Unbalanced Spatial Panel Data Models with Fixed Effects

Xiaoyu Meng and Zhenlin Yang*

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

July 22, 2021

Abstract

We consider estimation and inferences for fixed effects spatial panel data models based on unbalanced panels that result from randomly missing spatial units. The unbalanced nature of the panel data renders the standard method of estimation inapplicable. In this paper, we proposed an M-estimation method where the estimating functions are obtained by adjusting the concentrated quasi scores to account for the estimation of fixed effects and/or the presence of unknown spatiotemporal heteroskedasticity. The method allows for general time-varying spatial weight matrices without row-normalization, and is able to give full control of the individual and time specific effects for all the spatial units involved in the data. Consistency and asymptotic normality of the proposed estimators are established. Inference methods are introduced and their consistency is proved. Monte Carlo results show excellent finite sample performance of the proposed methods.

Key Words: Adjusted quasi score; Fixed effects; Spatial effects; Time-varying spatial weights; Unbalanced panel; Spatiotemporal heteroskedasticity.

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

1. Introduction

The literature on spatial panel data (SPD) models has been fast-growing since Anselin (1988), due to the facts that the SPD models are able to take into account the spatial interaction effects and control for the unobservable heterogeneity. Most of the works on SPD models are based on "complete" or "balanced" panels, i.e., a set of observations collected on n spatial units over the entire T periods in time (e.g., Baltagi et al., 2003; Lee and Yu, 2010; Baltagi

*Corresponding author. Email: zlyang@smu.edu.sg

and Yang, 2013a,b; Yang et al., 2016; Liu and Yang, 2020, to mention a few); only a few on "incomplete" or "unbalanced" panels (Wang and Lee, 2013b; Egger et al., 2005; Baltagi et al., 2007; Baltagi et al., 2015). This is in stark contrast to the usual panel model literature, which contains a sizable portion of works on unbalanced panels (e.g., Wansbeek and Kapteyn, 1989; Baltagi and Chang, 1994; Davis, 2001; Baltagi et al., 2001; Antweiler, 2001; Baltagi and Song, 2006; Bai et al., 2015; Wooldridge, 2019, among others), textbook treatments (Baltagi, 2013; Hsiao, 2014; Greene, 2018), and software implementations (`STATA`, `SAS`, and `R`).

Unbalanced panels are likely to be the norm in typical economic empirical settings (Baltagi and Song, 2006), so are the unbalanced spatial panels. Unbalancedness may be the result of "randomly missing" observations such as early drop-outs, late entrants and lack of economic activities, or "nonrandomly missing" observations such as attrition and sample selection (Baltagi and Song, 2006; Baltagi, 2013, Ch. 9). The key difference between the two missing mechanisms is that in the former analyses can simply be done based on the actual observed data, but in the latter "imputation" may be necessary before formal analyses. Under the random missing mechanism, most of the methods and techniques developed for balanced panels can be adapted to suit the unbalanced panels, but these may not be true or cannot be easily done for spatial panels. Under the nonrandom missing mechanism, the treatments become much more complicated for both regular and spatial panels, in particular the latter.

The limited literature on unbalanced spatial panels contains three interesting empirical studies under the randomly missing mechanism (RMM). Baltagi et al. (2007) studied the third-country effects on foreign direct investment (FDI) based on an unbalanced SPD model with only spatial error effects. Egger et al. (2005) studied US state tax competition based on an unbalanced SPD model with both spatial lag and spatial error. Both papers focus on random effects model and adapt the GMM approach of Kapoor et al. (2007). There are no theoretical studies being given on the properties of these methods and no formal considerations being given on the models with fixed effects. Baltagi et al. (2015) studied hedonic housing prices based on an unbalanced spatial lag pseudo-panel data model with nested random effects by adapting the ML approach of Antweiler (2001). The sole theoretical work in this literature is Wang and Lee (2013b) who studied SPD models with (correlated) random effects where missing data occur only on the response variable.[1] Many important and common issues remain

[1]The fixed effects model was treated in an appendix. It requires the spatial weights matrices to be time-invariant that is clearly not satisfied by the unbalanced SPD models. Related works but under spatial cross-sectional setup include Kelejian and Prucha (2010), LeSage and Pace (2004) and Wang and Lee (2013a).

for the unbalanced SPD models even under the simpler random missing mechanism, such as fixed effects, heteroskedasticity (spatial and temporal), and serial correlation. It is therefore highly desirable to develop general estimation and inference methods to address these issues.

In this paper, we consider the unbalanced SPD models with RMM. In a spatial panel framework, by RMM we mean specifically "randomly missing spatial units" in the sense that the spatial units not present in the tth time period did not make impacts on their 'neighbors' at that time so that analyses can simply be done based on the observed spatial units and their spatial interactions. The popular transformation method (Lee and Yu, 2010) cannot be applied to handle the fixed effects due to the fact that spatial weight matrices are time-varying, and may not be row-normalizable (Liu and Lee, 2010). The heteroskedasticity-robust method of Liu and Yang (2020) cannot be applied either due to a similar reason.[2] Allowing serial correlation in the error term is interesting but has not been considered. We focus on the unbalanced SPD models with both unit- and time-specific fixed effects, where the errors can be homoskedastic or heteroskedastic of unknown form in both cross-sectional and time dimensions, leaving the issue of serial correlation to the discussion section.

We propose a general M-, or *adjusted quasi score* (AQS), estimation method, for estimating the unbalanced SPD models. The method starts from the joint quasi score functions of both the common parameters and fixed effects, then concentrates out the fixed effects to give the concentrated quasi score functions, and then adjusts these concentrated score functions to give a set of unbiased estimating functions for the common parameters – the AQS functions. Solving the AQS equations gives the AQS estimators that are shown to be consistent and asymptotically unbiased. We first consider an FE-SPD model with both spatial lag and spatial error effects under homoskedasticity to fix the main ideas behind the proposed methodology. Then, we make a full extension of the methods to allow for unknown heteroskedasticity in the errors across both space and time. For this, a new way of adjusting the concentrated quasi score functions is required to make them robust against the unknown heteroskedasticity. Consistency and asymptotic normality of all these proposed estimators are established. Simple methods of inference are introduced under both homoskedastic and heteroskedastic errors. Monte Carlo results show excellent finite sample performance of the proposed methods. The proposed methods are simple and reliable, and yet quite general, having a great extendibility

[2]The GMM-type methods of Moscone and Tosetti (2011) and Badinger and Egger (2015) cannot be easily adapted either, besides the issues on efficiency, time fixed effects, and incidental parameters problem.

for extra features in the model (e.g., serial correlation and time-varying coefficients), and for different types of models (e.g., models with random effects and interactive fixed effects).

The rest of the paper is organized as follows. Section 2 introduces the M-estimation method for estimating an unbalance SPD model with two-way FE under homoskedasticity, studies the consistency and asymptotic normality of the AQS estimators, and presents a simple method for standard errors estimation. Section 3 makes a full extension of the AQS methods in Section 2 by allowing the errors to be heteroskedastic across both space and time. Section 4 presents Monte Carlo results and Section 5 concludes. Proofs of the main results are given in Appendices B and C. Additional proofs and some extensions are given in Appendix D.

2. Unbalanced FE-SPD Model with Homoskedasticity

2.1. The Model

Consider a study that lasts T periods and involves a total of n spatial units. At time t, only n_t of these n spatial units are available to give observations on their responses and explanatory variables, and the rest are not due to random missing, e.g., early drop-outs, late entries, lack of economic activities, etc., as discussed in the introduction. These spatial units are interconnected with their 'connectivity' changing over time;[3] they typically vary in size, causing the error distributions to be heteroskedastic; and certain unit- and time-specific features may not be observed but must be acknowledged. These give rise to a spatial panel data (SPD) model with unbalanced panels, time-varying spatial weight matrices, unknown heteroskedasticity, and unit- and time-specific fixed effects (FE):

$$Y_t = \lambda_0 W_t Y_t + X_t \beta_0 + D_t \mu_0 + \alpha_{t0} l_{n_t} + U_t, \quad U_t = \rho_0 M_t U_t + V_t, \quad t = 1, \ldots, T, \tag{2.1}$$

where Y_t is a vector of observations on n_t spatial units at time t, X_t is an $n_t \times k$ matrices containing values of k time-varying exogenous regressors, and $U_t = (u_{1t}, u_{2t}, \ldots, u_{n_t t})'$ and $V_t = (v_{1t}, v_{2t}, \ldots, v_{n_t t})'$ are $n_t \times 1$ vectors of disturbances and idiosyncratic errors, respectively. W_t and M_t are given $n_t \times n_t$ spatial weight matrices. λ_0 and ρ_0 are spatial coefficients, which together with W_t and M_t characterize the spatial lag (`SL`) effects and the spatial error (`SE`) effects, respectively.[4] β_0 is a $k \times 1$ vector of regression coefficients. $\mu_0 = \{\mu_{i0}\}_{i=1}^n$ denotes an

[3]Changes in connectivity may be simply due to the changes in the available spatial units in each period or due to more fundamental changes in connectivity among spatial units over time.

[4]Spatial Durbin terms, $W_t X_t$, can be added. However, there might be overfitting identification problem if the model contains all three spatial effects (Anselin et al., 2008; Lee and Yu, 2016).

$n \times 1$ vector of unit-specific effects and $\alpha_0 = \{\alpha_{t0}\}_{t=1}^T$ a $T \times 1$ vector of time-specific effects.[5] D_t is an $n_t \times n$ 'selection' matrix obtained from the $n \times n$ identity matrix I_n by deleting its rows that correspond to the missing units at time t, and l_{n_t} is an $n_t \times 1$ vector of ones.

Both μ_0 and α_0 are allowed to correlate with the time-varying regressors in an arbitrary manner and hence are considered as fixed effects. When the change in W_t and M_t is due only to the missing spatial units, they can be written as $W_t = D_t W D_t'$ and $M_t = D_t M D_t'$, where W and M are the spatial weight matrices for all the n spatial units involved in the study. The idiosyncratic errors $\{v_{it}\}$ are first treated as independent and identically distributed (iid) across i and over t, and then extended to be independent but not identically distributed (inid).

An important advantage of the modeling strategy of (2.1) is that it allows the full control of the unobserved heterogeneity of all n spatial units, as long as each of the n spatial units is observed at least twice over the entire period of study so that all the n units remain in the model after the fixed effects being concentrated out. Moreover, the spatial weight matrices W_t and M_t are not necessarily row-normalized, and they are allowed to be generally time-varying, catering to both the random-missing mechanism and the genuine time-varying features.

Some generic notations and conventions will be followed. For a square matrix, $|\cdot|$ denotes its determinant and $\mathtt{tr}(\cdot)$ its trace. For a real symmetric matrix, $\gamma_{\min}(\cdot)$ and $\gamma_{\max}(\cdot)$ denote its smallest and largest eigenvalues. For a real $n \times m$ matrix A, A' denotes its transpose, $\|A\|_F$ its Frobenius norm, $\|A\|_1$ its maximum column sum norm, $\|A\|_\infty$ its maximum row sum norm, and $A^\circ = A + A'$. For a real $n \times m$ matrix A with a full column rank, $\mathbb{P}_A = A(A'A)^{-1}A'$ and $\mathbb{Q}_A = I_n - \mathbb{P}_A$ are the two orthogonal projection matrices. The operator $\mathtt{diag}(\cdot)$ forms a diagonal matrix by diagonal elements of a square matrix or elements of a given vector, $\mathtt{diagv}(\cdot)$ forms a column vector using diagonal elements of a square matrix, and $\mathtt{blkdiag}(\cdots)$ forms a block-diagonal matrix by the given submatrices. The usual expectation and variance operators, $\mathrm{E}(\cdot)$ and $\mathrm{Var}(\cdot)$, correspond to true parameter values with a subscript 0.

2.2. Quasi-Maximum Likelihood Estimation

Define $\mathbf{W} = \mathtt{blkdiag}(W_1, \ldots, W_T)$, $\mathbf{M} = \mathtt{blkdiag}(M_1, \ldots, M_T)$, $\mathbf{D}_\mu = (D_1', \ldots, D_T')'$, and $\mathbf{D}_\alpha = \mathtt{blkdiag}(l_{n_1}, \ldots, l_{n_T})$. Denote $N = \sum_{t=1}^T n_t$, $\mathbf{Y} = (Y_1', \ldots, Y_T')'$, $\mathbf{X} = (X_1', \ldots, X_T')'$, $\mathbf{U} = (U_1', \ldots, U_T')'$, and $\mathbf{V} = (V_1', \ldots, V_T')'$. Model (2.1) is written in the matrix form:

[5] They may appear in the model in the *interactive form* (Bai, 2009; Bai and Ng, 2013) to capture heterogenous responses on multiple shocks in time. This paper focuses on the case of additive fixed effects.

$\mathbf{Y} = \lambda_0 \mathbf{W}\mathbf{Y} + \mathbf{X}\beta_0 + \mathbf{D}_\mu \mu_0 + \mathbf{D}_\alpha \alpha_0 + \mathbf{U}$ and $\mathbf{U} = \rho_0 \mathbf{M}\mathbf{U} + \mathbf{V}$. The existing method of estimating an SPD model with fixed effects is to apply orthogonal transformations to wipe out the fixed effects so that the transformed model remains in the same spatial structure and the (quasi) likelihood can be formed (see, e.g., Lee and Yu, 2010; Yang et al., 2016). This method requires that the panel is balanced, spatial weight matrices are time-invariant and row-normalized, and idiosyncratic errors are homoskedastic. However, none of these is met in the current model specification. To overcome this difficulty, we start with the quasi maximum likelihood (QML) method that estimates the common parameters and the fixed effects together. To eliminate the effects of estimating the fixed effects on the estimation of the common parameters, we in next subsection modify the quasi score functions to produce a set of unbiased and consistent estimating equations. For QML estimation, first note that there are $n+T$ fixed effects parameters but only $n+T-1$ of them are identifiable. A zero-sum constraint is put on the $\alpha_t's$ and the QML estimation is based on the following model form:

$$\mathbf{Y} = \lambda_0 \mathbf{W}\mathbf{Y} + \mathbf{X}\beta_0 + \mathbf{D}_\mu \mu_0 + \mathbf{D}_\alpha^\star \alpha_0^\star + \mathbf{U}, \quad \mathbf{U} = \rho_0 \mathbf{M}\mathbf{U} + \mathbf{V}. \tag{2.2}$$

where $\alpha_0^\star = (\alpha_{20}^\star, \ldots, \alpha_{T0}^\star)'$, and $\mathbf{D}_\alpha^\star = [-l_{n_1} l_{T-1}';\ \texttt{blkdiag}(l_{n_2}, \ldots, l_{n_T})]$.

Denote the set of *common* parameters by $\theta = (\beta', \sigma_v^2, \delta')'$ where $\delta = (\lambda, \rho)'$, and the set of *incidental* parameters by $\phi = (\mu', \alpha^{\star\prime})'$. Define $\mathbf{A}_N(\lambda) = I_N - \lambda\mathbf{W}$ and $\mathbf{B}_N(\rho) = I_N - \rho\mathbf{M}$. We have the quasi Gaussian loglikelihood function of θ and ϕ:

$$\ell_N(\theta, \phi) = -\tfrac{N}{2}\ln 2\pi - \tfrac{N}{2}\ln\sigma_v^2 + \ln|\mathbf{A}_N(\lambda)| + \ln|\mathbf{B}_N(\rho)| - \tfrac{1}{2\sigma_v^2}\mathbf{V}'(\beta,\delta,\phi)\mathbf{V}(\beta,\delta,\phi), \tag{2.3}$$

where $\mathbf{V}(\beta,\delta,\phi) = \mathbf{B}_N(\rho)[\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\beta - \mathbf{D}\phi]$, and $\mathbf{D} = [\mathbf{D}_\mu,\ \mathbf{D}_\alpha^\star]$.

Let $\mathbb{D}(\rho) = \mathbf{B}_N(\rho)\mathbf{D}$. Given θ, $\ell_N(\theta,\phi)$ is partially maximized at

$$\hat{\phi}_N(\beta,\delta) = [\mathbb{D}'(\rho)\mathbb{D}(\rho)]^{-1}\mathbb{D}'(\rho)\mathbf{B}_N(\rho)[\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\beta]. \tag{2.4}$$

Substituting $\hat{\phi}_N(\beta,\delta)$ into $\ell_N(\theta,\phi)$ gives the concentrated quasi loglikelihood function for θ:

$$\ell_N^{c}(\theta) = -\tfrac{N}{2}\ln 2\pi - \tfrac{N}{2}\ln\sigma_v^2 + \ln|\mathbf{A}_N(\lambda)| + \ln|\mathbf{B}_N(\rho)| - \tfrac{1}{2\sigma_v^2}\tilde{\mathbf{V}}'(\beta,\delta)\tilde{\mathbf{V}}(\beta,\delta), \tag{2.5}$$

where $\tilde{\mathbf{V}}(\beta,\delta) = \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho)[\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\beta]$ and $\mathbb{Q}_{\mathbb{D}}(\rho)$ is the projection matrix based on $\mathbb{D}(\rho)$. The direct quasi maximum likelihood (QML) estimator $\hat{\theta}_{\texttt{QML}}$ of θ maximizes $\ell_N^{c}(\theta)$.

However, such a direct estimation of the common parameters θ completely ignores the impact from the estimation of the incidental parameters ϕ, rendering $\hat{\theta}_{\texttt{QML}}$ be inconsistent or asymptotically biased – the well known *incidental parameters problem* of Neyman and Scott

(1948). In their study for a balanced FE-SPD model, Lee and Yu (2010) show that the direct QMLEs of β and δ are consistent no matter T is large or small, but their distributions are asymptotically centered only when T is small relative to n. They further show that the QMLE of σ_v^2 is inconsistent and its limiting distribution is degenerate due to the incidental parameters problem when T is finite. Therefore, if the direct QML approach were followed, a bias correction needs to be done to remove the asymptotic bias for valid statistical inferences, which needs one additional condition that $\frac{T}{n^3} \to 0$. To overcome these problems, Lee and Yu (2010) propose a transformation approach to wipe out the fixed effects, taking advantages of the panel being balanced and spatial weight matrices being time-invariant and row-normalized. In our model specification, none of these features holds and the transformation approach fails to work. Therefore, an alternative (and more general) approach is highly desirable.

2.3. Adjusted Quasi Score Estimation

The root cause of inconsistency or asymptotic bias for the direct QML estimation is that a necessary condition for consistency of QML estimators, $\text{plim}\frac{1}{N}S_N^{\text{c}}(\theta_0) = 0$, is violated due to the concentration/estimation of the incidental parameters μ and α, where θ_0 denotes the true value of the parameter vector θ, and $S_N^{\text{c}}(\theta) = \frac{\partial}{\partial\theta}\ell_N^{\text{c}}(\theta)$ is a set of the concentrated quasi score (CQS) functions given as (see Appendix B)

$$S_N^{\text{c}}(\theta) = \begin{cases} \frac{1}{\sigma_v^2}\mathbf{X}'\mathbf{B}_N'(\rho)\tilde{\mathbf{V}}(\beta,\delta), \\ \frac{1}{2\sigma_v^4}[\tilde{\mathbf{V}}'(\beta,\delta)\tilde{\mathbf{V}}(\beta,\delta) - N\sigma_v^2], \\ \frac{1}{\sigma_v^2}\mathbf{Y}'\mathbf{W}'\mathbf{B}_N'(\rho)\tilde{\mathbf{V}}(\beta,\delta) - \texttt{tr}[\mathbf{F}_N(\lambda)], \\ \frac{1}{\sigma_v^2}\tilde{\mathbf{V}}'(\beta,\delta)\mathbf{G}_N(\rho)\tilde{\mathbf{V}}(\beta,\delta) - \texttt{tr}[\mathbf{G}_N(\rho)], \end{cases} \tag{2.6}$$

where $\mathbf{F}_N(\lambda) = \mathbf{W}\mathbf{A}_N^{-1}(\lambda)$ and $\mathbf{G}_N(\rho) = \mathbf{M}\mathbf{B}_N^{-1}(\rho)$.

Under mild conditions, maximizing $\ell_N^{\text{c}}(\theta)$ is equivalent to solving $S_N^{\text{c}}(\theta) = 0$. It is easy to show that at the true value θ_0 of θ,

$$\text{E}[S_N^{\text{c}}(\theta_0)] = \begin{cases} 0_k, \\ -\frac{n+T-1}{2\sigma_{v0}^2}, \\ \texttt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho_0)\mathbf{B}_N(\rho_0)\mathbf{F}_N(\lambda_0)\mathbf{B}_N^{-1}(\rho_0)] - \texttt{tr}[\mathbf{F}_N(\lambda_0)], \\ \texttt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho_0)\mathbf{G}_N(\rho_0)] - \texttt{tr}[\mathbf{G}_N(\rho_0)], \end{cases} \tag{2.7}$$

and that $\lim_{N\to\infty}\frac{1}{N}\text{E}[S_N^{\text{c}}(\theta_0)] \neq 0$ with a fixed T. This suggests that $\text{plim}_{N\to\infty}\frac{1}{N}S_N^{\text{c}}(\theta_0) \neq 0$, and therefore $\hat{\theta}_{\texttt{QML}}$ cannot be consistent when T is fixed. When T goes large with n, consistency can be achieved but one can show that the limiting distribution of $\sqrt{N}(\hat{\theta}_{\texttt{QML}} - \theta_0)$ is a non-centered normal, suggesting that $\hat{\theta}_N$ has a bias of order $\frac{1}{\sqrt{N}}$.

Note that $\mathrm{E}[S_N^c(\theta_0)]$ depends only on the common parameters θ_0 and the observables. It therefore offers a feasible way to analytically correct the CQS functions to give a set of unbiased estimating functions, or the *adjusted quasi score* (AQS) functions, as $S_N^*(\theta_0) = S_N^c(\theta_0) - \mathrm{E}[S_N^c(\theta_0)]$, which takes the form at the general θ:

$$S_N^*(\theta) = \begin{cases} \frac{1}{\sigma_v^2}\mathbf{X}'\mathbf{B}_N'(\rho)\tilde{\mathbf{V}}(\beta,\delta), \\ \frac{1}{2\sigma_v^4}\big[\tilde{\mathbf{V}}'(\beta,\delta)\tilde{\mathbf{V}}(\beta,\delta) - N_1\sigma_v^2\big], \\ \frac{1}{\sigma_v^2}\mathbf{Y}'\mathbf{W}'\mathbf{B}_N'(\rho)\tilde{\mathbf{V}}(\beta,\delta) - \mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho)\mathbf{F}_N(\lambda)\mathbf{B}_N^{-1}(\rho)], \\ \frac{1}{\sigma_v^2}\tilde{\mathbf{V}}'(\beta,\delta)\mathbf{G}_N(\rho)\tilde{\mathbf{V}}(\beta,\delta) - \mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{G}_N(\rho)], \end{cases} \tag{2.8}$$

where $N_1 = N - n - T + 1$, the *effective* sample size after taking into account the estimation of fixed effects. Solving the AQS equations: $S_N^*(\theta) = 0$, gives the AQS estimator of θ, i.e., $\hat{\theta}_N^* = \arg\{S_N^*(\theta) = 0\}$. It is easy to verity that $\mathrm{E}[S_N^*(\theta_0)] = 0$ and $\mathrm{plim}\frac{1}{N}S_N^*(\theta_0) = 0$, making it possible for $\hat{\theta}_N^*$ to be $\sqrt{N_1}$-consistent with a proper limiting distribution.

The AQS approach falls in spirit to the "*Modified Equations of Maximum Likelihood*" of Neyman and Scott (1948, Sec. 5), in searching for a potential method to handle the incidental parameters problem. Its generality and versatility in dealing with the incidental parameters problems have been demonstrated by recent works: Baltagi and Yang (2013a,b), Liu and Yang (2015, 2020), Yang (2018), Li and Yang (2020a,b) and Xu and Yang (2020).[6] In the special case of a balanced panel with time-invariant and row-normalized spatial weight matrices, our AQS method is equivalent to the QML method of Lee and Yu (2010) based on orthonormal transformations, with effective sample size $N_1 = N - n - T + 1 = (n-1)(T-1)$.

