tilde{\Phi}_3);$$

$$\tilde{\Xi}_{nT,(k+1,k+4)} = \frac{\mu_3}{nT}(\tilde{\Pi}_2 + \tilde{\Psi}_1 \mathbf{y}_0)'\mathtt{diag}(\Phi_4) + \frac{\sigma^4_{v0}}{nT}\operatorname{tr}(\tilde{\Phi}_1\Phi_4 + \tilde{\Phi}_1\Phi_4') + \frac{(\mu_4 - 3\sigma^4_{v0})}{nT}\mathtt{diag}(\tilde{\Phi}_1)'\mathtt{diag}(\Phi_4);$$

$$\tilde{\Xi}_{nT,(k+1,k+5)} = \tfrac{\mu_3}{nT}(\tilde{\Pi}_2+\tilde{\Psi}_1\mathbf{y}_0)'\mathtt{diag}(\Phi_5)+\tfrac{\sigma_{v0}^4}{nT}\operatorname{tr}(\tilde{\Phi}_1\Phi_5+\tilde{\Phi}_1\Phi_5')+\tfrac{(\mu_4-3\sigma_{v0}^4)}{nT}\mathtt{diag}(\tilde{\Phi}_1)'\mathtt{diag}(\Phi_5);$$

$$\tilde{\Xi}_{nT,(k+2,k+2)} = \tfrac{nT}{2\mu_3}(\tilde{\Pi}_3 + \tilde{\Psi}_2\mathbf{y}_0)'\mathtt{diag}(\tilde{\Phi}_2) + \tfrac{nT}{\sigma_{v0}^4}\operatorname{tr}(\tilde{\Phi}_2\tilde{\Phi}_2) + \tfrac{nT}{(\mu_4-3\sigma_{v0}^4)}\|\mathtt{diag}(\tilde{\Phi}_2)\|^2;$$

$$\tilde{\Xi}_{nT,(k+2,k+3)} = \tfrac{\mu_3}{nT}(\tilde{\Pi}_3 + \tilde{\Psi}_2\mathbf{y}_0)'\mathtt{diag}(\tilde{\Phi}_3) + \tfrac{\mu_3}{nT}(\tilde{\Pi}_4 + \tilde{\Psi}_3\mathbf{y}_0)'\mathtt{diag}(\tilde{\Phi}_2) + \tfrac{\sigma_{v0}^4}{nT}\operatorname{tr}(\tilde{\Phi}_2\tilde{\Phi}_3) + \tfrac{(\mu_4-3\sigma_{v0}^4)}{nT}\mathtt{diag}(\tilde{\Phi}_2)'\mathtt{diag}(\tilde{\Phi}_3);$$

$$\tilde{\Xi}_{nT,(k+2,k+4)} = \tfrac{\mu_3}{nT}(\tilde{\Pi}_4+\tilde{\Psi}_3\mathbf{y}_0)'\mathtt{diag}(\Phi_4)+\tfrac{\sigma_{v0}^4}{nT}\operatorname{tr}(\tilde{\Phi}_3\Phi_4+\tilde{\Phi}_3\Phi_4')+\tfrac{(\mu_4-3\sigma_{v0}^4)}{nT}\mathtt{diag}(\tilde{\Phi}_3)'\mathtt{diag}(\Phi_4);$$

$$\tilde{\Xi}_{nT,(k+2,k+5)} = \tfrac{\mu_3}{nT}(\tilde{\Pi}_4+\tilde{\Psi}_3\mathbf{y}_0)'\mathtt{diag}(\Phi_5)+\tfrac{\sigma_{v0}^4}{nT}\operatorname{tr}(\tilde{\Phi}_3\Phi_5+\tilde{\Phi}_3\Phi_5')+\tfrac{(\mu_4-3\sigma_{v0}^4)}{nT}\mathtt{diag}(\tilde{\Phi}_3)'\mathtt{diag}(\Phi_5);$$

$$\tilde{\Xi}_{nT,(k+3,k+3)} = \tfrac{2\mu_3}{nT}(\tilde{\Pi}_4 + \tilde{\Psi}_3\mathbf{y}_0)'\mathtt{diag}(\tilde{\Phi}_3) + \tfrac{\sigma_{v0}^4}{nT}\operatorname{tr}(\tilde{\Phi}_3\tilde{\Phi}_3) + \tfrac{(\mu_4-3\sigma_{v0}^4)}{nT}\|\mathtt{diag}(\tilde{\Phi}_3)\|^2;$$

$$\tilde{\Xi}_{nT,(k+3,k+4)} = \tfrac{\mu_3}{nT}(\tilde{\Pi}_4+\tilde{\Psi}_3\mathbf{y}_0)'\mathtt{diag}(\Phi_4)+\tfrac{\sigma_{v0}^4}{nT}\operatorname{tr}(\tilde{\Phi}_3\Phi_4+\tilde{\Phi}_3\Phi_4')+\tfrac{(\mu_4-3\sigma_{v0}^4)}{nT}\mathtt{diag}(\tilde{\Phi}_3)'\mathtt{diag}(\Phi_4);$$

$$\tilde{\Xi}_{nT,(k+3,k+5)} = \tfrac{\mu_3}{nT}(\tilde{\Pi}_4+\tilde{\Psi}_3\mathbf{y}_0)'\mathtt{diag}(\Phi_5)+\tfrac{\sigma_{v0}^4}{nT}\operatorname{tr}(\tilde{\Phi}_3\Phi_5+\tilde{\Phi}_3\Phi_5')+\tfrac{(\mu_4-3\sigma_{v0}^4)}{nT}\mathtt{diag}(\tilde{\Phi}_3)'\mathtt{diag}(\Phi_5);$$

$$\tilde{\Xi}_{nT,(k+4,k+4)} = \tfrac{1}{nT}\operatorname{tr}(\Phi_4\Phi_4') + \tfrac{(\mu_4-3\sigma_{v0}^4)}{nT}\|\mathtt{diag}(\Phi_4)\|^2;$$

$$\tilde{\Xi}_{nT,(k+4,k+5)} = \tfrac{\sigma_{v0}^4}{nT}\operatorname{tr}(\Phi_4\Phi_5 + \Phi_4\Phi_5') + \tfrac{(\mu_4-3\sigma_{v0}^4)}{nT}\mathtt{diag}(\Phi_4)'\mathtt{diag}(\Phi_5);$$

$$\tilde{\Xi}_{nT,(k+5,k+5)} = \tfrac{(\mu_4-3\sigma_{v0}^4)}{nT}\|\mathtt{diag}(\Phi_5)\|^2.$$

In Section 3.4, we propose an estimate for $\Sigma_{nT}(\psi_0)$, the VC matrix of the AQS function $S^*_{nT}(\psi_0)$. In the following, we use $\Sigma_{nT} = \Sigma_{nT}(\psi_0)$ for notation simplicity. The next Lemma finds the asymptotic leading term of $\Sigma_{nT}$.

**Lemma C.5.** *Suppose Assumptions A-H hold. Then, as $T/n \to 0$, we have,*

*(i)* $\Sigma_{nT,(k+a,k+b))} = \frac{1}{nT\sigma_{v0}^2}\mathtt{vec}(\dot{F}_a')'(M_{F_0}\otimes\dot{\Gamma}'\dot{\Gamma})\mathtt{vec}(\dot{F}_b') + O_p(T^{-2})$, *for* $a, b = 1, \ldots, k_\phi$, $\|\Sigma_{nT,\phi\phi} - H_{nT,\phi\phi}\|_{\mathtt{sp}} = o_p(T^{-1})$, *and* $\|\hat{\Sigma}_{nT,\phi\phi} - \Sigma_{nT,\phi\phi}\|_{\mathtt{sp}} = o_p(T^{-1})$.

*(ii)* $\hat{\Sigma}_{nT,(k+5+a,j)} - \Sigma_{nT,(k+5+a,j)} = o_p(T^{-1})$ *uniformly,* $j \le k+5$, $a = 1, \ldots, k_\phi$;

$$\Sigma_{nT,(k+5+a,j)} = \tfrac{1}{\sigma_{v0}^2 nT}\mathtt{vec}(\dot{X}_j)'(M_{F_0}\otimes\dot{\Gamma})\mathtt{vec}(\dot{F}_a') + o_p(T^{-1}),\ j = 1, \ldots, k;$$

$$\Sigma_{nT,(k+5+a,j)} = \tfrac{1}{\sigma_{v0}^2 nT}\mathtt{vec}(\dot{X}_j)'(M_{F_0}\otimes\dot{\Gamma})\mathtt{vec}(\dot{F}_a') + \tfrac{\mu_3}{nT}\mathtt{diag}(\Phi_{j-k})'\Pi_{4+a} + o_p(T^{-1}),\ j = k+1, \ldots, k+3;$$

$$\Sigma_{nT,(k+5+a,j)} = \tfrac{\mu_3}{nT}\mathtt{diag}(\Phi_{j-k})'\Pi_{4+a} + o_p(T^{-1}),\ j = k+4, k+5;$$

*(iii)* $\Sigma_{nT,\psi\psi} = H_{nT,\psi\psi} + \Xi_{nT}$, *where $\Xi$ is the same as $\tilde{\Xi}$ with quantities $\tilde{\Pi}_a$'s, $\tilde{\Psi}_b$'s and $\tilde{\Phi}_c$'s replaced by $\Pi_a$'s, $\Psi_b$'s and $\Phi_c$'s.*

For the next lemma, we derive detailed expressions of the elements $H_{nT,(j,q)}$ of $H_{nT}$.

For $j = 1, \ldots, k$,

$$H_{nT,(j,q)} = \tfrac{1}{nT\sigma_{v0}^2}\operatorname{tr}(\dot{X}_j'\dot{X}_q M_{F_0}),\quad q \le k+3,$$

$$H_{nT,(j,q)} = \frac{1}{nT\sigma_{v0}^2} \operatorname{tr}[X_j'(B_{30}'W_3 + W_3'B_{30})\mathbb{Z}(\theta_0)M_{F_0}], \quad q = k+4,$$
$$H_{nT,(j,q)} = \frac{1}{nT\sigma_{v0}^4} \operatorname{tr}[\dot{X}_j'B_{30}\mathbb{Z}(\theta_0)M_{F_0}], \quad q = k+45,$$
$$H_{nT,(j,q)} = \frac{1}{nT\sigma_{v0}^2} \operatorname{tr}[\dot{X}_j'B_{30}\mathbb{Z}(\theta_0)\dot{P}_{F,a}], \quad q = k+6,\ldots,k+5+k_\phi,$$

For $j = k+1$,

$$H_{nT,(j,q)} = \frac{1}{nT\sigma_{v0}^2} \operatorname{tr}(\dot{X}_j'\dot{X}_q M_{F_0}), \quad q \le k,$$
$$H_{nT,(j,q)} = \frac{1}{nT\sigma_{v0}^2} \operatorname{tr}(\dot{X}_j'\dot{X}_q M_{F_0}) - \frac{1}{nT} \operatorname{tr}[(M_{F_0} \otimes I_n)\dot{\mathbf{D}}_{-1,\psi_q}], \quad q = k+1, k+2, k+3$$
$$H_{nT,(j,q)} = \frac{1}{nT\sigma_{v0}^2} \operatorname{tr}[Y_{-1}'(B_{30}'W_3 + W_3'B_{30})\mathbb{Z}(\theta_0)M_{F_0}], \quad q = k+4,$$
$$H_{nT,(j,q)} = \frac{1}{nT\sigma_{v0}^4} \operatorname{tr}[\dot{X}_j'B_{30}\mathbb{Z}(\theta_0)M_{F_0}], \quad q = k+5,$$
$$H_{nT,(j,q)} = \frac{1}{nT\sigma_{v0}^2} \operatorname{tr}[\dot{X}_j'B_{30}\mathbb{Z}(\theta_0)\dot{P}_{F,s}] - \frac{1}{nT} \operatorname{tr}[(\dot{P}_{F,s} \otimes I_n)\mathbf{D}_{-1}], \quad q = k+6,\ldots,k+5+k_\phi.$$

For $j = k+2$,

$$H_{nT,(j,q)} = \frac{1}{nT\sigma_{v0}^2} \operatorname{tr}(\dot{X}_j'\dot{X}_q M_{F_0}), \quad q \le k,$$
$$H_{nT,(j,q)} = \frac{1}{nT\sigma_{v0}^2} \operatorname{tr}(\dot{X}_j'\dot{X}_q M_{F_0}) - \frac{1}{nT} \operatorname{tr}[(M_{F_0} \otimes W_1)\dot{\mathbf{D}}_{\psi_q}], \quad q = k+1, k+2, k+3$$
$$H_{nT,(j,q)} = \frac{1}{nT\sigma_{v0}^2} \operatorname{tr}[YW_1'(B_{30}'W_3 + W_3B_{30})\mathbb{Z}(\theta_0)M_{F_0}], \quad q = k+4,$$
$$H_{nT,(j,q)} = \frac{1}{nT\sigma_{v0}^4} \operatorname{tr}[\dot{X}_j'B_{30}\mathbb{Z}(\theta_0)M_{F_0}], \quad q = k+5,$$
$$H_{nT,(j,q)} = \frac{1}{nT\sigma_{v0}^2} \operatorname{tr}[Y'W_1'B_{30}'B_{30}\mathbb{Z}(\theta_0)\dot{P}_{F,s}] - \frac{1}{nT} \operatorname{tr}[(\dot{P}_{F,s} \otimes W_1)\mathbf{D}], \quad q = k+6,\ldots,k+5+k_\phi.$$

