3-2020

Estimation of fixed effects spatial dynamic panel data models with small T and unknown heteroskedasticity

Liyao LI
Singapore Management University, liyao.li.2015@phdecons.smu.edu.sg

Zhenlin YANG
Singapore Management University, zlyang@smu.edu.sg

To appear in Regional Science and Urban Economics

Estimation of Fixed Effects Spatial Dynamic Panel Data Models with Small T and Unknown Heteroskedasticity*

Liyao Li and Zhenlin Yang†

School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903

Emails: liyao.li.2015@phdecons.smu.edu.sg; zlyang@smu.edu.sg

January 12, 2020

Abstract

We consider the estimation and inference of fixed effects (FE) spatial dynamic panel data (SDPD) models under small T and unknown heteroskedasticity by extending the M-estimation strategy for homoskedastic FE-SDPD model of Yang (2018, *Journal of Econometrics*). Unbiased estimating equations are obtained by adjusting the conditional quasi-score functions given the initial observations, leading to M-estimators that are free from the initial conditions and robust against unknown cross-sectional heteroskedasticity. Consistency and asymptotic normality of the proposed M-estimator are established. The standard errors are obtained by representing the estimating equations as sums of martingale differences. Monte Carlo results show that the proposed M-estimators have good finite sample performance. The practical importance and relevance of allowing for heteroskedasticity in the model is illustrated using data on sovereign risk spillover.

Key Words: Adjusted quasi score; Dynamic panels; Fixed effects; Initial-condition; Martingale difference; Short panels; Spatial effects; Unknown heteroskedasticity.

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

1. Introduction

The spatial dynamic panel data (SDPD) models have become over the years more and more popular among the theoretical and applied researchers for being able to capture the dynamic effects as well as the effects of spatial interactions. Much attention has been paid to the SDPD models under large n and large T scenarios; see, e.g., Mutl (2006), Yang et al.

*We thank Co-Editor, Stephen Ross, and the two anonymous referees for their constructive comments that have led to improvements in the paper. Thanks are also due to the participants of the Conference on Frontier of New Economic Geography, Southeast University, China, Nov. 2018, and the Asian Meeting of the Econometrics Society, Xiamen University, China, June 2019, for their helpful comments. Zhenlin Yang gratefully acknowledges the financial support from Singapore Management University under Grant C244/MSS16E003.

†Corresponding Author: 90 Stamford Road, Singapore 178903. Phone: +65-6828-0852; Fax: +65-6828-0833. E-mail addresses: zlyang@smu.edu.sg.

(2006), Yu et al. (2008), Korniotis (2010), Lee and Yu (2014), Shi and Lee (2017), and Bai and Li (2018). Relatively lesser attention has been paid to the SDPD models under large n and small T setup: Elhorst (2010) considered the fixed effects (FE) SDPD model with spatial lag; Su and Yang (2015) studied the quasi maximum likelihood (QML) estimation of the SDPD model with spatial errors and fixed or random effects, where the initial observations are modelled; Yang (2018) proposed a unified M-estimation method for the FE-SDPD model with spatial lag, space-time lag as well as spatial error, which is free from the specification of initial conditions; Kuersteiner and Prucha (2018) considered GMM estimation of a model similar to that in Yang (2018), but allowing endogenous spatial weights, higher-order spatial effects, weakly exogenous covariates, interactive fixed effects, and heteroskedastic errors.

All of these estimators of the SDPD models are obtained under the assumption that the disturbances are homoskedastic, except the QML estimator of Bai and Li (2018) and the GMM estimator of Kuersteiner and Prucha (2018). The former is under large n and large T setup and the latter is under large n and small T setup and hence is most closely related to the model we study in this paper under the alternative M-estimation approach. As it is well known, the GMM method may face the issues of simplicity and efficiency; the majority of empirical microeconometric research involves panel data with a large number of cross-sectional units and a small number of time periods, called *short panels*; and in spatial panels, the homoskedasticity assumption may not hold in many situations as spatial units are often heterogeneous in important characteristics such as size, location, population, number of neighbors, etc. Anselin (1988) identifies that heteroskedasticity can occur due to the idiosyncrasies in the model specification that feeds to the disturbances. Different aggregations of data or mixture of an aggregated and non-aggregated data may also cause the errors to be heteroskedastic. Interactions between spatial units may further complicate the variance structure of the aggregated data.[1] It is therefore of great interest to develop a set of methods that are able to address all these issues associated with the SDPD models.

This paper contributes to the literature by proposing estimation and inference methods for the FE-SDPD model with spatial lag (`SL`), space-time lag (`STL`), and spatial error (`SE`) under large n and small T setup, allowing for the existence of cross-sectional heteroskedasticity (CH) of unknown form in the idiosyncratic errors. We extend the M-estimation strategy for the homoskedastic FE-SDPD model of Yang (2018) to give an M-estimator that is not only free from the specification of initial conditions, but also robust against the unknown CH. For inferences, we adopt the *outer-product-of-martingale-differences* (OPMD) method in Yang (2018) to give a consistent estimator of the variance covariance (VC) matrix of the M-estimator that is also free from the initial conditions and robust against unknown CH.

[1] See Lin and Lee (2010), Kelejian and Prucha (2010), Liu and Yang (2015), Breitung and Wigger (2018), and Taspınar et al. (2919) for more discussions on heteroskedasticity based on spatial cross-sectional models, and Moscone and Tosetti (2011) based on static spatial panel data models.

The likelihood-based estimation of an FE-SDPD model with short panels encounters three sets of *incidental parameters* in the sense of Neyman and Scott (1948), arising from the unobserved individual-specific FE, unspecified initial observations and unknown CH. The individual FE alone causes the direct QML estimation to encounter the incidental parameters bias of Neyman and Scott (1948) due to the fact that the number of parameters increases with sample size. However, with a balanced panel, this problem can easily be dealt with by first-differencing (as we do in this paper) or other transformations. The *initial values problem* alone renders the QML estimation to be inconsistent when T is small or asymptotically biased when T is large due to the fact that the conditional likelihood given the initial differences is used and the information contained in the initial differences about the structural parameters is thus ignored. In a dynamic panel, the distribution of the initial observations depends on the *unobservables*, i.e., the process starting positions and the past values of the time-varying variables, and hence is unspecified. Conditional on the initial differences is equivalent to ignoring the information contained in them about the structural parameters. In the case of a fixed T, the proportion of such ignorance is fixed, and hence consistency of parameter estimation cannot be achieved. In this case, one may model the initial observations as in Hsiao et al. (2002) and Su and Yang (2015) to give a full likelihood, but this approach depends on a linear relation between response and regressors and hence does not apply to the SDPD models with spatial lag terms as pointed out by Yang (2018). When T is large, a common practice is to estimate the structural parameters based on the conditional quasi likelihood, and then perform bias correction on the conditional QML estimators to eliminate the asymptotic bias as in Hahn and Kuersteiner (2002) and Yu et al. (2008). The proposed M-estimation strategy, however, works directly on the conditional quasi scores (CQS), making adjustments on the CQS functions to give a set of unbiased and consistent *estimating functions*. As the root-cause of inconsistency or asymptotic bias of the conditional QML estimators is the inconsistency or asymptotic bias of the CQS functions, such adjustments would eliminate the inconsistency or asymptotic bias of an estimator from its 'root'. Finally, the QML estimation of a spatial model is often based on the quasi Gaussian likelihood formulated under the assumption that the errors are homoskedastic. The score components of the spatial parameters are typically linear-quadratic in error vector, which are not robust against heteroskedasticity. As a result, the estimation of the structural parameters cannot be consistent, even if the initial values problem has been resolved as in Yang (2018). To take care of unknown CH on top of initial values problem, the method of *centering* of Yang (2018) no longer applies, and we thus proposed an entirely different way of adjusting the CQS functions to obtain a set of estimating functions that are not only free from the initial values but also robust against unknown CH.

Consistency and asymptotic normality of the proposed M-estimator are established. The consistency of the VC matrix estimator is also proved. Our M-estimation method for estimating model parameters remains valid if T goes large with n, but the OPMD method for VC matrix estimation does not. In this case, the usual plug-in method based on the conditional variance of the adjusted quasi score functions, given the initial differences, can be used. Our proposed M-estimation strategy is likelihood based, and hence is simpler and potentially more efficient than the corresponding GMM method. Monte Carlo results show that the proposed M-estimators have good finite sample performance, and that it is more efficient than the GMM estimator of Kuersteiner and Prucha (2018).

The practical importance and relevance of allowing for heteroskedasticity are illustrated by investigating sovereign risk spillovers among 51 countries over the periods 2007–2012. Based on a fitted SDPD model with `SL` and `STL`, the CH-robust estimation shows a much larger (positive) dynamic effect in sovereign risk than the regular M-estimation, with the estimated dynamic parameter being about three times larger. The CH-robust estimation results show positive and significant spillovers of the sovereign risk among countries through different channels, but the regular M-estimates of the spatial parameters are smaller and insignificant under certain risk transmission channels. Based on another fitted SDPD model with only `SE`, we also find positive and significant dynamic and spatial effects of sovereign risk by allowing for heteroskedasticity. However, we find that, under homoskedasticity assumption, neither dynamic nor spatial effect is significant through any channel.[2]

The rest of paper is as follows. Section 2 introduces the FE-SDPD model with small T and unknown heteroskedasticity and presents the conditional QML estimation of it. Section 3 introduces the heteroskedasticity robust M-estimation for the model, studies the asymptotic properties of the proposed estimators, and presents the OPMD estimator of VC matrix. Monte Carlo results are presented in Section 4. Section 5 empirically examines the sovereign risk spillovers. Section 6 concludes the paper. Technical proofs are collected in Appendix.

2. Model and Conditional QML Estimation

Consider the following general spatial dynamic panel data (SDPD) model with `SL`, `STL` and `SE` effects or in short `STLE` effects:

$$\begin{aligned} y_t &= \rho y_{t-1} + \lambda_1 W_1 y_t + \lambda_2 W_2 y_{t-1} + X_t\beta + Z\gamma + \mu + \alpha_t \mathbf{1}_n + u_t, \\ u_t &= \lambda_3 W_3 u_t + v_t, \quad t = 1, 2, \ldots, T, \end{aligned} \tag{2.1}$$

where $y_t = (y_{1t}, y_{2t}, \ldots, y_{nt})'$ is an $n \times 1$ vector of response variables, $\{X_t\}$ are $n \times p$ matrices of time-varying exogenous regressors, Z is an $n \times q$ matrix of time-invariant exogenous variables, μ is an $n \times 1$ vector of unobserved individual-specific effects, α_t are time-specific effects with $\mathbf{1}_n$ being an $n \times 1$ vector of ones, and $v_t = (v_{1t}, v_{2t}, \ldots, v_{nt})'$ is an $n \times 1$ vector of idiosyncratic errors with its elements $\{v_{it}\}$ being independent and identically distributed (iid) across t for

[2] Matlab codes for running these applications are available at `http://www.mysmu.edu/faculty/zlyang/`.

each i, and independent but not necessarily identically distribute (*inid*) across i for each t such that $\mathrm{E}(v_{it}) = 0$ and $\mathrm{Var}(v_{it}) = \sigma_v^2 h_{n,i}$, $i = 1, \ldots, n$, where $h_{n,i} > 0$ and $\frac{1}{n}\sum_{i=1}^{n} h_{n,i} = 1$. Note that σ_v^2 is the average of $\mathrm{Var}(v_{it})$, which can be consistently estimated along with the other model parameters. The scalar parameter ρ characterizes the dynamic effect, λ_1 the spatial lag effect, λ_2 the space-time effect, and λ_3 the spatial error effect, β and γ are the usual regression coefficients, $W_r, r = 1, 2, 3$, are the given $n \times n$ spatial weight matrices.[3]

When μ is considered as fixed effects in the sense that it can be correlated with the time-varying regressors in an arbitrary manner, it is treated as a vector of parameters. As we assume n is large and T is small and fixed, we eliminate μ by taking first-difference in (2.1) to avoid the incidental parameters problem,

$$\begin{aligned} \Delta y_t &= \rho \Delta y_{t-1} + \lambda_1 W_1 \Delta y_t + \lambda_2 W_2 \Delta y_{t-1} + \Delta X_t \beta + \Delta\alpha_t \mathbf{1}_n + \Delta u_t, \\ \Delta u_t &= \lambda_3 W_3 \Delta u_t + \Delta v_t, \quad t = 2, 3, \ldots, T. \end{aligned} \tag{2.2}$$

We note that the time-invariant variables Z is also differenced away. The parameters $\{\alpha_t\}$ or $\{\Delta\alpha_t\}$ are also considered as fixed effects. However, as T is fixed, they can be consistently estimated along with the other model parameters. Define $B_r \equiv B_r(\lambda_r) = I_n - \lambda_r W_r, r = 1, 3$ and $B_2 \equiv B_2(\rho, \lambda_2) = \rho I_n + \lambda_2 W_2$. Model (2.2) has reduced form:

$$\Delta y_t = B_1^{-1} B_2 \Delta y_{t-1} + B_1^{-1}(\Delta X_t \beta + \Delta\alpha_t 1_n) + B_1^{-1} B_3^{-1} \Delta v_t, \;\; t = 2, \ldots, T, \tag{2.3}$$

Let $\Delta Y = \{\Delta y_2', \ldots, \Delta y_T'\}'$, $\Delta Y_{-1} = \{\Delta y_1', \ldots, \Delta y_{T-1}'\}'$, and $\Delta X = \{\Delta X_2', \ldots, \Delta X_T'\}'$. Define $D = (I_{T-2} \otimes 1_n',\; 0_{(T-2)} 0_n')'$ where 0_m is an $m \times 1$ vector of zeros, $\Delta\mathbf{X} = (\mathbf{1}_{n(T-1)}, D, \Delta X)$, $\Delta\mathbf{v} = \{\Delta v_2', \ldots, \Delta v_T'\}'$, $\Delta\mathbf{u} = \{\Delta u_2', \ldots, \Delta u_T'\}'$, $\mathbf{W}_r = I_{T-1} \otimes W_r$, and $\mathbf{B}_r = I_{T-1} \otimes B_r, r = 1, 2, 3$, where $\otimes$ denotes the Kronecker product and I_k an $k \times k$ identity matrix. The reduced form (2.3) can be written in matrix form:

$$\Delta Y = \mathbf{B}_1^{-1} \mathbf{B}_2 \Delta Y_{-1} + \mathbf{B}_1^{-1} \Delta\mathbf{X}\boldsymbol{\beta} + \mathbf{B}_1^{-1} \mathbf{B}_3^{-1} \Delta\mathbf{v}, \tag{2.4}$$

where $\boldsymbol{\beta} = (\breve{\alpha}', \beta')'$, and $\breve{\alpha} = (\Delta\alpha_T, \Delta\alpha_2 - \Delta\alpha_T, \ldots, \Delta\alpha_{T-1} - \Delta\alpha_T)'$.

Let $\mathcal{H} = \mathtt{diag}(h_{n,1}, \ldots, h_{n,n})$, where $\mathtt{diag}(\cdot)$ forms a diagonal matrix based on the given the elements or based on the diagonal elements of a given matrix. It is easy to see that

$$\mathrm{Var}(\Delta\mathbf{u}) = \sigma_v^2 [C \otimes (B_3^{-1} \mathcal{H} B_3'^{-1})],$$

where C is a $(T-1) \times (T-1)$ constant matrix of the form,

$$C = \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 & 0 & 0 \\ -1 & 2 & -1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 2 & -1 \\ 0 & 0 & 0 & \cdots & 0 & -1 & 2 \end{pmatrix}. \tag{2.5}$$

[3]Spatial lags of time-varying regressors or spatial Durbin effects (Halleck Vega and Elhorst, 2015) can be added in the model, which are simply treated as additional exogenous regressors without additional technical complications. On the related parameter identification issue, see Elhorst (2012) and Lee and Yu (2016).

Under homoskedasticity, $\mathcal{H}$ reduces to I_n and the variance-covariance (VC) matrix of $\Delta\mathbf{u}$ becomes $\text{Var}(\Delta\mathbf{u}) = \sigma_v^2[C \otimes (B_3'B_3)^{-1}] \equiv \sigma_v^2\Omega$. Denote $\psi = (\boldsymbol{\beta}', \sigma_v^2, \rho, \lambda')'$ and $\lambda = (\lambda_1, \lambda_2, \lambda_3)'$. The conditional quasi-Gaussian loglikelihood of ψ in terms of $\Delta y_2, \dots, \Delta y_T$ treating Δy_1 as exogenous and v_{it} as normally distributed and homoskedastic is, ignoring the constant term,

$$\ell_{\texttt{STLE}}(\psi) = -\tfrac{n(T-1)}{2}\log(\sigma_v^2) - \tfrac{1}{2}\log|\Omega| + \log|\mathbf{B}_1| - \tfrac{1}{2\sigma_v^2}\Delta\mathbf{u}(\theta)'\Omega^{-1}\Delta\mathbf{u}(\theta), \tag{2.6}$$

where $\theta = (\boldsymbol{\beta}', \rho, \lambda_1, \lambda_2)'$, $\Delta\mathbf{u}(\theta) = \mathbf{B}_1\Delta Y - \mathbf{B}_2\Delta Y_{-1} - \Delta X\boldsymbol{\beta}$, and $|\cdot|$ denotes the determinant of a square matrix. Maximizing $\ell_{\texttt{STLE}}(\psi)$ gives the conditional QML (CQML) estimator $\hat{\psi}_c$ of ψ. It is well known that the QML estimation of a dynamic panel data model with short panels faces the **initial values problem**: Δy_1 is not exogenous but is treated so, and therefore $\ell_{\texttt{STLE}}(\psi)$ cannot be a genuine loglikelihood function even if v_{it} are homoskedastic and normal. Treating Δy_1 as exogenous ignores useful information (about ψ) contained in Δy_1, when T is fixed the degree of such ignorance is unchanged as n goes large. Hence, the CQML method cannot give a consistent estimate of ψ. When T is also large, ignoring the information contained in Δy_1 is asymptotically negligible, and the CQML estimator can be consistent. However, such consistency may not hold under unknown heteroskedasticity. Assuming homoskedasticity, Yang (2018) proposed an initial-condition free approach to consistently estimate the model by adjusting the quasi score function. In this paper, we extend the idea of Yang (2018) to allow for cross-sectional heteroskedasticity of unknown forms.

3. M-estimation of FE-SDPD Model with Heteroskedasticity

3.1. The Robust M-estimator

Recall $\psi = (\boldsymbol{\beta}', \sigma_v^2, \rho, \lambda')'$ where $\lambda' = (\lambda_1, \lambda_2, \lambda_3)$. Consider the conditional quasi score (CQS) function of ψ, $S_{\texttt{STLE}}(\psi) = \frac{\partial}{\partial\psi}\ell_{\texttt{STLE}}(\psi)$, where $\ell_{\texttt{STLE}}(\psi)$ is given in (2.6),

$$S_{\texttt{STLE}}(\psi) = \begin{cases} \frac{1}{\sigma_v^2}\Delta X'\Omega^{-1}\Delta\mathbf{u}(\theta), \\ \frac{1}{2\sigma_v^4}\Delta\mathbf{u}(\theta)'\Omega^{-1}\Delta\mathbf{u}(\theta) - \frac{n(T-1)}{2\sigma_v^2}, \\ \frac{1}{\sigma_v^2}\Delta\mathbf{u}(\theta)'\Omega^{-1}\Delta Y_{-1}, \\ \frac{1}{\sigma_v^2}\Delta\mathbf{u}(\theta)'\Omega^{-1}\mathbf{W}_1\Delta Y - \text{tr}(\mathbf{B}_1^{-1}\mathbf{W}_1), \\ \frac{1}{\sigma_v^2}\Delta\mathbf{u}(\theta)'\Omega^{-1}\mathbf{W}_2\Delta Y_{-1}, \\ \frac{1}{2\sigma_v^2}\Delta\mathbf{u}(\theta)'(C^{-1}\otimes\mathcal{A})\Delta\mathbf{u}(\theta) - (T-1)\text{tr}(W_3B_3^{-1}), \end{cases} \tag{3.1}$$

where $\mathcal{A} = W_3'B_3 + B_3'W_3$, and $\text{tr}(\cdot)$ is the trace of a square matrix. A necessary condition for an extremum estimator such as QMLE to be consistent is: $\text{plim}_{n\to\infty}\frac{1}{nT}S_{\texttt{STLE}}(\psi_0) = 0$ at the true parameter ψ_0 (van der Vaart, 1998). This is always the case for the $\boldsymbol{\beta}$ and σ_v^2 components of the score functions whether or not the errors are homoskedastic or the initial condition

Δy_1 is exogenous, but may not be the case for the ρ and λ components. We first derive the ρ and λ components of $\mathrm{E}[S_{\texttt{STLE}}(\psi_0)]$ under unknown heteroskedasticity $\mathcal{H}$ and show that their limits (upon dividing by nT) are generally not zero but free from the initial conditions. Then based on these mean expressions, we find the adjustments to the quasi-score functions so that the *adjusted quasi score function* $S^*_{\texttt{STLE}}(\psi_0)$ has a mean zero and $\mathrm{plim}_{n\to\infty}\frac{1}{nT}S^*_{\texttt{STLE}}(\psi_0) = 0$.

Denote a parametric quantity evaluated at the true parameter values, ψ_0, by adding a subscript 0, e.g., $B_{10} \equiv B_1$, $\Omega_0 \equiv \Omega$. The usual expectation, variance and covariance operators, 'E', 'Var' and 'Cov', correspond to the true parameter values. We do not differentiate $\mathcal{H}$ in true and general values as inferences concern only ψ_0. As in Yang (2018), we have the following very minimum requirements on the process at and before time 0.

Assumption A: *Under Model* (2.1)*, (i) the processes started m periods before the start of data collection, the* 0*th period, and (ii) if $m \geq 1$, Δy_0 is independent of future errors $\{v_t, t \geq 1\}$; if $m = 0$, y_0 is independent of future errors $\{v_t, t \geq 1\}$.*

Lemma 3.1. *Suppose Assumption A holds. Assume further that, for $i = 1, \ldots, n$ and $t = 0, 1, \ldots, T$, (i) the idiosyncratic errors $\{v_{it}\}$ are iid across t and inid across i with mean 0 and variance $\sigma^2_{v0}h_{n,i}$, where $h_{n,i} > 0$ and $\frac{1}{n}\sum_{i=1}^n h_{n,i} = 1$, (ii) the time-varying regressors X_t are exogenous, and (iii) both B_{10}^{-1} and B_{30}^{-1} exist. We have,*

$$\mathrm{E}(\Delta Y_{-1}\Delta \mathbf{v}') = -\sigma^2_{v0}\mathbf{D}_{-10}\mathbf{B}_{30}^{-1}\mathbf{H}, \tag{3.2}$$

$$\mathrm{E}(\Delta Y\Delta \mathbf{v}') = -\sigma^2_{v0}\mathbf{D}_{0}\mathbf{B}_{30}^{-1}\mathbf{H}, \tag{3.3}$$

where $\mathbf{H} = (I_{T-1} \otimes \mathcal{H})$, and $\mathbf{D}_{-1} \equiv \mathbf{D}_{-1}(\rho, \lambda_1, \lambda_2)$ and $\mathbf{D} \equiv \mathbf{D}(\rho, \lambda_1, \lambda_2)$ defined as:

$$\mathbf{D}_{-1} = \begin{pmatrix} I_n, & 0, & \ldots & 0, & 0 \\ \mathcal{B} - 2I_n, & I_n, & \ldots & 0, & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathcal{D}_{T-4}, & \mathcal{D}_{T-5}, & \ldots & \mathcal{B} - 2I_n, & I_n \end{pmatrix}\mathbf{B}_1^{-1},$$

$$\mathbf{D} = \begin{pmatrix} \mathcal{B} - 2I_n, & I_n, & \ldots & 0 \\ \mathcal{D}_0, & \mathcal{B} - 2I_n, & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{D}_{T-3}, & \mathcal{D}_{T-4}, & \ldots & \mathcal{B} - 2I_n \end{pmatrix}\mathbf{B}_1^{-1},$$

where $\mathcal{D}_t = \mathcal{B}^t(I_n - \mathcal{B})^2$, $\mathcal{B} = B_1^{-1}B_2$, and both $\mathbf{D}_{-1}$ and $\mathbf{D}$ are $n(T-1) \times n(T-1)$.