The root-finding process for the AQS estimation can be simplified by first solving the equations for β and σ_v^2, giving the constrained AQS estimators of β and σ_v^2:

$$\hat{\beta}_N^*(\delta) = [\mathbb{X}'(\rho)\mathbb{X}(\rho)]^{-1}\mathbb{X}'(\rho)\mathbf{C}_N(\delta)\mathbf{Y} \quad \text{and} \quad \hat{\sigma}_{v,N}^{*2}(\delta) = \tfrac{1}{N_1}\hat{\mathbf{V}}'(\delta)\hat{\mathbf{V}}(\delta), \tag{2.9}$$

where $\mathbb{X}(\rho) = \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho)\mathbf{X}$, $\mathbf{C}_N(\delta) = \mathbf{B}_N(\rho)\mathbf{A}_N(\lambda)$ and $\hat{\mathbf{V}}(\delta) = \tilde{\mathbf{V}}(\hat{\beta}_N^*(\delta),\delta)$. Substituting $\hat{\beta}_N^*(\delta)$ and $\hat{\sigma}_{v,N}^{*2}(\delta)$ back into (2.8) gives the concentrated AQS functions of δ:

[6] Clearly, this approach falls into the M-estimation method, and it is also a method of moments under the '*just identified*' scenario. Therefore, the resulting estimator is also called the M-estimator or MM estimator. The AQS approach offers a special way of finding the 'right set' of estimating equations or moment conditions.

$$S_N^{*c}(\delta) = \begin{cases} \frac{1}{\hat{\sigma}_{v,N}^{*2}(\delta)} \mathbf{Y}'\mathbf{W}'\mathbf{B}_N'(\rho)\hat{\mathbf{V}}(\delta) - \mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho)\mathbf{F}_N(\lambda)\mathbf{B}_N^{-1}(\rho)], \\ \frac{1}{\hat{\sigma}_{v,N}^{*2}(\delta)} \hat{\mathbf{V}}'(\delta)\mathbf{G}_N(\rho)\hat{\mathbf{V}}(\delta) - \mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{G}_N(\rho)]. \end{cases} \tag{2.10}$$

Solving the concentrated estimating (or AQS) equations, $S_N^{*c}(\delta) = 0$, we obtain the unconstrained AQS estimator $\hat{\delta}_N^*$ of δ. Thus the unconstrained AQS estimators of β and σ_v^2 are $\hat{\beta}_N^* \equiv \hat{\beta}_N^*(\hat{\delta}_N^*)$ and $\hat{\sigma}_{v,N}^{*2} \equiv \hat{\sigma}_{v,N}^{*2}(\hat{\delta}_N^*)$. The AQS estimator of θ is thus $\hat{\theta}_N^* = (\hat{\beta}_N^{*\prime}, \hat{\sigma}_{v,N}^{*2}, \hat{\delta}_N^*)'$.

From the above developments, we see that a big advantage of this method is that it provides a consistent estimation of all estimators including σ_v^2 with the joint asymptotic distribution of the AQS estimators being centered as long as N is large. Therefore, all the problems associated with the incidental parameters are gone. Furthermore, we do not have any restriction on the proportion of n and T as they go to infinity, and T (or n) can be even fixed. As this method is based on the adjusted quasi score functions, it may inherit the nice properties from the maximum likelihood estimation. It is well known that ML estimators often have better finite-samples properties than GMM/IV estimators. See also Hsiao (2018) for more discussions on the advantages of the quasi-likelihood approach compared with GMM estimation.

2.4. Asymptotic Properties of the AQS Estimators

Denote a parametric quantity evaluated at the true parameter values by dropping its argument(s), e.g., $\mathbf{A}_N \equiv \mathbf{A}_N(\lambda_0)$, $\mathbf{B}_N \equiv \mathbf{B}_N(\rho_0)$, and $\mathbf{C}_N \equiv \mathbf{C}_N(\delta_0)$. Let Δ be the parameter space for δ, and Δ_λ and Δ_ρ be the sub-spaces for λ and ρ, respectively. Consistency and asymptotic normality of the proposed AQS estimators for the unbalanced FE-SPD model are established under the following set of regularity conditions.

Assumption A: *The innovations v_{it} are iid for all i and t with mean zero, variance σ_{v0}^2, and $E|v_{it}|^{4+\epsilon_0} < \infty$ for some $\epsilon_0 > 0$.*

Assumption B: *The space Δ is compact, and the true parameters δ_0 lie in its interior.*

Assumption C: *(i) The elements of $\mathbf{X}$ are non-stochastic and bounded, uniformly in i and t, and (ii) $\lim_{N\to\infty} \frac{1}{N}\mathbb{X}'(\rho)\mathbb{X}(\rho)$ exists and is non-singular, uniformly in $\rho \in \Delta_\rho$.*

Assumption D: *$\{W_t\}$ and $\{M_t\}$ are known time-varying matrices. $\mathbf{W}$ and $\mathbf{M}$ are such that (i) their elements are at most of uniform order h_n^{-1} such that $\frac{h_n}{n} \to 0$, as $n \to \infty$; (ii) their diagonal elements are zero; and (iii) $\|\mathbf{W}\|_\infty$, $\|\mathbf{W}\|_1$, $\|\mathbf{M}\|_\infty$, and $\|\mathbf{M}\|_1$ are all bounded.*

Assumption E: *For $\mathbb{A}(\varpi) = \mathbf{A}_N(\lambda)$ or $\mathbf{B}_N(\rho)$ with $\varpi = \lambda$ or ρ,*

(i) both $\|\mathbb{A}^{-1}(\varpi_0)\|_\infty$ and $\|\mathbb{A}^{-1}(\varpi_0)\|_1$ are bounded;

(*ii*) *either* $\|\mathbb{A}^{-1}(\varpi)\|_\infty$ *or* $\|\mathbb{A}^{-1}(\varpi)\|_1$ *is bounded, uniformly in* $\varpi \in \Delta_\varpi$*;*

(*iii*) $0 < \underline{c}_\varpi \leq \inf_{\varpi\in\Delta_\varpi} \gamma_{\min}[\mathbb{A}'(\varpi)\mathbb{A}(\varpi)] \leq \sup_{\varpi\in\Delta_\varpi} \gamma_{\max}[\mathbb{A}'(\varpi)\mathbb{A}(\varpi)] \leq \bar{c}_\varpi < \infty$*;*

(*iv*) $B_s(\rho)D_s[\frac{1}{T}\sum_{t=1}^T D_t'B_t'(\rho)J_t(\rho)B_t(\rho)D_t]^{-1}D_t'B_t'(\rho)$ *is bounded in both row and column sum norms, uniformly in* $\rho \in \Delta_\rho$ *for all* s *and* t*, where* $B_t(\rho) = I_{n_t} - \rho M_t$ *for* $t = 1, \ldots, T$*, and* $J_t(\rho) = I_{n_1}$ *for* $t = 1$*, and* $I_{n_t} - B_t(\rho)l_{n_t}[l_{n_t}'B_t'(\rho)B_t(\rho)l_{n_t}]^{-1}l_{n_t}'B_t'(\rho)$ *for* $t = 2, \ldots, T$*.*

Assumption F: (*i*) *n is large (T is large or small),* (*ii*) *$\forall t$, n_t increases with n in the same rate, and* (*iii*) *all spatial units are observed at least twice over a total of T periods.*

Assumptions A-E are standard in the spatial econometrics literature (see, e.g., Lee, 2004a; Lee and Yu, 2010; Yang, 2018) except Assumption E(*iv*). With this additional condition, Lemma A.3 shows that $\|\mathbb{Q}_{\mathbb{D}}(\rho)\|_1$ and $\|\mathbb{Q}_{\mathbb{D}}(\rho)\|_\infty$ are bounded uniformly in $\rho \in \Delta_\rho$, which is necessary to facilitate the study of the asymptotic properties of the spatial parameter estimators. Assumption E(*iv*) is not restrictive as it holds for a special balanced panel.[7] Assumption F(*i*) allows (*a*) both n and T are large and (*b*) n is large and T is finite. Both scenarios encounter the so-called incidental parameters problem of Neyman and Scott (1948) due to the direct estimation of the individual and time fixed effects. The former leads to the asymptotic bias and the latter the inconsistency in the estimation of the structural parameters. As the usual transformation method is inapplicable to handle the incidental parameters problem in the unbalanced panels, a new (AQS) method is therefore introduced. Assumption F(*ii*) requires that each n_t increases with n, indicating that the number of observed individual should not be too small relative to n in each period. Assumption F(*iii*) ensures that the spatial structure is complete after μ being concentrated out.

We first prove the consistency of $\hat{\delta}_N^*$. The key step in the proof is to compare $S_N^{*c}(\delta)$ with its population counterpart. Let $\bar{S}_N^*(\theta) = \mathrm{E}[S_N^*(\theta)]$. Given δ, $\bar{S}_N^*(\theta) = 0$ is partially solved at

$$\bar{\beta}_N^*(\delta) = [\mathbb{X}'(\rho)\mathbb{X}(\rho)]^{-1}\mathbb{X}'(\rho)\mathbf{C}_N(\delta)\mathrm{E}(\mathbf{Y}) \quad \text{and} \quad \bar{\sigma}_{v,N}^{*2}(\delta) = \tfrac{1}{N_1}\mathrm{E}[\bar{\mathbf{V}}'(\delta)\bar{\mathbf{V}}(\delta)], \tag{2.11}$$

where $\bar{\mathbf{V}}(\delta) = \tilde{\mathbf{V}}(\bar{\beta}_N^*(\delta), \delta) = \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho)[\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\bar{\beta}_N^*(\delta)]$. Substituting $\bar{\beta}_N^*(\delta)$ and $\bar{\sigma}_{v,N}^{*2}(\delta)$ into the δ-component of $\bar{S}_N^*(\theta)$, we obtain the population counterpart of $S_N^{*c}(\delta)$ as

$$\bar{S}_N^{*c}(\delta) = \begin{cases} \frac{1}{\bar{\sigma}_{v,N}^{*2}(\delta)}\mathrm{E}[\mathbf{Y}'\mathbf{W}'\mathbf{B}_N'(\rho)\bar{\mathbf{V}}(\delta)] - \mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho)\mathbf{F}_N(\lambda)\mathbf{B}_N^{-1}(\rho)], \\ \frac{1}{\bar{\sigma}_{v,N}^{*2}(\delta)}\mathrm{E}[\bar{\mathbf{V}}'(\delta)\mathbf{G}_N(\rho)\bar{\mathbf{V}}(\delta)] - \mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{G}_N(\rho)]. \end{cases} \tag{2.12}$$

[7] For a balanced panel with a time-invariant and row-normalized spatial weight matrix, we have for all t, $n_t = n$, $D_t = I_n$, $M_t = M$, and $B_t(\rho) = I_n - \rho M \equiv B(\rho)$. As $M \times l_n = l_n$, $J_t(\rho) = I_n - \frac{1}{n}l_nl_n'$, $t = 2, \ldots, T$. Thus, $B_s(\rho)D_s[\frac{1}{T}\sum_{t=1}^T D_t'B_t'(\rho)J_t(\rho)B_t(\rho)D_t]^{-1}D_t'B_t'(\rho) = (I_n - \frac{T-1}{nT}l_nl_n')^{-1}$. As $I_n - \frac{T-1}{nT}l_nl_n'$ is strictly diagonally dominant in rows and columns, its inverse is bounded in row and column sum norms (Varah, 1975).

Clearly, $S_N^{*c}(\hat{\delta}_N^*) = 0$ by construction. Also, it is easy to see that $\bar{S}_N^{*c}(\delta_0) = 0$ as $\bar{\beta}_N^*(\delta_0) = \beta_0$ and $\bar{\sigma}_{v,N}^{*2}(\delta_0) = \sigma_{v0}^2$. Thus, by theorem 5.9 of van der Vaart (1988), $\hat{\delta}_N^*$ will be consistent for δ_0 if $\sup_{\delta\in\Delta} \frac{1}{N_1} \left\| S_N^{*c}(\delta) - \bar{S}_N^{*c}(\delta) \right\| \xrightarrow{p} 0$ and the following identification condition holds:

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

Assumption G is a high level assumption being put up for simplicity of presentation. It can be shown to be true under some low level conditions. We have (see (B.5), Appendix B),

$$\bar{\sigma}_{v,N}^{*2}(\delta) = \tfrac{1}{N_1}\eta'\mathbf{A}_N'^{-1}\mathcal{Q}_N'(\delta)\mathcal{Q}_N(\delta)\mathbf{A}_N^{-1}\eta + \tfrac{\sigma_{v0}^2}{N_1}\mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho)\mathcal{C}_N(\delta)],$$

where $\eta = \mathbf{X}\beta_0 + \mathbf{D}\phi_0$, $\mathcal{Q}_N(\delta) = \mathbb{Q}_{\mathbb{X}}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{C}_N(\delta)$, and $\mathcal{C}_N(\delta) = \mathbf{C}_N(\delta)(\mathbf{C}_N'\mathbf{C}_N)^{-1}\mathbf{C}_N'(\delta)$. A sufficient condition for Assumption G to hold is either (a) or (b) holds, where

(a) $\frac{1}{\bar{\sigma}_{v,N}^{*2}(\delta)}\eta'\mathbf{F}_N'\mathbf{B}_N'(\rho)\mathcal{Q}_N(\delta)\mathbf{A}_N^{-1}\eta + \mathtt{tr}[\frac{\sigma_{v0}^2}{\bar{\sigma}_{v,N}^{*2}(\delta)}\mathcal{P}_1(\delta) - \mathcal{P}_2(\delta)] \neq 0$, for $\delta \neq \delta_0$,

(b) $\frac{1}{\bar{\sigma}_{v,N}^{*2}(\delta)}\eta'\mathbf{A}_N'^{-1}\mathcal{Q}_N'(\delta)\mathbf{G}_N(\rho)\mathcal{Q}_N(\delta)\mathbf{A}_N^{-1}\eta + \mathtt{tr}[\frac{\sigma_{v0}^2}{\bar{\sigma}_{v,N}^{*2}(\delta)}\mathcal{P}_3(\rho)\mathcal{C}_N(\delta) - \mathcal{P}_3(\rho)] \neq 0$, for $\delta \neq \delta_0$,

with $\mathcal{P}_1(\delta) = \mathbf{C}_N'^{-1}\mathbf{C}_N'(\delta)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho)\mathbf{F}_N\mathbf{B}_N^{-1}$, $\mathcal{P}_2(\delta) = \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho)\mathbf{F}_N(\lambda)\mathbf{B}_N^{-1}(\rho)$, and $\mathcal{P}_3(\rho) = \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{G}_N(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)$. It is easy to see that $\mathcal{Q}_N(\delta_0)\mathbf{A}_N^{-1}\eta = 0$, $\mathcal{C}_N(\delta_0) = I_N$ and $\bar{\sigma}_{v,N}^{*2}(\delta_0) = \sigma_{v0}^2$. Hence the two quantities in (a) and (b) are identical 0 at the true parameter values. Once the consistency of $\hat{\delta}_N^*$ is established, the consistency of $\hat{\beta}_N^*$ and $\hat{\sigma}_{v,N}^{*2}$ follows by Assumptions C-E.

Theorem 2.1. *Suppose Assumptions A-G hold. We have, as $N \to \infty$, $\hat{\theta}_N^* \xrightarrow{p} \theta_0$.*

To derive the asymptotic distribution of $\hat{\theta}_N^*$, we apply the mean value theorem: $0 = S_N^*(\hat{\theta}_N^*) = S_N^*(\theta_0) + \frac{\partial}{\partial\theta'}S_N^*(\bar{\theta})(\hat{\theta}_N^* - \theta_0)$, where $\bar{\theta}$ lies between $\hat{\theta}_N^*$ and θ_0, and its value varies over the rows of $\frac{\partial}{\partial\theta'}S_N^*(\bar{\theta})$. Using $\tilde{\mathbf{V}}(\beta_0, \delta_0) = \mathbb{Q}_{\mathbb{D}}\mathbf{V}$ and $\mathbf{Y} = \mathbf{A}_N^{-1}(\eta + \mathbf{B}_N^{-1}\mathbf{V})$,

$$S_N^*(\theta_0) = \begin{cases} \frac{1}{\sigma_{v0}^2}\mathbb{X}'\mathbf{V}, \\ \frac{1}{2\sigma_{v0}^4}(\mathbf{V}'\mathbb{Q}_{\mathbb{D}}\mathbf{V} - N_1\sigma_v^2), \\ \frac{1}{\sigma_{v0}^2}\mathbf{V}'\mathcal{P}_2\mathbf{B}_N\eta + \frac{1}{\sigma_{v0}^2}\mathbf{V}'\mathcal{P}_2\mathbf{V} - \mathtt{tr}(\mathcal{P}_2), \\ \frac{1}{\sigma_{v0}^2}\mathbf{V}'\mathcal{P}_3\mathbf{V} - \mathtt{tr}(\mathcal{P}_3), \end{cases} \tag{2.13}$$

and its asymptotic normality is proved by the central limit theorem (CLT) for linear-quadratic (LQ) forms of Kelejian and Prucha (2001). This together with the proper asymptotic behavior of the 'Hessian' matrix, $\frac{\partial}{\partial\theta'}S_N^*(\theta)$ (given in (B.4), Appendix B), lead to the following theorem.

Theorem 2.2. *Under Assumptions A-G, we have, as $N \to \infty$,*

$$\sqrt{N_1}(\hat{\theta}_N^* - \theta_0) \xrightarrow{D} N\Big(0, \lim_{N\to\infty} \Sigma_N^{*-1}(\theta_0)\Gamma_N^*(\theta_0)\Sigma_N^{*-1}(\theta_0)\Big),$$

where $\Sigma_N^*(\theta_0) = -\frac{1}{N_1}\mathrm{E}[\frac{\partial}{\partial\theta'}S_N^*(\theta_0)]$ *and* $\Gamma_N^*(\theta_0) = \frac{1}{N_1}\mathrm{Var}[S_N^*(\theta_0)]$*, both assumed to exist and* $\Sigma_N^*(\theta_0)$ *assumed to be positive definite for sufficiently large* N*.*

2.5. Inference based on AQS estimation

To conduct inferences for θ based on the proposed AQS estimators, consistent estimates of $\Sigma_N^*(\theta_0)$ and $\Gamma_N^*(\theta_0)$ are needed. The analytical expression of $\Sigma_N^*(\theta)$ can easily be obtained from the Hessian matrix $\frac{\partial}{\partial\theta'}S_N^*(\theta)$ that is given in (B.4). As it depends only on the common parameters θ, a simple plug-in estimator $\Sigma_N^*(\hat\theta_N^*)$ can be used to consistently estimate $\Sigma_N^*(\theta_0)$. Alternatively, a simpler sample analogue of $\Sigma_N^*(\theta)$ also provides a consistent estimator:

$$\widehat{\Sigma}_N^* = -\frac{1}{N_1}\frac{\partial}{\partial\theta'}S_N^*(\theta)\Big|_{\theta=\hat\theta_N^*}. \tag{2.14}$$

The consistency of $\Sigma_N^*(\hat\theta_N^*)$ or $\widehat{\Sigma}_N^*$ is proved in the proof of Theorem 2.2.

Now, using Lemma A.5 with iid errors, one derives $\Gamma_N^*(\theta_0)$, which has the distinct elements:

$$\begin{aligned}
N_1\Gamma^*_{\beta\theta} &= \big[\tfrac{1}{\sigma_{v0}^2}\mathbb{X}'\mathbb{X},\ \tfrac{\gamma}{2\sigma_{v0}^3}\mathbb{X}'q,\ \tfrac{\gamma}{\sigma_{v0}}\mathbb{X}'p_2 + \tfrac{1}{\sigma_{v0}^2}\mathbb{X}'\mathcal{P}_2\mathbf{B}_N\eta,\ \tfrac{\gamma}{2\sigma_{v0}}\mathbb{X}'p_3\big],\\
N_1\Gamma^*_{\sigma_v^2\sigma_v^2} &= \tfrac{1}{4\sigma_{v0}^4}(2N_1 + \kappa q'q),\\
N_1\Gamma^*_{\sigma_v^2\lambda} &= \tfrac{\gamma}{2\sigma_{v0}^3}q'\mathcal{P}_2\mathbf{B}_N\eta + \tfrac{1}{2\sigma_{v0}^2}[2\mathtt{tr}(\mathcal{P}_2\mathbb{Q}_{\mathbb{D}}) + \kappa q'p_2],\\
N_1\Gamma^*_{\sigma_v^2\rho} &= \tfrac{1}{2\sigma_{v0}^2}[2\mathtt{tr}(\mathcal{P}_3\mathbb{Q}_{\mathbb{D}}) + \kappa q'p_3],\\
N_1\Gamma^*_{\lambda\lambda} &= \tfrac{1}{\sigma_{v0}^2}\eta'\mathbf{B}_N'\mathcal{P}_2'\mathcal{P}_2\mathbf{B}_N\eta + \tfrac{2\gamma}{\sigma_{v0}}p_2'\mathcal{P}_2\mathbf{B}_N\eta + \mathtt{tr}(\mathcal{P}_2\mathcal{P}_2^\circ) + \kappa p_2'p_2,\\
N_1\Gamma^*_{\lambda\rho} &= \mathtt{tr}(\mathcal{P}_3\mathcal{P}_2^\circ) + \kappa p_2'p_3 + \tfrac{\gamma}{\sigma_{v0}}p_3'\mathcal{P}_2\mathbf{B}_N\eta,\\
N_1\Gamma^*_{\rho\rho} &= \mathtt{tr}(\mathcal{P}_3\mathcal{P}_3^\circ) + \kappa p_3'p_3,
\end{aligned} \tag{2.15}$$

where $p_r = \mathtt{diagv}(\mathcal{P}_r), r = 2, 3$, and $q = \mathtt{diagv}(\mathbb{Q}_{\mathbb{D}})$. This shows clearly that the estimation of $\Gamma_N^*(\theta_0)$ is more complicated as $\Gamma_N^*(\theta_0)$ contains not only the common parameters θ, but also the fixed effects ϕ embedded in η, and the skewness γ and the excess kurtosis κ of the idiosyncratic errors. Thus, the common plug-in approach may not provide a valid estimate.

Let $\hat\phi_N^*$ be the AQS estimator of ϕ, obtained through (2.4), i.e., $\hat\phi_N^* = \hat\phi_N(\hat\beta_N^*, \hat\delta_N^*)$. Let $\hat\gamma_N$ and $\hat\kappa_N$ be consistent estimators of γ and κ. Let $\Gamma_N^*(\hat\theta_N^*) = \Gamma_N^*(\theta)|_{(\theta=\hat\theta_N^*, \phi=\hat\phi_N^*, \gamma=\hat\gamma_N, \kappa=\hat\kappa_N)}$ be the plug-in estimator. When both n and T are large, $\Gamma_N^*(\hat\theta_N^*)$ would be consistent as $\hat\phi_N^*$ is. However, when n is large but T is fixed, $\hat\phi_N^*$ (its component $\hat\mu_N^*$) is not consistent. Plugging $\hat\mu_N^*$ into $\Gamma_N^*(\theta)$ will induce a bias (inconsistency), and a bias correction is necessary.

From the expression of $\Gamma_N^*(\theta_0)$ given above, we see that only the λ-components involve ϕ through η, which may not be consistently estimated by the plug-in method. We can further show that the components of $\Gamma_N^*(\theta_0)$ linear in ϕ can also be consistently estimated by the plug-in method. Therefore, the only term that cannot be consistently estimated by the plug-in method is $\frac{1}{\sigma_{v0}^2}\eta'\mathbf{B}_N'\mathcal{P}_2'\mathcal{P}_2\mathbf{B}_N\eta$ associated with the λ-λ component of $\Gamma_N^*(\theta_0)$. We have the following corollary. See its proof in Appendix D for details on these discussions.

Corollary 2.1. *Under the assumptions of Theorem 2.2, we have,*

$$\Gamma_N^*(\hat\theta_N^*) = \Gamma_N^*(\theta_0) + \text{Bias}^*(\delta_0) + o_p(1),$$

where $\text{Bias}^*(\delta_0)$ *is a* $(k+3)\times(k+3)$ *matrix having zero entries everywhere except the* λ-λ *entry, which takes the form* $\frac{1}{N_1}\mathtt{tr}(\mathcal{P}_2'\mathcal{P}_2\mathbb{P}_{\mathbb{D}})$.