For $j = k+3$,

$$H_{nT,(j,q)} = \frac{1}{nT\sigma_{v0}^2} \operatorname{tr}(\dot{X}_j'\dot{X}_q M_{F_0}), \quad q \le k,$$
$$H_{nT,(j,q)} = \frac{1}{nT\sigma_{v0}^2} \operatorname{tr}(\dot{X}_j'\dot{X}_q M_{F_0}) - \frac{1}{nT} \operatorname{tr}[(M_{F_0} \otimes W_2)\dot{\mathbf{D}}_{-1,\rho}], \quad q = k+1,$$
$$H_{nT,(j,q)} = \frac{1}{nT\sigma_{v0}^2} \operatorname{tr}(\dot{X}_j'\dot{X}_q M_{F_0}) - \frac{1}{nT} \operatorname{tr}[(M_{F_0} \otimes W_2)\dot{\mathbf{D}}_{-1,\lambda_1}], \quad q = k+2,$$
$$H_{nT,(j,q)} = \frac{1}{nT\sigma_{v0}^2} \operatorname{tr}(\dot{X}_j'\dot{X}_q M_{F_0}) - \frac{1}{nT} \operatorname{tr}[(M_{F_0} \otimes W_2)\dot{\mathbf{D}}_{-1,\lambda_2}], \quad q = k+3,$$
$$H_{nT,(j,q)} = \frac{1}{nT\sigma_{v0}^2} \operatorname{tr}[Y_{-1}'W_2'(B_{30}'W_3 + W_3'B_{30})\mathbb{Z}(\theta_0)M_{F_0}], \quad q = k+4,$$
$$H_{nT,(j,q)} = \frac{1}{nT\sigma_{v0}^4} \operatorname{tr}[\dot{X}_j'B_{30}\mathbb{Z}(\theta_0)M_{F_0}], \quad q = k+5,$$
$$H_{nT,(j,q)} = \frac{1}{nT\sigma_{v0}^2} \operatorname{tr}[\dot{X}_j'B_{30}\mathbb{Z}(\theta_0)\dot{P}_{F,s}] - \frac{1}{nT} \operatorname{tr}[(\dot{P}_{F,s} \otimes W_2)\mathbf{D}_{-1}], \quad q = k+6,\ldots,k+5+k_\phi.$$

For $j = k+4$.

$$H_{nT,(j,q)} = \frac{1}{nT\sigma_{v0}^2} \operatorname{tr}[X_j'(B_{30}'W_3 + W_3'B_{30})\mathbb{Z}(\theta_0)M_{F_0}], \quad q \le k,$$
$$H_{nT,(j,q)} = \frac{1}{nT\sigma_{v0}^2} \operatorname{tr}[Y_{-1}'(B_{30}'W_3 + W_3'B_{30})\mathbb{Z}(\theta_0)M_{F_0}], \quad q = k+1,$$
$$H_{nT,(j,q)} = \frac{1}{nT\sigma_{v0}^2} \operatorname{tr}[YW_1'(B_{30}'W_3 + W_3B_{30})\mathbb{Z}(\theta_0)M_{F_0}], \quad q = k+2,$$
$$H_{nT,(j,q)} = \frac{1}{nT\sigma_{v0}^2} \operatorname{tr}[Y_{-1}'W_2'(B_{30}'W_3 + W_3'B_{30})\mathbb{Z}(\theta_0)M_{F_0}], \quad q = k+3,$$
$$H_{nT,(j,q)} = \frac{1}{nT\sigma_{v0}^2} \operatorname{tr}[\mathbb{Z}(\theta_0)'W_3'W_3\mathbb{Z}(\theta_0)M_{F_0}], \quad q = k+4,$$
$$H_{nT,(j,q)} = \frac{1}{nT\sigma_{v0}^4} \operatorname{tr}[\mathbb{Z}(\theta_0)'B_{30}'W_3\mathbb{Z}(\theta_0)M_{F_0}] + \frac{(T-r)}{nT} \operatorname{tr}[W_3B_{30}^{-1}W_3B_{30}^{-1}], \quad q = k+5,$$
$$H_{nT,(j,q)} = \frac{1}{nT\sigma_{v0}^2} \operatorname{tr}[\mathbb{Z}(\theta_0)'B_{30}'W_3\mathbb{Z}(\theta_0)\dot{P}_{F,s}], \quad q = k+6,\ldots,k+5+k_\phi.$$

For $j = k+5$,

$$H_{nT,(j,q)} = \frac{1}{nT\sigma_{v0}^2} \operatorname{tr}[\dot{X}_j'B_{30}\mathbb{Z}(\theta_0)\dot{P}_{F,s}], \quad q \le k+3,$$
$$H_{nT,(j,q)} = \frac{1}{nT\sigma_{v0}^4} \operatorname{tr}[\mathbb{Z}(\theta_0)'B_{30}'B_{30}\mathbb{Z}(\theta_0)M_{F_0}] - \frac{T-r}{2T\sigma_{v0}^2}, \quad q = k+4,$$
$$H_{nT,(j,q)} = o_p(1), \quad q = k+5,\ldots,k+5+k_\phi.$$

For $j = k+5+s$,

$H_{nT,(j,q)} = \frac{1}{nT\sigma_{v0}^2}\,\mathrm{tr}[\dot{X}_j' B_{30}\mathbb{Z}(\theta_0)\dot{P}_{F,s}],\ \ q \leq k+3,$
$H_{nT,(j,q)} = \frac{1}{nT\sigma_{v0}^2}\,\mathrm{tr}[\mathbb{Z}(\theta_0)' B_{30}' W_3 \mathbb{Z}(\theta_0)\dot{P}_{F,s}],\ \ q = k+4,$
$H_{nT,(j,q)} = o_p(1),\ \ q \leq k+5,$
$H_{nT,(j,q)} = \frac{1}{2nT\sigma_{v0}^2}\,\mathrm{tr}[\mathbb{Z}(\theta_0)' B_{30}' B_{30}\mathbb{Z}(\theta_0)\ddot{P}_{F,sq}].$

**Lemma C.6.** *Suppose Assumptions A-H hold. Then, as $T/n \to 0$, we have,*

*(i)* $H_{nT,(j,q)} = \frac{1}{nT\sigma_{v0}^2}\,\mathrm{tr}(\dot{X}_j'\dot{X}_q M_{F_0}) + o_p(1),\ j,q \leq k+3$ *except* $j = q = k+2$*;*
$H_{nT,(k+2,k+2)} = \frac{1}{nT\sigma_{v0}^2}\,\mathrm{tr}(\dot{X}_j'\dot{X}_j M_{F_0}) + \frac{T-r}{nT}\,\mathrm{tr}(W_1 B_{10}^{-1} W_1 B_{10}^{-1}) + o_p(1), ;$
$H_{nT,(k+4,k+4)} = \frac{1}{n}\,\mathrm{tr}(B_{30}^{-1\prime} W_3' W_3 B_{30}^{-1} + W_3 B_{30}^{-1} W_3 B_{30}^{-1}) + o_p(1);$
$H_{nT,(k+5,k+5)} = \frac{1}{2\sigma_{v0}^2} + o_p(1);\ H_{nT,(jq)} = o_p(1),$ *for other cases with* $j,q \leq k+5$.

*(ii)* $H_{nT,(k+5+a,k+5+b)} = \frac{1}{nT\sigma_{v0}^2}\mathtt{vec}(\dot{F}_a')'(M_{F_0}\otimes\dot{\Gamma}'\dot{\Gamma})\mathtt{vec}(\dot{F}_b') + o_p(T^{-1}),\ a,b = 1,\ldots,k_\phi.$

*(iii)* $H_{nT,(j,k+5+a)} = \frac{1}{nT\sigma_{v0}^2}\,\mathrm{tr}(\dot{X}_j'\dot{\Gamma}\dot{F}_a' M_{F_0}) + o_p(T^{-1}),\ j \leq k+3;$
$H_{nT,(j,k+5+a)} = o_p(T^{-1}),\ j = k+4, k+5,\ a = 1,\ldots,k_\phi.$

*(iv)* $H_{nT,(k+5+a,q)} = \frac{1}{nT\sigma_{v0}^2}\,\mathrm{tr}(\dot{X}_q'\dot{\Gamma}\dot{F}_a' M_{F_0}) + o_p(T^{-1}),\ q \leq k+3;$
$H_{nT,(k+5+a,q)} = o_p(T^{-1}),\ q = k+4, k+5,\ a = 1\ldots,k_\phi.$

*(v)* $H_{*,nT} = \tilde{H}_{nT} + o_p(1)$ *and* $H_{nT,\phi\phi}^{-1}$ *exists with a leading term stated in the proof.*

The next Lemma calculates the matrix $[H_{nT}^{-1}\Sigma_{nT}H_{nT}^{-1\prime}]_{\psi\psi}$. Recall the partition matrix expression of $H_{nT}^{-1}$ and write $\Sigma_{nT} = [\Sigma_{nT,\psi\psi}, \Sigma_{nT,\psi\phi}; \Sigma_{nT,\phi\psi}, \Sigma_{nT,\phi\phi}]$. We can verify that

$$[H_{nT}^{-1}\Sigma_{nT}H_{nT}'^{-1}]_{\psi\psi} = H_{*,nT}^{-1}\Sigma_{nT}^{*}H_{*,nT}^{-1},$$

where $\Sigma_{nT}^{*} = \Sigma_{nT,\psi\psi} - H_{nT,\psi\phi}H_{nT,\phi\phi}^{-1}\Sigma_{nT,\phi\psi} - \Sigma_{nT,\psi\phi}H_{nT,\phi\phi}^{-1}H_{nT,\phi\psi} + H_{nT,\psi\phi}H_{nT,\phi\phi}^{-1}H_{nT,\phi\psi}.$

**Lemma C.7.** *Suppose Assumptions A-H hold and $T/n \to 0$. We have $\Sigma_{nT}^{*} = \tilde{\Sigma}_{nT} + o_p(1)$ and $H_{*,nT}^{-1}\Sigma_{nT}^{*}H_{*,nT}^{-1} - \tilde{H}_{nT}^{-1}\tilde{\Sigma}_{nT}\tilde{H}_{nT}^{-1} = o_p(1)$.*

## Proofs of Lemmas C1-C7

**Proof of Lemma C.1:** To prove this Lemma, we employ Lemma D.6 (i), with $\psi = \psi_0$. Next we directly use notations defined in Lemma D.6. Under assumption H, the $r$th eigenvalue of $(F_0'F_0)(\Gamma' B_{30}' B_{30}\Gamma)/(NT)$, $C_{min}$ is bounded below by a constant $c > 0$ with probability approaching one. With $(n,T) \to \infty$, we have $\sigma_{v0}^2/T + ||\mathbb{T}_2||_{sp} \xrightarrow{p} 0$. Specifically, we have $\epsilon_1 = O_P(T^{-1})$ and $\epsilon_2 = O_P(n^{-1/2})$, where the second equality is by

$$\frac{||F_0\dot{\Gamma}'\mathbb{V}||_{sp}}{nT} \leq \frac{1}{\sqrt{n}}\frac{||F_0||}{\sqrt{T}}\frac{||\dot{\Gamma}'\mathbb{V}||_F}{\sqrt{nT}} = o_P(n^{-1/2})$$

and the assumption that $||\mathbb{V}'\mathbb{V} - n\sigma_{v0}^2 I_T||_{sp}/(nT) = O_P((nT)^{-1/2})$. Thus $\epsilon_1 + \epsilon_2 < C_{min}/2$ with probability approaching one, and Lemma D.6 (i) holds. Next we go into the details of terms of $P(j)$'s.

(a) $P^{(1)} = -M_{F_0}(\mathbb{T}_1 + \mathbb{T}_2)S^{(1)} - S^{(1)}(\mathbb{T}_1 + \mathbb{T}_2)M_{F_0}$. Note that $M_{F_0}\mathbb{T}_1 S^{(1)} = 0$. Plug in expressions of the terms, we have

$$\begin{aligned} P^{(1)} =& - M_{F_0}\mathbb{V}'(F_0\dot{\Gamma}')^\dagger - (\dot{\Gamma}F_0')^\dagger \mathbb{V} M_{F_0} \\ & - M_{F_0}(\mathbb{V}'\mathbb{V} - n\sigma_{v0}^2 I_T)(\dot{\Gamma}F_0')^\dagger (F_0\dot{\Gamma}')^\dagger - (F_0\dot{\Gamma}')^\dagger(\dot{\Gamma}F_0')^\dagger(\mathbb{V}'\mathbb{V} - n\sigma_{v0}^2 I_T)M_{F_0}. \end{aligned}$$

(b) $|| \sum_{j=2}^{\infty} P^{(j)}||_{sp} \le 4e^2 \frac{\epsilon_2(\epsilon_1+\epsilon_2)}{C_{min}(C_{min}-2e(\epsilon_1+\epsilon_2))} = O_P(n^{-1} + n^{-1/2}T^{-1}) = o_P((nT)^{-1/2})$. ■

**Proof of Lemma C.2:** First, consider the cases of $j \le k+3$. Using the identity $B_{30}\mathbb{Z}(\theta_0) = \dot{\Gamma}F_0' + \mathbb{V}$ and the fact that $M_{F_0}F_0 = 0$, we have

$$\mathrm{tr}[\dot{X}_j' B_{30}\mathbb{Z}(\theta_0)M_{\tilde{F}}] = \mathrm{tr}(\dot{X}_j'\mathbb{V}M_{F_0}) + \mathrm{tr}[\dot{X}_j'(\mathbb{V} + \dot{\Gamma}F_0')(M_{\tilde{F}} - M_{F_0})].$$

By Lemma D.1 (i-ii), we can readily show that $S_{j,nT}^p(\psi_0) = \tilde{S}_{j,nT} + o_p(\sqrt{nT})$, for $j \le k+3$. Second, consider the cases of $j = k+4, k+5$. We have,

$$\begin{aligned} \tilde{S}_{k+4,nT}^*(\psi_0) =& \frac{1}{\sigma_{v0}^2}\,\mathrm{tr}[(\dot{\Gamma}F_0' + \mathbb{V})'W_3 B_{30}^{-1}(\dot{\Gamma}F_0' + \mathbb{V})M_{\tilde{F}}] - (T-r)\,\mathrm{tr}(W_3 B_{30}^{-1}) \\ =& \frac{1}{\sigma_{v0}^2}\,\mathrm{tr}(\mathbb{V}'W_3 B_{30}^{-1}\mathbb{V}M_{F_0}) - (T-r)\,\mathrm{tr}(W_3 B_{30}^{-1}) \\ & + \frac{1}{\sigma_{v0}^2}\,\mathrm{tr}[[(\dot{\Gamma}F_0' + \mathbb{V})'W_3 B_{30}^{-1}[(\dot{\Gamma}F_0' + \mathbb{V})(M_{\tilde{F}} - M_{F_0})]. \end{aligned}$$

Lemma D.1 (iii) has shown that the last term is $o_p(\sqrt{nT})$.
To show $E(\tilde{S}_{j,nT}) = 0$, one can follow an analysis similar to the proof of Theorem 3.1. ■

**Proof of Lemma C.3:** (i) There are three sub-cases: (i-a) $j \le k$ and $q \le k+3$; (i-b) $j \le k$ and $q \le k+4, \ldots, k+5$; and (i-c) $j = k+1, \ldots, k+3$.