Lemma 3.1 presents very useful results, which is obtained by recursive backward substitution on the reduced form (2.3). Using these results, we immediately obtain for (3.1):

$$\mathrm{E}(\Delta \mathbf{u}'\Omega_0^{-1}\Delta Y_{-1}) = -\sigma^2_{v0}\mathrm{tr}(\mathbf{H}\mathbf{C}_{b0}\mathbf{D}_{-10}\mathbf{B}_{30}^{-1}), \tag{3.4}$$

$$\mathrm{E}(\Delta \mathbf{u}'\Omega_0^{-1}\mathbf{W}_1\Delta Y) = -\sigma^2_{v0}\mathrm{tr}(\mathbf{H}\mathbf{C}_{b0}\mathbf{W}_1\mathbf{D}_{0}\mathbf{B}_{30}^{-1}), \tag{3.5}$$

$$\mathrm{E}(\Delta \mathbf{u}'\Omega_0^{-1}\mathbf{W}_2\Delta Y_{-1}) = -\sigma^2_{v0}\mathrm{tr}(\mathbf{H}\mathbf{C}_{b0}\mathbf{W}_2\mathbf{D}_{-10}\mathbf{B}_{30}^{-1}), \tag{3.6}$$

$$\mathrm{E}\big[\Delta\mathbf{u}'(C^{-1}\otimes\mathcal{A}_0)\Delta\mathbf{u}\big]=2\sigma_{v0}^2\mathrm{tr}(\mathbf{H}\mathbf{W}_3\mathbf{B}_{30}^{-1}),\tag{3.7}$$

where $\mathbf{C}_b = C^{-1}\otimes B_3$, and the expression $\Omega = C\otimes(B_3'B_3)^{-1}$ defined below (2.5) has been used. These results show that the ρ and λ components of $\mathrm{plim}_{n\to\infty}\frac{1}{nT}S_{\mathtt{STLE}}(\psi_0)$ are not zero even under homoskedasticity, i.e., $\mathcal{H}=I_n$, and hence the CQMLE cannot be consistent even under homoskedasticity as shown by Yang (2018). Under homoskedasticity, the results in (3.4)-(3.7) reduce to functions of common parameters ψ_0 only and hence can directly be used to adjust $S_{\mathtt{STLE}}(\psi_0)$ to give a set of unbiased estimating functions, $S^{\diamond}_{\mathtt{STLE}}(\psi_0) = S_{\mathtt{STLE}}(\psi_0) - E[S_{\mathtt{STLE}}(\psi_0)|\mathcal{H}=I_n]$, leading to the M-estimator of Yang (2018). This M-estimator is consistent under homoskedasticity, but not under unknown heteroskedasticity since the ρ and λ components of $\mathrm{plim}_{n\to\infty}\frac{1}{nT}S^{\diamond}_{\mathtt{STLE}}(\psi_0)$ are not zero in general as shown by the results in (3.4)-(3.7) (see the end of Sec. 3.1 for some formal arguments).

The problem caused by the unknown $\mathcal{H}$ is that the results in (3.4)-(3.7) cannot be directly used to adjust $S_{\mathtt{STLE}}(\psi_0)$. Instead of directly subtracting the expectation, we find quadratic terms (in $\Delta\mathbf{u}$) with expectations being identical to (3.4)-(3.7) (except the sign):

$$\mathrm{E}(\Delta\mathbf{u}'\Omega_0^{-1}\mathbf{C}^{-1}\mathbf{D}_{-10}\Delta\mathbf{u})=\sigma_{v0}^2\mathrm{tr}(\mathbf{H}\mathbf{C}_{b0}\mathbf{D}_{-10}\mathbf{B}_{30}^{-1}),\tag{3.8}$$
$$\mathrm{E}(\Delta\mathbf{u}'\Omega_0^{-1}\mathbf{C}^{-1}\mathbf{W}_1\mathbf{D}_0\Delta\mathbf{u})=\sigma_{v0}^2\mathrm{tr}(\mathbf{H}\mathbf{C}_{b0}\mathbf{W}_1\mathbf{D}_0\mathbf{B}_{30}^{-1}),\tag{3.9}$$
$$\mathrm{E}(\Delta\mathbf{u}'\Omega_0^{-1}\mathbf{C}^{-1}\mathbf{W}_2\mathbf{D}_{-10}\Delta\mathbf{u})=\sigma_{v0}^2\mathrm{tr}(\mathbf{H}\mathbf{C}_{b0}\mathbf{W}_2\mathbf{D}_{-10}\mathbf{B}_{30}^{-1}),\tag{3.10}$$
$$\mathrm{E}\big[2\Delta\mathbf{u}'\big(C^{-1}\otimes B_{30}'\mathtt{diag}(W_3B_{30}^{-1})G_{30}\big)\Delta\mathbf{u}\big]=2\sigma_{v0}^2\mathrm{tr}(\mathbf{H}\mathbf{W}_3\mathbf{B}_{30}^{-1}),\tag{3.11}$$

where $\mathbf{C}^{-1}=C^{-1}\otimes I_n$ and $G_3=\mathrm{diag}(B_3^{-1})^{-1}$. Combining the terms inside the expectations in (3.4)-(3.7) with the corresponding terms inside the expectations in (3.8)-(3.11), we obtain a set of adjusted quasi score (AQS) functions for (ρ,λ), having zero expectations at ψ_0 even if the errors are heteroskedastic. The $\boldsymbol{\beta}$ and σ_v^2 components of $\frac{1}{nT}S_{\mathtt{STLE}}(\psi_0)$ have zero expectation and zero probability limit under $\mathcal{H}$, and hence do not need to be further adjusted. We have a set of unbiased and robust estimating functions for $\psi=(\boldsymbol{\beta}',\sigma_v^2,\rho,\lambda')'$:

$$S^{*}_{\mathtt{STLE}}(\psi)=\begin{cases}\frac{1}{\sigma_v^2}\Delta X'\Omega^{-1}\Delta\mathbf{u}(\theta),\\ \frac{1}{2\sigma_v^4}\Delta\mathbf{u}(\theta)'\Omega^{-1}\Delta\mathbf{u}(\theta)-\frac{n(T-1)}{2\sigma_v^2},\\ \frac{1}{\sigma_v^2}\Delta\mathbf{u}(\theta)'\Omega^{-1}\Delta Y_{-1}+\frac{1}{\sigma_v^2}\Delta\mathbf{u}(\theta)'\mathbf{E}_\rho\Delta\mathbf{u}(\theta),\\ \frac{1}{\sigma_v^2}\Delta\mathbf{u}(\theta)'\Omega^{-1}\mathbf{W}_1\Delta Y+\frac{1}{\sigma_v^2}\Delta\mathbf{u}(\theta)'\mathbf{E}_{\lambda_1}\Delta\mathbf{u}(\theta),\\ \frac{1}{\sigma_v^2}\Delta\mathbf{u}(\theta)'\Omega^{-1}\mathbf{W}_2\Delta Y_{-1}+\frac{1}{\sigma_v^2}\Delta\mathbf{u}(\theta)'\mathbf{E}_{\lambda_2}\Delta\mathbf{u}(\theta),\\ \frac{1}{2\sigma_v^2}\Delta\mathbf{u}(\theta)'\left[C^{-1}\otimes(\mathcal{A}-\mathbf{E}_{\lambda_3})\right]\Delta\mathbf{u}(\theta),\end{cases}\tag{3.12}$$

where $(\mathbf{E}_\rho,\mathbf{E}_{\lambda_1},\mathbf{E}_{\lambda_2})=\Omega^{-1}\mathbf{C}^{-1}(\mathbf{D}_{-1},\ \mathbf{W}_1\mathbf{D},\ \mathbf{W}_2\mathbf{D}_{-1})$, and $\mathbf{E}_{\lambda_3}=2B_3'\mathtt{diag}(W_3B_3^{-1})G_3$.

Solving the estimating equations, $S^{*}_{\mathtt{STLE}}(\psi)=0$, gives the robust M-estimator $\hat\psi_{\mathtt{M}}$. This can be done by first solving the equations for $\boldsymbol{\beta}$ and σ_v^2, given $\delta=(\rho,\lambda')'$, to give

$$\hat{\boldsymbol{\beta}}_{\mathtt{M}}(\delta)=(\Delta X'\Omega^{-1}\Delta X)^{-1}\Delta X'\Omega^{-1}(\mathbf{B}_1\Delta Y-\mathbf{B}_2\Delta Y_{-1}),\tag{3.13}$$

$$\hat{\sigma}^2_{v,\mathtt{M}}(\delta) = \tfrac{1}{n(T-1)}\Delta\hat{\mathbf{u}}(\delta)'\Omega^{-1}\Delta\hat{\mathbf{u}}(\delta), \tag{3.14}$$

where $\Delta\hat{\mathbf{u}}(\delta) = \Delta\mathbf{u}(\hat{\boldsymbol{\beta}}(\delta), \rho, \lambda_1, \lambda_2)$. Then, substituting $\hat{\boldsymbol{\beta}}_{\mathtt{M}}(\delta)$ and $\hat{\sigma}^2_{v,\mathtt{M}}(\delta)$ back into the last four components of the AQS function in (3.12) gives the concentrated AQS functions:

$$S^{*c}_{\mathtt{STLE}}(\delta) = \begin{cases} \frac{1}{\hat{\sigma}^2_{v,\mathtt{M}}(\delta)}\Delta\hat{\mathbf{u}}(\delta)'\Omega^{-1}\Delta Y_{-1} + \frac{1}{\hat{\sigma}^2_{v,\mathtt{M}}(\delta)}\Delta\hat{\mathbf{u}}(\delta)'\mathbf{E}_{\rho}\Delta\hat{\mathbf{u}}(\delta), \\ \frac{1}{\hat{\sigma}^2_{v,\mathtt{M}}(\delta)}\Delta\hat{\mathbf{u}}(\delta)'\Omega^{-1}\mathbf{W}_1\Delta Y + \frac{1}{\hat{\sigma}^2_{v,\mathtt{M}}(\delta)}\Delta\hat{\mathbf{u}}(\delta)'\mathbf{E}_{\lambda_1}\Delta\hat{\mathbf{u}}(\delta), \\ \frac{1}{\hat{\sigma}^2_{v,\mathtt{M}}(\delta)}\Delta\hat{\mathbf{u}}(\delta)'\Omega^{-1}\mathbf{W}_2\Delta Y_{-1} + \frac{1}{\hat{\sigma}^2_{v,\mathtt{M}}(\delta)}\Delta\hat{\mathbf{u}}(\delta)'\mathbf{E}_{\lambda_2}\Delta\hat{\mathbf{u}}(\delta), \\ \frac{1}{2\hat{\sigma}^2_{v,\mathtt{M}}(\delta)}\Delta\hat{\mathbf{u}}(\delta)'\left[C^{-1}\otimes(\mathcal{A}-\mathbf{E}_{\lambda_3})\right]\Delta\hat{\mathbf{u}}(\delta). \end{cases} \tag{3.15}$$

Note that $\hat{\sigma}^2_{v,\mathtt{M}}(\delta)$ can be dropped from the expression for $S^{*c}_{\mathtt{STLE}}(\delta)$. Solving the resulted concentrated estimating equations, $S^{*c}_{\mathtt{STLE}}(\delta) = 0$, we obtain the robust M-estimator $\hat{\delta}_{\mathtt{M}}$ of δ, which give the robust M-estimators of $\boldsymbol{\beta}$ and σ^2_v as $\hat{\boldsymbol{\beta}}_{\mathtt{M}} \equiv \hat{\boldsymbol{\beta}}_{\mathtt{M}}(\hat{\delta}_{\mathtt{M}})$ and $\hat{\sigma}^2_{v,\mathtt{M}} \equiv \hat{\sigma}^2_{v,\mathtt{M}}(\hat{\delta}_{\mathtt{M}})$. Thus, $\hat{\psi}_{\mathtt{M}} = (\hat{\boldsymbol{\beta}}'_{\mathtt{M}}, \hat{\sigma}^2_{v,\mathtt{M}}, \hat{\rho}_{\mathtt{M}}, \hat{\lambda}'_{\mathtt{M}})'$.[4] Many submodels can be easily obtained from Model (2.1) by setting one or two spatial parameters to zero. For example, setting λ_1 and λ_2 to zero, Model (2.1) reduces to an SDPD model with SE only, setting λ_2 and λ_3 to zero, Model (2.1) becomes an SDPD model with SL only, and setting λ_3 to zero, Model (2.1) reduces to an SDPD model with SL and STL. Estimation of these submodels proceeds simply by setting the specific parameters to zeros and excluding the corresponding components in the AQS functions.

An alternative way of adjusting the CQS functions is as follows. The $(\rho, \lambda_1, \lambda_2)$ components in the expected CQS functions given by (3.4)-(3.6) at ψ_0 can be rewritten as:

$$\mathrm{E}(\Delta\mathbf{u}'\Omega_0^{-1}\Delta Y_{-1}) = -\sigma^2_{v0}\mathrm{tr}\big(\mathtt{diag}(\mathcal{G}_{\rho 0})\mathbf{H}\big), \tag{3.16}$$

$$\mathrm{E}(\Delta\mathbf{u}'\Omega_0^{-1}\mathbf{W}_1\Delta Y) = -\sigma^2_{v0}\mathrm{tr}\big(\mathtt{diag}(\mathcal{G}_{\lambda_1 0})\mathbf{H}\big), \tag{3.17}$$

$$\mathrm{E}(\Delta\mathbf{u}'\Omega_0^{-1}\mathbf{W}_2\Delta Y) = -\sigma^2_{v0}\mathrm{tr}\big(\mathtt{diag}(\mathcal{G}_{\lambda_2 0})\mathbf{H}\big), \tag{3.18}$$

where $\mathcal{G}_{\rho} = \mathbf{C}_b\mathbf{D}_{-1}\mathbf{B}_3^{-1}$, $\mathcal{G}_{\lambda_1} = \mathbf{C}_b\mathbf{W}_1\mathbf{D}\mathbf{B}_3^{-1}$, and $\mathcal{G}_{\lambda_2} = \mathbf{C}_b\mathbf{W}_2\mathbf{D}_{-1}\mathbf{B}_3^{-1}$. Let $G_c = \mathrm{diag}^{-1}(\mathbf{C}_b)$, the expectations of the following quadratic terms are the negative of the expectations of the $(\rho, \lambda_1, \lambda_2)$ components of the CQS functions at true parameter values,

$$\mathrm{E}\left[\Delta\mathbf{u}'\left(\mathbf{B}'_{30}\mathtt{diag}(\mathcal{G}_{\rho 0})G_{c0}\right)\Delta\mathbf{u}\right] = \sigma^2_{v0}\mathrm{tr}\left(\mathtt{diag}(\mathcal{G}_{\rho 0})\mathbf{H}\right), \tag{3.19}$$

$$\mathrm{E}\left[\Delta\mathbf{u}'\left(\mathbf{B}'_{30}\mathtt{diag}(\mathcal{G}_{\lambda_1 0})G_{c0}\right)\Delta\mathbf{u}\right] = \sigma^2_{v0}\mathrm{tr}\big(\mathtt{diag}(\mathcal{G}_{\lambda_1 0})\mathbf{H}\big), \tag{3.20}$$

$$\mathrm{E}\left[\Delta\mathbf{u}'\left(\mathbf{B}'_{30}\mathtt{diag}(\mathcal{G}_{\lambda_2 0})G_{c0}\right)\Delta\mathbf{u}\right] = \sigma^2_{v0}\mathrm{tr}\big(\mathtt{diag}(\mathcal{G}_{\lambda_2 0})\mathbf{H}\big). \tag{3.21}$$

[4] It is well known that when a linear regression model $Y = X\beta + \varepsilon$ is estimated under homoskedasticity, the OLS estimator $\hat{\beta}$ is consistent whether ε is homoskedastic or heteroskedastic. The heteroskedasticity only alters the variance of $\hat{\beta}$. This is because $\hat{\beta}$ solves linear equations. In our model, it is still that, if δ_0 is given, $\hat{\boldsymbol{\beta}}(\delta_0)$ defined in (3.13) solves a set of linear equations, and hence is consistent under unknown $\mathcal{H}$. However, the ultimate estimate of $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}} \equiv \hat{\boldsymbol{\beta}}(\hat{\delta})$, of which consistency depends crucially on the consistency of $\hat{\delta}$. Take a simpler model where λ_3 term is dropped for illustration. We have $\delta = (\rho, \lambda_1, \lambda_2)'$ and $\Omega = \mathbf{C}$. From (3.13), $\hat{\boldsymbol{\beta}}(\hat{\delta}) = \hat{\boldsymbol{\beta}}(\delta_0) - (\Delta X'\mathbf{C}^{-1}\Delta X)^{-1}\Delta X'\mathbf{C}^{-1}[\mathbf{W}_1\Delta Y,\ \Delta Y_{-1},\ \mathbf{W}_2\Delta Y_{-1}](\hat{\delta}-\delta_0) = \boldsymbol{\beta}_0 - O(1)(\hat{\delta}-\delta_0) + o_p(1)$. If $\hat{\delta}$ is the CQMLE or the M-estimator of Yang (2018), then $\mathrm{plim}_{n\to\infty}\hat{\delta} \neq \delta_0$ under $\mathcal{H}$, and therefore, $\mathrm{plim}_{n\to\infty}\hat{\boldsymbol{\beta}} \neq \boldsymbol{\beta}_0$. The inconsistency of $\hat{\delta}$ spills over to the estimation of $\boldsymbol{\beta}$, making $\hat{\boldsymbol{\beta}}(\hat{\delta})$ inconsistent as well.

Therefore we obtain an alternative set of AQS functions which take similar forms as (3.12) with $\mathbf{E}_\rho = \mathbf{B}_3'\mathtt{diag}(\mathcal{G}_\rho)G_c$, $\mathbf{E}_{\lambda_1} = \mathbf{B}_3'\mathtt{diag}(\mathcal{G}_{\lambda_1})G_c$, $\mathbf{E}_{\lambda_2} = \mathbf{B}_3'\mathtt{diag}(\mathcal{G}_{\lambda_2})G_c$, and $\mathbf{E}_{\lambda_3}$ remains.

The subsequent developments and the proofs of the results are based on the first set of AQS functions. However, the results and proofs can be easily modified to fit the second set of AQS functions. Monte Carlo experiments are conducted using both sets of modifications and the results show that their performances are almost the same.

Inconsistency of M-estimator. Before moving into the formal study of the asymptotic properties of the proposed robust M-estimator, a final note is given to the M-estimator of Yang (2018) – it is generally not robust against unknown heteroskedasticity. To see this, we further let $\mathcal{G}_{\lambda_3} = \mathbf{W}_3\mathbf{B}_3^{-1}$. Let $g_{\varpi,t} = \mathtt{diag}(\mathcal{G}_{\varpi,t})$, for $\varpi = \rho_0, \lambda_{10}, \lambda_{20}, \lambda_{30}$, where $\mathcal{G}_{\varpi,t}$ is the tth diagonal block of $\mathcal{G}_\varpi$. Let $h = (h_{10}, \ldots, h_{n0})'$. For the M-estimator to be consistent under unknown heteroskedasticity, it is necessary that $\text{plim}_{n\to\infty}\frac{1}{n(T-1)}S^\diamond_{\mathtt{STLE}}(\psi_0) = 0$, where $S^\diamond_{\mathtt{STLE}}(\psi_0)$ is defined below (3.7). Under mild conditions, we have $\frac{1}{n(T-1)}S^\diamond_{\mathtt{STLE}}(\psi_0) = \frac{1}{n(T-1)}\text{E}[S^\diamond_{\mathtt{STLE}}(\psi_0)] + o_p(1)$. Under unknown $\mathcal{H}$, it is easy to see from (3.16)-(3.18) that the ϖ-component of $\frac{1}{n(T-1)}S^\diamond_{\mathtt{STLE}}(\psi_0)$ can be written as

$$\tfrac{1}{n(T-1)}S^\diamond_{\mathtt{STLE},\varpi}(\psi_0) = \tfrac{1}{n(T-1)}\text{tr}(\mathbf{H}\mathcal{G}_{\varpi 0} - \mathcal{G}_{\varpi 0}) + o_p(1) = \tfrac{1}{T-1}\textstyle\sum_{t=1}^{T-1}\text{Cov}(g_{\varpi,t}, h) + o_p(1),$$

for $\varpi = \rho_0, \lambda_{10}, \lambda_{20}, \lambda_{30}$, where $\text{Cov}(g_{\varpi,t}, h)$ is the sample covariance between the two vectors. Therefore, the necessary conditions for the M-estimator to be consistent are

$$\lim_{n\to\infty}\text{Cov}(g_{\varpi,t}, h) = 0, \quad \text{for } \varpi = \rho_0, \lambda_{10}, \lambda_{20}, \lambda_{30};\ t = 1, \ldots, T-1.$$

Obviously, under a general unknown $\mathcal{H}$ these cannot be true for all terms in general, and therefore the M-estimator of Yang (2018) cannot be consistent in general.[5]

3.2. Asymptotic Properties of Robust M-estimators

In this section, we study the consistency and asymptotic normality of the proposed M-estimator for the FE-SDPD model with the general spatial dependence structure and unknown heteroskedasticity. Some general notations are followed: $\|\cdot\|$ denotes the Frobenius norm, $\gamma_{\min}(\cdot)$ and $\gamma_{\max}(\cdot)$ denote, respectively, the minimum and maximum eigenvalues of a real symmetric matrix, besides the notations used earlier: $|\cdot|$ for determinant, $\text{tr}(\cdot)$ for trace, and $\mathtt{diag}(\cdot)$ for forming a diagonal matrix. The following assumptions are adapted from Yang (2018), allowing for cross-sectional heteroskedasticity of unknown form.

Assumption B: *The innovations v_{it} are (i) inid across $i = 1, \ldots, n$ and iid across*

[5]In fact, this condition may not hold even for the simplest term related to $\mathcal{G}_{\lambda_3} = I_{T-1}\otimes(W_3B_3^{-1})$. Note that $W_3B_3^{-1} = W_3 + \lambda_3W_3^2 + \lambda_3^2W_3^3 + \ldots$, if $\|\lambda_3W_3\| < 1$. According to Anselin (2003), the diagonal elements of W_3^r, $r \geq 2$, are inversely related to the elements k_{ni} of k_n, the vector of number of neighbours for each unit. If $h_{i0} \propto k_{ni}^{-1}$ and $\text{Var}(k_n) = O(1)$, then $\frac{1}{T-1}\sum_{t=1}^{T-1}\text{Cov}(g_{\lambda_3}, h) = \text{Cov}(g_{\lambda_3}, h) = O(1)$. The case of non-vanishing $\text{Var}(k_n)$ can occur in group interaction schemes where sizes of groups are of 'fixed' magnitude but number of groups increases with n. See Yang (2010) and Liu and Yang (2015) for more discussions on this issue.

$t = 0, 1, \ldots, T$ *with* $\mathrm{E}(v_{it}) = 0$ *and* $\mathrm{Var}(v_{it}) = \sigma_v^2 h_{n,i}$, $0 < h_{n,i} \leq c < \infty$ *and* $\frac{1}{n}\sum_{i=1}^n h_{n,i} = 1$; *(ii)* $\mathrm{E}|v_{it}|^{4+\epsilon_0} < \infty$ *for some* $\epsilon_0 > 0$.

Assumption C: *The space parameter space* $\mathbf{\Delta}$ *for* δ *is compact, and the true parameter* δ_0 *lies in its interior.*

Assumption D: *The time-varying regressors* $\{X_t,\ t = 0, 1, \ldots, T\}$ *are exogenous, their values are uniformly bounded, and* $\lim_{n\to\infty} \frac{1}{nT}\Delta X'\Delta X$ *exists and is nonsingular.*

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

Assumption F: *For an* $n\times n$ *matrix* Φ *uniformly bounded in either row or column sums, with elements of uniform order* ι_n^{-1}, *and an* $n \times 1$ *vector* ϕ *with elements of uniform order* $\iota_n^{-1/2}$, *(i)* $\frac{\iota_n}{n}\Delta y_1'\Phi\Delta y_1 = O_p(1)$ *and* $\frac{\iota_n}{n}\Delta y_1'\Phi\Delta v_2 = O_p(1)$; *(ii)* $\frac{\iota_n}{n}[\Delta y_1 - \mathrm{E}(\Delta y_1)]'\phi = o_p(1)$; *(iii)* $\frac{\iota_n}{n}[\Delta y_1'\Phi\Delta y_1 - \mathrm{E}(\Delta y_1'\Phi\Delta y_1)] = o_p(1)$, *and (iv)* $\frac{\iota_n}{n}[\Delta y_1'\Phi\Delta v_2 - \mathrm{E}(\Delta y_1'\Phi\Delta v_2)] = o_p(1)$.

To establish the consistency of $\hat{\delta}_{\mathtt{M}}$, and hence the consistency of $\hat{\boldsymbol{\beta}}_{\mathtt{M}}$ and $\hat{\sigma}^2_{v,\mathtt{M}}$, define the population AQS function: $\bar{S}^*_{\mathtt{STLE}}(\psi) = \mathrm{E}[S^*_{\mathtt{STLE}}(\psi)]$. Given δ, $\bar{S}^*_{\mathtt{STLE}}(\psi) = 0$ is solved at

$$\bar{\boldsymbol{\beta}}_{\mathtt{M}}(\delta) = (\Delta X'\Omega^{-1}\Delta X)^{-1}\Delta X'\Omega^{-1}(\mathbf{B}_1\mathrm{E}\Delta Y - \mathbf{B}_2\mathrm{E}\Delta Y_{-1}), \tag{3.22}$$

$$\bar{\sigma}^2_{v,\mathtt{M}}(\delta) = \tfrac{1}{n(T-1)}\mathrm{E}[\Delta\bar{u}(\delta)'\Omega^{-1}\Delta\bar{\mathbf{u}}(\delta)], \tag{3.23}$$

where $\Delta\bar{\mathbf{u}}(\delta) = \Delta\mathbf{u}(\theta)|_{\boldsymbol{\beta}=\bar{\boldsymbol{\beta}}(\delta)} = \mathbf{B}_1\Delta Y - \mathbf{B}_2\Delta Y_{-1} - \Delta X\bar{\boldsymbol{\beta}}(\delta)$. Substituting $\bar{\beta}_{\mathtt{M}}(\delta)$ and $\bar{\sigma}^2_{v,\mathtt{M}}(\delta)$ back into the $\boldsymbol{\delta}$-component of $\bar{S}^*_{\mathtt{STLE}}(\psi)$, we obtain

$$\bar{S}^{*c}_{\mathtt{STLE}}(\delta) = \begin{cases} \frac{1}{\bar{\sigma}^2_{v,\mathtt{M}}(\delta)}\mathrm{E}[\Delta\bar{\mathbf{u}}(\delta)'\Omega^{-1}\Delta Y_{-1}] + \frac{1}{\bar{\sigma}^2_v(\delta)}\mathrm{E}[\Delta\bar{\mathbf{u}}'(\delta)\mathbf{E}_\rho\Delta\bar{\mathbf{u}}(\delta)], \\ \frac{1}{\bar{\sigma}^2_{v,\mathtt{M}}(\delta)}\mathrm{E}[\Delta\bar{\mathbf{u}}(\delta)'\Omega^{-1}\mathbf{W}_1\Delta Y] + \frac{1}{\bar{\sigma}^2_{v,\mathtt{M}}(\delta)}\mathrm{E}[\Delta\bar{\mathbf{u}}(\delta)'\mathbf{E}_{\lambda_1}\Delta\bar{\mathbf{u}}(\delta)], \\ \frac{1}{\bar{\sigma}^2_{v,\mathtt{M}}(\delta)}\mathrm{E}[\Delta\bar{\mathbf{u}}(\delta)'\Omega^{-1}\mathbf{W}_2\Delta Y_{-1}] + \frac{1}{\bar{\sigma}^2_{v,\mathtt{M}}(\delta)}\mathrm{E}[\Delta\bar{\mathbf{u}}(\delta)'\mathbf{E}_{\lambda_2}\Delta\bar{\mathbf{u}}(\delta)], \\ \frac{1}{2\bar{\sigma}^2_{v,\mathtt{M}}(\delta)}\mathrm{E}[\Delta\bar{\mathbf{u}}(\delta)'\left[C^{-1}\otimes(\mathcal{A} - \mathbf{E}_{\lambda_3})\right]\Delta\bar{\mathbf{u}}(\delta)], \end{cases} \tag{3.24}$$

which gives the population counter part of the concentrated AQS function given in (3.15).