The result of Corollary 2.1 leads immediately a general consistent estimator of $\Gamma_N^*(\theta_0)$:

$$\widehat\Gamma_N^* = \Gamma_N^*(\hat\theta_N^*) - \text{Bias}^*(\hat\delta_N^*). \tag{2.16}$$

Then, it is only left to find consistent estimators for γ and κ. Since we cannot 'consistently' estimate $\mathbf{V} = \mathbf{B}_N(\mathbf{A}_N\mathbf{Y} - \eta)$ due to the incidental parameters problem, we start from $\tilde{\mathbf{V}} = \mathbb{Q}_{\mathbb{D}}\mathbf{V}$, which can be 'consistently' estimated by $\hat{\mathbf{V}} = \mathbb{Q}_{\mathbb{D}}(\hat\rho_N^*)\mathbf{B}_N(\hat\rho_N^*)[\mathbf{A}_N(\hat\lambda_N^*)\mathbf{Y} - \mathbf{X}\hat\beta_N^*]$. Let q_{jk} be the (j,k)th element of $\mathbb{Q}_{\mathbb{D}}$. Denote the elements of $\mathbf{V}$ by v_j, and the elements of $\tilde{\mathbf{V}}$ by $\tilde v_j, j = 1,\dots,N$, where j is the combined index for $i = 1,\dots,n_t$ and $t = 1,\dots,T$. Then, $\tilde v_j = q_{j1}v_1 + q_{j2}v_2 + \cdots + q_j v_N$, and thus,

$$\text{E}(\tilde v_j^3) = \textstyle\sum_{k=1}^N q_{jk}^3 \text{E}(v_k^3) = \sigma_v^3\gamma\sum_{k=1}^N q_{jk}^3,\ j = 1,\dots,N.$$

Summing $\text{E}(\tilde v_j^3)$ over j, we obtain $\gamma = \big(\sum_{j=1}^N \text{E}(\tilde v_j^3)\big)\big(\sigma_v^3\sum_{j=1}^N\sum_{k=1}^N q_{jk}^3\big)^{-1}$, and its sample analogue gives a consistent estimator of γ:

$$\hat\gamma_N = \frac{\sum_{j=1}^N \hat v_j^3}{\hat\sigma_{v,N}^{*3}\sum_{j=1}^N\sum_{k=1}^N \hat q_{jk}^3}. \tag{2.17}$$

where $\hat v_j$ is the jth element of $\hat{\mathbf{V}}(\hat\beta_N^*, \hat\lambda_N^*)$ and $\hat q_{jk}$ is the (j,k)th element of $\mathbb{Q}_{\mathbb{D}}(\hat\rho_N^*)$. Similarly,

$$\begin{aligned}\text{E}(\tilde v_j^4) &= \textstyle\sum_{k=1}^N q_{jk}^4\text{E}(v_k^4) + 3\sigma_v^4\sum_{k=1}^N\sum_{l=1}^N q_{jk}^2 q_{jl}^2 - 3\sigma_v^4\sum_{k=1}^N q_{jk}^4 \\ &= \textstyle\sum_{k=1}^N q_{jk}^4\kappa\sigma_v^4 + 3\sigma_v^4\sum_{k=1}^N\sum_{l=1}^N q_{jk}^2 q_{jl}^2,\end{aligned}$$

which gives $\kappa = \big(\sum_{j=1}^N \text{E}(\tilde v_j^4) - 3\sigma_v^4\sum_{j=1}^N\sum_{k=1}^N\sum_{l=1}^N q_{jk}^2q_{jl}^2\big)\big(\sigma_v^4\sum_{j=1}^N\sum_{k=1}^N q_{jk}^4\big)^{-1}$ by summing $\text{E}(\tilde v_j^4)$ over j. Hence, a consistent estimator for κ is

$$\hat\kappa_N = \frac{\sum_{j=1}^N \hat v_j^4 - 3\hat\sigma_{v,N}^{*4}\sum_{j=1}^N\sum_{k=1}^N\sum_{l=1}^N \hat q_{jk}^2\hat q_{jl}^2}{\hat\sigma_{v,N}^{*4}\sum_{j=1}^N\sum_{k=1}^N \hat q_{jk}^4}. \tag{2.18}$$

Corollary 2.2. *Under Assumptions A-G, we have, as $N \to \infty$,*

(i) $\hat{\gamma}_N \xrightarrow{p} \gamma_0$ *and* $\hat{\kappa}_N \xrightarrow{p} \kappa_0$*; (ii)* $\widehat{\Sigma}_N^* - \Sigma_N^*(\theta_0) \xrightarrow{p} 0$ *and* $\widehat{\Gamma}_N^* - \Gamma_N^*(\theta_0) \xrightarrow{p} 0$*; and therefore* $\widehat{\Sigma}_N^{*-1}\widehat{\Gamma}_N^*\widehat{\Sigma}_N^{*-1} - \Sigma_N^{*-1}(\theta_0)\Gamma_N^*(\theta_0)\Sigma_N^{*-1}(\theta_0) \xrightarrow{p} 0$.

3. Unbalanced FE-SPD Model with Heteroskedasticity

Cross-sectional heteroskedasticity is rather common in spatial regression models due to misspecification, peer interaction, aggregation, clustering, etc. (Anselin, 1988; Liu and Yang, 2015). The same is true for SPD or unbalanced SPD models. Robust methods have been introduced for SPD models, but are limited to balanced panels with cross-sectional heteroskedasticity only (Moscone and Tosetti, 2011; Baltagi and Yang, 2013b; Badinger and Egger, 2015; Liu and Yang, 2020). Time-series heteroskedasticity is also important, in particular in short panels (Alvarez and Arellano, 2004; Bai, 2013). Therefore, it is highly desirable to extend the set of estimation and inference methods introduced in Section 2 to allow for unknown spatiotemporal heteroskedasticity as specified in the extended assumption below.

Assumption A$'$: The innovations v_j (or v_{it}) are independently but not identically distributed (inid), i.e., $\{v_j\} \sim inid(0, \sigma_j^2)$, and $E|v_j|^{4+\epsilon_0} < \infty$ for some $\epsilon_0 > 0$.

Assumption A$'$ relaxes Assumption A by allowing the variance of the idiosyncratic error to vary freely across cross-section and over time. As $\mathrm{E}[S_N^*(\theta_0)] \neq 0$ under Assumption A$'$, we need to readjust score functions (2.6) to make it centered under unknown heteroskedasticity.

3.1. AQS Estimation under Unknown Heteroskedasticity

Denote $\mathbf{H} = \mathtt{diag}(\sigma_1^2, \sigma_2^2, \cdots, \sigma_N^2)$, and hence $\mathrm{Var}(\mathbf{V}) = \mathbf{H}$. As in Liu and Yang (2015, 2020), we modify the relevant components of the CQS vector $S_N^{\mathrm{c}}(\theta)$ given in (2.6), so that their expectations at the true parameter θ_0 are zero under unknown heteroskedasticity.

First, consider the stochastic element of the λ-component of $S_N^{\mathrm{c}}(\theta)$ given in (2.6). Define $\bar{\mathbf{F}}_N(\delta) = \mathbf{B}_N(\rho)\mathbf{F}_N(\lambda)\mathbf{B}_N^{-1}(\rho)$, and as usual denote $\bar{\mathbf{F}}_N = \bar{\mathbf{F}}_N(\delta_0)$. Using $\tilde{\mathbf{V}}(\beta_0, \delta_0) = \mathbb{Q}_{\mathbb{D}}\mathbf{V}$ and $\mathbf{B}_N\mathbf{W}\mathbf{Y} = \bar{\mathbf{F}}_N\mathbf{C}_N\mathbf{Y}$, and noting $\mathbf{C}_N\mathbf{Y} = \mathbf{B}_N\eta + \mathbf{V}$ and $\eta = \mathbf{X}\beta_0 + \mathbf{D}\phi_0$, we have,

$$\begin{aligned}
\mathrm{E}[\mathbf{Y}'\mathbf{W}'\mathbf{B}_N'\tilde{\mathbf{V}}(\beta_0, \delta_0)] &= \mathrm{E}(\mathbf{Y}'\mathbf{C}_N'\bar{\mathbf{F}}_N'\mathbb{Q}_{\mathbb{D}}\mathbf{V}) = \mathtt{tr}(\mathbf{H}\bar{\mathbf{F}}_N'\mathbb{Q}_{\mathbb{D}}) = \mathtt{tr}[\mathbf{H}\ \mathtt{diag}(\bar{\mathbf{F}}_N'\mathbb{Q}_{\mathbb{D}})] \\
&= \mathtt{tr}[\mathbf{H}\ \mathtt{diag}(\bar{\mathbf{F}}_N'\mathbb{Q}_{\mathbb{D}})\ \mathtt{diag}(\mathbb{Q}_{\mathbb{D}})^{-1}\mathbb{Q}_{\mathbb{D}}] = \mathrm{E}(\mathbf{Y}'\mathbf{C}_N'\bar{\mathbb{F}}_N'\mathbb{Q}_{\mathbb{D}}\mathbf{V}),
\end{aligned}$$

where $\bar{\mathbb{F}}_N' = \bar{\mathbb{F}}_N'(\delta_0)$ and $\bar{\mathbb{F}}_N'(\delta) = \mathtt{diag}[\bar{\mathbf{F}}_N'(\delta)\mathbb{Q}_{\mathbb{D}}(\rho)]\mathtt{diag}[\mathbb{Q}_{\mathbb{D}}(\rho)]^{-1}$. Taking the difference between the quantities inside the second expectation and the last expectation, we obtain:

$$\mathbf{Y}'\mathbf{C}_N'(\delta)[\bar{\mathbf{F}}_N'(\delta) - \bar{\mathbb{F}}_N'(\delta)]\tilde{\mathbf{V}}(\beta, \delta), \tag{3.1}$$

the adjusted λ-component of the CQS functions, having a zero expectation and a zero probability limit upon dividing by N at θ_0 under unknown heteroskedasticity.

Now, consider the stochastic element of the ρ-component of the CQS vector $S_N^c(\theta)$ given in (2.6). Similar to the above, we have,

$$\begin{aligned}
&\mathrm{E}(\tilde{\mathbf{V}}'\mathbf{G}_N\tilde{\mathbf{V}}) = \mathrm{E}(\mathbf{V}'\mathbb{Q}_{\mathbb{D}}\mathbf{G}_N\mathbb{Q}_{\mathbb{D}}\mathbf{V}) = \mathtt{tr}(\mathbf{H}\bar{\mathbf{G}}_N\mathbb{Q}_{\mathbb{D}}) = \mathtt{tr}[\mathbf{H}\ \mathtt{diag}(\bar{\mathbf{G}}_N\mathbb{Q}_{\mathbb{D}})] \\
&= \mathtt{tr}[\mathbf{H}\ \mathtt{diag}(\bar{\mathbf{G}}_N\mathbb{Q}_{\mathbb{D}})\ \mathtt{diag}(\mathbb{Q}_{\mathbb{D}})^{-1}\mathbb{Q}_{\mathbb{D}}] = \mathrm{E}(\mathbf{V}'\bar{\mathbb{G}}_N\mathbb{Q}_{\mathbb{D}}\mathbf{V}),
\end{aligned}$$

where $\bar{\mathbf{G}}_N(\rho) = \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{G}_N(\rho)$ and $\bar{\mathbb{G}}_N(\rho) = \mathtt{diag}[\bar{\mathbf{G}}_N(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)]\mathtt{diag}[\mathbb{Q}_{\mathbb{D}}(\rho)]^{-1}$. Replacing the $\mathbf{V}'$ in the second and last expectations by $[\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\beta]'\mathbf{B}_N'(\rho)$, and taking the difference between the two quantities inside the expectations, we obtain a robust AQS function for ρ:

$$[\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\beta]'\mathbf{B}_N'(\rho)[\bar{\mathbf{G}}_N(\rho) - \bar{\mathbb{G}}_N(\rho)]\tilde{\mathbf{V}}(\beta, \delta). \tag{3.2}$$

The β-component of $S_N^c(\theta)$ is automatically robust against the unknown heteroskedasticity. Therefore, the desired AQS functions robust against the unknown heteroskedasticity $\mathbf{H}$ are,

$$S_N^{\diamond}(\beta, \delta) = \begin{cases} \mathbb{X}'(\rho)\tilde{\mathbf{V}}(\beta, \delta), \\ \mathbf{Y}'\mathbf{C}_N'(\delta)[\bar{\mathbf{F}}_N'(\delta) - \bar{\mathbb{F}}_N'(\delta)]\tilde{\mathbf{V}}(\beta, \delta), \\ [\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\beta]'\mathbf{B}_N'(\rho)[\bar{\mathbf{G}}_N(\rho) - \bar{\mathbb{G}}_N(\rho)]\tilde{\mathbf{V}}(\beta, \delta). \end{cases} \tag{3.3}$$

Solving $S_N^{\diamond}(\beta, \delta) = 0$ gives the robust AQS (RAQS) estimators, $\hat{\beta}_N^{\diamond}$ and $\hat{\delta}_N^{\diamond}$, of β and δ.

Similarly, this root-finding process can be simplified by first solving for β given δ, to give the constrained estimator $\hat{\beta}_N^{\diamond}(\delta)$ and the concentrated RAQS functions:

$$S_N^{\diamond c}(\delta) = \begin{cases} \mathbf{Y}'\mathbf{C}_N'(\delta)[\bar{\mathbf{F}}_N'(\delta) - \bar{\mathbb{F}}_N'(\delta)]\hat{\mathbf{V}}(\delta), \\ [\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\hat{\beta}_N^{\diamond}(\delta)]'\mathbf{B}_N'(\rho)[\bar{\mathbf{G}}_N(\rho) - \bar{\mathbb{G}}_N(\rho)]\hat{\mathbf{V}}(\delta), \end{cases} \tag{3.4}$$

where $\hat{\beta}_N^{\diamond}(\delta) = \hat{\beta}_N^{*}(\delta)$ given in (2.9), and $\hat{\mathbf{V}}(\delta) = \tilde{\mathbf{V}}(\hat{\beta}_N^{\diamond}(\delta), \delta)$. Then, solving $S_N^{\diamond c}(\delta) = 0$, we obtain the RAQS estimator $\hat{\delta}_N^{\diamond}$ of δ, and thus the RAQS estimator $\hat{\beta}_N^{\diamond} \equiv \hat{\beta}_N^{\diamond}(\hat{\delta}_N^{\diamond})$ of β.

3.2. Asymptotic Properties of the Robust AQS Estimators

Similar to the case of the homoskedastic model, we first establish the consistency of $\hat{\delta}_N^{\diamond}$. Then, the consistency of $\hat{\beta}_N^{\diamond}$ follows. Let $\bar{S}_N^{\diamond}(\beta, \delta) = \mathrm{E}[S_N^{\diamond}(\beta, \delta)]$ be the population RAQS functions. Then, the β-component of $\bar{S}_N^{\diamond}(\beta, \delta) = 0$ is solved at $\bar{\beta}_N^{\diamond}(\delta) = \bar{\beta}_N^{*}(\delta)$ given in (2.11).

Upon substitution, we obtain the population counterpart of $S_N^{\diamond c}(\delta)$:

$$\bar{S}_N^{\diamond c}(\delta) = \begin{cases} \mathrm{E}[\mathbf{Y}'\mathbf{C}'_N(\delta)[\bar{\mathbf{F}}'_N(\delta) - \bar{\mathbb{F}}'_N(\delta)]\bar{\mathbf{V}}(\delta)], \\ \mathrm{E}\big\{\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\bar{\beta}_N^{\diamond}(\delta)]'\mathbf{B}'_N(\rho)[\bar{\mathbf{G}}_N(\rho) - \bar{\mathbb{G}}_N(\rho)]\bar{\mathbf{V}}(\delta)\big\}, \end{cases} \tag{3.5}$$

where $\bar{\mathbf{V}}(\delta) = \tilde{\mathbf{V}}(\bar{\beta}_N^{\diamond}(\delta), \delta)$. As $S_N^{\diamond c}(\hat{\delta}_N^{\diamond})$ and $\bar{S}_N^{\diamond c}(\delta_0)$ are both zero, by theorem 5.9 of van der Vaart (1988) $\hat{\delta}_N^{\diamond}$ will be consistent for δ_0 if $\sup_{\delta\in\Delta}\frac{1}{N_1}\left\|S_N^{\diamond c}(\delta) - \bar{S}_N^{\diamond c}(\delta)\right\| \xrightarrow{p} 0$ and the following identification condition holds:

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

Again, Assumption G′ is put up for simplicity. More primitive conditions under which Assumption G′ holds are that for $\delta \neq \delta_0$ either of the following conditions holds:
(*a*) $\eta'\mathbf{A}_N'^{-1}\mathbf{C}'_N(\delta)[\bar{\mathbf{F}}'_N(\delta) - \bar{\mathbb{F}}'_N(\delta)]\mathcal{Q}_N(\delta)\mathbf{A}_N^{-1}\eta + \mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho)\mathcal{C}_N^h(\delta)\big(\bar{\mathbf{F}}'_N(\delta) - \bar{\mathbb{F}}'_N(\delta)\big)] \neq 0$; or
(*b*) $\eta'\mathbf{A}_N'^{-1}\mathbf{C}'_N(\delta)\mathbb{M}'_N(\rho)\big[\bar{\mathbf{G}}_N(\rho) - \bar{\mathbb{G}}_N(\rho)\big]\mathcal{Q}_N(\delta)\mathbf{A}_N^{-1}\eta + \mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho)\mathcal{C}_N^h(\delta)\big(\bar{\mathbf{G}}_N(\rho) - \bar{\mathbb{G}}_N(\rho)\big)] \neq 0$,
where $\mathcal{C}_N^h(\delta) = \mathbf{C}_N(\delta)\mathbf{C}_N^{-1}\mathbf{H}\mathbf{C}_N^{-1\prime}\mathbf{C}'_N(\delta)$ and $\mathbb{M}_N(\rho) = \mathbf{I}_N - \mathbf{B}_N(\rho)\mathbf{X}[\mathbb{X}'(\rho)\mathbb{X}(\rho)]^{-1}\mathbb{X}'(\rho)$. Similarly, as $\mathcal{C}_N^h(\delta_0) = \mathbf{H}$ and $\mathcal{Q}_N(\delta_0)\mathbf{A}_N^{-1}\eta = 0$, the two quantities in (*a*) and (*b*) are 0 at δ_0.

Denote $\xi = (\beta', \delta')'$ and $\hat{\xi}_N^{\diamond} = (\hat{\beta}_N^{\diamond\prime}, \hat{\delta}_N^{\diamond\prime})'$. We have the following consistency theorem.

Theorem 3.1. *Under Assumptions A′, B-F and G′, we have, as $N \to \infty$, $\hat{\xi}_N^{\diamond} \xrightarrow{p} \xi_0$.*

Similarly, the asymptotic normality of $\hat{\xi}_N^{\diamond}$ can be established, by applying the mean value theorem to each element of $S_N^{\diamond}(\hat{\xi}_N^{\diamond}) = 0$ at ξ_0. The robust AQS function at ξ_0 is

$$S_N^{\diamond}(\xi_0) = \begin{cases} \mathbb{X}'\mathbf{V}, \\ \eta'\mathbf{B}'_N(\bar{\mathbf{F}}'_N - \bar{\mathbb{F}}'_N)\mathbb{Q}_{\mathbb{D}}\mathbf{V} + \mathbf{V}'(\bar{\mathbf{F}}'_N - \bar{\mathbb{F}}'_N)\mathbb{Q}_{\mathbb{D}}\mathbf{V}, \\ \phi_0'\mathbb{D}'_N(\bar{\mathbf{G}}_N - \bar{\mathbb{G}}_N)\mathbb{Q}_{\mathbb{D}}\mathbf{V} + \mathbf{V}'(\bar{\mathbf{G}}_N - \bar{\mathbb{G}}_N)\mathbb{Q}_{\mathbb{D}}\mathbf{V}, \end{cases} \tag{3.6}$$

which can also be verified to be asymptotically normal by using the CLT for LQ forms of Kelejian and Prucha (2001). The adjusted Hessian $\frac{\partial}{\partial\xi'}S_N^{\diamond}(\bar{\xi})$, shown in (C.1) in Appendix C, has a proper asymptotic behavior, for some $\bar{\xi}$ lying between $\hat{\xi}_N^{\diamond}$ and ξ_0 elementwise. Consequently, the asymptotic distribution for $\hat{\xi}_N^{\diamond}$ can be established in the following theorem.

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

$$\sqrt{N_1}(\hat{\xi}_N^{\diamond} - \xi_0) \xrightarrow{D} N\Big(0, \lim_{N\to\infty}\Sigma_N^{\diamond-1}(\xi_0)\Gamma_N^{\diamond}(\xi_0)\Sigma_N^{\diamond-1}(\xi_0)\Big),$$

where $\Sigma_N^{\diamond}(\xi_0) = -\frac{1}{N_1}\mathrm{E}\big[\frac{\partial}{\partial\xi'}S_N^{\diamond}(\xi_0)\big]$ and $\Gamma_N^{\diamond}(\xi_0) = \frac{1}{N_1}\mathrm{Var}\big[S_N^{\diamond}(\xi_0)\big]$, both assumed to exist and $\Sigma_N^{\diamond}(\xi_0)$ assumed to be positive definite for sufficiently large N.

3.3. Heteroskedasticity Robust Inferences

Robust inferences for ξ_0 depends on the availability of consistent estimators of $\Sigma_N^\diamond(\xi_0)$ and $\Gamma_N^\diamond(\xi_0)$. Similar to the case of homoskedastic model, $\Sigma_N^\diamond(\xi_0)$ can be estimated by its observed counterpart $\widehat{\Sigma}_N^\diamond = -\frac{1}{N_1}\frac{\partial}{\partial\xi'}S_N^\diamond(\xi)|_{\xi=\hat{\xi}_N^\diamond}$, with detailed expression of $\frac{\partial}{\partial\xi'}S_N^\diamond(\xi)$ being given in (C.1), Appendix C. The consistency of $\widehat{\Sigma}_N^\diamond$ is proved in the proof of Theorem 3.2.

However, the VC matrix $\Gamma_N^\diamond(\xi_0)$ involves the common parameters ξ_0, the fixed effects ϕ_0, and the unknown $\mathbf{H}$, as seen from its distinct elements derived by Lemma A.5:

$$\begin{aligned}
N_1\Gamma_{\beta\xi}^\diamond &= [\mathbb{X}'\mathbf{H}\mathbb{X},\ \ \mathbb{X}'\mathbf{H}\mathbb{L}_\lambda\mathbf{B}_N\eta,\ \ \mathbb{X}'\mathbf{H}\mathbb{L}_\rho\mathbb{D}_N\phi_0],\\
N_1\Gamma_{\lambda\lambda}^\diamond &= \eta'\mathbf{B}_N'\mathbb{L}_\lambda'\mathbf{H}\mathbb{L}_\lambda\mathbf{B}_N\eta + \mathtt{tr}(\mathbf{H}\mathbb{L}_\lambda\mathbf{H}\mathbb{L}_\lambda^\circ),\\
N_1\Gamma_{\lambda\rho}^\diamond &= \eta'\mathbf{B}_N'\mathbb{L}_\lambda'\mathbf{H}\mathbb{L}_\rho\mathbb{D}_N\phi_0 + \mathtt{tr}(\mathbf{H}\mathbb{L}_\lambda\mathbf{H}\mathbb{L}_\rho^\circ),\\
N_1\Gamma_{\rho\rho}^\diamond &= \phi_0'\mathbb{D}_N'\mathbb{L}_\rho'\mathbf{H}\mathbb{L}_\rho\mathbb{D}_N\phi_0 + \mathtt{tr}(\mathbf{H}\mathbb{L}_\rho\mathbf{H}\mathbb{L}_\rho^\circ),
\end{aligned} \tag{3.7}$$

where $\mathbb{L}_\lambda(\delta) = \mathbb{Q}_\mathbb{D}(\rho)[\bar{\mathbf{F}}_N(\delta) - \bar{\mathbb{F}}_N(\delta)]$ and $\mathbb{L}_\rho(\rho) = \mathbb{Q}_\mathbb{D}(\rho)[\bar{\mathbf{G}}_N'(\rho) - \bar{\mathbb{G}}_N'(\rho)]$. This makes the estimation of $\Gamma_N^\diamond(\xi_0)$ more challenging than the case of homoskedastic model as the number of unknown elements (parameters) in ϕ and $\mathbf{H}$ both grow with the sample size N (a more serious incidental parameters problem). A nice feature of the analytical expression of $\Gamma_N^\diamond(\xi_0)$ is that it does not involve 3rd and 4th moments of the errors due to the fact that the key matrices, $\mathbb{L}_\lambda(\delta)$ and $\mathbb{L}_\rho(\delta)$, have zero diagonals. This makes it possible to adopt again the approach of 'plug-in' and 'bias-correction' as in the case of homoskedastic model.

To facilitate the discussion, write $\Gamma_N^\diamond(\xi_0)$ as $\Gamma_N^\diamond(\xi_0, \phi, \mathbf{H})$. Let $\hat{\phi}_N^\diamond$ be the estimator of ϕ obtained from the RAQS estimator $\hat{\xi}_N^\diamond$ through (2.4). Let $\Gamma_N^\diamond(\hat{\xi}_N^\diamond, \hat{\phi}_N^\diamond, \mathbf{H})$ be the plug-in estimator of $\Gamma_N^\diamond(\xi_0)$ for a given $\mathbf{H}$. We have (similar to Corollary 2.1) the following corollary.

Corollary 3.1. *Under the assumptions of Theorem 3.2, we have,*

$$\Gamma_N^\diamond(\hat{\xi}_N^\diamond, \hat{\phi}_N^\diamond, \mathbf{H}) = \Gamma_N^\diamond(\xi_0) + \mathrm{Bias}_\phi^\diamond(\delta_0, \mathbf{H}) + o_p(1),$$

where $\mathrm{Bias}_\phi^\diamond(\delta_0, \mathbf{H})$ *is a* $(k+2)\times(k+2)$ *matrix with all the* β*-related entries being zero and the* δ *entry of the elements:* $\frac{1}{N_1}\mathtt{tr}(\mathbf{H}\mathbb{P}_\mathbb{D}\mathbb{L}_a'\mathbf{H}\mathbb{L}_b\mathbb{P}_\mathbb{D})$*, for* $a, b = \lambda, \rho$*.*

To estimate $\mathbf{H}$ and thus to give a full estimate of $\Gamma_N^\diamond(\xi_0, \phi, \mathbf{H})$, note that $\tilde{\mathbf{V}} = \mathbb{Q}_\mathbb{D}\mathbf{V}$, which can be 'consistently' estimated by $\hat{\mathbf{V}} = \mathbb{Q}_\mathbb{D}(\hat{\rho}_N^\diamond)\mathbf{B}_N(\hat{\rho}_N^\diamond)[\mathbf{A}_N(\hat{\lambda}_N^\diamond)\mathbf{Y} - \mathbf{X}\hat{\beta}_N^\diamond]$. Note also that

$$\mathrm{E}(\tilde{\mathbf{V}} \odot \tilde{\mathbf{V}}) = [\mathbb{Q}_\mathbb{D} \odot \mathbb{Q}_\mathbb{D}](\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2)',$$

where $\odot$ denotes the Hadamard (elementwise) product. A natural set of estimates of the heteroskedasticity parameters $(\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2)$ is therefore given as follows:

$$(\hat{\sigma}_1^2, \hat{\sigma}_2^2, \ldots, \hat{\sigma}_N^2)' = [\mathbb{Q}_{\mathbb{D}}(\hat{\rho}_N^{\diamond}) \odot \mathbb{Q}_{\mathbb{D}}(\hat{\rho}_N^{\diamond})]^{-}(\hat{\mathbf{V}} \odot \hat{\mathbf{V}}),$$

where $[\,\cdot\,]^{-}$ denotes a generalized inverse. An estimate of $\mathbf{H}$ is $\widehat{\mathbf{H}} = \mathtt{diag}(\hat{\sigma}_1^2, \hat{\sigma}_2^2, \ldots, \hat{\sigma}_N^2)$.