First, consider the cases (i-a) and (i-b). We have $-\frac{\partial}{\partial \psi_q}\tilde{S}_{j,nT}^*(\psi)|_{\psi_0} =$

$$\begin{cases} \frac{1}{\sigma_{v0}^2}\,\mathrm{tr}(\dot{X}_j'\dot{X}_q M_{\tilde{F}}) - \frac{1}{\sigma_{v0}^2}\,\mathrm{tr}[\dot{X}_j'(\dot{\Gamma}F_0' + \mathbb{V})\frac{\partial}{\partial\psi_q}M_{\hat{F}(\psi)}|_{\psi_0}], \\ \frac{1}{\sigma_{v0}^2}\,\mathrm{tr}\{\dot{X}_j'[W_3'B_{30}^{-1} + (B_{30}^{-1})'W_3](\dot{\Gamma}F_0' + \mathbb{V})M_{\tilde{F}}\} - \frac{1}{\sigma_{v0}^2}\,\mathrm{tr}[\dot{X}_j'(\dot{\Gamma}F_0' + \mathbb{V})\frac{\partial}{\partial\lambda_3}M_{\hat{F}(\psi)}|_{\psi_0}], \\ \frac{1}{\sigma_{v0}^4}\,\mathrm{tr}[\dot{X}_j'(\dot{\Gamma}F_0' + \mathbb{V})M_{\tilde{F}}]. \end{cases}$$

By Lemma D.2, we can readily show $\frac{1}{nT\sigma_{v0}^2}\,\mathrm{tr}(\dot{X}_j'\dot{X}_q M_{\tilde{F}}) = \frac{1}{nT\sigma_{v0}^2}\,\mathrm{tr}(\dot{X}_j'\dot{X}_q M_{F_0}) + o_p(1)$ and

$$\frac{1}{nT\sigma_{v0}^2}\,\mathrm{tr}[\dot{X}_j'(\dot{\Gamma}F_0' + \mathbb{V})\frac{\partial}{\partial\psi_q}M_{\hat{F}(\psi)}|_{\psi_0}] = \frac{1}{nT\sigma_{v0}^2}\,\mathrm{tr}(\dot{X}_j'P_{\dot{\Gamma}}\dot{X}_q M_{F_0}) + o_p(1).$$

It follows that $-\frac{1}{nT}\frac{\partial}{\partial\psi_q}\tilde{S}_{j,nT}^*(\psi)|_{\psi_0} = \frac{1}{nT\sigma_{v0}^2}\,\mathrm{tr}(\dot{X}_j'P_{\dot{\Gamma}}\dot{X}_q M_{F_0}) + o_p(1)$. For the case $q = k+4$ and $k+5$, we can show it is $o_p(1)$.

Next, consider the case (i-c). For $j = k+1, \ldots, k+3$, derivatives of $S_{j,nT}^p(\psi)$ can be similarly studied. In addition, we need to consider terms like $\frac{\partial}{\partial \beta_q} \operatorname{tr}[\mathbf{M}_{\hat{F}(\psi)} \mathbf{W}_1 \mathbf{D}(\rho, \lambda_1, \lambda_2)]|_{\psi_0}$. Lemma D.2 has found the asymptotic leading terms. Some algebra leads to the desired results.

(ii) For the case $j = k+4$ and $q \leq k+3$, we have

$$\begin{aligned}
-\frac{1}{nT} \frac{\partial \tilde{S}^*_{j,nT}(\psi)}{\partial \psi_q} |_{\psi_0} = & \frac{1}{nT\sigma^2_{v0}} \operatorname{tr}[\dot{X}'_q (W_3 B^{-1}_{30} + (B'_{30})^{-1} W'_3)(\dot{\Gamma} F'_0 + \mathbb{V}) M_{\tilde{F}}] \\
& - \frac{1}{nT\sigma^2_{v0}} \operatorname{tr}[(\dot{\Gamma} F'_0 + \mathbb{V})' W_3 B^{-1}_{30} (\dot{\Gamma} F'_0 + \mathbb{V}) \frac{\partial M_{\hat{F}(\psi)}}{\partial \psi_q} |_{\psi_0}].
\end{aligned}$$

We can verity that both terms on the right hand side are $o_p(1)$. The same result holds for the case $q = k+5$.

For the case $j = q = k+4$, we have

$$\begin{aligned}
-\frac{1}{nT} \frac{\partial \tilde{S}^*_{j,nT}(\psi)}{\partial \psi_q} |_{\psi_0} = & \frac{1}{nT\sigma^2_{v0}} \operatorname{tr}[(\dot{\Gamma} F'_0 + \mathbb{V})'(B'_{30})^{-1} W'_3 W_3 B^{-1}_{30} (\dot{\Gamma} F'_0 + \mathbb{V}) M_{\tilde{F}}] \\
& - \frac{1}{nT\sigma^2_{v0}} \operatorname{tr}[(\dot{\Gamma} F'_0 + \mathbb{V})' W_3 B^{-1}_{30} (\dot{\Gamma} F'_0 + \mathbb{V}) \frac{\partial M_{\hat{F}(\psi)}}{\partial \psi_q} |_{\psi_0}] + \frac{T-r}{nT} \operatorname{tr}(W_3 B^{-1}_{30} W_3 B^{-1}_{30}) \\
= & \frac{1}{nT} \operatorname{tr}(\mathbb{V}' B^{-1\prime}_{30} W'_3 W_3 B^{-1}_{30} \mathbb{V} M_{F_0}) + \frac{T-r}{nT\sigma^2_{v0}} \operatorname{tr}(W_3 B^{-1}_{30} W_3 B^{-1}_{30}) + o_p(1) \\
= & \frac{1}{n} \operatorname{tr}(B^{-1\prime}_{30} W'_3 W_3 B^{-1}_{30} + W_3 B^{-1}_{30} W_3 B^{-1}_{30}) + o_p(1).
\end{aligned}$$

(iii) The derivation is similar to parts (i-ii). ■

**Proof of Lemma C.4:** First, consider the case $j, q \leq k$. By Lemma D.3, we can show that

$$\tilde{\Sigma}_{nT,(jq)} = \frac{\sigma^2_{v0}}{nT} \tilde{\Pi}'_j \tilde{\Pi}_q = \frac{1}{nT\sigma^2_{v0}} \operatorname{tr}(\dot{X}'_j M_{\dot{\Gamma}} \dot{X}_q M_{F_0}) = \tilde{H}_{nT,(jq)}.$$

Second, consider the case $j \leq k, q = k+1$. We have

$$\begin{aligned}
\operatorname{Cov}(\tilde{S}_{j,nT}, \tilde{S}_{q,nT} \mid \mathcal{D}) = & \sigma^2_{v0} \tilde{\Pi}'_{1,j} (\tilde{\Pi}_2 + \tilde{\Psi}_1 \mathbf{y}_0) + \mu_3 \tilde{\Pi}'_{1,j} \texttt{diag}(\tilde{\Phi}_1) \\
= & \sigma^2_{v0} \tilde{\Pi}'_{1,j} [(\tilde{\Pi}_2 + \tilde{\Psi}_1 \mathbf{y}_0) + \tilde{\Phi}_1 \mathbf{v}] + \mu_3 \tilde{\Pi}'_{1,j} \texttt{diag}(\tilde{\Phi}_1) + o_p(nT) \\
= & \frac{1}{\sigma^2_{v0}} \operatorname{tr}(\dot{X}'_j M_{\dot{\Gamma}} \dot{X}_{k+1} M_{F_0}) + \mu_3 \tilde{\Pi}'_{1,j} \texttt{diag}(\tilde{\Phi}_1) + o_p(nT).
\end{aligned}$$

It follows that $\tilde{\Sigma}_{nT,(jq)} - \tilde{H}_{nT,(jq)} = \mu_3 \tilde{\Pi}'_{1,j} \texttt{diag}(\tilde{\Phi}_1)/(nT) + o_p(1)$. For the other cases where $j \leq k, q = k+1, \ldots, k+5$, the analysis is identical to that of $q = k+1$ case.
Third, consider the case $j = q = k+1$. We have,

$$\begin{aligned}
\operatorname{Var}(\tilde{S}_{k+1,nT} \mid \mathcal{D}) = & \sigma^2_{v0} \|\tilde{\Pi}_2 + \tilde{\Psi}_1 \mathbf{y}_0\|^2 + 2\mu_3 (\tilde{\Pi}_2 + \tilde{\Psi}_1 \mathbf{y}_0)' \texttt{diag}(\tilde{\Phi}_1) \\
& + \sigma^4_{v0} \operatorname{tr}(\tilde{\Phi}_1 \tilde{\Phi}_1 + \tilde{\Phi}_1 \tilde{\Phi}'_1) + (\mu_4 - 3\sigma^4_{v0}) \|\texttt{diag}(\tilde{\Phi}_1)\|^2.
\end{aligned}$$

And we have the identity that

$$
\begin{aligned}
\tilde{H}_{nT,(jj)} =& \frac{1}{\sigma_{v0}^2}\texttt{vec}(\dot{X}_{k+1})'(M_{F_0}\otimes M_{\dot{\Gamma}})\texttt{vec}(\dot{X}_{k+1}) \\
=& \sigma_{v0}^2\|(M_{F_0}\otimes M_{\dot{\Gamma}}B_{30})(\mathbf{Q}_{-1}\mathbf{y}_0+\boldsymbol{\eta}_{-\mathbf{1}}+\mathbf{D}_{-1}\mathbf{B}_{30}^{-1}\texttt{vec}(\dot{\Gamma}F_0'+\mathbb{V}))/\sigma_{v0}^2\|^2 \\
=& \sigma_{v0}^2\|\tilde{\Pi}_2+\tilde{\Psi}_1\mathbf{y}_0+\tilde{\Phi}_1\mathbf{v}\|^2 \\
=& \sigma_{v0}^2\|\tilde{\Pi}_2+\tilde{\Psi}_1\mathbf{y}_0\|^2+\sigma_{v0}^2\mathbf{v}'\tilde{\Phi}_1'\tilde{\Phi}_1\mathbf{v}'+2\sigma_{v0}^2\mathbf{v}'\tilde{\Phi}_1'(\tilde{\Pi}_2+\tilde{\Psi}_1\mathbf{y}_0) \\
=& \sigma_{v0}^2\|\tilde{\Pi}_2+\tilde{\Psi}_1\mathbf{y}_0\|^2+\sigma_{v0}^4\,\mathrm{tr}(\tilde{\Phi}_1\tilde{\Phi}_1')+o_p(nT).
\end{aligned}
$$

Combining the above two identities, we can find the leading term of $\tilde{\Xi}_{nT,(jj)}$. The rest of the results can be verified by similar analysis. ∎

**Proof of Lemma C.5:** (i) By Lemma D.3 we have $\mathrm{Var}(\Pi_{4+a}'\mathbf{v}) = \sigma_{v0}^2\|\Pi_{4+a}\|^2 = O_p(n)$ and $\mathrm{Var}(\mathbf{v}'\Phi_{5+a}\mathbf{v}) = (\mu_4-3\sigma_{v0}^2)\|\texttt{diag}(\Phi_{5+a})\|^2 = O_p(n/T^2)$. It follows that $\Pi_{4+a}'\mathbf{v}/\sqrt{nT} = O_p(T^{-1/2})$ and $\mathbf{v}'\Phi_{5+a}\mathbf{v}/\sqrt{nT} = O_p(T^{-3/2})$. Hence, one can easily see that

$$
\begin{aligned}
\Sigma_{nT,(k+5+a,k+5+b)} =& \frac{\sigma_{v0}^2}{nT}\Pi_{4+a}'\Pi_{4+b}+O_p(T^{-2}) = \frac{1}{\sigma_{v0}^2 nT}\,\mathrm{tr}(\dot{\Gamma}'\dot{\Gamma}\dot{F}_a'M_{F_0}\dot{F}_b)+O_p(T^{-2}) \\
=& \frac{1}{\sigma_{v0}^2 nT}\texttt{vec}(\dot{F}_a')'(M_{F_0}\otimes\dot{\Gamma}'\dot{\Gamma})\texttt{vec}(\dot{F}_b')+O_p(T^{-2}).
\end{aligned}
$$

In addition, we show that the absolute column sums of $\Sigma_{nT,\phi\phi}$ and $H_{nT,\phi\phi}$ is $o_p(T^{-1})$. Therefore, we have the spectrum norm bound of their difference. For the estimator $\hat{\Sigma}_{nT,\phi\phi}$, we only need to find the asymptotic orders of $\frac{1}{nT}\sum_{i=1}^n(\hat{\mathbf{g}}_i\hat{\mathbf{g}}_i'-\mathbf{g}_i\mathbf{g}_i')$. The derivation is tedious but straightforward, and thus is omitted.