The M-estimator $\hat{\delta}_{\mathtt{M}}$ of δ_0 is a zero of $S^{*c}_{\mathtt{STLE}}(\delta)$ and δ_0 is a zero of $\bar{S}^{*c}_{\mathtt{STLE}}(\delta)$ (as $\bar{\boldsymbol{\beta}}(\delta_0) = \boldsymbol{\beta}_0$ and $\bar{\sigma}^2_v(\delta_0) = \sigma^2_{v0}$). Thus, by Theorem 5.9 of van der Vaart (1998), $\hat{\delta}_{\mathtt{M}}$ will be consistent for δ_0 if $\sup_{\delta\in\Delta}\frac{1}{n(T-1)}\left\|S^{*c}_{\mathtt{STLE}}(\delta) - \bar{S}^{*c}_{\mathtt{STLE}}(\delta)\right\| \xrightarrow{p} 0$, and the following identification condition holds.

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

Theorem 3.1. *Suppose Assumptions A-G hold. Assume further that* (i) $\gamma_{\max}[\mathrm{Var}(\Delta Y)]$ *and* $\gamma_{\max}[\mathrm{Var}(\Delta Y_{-1})]$ *are bounded, and* (ii) $\inf_{\delta\in\Delta}\gamma_{\min}\big(\mathrm{Var}(\mathbf{B}_1\Delta Y - \mathbf{B}_2\Delta Y_{-1})\big) \geq \underline{c}_y > 0$. *We have, as* $n \to \infty$, $\hat{\psi}_{\mathtt{M}} \xrightarrow{p} \psi_0$.

To establish the asymptotic normality of $\hat{\psi}_{\mathtt{M}}$, by backward substitutions on the reduced form (2.3) we represent ΔY and ΔY_{-1} in terms of $\Delta \mathbf{y}_1 = 1_{T-1} \otimes \Delta y_1$ and $\Delta \mathbf{v}$:

$$\Delta Y = \mathbb{R}\, \Delta \mathbf{y}_1 + \boldsymbol{\eta} + \mathbb{S}\Delta \mathbf{v}, \tag{3.25}$$

$$\Delta Y_{-1} = \mathbb{R}_{-1}\Delta \mathbf{y}_1 + \boldsymbol{\eta}_{-1} + \mathbb{S}_{-1}\Delta \mathbf{v}, \tag{3.26}$$

where $\mathbb{R} = \text{blkdiag}(\mathcal{B}_0, \mathcal{B}_0^2, \ldots, \mathcal{B}_0^{T-1})$, $\mathbb{R}_{-1} = \text{blkdiag}(I_n, \mathcal{B}_0, \ldots, \mathcal{B}_0^{T-2})$, $\boldsymbol{\eta} = \mathbb{B}\mathbf{B}_{10}^{-1}\Delta X \boldsymbol{\beta}_0$, $\boldsymbol{\eta}_{-1} = \mathbb{B}_{-1}\mathbf{B}_{10}^{-1}\Delta X\boldsymbol{\beta}_0$, $\mathbb{S} = \mathbb{B}\mathbf{B}_{10}^{-1}\mathbf{B}_{30}^{-1}$, $\mathbb{S}_{-1} = \mathbb{B}_{-1}\mathbf{B}_{10}^{-1}\mathbf{B}_{30}^{-1}$,

$$\mathbb{B} = \begin{pmatrix} I_n & 0 & \ldots & 0 & 0 \\ \mathcal{B}_0 & I_n & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathcal{B}_0^{T-2} & \mathcal{B}_0^{T-3} & \ldots & \mathcal{B}_0 & I_n \end{pmatrix}, \quad \text{and} \quad \mathbb{B}_{-1} = \begin{pmatrix} 0 & 0 & \ldots & 0 & 0 \\ I_n & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathcal{B}_0^{T-3} & \mathcal{B}_0^{T-4} & \ldots & I_n & 0 \end{pmatrix}.$$

By representations (3.25) and (3.26), the AQS function at ψ_0 can be written as

$$S^*_{\mathtt{STLE}}(\psi_0) = \begin{cases} \Pi_1'\Delta \mathbf{v}, \\ \Delta \mathbf{v}'\Phi_1\Delta \mathbf{v} - \frac{n(T-1)}{2\sigma_{v0}^2}, \\ \Delta \mathbf{v}'\Psi_1\Delta \mathbf{y}_1 + \Pi_2'\Delta \mathbf{v} + \Delta \mathbf{v}'\Phi_2\Delta \mathbf{v}, \\ \Delta \mathbf{v}'\Psi_2\Delta \mathbf{y}_1 + \Pi_3'\Delta \mathbf{v} + \Delta \mathbf{v}'\Phi_3\Delta \mathbf{v}, \\ \Delta \mathbf{v}'\Psi_3\Delta \mathbf{y}_1 + \Pi_4'\Delta \mathbf{v} + \Delta \mathbf{v}'\Phi_4\Delta \mathbf{v}, \\ \Delta \mathbf{v}'\Phi_5\Delta \mathbf{v}, \end{cases} \tag{3.27}$$

where $\Pi_1 = \frac{1}{\sigma_{v0}^2}\mathbf{C}_{b0}\Delta X$, $\Pi_2 = \frac{1}{\sigma_{v0}^2}\mathbf{C}_{b0}\boldsymbol{\eta}_{-1}$, $\Pi_3 = \frac{1}{\sigma_{v0}^2}\mathbf{C}_{b0}\mathbf{W}_1\boldsymbol{\eta}$, $\Pi_4 = \frac{1}{\sigma_{v0}^2}\mathbf{C}_{b0}\mathbf{W}_2\boldsymbol{\eta}_{-1}$,
$\Phi_1 = \frac{1}{2\sigma_{v0}^4}\mathbf{C}^{-1}$, $\Phi_2 = \frac{1}{\sigma_{v0}^2}(\mathbf{C}_{b0}\mathbb{S}_{-1} + \mathbf{B}_{30}^{-1\prime}\mathbf{E}_{\rho 0}\mathbf{B}_{30}^{-1})$, $\Phi_3 = \frac{1}{\sigma_{v0}^2}(\mathbf{C}_{b0}\mathbf{W}_1\mathbb{S} + \mathbf{B}_{30}^{-1\prime}\mathbf{E}_{\lambda_{10}}\mathbf{B}_{30}^{-1})$,
$\Phi_4 = \frac{1}{\sigma_{v0}^2}(\mathbf{C}_{b0}\mathbf{W}_2\mathbb{S}_{-1} + \mathbf{B}_{30}^{-1\prime}\mathbf{E}_{\lambda_{20}}\mathbf{B}_{30}^{-1})$, $\Phi_5 = \frac{1}{2\sigma_{v0}^2}\left[C^{-1} \otimes \left(B_{30}^{-1\prime}(\mathcal{A}_0 - \mathbf{E}_{\lambda_{30}})B_{30}^{-1}\right)\right]$,
$\Psi_1 = \frac{1}{\sigma_{v0}^2}\mathbf{C}_{b0}\mathbb{R}_{-1}$, $\Psi_2 = \frac{1}{\sigma_{v0}^2}\mathbf{C}_{b0}\mathbf{W}_1\mathbb{R}$, and $\Psi_3 = \frac{1}{\sigma_{v0}^2}\mathbf{C}_{b0}\mathbf{W}_2\mathbb{R}_{-1}$.

Theorem 3.2. *Assume Assumptions A-G hold. We have, as $n \to \infty$,*

$$\sqrt{n(T-1)}(\hat{\psi}_{\mathtt{M}} - \psi_0) \xrightarrow{D} N\left[0, \lim_{n\to\infty} \Sigma^{*-1}_{\mathtt{STLE}}(\psi_0)\Gamma^*_{\mathtt{STLE}}(\psi_0)\Sigma^{*-1}_{\mathtt{STLE}}(\psi_0)\right],$$

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

3.3. Robust Estimation of VC Matrix

As $\Sigma^*_{\mathtt{STLE}}(\psi_0)$ is the expected negative Hessian, it can be consistently estimated by its observed counter part, $\widehat{\Sigma}^*_{\mathtt{STLE}} = -\frac{1}{n(T-1)}\frac{\partial}{\partial \psi'}S^*_{\mathtt{STLE}}(\psi)\big|_{\psi=\hat{\psi}_{\mathtt{M}}}$, with the detailed expression and the proof of consistency being given in the proof of Theorem 3.3 in Appendix B.

However, the traditional methods of estimating $\Gamma^*_{\mathtt{STLE}}(\psi_0)$ are not applicable as (i) the initial differences Δy_1 need to be 'specified' or modelled when T is fixed and small, of which a valid modelling strategy is unknown for the general FE-SDPD model, and (ii) the analytical expression of $\Gamma^*_{\texttt{STLE}}(\psi_0)$, if it is available, cannot be used as it contains unobservables. We follow the idea of Yang (2018) to decompose the AQS function into a sum of vector martingale difference (MD) sequences so that the average of the outer products of the MDs (OPMD) gives a consistent estimate of the VC matrix of the AQS function.

From (3.27) we see that the AQS function $S^*_{\texttt{STLE}}(\psi_0)$ contains three types of elements:

$$\Pi'\Delta\mathbf{v}, \quad \Delta\mathbf{v}'\Phi\Delta\mathbf{v}, \quad \text{and } \Delta\mathbf{v}'\Psi\Delta\mathbf{y}_1,$$

where Π, Φ and Ψ are nonstochastic matrices (depending on ψ_0) with Π being $n(T-1)\times p$ or $n(T-1)\times 1$, and Φ and Ψ being $n(T-1)\times n(T-1)$.

As our asymptotics depend only on n and the transformed errors remain independent across i, the above linear, quadratic and bilinear terms can be written as sums of n uncorrelated terms, so that their variance can be estimated by the averages of the outer products of the summands. For a square matrix A, let A^u, A^l and A^d be, respectively, its upper-triangular, lower-triangular, and diagonal matrix such that $A = A^u + A^l + A^d$. Denote by Π_t, Φ_{ts} and Ψ_{ts} the submatrices of Π, Φ and Ψ partitioned according to $t, s = 2, \ldots, T$. Let $\{\mathcal{G}_{n,i}\}$ be the increasing sequence of σ-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 σ-field generated by $(v_0, \Delta y_0)$, and define $\mathcal{F}_{n,i} = \mathcal{F}_{n,0} \otimes \mathcal{G}_{n,i}$. Clearly, $\mathcal{F}_{n,i-1} \subseteq \mathcal{F}_{n,i}$, i.e., $\{\mathcal{F}_{n,i}\}_{i=1}^n$ is an increasing sequence of σ-fields, for each $n \geq 1$.

First, for the terms linear in $\Delta\mathbf{v}$, we have,

$$\Pi'\Delta\mathbf{v} = \textstyle\sum_{t=2}^T \Pi_t'\Delta v_t = \sum_{t=2}^T \sum_{i=1}^n \Pi_{it}'\Delta v_{it} = \sum_{i=1}^n \sum_{t=2}^T \Pi_{it}'\Delta v_{it} \equiv \sum_{i=1}^n g_{\Pi,i}. \tag{3.28}$$

Clearly, $\{g_{\Pi,i}\}$ are independent with mean zero, and thus form a vector M.D. sequence.

Second, the terms quadratic in $\Delta\mathbf{v}$ are decomposed as follows,

$$\begin{aligned}
\Delta\mathbf{v}'\Phi\Delta\mathbf{v} &= \textstyle\sum_{t=2}^T \sum_{s=2}^T \Delta v_t'\Phi_{ts}\Delta v_s \\
&= \textstyle\sum_{t=2}^T \sum_{s=2}^T \Delta v_t'(\Phi_{ts}^u + \Phi_{ts}^l + \Phi_{ts}^d)\Delta v_s \\
&= \textstyle\sum_{t=2}^T \sum_{s=2}^T \big(\Delta v_s'\Phi_{ts}^{u\prime}\Delta v_t + \Delta v_t'\Phi_{ts}^l\Delta v_s + \Delta v_t'\Phi_{ts}^d\Delta v_s\big) \\
&= \textstyle\sum_{t=2}^T \sum_{s=2}^T \big(\Delta v_s'\Phi_{ts}^{u\prime}\Delta v_t + \Delta v_s'\Phi_{ts}^l\Delta v_t + \Delta v_t'\Phi_{ts}^d\Delta v_s\big) \\
&= \textstyle\sum_{t=2}^T \sum_{s=2}^T \big[\Delta v_t'(\Phi_{st}^{u\prime} + \Phi_{ts}^l)\Delta v_s + \Delta v_t'\Phi_{ts}^d\Delta v_s\big] \\
&= \textstyle\sum_{t=2}^T \Delta v_t'\Delta\xi_t + \sum_{t=2}^T \Delta v_t'\Delta v_t^* \\
&= \textstyle\sum_{i=1}^n \sum_{t=2}^T \big(\Delta v_{it}\Delta\xi_{it} + \Delta v_{it}\Delta v_{it}^*\big),
\end{aligned}$$

where $\Delta\xi_t = \sum_{s=2}^T(\Phi_{st}^{u\prime} + \Phi_{ts}^l)\Delta v_s$ and $\Delta v_t^* = \sum_{s=2}^T \Phi_{ts}^d\Delta v_s$.

Letting $\{c_{ts}\} = C$ and $\{\Phi_{ii,ts}\} = \text{diag}(\Phi_{ts})$, we have

$$\begin{aligned}
\mathrm{E}(\Delta\mathbf{v}'\Phi\Delta\mathbf{v}) &= \sigma_{v0}^2 \text{tr}[(C\otimes\mathcal{H})\Phi] \\
&= \textstyle\sigma_{v0}^2 \sum_{t=2}^T \sum_{s=2}^T \text{tr}(c_{ts}\Phi_{st}\mathcal{H}) \\
&= \textstyle\sum_{i=1}^n \sigma_{v0}^2 h_{n,i} \sum_{s=2}^T \sum_{s=2}^T (c_{ts}\Phi_{ii,st}) \\
&\equiv \textstyle\sum_{i=1}^n \sum_{t=2}^T d_{\Phi,it}.
\end{aligned}$$

Thus, $\Delta\mathbf{v}'\Phi\Delta\mathbf{v} - \mathrm{E}(\Delta\mathbf{v}'\Phi\Delta\mathbf{v}) = \sum_{i=1}^{n} g_{\Phi,i}$, where,

$$g_{\Phi,i} = \sum_{t=2}^{T}(\Delta v_{it}\Delta\xi_{it} + \Delta v_{it}\Delta v_{it}^{*} - d_{\Phi,it}), \tag{3.29}$$

$\{g_{\Phi,i}, \mathcal{G}_{n,i}\}$ form an M.D. sequence as $\Delta\xi_{it}$ is $\mathcal{G}_{n,i-1}$-measurable,and $\mathrm{E}(g_{\Phi,i}|\mathcal{G}_{n,i-1}) = 0$.

Finally, we decompose the terms bilinear in Δv and $\Delta\mathbf{y}_1$. First, we transform Δy_1 into

$$\Delta y_1^{\circ} = B_{30}B_{10}\Delta y_1 = B_{30}B_{20}\Delta y_0 + B_{30}\Delta\mathbf{X}_1\beta_0 + \Delta v_1.$$

Letting $\Psi_{t+} = \sum_{s=2}^{T}\Psi_{ts}$, $t = 2,\ldots,T$, $\Theta = \Psi_{2+}(B_{30}B_{10})^{-1}$ and $\{\Theta_{ii}\} = \mathtt{diag}(\Theta)$ we have,

$$\begin{aligned}
\Delta\mathbf{v}'\Psi\Delta\mathbf{y}_1 &= \sum_{t=2}^{T}\sum_{s=2}^{T}\Delta v_t'\Psi_{ts}\Delta y_1 \\
&= \sum_{t=2}^{T}\Delta v_t'\big(\sum_{s=2}^{T}\Psi_{ts}\big)\Delta y_1 \\
&= \sum_{t=2}^{T}\Delta v_t'\Psi_{t+}\Delta y_1 \\
&= \Delta v_2'\Theta\Delta y_1^{\circ} + \sum_{t=3}^{T}\Delta v_t'\Delta y_{1t}^{*} \\
&= \Delta v_2'(\Theta^{u} + \Theta^{l} + \Theta^{d})\Delta y_1^{\circ} + \sum_{t=3}^{T}\Delta v_t'\Delta y_{1t}^{*} \\
&= \Delta v_2'(\Theta^{u} + \Theta^{l})\Delta y_1^{\circ} + \Delta v_2'\Theta^{d}\Delta y_1^{\circ} + \sum_{t=3}^{T}\Delta v_t'\Delta y_{1t}^{*} \\
&= \sum_{i=1}^{n}\Delta v_{2i}\Delta\zeta_i + \sum_{i=1}^{n}\Theta_{ii}\Delta v_{2i}\Delta y_{1i}^{\circ} + \sum_{i=1}^{n}\sum_{t=3}^{T}\Delta v_{it}'\Delta y_{1it}^{*},
\end{aligned}$$

where $\Delta y_{1t}^{*} = \Psi_{t+}\Delta y_1$ and $\{\Delta\zeta_i\} = \Delta\zeta = (\Theta^{u} + \Theta^{l})\Delta y_1^{\circ}$. It can be easily seen that $\mathrm{E}(\Delta v_{2i}\Delta\zeta_i|\mathcal{F}_{n,i-1}) = 0$, therefore the first term is the sum of an M.D. sequence. The third term is the sum of n uncorrelated terms of mean zero as Δy_1 is independent of $\Delta v_t, t \geq 3$. By Assumption A, Δy_0 is independent of $v_t, t \geq 1$, so we have $\mathrm{E}(\Delta v_2'\Theta^{d}\Delta y_1^{\circ}) = -\sigma_{v0}^{2}\mathrm{tr}(\Theta\mathcal{H})$. Therefore, $\Delta\mathbf{v}'\Psi\mathbf{y}_1 - \mathrm{E}(\Delta\mathbf{v}'\Psi\mathbf{y}_1) = \sum_{i=1}^{n} g_{\Psi,i}$ where,

$$g_{\Psi,i} = \Delta v_{2i}\Delta\zeta_i + \Theta_{ii}(\Delta v_{2i}\Delta y_{1i}^{\circ} + \sigma_{v0}^{2}h_{n,i}) + \sum_{t=3}^{T}\Delta v_{it}\Delta y_{1it}^{*}. \tag{3.30}$$

$\mathrm{E}(g_{\Psi,i}|\mathcal{F}_{n,i-1}) = 0$ and hence $\{g_{3i}, \mathcal{F}_{n,i}\}$ form an M.D. sequence. It is then easy to see that $\{(g_{\Pi,i}', g_{\Phi,i}, g_{\Psi,i})', \mathcal{F}_{n,i}\}$ form a vector M.D. sequence.

Using (3.28)-(3.30), $S_{\mathtt{STLE}}^{*}(\psi_0)$ can be written as a sum of vector M.D.s. For each $\Pi_r, r = 1,2,3,4$, in (3.27), define $g_{\Pi r,i}$ according to (3.28); for each $\Phi_r, r = 1,\ldots,5$, define $g_{\Phi r,i}$ according to (3.29); and for each $\Psi_r, r = 1,2,3$, define $g_{\Psi r,i}$ according to (3.30). Define

$$g_i = (g_{\Pi 1i}',\ g_{\Phi 1i},\ g_{\Psi 1i} + g_{\Pi 2i} + g_{\Phi 2i},\ g_{\Psi 2i} + g_{\Pi 3i} + g_{\Phi 3i},\ g_{\Psi 3i} + g_{\Pi 4i} + g_{\Phi 4i},\ g_{\Phi 5i})'.$$

Then, $S_{\mathtt{STLE}}^{*}(\psi_0) = \sum_{i=1}^{n} g_i$, and $\{g_i, \mathcal{F}_{n,i}\}$ form a vector M.D. sequence and $\mathrm{Var}[S_{\mathtt{STLE}}^{*}(\psi_0)] = \sum_{i=1}^{n}\mathrm{E}(g_i g_i')$. Therefore the OPMD estimator of $\Gamma_{\mathtt{STLE}}^{*}$ is given as:

$$\widehat{\Gamma}_{\mathtt{STLE}}^{*} = \tfrac{1}{n(T-1)}\sum_{i=1}^{n}\hat{g}_i\hat{g}_i', \tag{3.31}$$

where $\hat{g}_i$ is obtained by replacing ψ_0, $\Delta\mathbf{v}$, and $h_{n,i}$ in g_i by $\hat{\psi}_{\mathtt{M}}$, $\widehat{\Delta}v$, and $\hat{h}_{n,i}$, noting that Δy_1 is observed. Finally, to estimate $\{h_{n,i}\}$, using the expression for $\mathrm{Var}(\Delta\mathbf{u})$ given below (2.4), we can write $\mathbf{H} = \frac{1}{\sigma_v^2}\mathbf{C}_b\mathrm{E}(\Delta\mathbf{u}\Delta\mathbf{u}')\mathbf{B}_3'$, of which a natural estimator would be $\widehat{\mathbf{H}} = \hat{\sigma}_{v,\mathtt{M}}^{-2}\widehat{\mathbf{C}}_b\widehat{\Delta}\mathbf{u}\widehat{\Delta}\mathbf{u}'\widehat{\mathbf{B}}_3'$ in line with the idea of White (1980), where $\widehat{\Delta}\mathbf{u}$ is the residual vector from the robust M-estimation. Alternatively, $h_{n,i}$ can be simply estimated as follows. Noting that $E[(\Delta v_{it})^2] = 2\sigma_v^2 h_{n,i}$, for $t = 2, \ldots, T$, averaging over t gives $\hat{h}_{n,i} = \frac{1}{2(T-1)}\hat{\sigma}_{v,\mathtt{M}}^{-2}\sum_{t=2}^{T}(\Delta\hat{v}_{it})^2$.

Theorem 3.3. *Under the assumptions of Theorem* (3.1)*, we have, as* $n \to \infty$*,*

$$\widehat{\Gamma}^*_{\mathtt{STLE}} - \Gamma^*_{\mathtt{STLE}}(\psi_0) = \tfrac{1}{n(T-1)}\sum_{i=1}^{n}\left[\hat{g}_i\hat{g}_i' - \mathrm{E}(g_i g_i')\right] \xrightarrow{p} 0,$$

and hence, $\widehat{\Sigma}^{*-1}_{\mathtt{STLE}}\widehat{\Gamma}^*_{\mathtt{STLE}}\widehat{\Sigma}^{*-1}_{\mathtt{STLE}} - \Sigma^{*-1}_{\mathtt{STLE}}(\psi_0)\Gamma^*_{\mathtt{STLE}}(\psi_0)\Sigma^{*-1}_{\mathtt{STLE}}(\psi_0) \xrightarrow{p} 0.$

4. Monte Carlo Study

Monte Carlo experiments are carried out to investigate the finite sample performance of the proposed M-estimator of the FE-SDPD model with unknown heteroskedasticity and the finite sample performance of the corresponding VC matrix estimator. Further, the proposed robust M-estimator is compared with the M-estimator of Yang (2018) under homoskedasticity for efficiency purpose. It is also compared with the robust optimal GMM estimator of Kuersteiner and Prucha (2018). Finally, an investigation is given for a model with spatial Durbin terms. We use the following three data generating processes (DGPs):

$$\text{DGP1: } y_t = \rho y_{t-1} + \lambda_1 W_1 y_t + \lambda_2 W_2 y_{t-1} + X_t\beta + Z\gamma + \mu + u_t,$$
$$\text{DGP2: } y_t = \rho y_{t-1} + \lambda_1 W_1 y_t + \lambda_2 W_2 y_{t-1} + X_t\beta + W_d X_t \beta_d + Z\gamma + \mu + u_t,$$
$$\text{DGP3: } y_t = \rho y_{t-1} + \lambda_1 W_1 y_t + \lambda_2 W_2 y_{t-1} + X_t\beta + Z\gamma + \mu + v_t,$$

where $u_t = \lambda_3 W_3 u_t + v_t$ for DGP1 and DGP2, μ is the vector of fixed effects and v_t the vector of idiosyncratic errors.