To 'see' the invertibility of $\mathbb{Q}_{\mathbb{D}}(\rho) \odot \mathbb{Q}_{\mathbb{D}}(\rho)$, we have, $\mathbb{Q}_{\mathbb{D}}(\rho) \odot \mathbb{Q}_{\mathbb{D}}(\rho) = I_N - 2I_N \odot \mathbb{P}_{\mathbb{D}}(\rho) + \mathbb{P}_{\mathbb{D}}(\rho) \odot \mathbb{P}_{\mathbb{D}}(\rho)$. By Schur product theorem, the last term is positive semi-definite. In addition, when T is not too small, $I_N - 2I_N \odot \mathbb{P}_{\mathbb{D}}(\rho)$ is positive definite, because the diagonal elements of $\mathbb{P}_{\mathbb{D}}(\rho)$ are of order $O_p(1/T)$ (See the proof of Lemma A.3). Thus, $\mathbb{Q}_{\mathbb{D}}(\rho) \odot \mathbb{Q}_{\mathbb{D}}(\rho)$ is typically invertible, for ρ in a neighborhood of ρ_0, which is assumed to facilitate the proof of theoretical results. In practice, however, one may just use the generalized inverse of $\mathbb{Q}_{\mathbb{D}}(\rho) \odot \mathbb{Q}_{\mathbb{D}}(\rho)$.

From (3.7), we see that the elements of $\Gamma(\xi_0, \phi, \mathbf{H})$ take either of the forms: $\mathtt{tr}(\mathbf{H}C_N)$ and $\mathtt{tr}(\mathbf{H}A_N\mathbf{H}B_N)$. It is important to know the effects of replacing $\mathbf{H}$ by $\hat{\mathbf{H}}$ in these two forms.

Lemma 3.1. *Assume $\Pi_N(\rho) = [\mathbb{Q}_{\mathbb{D}}(\rho) \odot \mathbb{Q}_{\mathbb{D}}(\rho)]^{-1}$ exists for ρ in a neighborhood of ρ_0, and is bounded in both row and column sum norms.*[8] *Let $A_N = [a_{ij}]$ and $B_N = [b_{ij}]$ be square matrices of dimension N with zero diagonals and bounded row and column sum norms. Let $C_N = [c_{ij}]$ be an $N \times N$ matrix with diagonal elements being uniformly bounded. We have,*

(i) $\frac{1}{N}\mathtt{tr}(\widehat{\mathbf{H}}C_N) - \frac{1}{N}\mathtt{tr}(\mathbf{H}C_N) = o_p(1)$,

(ii) $\frac{1}{N}\mathtt{tr}(\widehat{\mathbf{H}}A_N\widehat{\mathbf{H}}B_N) - \frac{2}{N}\mathtt{tr}((A_N \odot B_N)\Pi_N\Lambda(\mathbf{H})\Pi_N) - \frac{1}{N}\mathtt{tr}(\mathbf{H}A_N\mathbf{H}B_N) = o_p(1)$,

where $\Pi_N = \Pi_N(\rho_0)$, $\Lambda(\mathbf{H}) = \{(q_j'\mathbf{H}q_k)^2\}_{j,k=1}^N$, and q_j' is the jth row of $\mathbb{Q}_{\mathbb{D}}$.

The bias term in Corollary 3.1 needs a further correction when $\mathbf{H}$ is replaced by $\hat{\mathbf{H}}$ as it contains elements of the form $\mathtt{tr}(\mathbf{H}A_N\mathbf{H}B_N)$ with diagonal elements of A_N and B_N not strictly zero. However, the effect of non-zero diagonals is shown to be negligible due to the existence of a lower ranked matrix $\mathbb{P}_{\mathbb{D}}$ and its orthogonality with $\mathbb{Q}_{\mathbb{D}}$. Combining the results of Corollary 3.1 and Lemma 3.1, we have the full estimate of $\Gamma_N^{\diamond}(\xi_0)$:

$$\widehat{\Gamma}_N^{\diamond} = \Gamma_N^{\diamond}(\hat{\xi}_N^{\diamond}, \hat{\phi}_N^{\diamond}, \hat{\mathbf{H}}) - \text{Bias}_{\phi}^{\diamond}(\hat{\delta}_N^{\diamond}, \widehat{\mathbf{H}}) - \text{Bias}_{\mathbf{H}}^{\diamond}(\hat{\delta}_N^{\diamond}, \widehat{\mathbf{H}}), \tag{3.8}$$

where $\text{Bias}_{\mathbf{H}}^{\diamond}(\delta_0, \mathbf{H})$ has entries 0, or $\frac{2}{N_1}\mathtt{tr}((\mathbb{L}_a \odot \mathbb{L}_b^{\circ} - \mathbb{P}_{\mathbb{D}}\mathbb{L}_a' \odot \mathbb{L}_b\mathbb{P}_{\mathbb{D}})\Pi_N\Lambda(\mathbf{H})\Pi_N), a, b = \lambda, \rho$.

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

$$\widehat{\Sigma}_N^{\diamond} - \Sigma_N^{\diamond}(\xi_0) \xrightarrow{p} 0 \quad \textit{and} \quad \widehat{\Gamma}_N^{\diamond} - \Gamma_N^{\diamond}(\xi_0) \xrightarrow{p} 0,$$

and therefore, $\widehat{\Sigma}_N^{\diamond-1}\widehat{\Gamma}_N^{\diamond}\widehat{\Sigma}_N^{\diamond-1} - \Sigma_N^{\diamond-1}(\xi_0)\Gamma_N^{\diamond}(\xi_0)\Sigma_N^{\diamond-1}(\xi_0) \xrightarrow{p} 0.$

[8] These assumptions hold for a balanced SPD model. Following Footnote 7, $\mathbb{Q}_{\mathbb{D}}(\rho) = (I_T - \frac{l_T l_T'}{T}) \otimes (I_n - \frac{l_n l_n'}{n})$, where $\otimes$ denotes the Kronecker product. Thus, $[\mathbb{Q}_{\mathbb{D}}(\rho) \odot \mathbb{Q}_{\mathbb{D}}(\rho)]^{-1}$ exists if $T > 2$ by Schur product theorem. Further, $|(\mathbb{Q}_{\mathbb{D}}(\rho) \odot \mathbb{Q}_{\mathbb{D}}(\rho))_{ii}| - \sum_{j \neq i} |(\mathbb{Q}_{\mathbb{D}}(\rho) \odot \mathbb{Q}_{\mathbb{D}}(\rho))_{ij}| = \frac{(n-1)(T-1)[(n-2)(T-2)-2]}{n^2T^2} > c > 0, \forall i, T > 2$. As $\mathbb{Q}_{\mathbb{D}}(\rho) \odot \mathbb{Q}_{\mathbb{D}}(\rho)$ is symmetric, we conclude it is strictly diagonally dominant in both rows and columns. Hence, Theorem 1 and Corollary 1 of Varah (1975) imply that $\|\Pi_N(\rho)\|_1$ and $\|\Pi_N(\rho)\|_\infty$ are both bounded.

4. Monte Carlo Study

Extensive Monte Carlo experiments are carried out to investigate the finite sample performance of the proposed AQS estimators, the robust AQS (RAQS) estimators, and the corresponding standard error estimators of the unbalanced SPD models with two-way fixed effects. To see the effectiveness of the adjustments on the concentrated quasi scores in controlling the effects of estimating the fixed effects, we also include the direct QML estimators in the Monte Carlo study. We choose different values of n and T, and fix the percentage of randomly missing observations at 10%, and make sure that each individual is observed at least twice over the entire period. We consider two data generating processes: unbalanced FE-SPD models with `SL` and `SE` effects or with `SL` and `SD` (spatial Durbin) effects:

$$\texttt{SL-SE Model}: \ Y_t = \lambda W_t Y_t + X_t\beta_1 + D_t\mu + \alpha_t l_{n_t} + U_t, \quad U_t = \rho M_t U_t + V_t, \tag{4.1}$$

$$\texttt{SL-SD Model}: \ Y_t = \lambda W_t Y_t + X_t\beta_1 + W_t X_t \beta_2 + D_t\mu + \alpha_t l_{n_t} + V_t, \tag{4.2}$$

for $t = 1, \ldots, T$. Note that we consider Durbin effects, $W_t X_t$, only in the `SL` model due to the identification issue mentioned earlier. We choose $\beta_1 = 1$, $\beta_2 = 0$ or 0.5, $\lambda = 0.2$ and $\rho = 0.2$. Generate $X_t's$ independently from $N(0, 2^2 I_n)$, and set the individual effects $\mu = \frac{1}{T}\Sigma_{t=1}^T X_t + e$, where $e \sim N(0, I_n)$. Then, omit the "missing" elements of X_t. The time fixed effects α are generated from $N(0, I_T)$. The error (v_{it}) distributions can be (i) normal, (ii) normal mixture (10% $N(0, 4^2)$ and 90% $N(0,1)$), or (iii) chi-square with 3 degrees of freedom.[9] For the purpose of comparison, we set $\sigma_{v0}^2 = 1$ for homoskedastic case, and set the average of error variances in the heteroskedastic case to 1. Monte Carlo (empirical) means and standard deviations (shown in the brackets) are reported for QMLE, AQSE and RAQSE. Further, empirical averages of the standard error estimates (shown in the square brackets) are also reported for AQSE and RAQSE, based on the robust VC matrix estimates, $\widehat{\Sigma}_N^{*-1}\widehat{\Gamma}_N^{*}\widehat{\Sigma}_N^{*-1}$ for the AQSE and $\widehat{\Sigma}_N^{\diamond-1}\widehat{\Gamma}_N^{\diamond}\widehat{\Sigma}_N^{\diamond-1}$ for the RAQSE. The number of Monte Carlo runs is 1000.

The spatial weights W_t and M_t are first generated as time-varying $n \times n$ matrices according to `rook` contiguity, `queen` contiguity, or `group` interaction scheme, and then their rows and columns corresponding to the missing spatial units are deleted. The groups' sizes in the `group` interaction scheme can be either increasing or fixed as n increases. In the latter case, the variation in group sizes does not shrink to zero as n increases. As a result, the AQSE

[9] In the cases (ii) and (iii), the generated errors are standardized to have mean zero and variance σ_v^2.

would not be consistent under heteroskedasticity (Liu and Yang, 2015, 2020).[10] In this case, the heteroskedasticity is generated as follows: for each group, if the group size is larger than the mean group size, then the variance is set to be the same as the group size, otherwise, the variance is the square of the inverse of the group size (Lin and Lee, 2010).

Tables 1a and 1b report partial Monte Carlo results for the unbalanced FE-SPD model with `SL` and `SE` effects and homoskedastic errors, for $T = 5$ and 10, respectively. The results show an excellent finite performance of the proposed AQS and RAQS estimators, as well as their standard error estimators. The proposed AQS method performs uniformly much better than the QML method in the estimation of σ_v^2, λ and ρ, irrespective of the choices of the spatial weight matrices and the values of n and T. Our AQS estimators exhibit a good performance even when the sample size is as small as $n = 50$ and $T = 5$, and improve on average when the sample expands, regardless of the error distributions. The $\sqrt{N_1}$-consistency of the AQSEs is clearly demonstrated by the Monte Carlo sds. Moreover, the robust estimates of standard errors $\hat{sd}$'s are on average very close to the corresponding Monte Carlo standard errors. By comparing the results of AQS and RAQS, we cannot see which one beats the other in terms of bias and efficiency for these homoskedastic models.

Tables 2a and 2b present partial Monte Carlo results for the unbalanced FE-SPD model with `SL` and `SD` effects and homoskedastic errors, for $T = 5$ and 10, respectively. The results again show an excellent performance of the proposed set of estimation and inference methods. As in the case of the `SL-SE` model, the AQSE and RAQSE give quite similar results, and both show a clear convergence as sample size increases. Their corresponding standard error estimates also perform very well. In contrast, the QMLE can perform poorly.

Tables 3a and 3b report partial Monte Carlo results for the unbalanced FE-SPD model with `SL` and `SE` effects and heteroskedastic errors, for $T = 5$ and 10, respectively. The results show an excellent finite sample performance of the proposed RAQSE and its estimated standard error. In contrast, the QMLE and AQSE typically provide worse estimates for spatial parameters than RAQSE. Our RAQSEs perform well even when sample size is quite small, and show convergence to the true value as sample size increases. In addition, $\hat{sd}$s are very closed to sds for our RAQSE, consistent with our theoretical expectation.

Tables 4 presents partial Monte Carlo results for the unbalanced FE-SPD model with `SL`

[10]In the former, we let $K(n) = Round(n^{0.5})$ be the number of groups and then generate $K(n)$ group sizes according to a uniform distribution, and in the latter, we start with six groups of sizes (3,5,7,9,11,15) and then replicate to give a n to be multiples of 50. See Yang (2015) for details in generating these spatial layouts.

and `SD` effects and heteroskedastic errors, for $T = 5$ and 10, respectively. The weight matrix is specified as group interaction with a fixed group sizes scheme. We can see a much better finite sample performance for our RAQSE than QMLE and AQSE, and the corresponding standard error estimates also have a good performance.

5. Conclusion and Discussion

We consider estimation and inference for an unbalanced spatial panel data model with both individual and time fixed effects, where the unbalancedness is caused by, e.g., late entries, early dropouts, lack of economic activities, such that the missing spatial units at a given time period do not generate any spillover effects on their 'neighbors". Unbalanced spatial panels with fixed effects render the commonly adopted approach, the orthogonal transformation, inapplicable. An *adjusted quasi score* (AQS) is proposed, which adjusts the concentrated quasi scores (with the fixed effects being concentrated out) to remove the effects of estimating these *incidental parameters*. For the statistical inferences, the main difficulty lies with the fact that 'consistent' estimates of the idiosyncratic errors are unavailable due to the incidental parameters problem. A 'plug-in and then bias-correction' method is proposed to give consistent estimates of the standard errors of the AQS estimators. The proposed methods are then extended to allow for unknown heteroskedasticity along both the cross-sectional and time dimensions. Monte Carlo results show excellent performance of the proposed estimation and inference methods.

The proposed methods are seen to be very general in handling the unbalanced SPD models in the presence of incidental parameters such as fixed effects and unknown heteroskedasticity, allowing the spatial weight matrices to be time-varying and without row-normalizations. The generality of the proposed methods is further demonstrated in Appendix D by considering the following extensions: the unbalanced SPD model with (*i*) two-way fixed effects (FE) and serial correlation, (*ii*) two-way FE, heteroskedasticity and serial correlation, (*iii*) two-way random effects (RE) and serial correlation, and (*iv*) two-way RE, heteroskedasticity and serial correlation. The current study also sheds light on an interesting but challenging extension: unbalanced SPD models with interactive fixed effects in the spirit of Bai et al. (2015). However, rigorous studies on these extensions can only be done in future works.

Appendix A: Some Basic Lemmas

The following lemmas are essential to the proofs of the main results in this paper. Lemmas A.1, A.2 and A.5 are taken from the literature. The other lemmas contain new elements and their proofs are given in Appendix D.

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

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

(ii) the elements of A_N are uniformly bounded and tr$(A_N) = O(N)$, and

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

Lemma A.2. *(Lemma A.3, Lee, 2004b): For $\mathbf{W}$ and $\mathbf{A}_N(\lambda)$ defined in Model* (2.2)*, if $\|\mathbf{W}\|$ and $\|\mathbf{A}_N^{-1}\|$ are uniformly bounded, where $\|\cdot\|$ is a matrix norm, then $\|\mathbf{A}_N^{-1}(\lambda)\|$ is uniformly bounded in a neighborhood of λ_0.*

Lemma A.3. *Under Assumptions C-E, we have*

(i) $\mathbb{Q}_{\mathbb{D}}(\rho)$ is uniformly bounded in both row and column sums, uniformly in $\rho \in \Delta_\rho$;

(ii) $\mathbb{Q}_{\mathbb{X}}(\rho)$ is uniformly bounded in both row and column sums, uniformly in $\rho \in \Delta_\rho$.

Lemma A.4. *Suppose that $\{A_N\}$ and $\{B_N\}$ are two sequences of $N \times N$ matrices that are uniformly bounded in either row or column sums. Under Assumptions C-E, $\mathtt{tr}[A_N \mathbb{P}_{\mathbb{X}}(\rho) B_N] = O(1)$, uniformly in $\rho \in \Delta_\rho$.*

Lemma A.5. *(Lemma A.2, Lin and Lee, 2010; Lemma A.3, Liu and Yang, 2015): Let $A_N = [a_{ij}]$ and $B_N = [b_{ij}]$ be two square matrices of dimension N and c_N be an $N \times 1$ vector of elements c_i. Assume that innovations $\{v_j\}$ have zero mean and are mutually independent, i.e. $v_j \sim inid(0, \sigma_j^2)$. Letting $\mathbf{H} = \mathtt{diag}\{\sigma_1^2, \cdots, \sigma_N^2\}$ and $\mathbf{V} = (v_1, \cdots, v_N)'$, we have,*

(i) $\mathrm{E}(\mathbf{V}' A_N \mathbf{V}) = \mathtt{tr}(\mathbf{H} A_N) = \sum_{i=1}^N a_{ii} \sigma_i^2$,

(ii) $\mathrm{E}(\mathbf{V}' A_N \mathbf{V} \cdot c_N' \mathbf{V}) = \sum_{i=1}^N a_{ii} c_i \mathrm{E}(v_i^3)$,

(iii) $\mathrm{E}(\mathbf{V}' A_N \mathbf{V} \cdot \mathbf{V}' B_N \mathbf{V}) = \sum_{i=1}^N a_{ii} b_{ii} [\mathrm{E}(v_i^4) - 3\sigma_i^4] + \mathtt{tr}(\mathbf{H} A_N) \mathtt{tr}(\mathbf{H} B_N) + \mathtt{tr}(\mathbf{H} A_N \mathbf{H} B_N^\circ)$,

(iv) $\mathrm{Var}(\mathbf{V}' A_N \mathbf{V}) = \sum_{i=1}^N a_{ii}^2 [\mathrm{E}(v_i^4) - 3\sigma_i^4] + \mathtt{tr}(\mathbf{H} A_N \mathbf{H} A_N^\circ)$.

Lemma A.6. *(Lemma A.3, Lin and Lee, 2010, extended): Let $\{A_N\}$ be a sequence of $N \times N$ matrices such that either $\|A_N\|_\infty$ or $\|A_N\|_1$ is bounded. Suppose that the elements of*

A_N are $O(h_n^{-1})$ uniformly in all i and j. Let innovation vector $\mathbf{V}$ be defined as in Lemma A.5. Let c_N be an $N \times 1$ vector with elements of uniform order $O(h_n^{-1/2})$. Then

(i) $\mathrm{E}(\mathbf{V}' A_N \mathbf{V}) = O(\frac{N}{h_n})$, (ii) $\mathrm{Var}(\mathbf{V}' A_N \mathbf{V}) = O(\frac{N}{h_n})$,

(iii) $\mathbf{V}' A_N \mathbf{V} = O_p(\frac{N}{h_n})$, (iv) $\mathbf{V}' A_N \mathbf{V} - \mathrm{E}(\mathbf{V}' A_N \mathbf{V}) = O_p((\frac{N}{h_n})^{\frac{1}{2}})$,

(v) $c_N' A_N \mathbf{V} = O_p((\frac{N}{h_n})^{\frac{1}{2}})$, **if** $\|A_N\|_1$ *is bounded.*

Appendix B: Proofs for Section 2

In proving the theorems, the following facts are used: (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) $\gamma_{min}(A)\mathtt{tr}(B) \leq \mathtt{tr}(AB) \leq \gamma_{max}(A)\mathtt{tr}(B)$ for symmetric matrix A and positive semi-definite (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.

Derivation of the AQS functions and the Hessian matrix:

Writing the key quantity in the concentrated quasi loglikelihood (2.5) as $\tilde{\mathbf{V}}'(\beta, \delta)\tilde{\mathbf{V}}(\beta, \delta) = [\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\beta]'\mathbf{B}_N'(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho)[\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\beta]$, and using the facts that for an invertible matrix $A(\lambda)$, $\frac{\partial}{\partial\lambda}\ln|A(\lambda)| = \mathtt{tr}[A^{-1}(\lambda)\frac{\partial}{\partial\lambda}A(\lambda)]$ and $\frac{\partial}{\partial\lambda}A^{-1}(\lambda) = -A^{-1}(\lambda)[\frac{\partial}{\partial\lambda}A(\lambda)]A^{-1}(\lambda)$, it is straightforward to derive $S_N^c(\theta)$. However, the derivation of the ρ-component is complicated and some intermediate results are useful. First, $\frac{\partial}{\partial\rho}[\mathbf{B}_N'(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho)] = -\mathbf{M}'\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho) - \mathbf{B}_N'(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{M} + \mathbf{B}_N'(\rho)\dot{\mathbb{Q}}_{\mathbb{D}}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho) + \mathbf{B}_N'(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\dot{\mathbb{Q}}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho)$, where $\dot{\mathbb{Q}}_{\mathbb{D}}(\rho) = \frac{\partial}{\partial\rho}\mathbb{Q}_{\mathbb{D}}(\rho)$. With $\frac{\partial}{\partial\rho}\mathbb{D}_N(\rho) = -\mathbf{M}[\mathbf{D}, \mathbf{D}_\alpha^\star] = -\mathbf{G}_N(\rho)\mathbb{D}_N(\rho)$, we have

$$\dot{\mathbb{Q}}_{\mathbb{D}}(\rho) = \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{G}_N(\rho)\mathbb{P}_{\mathbb{D}}(\rho) + \mathbb{P}_{\mathbb{D}}(\rho)\mathbf{G}_N'(\rho)\mathbb{Q}_{\mathbb{D}}(\rho). \tag{B.1}$$

This leads to $-\frac{\partial}{\partial\rho}[\mathbf{B}_N'(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho)] = \mathbf{B}_N'(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{G}_N^\circ(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho) \equiv \Psi(\rho)$, the ρ-component of the CQS function (2.6), and the ρ-component of the AQS function (2.8):

$$S_\rho^*(\theta) = \tfrac{1}{2\sigma_v^2}[\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\beta]'\Psi(\rho)[\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\beta] - \mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{G}_N(\rho)]. \tag{B.2}$$

This is expressed in terms of $\Psi(\rho)$ and $\mathbf{G}_N^\circ(\rho)$ to facilitate the derivations of the ρ-related terms of the Hessian matrix $\frac{\partial}{\partial\rho}\Psi(\rho)$. Again, the (ρ, ρ) term of $\frac{\partial}{\partial\rho}\Psi(\rho)$ is most complicate. For a comformable vector a, we have by taking use of (B.1) and after some tedious algebra,

$$a'[\tfrac{\partial}{\partial\rho}\Psi(\rho)]a = 2a'\mathbf{B}_N'(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)[\mathbf{G}_N^\circ(\rho)\mathbb{P}_{\mathbb{D}}(\rho)\mathbf{G}_N^\circ(\rho) - \mathbf{G}_N'(\rho)\mathbf{G}_N(\rho)]\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho)a. \tag{B.3}$$

With the set of AQS functions $S_N^*(\theta)$ given in (2.8) and (B.1)-(B.3), we obtain the components of the Hessian matrix $H_N^*(\theta) = \frac{\partial}{\partial\theta'} S_N^*(\theta)$:

$$\begin{aligned}
H^*_{\beta\beta}(\theta) &= -\tfrac{1}{\sigma_v^2}\mathbb{X}'(\rho)\mathbb{X}(\rho),\\
H^*_{\beta\sigma_v^2}(\theta) &= -\tfrac{1}{\sigma_v^4}\mathbb{X}'(\rho)\tilde{\mathbf{V}}(\beta,\delta) \;=\; H^{*\prime}_{\sigma_v^2\beta},\\
H^*_{\beta\lambda}(\theta) &= -\tfrac{1}{\sigma_v^2}\mathbb{X}'(\rho)\mathbb{Y}(\rho) \;=\; H^{*\prime}_{\lambda\beta},\\
H^*_{\beta\rho}(\theta) &= -\tfrac{1}{\sigma_v^2}\mathbb{X}'(\rho)\mathbf{G}^\circ_N(\rho)\tilde{\mathbf{V}}(\beta,\delta) \;=\; H^{*\prime}_{\rho\beta},\\
H^*_{\sigma_v^2\sigma_v^2}(\theta) &= -\tfrac{1}{\sigma_v^6}\tilde{\mathbf{V}}'(\beta,\delta)\tilde{\mathbf{V}}(\beta,\delta) + \tfrac{1}{2\sigma_v^4}N_1,\\
H^*_{\sigma_v^2\lambda}(\theta) &= -\tfrac{1}{\sigma_v^4}\mathbb{Y}'(\rho)\tilde{\mathbf{V}}(\beta,\delta) \;=\; H^{*\prime}_{\lambda\sigma_v^2},\\
H^*_{\sigma_v^2\rho}(\theta) &= -\tfrac{1}{2\sigma_v^4}\tilde{\mathbf{V}}'(\beta,\delta)\mathbf{G}^\circ_N(\rho)\tilde{\mathbf{V}}(\beta,\delta) \;=\; H^{*\prime}_{\rho\sigma_v^2},\\
H^*_{\lambda\lambda}(\theta) &= -\tfrac{1}{\sigma_v^2}\mathbb{Y}'(\rho)\mathbb{Y}(\rho) - \mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho)\mathbf{F}_N^2(\lambda)\mathbf{B}_N^{-1}(\rho)],\\
H^*_{\lambda\rho}(\theta) &= -\tfrac{1}{\sigma_v^2}\mathbb{Y}'(\rho)\mathbf{G}^\circ_N(\rho)\tilde{\mathbf{V}}(\beta,\delta) - \mathtt{tr}[\mathbf{F}_N(\lambda)\mathbb{R}_N(\rho)],\\
H^*_{\rho\lambda}(\theta) &= -\tfrac{1}{\sigma_v^2}\mathbb{Y}'(\rho)\mathbf{G}^\circ_N(\rho)\tilde{\mathbf{V}}(\beta,\delta),\\
H^*_{\rho\rho}(\theta) &= \tfrac{1}{\sigma_v^2}\tilde{\mathbf{V}}'(\beta,\delta)\mathcal{R}_{1N}(\rho)\tilde{\mathbf{V}}(\beta,\delta) - \mathtt{tr}[\mathcal{R}_{2N}(\rho)],
\end{aligned} \tag{B.4}$$

where $\mathbb{Y}(\rho) = \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho)\mathbf{W}\mathbf{Y}$, $\mathbb{R}_N(\rho) = \mathbf{B}_N^{-1}(\rho)\mathbb{P}_{\mathbb{D}}(\rho)\mathbf{G}^\circ_N(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho)$, $\mathcal{R}_{1N}(\rho) = \mathbf{G}^\circ_N(\rho)\mathbb{P}_{\mathbb{D}}(\rho)\mathbf{G}^\circ_N(\rho) - \mathbf{G}'_N(\rho)\mathbf{G}_N(\rho)$ and $\mathcal{R}_{2N}(\rho) = \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{G}_N(\rho)[\mathbb{P}_{\mathbb{D}}(\rho)\mathbf{G}^\circ_N(\rho) + \mathbf{G}_N(\rho)]$.