(ii) First consider the $j\le k$. Similar to part (i), we have that

$$
\Sigma_{nT,(k+5+a,j)} = \frac{\sigma_{v0}^2}{nT}\Pi_{1,j}'\Pi_{4+a}+o_p(T^{-3/2})\frac{1}{\sigma_{v0}^2 nT}\texttt{vec}(\dot{X}_j)'(M_{F_0}\otimes\dot{\Gamma})\texttt{vec}(\dot{F}_a')+o_p(T^{-3/2}).
$$

Next, consider the case $j=k+1$. We have that

$$
\begin{aligned}
\Sigma_{nT,(k+5+a,j)} =& \frac{\sigma_{v0}^2}{nT}(\Pi_2+\Psi_1\mathbf{y}_0)'\Pi_{4+a}+\frac{\mu_3}{nT}\texttt{diag}(\Phi_1)'\Pi_{4+a}+O_p(T^{-3/2}) \\
=& \frac{\sigma_{v0}^2}{nT}(\Pi_2+\Psi_1\mathbf{y}_0+\Phi_1\mathbf{v})'\Pi_{4+a}+\frac{\mu_3}{nT}\texttt{diag}(\Phi_1)'\Pi_{4+a}+o_p(T^{-1}) \\
=& \frac{1}{\sigma_{v0}^2 nT}\texttt{vec}(\dot{X}_j)'(M_{F_0}\otimes\dot{\Gamma})\texttt{vec}(\dot{F}_a')+\frac{\mu_3}{nT}\texttt{diag}(\Phi_1)'\Pi_{4+a}+o_p(T^{-1}).
\end{aligned}
$$

The cases $j=k+2,\dots,k+5$ can be similarly studied.

(iii) The proof is identical to that of Lemma C.4. ∎

**Proof of Lemma C.6:** (i) It suffices to study that terms like $\frac{1}{nT}\,\mathrm{tr}[(M_{F_0}\otimes I_n)\dot{\mathbf{D}}_{-1,\rho}]$. By

Lemma D.4, we have

$$\begin{aligned}\frac{1}{nT}\operatorname{tr}[(M_{F_0}\otimes I_n)\dot{\mathbf{D}}_{-1,\psi_j}] =& o_p(1),\ \text{ for } j=k+1,k+2,k+3\\ \frac{1}{nT}\operatorname{tr}[(M_{F_0}\otimes I_n)\dot{\mathbf{D}}_{\psi_j}] =& o_p(1),\ \text{ for } j=k+1,k+3\\ \frac{1}{nT}\operatorname{tr}[(M_{F_0}\otimes W_2)\dot{\mathbf{D}}_{-1,\psi_j}] =& o_p(1),\ \text{ for } j=k+1,k+2,k+3,\end{aligned}$$

where $\psi_{k+1}=\rho$, $\psi_{k+2}=\lambda_1$, and $\psi_{k+3}=\lambda_2$.

(ii) Note that $H_{nT,(k+5+a,k+5+b)}=\frac{1}{2nT\sigma_{v0}^2}\operatorname{tr}[\mathbb{Z}(\theta_0)'B_{30}'B_{30}\mathbb{Z}(\theta_0)\ddot{P}_{F,ab}]$. We have,

$$\begin{aligned}\ddot{P}_{F,ab}=\frac{\partial^2 P_F}{\partial\phi_a\partial\phi_b}|_{\psi_0} =& F_0(F_0'F_0)^{-1}(\dot{F}_a'M_{F_0}\dot{F}_b+\dot{F}_b'M_{F_0}\dot{F}_a)(F_0'F_0)^{-1}F_0'\\ &-M_{F_0}[\dot{F}_a(F_0'F_0)^{-1}\dot{F}_b'+\dot{F}_b(F_0'F_0)^{-1}\dot{F}_a']M_{F_0}\\ &+M_{F_0}\dot{F}_a(F_0'F_0)^{-1}F_0'\dot{F}_b(F_0'F_0)^{-1}F_0'+F_0(F_0'F_0)^{-1}\dot{F}_b'F_0(F_0'F_0)^{-1}\dot{F}_a'M_{F_0}\\ &+M_{F_0}\dot{F}_b(F_0'F_0)^{-1}F_0'\dot{F}_a(F_0'F_0)^{-1}F_0'+F_0(F_0'F_0)^{-1}\dot{F}_a'F_0(F_0'F_0)^{-1}\dot{F}_b'M_{F_0}.\end{aligned}$$

Plugging in the expression of $\ddot{P}_{F,ab}$ into $H_{nT,(k+5+a,k+5+b)}=\frac{1}{2nT\sigma_{v0}^2}\operatorname{tr}[\mathbb{Z}(\theta_0)'B_{30}'B_{30}\mathbb{Z}(\theta_0)\ddot{P}_{F,ab}]$, we obtain the results after some tedious algebra.

(iii) First consider the $j\le k$ case. We have,

$$\begin{aligned}H_{nT,(j,k+5+a)}=&-\frac{1}{nT\sigma_{v0}^2}\operatorname{tr}(\dot{X}_j'\dot{\Gamma}F_0'\dot{P}_{F,a})-\frac{1}{nT\sigma_{v0}^2}\operatorname{tr}(\dot{X}_j'\mathbb{V}\dot{P}_{F,a})\\ =&\frac{1}{nT\sigma_{v0}^2}\operatorname{tr}(\dot{X}_j'\dot{\Gamma}\dot{F}_a'M_{F_0})+O_p(n^{-1/2}T^{-1})\\ =&\frac{1}{nT\sigma_{v0}^2}\mathtt{vec}(\dot{F}_a')'(M_{F_0}\otimes\dot{\Gamma}')\mathtt{vec}(\dot{X}_j)+O_p(n^{-1/2}T^{-1}).\end{aligned}$$

Next consider the cases $j=k+1$, we have,

$$H_{nT,(j,k+5+a)}=\frac{1}{nT\sigma_{v0}^2}\operatorname{tr}[\dot{X}_j'B_{30}\mathbb{Z}(\theta_0)\dot{P}_{F,a}]+\frac{1}{nT}\operatorname{tr}[(\dot{P}_{F,a}\otimes I_n)\mathbf{D}_{-1}]$$

The analysis is similar to the $j\le k$ case. However, one difference is that now we need to use the result $-\frac{1}{nT\sigma_{v0}^2}\operatorname{tr}(\dot{X}_{k+1}'\mathbb{V}\dot{P}_{F,a})+\frac{1}{nT}\operatorname{tr}[(\dot{P}_{F,a}\otimes I_n)\mathbf{D}_{-1}]=o_p(T^{-1})$. The key point is that $\dot{X}_{k+1}$ is $B_{30}Y_{-1}$ which is no longer exogenous. Its mean is $\frac{1}{nT}\operatorname{tr}[(\dot{P}_{F,a}\otimes I_n)\mathbf{D}_{-1}]$, which is not of small order. But the recentered term can be verified to be small. The same analysis procedure gives us the result for cases $j=k+2,\ldots,k+5$.

(iv) It suffices to find the exact expression of $-\frac{1}{nT\sigma_{v0}^2}\operatorname{tr}(\dot{X}_j'\mathbb{V}\dot{P}_{F,a})$:

$$\begin{aligned}-\frac{1}{nT\sigma_{v0}^2}\operatorname{tr}(\dot{X}_j'\mathbb{V}\dot{P}_{F,a})=&\frac{1}{nT\sigma_{v0}^2}\operatorname{tr}[\dot{X}_j'\mathbb{V}(M_{F_0}\dot{F}_a(F_0'F_0)^{-1}F_0'+F_0(F_0'F_0)^{-1}\dot{F}_a'M_{F_0})]\\ =&\frac{1}{nT\sigma_{v0}^2}\operatorname{tr}[(F_0'F_0)^{-1}\dot{F}_a'M_{F_0}(\dot{X}_j'\mathbb{V}+\mathbb{V}'\dot{X}_j)F_0]\\ =&\frac{1}{nT\sigma_{v0}^2}\mathtt{vec}[F_0'(\dot{X}_j'\mathbb{V}+\mathbb{V}'\dot{X}_j)]'[M_{F_0}\otimes(F_0'F_0)^{-1}]\mathtt{vec}(\dot{F}_a').\end{aligned}$$

This term is $O_p(T^{-1})$, for $j=k+1,\ldots,k+3$.

(v) Recall $H_{nT,*} = H_{nT,\psi\psi} - H_{nT,\psi\phi}H_{nT,\phi\phi}^{-1}H_{nT,\phi\psi}$, in the inverse of the partitioned $H_{nT}$:

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

The key step to establishing the result is to study the second term $H_{nT,\psi\phi}(H_{nT,\phi\phi})^{-1}H_{nT,\phi\psi}$, which involves high-dimensional matrices.

First, consider $(H_{nT,\phi\phi})^{-1}$, we can show that $\mathcal{S} = [\texttt{vec}(\dot{F}_1'), \ldots, \texttt{vec}(\dot{F}_{r(T-r)'})] = S_* \otimes L$, where $S_* = [I_{T-r}, 0'_{r\times(T-r)}]'$, $L = [e_{r,r}, \ldots, e_{1,r}]$, and $e_{j,r}$ is the $j$th unit vector of dimension $r$. Using the result $M_{F_0} = M_{F_0}\mathcal{S}_*(\mathcal{S}_*'M_{F_0}\mathcal{S}_*)^{-1}\mathcal{S}_*'M_{F_0}$ in Lemma D.5. We can obtain,

$$\|H_{nT,\phi\phi} - \tfrac{1}{nT\sigma_{v0}^2}(S_*'M_{F_0}S_* \otimes L\dot{\Gamma}'\dot{\Gamma}L)\|_{\texttt{sp}} = o_p(T^{-1}).$$

It follows that $\|(H_{nT,\phi\phi})^{-1} - nT\sigma_{v0}^2[(S_*'M_{F_0}S_*)^{-1} \otimes L(\dot{\Gamma}'\dot{\Gamma})^{-1}L]\|_{\texttt{sp}} = o_p(T)$.

Next, we consider $H_{nT,\psi\phi}$. For its $j$th row, $H_{nT,j\phi}$, we obtain

$$\|H_{nT,j\phi} - \tfrac{1}{nT\sigma_{v0}^2}\texttt{vec}(\dot{X}_j)'(M_{F_0}\mathcal{S}_* \otimes \dot{\Gamma}L)\| = o_p(T^{-1/2}), \qquad \text{for } j = 1, \ldots, k+3;$$
$$\|H_{nT,j\phi}\| = o_p(T^{-1/2}), \qquad \text{for } j = k+4, k+5.$$

Finally, consider $H_{nT,\phi\psi}$. Its $j$th column has the following property,
$H_{nT,\phi j} = \frac{1}{nT\sigma_{v0}^2}\{(\mathcal{S}_*'M_{F_0} \otimes L\dot{\Gamma}')\texttt{vec}(\dot{X}_j) + [\mathcal{S}_*'M_{F_0} \otimes L(F_0'F_0)^{-1}]\texttt{vec}[F_0'(\dot{X}_j'\mathbb{V} + \mathbb{V}'\dot{X}_j)]\}$.
One can verify that

$$\begin{aligned} &\tfrac{1}{nT\sigma_{v0}^2}\texttt{vec}(\dot{X}_q)'[M_{F_0} \otimes \dot{\Gamma}(\dot{\Gamma}'\dot{\Gamma})^{-1}(F_0'F_0)^{-1}]\texttt{vec}[F_0'(\dot{X}_j'\mathbb{V} + \mathbb{V}'\dot{X}_j)] \\ &= \tfrac{1}{nT\sigma_{v0}^2}\,\text{tr}[\dot{X}_q'\dot{\Gamma}(\dot{\Gamma}'\dot{\Gamma})^{-1}(F_0'F_0)^{-1}F_0'(\dot{X}_j'\mathbb{V} + \mathbb{V}'\dot{X}_j)] = o_p(1). \end{aligned}$$

Combine the above results, we have, the $(j,q)$ element of $H_{nT,\psi\phi}(H_{nT,\phi\phi})^{-1}H_{nT,\phi\psi}$ is

$$\tfrac{1}{nT\sigma_{v0}^2}\texttt{vec}(\dot{X}_j)'(M_{F_0} \otimes P_{\dot{\Gamma}})\texttt{vec}(\dot{X}_q) + o_p(1) \text{ for } j, q \le k+3$$

and is $o_p(1)$ for $j, q = k+4, k+5$. Summing up, we have shown that $H_{nT,*} = \tilde{H}_{nT} + o_p(1)$. ■

**Proof of Lemma C.7:** We take five steps to establish the result and within each step we discuss several cases.

(i) First, we consider the term $H_{nT,\psi\phi}(H_{nT,\phi\phi})^{-1}\Sigma_{nT,\phi\psi}$. The leading term of its $(j,q)$th entire can be summarized as follows:

- (i-1) $\frac{1}{\sigma_{v0}^2 nT}\texttt{vec}(\dot{X}_j)'(M_{F_0} \otimes P_{\dot{\Gamma}})\texttt{vec}(\dot{X}_q)$, for $j, q \le k+3$;
- (i-2) $\frac{1}{nT\sigma_{v0}^2}\,\text{tr}(\dot{X}_j'P_{\dot{\Gamma}}\dot{X}_qM_{F_0}) + \frac{\mu_3}{nT}\texttt{vec}(\dot{X}_j)'(M_{F_0} \otimes P_{\dot{\Gamma}})\texttt{diag}(\Phi_a) + o_p(1)$, for $j \le k+3, q = k+a, a = 1, 2, 3$;
- (i-3) $\frac{\mu_3}{nT}\texttt{vec}(\dot{X}_j)'(M_{F_0} \otimes P_{\dot{\Gamma}})\texttt{diag}(\Phi_a)$, for $j \le k+3, q = k+a, a = 4, 5$;
- (i-4) $o_p(1)$, for $j = k+4, k+5$.