The elements of X_t are generated in a similar fashion as in Hsiao et al. (2002),[6] and the elements of Z are randomly generated from Bernoulli (0.5). The spatial weight matrices are generated according to group interaction schemes where group sizes change across the groups but not with respect to the sample size.[7] The idiosyncratic errors are generated as $v_t = \sigma_v^2\mathcal{H}e_t$. Similar to Lin and Lee (2010), the heteroskedasticity $\mathcal{H}$ is generated in two different ways: H-I, for each i, if the number of neighbors is smaller than the average number of neighbors, then $h_{n,i}$ equals to the number of its neighbors, otherwise it is the square of the inverse of the number of its neighbors; and H-II, for each i, if the number of neighbors is larger than the average number of neighbor, then $h_{n,i}$ equals to the number of its neighbors, otherwise it is the square of the inverse of the number of its neighbors. The variance structure is nonlinear in the number of neighbors. In the first case, the error variance increases and then decreases with the number of neighbors, and in the second case, the variance decreases and then increases with the number of neighbors. The $h_{n,i}$'s are normalized to have mean

[6] The detail is: $X_t = \mu_x + gt1_n + \zeta_t$, $(1-\phi_1 L)\zeta_t = \varepsilon_t + \phi_2\varepsilon_{t-1}$, $\varepsilon_t \sim N(0, \sigma_1^2 I_n)$, $\mu_x = e + \frac{1}{T+m+1}\sum_{t=-m}^{T}\varepsilon_t$, and $e \sim N(0, \sigma_2^2 I_n)$. Let $\theta_x = (g, \phi_1, \phi_2, \sigma_1, \sigma_2)$. Thus, σ_1/σ_v represents the signal-to-noise ratio (SNR).

[7] It can be generated as follows, first generate a vector of possible group sizes randomly $(m_1, \ldots, m_k)$ such that $\sum_{j=1}^{k} m_j = n_1$. Then replicate the groups r times such that $rn_1 = n$.

one. The distribution of e_t can be (i) normal, (ii) normal mixture with 10% of the values generated from $N(0,4)$ and 90% from $N(0,1)$, or (iii) chi-squared with degree of freedom of 3. In both (ii) and (iii), the generated errors are standardized to have mean zero and variance 1. We choose $\beta = 1$, $\beta_d = 0.2$ and $\sigma_v^2 = 1$. We use a set of values for ρ ranging from -0.9 to 0.9, a set of values for $(\lambda_1, \lambda_2, \lambda_3)$ in the similar range, $m = 10$, $T = 3$ or 7, and $N = 50, 100, 200$ or 400. Each set of Monte Carlo results, corresponding to a combination of the values of $(n, T, m, \rho, \lambda_1, \lambda_2, \lambda_3)$ is based on 2000 samples.

Monte Carlo (empirical) means and standard deviations (sds) are reported for the CQML estimators (CQMLEs), the M-estimators, and the robust M-estimators. Empirical averages of the robust standard errors (rses) based on the VC matrix estimate $\widehat{\Sigma}^{*-1}_{\mathtt{SDPD}}\widehat{\Gamma}^{*}_{\mathtt{SDPD}}\widehat{\Sigma}^{*-1}_{\mathtt{SDPD}}$ are also reported for the robust M-estimators, which should be compared with the corresponding empirical sds. The ses of the M-estimator based only on $\widehat{\Sigma}^{*}_{\mathtt{SDPD}}$ or $\widehat{\Gamma}^{*}_{\mathtt{SDPD}}$ are also computed, and the results (unreported to conserve space) show that they are not robust. These results, together with additional unreported Monte Carlo results are available upon request.

Tables 1-3 present the results based on DGP1, the FE-SDPD model with all three types of spatial effects. Tables 1a and 1b present the results when $\mathcal{H}$ is generated by H-I and $T = 3$, with SNR being 1 and 3 (see Footnote 6), respectively. The results show that the proposed robust M-estimators perform quite well. The inconsistency of CQMLEs and M-estimators is shown clearly in Table 1a. The CQMLEs and M-estimators for ρ, λ_1, λ_2 and λ_3 are severely biased, and they do not show a sign of convergence as n increases. Meanwhile, the robust M-estimators perform much better and show a clear sign of convergence. The rses are close to the corresponding Monte Carlo sds in general, showing the robustness and good finite sample performance of the proposed OPMD estimate of VC matrix. The rses of λ_3 is slightly biased. However, this bias is reduced when SNR is larger. The results presented in Table 1b show clearly the biases of CQMLEs and M-estimators are reduced but still persistent, especially for the case of λ_3. The robust M-estimators still perform very well, and the bias encountered in rses of λ_3 reduced. When T is set to 7, the results (unreported for brevity) show that the bias of the CQMLEs and M-estimators are reduced, but the pattern of inconsistency still remains, whereas the robust M-estimators and rses are nearly unbiased.

Tables 2a and 2b present the results when $\mathcal{H}$ is generated by H-II and $T = 3$, with SNR being 1 and 3, respectively. In this case, both the M-estimators and the robust M-estimators of ρ, λ_1 and λ_2 perform quite well, but the M-estimator of λ_3 has a larger bias. Comparing with the results in Table 1a and 1b we see that when sample size is not large, the CQMLEs and the M-estimators can be very sensitive to the way heteroskedasticity is generated and to the magnitude of SNR. The rses under H-II perform better than those under H-I and are generally quite close to the corresponding Monte Carlo sds, except that there is some bias in rses of $\hat{\sigma}^2_{\mathtt{M},v}$ when error terms are not normally distributed.

Table 3 presents the results when the idiosyncratic errors are homoskedastic ($\mathcal{H} = I_n$). In this case, both the M-estimation of Yang (2018) and the proposed robust M-estimator are consistent. As expected, the results show that the proposed robust M-estimator is less efficient than the M-estimator of Yang (2018). However, this efficiency loss is quite marginal.

Table 4 presents the results based on DGP2, the FE-SDPD model with all three types of spatial effects and a spatial Durbin term, where $\mathcal{H}$ = H-I, $T = 3$ and SNR = 3. As discussed earlier, the spatial Durbin terms can simply be included in the model as additional exogenous regressors, and the proposed set of estimation and inference methods perform well.[8]

Table 5 presents the results based on DGP3, the FE-SDPD model with SL and STL, where $\mathcal{H}$ = H-I, $T = 3$ and SNR = 3. As the main focus of this set of Monte Carlo experiments is to compare the proposed robust M-estimator with the GMM estimator of Kuersteiner and Prucha (2018), we only report the empirical means and sds for these two estimators.[9] The results show clear convergence patterns of both estimators: as sample size grows both estimators become less biased and less variable. The robust M-estimator has a smaller bias than the GMM estimator for all sample sizes and all three different error distributions. The empirical sds show that the robust M-estimator is much more efficient than the GMM estimator.

5. Empirical Application: Sovereign Risk Spillover

This section presents an empirical application of the proposed M-estimator for the FE-SDPD model under small T and unknown heteroskedasticity. We investigate international spillover of the sovereign bond spreads of 51 countries from 2007 to 2012, and we find that it is important to allow for heteroskedasticity in the estimation.

The increasing economic and financial integration worldwide has led to a continuous discussion of global transmission of risk in the past two decades, especially after the European sovereign debt crisis from 2010 to 2012. Many studies have applied the spatial econometrics frameworks to analyse risk spillovers. Saldías (2013) uses a spatial error model to identify sector risk determinants. Favero (2013) uses a GVAR approach that incorporates the space-time lag to model the government bond spreads in the Euro area. Keiler and Eder (2015) use a spatial lag model to model the credit default swap (CDS) spreads of financial institutions, whereas Tonzer (2015) uses a spatial lag model to analyse the banking sector risk. Blasques et al. (2016) model sovereign CDS spreads using spatial Durbin panel data model with time-varying spatial dependence parameter. Debarsy et al. (2018) use a spatial dynamic panel data model to measure sovereign risk spillover considering different channels of risk transmission.

[8] Halleck Vega and Elhorst (2015) advocate spatial Durbin models. There can be issues of identification and overfitting with Durbin models, especially when all spatial terms are included in the model and the same spatial weight matrix is used for all. See Elhorst (2012) and Lee and Yu (2016) for more details.

[9] To implement the GMM in Kuersteiner and Prucha (2018) we use the code provided by the authors at: `http://econweb.umd.edu/%7Ekuersteiner/research_UMD.html`

All of these works are under the assumption that disturbances are homoskedastic. However, as different financial sectors vary greatly in size and depth, different countries vary greatly in so many aspects such as population, location, completeness of financial market, bureaucratic quality, government stability, openness to trade, and other social-economical characteristics, it is natural to think that we should allow the innovations to be heteroskedastic.

Our data covers 51 countries, including both advanced and emerging markets over six years from 2007 to 2012. The list of countries included in our analysis is presented in Table E1. Bond yield spread, credit default swap and credit ratings are three commonly used measures of sovereign risk in the literature. We follow the main body of the literature to measure the sovereign risk by sovereign bond yields spreads. For advanced economies, the spread is computed as the difference between the 10-year bond yields on the secondary market and the 10-year US treasury bond yield. We obtain the data from `Datastream`. For emerging markets, we use Emerging Market Bond Index Global (EMBIG) obtained from Global Economic Monitor of the World Bank database to measure the spreads in order to have a consistent measure for both advanced and emerging economies as in Beirne and Frazscher (2013) and Debarsy et al. (2018). In line with the literature, the set of exogenous explanatory variables we use contains debt-to-GDP ratio, deficit-to-GDP ratio, current account balance (CA) to GDP ratio, real GDP growth rate, inflation (CPI), real effective exchange rate and the volatility index (VIX). The first five variables control the macroeconomic and financial fundamentals of each country and the last variable controls the general market conditions. The data for these variables are collected from IMF World Economic Outlook (WEO). We use yearly data because it is the original frequency for most of the variables we consider. The original frequencies are daily for VIX and the bond yield spread, and monthly for real effective exchange rate. We use the average values over a year for those variables.

Table E1. List of Countries

Argentina	Australia	Austria	Belgium	Brazil	Bulgaria	Canada
Chile	China	Colombia	Czech Republic	Denmark	Ecuador	Egypt
Finland	France	Germany	Ghana	Greece	Hungary	Indonesia
Ireland	Italy	Jamaica	Japan	Kazakhstan	Korea	Malaysia
Mexico	Netherlands	New Zealand	Norway	Pakistan	Panama	Peru
Philippines	Poland	Portugal	Russia	Singapore	South Africa	Spain
Sweden	Switzerland	Tunisia	Turkey	Ukraine	United Kingdom	Uruguay
Venezuela	Vietnam					

We fit the data to the general model (2.1) and several sub-models, and report the results corresponding to the following two models that best fit the data: the SDPD model with `SL` and `STL`: $y_t = \rho y_{t-1} + \lambda_1 W y_t + \lambda_2 W y_{t-1} + X_t\beta + Z\gamma + \mu + v_t$, and the SDPD model with `SE` only: $y_t = \rho y_{t-1} + X_t\beta + Z\gamma + \mu + u_t,\;\; u_t = \lambda_3 W_3 u_t + v_t$.

Both specifications are in fact widely used in empirical studies. In both models, y_t is an $n \times 1$ vector of government bond yields spreads, X_t is an $n \times k$ matrix containing the observed time varying exogenous variables, Z is an $n \times k_z$ matrix containing the observed time invariant variables, μ is an $n \times 1$ vector of unobserved country-specific fixed effects, and the elements $\{v_{it}\}$ of v_t are assumed to be *iid* across time but *inid* across country with mean zero and variance $\sigma_v^2 h_i$. We consider three different weight matrices to investigate three risk transmission channels. The first weight matrix, W_{trade}, represents the real linkage between economies, and it is constructed using bilateral trade flow to measure the connectivity between countries. The (i, j) element of $W_{\text{trade},t}$ is $W_{ijt} = \frac{M_{ijt}+X_{ijt}}{GDP_{it}+GDP_{jt}}$, where M_{ijt} is the total import of country i from country j in year t represented in US Dollars, X_{ijt} is the total export of country i to country j in year t, GDP_{it} is the nominal gross domestic product for country i in year t and W_{trade} is the time average of $W_{\text{trade},t}$. The data for bilateral trade volume is collected from the World Integrated Trade Solution (WITS) database, and the data for GDP is available in the WEO database. The second and third weight matrices, W_{deficit} and W_{debt}, represent the information linkage between economies, and they are constructed using similarities in debt or deficit level to measure the connectivity of government risk. Elements of these two weight matrices are the time average of $W_{ijt} = \frac{1}{|A_{it}-A_{jt}|+1}$, where A_{it} is debt-to-GDP ratio or deficit-to-GDP ratio of country i at time t. See Favero (2013) and Debarsy et al. (2018) for more discussions on the transmission channels.

Table E2. Estimation Results of SDPD Model with SL and STL

	W_{trade}		W_{deficit}		W_{debt}	
Variables	`M-Est`	`RM-Est`	`M-Est`	`RM-Est`	`M-Est`	`RM-Est`
debt/GDP	.1190(.060)	.0782(.029)	.1181(.059)	.0671(.034)	.1179(.057)	.0675(.031)
deficit/GDP	-.0536(.089)	-.1756(.126)	-.0878(.098)	-.1637(.155)	-.0512(.097)	-.2118(.173)
CA/GDP	.0130(.040)	-.0266(.022)	.0247(.035)	-.0312(.025)	.0271(.031)	-.0110(.010)
CPI	.1519(.034)	.1765(.088)	.1627(.038)	.2518(.119)	.1751(.034)	.2177(.109)
GPD growth	-.1867(.103)	-.1411(.071)	-.1494(.113)	-.0317(.087)	-.1589(.089)	-.0770(.071)
VIX	.0443(.034)	.0204(.009)	.0466(.077)	.0849(.036)	.0707(.045)	.0393(.017)
Reer	.0267(.021)	.0340(.017)	.0202(.020)	.0309(.013)	.0277(.021)	.0381(.016)
Y_{t-1}	.1417(.059)	.5490(.108)	.2126(.046)	.6739(.112)	.1392(.061)	.6207(.114)
WY_t	.1815(.266)	.4955(.241)	.3668(.432)	.3441(.200)	.0831(.399)	.5767(.289)
WY_{t-1}	-.5713(.291)	-.5951(.194)	-.6918(.347)	-.4193(.176)	-.4008(.165)	-.8803(.422)

Table E2 shows the estimation results when the data is fit to `SL-STL` model. We compare the results from the M-estimation of Yang (2018) and the proposed robust M-estimation in this paper under all three weight matrices. Under the robust M-estimation, the signs for all parameters are as expected and the parameter estimates for debt/GDP, CPI, VIX and real effective exchange rate are significant regardless of which weight matrix is used. The results are in line with the previous studies. Under the M-estimation, the sign of parameter estimate for CA/GDP is not as expected although insignificant, and only the parameters of debt/GDP and CPI are significant. The parameter of time-lag variable is estimated to be positive and significant under both methods but the magnitudes are much larger for robust M-estimates. Under the robust M-estimation, the parameter of spatial-lag variable is estimated to be positive and significant when W_{trade} and W_{debt} are used and insignificant when W_{deficit} is used. Under the M-estimation, the parameter estimates for spatial-lag variable are positive but insignificant under all three weight matrices and the magnitudes are smaller than those of the robust M-estimates. Parameter estimates for space-time lag variable are negative and significant for all weight matrices under the robust M-estimation, whereas it is insignificant when W_{deficit} is used under the M-estimation.

Table E3 shows the estimation results when the data is fit to SE model. First, we observe that the signs of parameter estimates for all variables stay the same and the magnitudes remain similar for both methods. The parameter estimate of debt/GDP becomes insignificant whereas the parameter estimate of deficit/GDP becomes significant under the robust M-estimation for all weight matrices. Under the M-estimation, both debt/GDP and deficit/GDP are insignificant in this model. The results for other variables are similar to those from STL model. The spatial error parameter and the dynamic parameter are estimated to be positive and significant by the robust M-estimation, but insignificant under the M-estimation.

Table E3. Estimation Results of SDPD Model with SE

	W_{trade}		W_{deficit}		W_{debt}	
Variables	M-Est	RM-Est	M-Est	RM-Est	M-Est	RM-Est
debt/GDP	.1189(.191)	.0731(.073)	.1073(.760)	.0687(.076)	.1205(.376)	.0681(.077)
deficit/GDP	-.0503(.057)	-.1828(.093)	-.0744(.361)	-.2107(.085)	-.0264(.026)	-.1807(.091)
CA/GDP	.0182(.178)	-.0170(.060)	-.0047(.012)	-.0339(.056)	.0169(.035)	-.0284(.055)
CPI	.1407(.071)	.1859(.034)	.1674(.083)	.2149(.037)	.1703(.085)	.2187(.045)
GPD growth	-.1953(.784)	-.1033(.132)	-.1700(.234)	-.0725(.150)	-.1564(.651)	-.0731(.136)
VIX	.1025(.049)	.1356(.055)	.0960(.047)	.1739(.112)	.0997(.047)	.1323(.056)
Reer	.0248(.012)	.0350(.018)	.0157(.008)	.0206(.019)	.0264(.012)	.0349(.018)
Y_{t-1}	.1249(.179)	.5906(.257)	.1678(.269)	.6806(.277)	.1142(.056)	.6183(.284)
$W_3 u_t$	.4727(.561)	.5666(.124)	.5693(.568)	.7804(.164)	.2588(.129)	.4958(.194)

6. Conclusion and Discussion

This paper considers the M-estimation and inference methods for the SDPD models with fixed effects and unknown heteroskedasticity, based on short panels. The estimation method extends the idea of Yang (2018) to allow for unknown heteroskedasticity by using modification terms that are quadratic in disturbances. The modified quasi-score function gives unbiased estimating equations. The statistical inferences are based on the *outer-product-of-martingale-differences* (OPMD) method proposed by Yang (2018). The asymptotic properties of the M-estimators and the estimators of VC matrix are studied. Monte Carlo experiments show that both the robust M-estimators and the estimators of standard errors perform very well and that ignoring the heteroskedasticity would cause significant bias. We apply our methods to investigate the international government risk spillover through both real linkage and information channels. The results show that allowing for heteroskedastic disturbances can be important. The proposed methods, therefore, provide a useful set of econometrics tools for applied researchers.

We have studied the case where the disturbances are heteroskedastic across individuals. It would be interesting to further extend our method to allow for heteroskedasticity in both individual and time, and for serial correlation. It would also be interesting to extend our method to allow for endogenous regressors, interactive fixed effects, time varying weight matrices and time varying spatial parameters. These models would be more challenging and are beyond the scope of this paper, and will be studied in future works.

Appendix A: Some Basic Lemmas

The following lemmas are essential for the proofs of the main results in this paper, where Lemmas A.4 and A.5 extend those of Yang(2018) by allowing v_{it} to be inid across i.

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 $O(\iota_n^{-1})$. Then*

(i) the sequence $\{A_nB_n\}$ 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_nC_n and C_nA_n are uniformly $O(\iota_n^{-1})$.

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

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

Lemma A.4. *(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 $O(\iota^{-1})$ uniformly in all i and j. Let v_n be a random n-vector of inid elements satisfying Assumption B, and b_n a constant n-vector of elements of uniform order $O(\iota^{-1/2})$. Then*

(i) $\mathrm{E}(v_n'A_nv_n) = O(\frac{n}{\iota_n})$, *(ii)* $\mathrm{Var}(v_n'A_nv_n) = O(\frac{n}{\iota_n})$,

(iii) $\mathrm{Var}(v_n'A_nv_n + b_n'v_n) = O(\frac{n}{\iota_n})$, *(iv)* $v_n'A_nv_n = O_p(\frac{n}{\iota_n})$,

(v) $v_n'A_nv_n - \mathrm{E}(v_n'A_nv_n) = O_p((\frac{n}{\iota_n})^{\frac{1}{2}})$, *(vi)* $v_n'A_nb_n = O_p((\frac{n}{\iota_n})^{\frac{1}{2}})$,

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 $O(\iota_n^{-1})$.

Lemma A.5. *(Yang, 2018, Central Limit Theorem for bilinear quadratic forms): Let $\{\Phi_n\}$ be a sequence of $n \times n$ matrices with row and column sums uniformly bounded, and elements of uniform order $O(\iota_n^{-1})$. Let v_n be a random n-vector satisfying Assumption B. 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 $O(\iota_n^{-1})$, (ii) $sup_iE|b_{ni}|^{2+\epsilon_0} < \infty$, (iii) $\frac{\iota_n}{n}\sum_{i=1}^n[\phi_{n,ii}(b_{ni} - \mathrm{E}b_{ni})] = o_p(1)$ where $\{\phi_{n,ii}\}$ are the diagonal elements of Φ_n, and (iv) $\frac{\iota_n}{n}\sum_{i=1}^n[b_{ni}^2 - \mathrm{E}(b_{ni}^2)] = o_p(1)$. Let $\mathcal{H}_n = \mathrm{diag}(h_{n1}, \ldots, h_{nn})$. Define the bilinear-quadratic form:*

$$Q_n = b_n'v_n + v_n'\Phi_nv_n - \sigma_v^2 tr(\Phi_n\mathcal{H}_n),$$

and let $\sigma_{Q_n}^2$ be the variance of Q_n. If $\lim_{n\to\infty} \iota_n^{1+2/\epsilon_0}/n = 0$ and $\{\frac{\iota_n}{n}\sigma_{Q_n}^2\}$ are bounded away from zero, then $Q_n/\sigma_{Q_n} \xrightarrow{d} N(0,1)$.

Appendix B: Proofs of Theoretical Results

Proof of Lemma 3.1: By (2.3), $\Delta y_t = \mathcal{B}_0 \Delta y_{t-1} + B_{10}^{-1}\Delta X_t + B_{10}^{-1}B_{30}^{-1}\Delta v_t, t = 2, \ldots, T$, backward substitution leads to $\mathrm{E}(\Delta y_t \Delta v_t') = -\sigma_{v0}^2(\mathcal{B}_0 - 2I_n)B_{10}^{-1}B_{30}^{-1}\mathcal{H}$, $t = 2, \ldots, T$, $\mathrm{E}(\Delta y_t \Delta v_{t+1}') = -\sigma_{v0}^2 B_{10}^{-1}B_{30}^{-1}\mathcal{H}$, $t = 1, \ldots, T-1$, and $\mathrm{E}(\Delta y_t \Delta v_s') = 0$ if $s \geq t+2$; and

$$\begin{aligned}
&\mathrm{E}(\Delta y_t \Delta v_s') = \mathcal{B}_0 \mathrm{E}(\Delta y_{t-1}\Delta v_s') = \mathcal{B}_0^2 \mathrm{E}(\Delta y_{t-2}\Delta v_s') \ = \ \cdots \\
&= \mathcal{B}_0^{t-s}\mathrm{E}(\Delta y_s \Delta v_s') + \mathcal{B}_0^{t-s-1}B_{10}^{-1}B_{30}^{-1}\mathrm{E}(\Delta v_{s+1}\Delta v_s') \\
&= \mathcal{B}_0^{t-s+1}\mathrm{E}(\Delta y_{s-1}\Delta v_s') + \mathcal{B}_0^{t-s}B_{10}^{-1}B_{30}^{-1}\mathrm{E}(\Delta v_s \Delta v_s') + \mathcal{B}_0^{t-s-1}B_{10}^{-1}B_{30}^{-1}\mathrm{E}(\Delta v_{s+1}\Delta v_s') \\
&= \mathcal{B}_0^{t-s+1}B_{10}^{-1}B_{30}^{-1}\mathrm{E}(\Delta v_{s-1}\Delta v_s') + \mathcal{B}_0^{t-s}B_{10}^{-1}B_{30}^{-1}\mathrm{E}(\Delta v_s \Delta v_s') + \mathcal{B}_0^{t-s-1}B_{10}^{-1}B_{30}^{-1}\mathrm{E}(\Delta v_{s+1}\Delta v_s') \\
&= -\mathcal{B}_0^{t-s+1}B_{10}^{-1}B_{30}^{-1}\sigma_{v0}^2\mathcal{H} + 2\mathcal{B}_0^{t-s}B_{10}^{-1}B_{30}^{-1}\sigma_{v0}^2\mathcal{H} - \mathcal{B}_0^{t-s-1}B_{10}^{-1}B_{30}^{-1}\sigma_{v0}^2\mathcal{H} \\
&= -\sigma_{v0}^2\mathcal{B}_0^{t-s-1}(\mathcal{B}_0 - I_n)^2 B_{10}^{-1}B_{30}^{-1}\mathcal{H},
\end{aligned}$$

if $s \leq t-1$. Summarizing above, we obtain the results of Lemma (3.1). ■

Proofs of the theorems need the following matrix results: (i) the eigenvalues of a projection matrix are either 0 or 1; (ii) the eigenvalues of a positive definite (p.d.) matrix are strictly positive; (iii) $\gamma_{\min}(A)\mathrm{tr}(B) \leq \mathrm{tr}(AB) \leq \gamma_{\max}(A)\mathrm{tr}(B)$ for symmetric matrix A and positive semidefinite (p.s.d.) matrix B; (iv) $\gamma_{\max}(A+B) \leq \gamma_{\max}(A) + \gamma_{\max}(B)$ for symmetric matrices A and B; and (v) $\gamma_{\max}(AB) \leq \gamma_{\max}(A)\gamma_{\max}(B)$ for p.s.d. matrices A and B (Bernstein, 2009).