Proof of Theorem 2.1: By theorem 5.9 of van der Vaart (1988), we only need to show $\sup_{\delta\in\delta}\frac{1}{N_1}\left\|S_N^{*c}(\delta) - \bar{S}_N^{*c}(\delta)\right\| \xrightarrow{p} 0$ under the assumptions in Theorem 2.1. From (2.10) and (2.12), the consistency of $\hat{\delta}_N^*$ follows from:

- (a) $\inf_{\delta\in\Delta}\bar{\sigma}^{*2}_{v,N}(\delta)$ is bounded away from zero,
- (b) $\sup_{\delta\in\Delta}\left|\hat{\sigma}^{*2}_{v,N}(\delta) - \bar{\sigma}^{*2}_{v,N}(\delta)\right| = o_p(1)$,
- (c) $\sup_{\delta\in\Delta}\frac{1}{N_1}\left|\mathbf{Y}'\mathbf{W}'\mathbf{B}'_N(\rho)\hat{\mathbf{V}}(\delta) - \mathrm{E}[\mathbf{Y}'\mathbf{W}'\mathbf{B}'_N(\rho)\bar{\mathbf{V}}(\delta)]\right| = o_p(1)$,
- (d) $\sup_{\delta\in\Delta}\frac{1}{N_1}\left|\hat{\mathbf{V}}'(\delta)\mathbf{G}_N(\rho)\hat{\mathbf{V}}(\delta) - \mathrm{E}[\bar{\mathbf{V}}'(\delta)\mathbf{G}_N(\rho)\bar{\mathbf{V}}(\delta)]\right| = o_p(1)$.

Proof of (a). From (2.11), we can write $\bar{\beta}_N^*(\delta) = [\mathbb{X}'(\rho)\mathbb{X}(\rho)]^{-1}\mathbb{X}'(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{C}_N(\delta)\mathrm{E}(\mathbf{Y})$ as $\mathbb{X}(\rho) = \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho)\mathbf{X}$ and $\mathbb{Q}_{\mathbb{D}}(\rho)$ is idempotent. Thus, $\bar{\mathbf{V}}(\delta) = \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{C}_N(\delta)\mathbf{Y} - \mathbb{X}(\rho)\bar{\beta}_N^*(\delta) = \mathbb{Q}_{\mathbb{X}}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{C}_N(\delta)\mathbf{Y} + \mathbb{P}_{\mathbb{X}}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{C}_N(\delta)[\mathbf{Y} - \mathrm{E}(\mathbf{Y})]$. By the orthogonality between $\mathbb{Q}_{\mathbb{D}}(\rho)$ and $\mathbb{P}_{\mathbb{D}}(\rho)$ and using $\mathbf{Y} = \mathbf{A}_N^{-1}(\eta + \mathbf{B}_N^{-1}\mathbf{V})$, we have,

$$\begin{aligned}
\bar{\sigma}^{*2}_{v,N}(\delta) &= \tfrac{1}{N_1}\mathrm{E}[\bar{\mathbf{V}}'(\delta)\bar{\mathbf{V}}(\delta)]\\
&= \tfrac{1}{N_1}\mathrm{E}[\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y}] + \tfrac{1}{N_1}\mathrm{E}\big\{[\mathbf{Y} - \mathrm{E}(\mathbf{Y})]'\mathbf{P}(\delta)[\mathbf{Y} - \mathrm{E}(\mathbf{Y})]\big\} \qquad \text{(B.5)}\\
&= \tfrac{1}{N_1}\mathrm{E}(\mathbf{Y})'\mathbf{Q}(\delta)\mathrm{E}(\mathbf{Y}) + \tfrac{1}{N_1}\mathrm{E}\big\{[\mathbf{Y} - \mathrm{E}(\mathbf{Y})]'[\mathbf{Q}(\delta) + \mathbf{P}(\delta)][\mathbf{Y} - \mathrm{E}(\mathbf{Y})]\big\}\\
&= \tfrac{1}{N_1}\mathrm{E}(\mathbf{Y})'\mathbf{Q}(\delta)\mathrm{E}(\mathbf{Y}) + \tfrac{1}{N_1}\mathrm{E}\big\{[\mathbf{Y} - \mathrm{E}(\mathbf{Y})]'\mathbf{C}'_N(\delta)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{C}_N(\delta)[\mathbf{Y} - \mathrm{E}(\mathbf{Y})]\big\}\\
&= \tfrac{1}{N_1}\eta'\mathbf{A}_N'^{-1}\mathbf{Q}(\delta)\mathbf{A}_N^{-1}\eta + \tfrac{\sigma_{v0}^2}{N_1}\mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho)\mathcal{C}_N(\delta)],
\end{aligned}$$

where $\mathbf{Q}(\delta) = \mathbf{C}_N'(\delta)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbb{Q}_{\mathbb{X}}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{C}_N(\delta)$ and $\mathbf{P}(\delta) = \mathbf{C}_N'(\delta)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbb{P}_{\mathbb{X}}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{C}_N(\delta)$. The first term can be written in the form of $a'(\delta)a(\delta)$ for an $N \times 1$ vector function of δ, and thus is non-negative, uniformly in $\delta \in \Delta$. For the second term,

$$\begin{aligned}&\tfrac{\sigma_{v0}^2}{N_1}\mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho)\mathcal{C}_N(\delta)] \geq \tfrac{\sigma_{v0}^2}{N_1}\gamma_{\min}[\mathcal{C}_N(\delta)]\mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho)] = \sigma_{v0}^2\gamma_{\min}[\mathcal{C}_N(\delta)]\\ &\geq \sigma_{v0}^2\gamma_{\max}(\mathbf{A}_N'\mathbf{A}_N)^{-1}\gamma_{\max}(\mathbf{B}_N'\mathbf{B}_N)^{-1}\gamma_{\min}[\mathbf{A}_N'(\lambda)\mathbf{A}_N(\lambda)]\gamma_{\min}[\mathbf{B}_N'(\rho)\mathbf{B}_N(\rho)] > 0,\end{aligned}$$

uniformly in $\delta \in \Delta$, by Assumption E(iii). It follows that $\inf_{\delta\in\Delta}\bar{\sigma}_{v,N}^{*2}(\delta) > 0$.

Proof of (b). From (2.9), we can write $\hat{\beta}_N^*(\delta) = [\mathbb{X}'(\rho)\mathbb{X}(\rho)]^{-1}\mathbb{X}'(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{C}_N(\delta)\mathbf{Y}$. Then, $\hat{\mathbf{V}}(\delta) = \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho)[\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\hat{\beta}_N^*(\delta)] = \mathbb{Q}_{\mathbb{X}}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{C}_N(\delta)\mathbf{Y}$ and $\hat{\sigma}_{v,N}^{*2}(\delta) = \frac{1}{N_1}\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y}$. From (B.5), $\bar{\sigma}_{v,N}^{*2}(\delta) = \frac{1}{N_1}\mathrm{E}[\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y}] + \frac{\sigma_{v0}^2}{N_1}\mathtt{tr}[\mathbf{C}_N'^{-1}\mathbf{P}(\delta)\mathbf{C}_N^{-1}]$. Thus,

$$\hat{\sigma}_{v,N}^{*2}(\delta) - \bar{\sigma}_{v,N}^{*2}(\delta) = \tfrac{1}{N_1}[\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y} - \mathrm{E}(\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y})] - \tfrac{\sigma_{v0}^2}{N_1}\mathtt{tr}[\mathbf{C}_N'^{-1}\mathbf{P}(\delta)\mathbf{C}_N^{-1}].$$

For the second term, we have, $0 \leq \frac{1}{N_1}\mathtt{tr}[\mathbf{C}_N'^{-1}\mathbf{P}(\delta)\mathbf{C}_N^{-1}] \leq \frac{1}{N_1}\gamma_{\max}[\mathcal{C}_N(\delta)]\gamma_{\max}^2[\mathbb{Q}_{\mathbb{D}}(\rho)]\mathtt{tr}[\mathbb{P}_{\mathbb{X}}(\rho)]$ $= o(1)$, because $\mathtt{tr}[\mathbb{P}_{\mathbb{X}}(\rho)] = k$, $\gamma_{\max}[\mathbb{Q}_{\mathbb{D}}(\rho)] = 1$ and, by Assumption E(iii), $\gamma_{\max}[\mathcal{C}_N(\delta)] \leq \gamma_{\min}(\mathbf{A}_N'\mathbf{A}_N)^{-1}\gamma_{\min}(\mathbf{B}_N'\mathbf{B}_N)^{-1}\gamma_{\max}[\mathbf{A}_N'(\lambda)\mathbf{A}_N(\lambda)]\gamma_{\max}[\mathbf{B}_N'(\rho)\mathbf{B}_N(\rho)] < \infty$. Therefore, one has $\sup_{\delta\in\Delta}|\frac{\sigma_{v0}^2}{N_1}\mathtt{tr}[\mathbf{C}_N'^{-1}\mathbf{P}(\delta)\mathbf{C}_N^{-1}]| = o(1)$. For the first term, we prove the uniform convergence result: $\sup_{\delta\in\Delta}|\frac{1}{N_1}[\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y} - \mathrm{E}(\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y})]| = o_p(1)$, which follows from pointwise convergence of $\frac{1}{N_1}[\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y} - \mathrm{E}(\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y})]$ to zero for each $\delta \in \Delta$ and the stochastic equicontinuity of $\frac{1}{N_1}\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y}$, according to Andrews (1992). We have,

$$\begin{aligned}&\tfrac{1}{N_1}[\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y} - \mathrm{E}(\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y})]\\ &= \tfrac{1}{N_1}(\eta + \mathbf{B}_N^{-1}\mathbf{V})'\mathbf{A}_N'^{-1}\mathbf{Q}(\delta)\mathbf{A}_N^{-1}(\eta + \mathbf{B}_N^{-1}\mathbf{V}) - \tfrac{1}{N_1}\mathrm{E}[(\eta + \mathbf{B}_N^{-1}\mathbf{V})'\mathbf{A}_N'^{-1}\mathbf{Q}(\delta)\mathbf{A}_N^{-1}(\eta + \mathbf{B}_N^{-1}\mathbf{V})]\\ &= \tfrac{2}{N_1}\mathbf{V}'\mathbf{C}_N^{-1\prime}\mathbf{Q}(\delta)\mathbf{A}_N^{-1}\eta + \tfrac{1}{N_1}[\mathbf{V}'\mathbf{C}_N^{-1\prime}\mathbf{Q}(\delta)\mathbf{C}_N^{-1}\mathbf{V} - \sigma_{v0}^2\mathtt{tr}(\mathbf{C}_N^{-1\prime}\mathbf{Q}(\delta)\mathbf{C}_N^{-1})].\end{aligned}$$

By Assumption E, and Lemmas A.1 and A.3, one shows that $\mathbf{C}_N^{-1\prime}\mathbf{Q}(\delta)\mathbf{A}_N^{-1}$and $\mathbf{C}_N^{-1\prime}\mathbf{Q}(\delta)\mathbf{C}_N^{-1}$ are bounded in both row and column sum norms, for each $\delta \in \Delta$. Further, the elements of η are uniformly bounded. Thus, the pointwise convergence of the first term follows from Lemma A.6 (v), and the pointwise convergence of the second term follows from Lemma A.6 (iv). Therefore, $\frac{1}{N_1}[\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y} - \mathrm{E}(\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y})] \xrightarrow{p} 0$, for each $\delta \in \Delta$.

Next, let δ_1 and δ_2 be in Δ. We have by the mean value theorem (MVT):

$$\tfrac{1}{N_1}\mathbf{Y}'\mathbf{Q}(\delta_2)\mathbf{Y} - \tfrac{1}{N_1}\mathbf{Y}'\mathbf{Q}(\delta_1)\mathbf{Y} = \tfrac{1}{N_1}\mathbf{Y}'[\tfrac{\partial}{\partial\delta'}\mathbf{Q}(\bar{\delta})]\mathbf{Y}(\delta_2 - \delta_1),$$

where $\bar{\delta}$ lies between δ_1 and δ_2. It follows that $\frac{1}{N_1}\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y}$ is stochastically equicontinuous if $\sup_{\delta\in\Delta}\frac{1}{N_1}\mathbf{Y}'[\frac{\partial}{\partial\varpi}\mathbf{Q}(\delta)]\mathbf{Y} = O_p(1)$, $\varpi = \lambda, \rho$. We only show $\sup_{\delta\in\Delta}\frac{1}{N_1}\mathbf{Y}'[\frac{\partial}{\partial\rho}\mathbf{Q}(\delta)]\mathbf{Y} = O_p(1)$

as the proof of $\sup_{\delta\in\Delta}\frac{1}{N_1}\mathbf{Y}'[\frac{\partial}{\partial\lambda}\mathbf{Q}(\delta)]\mathbf{Y}=O_p(1)$ is similar and simpler. Note that

$$\begin{aligned}\tfrac{\partial}{\partial\rho}\mathbf{Q}(\delta)= & -\mathbf{C}_N'(\delta)\mathbf{G}_N'(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbb{Q}_{\mathbb{X}}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{C}_N(\delta)+\mathbf{C}_N'(\delta)\dot{\mathbb{Q}}_{\mathbb{D}}(\rho)\mathbb{Q}_{\mathbb{X}}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{C}_N(\delta)\\ & +\mathbf{C}_N'(\delta)\mathbb{Q}_{\mathbb{D}}(\rho)\dot{\mathbb{Q}}_{\mathbb{X}}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{C}_N(\delta)+\mathbf{C}_N'(\delta)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbb{Q}_{\mathbb{X}}(\rho)\dot{\mathbb{Q}}_{\mathbb{D}}(\rho)\mathbf{C}_N(\delta)\\ & -\mathbf{C}_N'(\delta)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbb{Q}_{\mathbb{X}}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{G}_N(\rho)\mathbf{C}_N(\delta),\end{aligned}$$

where $\dot{\mathbb{Q}}_{\mathbb{X}}(\rho)=\frac{\partial}{\partial\rho}\mathbb{Q}_{\mathbb{X}}(\rho)$. Using (B.1), we have after some algebra, $\dot{\mathbb{X}}(\rho)=\frac{\partial}{\partial\rho}\mathbb{X}(\rho)=\mathbb{G}_N(\rho)\mathbb{X}(\rho)$ where $\mathbb{G}_N(\rho)=\mathbb{P}_{\mathbb{D}}(\rho)\mathbf{G}_N'(\rho)-\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{G}_N(\rho)$, which gives

$$\dot{\mathbb{Q}}_{\mathbb{X}}(\rho)=-\mathbb{P}_{\mathbb{X}}(\rho)\mathbb{G}_N'(\rho)\mathbb{Q}_{\mathbb{X}}(\rho)-\mathbb{Q}_{\mathbb{X}}(\rho)\mathbb{G}_N(\rho)\mathbb{P}_{\mathbb{X}}(\rho). \tag{B.6}$$

For a comformable vector a and taking use (B.1) and (B.6), we have after some algebra,

$$a'[\tfrac{\partial}{\partial\rho}\mathbf{Q}(\delta)]a=-2a'\bar{\mathbf{Q}}(\delta)a, \tag{B.7}$$

where $\bar{\mathbf{Q}}(\delta)=\mathcal{Q}_N'(\delta)\mathbf{G}_N(\rho)\mathcal{Q}_N(\delta)$ and $\mathcal{Q}_N(\delta)=\mathbb{Q}_{\mathbb{X}}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{C}_N(\delta)$. Some rearrangements lead to $\bar{\mathbf{Q}}(\delta)=\mathcal{Q}_N'(\delta)\mathbf{M}\bar{\mathbb{Q}}_{\mathbb{D}}(\rho)\bar{\mathbb{Q}}_{\mathbb{X}}(\rho)\mathbf{A}_N(\lambda)$, where $\bar{\mathbb{Q}}_{\mathbb{D}}(\rho)=I_N-\mathbf{D}[\mathbb{D}'(\rho)\mathbb{D}(\rho)]^{-1}\mathbb{D}'(\rho)\mathbf{B}_N(\rho)$ and $\bar{\mathbb{Q}}_{\mathbb{X}}(\rho)=I_N-\mathbf{X}[\mathbb{X}'(\rho)\mathbb{X}(\rho)]^{-1}\mathbb{X}'(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho)$. Following exactly the same way as we prove Lemma A.3, we show that $\bar{\mathbb{Q}}_{\mathbb{D}}(\rho)$ and $\bar{\mathbb{Q}}_{\mathbb{X}}(\rho)$ are also uniformly bounded in both row and column sums, uniformly in $\rho\in\Delta_\rho$. This implies that both $\|\bar{\mathbf{Q}}(\delta)\|_1$ and $\|\bar{\mathbf{Q}}(\delta)\|_\infty$ are bounded uniformly in $\delta\in\Delta$. As $\mathbf{Y}=\mathbf{A}_N^{-1}(\eta+\mathbf{B}_N^{-1}\mathbf{V})$, Lemma A.1 and Lemma A.6 imply

$$\begin{aligned}&\tfrac{1}{N_1}\mathbf{Y}'[\tfrac{\partial}{\partial\rho}\mathbf{Q}(\delta)]\mathbf{Y}=-\tfrac{2}{N_1}\mathbf{Y}'\bar{\mathbf{Q}}(\delta)\mathbf{Y}=-\tfrac{2}{N_1}(\eta+\mathbf{B}_N^{-1}\mathbf{V})'\mathbf{A}_N'^{-1}\bar{\mathbf{Q}}(\delta)\mathbf{A}_N^{-1}(\eta+\mathbf{B}_N^{-1}\mathbf{V})\\ &=-\tfrac{2}{N_1}\eta'\mathbf{A}_N'^{-1}\bar{\mathbf{Q}}(\delta)\mathbf{A}_N^{-1}\eta-\tfrac{4}{N1}\eta'\mathbf{A}_N'^{-1}\bar{\mathbf{Q}}(\delta)\mathbf{C}_N^{-1}\mathbf{V}-\tfrac{2}{N_1}\mathbf{V}'\mathbf{C}_N'^{-1}\bar{\mathbf{Q}}(\delta)\mathbf{C}_N^{-1}\mathbf{V}=O_p(1),\end{aligned}$$

uniformly in $\delta\in\Delta$. Thus, $\sup_{\delta\in\Delta}\frac{1}{N_1}\mathbf{Y}'[\frac{\partial}{\partial\rho}\mathbf{Q}(\delta)]\mathbf{Y}=O_p(1)$. Following the similar analysis, one also has $\sup_{\delta\in\Delta}\frac{1}{N_1}\mathbf{Y}'[\frac{\partial}{\partial\lambda}\mathbf{Q}(\delta)]\mathbf{Y}=O_p(1)$. Therefore, $\sup_{\delta\in\Delta}|\hat{\sigma}^{*2}_{v,N}(\delta)-\bar{\sigma}^{*2}_{v,N}(\delta)|=o_p(1)$.

Proof of (c). By the expressions of $\hat{\mathbf{V}}(\lambda)$ and $\bar{\mathbf{V}}(\delta)$ given above, we have

$$\begin{aligned}&\tfrac{1}{N_1}\mathbf{Y}'\mathbf{W}'\mathbf{B}_N'(\rho)\hat{\mathbf{V}}(\delta)-\tfrac{1}{N_1}\mathrm{E}[\mathbf{Y}'\mathbf{W}'\mathbf{B}_N'(\rho)\bar{\mathbf{V}}(\delta)]\\ &=\tfrac{1}{N_1}[\mathbf{Y}'\mathbf{W}'\mathbf{B}_N'(\rho)\mathcal{Q}_N(\delta)\mathbf{Y}-\mathrm{E}(\mathbf{Y}'\mathbf{W}'\mathbf{B}_N'(\rho)\mathcal{Q}_N(\delta)\mathbf{Y})]-\tfrac{\sigma^2_{v0}}{N_1}\mathtt{tr}[\mathbf{C}_N'^{-1}\mathbf{W}'\mathbf{B}_N'(\rho)\mathcal{P}_N(\delta)\mathbf{C}_N^{-1}],\end{aligned}$$

where $\mathcal{P}_N(\delta)=\mathbb{P}_{\mathbb{X}}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{C}_N(\delta)$. The first term is similar in form to $\frac{1}{N_1}[\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y}-\mathrm{E}(\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y})]$ from **(b)**, and its uniform convergence is shown in a similar way. Furthermore, by Lemma A.4, it is easy to see that the second term is $o(1)$ uniformly in $\delta\in\Delta$.

Proof of (d). Again, using the expressions of $\bar{\mathbf{V}}(\delta)$ and $\hat{\mathbf{V}}(\delta)$, we have

$$\begin{aligned}&\tfrac{1}{N_1}\hat{\mathbf{V}}'(\delta)\mathbf{G}_N(\rho)\hat{\mathbf{V}}(\delta)-\tfrac{1}{N_1}\mathrm{E}[\bar{\mathbf{V}}'(\delta)\mathbf{G}_N(\rho)\bar{\mathbf{V}}(\delta)]\\ &=\tfrac{1}{N_1}[\mathbf{Y}'\bar{\mathbf{Q}}(\delta)\mathbf{Y}-\mathrm{E}(\mathbf{Y}'\bar{\mathbf{Q}}(\delta)\mathbf{Y})]-\tfrac{\sigma^2_{v0}}{N_1}\mathtt{tr}[\mathbf{C}_N'^{-1}\mathcal{P}_N'(\delta)\mathbf{G}_N^{\circ}(\rho)\mathcal{Q}_N(\delta)\mathbf{C}_N^{-1}]\\ &\quad-\tfrac{\sigma^2_{v0}}{N_1}\mathtt{tr}[\mathbf{C}_N'^{-1}\mathcal{P}_N'(\delta)\mathbf{G}_N(\rho)\mathcal{P}_N(\delta)\mathbf{C}_N^{-1}].\end{aligned}$$

Therefore, the uniform convergence of the first term can also be shown similarly as we do for $\frac{1}{N_1}[\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y} - \mathrm{E}(\mathbf{Y}'\mathbf{Q}(\delta)\mathbf{Y})]$ since they have similar forms. By Lemma A.4, the remaining two terms are easily seen to be $o(1)$, uniformly in $\delta \in \Delta$. ■

Proof of Theorem 2.2: Applying the MVT to each element of $S_N^*(\hat{\theta}_N^*)$, we have

$$0 = \frac{1}{\sqrt{N_1}} S_N^*(\hat{\theta}_N^*) = \frac{1}{\sqrt{N_1}} S_N^*(\theta_0) + \left[\frac{1}{N_1} \frac{\partial}{\partial \theta'} S_N^*(\theta) \Big|_{\theta = \bar{\theta}_r \text{ in } r\text{th row}} \right] \sqrt{N_1}(\hat{\theta}_N^* - \theta_0), \qquad \text{(B.8)}$$

where $\{\bar{\theta}_r\}$ are on the line segment between $\hat{\theta}_N^*$ and θ_0. The result of the theorem follows if

(a) $\frac{1}{\sqrt{N_1}} S_N^*(\theta_0) \xrightarrow{D} N[0, \lim_{N\to\infty} \Gamma_N^*(\theta_0)]$,

(b) $\frac{1}{N_1}[\frac{\partial}{\partial \theta'} S_N^*(\theta)\big|_{\theta = \bar{\theta}_r \text{ in } r\text{th row}} - \frac{\partial}{\partial \theta'} S_N^*(\theta_0)] = o_p(1)$, and

(c) $\frac{1}{N_1}[\frac{\partial}{\partial \theta'} S_N^*(\theta_0) - \mathrm{E}(\frac{\partial}{\partial \theta'} S_N^*(\theta_0))] = o_p(1)$.