To see the results, we take $q = k + 1$ in case (i-2) as an example, the leading term of the $(j, q)$th entry of $H_{nT,\psi\phi}(H_{nT,\phi\phi})^{-1}\Sigma_{nT,\phi\psi}$ is in this case,

$$\begin{aligned} &\tfrac{1}{nT}\mathtt{vec}(\dot{X}_j)'(M_{F_0} \otimes P_{\dot{\Gamma}})[\Pi_2 + \Psi_1 \mathbf{y}_0 + \mu_3 \mathtt{diag}(\Phi_1)] \\ = &\tfrac{1}{nT\sigma_{v0}^2} \operatorname{tr}(\dot{X}_j' P_{\dot{\Gamma}} \dot{X}_q M_{F_0}) + \tfrac{\mu_3}{nT}\mathtt{vec}(\dot{X}_j)'(M_{F_0} \otimes P_{\dot{\Gamma}})\mathtt{diag}(\Phi_1) + o_p(1). \end{aligned}$$

The equality is by the fact that $\frac{1}{nT}\mathtt{vec}(\dot{X}_j)'(M_{F_0} \otimes P_{\dot{\Gamma}})\Phi_1 \mathbf{v} = 0$ and the decomposition of $\mathbf{Y}_{-1}$. Other cases can be shown similarly.

(ii) Second, we consider the term $\Sigma_{nT,\psi\phi}(H_{nT,\phi\phi})^{-1}H_{nT,\phi\psi}$. Compared to $H_{nT,\psi\phi}$, the $(*, k+1 : k+3)$ entries of $H_{nT,\phi\psi}$ has an additional term, as in Lemma C.6. However, we can show that these additional term only results in additional $o_p(1)$ terms in the final product $\Sigma_{nT,\psi\phi}(H_{nT,\phi\phi})^{-1}H_{nT,\phi\psi}$. The leading terms of the $(j, q)$th entry is summarized as:

(ii-1) $\frac{1}{nT\sigma_{v0}^2} \operatorname{tr}(\dot{X}_j' P_{\dot{\Gamma}} \dot{X}_q M_{F_0})$, for $j, q \le k + 3$;

(ii-2) $\frac{1}{nT\sigma_{v0}^2} \operatorname{tr}(\dot{X}_j' P_{\dot{\Gamma}} \dot{X}_q M_{F_0}) + \frac{\mu_3}{nT}\mathtt{vec}(\dot{X}_q)'(M_{F_0} \otimes P_{\dot{\Gamma}})\mathtt{diag}(\Phi_a)$,
for, $j = k + a, a = 1, 2, 3,\ q \le k + 3$;

(ii-3) $\frac{\mu_3}{nT}\mathtt{vec}(\dot{X}_q)'(M_{F_0} \otimes P_{\dot{\Gamma}})\mathtt{diag}(\Phi_a)$, for $j = k + a, a = 4, 5,\ q \le k + 3$;

(ii-4) $o_p(1)$, for $q = k + 4, k + 5$.

(iii) Third, consider the last term $H_{nT,\psi\phi}(H_{nT,\phi\phi})^{-1}H_{nT,\phi\psi}$. We have already found the leading term of its $(j, q)$th entry in Lemma C.6, which can be summarized as follows:

(iii-1) $\frac{1}{nT\sigma_{v0}^2} \operatorname{tr}(\dot{X}_j' P_{\dot{\Gamma}} \dot{X}_q M_{F_0})$, for $j, q \le k + 3$;

(iii-2) $o_p(1)$, for the other cases.

(iv) Then we derive the summation of three matrices studied in parts (i-iii).

(iv-1) For $j, q \le k$ the $(j, q)$th entry has leading term $-\frac{1}{nT\sigma_{v0}^2} \operatorname{tr}(\dot{X}_j' P_{\dot{\Gamma}} \dot{X}_q M_{F_0})$;

(iv-2) For $j \le k$, $q = k + a$, $a = 1, 2, 3$, the $(j, q)$th entry has the leading term

$$-\tfrac{1}{nT\sigma_{v0}^2} \operatorname{tr}(\dot{X}_j' P_{\dot{\Gamma}} \dot{X}_q M_{F_0}) - \tfrac{\mu_3}{nT}\mathtt{vec}(\dot{X}_j)'(M_{F_0} \otimes P_{\dot{\Gamma}})\mathtt{diag}(\Phi_a).$$

The leading term of $(q, j)$th entry is identical to the above one.

(iv-3) For $j \le k$, $q = k + a$, $a = 4, 5$, the $(j, q)$th entry has the leading term

$$-\tfrac{\mu_3}{nT}\mathtt{vec}(\dot{X}_j)'(M_{F_0} \otimes P_{\dot{\Gamma}})\mathtt{diag}(\Phi_a).$$

The leading term of $(q, j)$th entry is identical to that of the $(j, q)$th entry.

(iv-4) For $j = k + a$ and $q = k + b$, where $a, b = 1, 2, 3$, the leading term is

$$-\tfrac{1}{nT\sigma_{v0}^2} \operatorname{tr}(\dot{X}_j' P_{\dot{\Gamma}} \dot{X}_q M_{F_0}) - \tfrac{\mu_3}{nT}\mathtt{vec}(\dot{X}_j)'(M_{F_0} \otimes P_{\dot{\Gamma}})\mathtt{diag}(\Phi_b) - \tfrac{\mu_3}{nT}\mathtt{vec}(\dot{X}_q)'(M_{F_0} \otimes P_{\dot{\Gamma}})\mathtt{diag}(\Phi_a);$$

(iv-5) For $j = k + a$ and $q = k + b$, where $a, b = 4, 5$, the terms are $o_p(1)$.

(v) Finally, we look at the difference between $\tilde{\Sigma}_{nT}$ and $\Sigma_{nT}$. It suffices to compare the leading terms as in Lemmas C.4 and C.5.

(v-1) For $j \le k$, the leading term of $\tilde{\Sigma}_{jq,nT}$ is

$\frac{\sigma_{v0}^2}{nT}\tilde{\Pi}_{1,j}'\tilde{\Pi}_{1,q}$, for $q \le k$;

$\frac{\sigma_{v0}^2}{nT}\tilde{\Pi}_{1,j}'\tilde{\Pi}_{k+a} + \frac{\mu_3}{nT}\tilde{\Pi}_{1,j}'\texttt{diag}(\tilde{\Phi}_a)$, for $q = k+a$, $a = 1,2,3$;

$\frac{\mu_3}{nT}\tilde{\Pi}_{1,j}'\texttt{diag}(\tilde{\Phi}_a)$, for $q = k+a$, $a = 4,5$.

By Lemma D.5, the above terms can be rewritten as

$\Sigma_{jq,nT} - \frac{1}{nT\sigma_{v0}^2}\operatorname{tr}(\dot{X}_j' P_{\dot{\Gamma}}\dot{X}_q M_{F_0}) + o_p(1)$, for $q \le k$;

$\Sigma_{jq,nT} - \frac{1}{nT\sigma_{v0}^2}\operatorname{tr}(\dot{X}_j' P_{\dot{\Gamma}}\dot{X}_q M_{F_0}) - \frac{\mu_3}{nT}\Pi_{1,j}'(I_T \otimes P_{\dot{\Gamma}})\texttt{diag}(\Phi_a) + o_p(1)$, for $q = k+a, a = 1,2,3$;

$\Sigma_{jq,nT} - \frac{\mu_3}{nT}\Pi_{1,j}'(I_T \otimes P_{\dot{\Gamma}})\texttt{diag}(\Phi_a) + o_p(1)$, for $q = k+a$, $a = 4,5$.

(v-2) Consider cases with $j = k+a$, and $a = 1,2,3$. We show the case $q = k+1$.

$$\begin{aligned}
\tilde{\Sigma}_{jj,nT} &= \frac{\sigma_{v0}^2}{nT}\|\tilde{\Pi}_2 + \tilde{\Psi}_1\mathbf{y}_0\|^2 + \frac{2\mu_3}{nT}(\tilde{\Pi}_2 + \tilde{\Psi}_1\mathbf{y}_0)'\texttt{diag}(\tilde{\Phi}_1) + \frac{\sigma_{v0}^4}{nT}\operatorname{tr}(\tilde{\Phi}_1\tilde{\Phi}_1 + \tilde{\Phi}_1\tilde{\Phi}_1') + \frac{(\mu_4 - 3\sigma_{v0}^4)}{nT}\|\texttt{diag}(\tilde{\Phi}_1)\|^2 \\
&= \frac{1}{\sigma_{v0}^2 T}\operatorname{tr}[\dot{X}_j'\dot{X}_j M_{F_0}] + \frac{2\mu_3}{nT}(\Pi_2 + \Psi_1\mathbf{y}_0)'\texttt{diag}(\tilde{\Phi}_1) + \frac{\sigma_{v0}^4}{nT}\operatorname{tr}(\Phi_1\Phi_1) + \frac{(\mu_4 - 3\sigma_{v0}^4)}{nT}\|\texttt{diag}(\Phi_1)\|^2 \\
&\quad - \frac{1}{\sigma_{v0}^2 nT}\operatorname{tr}(\dot{X}_j' P_{\dot{\Gamma}}\dot{X}_j M_{F_0}) - \frac{2\mu_3}{nT}\sigma_{v0}^2(\Pi_2 + \Psi_1\mathbf{y}_0)'(M_{F_0} \otimes P_{\dot{\Gamma}})\texttt{diag}(\tilde{\Phi}_1) + o_p(1) \\
&= \Sigma_{jq,nT} - \frac{1}{\sigma_{v0}^2 nT}\operatorname{tr}(\dot{X}_j' P_{\dot{\Gamma}}\dot{X}_j M_{F_0}) - \frac{2\mu_3}{nT}\texttt{vec}(\dot{X}_j)'(M_{F_0} \otimes P_{\dot{\Gamma}})\texttt{diag}(\tilde{\Phi}_1) + o_p(1).
\end{aligned}$$

For other cases we have can follow a similar analysis to show the following results.

For $j = k+a, q = k+b$, $a, b \le 3$,

$$\begin{aligned}
\tilde{\Sigma}_{jq,nT} &= \Sigma_{jq,nT} - \frac{1}{\sigma_{v0}^2 nT}\operatorname{tr}(\dot{X}_j' P_{\dot{\Gamma}}\dot{X}_j M_{F_0}) - \frac{\mu_3}{nT}\texttt{vec}(\dot{X}_j)'(M_{F_0} \otimes P_{\dot{\Gamma}})\texttt{diag}(\tilde{\Phi}_b) \\
&\quad - \frac{\mu_3}{nT}\texttt{vec}(\dot{X}_q)'(M_{F_0} \otimes P_{\dot{\Gamma}})\texttt{diag}(\tilde{\Phi})_a + o_p(1);
\end{aligned}$$

For $j = k+a, q = k+b$, $a \le 3$, and $b = 4,5$,

$$\tilde{\Sigma}_{jq,nT} = \Sigma_{jq,nT} - \frac{\mu_3}{nT}\texttt{vec}(\dot{X}_j)'(M_{F_0} \otimes P_{\dot{\Gamma}})\texttt{diag}(\Phi_b) + o_p(1).$$

(v-3) Consider the cases with $j = k+4, k+5$. We can show.

$$\tilde{\Sigma}_{jq,nT} = \Sigma_{jq,nT} + o_p(1), \text{ for } q = k+4, k+5.$$

Therefore,we have shown that $\Sigma_{nT}^* = \tilde{\Sigma}_{nT} + o_p(1)$. ■

# Appendix D: Additional Technical Lemmas

Additonal technical lemmas D1-D5 are given in this appendix, which are used in the proof of Lemmas C1-C7. The following results are useful in their proofs: $\mathbf{D} = \sum_{j=0}^{T-1}(J_T^j \otimes \mathcal{B}_0^j B_{10}^{-1})$ and $\mathbf{D}_{-1} = \sum_{j=1}^{T-1}(J_T^j \otimes \mathcal{B}_0^{j-1} B_{10}^{-1})$, where $J_T = [0_{1\times(T-1)}, 0; I_{T-1}, 0_{(T-1)\times 1}]$; $\|\mathbf{D}\|_F = O(\sqrt{nT})$ and $\|\mathbf{D}_{-1}\|_F = O(\sqrt{nT})$. Let $\texttt{mat}(\cdot)$ be the reverse operator of $\texttt{vec}$, that is $\texttt{mat}(\texttt{vec}(X_j)) = X_j$.

**Lemma D.1.** *Suppose Assumptions A-H hold and $T/n \to 0$. Then we have*

*(i)* $\operatorname{tr}[\dot{X}_j'(\mathbb{V} + \dot{\Gamma}F_0')(M_{\tilde{F}} - M_{F_0})] = -\operatorname{tr}(\dot{X}_j' P_{\dot{\Gamma}}\mathbb{V}M_{F_0}) + o_p(\sqrt{nT})$,

*(ii)* $\operatorname{tr}[(M_{\tilde{F}} \otimes I_n)\mathbf{D}_{-1}] = \operatorname{tr}[(M_{F_0} \otimes M_{\dot{\Gamma}})\mathbf{D}_{-1}] + o_p(\sqrt{nT})$,
$\operatorname{tr}[(M_{\tilde{F}} \otimes W_1)\mathbf{D}] = \operatorname{tr}[(M_{F_0} \otimes M_{\dot{\Gamma}}W_1)\mathbf{D}] + o_p(\sqrt{nT})$
$\operatorname{tr}[(M_{\tilde{F}} \otimes W_2)\mathbf{D}_{-1}] = \operatorname{tr}[(M_{F_0} \otimes M_{\dot{\Gamma}}W_2)\mathbf{D}_{-1}] + o_p(\sqrt{nT})$

*(iii)* $\operatorname{tr}[(\mathbb{V} + \dot{\Gamma}F_0')'(\mathbb{V} + \dot{\Gamma}F_0')(M_{\tilde{F}} - M_{F_0})] = o_p(\sqrt{nT})$,
$\operatorname{tr}[(\mathbb{V} + \dot{\Gamma}F_0')'W_3B_{30}^{-1}(\mathbb{V} + \dot{\Gamma}F_0')(M_{\tilde{F}} - M_{F_0})] = o_p(\sqrt{nT})$ .