Proof of Theorem 3.1: Use $\Delta\hat{\mathbf{u}}(\delta)$ defined below (2.6) and $\Delta\bar{\mathbf{u}}(\delta)$ defined below (3.23). Let $\mathbf{B}_r^* = \Omega^{-\frac{1}{2}}\mathbf{B}_r$, where $\Omega^{\frac{1}{2}}$ is the square-root matrix of Ω. We can write $\Delta\hat{\mathbf{u}}^*(\delta) = \Omega^{-\frac{1}{2}}\Delta\hat{\mathbf{u}}(\delta) = \mathbf{M}(\mathbf{B}_1^*\Delta Y - \mathbf{B}_2^*\Delta Y_{-1})$ and $\Delta\bar{\mathbf{u}}^*(\delta) = \Omega^{-\frac{1}{2}}\Delta\bar{\mathbf{u}}(\delta) = \mathbf{M}(\mathbf{B}_1^*\Delta Y - \mathbf{B}_2^*\Delta Y_{-1}) + \mathbf{P}[\mathbf{B}_1^*(\Delta Y - \mathrm{E}(\Delta Y)) - \mathbf{B}_2^*(\Delta Y_{-1} - \mathrm{E}(\Delta Y_{-1}))]$, where $\mathbf{P} = \Omega^{-\frac{1}{2}}\Delta X(\Delta X'\Omega^{-1}\Delta X)^{-1}\Delta X'\Omega^{-\frac{1}{2}}$, and $\mathbf{M} = I_{n(T-1)} - \mathbf{P}$. By (3.14) and (3.23), we have

$$\begin{aligned}
\hat{\sigma}_{v,\mathtt{M}}^2(\delta) &= \tfrac{1}{n(T-1)}(\mathbf{B}_1^*\Delta Y - \mathbf{B}_2^*\Delta Y_{-1})'\mathbf{M}(\mathbf{B}_1^*\Delta Y - \mathbf{B}_2^*\Delta Y_{-1}), \\
\bar{\sigma}_{v,\mathtt{M}}^2(\delta) &= \tfrac{1}{n(T-1)}\mathrm{tr}[\mathrm{Var}(\mathbf{B}_1^*\Delta Y - \mathbf{B}_2^*\Delta Y_{-1})] \\
&\quad + \tfrac{1}{n(T-1)}(\mathbf{B}_1^*\mathrm{E}\Delta Y - \mathbf{B}_2^*\mathrm{E}\Delta Y_{-1})'\mathbf{M}(\mathbf{B}_1^*\mathrm{E}\Delta Y - \mathbf{B}_2^*\mathrm{E}\Delta Y_{-1}).
\end{aligned}$$

The second term in $\bar{\sigma}_{v,\mathtt{M}}^2(\delta)$ is nonnegative uniformly in $\delta \in \Delta$ as $\mathbf{M}$ is p.s.d. For the first term, by the definition of the matrix C, Assumption E(iv) and the assumption (ii) given in the theorem, $\frac{1}{n(T-1)}\mathrm{tr}[\Omega^{-1}\mathrm{Var}(\mathbf{B}_1\Delta Y - \mathbf{B}_2\Delta Y_{-1})] \geq \frac{1}{n(T-1)}\gamma_{\min}(C^{-1})\gamma_{\min}(B_3'B_3)\mathrm{tr}[\mathrm{Var}(\mathbf{B}_1\Delta Y - \mathbf{B}_2\Delta Y_{-1})] > c > 0$, uniformly in $\delta \in \Delta$, implying $\inf_{\delta\in\Delta}\bar{\sigma}_{v,\mathtt{M}}^2(\delta) > c > 0$. It is easy to show that $\sup_{\delta\in\Delta}|\hat{\sigma}_{v,\mathtt{M}}^2(\delta) - \bar{\sigma}_{v,\mathtt{M}}^2(\delta)| = o_p(1)$. Therefore, we can drop $\hat{\sigma}_{v,\mathtt{M}}^2(\delta)$ in the concentrated AQS function (3.15) and $\bar{\sigma}_{v,\mathtt{M}}^2(\delta)$ in its population counter part (3.24) and write:

$$S_{\mathtt{STLE}}^{*c}(\delta) - \bar{S}_{\mathtt{STLE}}^{*c}(\delta) =$$
$$\begin{cases}
\Delta\hat{\mathbf{u}}(\delta)'\Omega^{-1}\Delta Y_{-1} - \mathrm{E}[\Delta\bar{\mathbf{u}}(\delta)'\Omega^{-1}\Delta Y_{-1}] + \Delta\hat{\mathbf{u}}'(\delta)\mathbf{E}_\rho\Delta\hat{\mathbf{u}}(\delta) - \mathrm{E}[\Delta\bar{\mathbf{u}}'(\delta)\mathbf{E}_\rho\Delta\bar{\mathbf{u}}(\delta)], \\
\Delta\hat{\mathbf{u}}(\delta)'\Omega^{-1}\mathbf{W}_1\Delta Y - \mathrm{E}[\Delta\bar{\mathbf{u}}(\delta)'\Omega^{-1}\mathbf{W}_1\Delta Y] + \Delta\hat{\mathbf{u}}(\delta)'\mathbf{E}_{\lambda_1}\Delta\hat{\mathbf{u}}(\delta) - \mathrm{E}[\Delta\bar{\mathbf{u}}(\delta)'\mathbf{E}_{\lambda_1}\Delta\bar{\mathbf{u}}(\delta)], \\
\Delta\hat{\mathbf{u}}(\delta)'\Omega^{-1}\mathbf{W}_2\Delta Y_{-1} - \mathrm{E}[\Delta\bar{\mathbf{u}}(\delta)'\Omega^{-1}\mathbf{W}_2\Delta Y_{-1}] + \Delta\hat{\mathbf{u}}(\delta)'\mathbf{E}_{\lambda_2}\Delta\hat{\mathbf{u}}(\delta) - \mathrm{E}[\Delta\bar{\mathbf{u}}(\delta)'\mathbf{E}_{\lambda_2}\Delta\bar{\mathbf{u}}(\delta)] \\
\Delta\hat{\mathbf{u}}(\delta)'\Upsilon\Delta\hat{\mathbf{u}}(\delta) - \mathrm{E}[\Delta\bar{\mathbf{u}}(\delta)'\Upsilon\Delta\bar{\mathbf{u}}(\delta)],
\end{cases}$$

where $\Upsilon = \frac{1}{2}[C^{-1} \otimes (\mathcal{A} - \mathbf{E}_{\lambda_3})]$. With Assumption G, consistency of $\hat{\delta}_{\mathtt{M}}$ follows from:

(**a**) $\sup_{\delta\in\Delta} \frac{1}{n(T-1)}\big|\Delta\hat{\mathbf{u}}(\delta)'\Omega^{-1}\Delta Y_{-1} - \mathrm{E}[\Delta\bar{\mathbf{u}}(\delta)'\Omega^{-1}\Delta Y_{-1}]\big| = o_p(1)$,

(**b**) $\sup_{\delta\in\Delta} \frac{1}{n(T-1)}\big|\Delta\hat{\mathbf{u}}(\delta)'\Omega^{-1}\mathbf{W}_1\Delta Y - \mathrm{E}[\Delta\bar{\mathbf{u}}(\delta)'\Omega^{-1}\mathbf{W}_1\Delta Y]\big| = o_p(1)$,

(**c**) $\sup_{\delta\in\Delta} \frac{1}{n(T-1)}\big|\Delta\hat{\mathbf{u}}(\delta)'\Omega^{-1}\mathbf{W}_2\Delta Y_{-1} - \mathrm{E}[\Delta\bar{\mathbf{u}}(\delta)'\Omega^{-1}\mathbf{W}_2\Delta Y_{-1}]\big| = o_p(1)$,

(**d**) $\sup_{\delta\in\Delta} \frac{1}{n(T-1)}\big|\Delta\hat{\mathbf{u}}'(\delta)\mathbf{E}_{\rho}\Delta\hat{\mathbf{u}}(\delta) - \mathrm{E}[\Delta\bar{\mathbf{u}}'(\delta)\mathbf{E}_{\rho}\Delta\bar{\mathbf{u}}(\delta)]\big| = o_p(1)$,

(**e**) $\sup_{\delta\in\Delta} \frac{1}{n(T-1)}\big|\Delta\hat{\mathbf{u}}'(\delta)\mathbf{E}_{\lambda_1}\Delta\hat{\mathbf{u}}(\delta) - \mathrm{E}[\Delta\bar{\mathbf{u}}'(\delta)\mathbf{E}_{\lambda_1}\Delta\bar{\mathbf{u}}(\delta)]\big| = o_p(1)$,

(**f**) $\sup_{\delta\in\Delta} \frac{1}{n(T-1)}\big|\Delta\hat{\mathbf{u}}'(\delta)\mathbf{E}_{\lambda_2}\Delta\hat{\mathbf{u}}(\delta) - \mathrm{E}[\Delta\bar{\mathbf{u}}'(\delta)\mathbf{E}_{\lambda_2}\Delta\bar{\mathbf{u}}(\delta)]\big| = o_p(1)$,

(**g**) $\sup_{\delta\in\Delta} \frac{1}{n(T-1)}\big|\Delta\hat{\mathbf{u}}(\delta)'\Upsilon\Delta\hat{\mathbf{u}}(\delta) - \mathrm{E}[\Delta\bar{\mathbf{u}}(\delta)'\Upsilon\Delta\bar{\mathbf{u}}(\delta)]\big| = o_p(1)$.

Proof of (a). By the expressions of $\Delta\hat{\mathbf{u}}^*(\delta)$ and $\Delta\bar{\mathbf{u}}^*(\delta)$ given above, we can write

$$\tfrac{1}{n(T-1)}\{\Delta\hat{\mathbf{u}}(\delta)'\Omega^{-1}\Delta Y_{-1} - \mathrm{E}[\Delta\bar{\mathbf{u}}(\delta)'\Omega^{-1}\Delta Y_{-1}]\} = (Q_1 - \mathrm{E}Q_1) + (Q_2 - \mathrm{E}Q_2) - Q_3 - Q_4, \quad \text{(C.1)}$$

where $Q_1 = \frac{1}{n(T-1)}\Delta Y'\mathbf{B}_1^{*\prime}\mathbf{M}\Omega^{-\frac{1}{2}}\mathbf{W}_1\Delta Y$, $\quad Q_2 = \frac{1}{n(T-1)}\Delta Y_{-1}'\mathbf{B}_2^{*\prime}\mathbf{M}\Omega^{-\frac{1}{2}}\mathbf{W}_1\Delta Y$,
$Q_3 = \frac{1}{n(T-1)}\mathrm{tr}[\mathbf{B}_1^{*\prime}\mathbf{P}\Omega^{-\frac{1}{2}}\mathbf{W}_1\mathrm{Var}(\Delta Y)]$, $Q_4 = \frac{1}{n(T-1)}\mathrm{tr}[\mathbf{B}_2^{*\prime}\mathbf{P}\Omega^{-\frac{1}{2}}\mathbf{W}_1\mathrm{Cov}(\Delta Y, \Delta Y_{-1}')]$.
Let $\mathbf{M}^* = \Omega^{-\frac{1}{2}}\mathbf{M}\Omega^{-\frac{1}{2}}$. Using (3.25), Q_1 can be decomposed into:

$$Q_1 = \tfrac{1}{n(T-1)}\big[\Delta\mathbf{y}_1'\mathbb{R}'\mathbf{B}_1'\mathbf{M}^*\mathbf{W}_1\mathbb{R}\Delta\mathbf{y}_1 + 2\Delta\mathbf{y}_1'\mathbb{R}'\mathbf{B}_1'\mathbf{M}^*\mathbf{W}_1\boldsymbol{\eta} + 2\Delta\mathbf{y}_1'\mathbb{R}'\mathbf{B}_1'\mathbf{M}^*\mathbf{W}_1\mathbb{S}\Delta\mathbf{v} + \boldsymbol{\eta}'\mathbf{B}_1'\mathbf{M}^*\mathbf{W}_1\boldsymbol{\eta} + 2\boldsymbol{\eta}'\mathbf{B}_1'\mathbf{M}^*\mathbf{W}_1\mathbb{S}\Delta\mathbf{v} + \Delta\mathbf{v}'\mathbb{S}'\mathbf{B}_1'\mathbf{M}^*\mathbf{W}_1\mathbb{S}\Delta\mathbf{v}\big] \equiv \textstyle\sum_{\ell=1}^{6} Q_{1,\ell},$$

and further using (3.26), Q_2 can be decomposed into:

$$Q_2 = \tfrac{1}{n(T-1)}\big[\Delta\mathbf{y}_1'\mathbb{R}_1'\mathbf{B}_2'\mathbf{M}^*\mathbf{W}_1\mathbb{R}\Delta\mathbf{y}_1 + \Delta\mathbf{y}_1'\mathbb{R}_1'\mathbf{B}_2'\mathbf{M}^*\mathbf{W}_1\boldsymbol{\eta} + \Delta\mathbf{y}_1'\mathbb{R}_1'\mathbf{B}_2'\mathbf{M}^*\mathbf{W}_1\mathbb{S}\Delta\mathbf{v} + \boldsymbol{\eta}_1'\mathbf{B}_2'\mathbf{M}^*\mathbf{W}_1\mathbb{R}\Delta\mathbf{y}_1 + \boldsymbol{\eta}_1'\mathbf{B}_2'\mathbf{M}^*\mathbf{W}_1\boldsymbol{\eta} + \boldsymbol{\eta}_1'\mathbf{B}_2'\mathbf{M}^*\mathbf{W}_1\mathbb{S}\Delta\mathbf{v} + \Delta\mathbf{v}'\mathbb{S}_1'\mathbf{B}_2'\mathbf{M}^*\mathbf{W}_1\mathbb{R}\Delta\mathbf{y}_1 + \Delta\mathbf{v}'\mathbb{S}_1'\mathbf{B}_2'\mathbf{M}^*\mathbf{W}_1\boldsymbol{\eta} + \Delta\mathbf{v}'\mathbb{S}_1'\mathbf{B}_2'\mathbf{M}^*\mathbf{W}_1\mathbb{S}\Delta\mathbf{v}\big] \equiv \textstyle\sum_{\ell=1}^{9} Q_{2,\ell}.$$

Thus, both Q_1 and Q_2 are sums of terms of the forms: $\frac{1}{n(T-1)}\Delta\mathbf{y}_1'\Xi\Delta\mathbf{y}_1$, $\frac{1}{n(T-1)}\Delta\mathbf{v}'\Phi\Delta\mathbf{v}$, $\frac{1}{n(T-1)}\Delta\mathbf{y}_1'\Psi\Delta\mathbf{v}$, $\frac{1}{n(T-1)}\Delta\mathbf{y}_1'\varphi$, and $\frac{1}{n(T-1)}\Delta\mathbf{v}'\phi$, where the matrices Ξ, Φ and Ψ, and the vectors ϕ and φ are multiplicative in terms of $\mathbb{R}$, $\mathbb{R}_{-1}$, $\mathbb{S}$, $\mathbb{S}_{-1}$, $\boldsymbol{\eta}$, $\boldsymbol{\eta}_{-1}$, $\mathbf{B}_1$, $\mathbf{B}_2$, $\mathbf{M}^*$. Note that $\mathbb{R}$, $\mathbb{R}_{-1}$, $\mathbb{S}$, $\mathbb{S}_{-1}$, $\boldsymbol{\eta}$ and $\boldsymbol{\eta}_{-1}$ depend on true parameter values, and they are uniformly bounded in both row and column sums by Assumption E (*iii*), Lemma A.1 and Lemma A.3. $\mathbf{B}_1$ depends on λ_1, $\mathbf{B}_2$ depends on ρ and λ_2, and $\mathbf{M}$ depends on λ_3. They are uniformly bounded in either row or column sums for each $\delta \in \mathbf{\Delta}$ by Assumption E (*iv*), Lemma A.1 and Lemma A.3. Therefore, by Lemma A.1, we have for each $\delta \in \mathbf{\Delta}$, Ξ, Φ and Ψ are uniformly bounded in either row or column sums, and the elements of ϕ and φ are of uniform order $O(\iota^{-\frac{1}{2}})$.

Note that Ξ, Φ, Ψ, ϕ and φ depend on δ in general. Partition them according to $t, s = 2, \ldots, T$, and denote the partitioned matrices/vectors by Ξ_{ts}, Φ_{ts}, Ψ_{ts}, ϕ_t and φ_t. First, for the terms quadratic in $\Delta\mathbf{y}_1$, they can be written as $\frac{1}{n}\Delta y_1'\Xi_{++}\Delta y_1$, where $\Xi_{++} = \frac{1}{T-1}\sum_t\sum_s \Xi_{t,s}$. As Ξ is uniformly bounded in either row or column sums, we have Ξ_{++} is uniformly bounded in either row or column sums for any $\delta \in \mathbf{\Delta}$. The pointwise convergence of $\frac{1}{n}[\Delta y_1'\Xi_{++}(\delta)\Delta y_1 -$

$\mathrm{E}(\Delta y_1'\Xi_{++}\Delta y_1)]$ thus follows from Assumption F(iii). Second, the terms quadratic in Δv can be written as $\frac{1}{n}\sum_{t,s} v_t'\Pi_{ts}v_s$. The pointwise convergence of $\frac{1}{n}[v_t'\Pi_{ts}v_s - \mathrm{E}(v_t'\Pi_{ts}v_s)]$ follows from Lemma A.4 for each $t, s = 1, \ldots, T$. Third, the pointwise convergence of bilinear terms $\frac{1}{n}[\Delta\mathbf{y}_1'\Psi\Delta\mathbf{v} - \mathrm{E}(\Delta\mathbf{y}_1'\Psi\Delta\mathbf{v})]$ follows by writing $\Delta\mathbf{y}_1'\Psi\Delta\mathbf{v} = \sum_s \Delta\mathbf{y}_1'\Psi_{+s}v_s$ and applying Lemma A.4 (v), Lemma A.4 (vi), and Assumption F (iv). Finally, the pointwise convergence of $\frac{1}{n(T-1)}[\Delta\mathbf{y}_1'\varphi - \mathrm{E}(\Delta\mathbf{y}_1'\varphi)]$ follows from Assumption F(ii), and that of $\frac{1}{n(T-1)}\Delta\mathbf{v}'\phi$ from Chebyshev inequality. Thus, $Q_{k,\ell} - \mathrm{E}(Q_{k,\ell}) \xrightarrow{p} 0$, for each $\delta \in \mathbf{\Delta}$, and all k and ℓ.

Now, use notation $Q_{k,\ell}(\delta)$ and let δ_1 and δ_2 be in $\mathbf{\Delta}$. We have by the mean value theorem:

$$Q_{k,\ell}(\delta_2) - Q_{k,\ell}(\delta_1) = \tfrac{\partial}{\partial\delta'}Q_{k,\ell}(\bar{\delta})(\delta_2 - \delta_1), \quad k = 1, 2, \text{ and all corresponding } \ell,$$

where $\bar{\delta}$ lies between δ_1 and δ_2 elementwise. It is easy to verify that $\sup_{\delta\in\mathbf{\Delta}} |\frac{\partial}{\partial\omega}Q_{k,\ell}(\delta)| = O_p(1)$ for $\omega = \rho, \lambda_1, \lambda_2$ as $Q_{k,\ell}(\delta)$ is linear or quadratic in ρ, λ_1 and λ_2, and thus the corresponding partial derivatives take simple forms. Only the matrix $\mathbf{M}^*$ involves λ_3 and its derivative is $\frac{d}{d\lambda_3}\mathbf{M}^* = \mathbf{M}^*\Omega(C^{-1}\otimes\mathcal{A})\Omega\mathbf{M}^*$. Take $Q_{1,1}$ for example, noting that $\gamma_{max}(\mathbf{M}) = 1$, we have by definition of matrix C, Assumption E(iii) and Assumption F(i),

$$\begin{aligned}
\sup_{\delta\in\mathbf{\Delta}} |\tfrac{\partial}{\partial\lambda_3}Q_{1,1}(\delta)| &= \sup_{\delta\in\mathbf{\Delta}} \tfrac{1}{n(T-1)}|\Delta\mathbf{y}_1'\mathbb{R}'\mathbf{B}_1'\mathbf{M}^*\Omega(C^{-1}\otimes\mathcal{A})\Omega\mathbf{M}^*\mathbf{W}_1\mathbb{R}\Delta\mathbf{y}_1| \\
&\le \sup_{\delta\in\mathbf{\Delta}} \tfrac{1}{n(T-1)}|\Delta\mathbf{y}_1'\mathbb{R}'\mathbf{B}_1'(C^{-1}\otimes\mathcal{A})\mathbf{W}_1\mathbb{R}\Delta\mathbf{y}_1| \\
&\le \gamma_{min}^{-1}(C)\gamma_{max}(\mathcal{A})\gamma_{max}(\mathbf{B}_1)\tfrac{1}{n(T-1)}|\Delta\mathbf{y}_1'\mathbb{R}'\mathbf{W}_1\mathbb{R}\Delta\mathbf{y}_1| \\
&= O(1) \times O(1) \times O(1) \times O_p(1).
\end{aligned}$$

The results $\sup_{\delta\in\mathbf{\Delta}} |\frac{\partial}{\partial\lambda_3}Q_{k,\ell}(\delta)| = O_p(1)$ can be proved similarly for all other cases.

It follows that $Q_{k,\ell}(\delta)$ are stochastic equicontinuous, and by Theorem 1 of Andrews (1992), $Q_{k,\ell}(\delta) - \mathrm{E}Q_{k,\ell}(\delta) \xrightarrow{p} 0$, uniformly in $\delta \in \mathbf{\Delta}$. Thus, $Q_k(\delta) - \mathrm{E}Q_k(\delta) \xrightarrow{p} 0$, uniformly in $\delta \in \mathbf{\Delta}$, $k = 1, 2$. It left to show that $Q_k(\delta) \to 0, k = 3, 4$, uniformly in $\delta \in \mathbf{\Delta}$. We have,

$$\begin{aligned}
Q_3 = \;& \tfrac{1}{n(T-1)}\mathrm{tr}[\mathbf{B}_1'\Omega^{-1}\Delta X(\Delta X'\Omega^{-1}\Delta X)^{-1}\Delta X'\Omega^{-1}\mathbf{W}_1\mathrm{Var}(\Delta Y)] \\
\le \;& \tfrac{1}{n(T-1)}\gamma_{\max}(\Omega^{-2})\gamma_{\max}(\mathbf{B}_1)\gamma_{\max}(\mathbf{W}_1)\gamma_{\min}^{-1}(\Delta X'\Omega^{-1}\Delta X)\ \mathrm{tr}[\Delta X'\mathrm{Var}(\Delta Y)\Delta X] \\
\le \;& \tfrac{1}{n(T-1)}\gamma_{\max}^2(\Omega^{-1})\gamma_{\max}(\mathbf{B}_1)\gamma_{\max}(\mathbf{W}_1)\gamma_{\max}(\mathrm{Var}(\Delta Y))\gamma_{\min}^{-1}\big(\tfrac{\Delta X'\Omega^{-1}\Delta X}{n(T-1)}\big)\ \tfrac{\mathrm{tr}[\Delta X'\Delta X]}{n(T-1)}.
\end{aligned}$$

Recall $\Omega^{-1} = C^{-1} \otimes B_3'B_3$. By Assumption E$(iv)$, we have, $0 < \underline{c}_w \le \inf_{\lambda_3\in\Lambda_3} \gamma_{\min}(\Omega^{-1}) \le \sup_{\lambda_3\in\Lambda_3} \gamma_{\max}(\Omega^{-1}) \le \bar{c}_w < \infty$. By Assumption D, $0 < \underline{c}_x \le \inf_{\lambda_3\in\Lambda_3} \gamma_{\min}(\Omega^{-1})\gamma_{\min}\big(\frac{\Delta X'\Delta X}{n(T-1)}\big) \le \gamma_{\min}\big(\frac{\Delta X'\Omega^{-1}\Delta X}{n(T-1)}\big) \le \gamma_{\max}\big(\frac{\Delta X'\Omega^{-1}\Delta X}{n(T-1)}\big) \le \sup_{\lambda_3\in\Lambda_3} \gamma_{\max}(\Omega^{-1})\gamma_{\max}\big(\frac{\Delta X'\Delta X}{n(T-1)}\big) \le \bar{c}_x < \infty$.
It follows that by the assumptions in Theorem 3.1 and Assumption D,

$$Q_3 \le \tfrac{1}{n(T-1)}\bar{c}_w^2\underline{c}_x\bar{c}_y\bar{c}_{b_1}\bar{c}_{w_1}\tfrac{1}{n(T-1)}\mathrm{tr}[\Delta X'\Delta X] = O(n^{-1}).$$

The convergence of Q_4 can be proved similarly. Therefore,

$$\tfrac{1}{n(T-1)}\{\Delta\hat{\mathbf{u}}(\delta)'\Omega^{-1}\Delta Y_{-1} - \mathrm{E}[\Delta\bar{\mathbf{u}}(\delta)'\Omega^{-1}\Delta Y_{-1}]\} \xrightarrow{p} 0, \text{ uniformly in } \delta \in \mathbf{\Delta},$$

completing the proof of **(a)**.