Proof of (a). From (2.13), we see that the elements of $S_N^*(\theta_0)$ are linear-quadratic forms in $\mathbf{V}$. Thus, for every non-zero $(k+3) \times 1$ vector of constants a, $a' S_N^*(\theta_0)$ is of the form:

$$a' S_N^*(\theta_0) = b_N' \mathbf{V} + \mathbf{V}' \Phi_N \mathbf{V} - \sigma_v^2 \mathtt{tr}(\Phi_N),$$

for suitably defined non-stochastic vector b_N and matrix Φ_N. Based on Assumptions A-F, it is easy to verify (by Lemma A.1 and Lemma A.3(i)) that b_N and matrix Φ_N satisfy the conditions of the CLT for LQ form of Kelejian and Prucha (2001), and hence the asymptotic normality of $\frac{1}{\sqrt{N_1}} a' S_N^*(\theta_0)$ follows. By Cramér-Wold device, $\frac{1}{\sqrt{N_1}} S_N^*(\theta_0) \xrightarrow{D} N[0, \lim_{N\to\infty} \Gamma_N^*(\theta_0)]$, where elements of $\Gamma_N^*(\theta_0)$ are given in (2.15).

Proof of (b). The Hessian matrix $H_N^*(\theta) = \frac{\partial}{\partial \theta'} S_N^*(\theta)$ is given in (B.4). By Assumptions D and E, and Lemma A.1 and Lemma A.3(i), $\mathbb{R}_N(\rho_0)$, $\mathcal{R}_{1N}(\rho_0)$ and $\mathcal{R}_{2N}(\rho_0)$ are all bounded in row and column sum norms. With these and $\mathbf{Y} = \mathbf{A}_N^{-1}(\eta + \mathbf{B}_N^{-1}\mathbf{V})$, Lemma A.6 leads to $\frac{1}{N_1} H_N^*(\theta_0) = O_p(1)$. Thus, $\frac{1}{N_1} H_N^*(\bar{\theta}) = O_p(1)$ since $\bar{\theta} \xrightarrow{p} \theta_0$ due to $\hat{\theta}_N^* \xrightarrow{p} \theta_0$, where for ease of exposition, $H_N^*(\bar{\theta})$ is used to denote $\frac{\partial}{\partial \theta'} S_N^*(\theta)\big|_{\theta = \bar{\theta}_r \text{ in } r\text{th row}}$. As $\bar{\sigma}_v^2 \xrightarrow{p} \sigma_{v0}^2$, we have $\bar{\sigma}_v^{-r} = \sigma_{v0}^{-r} + o_p(1)$, for $r = 2, 4, 6$. As σ_v^{-r} appears in $H_N^*(\theta)$ multiplicatively, $\frac{1}{N_1} H_N^*(\bar{\theta}) = \frac{1}{N_1} H_N^*(\bar{\beta}, \bar{\lambda}, \bar{\rho}, \sigma_{v0}^2) + o_p(1)$. Thus, the proof of **(b)** is equivalent to the proof of

$$\frac{1}{N_1}[H_N^*(\bar{\beta}, \bar{\lambda}, \bar{\rho}, \sigma_{v0}^2) - H_N^*(\theta_0)] \xrightarrow{p} 0,$$

or the proofs of $\frac{1}{N_1}[H_N^{*\mathtt{S}}(\bar{\beta}, \bar{\lambda}, \bar{\rho}, \sigma_{v0}^2) - H_N^{*\mathtt{S}}(\theta_0)] \xrightarrow{p} 0$ and $\frac{1}{N_1}[H_N^{*\mathtt{NS}}(\bar{\delta}) - H_N^{*\mathtt{NS}}(\delta_0)] \xrightarrow{p} 0$, where $H_N^{*\mathtt{S}}$ and $H_N^{*\mathtt{NS}}$ denote, respectively, the stochastic and non-stochastic parts of H_N^*.

For the stochastic part, we see from (B.4) that all the components of $H_N^{*\mathtt{S}}(\beta, \lambda, \rho, \sigma_{v0}^2)$ are linear, bilinear or quadratic in β and λ, but nonlinear in ρ. Hence, with an application of the

MVT on $H_N^{*\mathtt{S}}(\bar\beta,\bar\lambda,\bar\rho,\sigma_{v0}^2)$ w.r.t $\bar\rho$ 'variable', we can write $\frac{1}{N_1}[H_N^{*\mathtt{S}}(\bar\beta,\bar\lambda,\bar\rho,\sigma_{v0}^2)-H_N^{*\mathtt{S}}(\theta_0)]$ as

$$\tfrac{1}{N_1}[\tfrac{\partial}{\partial\rho}H_N^{*\mathtt{S}}(\bar\beta,\bar\lambda,\dot\rho,\sigma_{v0}^2)](\bar\rho-\rho_0)+\tfrac{1}{N_1}[H_N^{*\mathtt{S}}(\bar\beta,\bar\lambda,\rho_0,\sigma_{v0}^2)-H_N^{*\mathtt{S}}(\theta_0)],$$

where $\dot\rho$ lies between $\bar\rho$ and ρ_0. Therefore, it suffices to show (i) $\frac{1}{N_1}\frac{\partial}{\partial\rho}H_N^{*\mathtt{S}}(\bar\beta,\bar\lambda,\dot\rho,\sigma_{v0}^2)=O_p(1)$, and (ii) $\frac{1}{N_1}[H_N^{*\mathtt{S}}(\bar\beta,\bar\lambda,\rho_0,\sigma_{v0}^2)-H_N^{*\mathtt{S}}(\theta_0)]=o_p(1)$.

We select one of the most complicated components, $H_{\rho\lambda}^{*\mathtt{S}}(\theta)=-\frac{1}{\sigma_v^2}\mathbb{Y}'(\rho)\mathbf{G}_N^\circ(\rho)\tilde{\mathbf{V}}(\beta,\delta)$, to illustrate the general idea in the proof. We have, after some algebra,

$$\tfrac{1}{N_1}\tfrac{\partial}{\partial\rho}H_{\rho\lambda}^{*\mathtt{S}}(\bar\beta,\bar\lambda,\dot\rho,\sigma_{v0}^2)=\tfrac{2}{N_1\sigma_{v0}^2}\mathbb{Y}'(\dot\rho)\mathcal{R}_{1N}(\dot\rho)\mathbb{Q}_{\mathbb{D}}(\dot\rho)\mathbf{B}_N(\dot\rho)(\mathbf{A}_N(\bar\lambda)\mathbf{Y}-\mathbf{X}\bar\beta),$$

$$\tfrac{1}{N_1}[H_N^{*\mathtt{S}}(\bar\beta,\bar\lambda,\rho_0,\sigma_{v0}^2)-H_N^{*\mathtt{S}}(\theta_0)]=\tfrac{1}{N_1\sigma_{v0}^2}\mathbb{Y}'\mathbf{G}_N^\circ\mathbb{Y}(\bar\lambda-\lambda_0)+\tfrac{1}{N_1\sigma_{v0}^2}\mathbb{Y}'\mathbf{G}_N^\circ\mathbb{X}(\bar\beta-\beta_0).$$

By Lemmas A.1 and A.6, it is easy to show that $\frac{1}{N_1}\mathbb{Y}'\mathbf{G}_N^\circ\mathbb{Y}=O_p(1)$ and $\frac{1}{N_1}\mathbb{Y}'\mathbf{G}_N^\circ\mathbb{X}=O_p(1)$. Therefore, (ii) holds. To prove (i), we have

$$\begin{aligned}&\mathbb{Y}'(\dot\rho)\mathcal{R}_{1N}(\dot\rho)\mathbb{Q}_{\mathbb{D}}(\dot\rho)\mathbf{B}_N(\dot\rho)(\mathbf{A}_N(\bar\lambda)\mathbf{Y}-\mathbf{X}\bar\beta)\\ =&(\mathbf{A}_N^{-1}\eta+\mathbf{C}_N^{-1}\mathbf{V})'\mathcal{H}_N(\dot\rho)[\mathbf{A}_N(\bar\lambda)\mathbf{A}_N^{-1}\eta+\mathbf{A}_N(\bar\lambda)\mathbf{C}_N^{-1}\mathbf{V}-\mathbf{X}\bar\beta]\end{aligned}$$

where $\mathcal{H}_N(\dot\rho)=\mathbf{W}'\mathbf{B}_N'(\dot\rho)\mathbb{Q}_{\mathbb{D}}(\dot\rho)\mathcal{R}_{1N}(\dot\rho)\mathbb{Q}_{\mathbb{D}}(\dot\rho)\mathbf{B}_N(\dot\rho)$. Lemma A.2 implies $\mathbf{B}_N^{-1}(\dot\rho)$ embedded in $\mathcal{H}_N(\dot\rho)$ is uniformly bounded in both row and column sums since $\dot\rho-\rho_0=o_p(1)$. Therefore, it is easy to see the above equation is $O_p(N)$ by Lemma A.6 and then result (i) follows.

For the non-stochastic part, we illustrate the proof using the most complicate $\lambda\lambda$-term. Noting that the non-stochastic part is nonlinear in both $\bar\lambda$ and $\bar\rho$, we have by the MVT,

$$\begin{aligned}&\tfrac{1}{N_1}[H_{\lambda\lambda}^{*\mathtt{NS}}(\bar\delta)-H_{\lambda\lambda}^{*\mathtt{NS}}(\delta_0)]=-\tfrac{1}{N_1}\mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\bar\rho)\mathbf{B}_N(\bar\rho)\mathbf{F}_N^2(\bar\lambda)\mathbf{B}_N^{-1}(\bar\rho)-\mathbb{Q}_{\mathbb{D}}\mathbf{B}_N\mathbf{F}_N^2\mathbf{B}_N^{-1}]\\ =&-(\bar\lambda-\lambda_0)\tfrac{1}{N_1}\mathtt{tr}[2\mathbb{Q}_{\mathbb{D}}(\dot\rho)\mathbf{B}_N(\dot\rho)\mathbf{F}_N^3(\dot\lambda)\mathbf{B}_N^{-1}(\dot\rho)]-(\bar\rho-\rho_0)\tfrac{1}{N_1}\mathtt{tr}[\mathbf{F}_N^2(\dot\lambda)\mathbb{R}_N(\dot\rho)],\end{aligned}$$

where $\dot\lambda$ lies between $\bar\lambda$ and λ_0 and $\dot\rho$ lies between $\bar\rho$ and ρ_0. Again, by Lemma A.2, we conclude that both $\mathbf{A}_N^{-1}(\dot\lambda)$ and $\mathbf{B}_N^{-1}(\dot\rho)$ are uniformly bounded in both row and column sums. Therefore, the terms inside the trace both have elements that are uniformly bounded. As $\bar\delta-\delta_0=o_p(1)$, we have $\frac{1}{N_1}[H_{\lambda\lambda}^{*\mathtt{NS}}(\bar\delta)-H_{\lambda\lambda}^{*\mathtt{NS}}(\delta_0)]=o_p(1)$.

Proof of (c). Since $\mathbf{Y}=\mathbf{A}_N^{-1}(\eta+\mathbf{B}_N^{-1}\mathbf{V})$, the Hessian matrix at true θ_0 are seen to be linear combinations of terms linear or quadratic in $\mathbf{V}$, and constants. The constant terms are canceled out. Other terms are shown to be $o_p(1)$ based on Lemma A.6. For example,

$$\tfrac{1}{N_1}[H_{\rho\rho}^*(\rho_0)-\mathrm{E}(H_{\rho\rho}^*(\rho_0))]=\tfrac{1}{N_1\sigma_{v0}^2}[\mathbf{V}'\mathbb{Q}_{\mathbb{D}}\mathcal{R}_{1N}\mathbb{Q}_{\mathbb{D}}\mathbf{V}-\mathrm{E}(\mathbf{V}'\mathbb{Q}_{\mathbb{D}}\mathcal{R}_{1N}\mathbb{Q}_{\mathbb{D}}\mathbf{V})]=o_p(1). \qquad \blacksquare$$

Proof of Corollary 2.1: The proof is given in Appendix D. ■

Proof of Corollary 2.2: The proof is given in Appendix D. ■

Appendix C: Proofs for Section 3

Derivation of the Hessian matrix for robust AQS functions:

With the set of robust AQS functions $S_N^\diamond(\xi)$ given in (3.3), we obtain the components of the Hessian matrix $H_N^\diamond(\xi) = \frac{\partial}{\partial \xi'} S_N^\diamond(\xi)$:

$$
\begin{aligned}
H^\diamond_{\beta\beta}(\xi) &= -\mathbb{X}'(\rho)\mathbb{X}(\rho), \qquad\qquad H^\diamond_{\beta\lambda}(\xi) = -\mathbb{X}'(\rho)\mathbb{Y}(\rho),\\
H^\diamond_{\beta\rho}(\xi) &= -\mathbb{X}'(\rho)\mathbf{G}^\circ_N(\rho)\tilde{\mathbf{V}}(\beta,\delta), \quad H^\diamond_{\lambda\beta}(\xi) = -\mathbf{Y}'\mathbf{C}'_N(\delta)\mathbb{L}'_\lambda(\delta)\mathbb{X}(\rho),\\
H^\diamond_{\lambda\lambda}(\xi) &= -\mathbb{Y}'(\rho)\mathbb{Y}(\rho) + \mathbf{Y}'\mathbf{W}'\mathbf{B}'_N(\rho)\bar{\mathbb{F}}'_N(\delta)\tilde{\mathbf{V}}(\beta,\delta)\\
&\quad -\mathbf{Y}'\mathbf{C}'_N(\delta)[\bar{\mathbb{F}}'_{N\lambda}(\delta)\tilde{\mathbf{V}}(\beta,\delta) - \bar{\mathbb{F}}'_N(\delta)\mathbb{Y}(\rho)],\\
H^\diamond_{\lambda\rho}(\xi) &= -\mathbb{Y}'(\rho)\mathbf{G}^\circ_N(\rho)\tilde{\mathbf{V}}(\beta,\delta) - \mathbf{Y}'\mathbf{C}'_N(\delta)[-\mathbf{G}'_N(\delta)\bar{\mathbb{F}}'_N(\delta)\\
&\quad +\bar{\mathbb{F}}'_{N\rho}(\delta) + \bar{\mathbb{F}}'_N(\delta)\mathbb{G}_N(\rho)]\tilde{\mathbf{V}}(\beta,\delta),\\
H^\diamond_{\rho\beta}(\xi) &= -[\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\beta]'\mathbf{B}'_N(\rho)\mathbb{L}'_\rho(\rho)\mathbb{X}(\rho) - \tilde{\mathbf{V}}'(\beta,\delta)\mathbb{L}_\rho(\rho)\mathbf{B}_N(\rho)\mathbf{X},\\
H^\diamond_{\rho\lambda}(\xi) &= -\mathbb{Y}'(\rho)\mathbf{G}^\circ_N(\rho)\tilde{\mathbf{V}}(\beta,\delta) + \mathbf{Y}'\mathbf{W}'\mathbf{B}'_N(\rho)\bar{\mathbb{G}}_N(\rho)\tilde{\mathbf{V}}(\beta,\delta)\\
&\quad +[\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\beta]'\mathbf{B}'_N(\rho)\bar{\mathbb{G}}_N(\rho)\mathbb{Y}(\rho),\\
H^\diamond_{\rho\rho}(\xi) &= \tilde{\mathbf{V}}'(\beta,\delta)\mathcal{R}_{1N}(\rho)\tilde{\mathbf{V}}(\beta,\delta) - [\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\beta]'\mathbf{B}'_N(\rho)[-\mathbf{G}'_N(\delta)\bar{\mathbb{G}}_N(\rho)\\
&\quad +\bar{\mathbb{G}}_{N\rho}(\rho) + \bar{\mathbb{G}}_N(\rho)\mathbb{G}_N(\rho)]\tilde{\mathbf{V}}(\beta,\delta),
\end{aligned}
\tag{C.1}
$$

where $\bar{\mathbb{F}}'_{N\lambda}(\delta) = \texttt{diag}[\mathbf{B}_N^{-1\prime}(\rho)\mathbf{F}'^2_N(\lambda)\mathbf{B}'_N(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)]\texttt{diag}[\mathbb{Q}_{\mathbb{D}}(\rho)]^{-1}$,

$\bar{\mathbb{F}}'_{N\rho}(\delta) = \texttt{diag}[\mathcal{K}_{1N}(\delta)]\texttt{diag}[\mathbb{Q}_{\mathbb{D}}(\rho)]^{-1} - \bar{\mathbb{F}}'_N(\delta)\texttt{diag}[\dot{\mathbb{Q}}_{\mathbb{D}}(\rho)]\texttt{diag}[\mathbb{Q}_{\mathbb{D}}(\rho)]^{-1}$,

$\mathcal{K}_{1N}(\delta) = \bar{\mathbf{F}}'_N(\delta)\mathbf{G}'_N(\rho)\mathbb{Q}_{\mathbb{D}}(\rho) - \bar{\mathbf{F}}'_N(\delta)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{G}'_N(\rho) + \bar{\mathbf{F}}'_N(\delta)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{G}^\circ_N(\rho)\mathbb{P}_{\mathbb{D}}(\rho)$,

$\bar{\mathbb{G}}_{N\rho}(\rho) = \texttt{diag}[\mathcal{K}_{2N}(\rho)]\texttt{diag}[\mathbb{Q}_{\mathbb{D}}(\rho)]^{-1} - \bar{\mathbb{G}}_N(\rho)\texttt{diag}[\dot{\mathbb{Q}}_{\mathbb{D}}(\rho)]\texttt{diag}[\mathbb{Q}_{\mathbb{D}}(\rho)]^{-1}$,

$\mathcal{K}_{2N}(\rho) = [\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{G}_N(\rho)\mathbb{P}_{\mathbb{D}}(\rho) + \mathbb{P}_{\mathbb{D}}(\rho)\mathbf{G}'_N(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)]\mathbf{G}^\circ_N(\rho)\mathbb{Q}_{\mathbb{D}}(\rho) + \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{G}^2_N(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)$.

Proof of Theorem 3.1. Since the consistency of $\hat{\beta}^\diamond_N$ follows almost immediately that of $\hat{\delta}^\diamond_N$ under Assumptions C and E, we only need to prove that $\hat{\delta}^\diamond_N$ is consistent to δ_0. By theorem 5.9 of van der Vaart (1988), $\hat{\delta}^\diamond_N$ will be consistent for δ_0 if $\sup_{\delta\in\Delta} \frac{1}{N_1}\left\|S^{\diamond c}_N(\delta) - \bar{S}^{\diamond c}_N(\delta)\right\| \xrightarrow{p} 0$.

Let $\mathbb{L}_\lambda(\delta) = \mathbb{Q}_{\mathbb{D}}(\rho)[\bar{\mathbf{F}}_N(\delta) - \bar{\mathbb{F}}_N(\delta)]$, $\mathbb{L}_\rho(\rho) = \mathbb{Q}_{\mathbb{D}}(\rho)[\bar{\mathbf{G}}'_N(\rho) - \bar{\mathbb{G}}'_N(\rho)]$ and $\mathbb{N}_N(\rho) = I_N - \mathbb{M}_N(\rho)$. Note that $\mathbf{B}_N(\rho)[\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\hat{\beta}^\diamond_N(\delta)] = \mathbb{M}_N(\rho)\mathbf{C}_N(\delta)\mathbf{Y}$ and $\mathbf{B}_N(\rho)[\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\bar{\beta}^\diamond_N(\delta)] = \mathbb{M}_N(\rho)\mathbf{C}_N(\delta)\mathbf{Y} + \mathbb{N}_N(\rho)\mathbf{C}_N(\delta)[\mathbf{Y} - \mathrm{E}(\mathbf{Y})]$. Recall $\hat{\mathbf{V}}(\delta) = \mathcal{Q}_N(\delta)\mathbf{Y}$ and $\bar{\mathbf{V}}(\delta) = \mathcal{Q}_N(\delta)\mathbf{Y} + \mathcal{P}_N(\delta)[\mathbf{Y} - \mathrm{E}(\mathbf{Y})]$. With Assumption G$'$, the consistency of $\hat{\delta}^\diamond_N$ follows if:

(i) $\sup_{\delta\in\Delta} \frac{1}{N_1}\left|\mathbf{Y}'\mathbf{Q}^h_r(\delta)\mathbf{Y} - \mathrm{E}[\mathbf{Y}'\mathbf{Q}^h_r(\delta)\mathbf{Y}]\right| = o_p(1)$, for $r = 1, 2$;

(ii) $\sup_{\delta\in\Delta} \frac{\sigma^2_{v0}}{N_1}\texttt{tr}[\mathbf{C}'^{-1}_N\mathbf{P}^h_s(\delta)\mathbf{C}^{-1}_N] = o(1)$, for $s = 1, 2, 3$;

where $\mathbf{Q}^h_1(\delta) = \mathbf{C}'_N(\delta)\mathbb{L}'_\lambda(\delta)\mathcal{Q}_N(\delta)$, $\mathbf{Q}^h_2(\delta) = \mathbf{C}'_N(\delta)\mathbb{M}'_N(\rho)\mathbb{L}'_\rho(\rho)\mathcal{Q}_N(\delta)$, $\mathbf{P}^h_1(\delta) = \mathbf{C}'_N(\delta)\mathbb{L}'_\lambda(\delta)\mathcal{P}_N(\delta)$,

$\mathbf{P}_2^h(\delta) = \mathbf{C}_N'(\delta)\mathbb{L}_\rho'(\rho)\mathcal{P}_N(\delta)$ and $\mathbf{P}_3^h(\delta) = \mathbf{C}_N'(\delta)\mathbb{N}_N'(\rho)\mathbb{L}_\rho'(\rho)\mathcal{Q}_N(\delta)$.

Note that $\mathbf{Q}_1^h(\delta) = \mathbf{C}_N'(\delta)[\bar{\mathbf{F}}_N'(\delta) - \bar{\mathbb{F}}_N'(\delta)]\mathcal{Q}_N(\delta) = \mathbf{W}'\mathbf{B}_N'(\rho)\mathcal{Q}_N(\delta) - \mathbf{C}_N'(\delta)\bar{\mathbb{F}}_N'(\delta)\mathcal{Q}_N(\delta)$. As $\bar{\mathbb{F}}_N'(\delta)$ is a diagonal matrix which is naturally bounded in both row and column sums, uniformly in $\delta \in \Delta$, we conclude $\mathbf{Q}_1^h(\delta)$ is bounded in both row and column sum norms, uniformly in $\delta \in \Delta$, by Lemma A.1. Similarly, $\mathbf{Q}_2^h(\delta) = \mathbf{C}_N'(\delta)\mathbb{M}_N'(\rho)[\bar{\mathbf{G}}_N(\rho) - \bar{\mathbb{G}}_N(\rho)]\mathcal{Q}_N(\delta) = \bar{\mathbf{Q}}(\delta) - \mathbf{C}_N'(\delta)\mathbb{M}_N'(\rho)\bar{\mathbb{G}}_N(\rho)\mathcal{Q}_N(\delta)$ is also bounded in both row and column sum norms, uniformly in $\delta \in \Delta$. Hence, $\mathbf{Q}_1^h(\delta)$ and $\mathbf{Q}_2^h(\delta)$ have forms similar to $\mathbf{Q}(\delta)$. The proof of (i) thus follows that of Theorem 2.1 **(b)**. For (ii), noting that $\mathcal{P}_N(\delta) = \mathbb{P}_{\mathbb{X}}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{C}_N(\delta)$, we have $\sup_{\delta\in\Delta} \frac{\sigma_{v0}^2}{N_1}\mathtt{tr}[\mathbf{C}_N'^{-1}\mathbf{P}_s^h(\delta)\mathbf{C}_N^{-1}] = o(1), s = 1, 2$, by Lemma A.4. For the final result, we have,

$$\begin{aligned}\tfrac{1}{N_1}\mathtt{tr}[\mathbf{C}_N'^{-1}\mathbf{P}_3^h(\delta)\mathbf{C}_N^{-1}] &= -\tfrac{1}{N_1}\mathtt{tr}[\mathbf{C}_N'(\delta)\mathbb{N}_N'(\rho)\mathbb{L}_\rho'(\rho)\mathcal{Q}_N(\delta)\mathrm{Var}(\mathbf{Y})] \\ &= -\tfrac{1}{N_1}\mathtt{tr}\big[(\tfrac{1}{N_1}\mathbb{X}'(\rho)\mathbb{X}(\rho))^{-1}(\tfrac{1}{N_1}\mathbf{X}'\mathbf{B}_N'(\rho)\mathbb{L}_\rho'(\rho)\mathbb{Q}_{\mathbb{X}}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathcal{C}_N^h(\delta)\mathbb{X}(\rho))\big].\end{aligned}$$

Assumption C implies that the elements of $[\frac{1}{N_1}\mathbb{X}'(\rho)\mathbb{X}(\rho)]^{-1}$ are uniformly bounded for large enough N, uniformly in $\rho \in \Delta_\rho$. Lemma A.1 and Lemma A.3 together imply the term between $\mathbf{X}'$ and $\mathbb{X}(\rho)$ are uniformly bounded in both row and column sums, uniformly in $\delta \in \Delta$. Hence, the elements of the second part in the trace are also uniformly bounded. As the number of regressors k is finite, the quantity $\frac{\sigma_{v0}^2}{N_1}\mathtt{tr}[\mathbf{C}_N'^{-1}\mathbf{P}_3^h(\delta)\mathbf{C}_N^{-1}]$ will shrink to zero as N goes large, uniformly in $\delta \in \Delta$. These complete the proof of the theorem. ∎

Proofs of Theorem 3.2, Corollary 3.1, Lemma 3.1 and Corollary 3.2 are in Appendix D.