**Lemma D.2.** *Suppose Assumptions A-H hold and $T/n \to 0$. Let $\dot{X}_{k+1} = B_{30}Y_{-1}$, $\dot{X}_{k+2} = B_{30}W_1Y$, $\dot{X}_{k+3} = B_{30}W_2Y_{-1}$, and $\dot{X}_{k+4} = W_3B_{30}^{-1}(\dot{\Gamma}F_0' + \mathbb{V})$. We have,*

*(i)* $\left\| \frac{\partial}{\partial \psi_q} M_{\hat{F}(\psi)}\big|_{\psi_0} - M_{F_0}\dot{X}_p'(F_0\dot{\Gamma}')^{\dagger} - (\dot{\Gamma}F_0')^{\dagger}\dot{X}_q M_{F_0} \right\|_{sp} = o_p(1)$, *for* $p = 1, \ldots, k+4$*;*

*(ii)* $\frac{\partial}{\partial \psi_p} \operatorname{tr}[\mathbf{M}_{\hat{F}(\psi)}\mathbf{D}_{-1}(\rho, \lambda_1, \lambda_2)]\big|_{\psi_0}$ *and* $\frac{\partial}{\partial \psi_p} \operatorname{tr}[\mathbf{M}_{\hat{F}(\psi)}\mathbf{W}_2\mathbf{D}_{-1}(\rho, \lambda_1, \lambda_2)]\big|_{\psi_0}$ *are both* $o_p(nT)$, *for* $p = 1, \ldots, k+4$*;*

*(iii)* $\frac{\partial}{\partial \psi_p} \operatorname{tr}[\mathbf{M}_{\hat{F}(\psi)}\mathbf{W}_1\mathbf{D}(\rho, \lambda_1, \lambda_2)]\big|_{\psi_0} = (T-r)\operatorname{tr}(W_1B_{10}^{-1}W_1B_{10}^{-1}) + o_p(nT)$, *for* $p = k+2$, *and* $\frac{\partial}{\partial \psi_p} \operatorname{tr}[\mathbf{M}_{\hat{F}(\psi)}\mathbf{W}_1\mathbf{D}(\rho, \lambda_1, \lambda_2)]\big|_{\psi_0} = o_p(nT)$, *for* $p \neq k+2$.

*(iv)* $\frac{\partial}{\partial \psi_p} \operatorname{tr}[W_3B_3^{-1}(\lambda_3)]\big|_{\psi_0} = \operatorname{tr}(W_3B_{30}^{-1}W_3B_{30}^{-1})$, *for* $p = k+4$ *and* $\frac{\partial}{\partial \psi_p} \operatorname{tr}[W_3B_3^{-1}(\lambda_3)]\big|_{\psi_0} = 0$, *for* $p \neq k+4$.

**Lemma D.3.** *For a random vector* $\mathbf{v}$ *with i.i.d. zero mean entries, whose third and fourth moments are denoted as* $\mu_3$ *and* $\mu_4$*, we can obtain that*

$$
\begin{aligned}
&cov(\Pi_a'\mathbf{v} + \mathbf{v}'\Phi_a\mathbf{v} - \sigma_{v0}^2\operatorname{tr}(\Phi_a), \Pi_b'\mathbf{v} + \mathbf{v}'\Phi_b\mathbf{v} - \sigma_{v0}^2\operatorname{tr}(\Phi_b)) \\
=&\sigma_{v0}^2\Pi_a'\Pi_b + \mu_3[\Pi_a'\mathtt{diag}(\Phi_b) + \Pi_b'\mathtt{diag}(\Phi_a)] + \sigma_{v0}^4\operatorname{tr}(\Phi_a\Phi_b + \Phi_a'\Phi_b) \\
&+ (\mu_4 - 3\sigma_{v0}^4)\mathtt{diag}(\Phi_a)'\mathtt{diag}(\Phi_b)
\end{aligned}
$$

**Lemma D.4.** *Suppose Assumptions A-H hold and $T/n \to 0$. We have,*

*(i)* $\frac{1}{nT}\operatorname{tr}[(M_{F_0} \otimes I_n)\dot{\mathbf{D}}_{-1,\psi_j}] = o_p(1)$ *and* $\frac{1}{nT}\operatorname{tr}[(M_{F_0} \otimes W_2)\dot{\mathbf{D}}_{-1,\psi_j}] = o_p(1)$, *for* $j = k+1, k+2, k+3$*;*

*(ii)* $\frac{1}{nT}\operatorname{tr}[(M_{F_0} \otimes I_n)\dot{\mathbf{D}}_{\psi_j}] = o_p(1)$, *for* $j = k+1, k+3$.

*(iii)* $\frac{1}{nT}\operatorname{tr}[(M_{F_0} \otimes W_1)\dot{\mathbf{D}}_{\lambda_1}] = \frac{T-r}{nT}\operatorname{tr}[W_1B_{10}^{-1}W_1B_{10}^{-1}] + o_p(1)$.

**Lemma D.5.** *Suppose Assumptions A-H hold and $T/n \to 0$.*

*(i)* $\tilde{\Pi}_a = (I_T \otimes M_{\dot{\Gamma}})\Pi_a = \Pi_a - (I_T \otimes P_{\dot{\Gamma}})\Pi_a$, *for* $a = 1, \ldots, 4$,
$\tilde{\Psi}_b = (I_T \otimes M_{\dot{\Gamma}})\Psi_b = \Psi_b - (I_T \otimes P_{\dot{\Gamma}})\Psi_b$, *for* $b = 1, 2, 3$, *and*
$\tilde{\Phi}_c = (I_T \otimes M_{\dot{\Gamma}})\Phi_c = \Phi_c - (I_T \otimes P_{\dot{\Gamma}})\Phi_c$, *for* $c = 1, 2, 3$*;*

*(ii)* $\|\mathtt{diag}[(I_T \otimes P_{\dot{\Gamma}})\Phi_a]\|_{\infty} = O_p(n^{-1/2})$, *and*
$\|\mathtt{diag}[(I_T \otimes P_{\dot{\Gamma}})\Phi_a]\| = O_p(\sqrt{T/n})$, *for* $a = 1, 2, 3$*;*

*(iii)* $\tilde{\Pi}_a'\tilde{\Pi}_b = \Pi_a'\Pi_b - \Pi_a'(I_T \otimes P_{\dot{\Gamma}})\Pi_b$,
$\tilde{\Pi}_a'\mathtt{diag}(\tilde{\Phi}_b) = \Pi_a'\mathtt{diag}(\Phi_b) - \Pi_a'(I_T \otimes P_{\dot{\Gamma}})\mathtt{diag}(\Phi_b) + o_p(nT)$, *and*
$\mathtt{diag}(\tilde{\Phi}_a)'\mathtt{diag}(\tilde{\Phi}_b) = diag(\Phi_a)'\mathtt{diag}(\Phi_b) + o_p(nT)$*;* $\mathrm{tr}(\tilde{\Phi}_a\tilde{\Phi}_b) = \mathrm{tr}(\Phi_a\Phi_b) + o_p(nT)$*;*

*(iv)* $\frac{\sigma_{v0}^2}{nT}\tilde{\Pi}_{1,j}'\tilde{\Pi}_{1,q} = \frac{1}{nT\sigma_{v0}^2}\,\mathrm{tr}(\dot{X}_j'M_{\dot{\Gamma}}\dot{X}_qM_{F_0})$*, for* $j, q \leq k$,
$\frac{\sigma_{v0}^2}{nT}\tilde{\Pi}_{1,j}'(\tilde{\Pi}_{a+1} + \tilde{\Psi}_a\mathbf{y}_0) = \frac{1}{nT\sigma_{v0}^2}\,\mathrm{tr}(\dot{X}_j'M_{\dot{\Gamma}}\dot{X}_{k+a}M_{F_0}) + o_p(1)$*, for* $j \leq k$*,* $a = 1, 2, 3$,
$\frac{\sigma_{v0}^2}{nT}(\tilde{\Pi}_{b+1} + \tilde{\Psi}_b\mathbf{y}_0)'(\tilde{\Pi}_{a+1} + \tilde{\Psi}_a\mathbf{y}_0) + \frac{\sigma_{v0}^4}{nT}\,\mathrm{tr}(\tilde{\Phi}_a'\tilde{\Phi}_b) = \frac{1}{nT\sigma_{v0}^2}\,\mathrm{tr}(\dot{X}_{k+a}'M_{\dot{\Gamma}}\dot{X}_{k+b}M_{F_0}) + o_p(1)$,
*for* $a, b = 1, 2, 3$*;*

*(v) Let* $\Pi_{4+s}$ *be defined as in Appendix C. We have* $[\Pi_{4+1}, \ldots, \Pi_{4+\phi_s}] = \frac{1}{\sigma_{v0}^2}(M_{F_0}\mathcal{S}_* \otimes \dot{\Gamma}L)$*. In addition, we have* $M_{F_0} = M_{F_0}\mathcal{S}_*(\mathcal{S}_*'M_{F_0}\mathcal{S}_*)^{-1}\mathcal{S}_*'M_{F_0}$.

**Lemma D.6.** *Let* $\mathbb{T}_0 = \frac{1}{nT}F_0\dot{\Gamma}'\dot{\Gamma}F_0'$, $\mathbb{T}_1 = \frac{\sigma_{v0}^2}{T}I_T$, $\mathbb{T}_2 = \frac{1}{nT}[F_0\dot{\Gamma}'\mathbb{V} + \mathbb{V}'\dot{\Gamma}F_0' + (\mathbb{V}'\mathbb{V} - n\sigma_{v0}^2I_T)]$*, and* $\mathbb{T}_3(\psi) = \frac{1}{nT}[Z(\theta)'\Omega^{-1}(\lambda_3)Z(\theta) - Z(\theta_0)'\Omega^{-1}(\lambda_{30})Z(\theta_0)]$*, with* $\epsilon_1 = \sigma_{v0}^2/T$*,* $\epsilon_2 = ||\mathbb{T}_2||_{sp}$ *and* $\epsilon_3(\psi) = ||\mathbb{T}_3(\psi)||_{sp}$*. Suppose* $\epsilon_1 + \epsilon_2 + \epsilon_3(\psi) < C_{min}/2$*, where* $C_{min}$ *is the minimum nonzero eigenvalue of* $\mathbb{T}_0$*. The following statements hold.*

*(i) We have that* $M_{\hat{F}(\psi)} - M_{F_0} = \sum_{j=1}^{\infty} P^{(j)}$*, with the expression given in the proof. In addition,*

$$||\sum_{j=s}^{\infty} P^{(j)}||_{sp} \leq \left(\frac{2e(\epsilon_1 + \epsilon_2 + \epsilon_3(\psi))}{C_{min}}\right)^{s-1} \frac{2e(\epsilon_2 + \epsilon_3(\psi))}{C_{min} - 2e(\epsilon_1 + \epsilon_2 + \epsilon_3(\psi))}.$$

*(ii) Let* $C = \max_j \left\|\frac{\partial \mathbb{T}_3(\psi)}{\partial \psi_j}\ |_{\psi_0}\right\|_{sp}$*. We have* $\frac{P^{(1)}}{\partial \psi_j}\ |_{\psi_0} = -M_{F_0}\frac{\partial \mathbb{T}_3(\psi)}{\partial \psi_j}\ |_{\psi_0}\ S^{(1)} - S^{(1)}\frac{\partial \mathbb{T}_3(\psi)}{\partial \psi_j}\ |_{\psi_0}\ M_{F_0}$ *and*

$$\left\|\sum_{k=2}^{\infty} \frac{P^{(j)}}{\partial \psi_j}\ |_{\psi_0}\right\|_{sp} \leq \frac{4eC(\epsilon_1 + \epsilon_2)}{C_{min}(C_{min} - 2e(\epsilon_1 + \epsilon_2))}.$$

## Proofs of Lemmas D1-D6

**Proof of Lemma D.1:** (i) First, we study $\mathrm{tr}[\dot{X}_j'\dot{\Gamma}F_0'(M_{\tilde{F}} - M_{F_0})]$, and $\mathrm{tr}[\dot{X}_j'\mathbb{V}(M_{\tilde{F}} - M_{F_0})]$.

(i-1) Consider the first term. The decomposition of $M_{\tilde{F}} - M_{F_0}$ in Lemma C.1 leads to

$$\mathrm{tr}[\dot{X}_j'\dot{\Gamma}F_0'(M_{\tilde{F}} - M_{F_0})] = \mathrm{tr}(\dot{X}_j'\dot{\Gamma}F_0'M_{\tilde{F},v}^{(1)}) + \mathrm{tr}(\dot{X}_j'\dot{\Gamma}F_0'M_{\tilde{F},v}^{(2)}) + \mathrm{tr}(\dot{X}_j'\dot{\Gamma}F_0'M_{\tilde{F},v}^{(rem)}),$$

where the first term $\mathrm{tr}(\dot{X}_j'\dot{\Gamma}F_0'M_{\tilde{F},v}^{(1)}) = -\,\mathrm{tr}(\dot{X}_j'P_{\dot{\Gamma}}\mathbb{V}M_{F_0})$, and the third term $\mathrm{tr}(\dot{X}_j'\dot{\Gamma}F_0'M_{\tilde{F},v}^{(rem)}) = O_p(n/\sqrt{T}) = o_p(\sqrt{nT})$.