Proofs of (b)-(g). Using the expressions of $\Delta\hat{\mathbf{u}}(\delta)$ and $\Delta\bar{\mathbf{u}}(\delta)$ given earlier, all the quantities inside $|\cdot|$ can be expressed in forms similar to (C.1). Then, using the expressions of ΔY, and ΔY_1, all the quantities can be further decomposed into sums of terms linear, quadratic or bilinear in $\Delta\mathbf{v}$ and $\Delta\mathbf{y}_1$. The proofs of (b) to (g) thus follow that of (a). ■

Proof of Theorem 3.2: We have by the mean value theorem,

$$0 = \tfrac{1}{\sqrt{n(T-1)}} S^*_{\texttt{STLE}}(\hat{\psi}_{\texttt{STLE}}) = \tfrac{1}{\sqrt{n(T-1)}} S^*_{\texttt{STLE}}(\psi_0) + \big[\tfrac{1}{n(T-1)} \tfrac{\partial}{\partial\psi'} S^*_{\texttt{STLE}}(\bar{\psi})\big] \sqrt{n(T-1)}(\hat{\psi}_{\texttt{M}} - \psi_0),$$

where $\bar{\psi}$ lies elementwise between $\hat{\psi}_{\texttt{M}}$ and ψ_0. The result of the theorem follows if

- **(a)** $\tfrac{1}{\sqrt{n(T-1)}} S^*_{\texttt{STLE}}(\psi_0) \xrightarrow{D} N\big[0, \lim_{n\to\infty} \Gamma^*_{\texttt{STLE}}(\psi_0)\big]$,
- **(b)** $\tfrac{1}{n(T-1)}\big[\tfrac{\partial}{\partial\psi'} S^*_{\texttt{STLE}}(\bar{\psi}) - \tfrac{\partial}{\partial\psi'} S^*_{\texttt{STLE}}(\psi_0)\big] \xrightarrow{p} 0$, and
- **(c)** $\tfrac{1}{n(T-1)}\big[\tfrac{\partial}{\partial\psi'} S^*_{\texttt{STLE}}(\psi_0) - \mathrm{E}\big(\tfrac{\partial}{\partial\psi'} S^*_{\texttt{STLE}}(\psi_0)\big)\big] \xrightarrow{p} 0$.

Proof of (a). From (3.27), we see that $S^*_{\texttt{STLE}}(\psi_0)$ consists of three types of elements: $\Pi'\Delta\mathbf{v}$, $\Delta\mathbf{v}'\Phi\Delta\mathbf{v}$ and $\Delta\mathbf{v}'\Psi\Delta\mathbf{y}_1$, which can be written as

$$\Pi'\Delta\mathbf{v} = \textstyle\sum_{t=1}^T \Pi_t^{*\prime} v_t, \;\; \Delta\mathbf{v}\Phi\Delta\mathbf{v} = \sum_{t=1}^T\sum_{s=1}^T v_t'\Phi^*_{ts} v_s, \;\text{ and } \Delta\mathbf{v}'\Psi\Delta\mathbf{y}_1 = \sum_{t=1}^T v_t'\Psi^*_t \Delta y_1,$$

where Π^*_t, Φ^*_{ts} and Ψ^*_t are formed by the elements of the partitioned Π, Φ and Ψ, respectively. By (2.1), $y_1 = B_{10}^{-1}B_{20}y_0 + \eta_1 + B_{10}^{-1}B_{30}^{-1}v_1$, leading to $\sum_{t=1}^T v_t'\Psi^*_t\Delta y_1 = \sum_{t=1}^T v_t'\Psi^{**}_t y_0 + \sum_{t=1}^T v_t'\Psi^{*+}_t v_1$, for suitably defined non-stochastic quantities η_1, Ψ^{**}_t and Ψ^{*+}_t. These show that, for every non-zero $(p+5)\times 1$ vector of constants c, $c'S^*_{\texttt{STLE}}(\psi_0)$ can be expressed as

$$c'S^*_{\texttt{STLE}}(\psi_0) = \sum_{t=1}^T\sum_{s=1}^T v_t'A_{ts}v_s + \sum_{t=1}^T v_t'B_t v_1 + \sum_{t=1}^T v_t' g(y_0),$$

for suitably defined non-stochastic matrices A_{ts} and B_t, and the function $g(y_0)$ linear in y_0. As, $\{y_0, v_1, \ldots, v_T\}$ are independent, the asymptotic normality of $\tfrac{1}{\sqrt{n(T-1)}} c'S^*_{\texttt{STLE}}(\psi_0)$ follows from Lemma A.5. Finally, the Cramér-Wold devise leads to the joint asymptotic normality.

Proof of (b). The Hessian matrix, $H^*_{\texttt{STLE}}(\psi) = \tfrac{\partial}{\partial\psi'} S^*_{\texttt{STLE}}(\psi)$, has the elements:

$H^*_{\beta\beta} = -\tfrac{1}{\sigma_v^2}\Delta X'\Omega^{-1}\Delta X$, $\quad H^*_{\beta\sigma_v^2} = -\tfrac{1}{\sigma_v^4}\Delta X'\Omega^{-1}\Delta\mathbf{u}(\theta)$, $\quad H^*_{\beta\rho} = -\tfrac{1}{\sigma_v^2}\Delta X'\Omega^{-1}\Delta Y_{-1}$,

$H^*_{\beta\lambda_1} = -\tfrac{1}{\sigma_v^2}\Delta X'\Omega^{-1}\mathbf{W}_1\Delta Y$, $\quad H^*_{\beta\lambda_2} = -\tfrac{1}{\sigma_v^2}\Delta X'\Omega^{-1}\mathbf{W}_2\Delta Y_{-1}$, $\quad H^*_{\beta\lambda_3} = \tfrac{1}{\sigma_v^2}\Delta X'\dot{\Omega}^{-1}\Delta\mathbf{u}(\theta)$,

$H^*_{\sigma_v^2\lambda_3} = \tfrac{1}{2\sigma_v^4}\Delta u(\theta)'\dot{\Omega}^{-1}\Delta\mathbf{u}(\theta)$, $\quad H^*_{\sigma_v^2\rho} = -\tfrac{1}{\sigma_v^4}\Delta Y_{-1}'\Omega^{-1}\Delta\mathbf{u}(\theta)$, $\quad H^*_{\sigma_v^2\lambda_1} = -\tfrac{1}{\sigma_v^4}\Delta Y'\mathbf{W}_1'\Omega^{-1}\Delta\mathbf{u}(\theta)$,

$H^*_{\sigma_v^2\lambda_2} = -\tfrac{1}{\sigma_v^4}\Delta Y_{-1}'\mathbf{W}_2'\Omega^{-1}\Delta\mathbf{u}(\theta)$, $\quad H^*_{\sigma_v^2\sigma_v^2} = -\tfrac{1}{\sigma_v^6}\Delta\mathbf{u}(\theta)'\Omega^{-1}\Delta\mathbf{u}(\theta) + \tfrac{n(T-1)}{2\sigma_v^4}$,

$H^*_{\rho\lambda_3} = \tfrac{1}{\sigma_v^2}\Delta Y_{-1}'\dot{\Omega}^{-1}\Delta\mathbf{u}(\theta) + \tfrac{1}{\sigma_v^2}\Delta\mathbf{u}'(\theta)\dot{\mathbf{E}}_{\rho\lambda_3}\Delta\mathbf{u}(\theta)$,

$H^*_{\lambda_1\lambda_3} = \tfrac{1}{\sigma_v^2}\Delta Y'\mathbf{W}_1'\dot{\Omega}^{-1}\Delta\mathbf{u}(\theta) + \tfrac{1}{\sigma_v^2}\Delta\mathbf{u}'(\theta)\dot{\mathbf{E}}_{\lambda_1\lambda_3}\Delta\mathbf{u}(\theta)$,

$H^*_{\lambda_2\lambda_3} = \tfrac{1}{\sigma_v^2}\Delta Y_{-1}'\mathbf{W}_2'\dot{\Omega}^{-1}\Delta\mathbf{u}(\theta) + \tfrac{1}{\sigma_v^2}\Delta\mathbf{u}'(\theta)\dot{\mathbf{E}}_{\lambda_2\lambda_3}\Delta\mathbf{u}(\theta)$,

$H^*_{\lambda_3\lambda_3} = -\tfrac{1}{2\sigma_v^2}\Delta\mathbf{u}(\theta)'[C^{-1}\otimes(2W_3'W_3 + \dot{\mathbf{E}}_{\lambda_3\lambda_3})]\Delta\mathbf{u}(\theta)$,

$$
\begin{aligned}
H^*_{\rho\rho} &= -\tfrac{1}{\sigma_v^2}\Delta Y'_{-1}\Omega^{-1}\Delta Y_{-1} + \tfrac{1}{\sigma_v^2}[\Delta\mathbf{u}'(\theta)\dot{\mathbf{E}}_{\rho\rho}\Delta\mathbf{u}(\theta) - \Delta Y'_{-1}(\mathbf{E}_\rho + \mathbf{E}'_\rho)\Delta\mathbf{u}(\theta)],\\
H^*_{\rho\lambda_1} &= -\tfrac{1}{\sigma_v^2}\Delta Y'_{-1}\Omega^{-1}\mathbf{W}_1\Delta Y + \tfrac{1}{\sigma_v^2}[\Delta\mathbf{u}'(\theta)\dot{\mathbf{E}}_{\rho\lambda_1}\Delta\mathbf{u}(\theta) - \Delta Y'\mathbf{W}'_1(\mathbf{E}_\rho + \mathbf{E}'_\rho)\Delta\mathbf{u}(\theta)],\\
H^*_{\rho\lambda_2} &= -\tfrac{1}{\sigma_v^2}\Delta Y'_{-1}\Omega^{-1}\mathbf{W}_2\Delta Y_{-1} + \tfrac{1}{\sigma_v^2}[\Delta\mathbf{u}'(\theta)\dot{\mathbf{E}}_{\rho\lambda_2}\Delta\mathbf{u}(\theta) - \Delta Y'_{-1}\mathbf{W}'_2(\mathbf{E}_\rho + \mathbf{E}'_\rho)\Delta\mathbf{u}(\theta)],\\
H^*_{\lambda_1\lambda_1} &= -\tfrac{1}{\sigma_v^2}\Delta Y'\mathbf{W}'_1\Omega^{-1}\mathbf{W}_1\Delta Y + \tfrac{1}{\sigma_v^2}[\Delta\mathbf{u}'(\theta)\dot{\mathbf{E}}_{\lambda_1\lambda_1}\Delta\mathbf{u}(\theta) - \Delta Y'\mathbf{W}'_1(\mathbf{E}_{\lambda_1} + \mathbf{E}'_{\lambda_1})\Delta\mathbf{u}(\theta)],\\
H^*_{\lambda_1\lambda_2} &= -\tfrac{1}{\sigma_v^2}\Delta Y'\mathbf{W}'_1\Omega^{-1}\mathbf{W}_2\Delta Y_{-1} + \tfrac{1}{\sigma_v^2}[\Delta\mathbf{u}'(\theta)\dot{\mathbf{E}}_{\lambda_1\lambda_2}\Delta\mathbf{u}(\theta) - \Delta Y'_{-1}\mathbf{W}'_2(\mathbf{E}_{\lambda_1} + \mathbf{E}'_{\lambda_1})\Delta\mathbf{u}(\theta)],\\
H^*_{\lambda_2\lambda_2} &= -\tfrac{1}{\sigma_v^2}\Delta Y'_{-1}\mathbf{W}'_2\Omega^{-1}\mathbf{W}_2\Delta Y_{-1} + \tfrac{1}{\sigma_v^2}[\Delta\mathbf{u}'(\theta)\dot{\mathbf{E}}_{\lambda_2\lambda_2}\Delta\mathbf{u}(\theta) - \Delta Y'_{-1}\mathbf{W}'_2(\mathbf{E}_{\lambda_2} + \mathbf{E}'_{\lambda_2})\Delta\mathbf{u}(\theta)],
\end{aligned}
$$

where $\dot{\Omega}^{-1} = \frac{\partial}{\partial\lambda_3}\Omega^{-1}$, $\dot{\mathbf{E}}_{r,v} = \frac{\partial}{\partial v}\mathbf{E}_r$, $r, v = \rho, \lambda_1, \lambda_2, \lambda_3$, and

$$
\begin{aligned}
&\dot{\mathbf{E}}_{\rho\rho} = \Omega^{-1}\mathbf{C}^{-1}\dot{\mathbf{D}}_{-1,\rho}, \quad \dot{\mathbf{E}}_{\rho\lambda_1} = \Omega^{-1}\mathbf{C}^{-1}\dot{\mathbf{D}}_{-1,\lambda_1}, \quad \dot{\mathbf{E}}_{\rho\lambda_2} = \Omega^{-1}\mathbf{C}^{-1}\dot{\mathbf{D}}_{-1,\lambda_2}\\
&\dot{\mathbf{E}}_{\rho\lambda_3} = \dot{\Omega}^{-1}\mathbf{C}^{-1}\mathbf{D}_{-1}, \quad \dot{\mathbf{E}}_{\lambda_1\lambda_1} = \Omega^{-1}\mathbf{C}^{-1}\mathbf{W}_1\dot{\mathbf{D}}_{\lambda_1}, \quad \dot{\mathbf{E}}_{\lambda_1\lambda_2} = \Omega^{-1}\mathbf{C}^{-1}\mathbf{W}_1\dot{\mathbf{D}}_{\lambda_2},\\
&\dot{\mathbf{E}}_{\lambda_1\lambda_3} = \dot{\Omega}^{-1}\mathbf{C}^{-1}\mathbf{W}_1\mathbf{D}, \quad \dot{\mathbf{E}}_{\lambda_2\lambda_2} = \Omega^{-1}\mathbf{C}^{-1}\mathbf{W}_2\dot{\mathbf{D}}_{-1,\lambda_2}, \quad \dot{\mathbf{E}}_{\lambda_2\lambda_3} = \dot{\Omega}^{-1}\mathbf{C}^{-1}\mathbf{W}_2\mathbf{D}_{-1},\\
&\dot{\mathbf{E}}_{\lambda_3\lambda_3} = 2[B'_3\text{diag}(W_3B_3^{-1}W_3B_3^{-1}) - W'_3\text{diag}(W_3B_3^{-1})]\text{diag}^{-1}(B_3^{-1}) + 2B'_3\text{diag}(W_3B_3^{-1})d_{3\lambda_3},\\
&d_{3\lambda_3} = \tfrac{d}{d\lambda_3}\text{diag}^{-1}(B_3^{-1}) = -\text{diag}^{-1}(B_3^{-1})\text{diag}(B_3^{-1}W_3B_3^{-1})\text{diag}^{-1}(B_3^{-1}).
\end{aligned}
$$

Noting that $\sigma_v^r, r = 2,4,6$ appears in $H^*_{\texttt{STLE}}(\psi)$ multiplicatively, we have $\frac{1}{n(T-1)}H^*_{\texttt{STLE}}(\bar{\psi}) = \frac{1}{n(T-1)}H^*_{\texttt{STLE}}(\bar{\beta}, \sigma^2_{v0}, \bar{\rho}, \bar{\lambda}) + o_p(1)$, as $\bar{\sigma}_v^2 \xrightarrow{p} \sigma^2_{v0}$, $\bar{\sigma}_v^{-r} = \sigma_{v0}^{-r} + o_p(1)$. Therefore the proof of (**b**) is thus equivalent to the proof of

$$\tfrac{1}{n(T-1)}\left[H^*_{\texttt{STLE}}(\bar{\beta}, \sigma^2_{v0}, \bar{\rho}, \bar{\lambda}) - H^*_{\texttt{STLE}}(\psi_0)\right] \xrightarrow{p} 0.$$

Writing $\Delta\mathbf{u}(\theta) = \Delta\mathbf{u} - (\lambda_1 - \lambda_{10})\mathbf{W}_1\Delta Y - (\rho - \rho_0)\Delta Y_{-1} - (\lambda_2 - \lambda_{20})\mathbf{W}_2\Delta Y_{-1} - \Delta X(\beta - \beta_0)$, and $\Delta\mathbf{u} = \mathbf{B}_{30}^{-1}\mathbf{F}\mathbf{v}$, where $\mathbf{F}$ is the first-difference matrix. By expressions (3.26) and (3.25), we can represent all the random elements of $H^*_{\texttt{STLE}}(\phi)$ as linear combinations of terms

$$
\begin{aligned}
&\text{quadratic in } \mathbf{v}: && (\varpi - \varpi_0)^j(\omega - \omega_0)^k\mathbf{v}'\mathbb{A}\mathbf{G}(\delta)\mathbb{B}\mathbf{v},\\
&\text{quadratic in } \Delta\mathbf{y}_1: && (\varpi - \varpi_0)^j(\omega - \omega_0)^k\Delta\mathbf{y}'_1\mathbb{A}\mathbf{G}(\delta)\mathbb{B}\Delta\mathbf{y}_1,\\
&\text{linear in } \mathbf{v}: && (\varpi - \varpi_0)^j\mathbf{v}'\mathbb{A}\mathbf{G}(\delta)\mathbb{B}\mathbb{Z},\\
&\text{linear in } \Delta\mathbf{y}_1: && (\varpi - \varpi_0)^j\Delta\mathbf{y}'_1\mathbb{A}\mathbf{G}(\delta)\mathbb{B}\mathbb{Z},\\
&\text{bylinear in } \mathbf{v} \text{ and } \Delta\mathbf{y}_1: && (\varpi - \varpi_0)^j(\omega - \omega_0)^k\mathbf{v}'\mathbb{A}\mathbf{G}(\delta)\mathbb{B}\Delta\mathbf{y}_1,
\end{aligned}
$$

for $j, k = 0, 1$, $\varpi, \omega = \rho, \lambda_1, \lambda_2$, where $\mathbb{A}$ and $\mathbb{B}$ denote $n(T-1) \times n(T-1)$ nonstochastic matrices, and $\mathbb{Z}$ $n(T-1) \times k$ nonstochastic vector or matrices, free from parameters; and $\mathbf{G}(\delta)$ can be Ω^{-1}, $\dot{\Omega}^{-1}$, $\mathbf{D}, \mathbf{D}_1, \dot{\mathbf{D}}_{\lambda_1}, \dot{\mathbf{D}}_{\lambda_2}, \dot{\mathbf{D}}_{-1\rho}, \dot{\mathbf{D}}_{-1,\lambda_1}, \dot{\mathbf{D}}_{-1,\lambda_2}$, and $\dot{\mathbf{E}}_{\lambda_3,\lambda_3}$.

Take a quadratic term of $\mathbf{v}$ for example. We have by MVT,

$$
\begin{aligned}
&\tfrac{1}{nT}[\mathbf{v}'\mathbb{A}\mathbf{G}(\bar{\rho}, \bar{\lambda}')\mathbb{B}\mathbf{v} - \mathbf{v}'\mathbb{A}\mathbf{G}(\rho_0, \lambda'_0)\mathbb{B}\mathbf{v}]\\
= \;&\tfrac{\bar{\rho}-\rho_0}{nT}\mathbf{v}'\mathbb{A}\dot{\mathbf{G}}_{\rho^*}\mathbb{B}\mathbf{v} + \tfrac{\bar{\lambda}_1-\lambda_{10}}{nT}\mathbf{v}'\mathbb{A}\dot{\mathbf{G}}_{\lambda_1^*}\mathbb{B}\mathbf{v} + \tfrac{\bar{\lambda}_2-\lambda_{20}}{nT}\mathbf{v}'\mathbb{A}\dot{\mathbf{G}}_{\lambda_2^*}\mathbb{B}\mathbf{v} + \tfrac{\bar{\lambda}_3-\lambda_{30}}{nT}\mathbf{v}'\mathbb{A}\dot{\mathbf{G}}_{\lambda_3^*}\mathbb{B}\mathbf{v},
\end{aligned}
$$

where $\dot{\mathbf{G}}_\rho$ and $\dot{\mathbf{G}}_{\lambda_r}$ are the partial derivatives of $\mathbf{G}$ evaluated at δ^*, which lies between $\bar{\delta}$ and δ_0. From the expression of the Hessian matrix given earlier, we see that $\mathbf{G}$ is the multiplications and linear combinations of matrices B_r, $B_r^{-1}$and $W_r, r = 1,2,3$. Therefore, its partial derivatives evaluated at δ are the multiplications and linear combinations of B_r, $B_r^{-1}$and $W_r, r = 1,2,3$, and hence are uniformly bounded in both row and column sums for

δ in a neighbourhood of δ_0, by Assumption E(iv) and Lemma A.2. By applying Lemma A.4 (i) and using the consistency of $\hat{\psi}_{\mathtt{M}}$, we have $\frac{1}{nT}[\mathbf{v}'\mathbb{A}\mathbf{G}(\bar{\delta})\mathbb{B}\mathbf{v} - \mathbf{v}'\mathbb{A}\mathbf{G}(\delta_0)\mathbb{B}\mathbf{v}] \xrightarrow{p} 0$. The convergence of all other terms can be shown similarly by using Lemma A.4, Assumption F, and the consistency of $\hat{\psi}_{\mathtt{M}}$.

Proof of (c). First, for the terms involving only $\Delta\mathbf{u}$ (linear or quadratic), the results follows Lemma A.4(v)-(vi). Second, by the representations (3.25) and (3.26) all the terms involving ΔY and ΔY_{-1} can be written as sums of the terms linear in $\Delta\mathbf{y}$, quadratic in $\Delta\mathbf{y}$, bilinear in $\Delta\mathbf{y}$ and $\Delta\mathbf{v}$, or quadratic in $\Delta\mathbf{v}$. Thus, the results follow by repeatedly applying Lemma A.1, Lemma A.4, and Assumption F. ■

Proof of Theorem 3.3: First, the result $\Sigma^*_{\mathtt{STLE}}(\hat{\psi}_{\mathtt{M}}) - \Sigma^*_{\mathtt{STLE}}(\psi_0) \xrightarrow{p} 0$ is implied by the result **(b)** in the proof of Theorem 3.2. The result $\frac{1}{n(T-1)}\sum_{i=1}^n[\hat{g}_i\hat{g}_i' - \mathrm{E}(g_ig_i')] \xrightarrow{p} 0$ follows from $\frac{1}{n(T-1)}\sum_{i=1}^n[\hat{g}_i\hat{g}_i' - g_ig_i'] \xrightarrow{p} 0$ and $\frac{1}{n(T-1)}\sum_{i=1}^n[g_ig_i' - \mathrm{E}(g_ig_i')] \xrightarrow{p} 0$. The proof of the former is straightforward by applying MVT. We focus on the proof of the latter result. As the elements of $S^*_{\mathtt{STLE}}(\psi_0)$ are mixtures of terms of the forms $\Pi'\Delta\mathbf{v} = \sum_{i=1}^n g_{\Pi i}$, $\Delta\mathbf{v}'\Phi\Delta\mathbf{v} = \sum_{i=1}^n g_{\Phi i}$ and $\Delta\mathbf{v}'\Psi\Delta\mathbf{y}_1 = \sum_{i=1}^n g_{\Psi i}$, it suffices to show that

$$\tfrac{1}{n(T-1)}\textstyle\sum_{i=1}^n[g_{\omega i}g_{\varpi i}' - \mathrm{E}(g_{\omega i}g_{\varpi i}')] = o_p(1),\ \omega, \varpi = \Pi, \Phi, \Psi.$$

For each i, let $\Delta v_{i\cdot}$ and $\Delta\xi_{i\cdot}$ be vectors that pick the elements of $\{\Delta v_{it}\}$ and $\{\Delta\xi_{it}\}$ for $t = 2, \ldots, T$, and Δv_{i-} and Δy^*_{1i-} be vectors that pick the elements of $\{\Delta v_{it}\}$ and $\{\Delta y^*_{1it}\}$ for $t = 3, \ldots, T$. Then (3.28), (3.29) and (3.30) can be written as $g_{\Pi i} = \Pi_{i\cdot}'\Delta v_{i\cdot}$, $g_{\Phi i} = \Delta v_{i\cdot}'\Delta\xi_{i\cdot} + \Delta v_{i\cdot}'\Delta v^*_{i\cdot} - 1_{T-1}'d_{i\cdot}$, and $g_{\Psi i} = \Delta v_{2i}\Delta\zeta_i + \Theta_{ii}(\Delta v_{2i}\Delta y^\circ_{1i} + \sigma^2_{v0}h_{n,i}) + \Delta v_{i-}'\Delta y^*_{1i-}$.

Without loss of generality, assume Π_{it} is a scalar, then $\frac{1}{n(T-1)}\sum_{i=1}^n[g_{\Pi i}g_{\Pi i}' - \mathrm{E}(g_{\Pi i}g_{\Pi i}')]$ can be written as $\frac{1}{n(T-1)}\sum_{i=1}^n \Pi_{i\cdot}'(\Delta v_{i\cdot}\Delta v_{i\cdot}' - \sigma^2_{v0}h_{n,i}C)\Pi_{i\cdot} \equiv \frac{1}{n(T-1)}\sum_{i=1}^n U_{n,i}$, where C is defined in (2.5). The $U_{n,i}$'s are independent across i as $\Delta v_{i\cdot}$'s are. It is easy to verify that $\mathrm{Var}(U_{n,i}) \le K_u < \infty$. Then we have $\frac{1}{n(T-1)}\sum_{i=1}^n U_{n,i} \xrightarrow{p} 0$ by Chebyshev's WLLN.