Appendix D: Supplementary data

Supplementary material, containing additional proofs and some extensions, can be found online at `http://www.mysmu.edu.sg/faculty/zlyang/SubPages/research.htm`

Acknowledgments

We thank Lung-Fei Lee, Ingmar Prucha, Xi Qu, and the participants of Asia Meeting of the Econometric Society (AMES) 2019, Xiamen, China, XIII Conference of Spatial Econometrics Association 2019, Pittsburgh, US, Shanghai Workshop of Econometrics 2019, SUFE, Shanghai, China, and AMES 2021, Curtin University, Malaysia, for their helpful comments. Zhenlin Yang gratefully acknowledges the financial support from Singapore Management University under Lee Kong Chian Fellowship.

References

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

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

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

[4] Anselin, L., Le Gallo, J., & Jayet, H., 2008. Spatial panel econometrics. In *The econometrics of panel data: Fundamentals and Recent Development in Theory and Practice* Springer, Berlin, Heidelberg. 625-660.

[5] Antweiler, W., 2001. Nested random effects estimation in unbalanced panel data. *Journal of econometrics* 101, 295-313.

[6] Badinger, H., Egger, P., 2015. Fixed effects and random effects estimation of higher-order spatial autoregressive models with spatial autoregressive and heteroscedastic disturbances. *Spatial Economic Analysis* 10, 11-35.

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

[8] Bai, J., 2013. Likelihood approach to dynamic panel models with interactive effects. *Working paper*, Columbia University.

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

[10] Bai, J., Liao, Y., & Yang, J., 2015. Unbalanced panel data models with interactive effects. In: Badi H. Baltagi (Eds). *The Oxford Handbook of Panel Data*.

[11] Baltagi, B. H., 2013. *Econometric Analysis of Panel Data*, 5th edition. John Wiley & Sons Ltd.

[12] Baltagi, B. H., Bresson, G., Etienne, J. M., 2015. Hedonic housing prices in Paris: an unbalanced spatial lag pseudo-panel model with nested random effects. *Journal of Applied Econometrics* 30, 509-528.

[13] Baltagi, B. H., Chang, Y. J., 1994. Incomplete panels: A comparative study of alternative estimators for the unbalanced one-way error component regression model. *Journal of Econometrics* 62, 67-89.

[14] Baltagi, B. H., Egger, P., Pfaffermayr, M., 2007. Estimating models of complex FDI: Are there third-country effects? *Journal of econometrics* 140, 260-281.

[15] Baltagi, B. H., Song, 2006. Unbalanced panel data: A survey. *Statistical Papers* 47, 493-523.

[16] Baltagi, B. H., Song, S. H., Jung, B. C., 2001. The unbalanced nested error component regression model. *Journal of econometrics* 101, 357-381.

[17] Baltagi, B. H., Song, S. H., Koh, W., 2003. Testing panel data regression models with spatial error correlation. *Journal of econometrics* 117, 123-150.

[18] Baltagi, B. H., Yang, Z. L., 2013a. Standardized LM tests for spatial error dependence in linear or panel regressions. *Econometrics Journal* 16, 103-134.

[19] Baltagi, B. H., Yang, Z. L., 2013b. Heteroskedasticity and non-normality robust LM tests of spatial dependence. *Regional Science and Urban Economics* 43, 725-739.

[20] Davis, P., 2001. Estimating multi-way error components models with unbalanced data structures using instrumental variables. *Journal of Econometrics* 106, 67-95.

[21] Egger, P., Pfaffermayr, M., Winner H., 2005. An unbalanced spatial panel data approach to US state tax competition. *Economics Letter* 88, 329-335.

[22] Greene, W. H., 2018. *Econometric Analysis*, 8th Edition. Pearson.

[23] Hsiao, C., 2014. *Analysis of panel data*. Cambridge university press.

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

[25] Kapoor, M., Kelejian, H. H., Prucha, I. R., 2007. Panel data models with spatially correlated error components. *Journal of econometrics* 140, 97-130.

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

[27] Kelejian, H. H., Prucha, I. R., 2001. On the asymptotic distribution of the Moran I test statistic with applications. *Journal of econometrics* 104, 219-257.

[28] Kelejian, H. H., Prucha, I. R., 2010. Spatial models with spatially lagged dependent variables and incomplete data. *Journal of geographical systems* 12, 241-257.

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

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

[31] Lee, L. F., 2004b. A supplement to 'Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models'. *Working paper*, Department of Economics, Ohio State University.

[32] Lee, L. F., Yu, J., 2010. Estimation of spatial autoregressive panel data models with fixed effects. *Journal of Econometrics* 154, 165-185.

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

[34] LeSage, J. P., Pace, R. K. 2004. Models for spatially dependent missing data. *The Journal of Real Estate Finance and Economics* 29, 233-254.

[35] Li, L. Y., Yang, Z. L., 2020a. M-Estimation of Fixed Effects Spatial Dynamic Panel Data Models with Small T and Unknown Heteroskedasticity. *Regional Science and Urban Economics*. In Press, https://doi.org/10.1016/j.regsciurbeco.2020.103520.

[36] Li, L. Y., Yang, Z. L., 2020b. Spatial Dynamic Panel Data Models with Correlated Random Effects. *Journal of Econometrics*, forthcoming.

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

[38] Liu, X., Lee, L. F., 2010. GMM estimation of social interaction models with centrality. *Journal of Econometrics* 159, 99-115.

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

[40] Liu, S. F. Yang, Z. L., 2020. Robust estimation and inference of spatial panel data models with fixed effects. *Japanese Journal of Statistics and Data Science*, forthcoming.

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

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

[43] Varah, J. M., 1975. A lower bound for the smallest singular value of a matrix. *Linear Algebra and its Applications* 11, 3-5.

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

[45] Wansbeek, T., Kapteyn, A., 1989. Estimation of the error-components model with incomplete panels. *Journal of Econometrics* 41, 341-361.

[46] Wang, W., Lee, L. F., 2013a. Estimation of spatial autoregressive models with randomly missing data in the dependent variable. *Econometrics Journal* 16, 73-102.

[47] Wang, W., Lee, L. F., 2013b. Estimation of spatial panel data models with randomly missing data in the dependent variable. *Regional Science and Urban Economics* 43, 521-538.

[48] Wooldridge, J. M., 2019. Correlated random effects models with unbalanced panels. *Journal of Econometrics* 211 137-150.

[49] Xu, Y. H., Yang, Z. L., 2020. Specification tests for temporal heterogeneity in spatial panel data models with fixed effects. *Regional Science and Urban Economics*. In Press, `https://doi.org/10.1016/j.regsciurbeco.2019.103488`.

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

[51] Yang, Z. L., 2018. Unified M-estimation of Fixed-Effects Spatial Dynamic Models with Short Panels. *Journal of Econometrics* 205, 423-447.

[52] Yang, Z. L., Yu, J. H, & Liu, S. F., 2016. Bias correction and refined inferences for fixed effects spatial panel data models. *Regional Science and Urban Economics* 61, 52-72.

Table 1a. Empirical mean(sd)[$\hat{sd}$] of the estimators for FE-SPD model with `SL-SE` effects 10% random missing, **homoskedasticity**, $(\beta_1, \lambda, \rho, \sigma_v^2) = (1, 0.2, 0.2, 1)$, **T=5**.

	W= Rook, M=Queen			W=Group-I, M=Queen		
	QMLE	AQSE	RAQSE	QMLE	AQSE	RAQSE
	$n = 50$; error = 1, 2, 3, for the three panels below					
β_1	.9998(.039)	1.0007(.039)[.039]	1.0007(.039)[.039]	.9976(.038)	.9986(.038)[.038]	.9986(.038)[.038]
λ	.1848(.063)	.1999(.063)[.062]	.1999(.063)[.062]	.1666(.077)	.1885(.075)[.075]	.1887(.076)[.074]
ρ	.1112(.152)	.1868(.146)[.148]	.1867(.146)[.146]	.1101(.147)	.1889(.141)[.150]	.1889(.141)[.147]
σ_v^2	.7394(.083)	.9829(.110)[.107]	–	.7390(.082)	.9828(.109)[.106]	–
β_1	.9981(.038)	.9989(.038)[.039]	.9989(.038)[.038]	.9980(.038)	.9989(.038)[.038]	.9989(.038)[.038]
λ	.1849(.061)	.1998(.061)[.062]	.1999(.061)[.060]	.1689(.076)	.1909(.074)[.074]	.1909(.074)[.072]
ρ	.1179(.149)	.1933(.143)[.148]	.1932(.144)[.140]	.1121(.148)	.1915(.143)[.150]	.1913(.143)[.143]
σ_v^2	.7358(.172)	.9780(.228)[.215]	–	.7420(.176)	.9867(.234)[.218]	–
β_1	.9981(.038)	.9990(.038)[.039]	.9990(.038)[.038]	.9980(.038)	.9990(.037)[.038]	.9990(.037)[.038]
λ	.1825(.061)	.1976(.061)[.062]	.1976(.061)[.061]	.1688(.078)	.1907(.076)[.074]	.1908(.076)[.073]
ρ	.1165(.150)	.1919(.144)[.148]	.1917(.144)[.143]	.1104(.150)	.1894(.145)[.150]	.1894(.145)[.146]
σ_v^2	.7421(.128)	.9864(.169)[.161]	–	.7380(.129)	.9814(.171)[.161]	–
	$n = 100$; error = 1, 2, 3, for the three panels below					
β_1	1.0010(.027)	1.0011(.026)[.027]	1.0011(.026)[.027]	1.0015(.029)	1.0009(.029)[.029]	1.0009(.029)[.028]
λ	.1922(.043)	.1993(.043)[.042]	.1994(.043)[.042]	.1842(.055)	.1960(.055)[.054]	.1961(.055)[.053]
ρ	.1565(.099)	.1906(.096)[.100]	.1906(.096)[.099]	.1626(.104)	.1954(.101)[.099]	.1954(.101)[.098]
σ_v^2	.7617(.060)	.9942(.078)[.076]	–	.7604(.058)	.9928(.076)[.076]	–
β_1	.9993(.028)	.9994(.028)[.027]	.9994(.028)[.027]	1.0015(.029)	1.0009(.029)[.029]	1.0009(.029)[.028]
λ	.1923(.042)	.1994(.042)[.042]	.1994(.042)[.042]	.1829(.055)	.1948(.054)[.054]	.1948(.054)[.053]
ρ	.1623(.102)	.1962(.099)[.099]	.1962(.099)[.096]	.1588(.100)	.1917(.097)[.099]	.1916(.097)[.097]
σ_v^2	.7624(.128)	.9951(.167)[.160]	–	.7674(.128)	1.0019(.167)[.161]	–
β_1	.9983(.027)	.9984(.027)[.027]	.9984(.027)[.027]	1.0000(.028)	.9994(.028)[.029]	.9994(.028)[.028]
λ	.1937(.043)	.2009(.043)[.042]	.2009(.043)[.042]	.1831(.056)	.1950(.055)[.054]	.1950(.055)[.053]
ρ	.1621(.100)	.1961(.097)[.099]	.1961(.097)[.098]	.1599(.098)	.1928(.095)[.099]	.1929(.096)[.097]
σ_v^2	.7625(.092)	.9951(.120)[.118]	–	.7636(.091)	.9970(.118)[.118]	–
	$n = 200$; error = 1, 2, 3, for the three panels below					
β_1	1.0002(.019)	1.0001(.019)[.019]	1.0001(.019)[.019]	1.0002(.020)	1.0001(.020)[.020]	1.0001(.020)[.019]
λ	.1964(.028)	.1998(.028)[.029]	.1998(.028)[.029]	.1856(.049)	.1955(.049)[.048]	.1955(.049)[.048]
ρ	.1805(.071)	.1947(.069)[.068]	.1948(.069)[.068]	.1829(.069)	.1970(.068)[.068]	.1970(.068)[.068]
σ_v^2	.7703(.042)	.9958(.054)[.053]	–	.7708(.040)	.9966(.052)[.053]	–
β_1	.9997(.019)	.9996(.019)[.019]	.9996(.019)[.019]	1.0001(.020)	.9999(.020)[.020]	.9999(.020)[.019]
λ	.1969(.029)	.2003(.029)[.029]	.2003(.029)[.028]	.1851(.049)	.1950(.049)[.048]	.1950(.049)[.048]
ρ	.1850(.069)	.1991(.067)[.068]	.1991(.067)[.067]	.1864(.068)	.2004(.066)[.068]	.2004(.066)[.067]
σ_v^2	.7679(.089)	.9927(.115)[.114]	–	.7701(.091)	.9956(.118)[.114]	–
β_1	1.0007(.019)	1.0006(.019)[.019]	1.0006(.019)[.019]	1.0002(.019)	1.0000(.019)[.020]	1.0000(.019)[.020]
λ	.1968(.028)	.2002(.028)[.029]	.2002(.028)[.029]	.1861(.049)	.1960(.048)[.048]	.1960(.048)[.048]
ρ	.1840(.069)	.1981(.067)[.068]	.1981(.067)[.067]	.1832(.070)	.1973(.068)[.068]	.1973(.068)[.067]
σ_v^2	.7688(.063)	.9939(.082)[.083]	–	.7736(.066)	1.0002(.085)[.085]	–
	$n = 400$; error = 1, 2, 3, for the three panels below					
β_1	1.0003(.014)	1.0003(.014)[.013]	1.0003(.014)[.013]	1.0003(.013)	1.0003(.013)[.013]	1.0003(.013)[.013]
λ	.1985(.019)	.2001(.019)[.019]	.2001(.019)[.019]	.1875(.041)	.1949(.040)[.042]	.1949(.040)[.042]
ρ	.1936(.049)	.1982(.048)[.047]	.1982(.048)[.047]	.1953(.049)	.1999(.048)[.047]	.1999(.048)[.047]
σ_v^2	.7738(.028)	.9966(.036)[.038]	–	.7734(.029)	.9961(.037)[.038]	–
β_1	1.0001(.013)	1.0000(.013)[.013]	1.0000(.013)[.013]	1.0007(.013)	1.0007(.013)[.013]	1.0007(.013)[.013]
λ	.1985(.019)	.2001(.019)[.020]	.2001(.019)[.019]	.1899(.041)	.1972(.041)[.042]	.1972(.041)[.042]
ρ	.1937(.048)	.1983(.047)[.048]	.1983(.047)[.047]	.1922(.048)	.1969(.047)[.047]	.1969(.047)[.047]
σ_v^2	.7782(.063)	1.0023(.081)[.082]	–	.7767(.062)	1.0004(.080)[.082]	–
β_1	1.0001(.013)	1.0001(.013)[.013]	1.0001(.013)[.013]	.9999(.013)	.9999(.013)[.013]	.9999(.013)[.013]
λ	.1972(.020)	.1988(.020)[.020]	.1987(.020)[.019]	.1921(.042)	.1994(.041)[.042]	.1994(.041)[.042]
ρ	.1944(.050)	.1990(.049)[.047]	.1990(.049)[.047]	.1924(.049)	.1970(.048)[.047]	.1970(.048)[.047]
σ_v^2	.7743(.049)	.9973(.063)[.060]	–	.7729(.046)	.9955(.059)[.060]	–

Note: error = 1(normal), 2(normal mixture), 3(chi-square); X_t values are generated from $N(0, 2^2)$.

Table 1b. Empirical mean(sd)[$\hat{sd}$] of the estimators for FE-SPD model with `SL-SE` effects 10% random missing, **homoskedasticity**, $(\beta_1, \lambda, \rho, \sigma_v^2) = (1, 0.2, 0.2, 1)$, **T=10**.

	`W=Rook, M=Queen`			`W=Group-I, M=Queen`		
	QMLE	AQSE	RAQSE	QMLE	AQSE	RAQSE
	$n = 50$; error = 1, 2, 3, for the three panels below					
β_1	1.0019(.026)	1.0008(.026)[.026]	1.0008(.026)[.026]	.9993(.025)	.9989(.025)[.026]	.9989(.025)[.026]
λ	.1820(.039)	.1976(.039)[.040]	.1976(.039)[.040]	.1780(.046)	.1971(.045)[.045]	.1971(.046)[.045]
ρ	.1239(.093)	.1974(.091)[.091]	.1974(.092)[.091]	.1210(.093)	.1927(.092)[.091]	.1927(.092)[.091]
σ_v^2	.8641(.062)	.9930(.071)[.071]	–	.8641(.063)	.9936(.072)[.071]	–
β_1	1.0006(.026)	.9995(.026)[.026]	.9995(.026)[.026]	.9986(.025)	.9982(.025)[.026]	.9982(.025)[.025]
λ	.1849(.039)	.2004(.039)[.040]	.2004(.039)[.039]	.1779(.047)	.1968(.046)[.045]	.1969(.047)[.045]
ρ	.1203(.093)	.1941(.091)[.092]	.1940(.091)[.089]	.1235(.093)	.1951(.091)[.091]	.1950(.091)[.089]
σ_v^2	.8625(.144)	.9912(.166)[.156]	–	.8641(.138)	.9935(.158)[.156]	–
β_1	1.0019(.026)	1.0008(.026)[.026]	1.0008(.026)[.026]	1.0003(.026)	.9999(.026)[.026]	.9999(.026)[.026]
λ	.1819(.040)	.1976(.040)[.040]	.1976(.040)[.040]	.1771(.046)	.1960(.045)[.045]	.1961(.045)[.045]
ρ	.1194(.094)	.1931(.093)[.091]	.1931(.093)[.090]	.1211(.093)	.1928(.091)[.091]	.1928(.092)[.090]
σ_v^2	.8667(.105)	.9962(.121)[.113]	–	.8615(.101)	.9906(.116)[.113]	–
	$n = 100$; error = 1, 2, 3, for the three panels below					
β_1	1.0001(.018)	.9997(.018)[.018]	.9997(.018)[.018]	1.0004(.018)	1.0004(.018)[.017]	1.0004(.018)[.017]
λ	.1924(.027)	.1993(.027)[.027]	.1993(.027)[.027]	.1817(.040)	.1963(.039)[.039]	.1964(.039)[.039]
ρ	.1600(.063)	.1952(.062)[.063]	.1952(.062)[.063]	.1638(.064)	.1988(.063)[.063]	.1989(.063)[.063]
σ_v^2	.8792(.044)	.9986(.050)[.050]	–	.8787(.046)	.9981(.052)[.050]	–
β_1	1.0005(.018)	1.0000(.018)[.018]	1.0000(.018)[.018]	1.0001(.018)	1.0000(.018)[.017]	1.0000(.018)[.017]
λ	.1932(.027)	.2001(.027)[.027]	.2000(.027)[.027]	.1838(.040)	.1983(.040)[.039]	.1983(.040)[.039]
ρ	.1634(.062)	.1985(.061)[.063]	.1985(.061)[.062]	.1601(.063)	.1952(.062)[.063]	.1952(.062)[.062]
σ_v^2	.8773(.102)	.9964(.116)[.112]	–	.8780(.101)	.9973(.115)[.113]	–
β_1	1.0005(.018)	1.0001(.018)[.018]	1.0001(.018)[.018]	.9998(.018)	.9998(.018)[.017]	.9998(.018)[.017]
λ	.1923(.027)	.1992(.027)[.027]	.1992(.027)[.027]	.1834(.041)	.1979(.040)[.039]	.1979(.040)[.039]
ρ	.1609(.064)	.1961(.063)[.063]	.1961(.063)[.063]	.1618(.064)	.1969(.063)[.063]	.1969(.063)[.062]
σ_v^2	.8782(.073)	.9975(.083)[.082]	–	.8763(.072)	.9954(.082)[.082]	–
	$n = 200$; error = 1, 2, 3, for the three panels below					
β_1	1.0004(.013)	1.0001(.013)[.013]	1.0001(.013)[.013]	.9996(.013)	.9996(.013)[.012]	.9996(.013)[.012]
λ	.1961(.018)	.1994(.018)[.019]	.1994(.018)[.019]	.1883(.033)	.1986(.033)[.033]	.1986(.033)[.033]
ρ	.1823(.044)	.1985(.044)[.044]	.1986(.044)[.044]	.1834(.044)	.1997(.043)[.044]	.1997(.043)[.044]
σ_v^2	.8826(.030)	.9973(.034)[.035]	–	.8836(.031)	.9986(.035)[.035]	–
β_1	1.0002(.013)	.9999(.013)[.013]	.9999(.013)[.013]	.9998(.013)	.9997(.013)[.012]	.9997(.013)[.012]
λ	.1960(.018)	.1993(.018)[.019]	.1993(.018)[.019]	.1876(.033)	.1979(.033)[.033]	.1980(.033)[.033]
ρ	.1821(.043)	.1984(.043)[.044]	.1984(.043)[.044]	.1808(.045)	.1972(.044)[.044]	.1972(.044)[.044]
σ_v^2	.8820(.071)	.9967(.080)[.080]	–	.8825(.075)	.9973(.084)[.081]	–
β_1	.9996(.012)	.9993(.012)[.013]	.9993(.012)[.013]	1.0005(.012)	1.0005(.012)[.012]	1.0005(.012)[.012]
λ	.1968(.019)	.2000(.019)[.019]	.2000(.019)[.019]	.1878(.033)	.1981(.033)[.033]	.1982(.033)[.033]
ρ	.1818(.044)	.1980(.043)[.044]	.1980(.043)[.044]	.1834(.046)	.1997(.046)[.044]	.1997(.046)[.044]
σ_v^2	.8842(.053)	.9992(.060)[.058]	–	.8829(.051)	.9978(.057)[.058]	–
	$n = 400$; error = 1, 2, 3, for the three panels below					
β_1	1.0004(.009)	1.0003(.009)[.009]	1.0003(.009)[.009]	1.0001(.009)	1.0000(.009)[.009]	1.0000(.009)[.009]
λ	.1982(.014)	.1998(.014)[.014]	.1999(.014)[.014]	.1922(.027)	.1987(.027)[.026]	.1989(.027)[.026]
ρ	.1918(.033)	.1989(.032)[.031]	.1989(.032)[.031]	.1915(.033)	.1986(.033)[.031]	.1986(.033)[.031]
σ_v^2	.8854(.022)	.9982(.024)[.025]	–	.8853(.022)	.9982(.024)[.025]	–
β_1	.9998(.009)	.9997(.009)[.009]	.9997(.009)[.009]	1.0001(.009)	1.0000(.009)[.009]	1.0000(.009)[.009]
λ	.1983(.013)	.1999(.013)[.014]	.2000(.013)[.014]	.1913(.027)	.1978(.027)[.026]	.1981(.027)[.026]
ρ	.1931(.031)	.2001(.030)[.031]	.2001(.030)[.031]	.1905(.032)	.1976(.032)[.031]	.1976(.032)[.031]
σ_v^2	.8847(.050)	.9974(.056)[.057]	–	.8851(.051)	.9979(.057)[.057]	–
β_1	.9995(.009)	.9994(.009)[.009]	.9994(.009)[.009]	.9997(.009)	.9996(.009)[.009]	.9996(.009)[.009]
λ	.1978(.013)	.1994(.013)[.014]	.1996(.013)[.014]	.1926(.026)	.1991(.026)[.026]	.1993(.026)[.026]
ρ	.1931(.031)	.2002(.031)[.031]	.2002(.031)[.031]	.1907(.032)	.1978(.031)[.031]	.1978(.031)[.031]
σ_v^2	.8873(.038)	1.0004(.043)[.042]	–	.8881(.036)	1.0013(.041)[.042]	–

Note: error = 1(normal), 2(normal mixture), 3(chi-square); X_t values are generated from $N(0, 2^2)$.

Table 2a. Empirical mean$(sd)[\hat{sd}]$ of estimators for FE-SPD model with `SL-SD` effects 10% random missing, **homoskedasticity**, $(\beta_1, \beta_2, \lambda, \sigma_v^2) = (1, 0.5, 0.2, 1)$, **T=5**.