For the second term, plugging the expression of $M_{\tilde{F},v}^{(2)}$ given in Lemma C.1 into $\mathrm{tr}(\dot{X}_j'\dot{\Gamma}F_0'M_{\tilde{F},v}^{(2)})$, we obtain

$$\begin{aligned}\mathrm{tr}(\dot{X}_j'\dot{\Gamma}F_0'M_{\tilde{F},v}^{(2)}) &= -\,\mathrm{tr}[\dot{X}_j'(F_0\dot{\Gamma}')^{\dagger}(\mathbb{V}'\mathbb{V} - n\sigma_{v0}^2I_T)M_{F_0}] \\ &= -\,\mathrm{tr}[(\dot{\Gamma}'\dot{\Gamma})^{-1}(F_0'F_0)^{-1}F_0'(\mathbb{V}'\mathbb{V} - n\sigma_{v0}^2I_T)M_{F_0}\dot{X}_j'\dot{\Gamma}].\end{aligned}$$

We can verify that $F_0'(\mathbb{V}'\mathbb{V} - n\sigma_{v0}^2 I_T)M_{F_0}\dot{X}_j'\dot{\Gamma} = O_P(n^{3/2}T)$ by calculating its second order moment. Then the second term is of order $O_P(\sqrt{n})$. The result terms can be similarly verified to be $o_P(\sqrt{nT})$.

(i-2) Consider the second term $\operatorname{tr}[\dot{X}_j'\mathbb{V}(M_{\tilde{F}} - M_{F_0})]$. Again we use the result in Lemma C.1 to split it into three terms. For the first term $\operatorname{tr}(\dot{X}_j'\mathbb{V}M_{\tilde{F},v}^{(1)})$, we have

$$\operatorname{tr}(\dot{X}_j'\mathbb{V}M_{\tilde{F},v}^{(1)}) = -\operatorname{tr}[\dot{X}_j'\mathbb{V}(\dot{\Gamma}F_0')^\dagger\mathbb{V}M_{F_0}] - \operatorname{tr}[\dot{X}_j'\mathbb{V}M_{F_0}\mathbb{V}'(F_0'\dot{\Gamma})^\dagger] = O_P(T) = o_P(\sqrt{nT}).$$

where we have used $||\dot{X}_j\mathbb{V}||_F = O_P(\sqrt{n}T)$ and $||\dot{\Gamma}'\mathbb{V}||_F = O_P(\sqrt{nT})$. For the last two terms, we also have $\operatorname{tr}[\dot{X}_j'\mathbb{V}(M_{\tilde{F},v}^{(2)} + M_{\tilde{F},v}^{(rem)})] = o_p(\sqrt{nT})$.

Summing up the terms, we have $\operatorname{tr}[\dot{X}_j'(\dot{\Gamma}F_0' + \mathbb{V})(M_{\tilde{F}} - M_{F_0})] = \operatorname{tr}(\dot{X}_j'P_{\dot{\Gamma}}\mathbb{V}M_{F_0}) + o_p(\sqrt{nT})$.

(ii) Recall that $\mathbf{D}_{-1} = \sum_{j=1}^{T-1}(J_T^j \otimes \mathcal{B}_0^{j-1}B_{10}^{-1})$. We can write

$$\begin{aligned}
\operatorname{tr}(\mathbf{M}_{\tilde{F}}\mathbf{D}_{-1}) &= \textstyle\sum_{j=1}^{T-1} \operatorname{tr}(M_{\tilde{F}}J_T^j \otimes \mathcal{B}_0^{j-1}B_{10}^{-1}) \\
&= \operatorname{tr}(\mathbf{M}_{F_0}\mathbf{D}_{-1}) + \textstyle\sum_{j=1}^{T-1} \operatorname{tr}[(M_{\tilde{F}} - M_{F_0})J_T^j \otimes \mathcal{B}_0^{j-1}B_{10}^{-1}] \\
&= \operatorname{tr}(\mathbf{M}_{F_0}\mathbf{D}_{-1}) + A_1 + A_2 + A_3,
\end{aligned}$$

where $\|M_{\tilde{F}} - M_{F_0}\|_{\mathtt{sp}} = o_p(1)$ and

$$\begin{aligned}
A_1 &= \textstyle\sum_{j=1}^{T-1} \operatorname{tr}(M_{\tilde{F},v}^{(1)}J_T^j \otimes \mathcal{B}_0^{j-1}B_{10}^{-1}) = \sum_{j=1}^{T-1} \operatorname{tr}(M_{\tilde{F},v}^{(1)}J_T^j)\operatorname{tr}(\mathcal{B}_0^{j-1}B_{10}^{-1}), \\
A_2 &= \textstyle\sum_{j=1}^{T-1} \operatorname{tr}(M_{\tilde{F},v}^{(2)}J_T^j \otimes \mathcal{B}_0^{j-1}B_{10}^{-1}) = \sum_{j=1}^{T-1} \operatorname{tr}(M_{\tilde{F},v}^{(2)}J_T^j)\operatorname{tr}(\mathcal{B}_0^{j-1}B_{10}^{-1}), \\
A_3 &= \textstyle\sum_{j=1}^{T-1} \operatorname{tr}(M_{\tilde{F},v}^{(rem)}J_T^j \otimes \mathcal{B}_0^{j-1}B_{10}^{-1}) = \sum_{j=1}^{T-1} \operatorname{tr}(M_{\tilde{F},v}^{(rem)}J_T^j)\operatorname{tr}(\mathcal{B}_0^{j-1}B_{10}^{-1}).
\end{aligned}$$

We have $|\operatorname{tr}(\mathcal{B}_0^{j-1}B_{10}^{-1})| \le \|\mathcal{B}_0^{j-2}B_{10}^{-1}\|\|\mathcal{B}_0\| < C\rho^j\|B_{10}\| \cdot \|\mathcal{B}_0\| = \rho^j O_p(n)$. Now,

$$\operatorname{tr}(M_{\tilde{F},v}^{(1)}J_T^j) = -\operatorname{tr}[M_{F_0}\mathbb{V}'(F_0\dot{\Gamma}')^\dagger J_T^j] - \operatorname{tr}[(\dot{\Gamma}F_0')^\dagger\mathbb{V}M_{F_0}J_T^j] = O_p(1/\sqrt{n}),$$

uniformly over $j$. Summing up, we have $A_1 = o_p(\sqrt{nT})$. By similar arguments, we can show $A_2 + A_3 = o_p(\sqrt{nT})$. We can also show that $\operatorname{tr}[(M_{F_0} \otimes P_{\dot{\Gamma}})\mathbf{D}_{-1}] = o_p(\sqrt{nT})$. Therefore, we have $\operatorname{tr}(\mathbf{M}_{\tilde{F}}\mathbf{D}_{-1}) = \operatorname{tr}[(M_{F_0} \otimes M_{\dot{\Gamma}})\mathbf{D}_{-1}] + o_p(\sqrt{nT})$. The analysis of the other two terms is symmetric and is therefore omitted.

(iii) We prove the first result. Again we use the result in Lemma C.1 to obtain

$$\operatorname{tr}[(\mathbb{V} + \dot{\Gamma}F_0')'(\mathbb{V} + \dot{\Gamma}F_0')(M_{\tilde{F}} - M_{F_0})] = A_1 + A_2 + A_3,$$

where $A_1 = \operatorname{tr}[(\mathbb{V} + \dot{\Gamma}F_0')'(\mathbb{V} + \dot{\Gamma}F_0')M_{F,v}^{(1)}]$, $A_2 = \operatorname{tr}[(\mathbb{V} + \dot{\Gamma}F_0')'(\mathbb{V} + \dot{\Gamma}F_0')M_{F,v}^{(2)}]$, and

$A_3 = \operatorname{tr}[(\mathbb{V} + \dot{\Gamma}F_0')'(\mathbb{V} + \dot{\Gamma}F_0')M_{F,v}^{(rem)}]$. We show that

$$\begin{aligned} A_1 =& -2\operatorname{tr}[(\mathbb{V} + \dot{\Gamma}F_0')'\mathbb{V}M_{F_0}\mathbb{V}'(F_0\dot{\Gamma})^{\dagger}] \\ =& -2\operatorname{tr}[\mathbb{V}'\mathbb{V}M_{F_0}\mathbb{V}'\dot{\Gamma}(\dot{\Gamma}'\dot{\Gamma})^{-1}(F_0'F_0)^{-1}F_0'] - 2\operatorname{tr}(P_{\dot{\Gamma}}\mathbb{V}M_{F_0}\mathbb{V}') \\ =& O_p(\sqrt{n} + T/n) = o_p(\sqrt{nT}); \\ A_2 =& -2\operatorname{tr}[\mathbb{V}'\mathbb{V}M_{F_0}(\mathbb{V}'\mathbb{V} - n\sigma_{v0}^2 I_T)\dot{\Gamma}(\dot{\Gamma}'\dot{\Gamma})^{-1}(F_0'F_0)^{-1}(\dot{\Gamma}'\dot{\Gamma})^{-1}\dot{\Gamma}'] \\ & -2\operatorname{tr}[(\dot{\Gamma}F_0')^{\dagger}\mathbb{V}M_{F_0}(\mathbb{V}'\mathbb{V} - n\sigma_{v0}^2 I_T)] \\ =& o_p(\sqrt{nT}); \end{aligned}$$

and $A_3 = o_p(\sqrt{nT})$. The first result thus follows. The second result is proved similarly. ■

**Proof of Lemma D.2:** (i) We use the result of Lemma D.6 (ii). Now we have

$$\left\| \frac{\partial M_{\hat{F}(\psi}}{\psi_j} - \frac{\partial P^{(1)}}{\psi_j} \min_{\psi_0} \right\|_{sp} = \left\| \sum_{k=2}^{\infty} \frac{P^{(k)}}{\partial \psi_j} |_{\psi_0} \right\|_{sp} \leq \frac{4eC(\epsilon_1 + \epsilon_2)}{C_{min}(C_{min} - 2e(\epsilon_1 + \epsilon_2))},$$

where $\epsilon_1 + \epsilon_2 = O_P(T^{-1} + n^{-1/2})$. It suffices to consider $\frac{P^{(1)}}{\partial \psi_j} |_{\psi_0} =:$

$$-M_{F_0}\frac{\partial \mathbb{T}_3(\psi)}{\partial \psi_j} |_{\psi_0} S^{(1)} - S^{(1)}\frac{\partial \mathbb{T}_3(\psi)}{\partial \psi_j} |_{\psi_0} M_{F_0},$$

where $|| \frac{\partial \mathbb{T}_3(\psi)}{\partial \psi_j} |_{\psi_0} + (\dot{X}_j'\dot{\Gamma}F_0' + F_0\dot{\Gamma}'\dot{X}_j)/(nT)||_{sp} = o_p(1)$. Plug this expression into $\frac{P^{(1)}}{\partial \psi_j} |_{\psi_0}$ and make some rearrangement, we can obtain the desired results.

(ii) Note $\mathbf{M}_{\hat{F}(\psi)}\mathbf{D}_{-1}(\rho, \lambda_1, \lambda_2) = \sum_{j=1}^{T-1} M_{\hat{F}(\psi)}J_T^j \otimes \mathcal{B}^{j-1}(\rho, \lambda_1, \lambda_2)B_1^{-1}(\lambda_1)$, then we have

$$\begin{aligned} \frac{\partial}{\partial \psi_p} \operatorname{tr}[\mathbf{M}_{\hat{F}(\psi)}\mathbf{D}_{-1}(\rho, \lambda_1, \lambda_2)]\big|_{\psi_0} &= \textstyle\sum_{j=1}^{T-1} \frac{\partial}{\partial \psi_p}\big\{\operatorname{tr}[M_{\hat{F}(\psi)}J_T^j]\operatorname{tr}[\mathcal{B}^{j-1}(\rho, \lambda_1, \lambda_2)B_1^{-1}(\lambda_1)]\big\}\big|_{\psi_0} \\ &= \textstyle\sum_{j=1}^{T-1} \frac{\partial}{\partial \psi_p}\operatorname{tr}(M_{\hat{F}(\psi)}J_T^j)\big|_{\psi_0}\operatorname{tr}(\mathcal{B}_0^{j-1}B_{10}^{-1}) \\ &\quad + \textstyle\sum_{j=1}^{T-1}\operatorname{tr}(M_{\hat{F}(\psi_0)}J_T^j)\frac{\partial}{\partial \psi_p}\operatorname{tr}[\mathcal{B}^{j-1}(\rho, \lambda_1, \lambda_2)B_1^{-1}(\lambda_1)]\big|_{\psi_0} \\ &\equiv A + B. \end{aligned}$$

For term $A$, we derive the case for $p = 1, \ldots, k+3$, we have,

$$\frac{\partial}{\partial \psi_p}\operatorname{tr}(M_{\hat{F}(\psi)}J_T^j)\big|_{\psi_0} = -\operatorname{tr}[M_{F_0}\dot{X}_k'(F_0\dot{\Gamma}')^{\dagger}J_T^j] - \operatorname{tr}[(\dot{\Gamma}F_0')^{\dagger}\dot{X}_k M_{F_0}J_T^j] = O_p(1)$$

uniformly over $j$, and $|\operatorname{tr}(\mathcal{B}_0^{j-1}B_{10}^{-1})| < \bar{\rho}^j O_p(n)$. It follows that $A = O_p(n)$. For term $B$, first note that $B = 0$ for $p = 1, \ldots, k$ and $p = k+4$. For $p = k+1, k+2$, and$k+3$, we can show that $\frac{\partial}{\partial \psi_p}\operatorname{tr}[\mathcal{B}^{j-1}(\rho, \lambda_1, \lambda_2)B_1^{-1}(\lambda_1)]\big|_{\psi_0} = \bar{\rho}^j O_p(n)$. For $\operatorname{tr}(M_{\tilde{F}}J_T^j)$, we have,

$$|\operatorname{tr}(M_{\tilde{F}}J_T^j)| = |\operatorname{tr}(P_{\tilde{F}}J_T^j)| \leq \|\tilde{F}'J_T^j\tilde{F}\| \cdot \|(\tilde{F}'\tilde{F})^{-1}\| = O_p(1), \text{ for any } j > 0.$$

The first equality is by the fact that $\operatorname{tr}(J_T^j) = 0$. Hence, we can conclude that $B = O_p(n)$.