For $g_{\Phi i}$, we can write $\frac{1}{n(T-1)}\sum_{i=1}^n[g^2_{\Phi i} - \mathrm{E}(g^2_{\Phi i})] \equiv \sum_{r=1}^5 H_r$, where

$$H_1 = \tfrac{1}{n(T-1)}\textstyle\sum_{i=1}^n \left\{(\Delta v_{i\cdot}'\Delta\xi_{i\cdot})^2 - \mathrm{E}[(\Delta v_{i\cdot}'\Delta\xi_{i\cdot})^2]\right\}, \quad H_2 = \tfrac{2}{n(T-1)}\textstyle\sum_{i=1}^n(\Delta v_{i\cdot}'\Delta\xi_{i\cdot})(\Delta v_{i\cdot}'\Delta v^*_{i\cdot}),$$
$$H_3 = \tfrac{1}{n(T-1)}\textstyle\sum_{i=1}^n \left\{(\Delta v_{i\cdot}'\Delta v^*_{i\cdot})^2 - \mathrm{E}[(\Delta v_{i\cdot}'\Delta v^*_{i\cdot})^2]\right\}, \quad H_4 = -\tfrac{2}{n(T-1)}\textstyle\sum_{i=1}^n(1_{T-1}'d_{i\cdot})(\Delta v_{i\cdot}'\Delta\xi_{i\cdot}),$$
$$H_5 = -\tfrac{2}{n(T-1)}\textstyle\sum_{i=1}^n \left\{(1_{T-1}'d_{i\cdot})[\Delta v_{i\cdot}'\Delta v^*_{i\cdot} - \mathrm{E}(\Delta v_{i\cdot}'\Delta v^*_{i\cdot})]\right\}.$$

The first term can be written as:

$$H_1 = \tfrac{1}{n(T-1)}\textstyle\sum_{i=1}^n[\Delta\xi_{i\cdot}'(\Delta v_{i\cdot}\Delta v_{i\cdot}' - \sigma^2_{v0}h_{n,i}C)\Delta\xi_{i\cdot}] + \tfrac{\sigma^2_{v0}}{n(T-1)}\textstyle\sum_{i=1}^n[\Delta\xi_{i\cdot}'Ch_{n,i}\Delta\xi_{i\cdot} - \mathrm{E}(\Delta\xi_{i\cdot}'Ch_{n,i}\Delta\xi_{i\cdot})].$$

Let $V_{n,i} = \Delta\xi_{i\cdot}'(\Delta v_{i\cdot}\Delta v_{i\cdot}' - \sigma^2_{v0}h_{n,i}C)\Delta\xi_{i\cdot}$. We have $\mathrm{E}(V_{n,i}|\mathcal{G}_{n,i-1}) = 0$ as $\Delta\xi_{i\cdot}$ is $\mathcal{G}_{n,i-1}$-measurable. So, $\{V_{n,i}, \mathcal{G}_{n,i}\}$ form an M.D. array. It is easy to see that $\mathrm{E}|V_{n,i}^{1+\epsilon}| \le K_v < \infty$, for some $\epsilon > 0$. Thus, $\{V_{n,i}\}$ is uniformly integrable. The other two conditions of the WLLN for M.D. arrays of Davidson are satisfied. Therefore, $\frac{1}{n(T-1)}\sum_{i=1}^n V_{n,i} \xrightarrow{p} 0$.

For the second term of H_1, we can write $\Delta\xi_{i\cdot}' Ch_{n,i}\Delta\xi_{i\cdot} = \sum_s\sum_t \Delta\xi_{it}^2 C_{ts}h_{n,i}$ where C_{ts} is the (t,s) element of C. Recall that $\Delta\xi_t = \sum_{s=2}^T(\Phi_{st}^{u\prime}+\Phi_{ts}^l)\Delta v_s$, so we have,

$$\Delta\xi_{it} = \sum_{s=2}^T\sum_{j=1}^{i-1}(\Phi_{js,it}+\Phi_{it,js})\Delta v_{js} = \sum_{j=1}^{i-1}\sum_{s=2}^T(\Phi_{js,it}+\Phi_{it,js})\Delta v_{js} = \sum_{j=1}^{i-1}\phi_{ijt}'\Delta v_{j\cdot},$$

where $\phi_{ijt} = (\Phi_{j\cdot,it}+\Phi_{it,j\cdot})$ and $\Phi_{it,j\cdot}$ is the $(T-1)\times 1$ subvector that picks up the element from the it th row corresponding to $s=2,\ldots,T$. Thus we can write,

$$\begin{aligned}\frac{1}{n(T-1)}\sum_{i=1}^n\{(\Delta\xi_{it})^2 - \mathrm{E}[(\Delta\xi_{it})^2]\} = &\ \frac{1}{n(T-1)}\sum_{i=1}^n\sum_{j=1}^{i-1}[\phi_{ijt}'(\Delta v_{j\cdot}\Delta v_{j\cdot}' - \sigma_{v0}^2 h_{n,i}C)\phi_{ijt}] \\ &+2\frac{1}{n(T-1)}\sum_{j=1}^{n-1}\Delta v_{j\cdot}'\{\sum_{i=j+1}^n\sum_{k=1}^{j-1}\phi_{ijt}\phi_{ikt}'\Delta v_{k\cdot}\}.\end{aligned}$$

The first term is the 'average' of $n-1$ independent terms. $\{\sum_{i=j+1}^n\sum_{k=1}^{j-1}\phi_{ijt}\phi_{ikt}'\Delta v_{k\cdot}\}$ is $G_{n,j-1}$-measurable so the second term is the 'average' of an M.D. array. Conditions of Theorem 19.7 of Davidson (1994) are easily verified, and hence $\frac{1}{n(T-1)}\sum_{i=1}^n\{(\Delta\xi_{it})^2 - \mathrm{E}[(\Delta\xi_{it})^2]\} = o_p(1)$. Similarly, it can be shown that $\frac{1}{n(T-1)}\sum_{i=1}^n\{\Delta\xi_{it}\Delta\xi_{is} - \mathrm{E}[(\Delta\xi_{it}\Delta\xi_{is})]\} = o_p(1)$ for $s\neq t$. By the definition of C and Assumption B(i), C_{ts} and $h_{n,i}$ are uniformly bounded. Thus, $\frac{\sigma_{v0}^2}{n(T-1)}\sum_{i=1}^n[\Delta\xi_{i\cdot}'h_{n,i}C\Delta\xi_{i\cdot} - \mathrm{E}(\Delta\xi_{i\cdot}'h_{n,i}C\Delta\xi_{i\cdot})] = o_p(1)$, and $H_1 = o_p(1)$. The proofs for H_2 to H_5 can be done in a similar manner as the proof for H_1.

Finally, for $g_{\Psi i}$, we have,

$$\begin{aligned}&\frac{1}{n(T-1)}\sum_{i=1}^n[g_{\Psi i}^2 - \mathrm{E}(g_{\Psi i}^2)] \\ =\ &\frac{1}{n(T-1)}\sum_{i=1}^n[(\Delta v_{2i}^2 - 2\sigma_{v0}^2h_{n,i})\Delta\zeta_i^2] + \frac{2\sigma_{v0}^2}{n(T-1)}\sum_{i=1}^n h_{n,i}\Theta_{ii}\Delta v_{2i}\Delta\zeta_i \\ &+\frac{2\sigma_{v0}^2}{n(T-1)}\sum_{i=1}^n h_{n,i}\Theta_{ii}^2[\Delta v_{2i}\Delta y_{1i}^\circ - \mathrm{E}(\Delta v_{2i}\Delta y_{1i}^\circ)] \\ &+\frac{1}{n(T-1)}\sum_{i=1}^n\Theta_{ii}^2[(\Delta v_{2i}\Delta y_{1i}^\circ)^2 - \mathrm{E}((\Delta v_{2i}\Delta y_{1i}^\circ)^2)] \\ &+\frac{2}{n(T-1)}\sum_{i=1}^n\Theta_{ii}[\Delta v_{2i}^2\Delta\zeta_i\Delta y_{1i}^\circ - \mathrm{E}(\Delta v_{2i}^2\Delta\zeta_i\Delta y_{1i}^\circ)] + \frac{2\sigma_{v0}^2}{n(T-1)}\sum_{i=1}^n h_{n,i}[\Delta\zeta_i^2 - \mathrm{E}(\Delta\zeta_i^2)] \\ &+\frac{1}{n(T-1)}\sum_{i=1}^n[(\Delta v_{i-}'\Delta y_{1i-}^*)^2 - \mathrm{E}((\Delta v_{i-}'\Delta y_{1i-}^*)^2)] \\ &+\frac{2\sigma_{v0}^2}{n(T-1)}\sum_{i=1}^n h_{n,i}\Theta_{ii}[\Delta v_{i-}'\Delta y_{1i-}^* - \mathrm{E}(\Delta v_{i-}'\Delta y_{1i-}^*)] \\ &+\frac{2}{n(T-1)}\sum_{i=1}^n[\Delta v_{2i}\Delta\zeta_i(\Delta v_{i-}'\Delta y_{1i-}^*) - \mathrm{E}(\Delta v_{2i}\Delta\zeta_i(\Delta v_{i-}'\Delta y_{1i-}^*))] \\ &+\frac{2}{n(T-1)}\sum_{i=1}^n\Theta_{ii}[(\Delta v_{2i}\Delta y_{1i}^\circ)(\Delta v_{i-}'\Delta y_{1i-}^*) - \mathrm{E}((\Delta v_{2i}\Delta y_{1i}^\circ)(\Delta v_{i-}'\Delta y_{1i-}^*))] \equiv \sum_{r=1}^{10}Q_r.\end{aligned}$$

As $\{\Delta v_{2i}\Delta\zeta_i, \mathcal{F}_{n,i-1}\}$ form an M.D sequence, the convergence of Q_1 and Q_2 immediately follow from WLLN for M.D. arrays. The convergence of Q_3, Q_4 and Q_5 can be proved by using the expression $\Delta y_1^\circ = B_{30}B_{10}\Delta y_0 + B_{30}\Delta x_1\beta_0 + \Delta v_1$, and Lemma A.4. Recall that $\Delta\zeta = (\Theta^u+\Theta^l)\Delta y_1^\circ = (\Theta^u+\Theta^l)B_{30}B_{10}\Delta y_1$. Then we can write $Q_6 = \frac{2\sigma_{v0}^2}{n(T-1)}[\Delta y_1'A\Delta y_1 - \mathrm{E}(\Delta y_1'A\Delta y_1)]$, where $A = [(\Theta^u+\Theta^l)B_{30}B_{10}]'\mathcal{H}[(\Theta^u+\Theta^l)B_{30}B_{10}]$. By Assumption E and Lemma 3.1, A is uniformly bounded in both row and column sums. Therefore we have $Q_6 = o_p(1)$ by Assumption F. The results for Q_7 and Q_8 are proved by the independence between $\Delta v_{i-}'$ and Δy_{1i-}^* and Assumption F. Finally, the results for Q_9 and Q_{10} can be proved by further writing $\Delta y_{1t}^* = \Phi_{t+}\Delta y_1 = \Phi_{t+}(B_{30}B_{10})^{-1}\Delta y_1^\circ \equiv q(\Delta y_0, v_0) + \Phi_{t+}(B_{30}B_{10})^{-1}v_1$ and using Assumption F and Lemma A.4.

Subsequently, the cross-product terms: $\frac{1}{n(T-1)}\sum_{i=1}^n[g_{\Pi i}g_{\Phi i} - \mathrm{E}(g_{\Pi i}g_{\Phi i})]$,

$\frac{1}{n(T-1)}\sum_{i=1}^{n}[g_{\Pi i}g_{\Psi i}-\mathrm{E}(g_{\Pi i}g_{\Psi i})]$, and $\frac{1}{n(T-1)}\sum_{i=1}^{n}[g_{\Phi i}g_{\Psi i}-\mathrm{E}(g_{\Phi i}g_{\Psi i})]$, can all be decomposed in a similar manner, and the convergence of each of the decomposed terms can be proved in a similar way. ■

References

[1] Andrews, D. W. K., 1992. Generic uniform convergence. *Econometric Theory* 8, 241-257.

[2] Anselin, L., 1988. *Spatial Econometrics: Methods and Models.* The Netherlands: Kluwer Academic Press.

[3] Anselin, L., 2003. Spatial externalities, spatial multipliers, and spatial econometrics. *International Regional Science Review* 26, 153-166.

[4] Bai, J. and Li, K., 2018. Spatial panel data models with common shocks.

[5] Beirne, J. and Fratzscher, M., 2013. The pricing of sovereign risk and contagion during the European sovereign debt crisis. J*Journal of International Money and Finance*, 34, pp.60-82.

[6] Bernstein, D. S., 2009. *Matrix Mathematics: Theory, Facts, and Formulas. 2nd edition.* Princeton University Press, Princeton.

[7] Blasques, F., Koopman, S. J., Lucas, A. and Schaumburg, J., 2016. Spillover dynamics for systemic risk measurement using spatial financial time series models. *Journal of Econometrics* 195, 211-223.

[8] Breitung, J. and Wigger, C., 2018. Alternative GMM estimators for spatial regression models. *Spatial Economic Analysis*, 13, 148-170.

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

[10] Debarsy, N., Dossougoin, C., Ertur, C. and Gnabo, J.Y., 2018. Measuring sovereign risk spillovers and assessing the role of transmission channels: A spatial econometrics approach. *Journal of Economic Dynamics and Control* 87, 21-45.

[11] Eder, A. and Keiler, S., 2015. CDS Spreads and Contagion Amongst Systemically Important Financial InstitutionsA Spatial Econometric Approach. *International Journal of Finance & Economics* 20, 291-309. Vancouver

[12] Elhorst, J.P., 2010. Dynamic panels with endogenous interaction effects when T is small. *Regional Science and Urban Economics*, 40, 272-282.

[13] Elhorst, J. P., 2012. Dynamic spatial panels: models, methods, and applications. *Journal of Geographical Systems* 14, 5-28.

[14] Favero, C.A., 2013. Modelling and forecasting government bond spreads in the euro area: a GVAR model. *Journal of Econometrics* 177, 343-356.

[15] Hahn, J., Kuersteiner, G., 2002. Asymptotically unbiased inference for a dynamic model with fixed effects when both n and T are large. *Econometrica* 70, 1639-1657.

[16] Halleck Vega, S. and Elhorst, J. P., 2015. The SLX model. *Journal of Regional Science* 55, 339-363.

[17] Hsiao, C., Pesaran, M.H., Tahmiscioglu, A.K., 2002. Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. *Journal of Econometrics* 109, 107-150.

[18] Kelejian, H.H., Prucha, I.R., 1999. A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model. International Economic Review 40, 509533.

[19] Kelejian, H.H. and Prucha, I.R., 2010. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics* 157, 53-67.

[20] Korniotis, G. M., 2010. Estimating panel models with internal and external habit formation. *Journal of Business and Economic Statistics* 28, 145-158.

[21] Kuersteiner, G.M. and Prucha, I.R., 2018. Dynamic spatial panel models: Networks, common shocks, and sequential exogeneity. arXiv preprint arXiv:1802.01755.

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

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

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

[25] Lee, L. F., Yu, J., 2016. Identification of spatial Durbin panel models. *Journal of Applied Econometrics* 31, 133-162.

[26] Lin, X. and Lee, L.F., 2010. GMM estimation of spatial autoregressive models with unknown heteroskedasticity. *Journal of Econometrics* 157, 34-52.

[27] Liu, S.F. and Yang, Z., 2015. Modified QML estimation of spatial autoregressive models with unknown heteroskedasticity and nonnormality. *Regional Science and Urban Economics* 52, 50-70.

[28] Moscone, F. and Tosetti, E., 2011. GMM estimation of spatial panels with fixed effects and unknown heteroskedasticity. *Regional Science and Urban Economics* 41, 487-497.

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

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

[31] Saldías, M., 2013. A market-based approach to sector risk determinants and transmission in the euro area. *Journal of Banking & Finance* 37, 4534-4555.

[32] Shi, W. and Lee, L.F., 2017. Spatial dynamic panel data models with interactive fixed effects. J *Journal of Econometrics*, 197(2), pp.323-347.

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

[34] Taspınar, S., Doğan, O. and Bera, A. K., 2019. Heteroskedasticity-consistent covariance matrix estimators for spatial autoregressive models. *Spatial Economic Analysis*, 14, 241-268.

[35] Tonzer, L., 2015. Cross-border interbank networks, banking risk and contagion. *Journal of Financial Stability* 18, 19-32. Vancouver

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

[37] White, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4), pp.817-838.

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

[39] Yang, Z.L., 2010. A robust LM test for spatial error components. *Regional Science and Urban Economics*, 40(5), pp.299-310.

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

[41] 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 1a. Empirical Mean(sd)[rse]* of CQMLE, M-estimator, and Robust M-estimator, **DGP1**, $T = 3$, $m = 10$, $\mathcal{H}$=H-I, SNR=1 ;

		Normal Error			Normal Mixture			Chi-Square		
n	ψ	CQMLE	M-Est	RM-Est	CQMLE	M-Est	RM-Est	CQMLE	M-Est	RM-Est
50	1	.9631(.081)	.9482(.076)	.9876(.053)[.090]	.9658(.084)	.9507(.080)	.9849(.058)[.159]	.9634(.084)	.9491(.079)	.9844(.061)[.085]
	1	.7363(.239)	.7630(.249)	.8747(.261)[.422]	.7199(.388)	.7488(.417)	.8754(.546)[1.281]	.7229(.304)	.7520(.324)	.8670(.397)[.625]
	.3	.2489(.066)	.3305(.073)	.3193(.078)[.151]	.2535(.072)	.3330(.080)	.3249(.108)[.388]	.2540(.075)	.3344(.086)	.3237(.109)[.191]
	.2	-.1749(.669)	-.1536(.639)	.1180(.317)[.540]	-.1544(.663)	-.1374(.635)	.1001(.350)[.934]	-.1650(.664)	-.1393(.632)	.1011(.346)[.496]
	.2	.3671(.413)	.4330(.524)	.3003(.398)[.625]	.3618(.405)	.4289(.523)	.3116(.427)[1.199]	.3767(.427)	.4407(.542)	.3223(.423)[.615]
	.2	.1244(.611)	.1123(.633)	.0856(.446)[.328]	.1325(.584)	.1229(.607)	.0983(.437)[.478]	.1034(.608)	.0897(.629)	.0728(.442)[.360]
100	1	.9415(.079)	.9303(.076)	.9911(.041)[.047]	.9423(.082)	.9312(.079)	.9901(.045)[.058]	.9417(.079)	.9305(.076)	.9932(.042)[.045]
	1	.7881(.181)	.8110(.187)	.9361(.188)[.211]	.7814(.302)	.8066(.319)	.9493(.384)[.392]	.7790(.246)	.8026(.257)	.9379(.297)[.257]
	.3	.2474(.040)	.3229(.043)	.3036(.047)[.055]	.2498(.045)	.3247(.046)	.3084(.069)[.101]	.2497(.040)	.3249(.041)	.3058(.055)[.063]
	.2	-.2755(.599)	-.2416(.602)	.1567(.222)[.243]	-.2683(.617)	-.2325(.620)	.1520(.247)[.293]	-.2865(.613)	-.2529(.615)	.1519(.241)[.221]
	.2	.4208(.287)	.4464(.368)	.2272(.171)[.174]	.4149(.293)	.4388(.376)	.2261(.197)[.217]	.4215(.291)	.4484(.373)	.2273(.190)[.172]
	.2	.3309(.475)	.2954(.508)	.1471(.275)[.162]	.3310(.467)	.2948(.501)	.1499(.273)[.161]	.3378(.475)	.3024(.510)	.1459(.276)[.160]
200	1	.9606(.062)	.9479(.063)	.9992(.030)[.029]	.9609(.067)	.9485(.069)	.9985(.030)[.030]	.9614(.064)	.9490(.066)	.9987(.029)[.032]
	1	.8614(.150)	.8845(.159)	.9806(.138)[.145]	.8517(.236)	.8756(.251)	.9854(.284)[.226]	.8490(.189)	.8720(.200)	.9712(.205)[.189]
	.3	.2437(.028)	.3186(.033)	.3023(.033)[.036]	.2441(.032)	.3180(.037)	.3044(.049)[.053]	.2434(.030)	.3169(.035)	.3025(.039)[.049]
	.2	-.1878(.453)	-.1597(.478)	.1915(.098)[.092]	-.1759(.482)	-.1469(.503)	.1961(.113)[.100]	-.1829(.461)	-.1548(.484)	.1877(.105)[.113]
	.2	.3342(.189)	.3862(.278)	.2090(.094)[.091]	.3289(.197)	.3790(.289)	.2071(.104)[.094]	.3316(.190)	.3822(.278)	.2086(.101)[.115]
	.2	.4064(.318)	.3785(.351)	.1816(.148)[.112]	.3910(.325)	.3625(.358)	.1777(.159)[.109]	.4041(.318)	.3770(.351)	.1871(.156)[.110]
400	1	.9914(.035)	.9853(.035)	.9996(.017)[.017]	.9871(.042)	.9820(.041)	.9994(.017)[.016]	.9893(.039)	.9839(.038)	1.0002(.017)[.017]
	1	.9321(.106)	.9563(.111)	.9894(.095)[.103]	.9162(.182)	.9419(.191)	.9845(.195)[.161]	.9198(.143)	.9448(.150)	.9843(.146)[.130]
	.3	.2345(.020)	.3073(.024)	.2999(.021)[.024]	.2377(.024)	.3097(.028)	.3006(.033)[.037]	.2372(.022)	.3096(.026)	.3024(.027)[.030]
	.2	.0709(.262)	.0735(.265)	.1970(.051)[.052]	.0463(.317)	.0548(.310)	.1964(.054)[.052]	.0542(.294)	.0599(.291)	.1978(.054)[.052]
	.2	.2450(.138)	.2852(.202)	.2042(.072)[.072]	.2548(.163)	.2945(.228)	.2047(.077)[.072]	.2520(.152)	.2944(.218)	.2064(.077)[.072]
	.2	.2912(.215)	.2957(.230)	.1875(.105)[.081]	.3000(.228)	.3013(.239)	.1882(.104)[.078]	.3044(.221)	.3081(.233)	.1927(.102)[.078]

Note: 1. $\psi = (\beta, \sigma_v^2, \rho, \lambda_1, \lambda_2, \lambda_3)'$; 2. Variances increase and then decrease with group size; 3. W is generated according to fixed group scheme;
4. X_t values are generated with $\theta_x = (g, \phi_1, \phi_2, \sigma_1, \sigma_2) = (.01, .5, .5, 1, 1)$. * [rse]: empirical average of rses, only for robust M-estimator.

Table 1b. Empirical Mean(sd)[rse]* of CQMLE, M-estimator, and Robust M-estimator, **DGP1**, $T = 3$, $m = 10$, $\mathcal{H}$=H-I, SNR=3 ;

		Normal Error			Normal Mixture			Chi-Square		
n	ψ	CQMLE	M-Est	RM-Est	CQMLE	M-Est	RM-Est	CQMLE	M-Est	RM-Est
50	1	.9890(.048)	.9804(.047)	.9960(.034)[.034]	.9828(.053)	.9747(.051)	.9912(.036)[.048]	.9866(.052)	.9784(.050)	.9938(.038)[.083]
	1	.8246(.240)	.8373(.245)	.8873(.252)[.260]	.8195(.430)	.8360(.447)	.9071(.506)[.412]	.8178(.338)	.8323(.350)	.8883(.380)[.456]
	.3	.2721(.044)	.3123(.047)	.3068(.047)[.050]	.2763(.047)	.3169(.050)	.3119(.057)[.080]	.2742(.048)	.3141(.052)	.3086(.060)[.102]
	.2	.0567(.377)	.0572(.364)	.1707(.171)[.148]	.0249(.432)	.0297(.414)	.1513(.215)[.252]	.0456(.402)	.0488(.389)	.1600(.202)[.514]
	.2	.2703(.263)	.2987(.309)	.2361(.218)[.201]	.2959(.293)	.3262(.348)	.2613(.258)[.337]	.2819(.277)	.3094(.325)	.2518(.233)[.398]
	.2	.0916(.493)	.0932(.507)	.0660(.372)[.240]	.1106(.485)	.1108(.498)	.0820(.371)[.254]	.0996(.477)	.0997(.490)	.0743(.361)[.352]
100	1	.9854(.040)	.9800(.039)	.9962(.027)[.027]	.9828(.045)	.9774(.045)	.9963(.028)[.028]	.9845(.043)	.9790(.043)	.9975(.027)[.027]
	1	.8959(.176)	.9095(.181)	.9471(.181)[.183]	.8788(.330)	.8929(.341)	.9454(.371)[.267]	.8919(.262)	.9059(.270)	.9520(.288)[.224]
	.3	.2705(.025)	.3068(.026)	.3022(.028)[.028]	.2712(.028)	.3071(.028)	.3018(.037)[.036]	.2706(.026)	.3071(.026)	.3014(.029)[.030]
	.2	.0472(.302)	.0673(.301)	.1795(.135)[.121]	.0305(.339)	.0488(.338)	.1796(.140)[.130]	.0355(.328)	.0546(.327)	.1807(.142)[.128]
	.2	.2794(.156)	.2727(.181)	.2116(.095)[.087]	.2855(.174)	.2830(.205)	.2102(.096)[.092]	.2809(.169)	.2779(.199)	.2072(.101)[.091]
	.2	.2302(.345)	.2091(.355)	.1348(.225)[.159]	.2503(.346)	.2312(.357)	.1478(.225)[.171]	.2396(.352)	.2196(.363)	.1406(.230)[.154]
200	1	.9988(.023)	.9937(.023)	.9993(.019)[.019]	.9991(.025)	.9940(.026)	.9997(.020)[.019]	.9990(.023)	.9939(.024)	.9995(.019)[.019]
	1	.9499(.131)	.9635(.135)	.9789(.135)[.135]	.9485(.257)	.9627(.265)	.9815(.272)[.202]	.9372(.185)	.9507(.190)	.9665(.194)[.164]
	.3	.2676(.017)	.3021(.017)	.3005(.019)[.019]	.2675(.019)	.3018(.019)	.3008(.025)[.025]	.2683(.018)	.3022(.018)	.3008(.021)[.022]
	.2	.1288(.137)	.1493(.140)	.1949(.060)[.057]	.1325(.153)	.1533(.156)	.1997(.059)[.057]	.1314(.142)	.1518(.144)	.1964(.059)[.056]
	.2	.2249(.070)	.2251(.088)	.2044(.054)[.053]	.2235(.076)	.2226(.096)	.2009(.052)[.052]	.2221(.071)	.2212(.089)	.2018(.052)[.051]
	.2	.2619(.196)	.2435(.203)	.1824(.135)[.112]	.2519(.195)	.2324(.202)	.1729(.132)[.107]	.2520(.195)	.2333(.202)	.1740(.133)[.119]
400	1	1.0012(.011)	.9987(.011)	.9993(.011)[.011]	1.0014(.011)	.9989(.011)	.9997(.011)[.011]	1.0014(.012)	.9989(.012)	.9997(.011)[.011]
	1	.9702(.092)	.9832(.094)	.9892(.094)[.096]	.9663(.186)	.9795(.191)	.9865(.193)[.178]	.9724(.144)	.9855(.148)	.9927(.149)[.123]
	.3	.2663(.011)	.3005(.011)	.3001(.012)[.013]	.2664(.013)	.3004(.012)	.2997(.017)[.018]	.2665(.012)	.3007(.012)	.3005(.015)[.015]
	.2	.1800(.045)	.1858(.046)	.1988(.034)[.034]	.1796(.050)	.1851(.052)	.1995(.034)[.034]	.1784(.053)	.1841(.054)	.1987(.034)[.034]
	.2	.2010(.041)	.2077(.047)	.2003(.043)[.043]	.2032(.043)	.2104(.052)	.2012(.042)[.042]	.2017(.043)	.2087(.051)	.2011(.042)[.042]
	.2	.2331(.114)	.2317(.119)	.1886(.091)[.080]	.2322(.120)	.2311(.125)	.1860(.094)[.079]	.2354(.115)	.2339(.119)	.1893(.089)[.080]

Note: 1. $\psi = (\beta, \sigma_v^2, \rho, \lambda_1, \lambda_2, \lambda_3)'$; 2. Variances increase and then decrease with group size; 3. W is generated according to fixed group scheme; 4. X_t values are generated with $\theta_x = (g, \phi_1, \phi_2, \sigma_1, \sigma_2) = (.01, .5, .5, 3, 1)$. * [rse]: empirical average of rses, only for robust M-estimator.

Table 2a. Empirical Mean(sd)[rse]* of CQMLE, M-estimator, and Robust M-estimator, **DGP1**, $T = 3$, $m = 10$, $\mathcal{H}$=H-II, SNR=1 ;

		Normal Error			Normal Mixture			Chi-Square		
n	ψ	CQMLE	M-Est	RM-Est	CQMLE	M-Est	RM-Est	CQMLE	M-Est	RM-Est
50	1	1.0037(.019)	1.0003(.019)	1.0000(.020)[.019]	1.0032(.020)	.9997(.020)	.9995(.021)[.018]	1.0028(.020)	.9993(.020)	.9990(.020)[.019]
	1	.9336(.167)	.9391(.169)	.9349(.170)[.161]	.9455(.339)	.9517(.344)	.9480(.343)[.230]	.9512(.252)	.9571(.255)	.9529(.256)[.195]
	.3	.2851(.019)	.2998(.019)	.3000(.020)[.019]	.2850(.020)	.2999(.020)	.3000(.022)[.021]	.2848(.019)	.2998(.019)	.3001(.020)[.019]
	.2	.1969(.055)	.1974(.056)	.1976(.057)[.048]	.1935(.057)	.1939(.058)	.1938(.059)[.047]	.1954(.055)	.1958(.055)	.1960(.057)[.048]
	.2	.2020(.031)	.2015(.030)	.2000(.031)[.032]	.1996(.032)	.1992(.031)	.1984(.032)[.033]	.2010(.032)	.2007(.032)	.1993(.033)[.033]
	.2	.0169(.228)	.0179(.229)	.0940(.338)[.263]	.0235(.231)	.0245(.232)	.0280(.341)[.253]	.0151(.237)	.0160(.238)	.0165(.366)[.265]
100	1	1.0031(.016)	1.0005(.016)	1.0003(.016)[.016]	1.0026(.015)	1.0001(.016)	.9999(.016)[.015]	1.0026(.016)	1.0001(.016)	.9999(.016)[.016]
	1	.9716(.121)	.9767(.122)	.9747(.122)[.120]	.9745(.250)	.9799(.253)	.9783(.253)[.181]	.9809(.184)	.9862(.186)	.9844(.186)[.150]
	.3	.2858(.014)	.2995(.014)	.2998(.014)[.014]	.2862(.014)	.3000(.014)	.3001(.015)[.015]	.2863(.013)	.3002(.013)	.3004(.014)[.014]
	.2	.1954(.029)	.2008(.029)	.1998(.030)[.030]	.1941(.028)	.1994(.028)	.1985(.030)[.030]	.1947(.029)	.2001(.029)	.1991(.030)[.030]
	.2	.2076(.023)	.2002(.023)	.2004(.024)[.024]	.2081(.023)	.2005(.023)	.2008(.024)[.024]	.2075(.023)	.1999(.023)	.2000(.024)[.024]
	.2	.1145(.153)	.1104(.153)	.1271(.200)[.169]	.1099(.149)	.1058(.149)	.1218(.193)[.163]	.1117(.150)	.1075(.149)	.1239(.193)[.165]
200	1	1.0026(.011)	1.0004(.011)	1.0003(.011)[.010]	1.0025(.011)	1.0003(.011)	1.0002(.011)[.010]	1.0023(.010)	1.0001(.010)	.9999(.010)[.010]
	1	.9840(.086)	.9889(.087)	.9873(.087)[.086]	.9887(.176)	.9938(.178)	.9923(.177)[.135]	.9884(.130)	.9934(.131)	.9918(.131)[.109]
	.3	.2866(.009)	.2995(.009)	.2996(.009)[.010]	.2873(.010)	.3003(.010)	.3003(.010)[.010]	.2872(.009)	.3002(.009)	.3002(.010)[.010]
	.2	.1943(.023)	.2004(.023)	.1996(.024)[.024]	.1943(.024)	.2004(.024)	.1995(.025)[.023]	.1935(.024)	.1997(.024)	.1988(.024)[.024]
	.2	.2045(.022)	.2004(.022)	.2002(.022)[.022]	.2056(.022)	.2014(.022)	.2013(.023)[.022]	.2049(.022)	.2007(.022)	.2005(.023)[.022]
	.2	.1337(.104)	.1293(.104)	.1563(.127)[.116]	.1327(.105)	.1283(.105)	.1550(.129)[.115]	.1339(.102)	.1295(.102)	.1565(.126)[.116]
400	1	1.0011(.007)	1.0001(.007)	1.0000(.007)[.007]	1.0010(.007)	1.0000(.007)	1.0000(.007)[.007]	1.0010(.007)	1.0000(.007)	1.0000(.007)[.007]
	1	.9886(.062)	.9935(.062)	.9918(.062)[.061]	.9846(.128)	.9895(.130)	.9878(.129)[.096]	.9910(.093)	.9960(.094)	.9943(.094)[.078]
	.3	.2871(.006)	.2999(.007)	.3000(.007)[.007]	.2871(.007)	.2999(.007)	.2999(.007)[.008]	.2871(.007)	.3000(.007)	.3001(.007)[.007]
	.2	.1976(.014)	.1999(.014)	.1992(.014)[.014]	.1978(.014)	.2001(.014)	.1994(.014)[.014]	.1980(.014)	.2004(.014)	.1996(.014)[.014]
	.2	.2012(.014)	.2008(.014)	.2005(.014)[.014]	.2007(.014)	.2004(.014)	.2002(.014)[.014]	.2005(.014)	.2002(.014)	.1999(.014)[.014]
	.2	.1493(.072)	.1482(.072)	.1810(.086)[.082]	.1528(.071)	.1517(.071)	.1852(.085)[.080]	.1527(.070)	.1516(.070)	.1852(.083)[.081]

Note: 1. $\psi = (\beta, \sigma_v^2, \rho, \lambda_1, \lambda_2, \lambda_3)'$; 2. Variances decrease and then increase with group size; 3. W is generated according to fixed group scheme;
4. X_t values are generated with $\theta_x = (g, \phi_1, \phi_2, \sigma_1, \sigma_2) = (.01, .5, .5, 1, 1)$. * [rse]: empirical average of rses, only for robust M-estimator.

Table 2b. Empirical Mean(sd)[rse]* of CQMLE, M-estimator, and Robust M-estimator, **DGP1**, $T = 3$, $m = 10$, $\mathcal{H}$=H-II, SNR=3 ;

		Normal Error			Normal Mixture			Chi-Square		
n	ψ	CQMLE	M-Est	RM-Est	CQMLE	M-Est	RM-Est	CQMLE	M-Est	RM-Est
50	1	1.0037(.019)	1.0003(.019)	1.0001(.020)[.019]	1.0033(.019)	.9999(.020)	.9996(.020)[.019]	1.0030(.020)	.9995(.020)	.9993(.020)[.019]
	1	.9342(.167)	.9397(.168)	.9355(.169)[.161]	.9417(.339)	.9478(.344)	.9443(.345)[.227]	.9391(.261)	.9450(.265)	.9404(.265)[.193]
	.3	.2851(.018)	.2998(.019)	.3000(.019)[.019]	.2853(.019)	.3001(.019)	.3005(.022)[.021]	.2848(.019)	.2996(.019)	.2999(.020)[.019]
	.2	.1970(.055)	.1975(.055)	.1977(.057)[.048]	.1963(.053)	.1968(.054)	.1971(.055)[.047]	.1973(.055)	.1977(.055)	.1979(.057)[.047]
	.2	.2019(.031)	.2015(.030)	.2000(.031)[.032]	.2013(.031)	.2006(.031)	.1994(.032)[.032]	.2015(.031)	.2012(.031)	.1999(.032)[.032]
	.2	.0169(.229)	.0179(.229)	.095(.338)[.263]	.0123(.235)	.0132(.236)	.0207(.353)[.255]	.0139(.240)	.0149(.240)	.0178(.361)[.264]
100	1	1.0028(.016)	1.0003(.016)	1.0001(.016)[.015]	1.0026(.015)	1.0001(.015)	.9999(.015)[.015]	1.0031(.016)	1.0005(.016)	1.0003(.016)[.015]
	1	.9729(.120)	.9780(.121)	.9759(.121)[.120]	.9647(.248)	.9699(.251)	.9681(.250)[.177]	.9752(.187)	.9804(.189)	.9787(.190)[.150]
	.3	.2857(.013)	.2995(.014)	.2996(.014)[.014]	.2864(.014)	.3000(.013)	.3000(.015)[.015]	.2863(.013)	.3001(.013)	.3004(.014)[.014]
	.2	.1953(.029)	.2007(.029)	.1997(.030)[.030]	.1953(.030)	.2006(.030)	.1995(.031)[.030]	.1951(.028)	.2005(.028)	.1995(.029)[.030]
	.2	.2086(.023)	.2010(.023)	.2013(.024)[.024]	.2075(.023)	.2001(.023)	.2005(.024)[.024]	.2072(.023)	.1997(.023)	.1999(.024)[.024]
	.2	.1084(.156)	.1044(.156)	.1196(.202)[.171]	.1138(.151)	.1097(.151)	.1265(.196)[.160]	.1122(.153)	.1081(.153)	.1245(.198)[.165]
200	1	1.0018(.010)	.9996(.010)	.9994(.010)[.010]	1.0021(.011)	.9999(.011)	.9998(.011)[.010]	1.0024(.010)	1.0002(.011)	1.0001(.011)[.010]
	1	.9835(.086)	.9884(.087)	.9868(.087)[.086]	.9885(.175)	.9936(.176)	.9921(.176)[.135]	.9859(.131)	.9908(.132)	.9893(.132)[.108]
	.3	.2870(.009)	.2999(.009)	.3001(.009)[.010]	.2871(.010)	.3001(.009)	.3002(.010)[.010]	.2868(.009)	.2998(.009)	.2998(.009)[.010]
	.2	.1942(.023)	.2003(.024)	.1993(.024)[.024]	.1946(.024)	.2008(.024)	.1999(.024)[.023]	.1936(.024)	.1998(.024)	.1988(.024)[.024]
	.2	.2045(.022)	.2004(.022)	.2000(.022)[.022]	.2046(.022)	.2004(.022)	.2002(.022)[.022]	.2041(.022)	.2000(.022)	.1998(.023)[.022]
	.2	.1325(.104)	.1281(.104)	.1549(.127)[.117]	.1318(.104)	.1273(.104)	.1537(.128)[.114]	.1336(.102)	.1292(.102)	.1562(.125)[.115]
400	1	1.0011(.008)	1.0001(.008)	1.0001(.008)[.007]	1.0009(.007)	.9999(.007)	.9999(.007)[.007]	1.0008(.007)	.9998(.007)	.9998(.007)[.007]
	1	.9912(.061)	.9962(.061)	.9945(.061)[.050]	.9916(.123)	.9966(.124)	.9950(.124)[.050]	.9922(.095)	.9972(.096)	.9955(.096)[.050]
	.3	.2873(.007)	.3001(.007)	.3002(.007)[.007]	.2870(.007)	.2999(.007)	.2999(.007)[.007]	.2871(.007)	.3000(.007)	.3000(.007)[.007]
	.2	.1987(.013)	.2011(.013)	.2004(.014)[.017]	.1980(.013)	.2003(.013)	.1996(.014)[.017]	.1986(.014)	.2009(.014)	.2002(.014)[.017]
	.2	.2006(.014)	.2001(.014)	.1998(.014)[.018]	.2010(.014)	.2006(.014)	.2002(.014)[.018]	.2003(.014)	.1998(.014)	.1996(.014)[.018]
	.2	.1474(.072)	.1463(.072)	.1788(.087)[.070]	.1470(.073)	.1459(.073)	.1781(.088)[.070]	.1494(.070)	.1483(.070)	.1813(.083)[.070]

Note: 1. $\psi = (\beta, \sigma_v^2, \rho, \lambda_1, \lambda_2, \lambda_3)'$; 2. Variances decrease and then increase with group size; 3. W is generated according to fixed group scheme;
4. X_t values are generated with $\theta_x = (g, \phi_1, \phi_2, \sigma_1, \sigma_2) = (.01, .5, .5, 3, 1)$. * [rse]: empirical average of rses, only for robust M-estimator.

Table 3. Empirical Mean(sd)[rse]* of CQMLE, M-estimator, and Robust M-estimator, $T = 3$, $m = 10$, $\mathcal{H} = I_n$, SNR=1 ;

		Normal Error			Normal Mixture			Chi-Square		
n	ψ	CQMLE	M-Est	RM-Est	CQMLE	M-Est	RM-Est	CQMLE	M-Est	RM-Est
50	1	1.0082(.031)	.9991(.032)	.9988(.034)[.033]	1.0056(.033)	.9961(.033)	.9963(.034)[.033]	1.0066(.032)	.9974(.033)	.9973(.035)[.032]
	1	.9235(.135)	.9375(.139)	.9397(.148)[.150]	.9310(.271)	.9457(.280)	.9512(.290)[.219]	.9308(.201)	.9451(.207)	.9484(.213)[.177]
	.3	.2616(.032)	.3008(.034)	.3020(.040)[.040]	.2624(.034)	.3022(.035)	.3033(.047)[.050]	.2626(.034)	.3022(.035)	.3032(.043)[.045]
	.2	.1877(.109)	.1869(.113)	.1863(.120)[.105]	.1803(.126)	.1777(.112)	.1813(.125)[.115]	.1850(.114)	.1836(.120)	.1847(.124)[.112]
	.2	.2016(.101)	.2102(.114)	.2115(.124)[.112]	.2032(.111)	.2143(.131)	.2139(.139)[.127]	.2025(.104)	.2115(.119)	.2121(.122)[.117]
	.2	.1020(.215)	.1055(.222)	.1067(.231)[.189]	.1051(.228)	.1102(.237)	.1074(.243)[.166]	.0975(.223)	.1011(.232)	.1001(.242)[.170]
100	1	1.0066(.027)	1.0003(.027)	1.0002(.031)[.031]	1.0044(.027)	.9981(.027)	.9982(.035)[.034]	1.0046(.025)	.9982(.026)	.9982(.031)[.031]
	1	.9559(.096)	.9695(.099)	.9704(.102)[.106]	.9659(.198)	.9803(.204)	.9827(.211)[.165]	.9589(.147)	.9728(.152)	.9745(.157)[.132]
	.3	.2641(.022)	.2999(.023)	.3003(.029)[.030]	.2645(.023)	.3007(.024)	.3012(.033)[.034]	.2648(.022)	.3008(.023)	.3015(.029)[.030]
	.2	.1768(.091)	.1924(.090)	.1925(.094)[.090]	.1718(.098)	.1877(.097)	.1894(.096)[.087]	.1774(.093)	.1928(.092)	.1940(.093)[.085]
	.2	.2216(.062)	.2050(.065)	.2049(.071)[.069]	.2221(.064)	.2053(.069)	.2044(.072)[.065]	.2196(.062)	.2028(.066)	.2020(.072)[.066]
	.2	.1505(.149)	.1551(.151)	.1560(.157)[.139]	.1507(.156)	.1552(.157)	.1553(.161)[.149]	.1447(.151)	.1496(.153)	.1495(.158)[.147]
200	1	1.0051(.018)	.9996(.018)	.9996(.023)[.023]	1.0052(.018)	.9998(.018)	.9998(.024)[.024]	1.0046(.017)	.9991(.017)	.9990(.024)[.024]
	1	.9738(.069)	.9871(.071)	.9877(.075)[.076]	.9715(.139)	.9850(.143)	.9858(.148)[.119]	.9729(.104)	.9863(.107)	.9869(.112)[.098]
	.3	.2656(.015)	.2997(.016)	.2999(.022)[.022]	.2661(.015)	.3002(.015)	.3001(.024)[.025]	.2657(.015)	.2999(.016)	.3000(.021)[.021]
	.2	.1776(.052)	.1953(.052)	.1956(.058)[.055]	.1803(.049)	.1980(.049)	.1982(.052)[.051]	.1788(.051)	.1966(.051)	.1967(.057)[.055]
	.2	.2102(.039)	.2032(.042)	.2035(.049)[.048]	.2085(.038)	.2013(.041)	.2016(.046)[.045]	.2093(.040)	.2020(.042)	.2018(.049)[.048]
	.2	.1755(.093)	.1817(.094)	.1811(.099)[.094]	.1725(.093)	.1776(.095)	.1785(.099)[.093]	.1735(.095)	.1787(.096)	.1790(.101)[.095]
400	1	1.0024(.012)	.9999(.012)	.9999(.015)[.015]	1.0025(.012)	.9999(.012)	.9999(.015)[.015]	1.0020(.012)	.9995(.012)	.9995(.015)[.015]
	1	.9811(.049)	.9943(.051)	.9946(.056)[.056]	.9775(.098)	.9907(.101)	.9911(.106)[.100]	.9792(.072)	.9924(.074)	.9928(.079)[.071]
	.3	.2659(.011)	.3000(.011)	.3000(.018)[.018]	.2660(.011)	.3000(.011)	.3000(.018)[.019]	.2661(.011)	.3001(.011)	.3001(.017)[.017]
	.2	.1939(.029)	.2002(.029)	.2002(.033)[.032]	.1928(.029)	.1991(.029)	.1991(.034)[.033]	.1928(.028)	.1991(.028)	.1992(.034)[.034]
	.2	.1976(.028)	.1997(.030)	.1997(.037)[.037]	.1988(.029)	.2008(.031)	.2009(.036)[.034]	.1978(.028)	.2000(.031)	.2003(.036)[.036]
	.2	.1807(.063)	.1881(.064)	.1882(.069)[.068]	.1836(.062)	.1906(.063)	.1911(.069)[.067]	.1834(.064)	.1904(.065)	.1911(.071)[.069]

Note: 1. $\psi = (\beta, \sigma_v^2, \rho, \lambda_1, \lambda_2, \lambda_3)'$; 2. W is generated according to fixed group scheme; 3. v_{it} is homoskedastic;
4. X_t values are generated with $\theta_x = (g, \phi_1, \phi_2, \sigma_1, \sigma_2) = (.01, .5, .5, 3, 1)$. * [rse]: empirical average of rses, only for robust M-estimator.

Table 4. Empirical Mean(sd)[rse] of Robust M-estimator, **DGP2** (Durbin), $T = 3$, $m = 10$, SNR=3

	$n = 100$			$n = 400$		
ψ	Normal Error	Normal Mixture	Chi-Square	Normal Error	Normal Mixture	Chi-Square
1	1.000(.012)[.012]	1.000(.012)[.012]	1.000(.012)[.011]	1.000(.007)[.007]	1.000(.007)[.007]	1.000(.007)[.007]
.2	.200(.040)[.039]	.199(.040)[.038]	.199(.040)[.038]	.200(.018)[.017]	.200(.017)[.017]	.200(.017)[.017]
1	.947(.181)[.175]	.952(.376)[.251]	.953(.281)[.213]	.985(.096)[.094]	.979(.188)[.144]	.981(.140)[.118]
.3	.300(.013)[.013]	.301(.015)[.014]	.300(.014)[.013]	.300(.007)[.008]	.300(.008)[.009]	.300(.008)[.008]
.2	.197(.056)[.049]	.197(.057)[.051]	.196(.056)[.050]	.199(.024)[.023]	.198(.024)[.023]	.200(.024)[.023]
.2	.201(.052)[.045]	.202(.054)[.048]	.204(.052)[.046]	.201(.028)[.027]	.201(.027)[.027]	.201(.027)[.027]
.2	.128(.201)[.162]	.141(.192)[.150]	.138(.199)[.155]	.189(.088)[.080]	.187(.086)[.078]	.186(.085)[.079]

Note: 1. $\psi = (\beta, \beta_d, \sigma_v^2, \rho, \lambda_1, \lambda_2, \lambda_3)'$; 2. W is generated according to fixed group scheme;
3. X_t values are generated with $\theta_x = (g, \phi_1, \phi_2, \sigma_1, \sigma_2) = (.01, .5, .5, 3, 1)$.

Table 5. Empirical Mean(sd) of Robust M-estimator and GMM-estimator
DGP3, $T = 3$, $m = 10$, $\mathcal{H}$ = H-I, SNR=3

		Normal Error		Normal Mixture		Chi-Square	
n	ψ	RM-Est	GMM-Est	RM-Est	GMM-Est	RM-Est	GMM-Est
50	1	.9989(.021)	1.0038(.103)	.9993(.022)	1.0038(.098)	.9979(.022)	.9994(.104)
	.3	.3008(.019)	.2620(.086)	.3002(.021)	.2646(.083)	.3006(.020)	.2611(.087)
	.2	.1951(.084)	.2279(.117)	.1919(.086)	.2268(.102)	.1941(.083)	.2294(.112)
	.2	.2021(.063)	.1976(.077)	.2055(.065)	.1969(.079)	.2053(.062)	.1988(.085)
100	1	.9998(.015)	.9942(.069)	1.0000(.016)	.9972(.066)	.9997(.016)	.9966(.072)
	.3	.2995(.015)	.2619(.088)	.2991(.018)	.2702(.078)	.2998(.016)	.2627(.081)
	.2	.1968(.069)	.2148(.115)	.1920(.070)	.2153(.102)	.1955(.069)	.2150(.106)
	.2	.1995(.039)	.1887(.086)	.2017(.040)	.1913(.082)	.2010(.039)	.1904(.081)
200	1	.9992(.010)	.9949(.042)	.9997(.010)	.9951(.040)	.9998(.010)	.9953(.041)
	.3	.3002(.010)	.2858(.049)	.3002(.012)	.2866(.047)	.2999(.010)	.2851(.050)
	.2	.1985(.026)	.2027(.055)	.1996(.027)	.2034(.054)	.1992(.027)	.2048(.055)
	.2	.2009(.029)	.2035(.049)	.2001(.029)	.2013(.048)	.1996(.029)	.2007(.050)
400	1	1.0000(.007)	1.0021(.042)	1.0000(.008)	1.001(.042)	1.0003(.007)	1.0013(.042)
	.3	.3002(.008)	.2890(.032)	.3002(.009)	.2923(.033)	.2999(.008)	.2897(.032)
	.2	.1998(.019)	.2059(.061)	.1998(.019)	.2078(.061)	.1988(.020)	.2108(.059)
	.2	.1999(.021)	.1979(.049)	.2000(.021)	.1983(.046)	.2006(.021)	.1960(.045)

Note: 1. $\psi = (\beta, \rho, \lambda_1, \lambda_2)'$; 2. W is generated according to fixed group scheme;
3. X_t values are generated with $\theta_x = (g, \phi_1, \phi_2, \sigma_1, \sigma_2) = (.01, .5, .5, 3, 1)$.