(iii) Recall the representation of $\mathbf{D}(\rho, \lambda_1, \lambda_2)$ given at the beginning of Appendix D.2. We have,

$$\begin{aligned}&\operatorname{tr}[\mathbf{M}_{\hat{F}(\psi)}\mathbf{W}_1\mathbf{D}(\rho,\lambda_1,\lambda_2)] = \textstyle\sum_{j=0}^{T-1}\operatorname{tr}(M_{\hat{F}(\psi)}J_T^j)\operatorname{tr}[W_1\mathcal{B}^{j-1}(\rho,\lambda_1,\lambda_2)B_1^{-1}(\lambda_1)]\\ =& (T-r)\operatorname{tr}[W_1B_1^{-1}(\lambda_1)] + \textstyle\sum_{j=1}^{T-1}\operatorname{tr}(M_{\hat{F}(\psi)}J_T^j)\operatorname{tr}[W_1\mathcal{B}^{j-1}(\rho,\lambda_1,\lambda_2)B_1^{-1}(\lambda_1)].\end{aligned}$$

The analysis of the second term is similar to part (ii). For the first term is a function of $\lambda_1$ only and its derivative equals $(T-r)\operatorname{tr}[W_1B_1^{-1}(\lambda_1)W_1B_1^{-1}(\lambda_1)]$. ■

**Proof of Lemma D.3**: This lemma can be proved by direct derivation. ■

**Proof of Lemma D.4:** The proof is similar to the proof of Lemma D.2. ■

**Proof of Lemma D.5:** (i) The results can be verified by direct calculation.
(ii) We take $\mathtt{diag}[(I_T \otimes P_{\dot{\Gamma}})\Phi_1]$ as an example. We have,

$$\begin{aligned}&(M_{F_0}\otimes P_{\dot{\Gamma}}B_{30})\mathbf{D}_{-1}\mathbf{B}_{30}^{-1}/\sigma_{v0}^2\\ =&\tfrac{1}{\sigma_{v0}^2}(M_{F_0}\otimes P_{\dot{\Gamma}})\begin{pmatrix}0 & 0 & \cdots & 0 & 0\\ B_{30} & 0 & \cdots & 0 & 0\\ B_{30}\mathcal{B}_0 & B_{30} & \cdots & 0 & 0\\ \vdots & \vdots & \ddots & \vdots & \vdots\\ B_{30}\mathcal{B}_0^{T-2} & B_{30}\mathcal{B}_0^{T-3} & \cdots & B_{30} & 0\end{pmatrix}(I_T\otimes B_{10}^{-1}B_{30}^{-1})I_{nT}.\end{aligned}$$

Its $j$th diagonal elements can be thought of product of $j$th row of $M_{F_0}\otimes P_{\dot{\Gamma}}$, the middle part matrix and $j$th column of $I_{nT}$. Rows of $M_{F_0}\otimes P_{\dot{\Gamma}}$ are uniformly $O_p(n^{-1/2})$ in Euclidean norm and each column of $I_{nT}$ is a unit vector. The middle part matrix is of spectrum norm $O_p(1)$. It follows that the diagonal elements are uniformly $O_p(n^{-1/2})$.
(iii) The results can be verified easily. The derivation is similar to that of part (ii).
(iv) The first two results can be verified easily. We prove the third result. Note that $(M_{F_0}\otimes M_{\dot{\Gamma}})\mathtt{vec}(\dot{X}_{k+a})/\sigma_{v0}^2 = \tilde{\Pi}_{1+a} + \tilde{\Psi}_a\mathbf{y}_0 + \tilde{\Phi}_a\mathbf{v}$. We can show

$$\begin{aligned}&\tfrac{1}{nT\sigma_{v0}^2}\operatorname{tr}[\dot{X}_{k+a}'M_{\dot{\Gamma}}\dot{X}_{k+b}M_{F_0}]\\ =&\tfrac{\sigma_{v0}^2}{nT}(\tilde{\Pi}_{b+1}+\tilde{\Psi}_b\mathbf{y}_0)'(\tilde{\Pi}_{a+1}+\tilde{\Psi}_a\mathbf{y}_0) + \tfrac{\sigma_{v0}^2}{nT}\mathbf{v}'\tilde{\Phi}_a'\tilde{\Phi}_b\mathbf{v} + \tfrac{\sigma_{v0}^2}{nT}(\tilde{\Pi}_{b+1}+\tilde{\Psi}_b\mathbf{y}_0)'\tilde{\Phi}_a\mathbf{v}\\ &+\tfrac{\sigma_{v0}^2}{nT}(\tilde{\Pi}_{a+1}+\tilde{\Psi}_a\mathbf{y}_0)'\tilde{\Phi}_b\mathbf{v}\\ =&\tfrac{\sigma_{v0}^2}{nT}(\tilde{\Pi}_{b+1}+\tilde{\Psi}_b\mathbf{y}_0)'(\tilde{\Pi}_{a+1}+\tilde{\Psi}_a\mathbf{y}_0) + \tfrac{\sigma_{v0}^2}{nT}\mathbf{v}'\tilde{\Phi}_a'\tilde{\Phi}_b\mathbf{v} + o_p(1).\end{aligned}$$

One can verify that $\frac{\sigma_{v0}^2}{nT}\mathbf{v}'\tilde{\Phi}_a'\tilde{\Phi}_b\mathbf{v} = \frac{\sigma_{v0}^4}{nT}\operatorname{tr}[\tilde{\Phi}_a'\tilde{\Phi}_b]$. Hence, the desired result follows.
(v) The first result follows from $\Pi_{4+s} = \frac{1}{\sigma_{v0}^2}(M_{F_0}\otimes\dot{\Gamma})\mathtt{vec}(\dot{F}_s')$.

Consider the second result. Recall that $\mathcal{S}_* = [I_{T-r}, 0_{(T-r)\times r}]'$. Write $M_{F_0}$ as a partitioned matrix $M_{F_0} = [M_1, M_{12}; M_{21}, M_2]$, where $M_1$ is $(T-r)\times(T-r)$ and $M_2$ is $r\times r$. We have that $\mathcal{S}_*'M_{F_0}\mathcal{S}_* = M_1$ and $M_{F_0}\mathcal{S}_* = [M_1', M_{21}']'$. It follows that $M_{F_0}\mathcal{S}_*(\mathcal{S}_*'M_{F_0}\mathcal{S}_*)^{-1}\mathcal{S}_*'M_{F_0} =$

$[M_1, M_{12}; M_{21}, M_{21}M_1^{-1}M_{12}]$. Then, it suffices to show $M_{21}M_1^{-1}M_{12} = M_2$.

As $F_0 = [F^{*\prime}, I_r]'$, one can easily see that $M_1 = I_{T-r} - F^*(I_r + F^{*\prime}F^*)^{-1}F^{*\prime}$, $M_{21} = M_{12}' = -(I_r + F^{*\prime}F^*)^{-1}F^{*\prime}$, $M_2 = I_r - (I_r + F^{*\prime}F^*)^{-1}$, and $M_1^{-1} = I_{T-r} + F^*F^{*\prime}$. It follows that

$$\begin{aligned} M_{21}M_1^{-1}M_{12} &= (I_r + F^{*\prime}F^*)^{-1}F^{*\prime}(I_{T-r} + F^*F^{*\prime})F^*(I_r + F^{*\prime}F^*)^{-1} \\ &= (I_r + F^{*\prime}F^*)^{-1}F^{*\prime}F^* = I_r - (I_r + F^{*\prime}F^*)^{-1} = M_2. \end{aligned}$$

■

**Proof of Lemma D.6**: (i) We focus on matrix $\frac{1}{nT}Z(\theta)'\Omega^{-1}(\lambda_3)Z(\theta)$. We can rewrite it as a summation of four components:

$$\frac{1}{nT}Z(\theta)'\Omega^{-1}(\lambda_3)Z(\theta) = \mathbb{T}_0 + \mathbb{T}_1 + \mathbb{T}_2 + \mathbb{T}_3(\psi).$$

Let $\epsilon_1 = \sigma_{v0}^2/T$, $\epsilon_2 = ||\mathbb{T}_2||_{sp}$ and $\epsilon_3(\psi) = ||\mathbb{T}_3(\psi)||_{sp}$. Denote the $r$th eigenvalue of $\mathbb{T}_0$ as $C_{min}$.

We view $\mathbb{T}_0 + \mathbb{T}_1 + \mathbb{T}_2 + \mathbb{T}_3(\psi) \equiv \mathbb{T}_0 + \mathbb{T}^{(1)}$ as a perturbation of $\mathbb{T}_0$. The unperturbed operator is $\mathbb{T}_0$. The projection matrix that corresponds to its zero eigenvalues is $M_{F_0}$. Suppose $\epsilon_1 + \epsilon_2 + \epsilon_3(\psi)$ is smaller than $C_{min}/2$. By Weyl's inequality, the first $r$ eigenvalues are greater than $C_{min}/2$ and the other $T-r$ eigenvalues are bounded above by $C_{min}/2$. The projection matrix that corresponds to the smallest $T-r$ eigenvalues of $\mathbb{T}_0 + x\mathbb{T}^{(1)}$ can be written as

$$P(x) = \sum_{j=0}^{\infty} x^j P^{(j)}$$

where $P^{(0)} = M_{F_0}$, $S^{(0)} = -M_{F_0}$, $S^{(j)} = (S^{(1)})^j$ for $j > 0$,

$$\begin{aligned} S^{(1)} =& NT \cdot F_0(F_0'F_0)^{-1}(\Lambda_0'\Lambda_0)^{-1}(F_0'F_0)^{-1}F_0' \\ P^{(j)} =& (-1)^{j+1} \sum_{k_1+\cdots+k_{p+1}=j,\ k_s\geq 0} S^{(k_1)}\mathbb{T}^{(1)}S^{(k_2)}\cdots S^{(k_j)}\mathbb{T}^{(1)}S^{(k_{j+1})} \end{aligned}$$

The above expansion holds for $|x| \leq \frac{C_{min}}{2||\mathbb{T}^{(1)}||_{sp}}$. For detailed analysis, one can check pages 74-77 of Kato (1995). For $x = 1$ to be in this area, we only need $\epsilon_1 + \epsilon_2 + \epsilon_3(\psi) \leq C_{min}/2$. Thus we have $P(1) = M_{\hat{F}(\psi)}$ and $M_{\hat{F}(\psi)} - M_{F_0} = \sum_{j=1}^{\infty} P^{(j)}$.

Next we further bound $||P^{(j)}||_{sp}$ and its summation. In any $S^{(k_1)}\mathbb{T}^{(1)}S^{(k_2)}\cdots S^{(k_j)}\mathbb{T}^{(1)}S^{(k_{j+1})}$, we can see one $S^{(0)}\mathbb{T}S^{(k)}$ or $S^{(k)}\mathbb{T}S^{(0)}$ and $S^{(0)}\mathbb{T}_1S^{(k)} = 0$. Note that $||S^{(1)}||_{sp} = 1/C_{min}$. We can obtain the following bounds:

$$\begin{aligned} ||S^{(k_1)}\mathbb{T}^{(1)}S^{(k_2)}\cdots S^{(k_j)}\mathbb{T}^{(1)}S^{(k_{j+1})}||_{sp} \leq& \frac{(\epsilon_2 + \epsilon_3(\psi))(\epsilon_1 + \epsilon_2 + \epsilon_3(\psi))^{j-1}}{C_{min}^j} \\ \sum_{k_1+\cdots+k_{p+1}=j,\ k_s\geq 0} S^{(k_1)}1 =& (2j)!/(j!)^2 \leq (2e)^j, \end{aligned}$$

It follows that $||P^{(j)}||_{sp} \leq \frac{2e(\epsilon_2+\epsilon_3(\psi))}{C_{min}} \left(\frac{2e(\epsilon_1+\epsilon_2+\epsilon_3(\psi))}{C_{min}}\right)^{j-1}$ and

$$||\sum_{j=s}^{\infty} P^{(j)}||_{sp} \leq \left(\frac{2e(\epsilon_1+\epsilon_2+\epsilon_3(\psi))}{C_{min}}\right)^{s-1} \frac{2e(\epsilon_2+\epsilon_3(\psi))}{C_{min}-2e(\epsilon_1+\epsilon_2+\epsilon_3(\psi))}.$$

(ii) Consider the derivative of $P^{(j)}$'s. Note that $\frac{\partial \mathbb{T}_3(\psi)}{\partial \psi_k} |_{\psi_0} = -\frac{1}{nT}[\dot{X}_j' B_{30} Z(\theta_0) + Z(\theta_0)' B_{30}' \dot{X}_j]$ and $||\frac{1}{nT}[\dot{X}_j' B_{30} Z(\theta_0) + Z(\theta_0)' B_{30}' \dot{X}_j]||_{sp} \leq C$. For $\frac{P^{(1)}}{\partial \psi_k} |_{\psi_0}$, we have $\frac{P^{(1)}}{\partial \psi_k} |_{\psi_0} = -M_{F_0} \frac{\partial \mathbb{T}_3(\psi)}{\partial \psi_k} |_{\psi_0} S^{(1)} - S^{(1)} \frac{\partial \mathbb{T}_3(\psi)}{\partial \psi_k} |_{\psi_0} M_{F_0}$. For the rest of the terms, we can follow a similar analysis to that of $||P^{(j)}||_{sp}$ and obtain

$$\left\|\sum_{j=2}^{\infty} \frac{P^{(j)}}{\partial \psi_k} |_{\psi_0}\right\|_{sp} \leq \frac{4eC(\epsilon_1+\epsilon_2)}{C_{min}(C_{min}-2e(\epsilon_1+\epsilon_2))}.$$

■

## Appendix E: Details of Variable Constructions

Details of variable constructions in the empirical study in Section 6 are provided below.

Table D2.1 Variable Constructions of Empirical Study

| Variables | Descriptions |
|---|---|
| R& D investment | R& D spending / total asset |
| Size | log(total asset) |
| Leverage | total leverage/total asset |
| ROA | return on assets |
| PPE | value of tangible asset/total asset |
| Growth | $\text{sales}_t - \text{sales}_{t-1}/\text{sales}_t$ |

**Note**:All variables are standardized.