	W=Queen			W=Group-I		
	QMLE	AQSE	RAQSE	QMLE	AQSE	RAQSE
	$n = 50$; error = 1, 2, 3, for the three panels below					
β_1	1.0056(.041)	.9999(.041)[.041]	.9999(.041)[.041]	1.0130(.043)	1.0060(.043)[.043]	1.0060(.043)[.042]
β_2	.5898(.194)	.5147(.195)[.194]	.5146(.195)[.193]	.6567(.251)	.5636(.241)[.231]	.5637(.242)[.228]
λ	.1276(.125)	.1862(.125)[.123]	.1863(.125)[.122]	.1103(.131)	.1644(.125)[.119]	.1643(.126)[.117]
σ_v^2	.7390(.082)	.9779(.109)[.106]	–	.7398(.081)	.9793(.108)[.106]	–
β_1	1.0081(.040)	1.0024(.040)[.041]	1.0024(.040)[.040]	1.0101(.044)	1.0032(.044)[.042]	1.0031(.044)[.042]
β_2	.5909(.196)	.5158(.197)[.195]	.5156(.197)[.189]	.6426(.238)	.5504(.229)[.229]	.5497(.230)[.220]
λ	.1235(.122)	.1822(.122)[.124]	.1823(.123)[.120]	.1130(.124)	.1668(.119)[.119]	.1673(.120)[.114]
σ_v^2	.7410(.180)	.9806(.238)[.216]	–	.7413(.171)	.9812(.226)[.215]	–
β_1	1.0066(.041)	1.0009(.041)[.041]	1.0009(.041)[.041]	1.0113(.043)	1.0044(.042)[.042]	1.0044(.042)[.042]
β_2	.5904(.192)	.5151(.193)[.195]	.5151(.194)[.192]	.6385(.242)	.5463(.232)[.229]	.5458(.234)[.225]
λ	.1252(.123)	.1840(.123)[.124]	.1841(.123)[.122]	.1183(.126)	.1719(.121)[.118]	.1723(.122)[.116]
σ_v^2	.7436(.128)	.9841(.169)[.160]	–	.7411(.130)	.9809(.172)[.158]	–
	$n = 100$; error = 1, 2, 3, for the three panels below					
β_1	1.0047(.030)	1.0016(.030)[.031]	1.0016(.030)[.031]	1.0038(.027)	1.0012(.027)[.027]	1.0012(.027)[.027]
β_2	.5502(.135)	.5114(.135)[.133]	.5114(.135)[.138]	.5879(.180)	.5307(.175)[.173]	.5306(.176)[.177]
λ	.1621(.085)	.1908(.085)[.084]	.1908(.085)[.088]	.1466(.095)	.1808(.093)[.092]	.1808(.093)[.094]
σ_v^2	.7618(.059)	.9903(.076)[.075]	–	.7630(.058)	.9921(.075)[.075]	–
β_1	1.0053(.031)	1.0021(.031)[.031]	1.0021(.031)[.031]	1.0043(.027)	1.0016(.027)[.027]	1.0016(.027)[.027]
β_2	.5551(.130)	.5163(.130)[.133]	.5162(.130)[.136]	.5956(.189)	.5384(.183)[.173]	.5385(.183)[.175]
λ	.1585(.084)	.1872(.084)[.084]	.1872(.084)[.087]	.1433(.100)	.1775(.097)[.092]	.1774(.097)[.093]
σ_v^2	.7675(.129)	.9977(.168)[.159]	–	.7644(.129)	.9940(.167)[.159]	–
β_1	1.0032(.030)	1.0001(.030)[.030]	1.0001(.030)[.031]	1.0044(.027)	1.0017(.027)[.027]	1.0017(.027)[.027]
β_2	.5535(.136)	.5149(.136)[.133]	.5150(.136)[.136]	.5859(.180)	.5285(.175)[.173]	.5283(.175)[.176]
λ	.1598(.086)	.1884(.085)[.084]	.1883(.085)[.087]	.1465(.096)	.1808(.093)[.092]	.1810(.093)[.094]
σ_v^2	.7616(.091)	.9900(.119)[.116]	–	.7676(.095)	.9981(.123)[.118]	–
	$n = 200$; error = 1, 2, 3, for the three panels below					
β_1	1.0020(.021)	1.0006(.021)[.020]	1.0006(.021)[.021]	1.0027(.020)	1.0011(.020)[.020]	1.0011(.020)[.020]
β_2	.5244(.096)	.5056(.096)[.094]	.5056(.096)[.097]	.5722(.170)	.5257(.165)[.157]	.5257(.165)[.160]
λ	.1824(.058)	.1962(.059)[.057]	.1962(.059)[.060]	.1597(.083)	.1858(.081)[.079]	.1859(.081)[.080]
σ_v^2	.7713(.041)	.9948(.053)[.053]	–	.7726(.041)	.9949(.053)[.053]	–
β_1	1.0013(.020)	.9999(.020)[.020]	.9999(.020)[.021]	1.0026(.020)	1.0011(.020)[.020]	1.0011(.020)[.020]
β_2	.5269(.093)	.5082(.093)[.093]	.5081(.093)[.097]	.5687(.160)	.5224(.156)[.157]	.5222(.156)[.158]
λ	.1808(.057)	.1945(.057)[.057]	.1946(.057)[.060]	.1605(.080)	.1866(.078)[.079]	.1867(.078)[.079]
σ_v^2	.7726(.091)	.9964(.118)[.114]	–	.7740(.089)	.9967(.114)[.113]	–
β_1	1.0017(.020)	1.0002(.020)[.020]	1.0002(.020)[.021]	1.0033(.019)	1.0018(.019)[.020]	1.0018(.019)[.020]
β_2	.5248(.094)	.5060(.094)[.094]	.5060(.094)[.097]	.5759(.164)	.5293(.160)[.158]	.5293(.160)[.159]
λ	.1826(.056)	.1963(.056)[.057]	.1963(.056)[.060]	.1564(.082)	.1827(.080)[.079]	.1827(.080)[.080]
σ_v^2	.7741(.065)	.9984(.083)[.084]	–	.7726(.067)	.9950(.087)[.083]	–
	$n = 400$; error = 1, 2, 3, for the three panels below					
β_1	1.0012(.014)	1.0005(.014)[.014]	1.0005(.014)[.014]	1.0012(.014)	1.0005(.014)[.014]	1.0005(.014)[.014]
β_2	.5120(.065)	.5028(.065)[.064]	.5028(.065)[.067]	.5526(.142)	.5197(.139)[.136]	.5196(.139)[.138]
λ	.1909(.041)	.1980(.041)[.041]	.1980(.041)[.043]	.1702(.069)	.1890(.067)[.067]	.1891(.067)[.068]
σ_v^2	.7761(.030)	.9992(.039)[.038]	–	.7759(.029)	.9989(.038)[.038]	–
β_1	1.0011(.014)	1.0004(.014)[.014]	1.0004(.014)[.014]	1.0010(.014)	1.0003(.014)[.014]	1.0003(.014)[.014]
β_2	.5116(.064)	.5024(.064)[.064]	.5024(.064)[.067]	.5553(.141)	.5224(.139)[.136]	.5225(.139)[.138]
λ	.1904(.042)	.1975(.042)[.041]	.1975(.042)[.043]	.1682(.070)	.1870(.069)[.067]	.1870(.069)[.068]
σ_v^2	.7730(.062)	.9952(.080)[.081]	–	.7756(.064)	.9986(.082)[.082]	–
β_1	1.0010(.014)	1.0003(.014)[.014]	1.0003(.014)[.014]	1.0012(.014)	1.0004(.014)[.014]	1.0004(.014)[.014]
β_2	.5101(.063)	.5009(.063)[.064]	.5009(.063)[.067]	.5551(.140)	.5223(.138)[.136]	.5222(.138)[.138]
λ	.1913(.040)	.1984(.040)[.041]	.1984(.040)[.043]	.1692(.068)	.1879(.066)[.067]	.1880(.067)[.068]
σ_v^2	.7764(.047)	.9996(.061)[.060]	–	.7763(.048)	.9996(.061)[.060]	–

Note: error = 1(normal), 2(normal mixture), 3(chi-square); X_t values are generated from $N(0, 2^2)$.

Table 2b. Empirical mean(sd)[$\hat{sd}$] of estimators for FE-SPD model with `SL-SD` effects 10% random missing, **homoskedasticity**, $(\beta_1, \beta_2, \lambda, \sigma_v^2) = (1, 0.5, 0.2, 1)$, **T=10**.

	W=Queen			W=Group-I		
	QMLE	AQSE	RAQSE	QMLE	AQSE	RAQSE
	$n = 50$; error = 1, 2, 3, for the three panels below					
β_1	1.0072(.026)	1.0014(.026)[.027]	1.0014(.026)[.027]	1.0083(.029)	1.0016(.029)[.028]	1.0016(.029)[.028]
β_2	.5823(.122)	.5074(.123)[.124]	.5073(.123)[.125]	.6090(.159)	.5238(.152)[.149]	.5237(.152)[.149]
λ	.1368(.078)	.1935(.078)[.080]	.1935(.079)[.081]	.1357(.081)	.1861(.078)[.075]	.1861(.078)[.075]
σ_v^2	.8604(.060)	.9889(.069)[.070]	–	.8607(.062)	.9894(.071)[.071]	–
β_1	1.0069(.027)	1.0011(.027)[.027]	1.0011(.027)[.027]	1.0099(.028)	1.0032(.027)[.028]	1.0032(.027)[.028]
β_2	.5850(.122)	.5104(.123)[.124]	.5104(.123)[.124]	.6074(.156)	.5223(.150)[.149]	.5224(.150)[.148]
λ	.1354(.080)	.1918(.080)[.079]	.1919(.080)[.079]	.1355(.079)	.1858(.075)[.075]	.1857(.075)[.073]
σ_v^2	.8674(.140)	.9970(.161)[.158]	–	.8686(.143)	.9984(.164)[.159]	–
β_1	1.0068(.027)	1.0010(.027)[.027]	1.0010(.027)[.027]	1.0069(.028)	1.0003(.028)[.028]	1.0003(.028)[.028]
β_2	.5827(.124)	.5082(.125)[.124]	.5079(.125)[.125]	.6057(.153)	.5208(.147)[.148]	.5209(.147)[.148]
λ	.1364(.080)	.1929(.080)[.079]	.1931(.080)[.080]	.1370(.078)	.1872(.074)[.074]	.1871(.074)[.074]
σ_v^2	.8607(.105)	.9893(.120)[.113]	–	.8575(.101)	.9857(.116)[.112]	–
	$n = 100$; error = 1, 2, 3, for the three panels below					
β_1	1.0024(.019)	.9997(.019)[.018]	.9997(.019)[.018]	1.0033(.019)	.9999(.019)[.019]	.9999(.019)[.019]
β_2	.5381(.083)	.5025(.083)[.083]	.5025(.083)[.086]	.5815(.135)	.5142(.130)[.127]	.5140(.131)[.128]
λ	.1697(.055)	.1982(.055)[.055]	.1982(.055)[.057]	.1528(.069)	.1927(.067)[.065]	.1928(.067)[.066]
σ_v^2	.8765(.043)	.9956(.049)[.050]	–	.8766(.044)	.9958(.050)[.050]	–
β_1	1.0036(.018)	1.0008(.018)[.018]	1.0008(.018)[.018]	1.0036(.019)	1.0002(.019)[.019]	1.0002(.019)[.019]
β_2	.5403(.083)	.5046(.083)[.083]	.5045(.083)[.086]	.5882(.133)	.5207(.128)[.128]	.5206(.128)[.128]
λ	.1678(.056)	.1963(.056)[.055]	.1963(.056)[.057]	.1480(.069)	.1881(.066)[.066]	.1882(.066)[.066]
σ_v^2	.8771(.101)	.9963(.115)[.113]	–	.8752(.101)	.9942(.114)[.113]	–
β_1	1.0029(.018)	1.0001(.018)[.018]	1.0001(.018)[.018]	1.0045(.019)	1.0011(.019)[.019]	1.0011(.019)[.019]
β_2	.5381(.083)	.5024(.083)[.083]	.5024(.083)[.086]	.5853(.131)	.5179(.127)[.127]	.5179(.126)[.128]
λ	.1686(.055)	.1971(.055)[.055]	.1971(.055)[.057]	.1488(.068)	.1889(.066)[.066]	.1889(.066)[.066]
σ_v^2	.8755(.074)	.9944(.084)[.081]	–	.8755(.075)	.9945(.086)[.081]	–
	$n = 200$; error = 1, 2, 3, for the three panels below					
β_1	1.0017(.013)	1.0003(.013)[.013]	1.0003(.013)[.013]	1.0023(.013)	1.0007(.013)[.013]	1.0007(.013)[.013]
β_2	.5218(.061)	.5034(.061)[.060]	.5034(.061)[.062]	.5592(.109)	.5124(.106)[.104]	.5124(.106)[.106]
λ	.1827(.039)	.1972(.039)[.039]	.1972(.039)[.040]	.1653(.055)	.1926(.054)[.054]	.1925(.054)[.055]
σ_v^2	.8835(.032)	.9979(.036)[.035]	–	.8837(.032)	.9983(.036)[.035]	–
β_1	1.0018(.013)	1.0005(.013)[.013]	1.0005(.013)[.013]	1.0021(.013)	1.0005(.013)[.013]	1.0005(.013)[.013]
β_2	.5202(.060)	.5017(.060)[.060]	.5018(.060)[.062]	.5580(.106)	.5112(.103)[.104]	.5110(.103)[.106]
λ	.1838(.040)	.1982(.040)[.039]	.1982(.040)[.040]	.1661(.055)	.1933(.053)[.054]	.1934(.053)[.055]
σ_v^2	.8816(.073)	.9958(.082)[.080]	–	.8839(.072)	.9985(.082)[.081]	–
β_1	1.0017(.013)	1.0003(.013)[.013]	1.0003(.013)[.013]	1.0019(.013)	1.0004(.013)[.013]	1.0004(.013)[.013]
β_2	.5220(.060)	.5035(.060)[.060]	.5035(.060)[.062]	.5581(.107)	.5113(.104)[.104]	.5112(.104)[.106]
λ	.1834(.039)	.1978(.039)[.039]	.1978(.039)[.040]	.1670(.054)	.1942(.053)[.054]	.1943(.053)[.055]
σ_v^2	.8837(.052)	.9981(.059)[.058]	–	.8844(.051)	.9991(.057)[.058]	–
	$n = 400$; error = 1, 2, 3, for the three panels below					
β_1	1.0007(.009)	1.0000(.009)[.009]	1.0000(.009)[.010]	1.0012(.009)	1.0004(.009)[.009]	1.0004(.009)[.009]
β_2	.5101(.043)	.5004(.043)[.044]	.5004(.043)[.046]	.5476(.100)	.5120(.098)[.091]	.5121(.098)[.093]
λ	.1918(.026)	.1990(.026)[.028]	.1989(.026)[.029]	.1732(.049)	.1935(.048)[.046]	.1934(.048)[.047]
σ_v^2	.8851(.022)	.9980(.025)[.025]	–	.8867(.023)	.9998(.026)[.025]	–
β_1	1.0006(.009)	.9998(.009)[.009]	.9998(.009)[.010]	1.0007(.009)	.9999(.009)[.009]	.9999(.009)[.009]
β_2	.5086(.043)	.4989(.043)[.044]	.5018(.060)[.062]	.5444(.093)	.5088(.091)[.091]	.5088(.091)[.092]
λ	.1934(.026)	.2006(.026)[.028]	.1982(.040)[.040]	.1753(.047)	.1955(.046)[.046]	.1956(.046)[.047]
σ_v^2	.8867(.052)	.9998(.059)[.058]	–	.8853(.051)	.9983(.058)[.057]	–
β_1	1.0009(.009)	1.0002(.009)[.009]	1.0003(.013)[.013]	1.0012(.009)	1.0003(.009)[.009]	1.0003(.009)[.009]
β_2	.5127(.042)	.5030(.042)[.044]	.5035(.060)[.062]	.5489(.094)	.5131(.092)[.092]	.5131(.092)[.093]
λ	.1913(.026)	.1984(.026)[.028]	.1978(.039)[.040]	.1726(.048)	.1930(.047)[.046]	.1930(.047)[.047]
σ_v^2	.8840(.037)	.9967(.042)[.041]	–	.8864(.036)	.9996(.041)[.042]	–

Note: error = 1(normal), 2(normal mixture), 3(chi-square); X_t values are generated from $N(0, 2^2)$.

Table 3a. Empirical mean(sd)[$\hat{sd}$] of estimators for FE-SPD model with `SL-SE` effects
10% random missing, **heteroskedasticity**, $(\beta_1, \lambda, \rho, \sigma_v^2) = (1, 0.2, 0.2, 1)$, **T=5**.

	`W=M=Group-II`			`W=Group-II, M=Queen`		
	QMLE	AQSE	RAQSE	QMLE	AQSE	RAQSE
	$n = 50$; error = 1, 2, 3, for the three panels below					
β_1	1.0001(.042)	1.0001(.042)[.041]	.9993(.042)[.042]	.9974(.039)	.9978(.039)[.040]	.9979(.039)[.039]
λ	.1878(.070)	.1944(.068)[.098]	.1973(.080)[.080]	.1311(.086)	.1735(.083)[.106]	.1894(.089)[.090]
ρ	-.0161(.209)	.0905(.177)[.199]	.1016(.272)[.247]	.1077(.147)	.1845(.141)[.145]	.1889(.141)[.142]
σ_v^2	.7664(.102)	1.0237(.136)[.150]	–	.7717(.102)	1.0264(.136)[.150]	–
β_1	1.0012(.042)	1.0013(.042)[.040]	1.0005(.043)[.041]	.9980(.039)	.9985(.039)[.040]	.9985(.039)[.038]
λ	.1873(.072)	.1938(.069)[.098]	.1956(.082)[.080]	.1343(.083)	.1757(.080)[.105]	.1915(.085)[.088]
ρ	-.0008(.198)	.1036(.168)[.198]	.1235(.248)[.232]	.1086(.146)	.1853(.140)[.146]	.1894(.140)[.135]
σ_v^2	.7606(.217)	1.0154(.290)[.274]	–	.7695(.221)	1.0234(.294)[.277]	–
β_1	.9986(.041)	.9986(.042)[.041]	.9978(.042)[.041]	.9980(.040)	.9984(.040)[.040]	.9985(.040)[.039]
λ	.1843(.070)	.1911(.067)[.099]	.1928(.081)[.080]	.1316(.084)	.1737(.081)[.106]	.1896(.087)[.089]
ρ	-.0072(.205)	.0980(.174)[.198]	.1144(.260)[.238]	.1113(.144)	.1878(.138)[.145]	.1919(.138)[.139]
σ_v^2	.7727(.158)	1.0319(.211)[.212]	–	.7744(.161)	1.0300(.214)[.210]	–
	$n = 100$; error = 1, 2, 3, for the three panels below					
β_1	1.0005(.028)	1.0005(.028)[.028]	1.0002(.028)[.028]	1.0003(.025)	1.0005(.025)[.026]	1.0006(.025)[.026]
λ	.1927(.049)	.1954(.048)[.064]	.1992(.056)[.054]	.1733(.045)	.1849(.045)[.054]	.1960(.046)[.048]
ρ	.0883(.131)	.1304(.120)[.127]	.1573(.167)[.160]	.1665(.095)	.1995(.093)[.095]	.1980(.092)[.094]
σ_v^2	.7572(.071)	.9925(.093)[.105]	–	.7706(.070)	.9985(.090)[.102]	–
β_1	1.0004(.029)	1.0004(.029)[.028]	1.0001(.029)[.028]	.9994(.026)	.9996(.026)[.026]	.9997(.026)[.026]
λ	.1921(.049)	.1948(.048)[.063]	.1985(.055)[.053]	.1736(.045)	.1851(.045)[.054]	.1960(.046)[.047]
ρ	.0884(.129)	.1305(.118)[.128]	.1578(.165)[.157]	.1668(.093)	.1997(.090)[.095]	.1984(.090)[.090]
σ_v^2	.7554(.155)	.9901(.203)[.199]	–	.7648(.153)	.9910(.198)[.194]	–
β_1	.9997(.029)	.9997(.029)[.028]	.9995(.029)[.028]	.9996(.026)	.9998(.026)[.026]	.9999(.026)[.026]
λ	.1914(.049)	.1941(.048)[.063]	.1979(.055)[.054]	.1739(.045)	.1854(.045)[.054]	.1964(.047)[.047]
ρ	.0877(.130)	.1299(.119)[.128]	.1566(.167)[.159]	.1656(.097)	.1985(.095)[.095]	.1971(.094)[.093]
σ_v^2	.7614(.115)	.9979(.150)[.152]	–	.7646(.112)	.9907(.145)[.146]	–
	$n = 200$; error = 1, 2, 3, for the three panels below					
β_1	.9991(.019)	.9991(.019)[.019]	.9990(.019)[.019]	.9987(.020)	.9988(.020)[.019]	.9989(.020)[.019]
λ	.1950(.034)	.1962(.034)[.044]	.2000(.040)[.040]	.1784(.035)	.1853(.034)[.042]	.1980(.036)[.037]
ρ	.1272(.086)	.1441(.082)[.084]	.1763(.112)[.107]	.1862(.071)	.2006(.070)[.068]	.1996(.069)[.069]
σ_v^2	.7657(.050)	.9907(.065)[.073]	–	.7647(.051)	.9883(.065)[.073]	–
β_1	.9996(.019)	.9996(.019)[.019]	.9995(.019)[.019]	.9992(.019)	.9993(.019)[.019]	.9994(.019)[.019]
λ	.1939(.035)	.1951(.035)[.044]	.1984(.041)[.040]	.1773(.035)	.1842(.034)[.042]	.1968(.036)[.037]
ρ	.1345(.083)	.1511(.079)[.083]	.1857(.106)[.105]	.1816(.071)	.1960(.069)[.068]	.1952(.069)[.068]
σ_v^2	.7704(.111)	.9967(.144)[.143]	–	.7660(.110)	.9899(.142)[.141]	–
β_1	.9996(.019)	.9996(.019)[.019]	.9996(.019)[.019]	.9987(.019)	.9988(.019)[.019]	.9990(.019)[.019]
λ	.1947(.035)	.1959(.035)[.044]	.1996(.041)[.040]	.1782(.035)	.1851(.034)[.042]	.1978(.036)[.037]
ρ	.1290(.085)	.1459(.081)[.084]	.1787(.110)[.106]	.1840(.070)	.1984(.069)[.068]	.1975(.068)[.068]
σ_v^2	.7646(.080)	.9893(.104)[.106]	–	.7664(.081)	.9905(.104)[.106]	–
	$n = 400$; error = 1, 2, 3, for the three panels below					
β_1	.9999(.014)	.9999(.014)[.013]	.9999(.014)[.014]	1.0000(.013)	1.0000(.013)[.013]	1.0000(.013)[.013]
λ	.1966(.026)	.1970(.026)[.031]	.1998(.030)[.030]	.1862(.024)	.1894(.024)[.028]	.2002(.025)[.026]
ρ	.1491(.060)	.1550(.058)[.058]	.1892(.075)[.074]	.1945(.047)	.1990(.046)[.047]	.1991(.046)[.047]
σ_v^2	.7849(.034)	1.0110(.044)[.052]	–	.7839(.034)	1.0096(.044)[.052]	–
β_1	.9998(.014)	.9998(.014)[.013]	.9998(.014)[.014]	.9994(.013)	.9994(.013)[.013]	.9994(.013)[.013]
λ	.1968(.027)	.1972(.027)[.031]	.1998(.031)[.030]	.1835(.025)	.1866(.025)[.028]	.1973(.026)[.026]
ρ	.1509(.061)	.1568(.059)[.058]	.1914(.075)[.074]	.1959(.048)	.2003(.047)[.047]	.2004(.047)[.047]
σ_v^2	.7878(.079)	1.0148(.102)[.103]	–	.7842(.080)	1.0100(.103)[.103]	–
β_1	1.0000(.013)	1.0000(.013)[.013]	1.0000(.014)[.014]	1.0005(.014)	1.0005(.014)[.013]	1.0005(.014)[.013]
λ	.1949(.027)	.1953(.027)[.031]	.1980(.031)[.030]	.1861(.024)	.1893(.024)[.028]	.2001(.025)[.026]
ρ	.1500(.059)	.1559(.057)[.058]	.1904(.073)[.074]	.1955(.047)	.1999(.046)[.047]	.2001(.046)[.047]
σ_v^2	.7869(.059)	1.0136(.076)[.078]	–	.7854(.059)	1.0116(.076)[.078]	–

Note: error = 1(normal), 2(normal mixture), 3(chi-square); X_t values are generated from $N(0, 2^2)$.

Table 3b. Empirical mean(sd)[$\hat{sd}$] of estimators for FE-SPD model with `SL-SE` effects 10% random missing, **heteroskedasticity**, $(\beta_1, \lambda, \rho, \sigma_v^2) = (1, 0.2, 0.2, 1)$, **T=10**.

	`W=M=Group-II`			`W=Group-II, M=Queen`		
	QMLE	AQSE	RAQSE	QMLE	AQSE	RAQSE
$n = 50$; error = 1, 2, 3, for the three panels below						
β_1	1.0015(.025)	1.0012(.025)[.025]	1.0009(.025)[.024]	1.0024(.026)	1.0007(.026)[.026]	.9999(.026)[.025]
λ	.1899(.039)	.1953(.038)[.054]	.1987(.043)[.042]	.1677(.036)	.1880(.035)[.047]	.1973(.037)[.037]
ρ	.0447(.119)	.1342(.106)[.117]	.1660(.146)[.139]	.1290(.092)	.2001(.091)[.090]	.1988(.091)[.089]
σ_v^2	.8647(.073)	.9940(.083)[.093]	–	.8737(.074)	1.0030(.085)[.094]	–
β_1	.9997(.025)	.9995(.025)[.025]	.9992(.025)[.024]	1.0015(.025)	.9998(.025)[.025]	.9989(.025)[.025]
λ	.1900(.039)	.1954(.037)[.054]	.1990(.042)[.042]	.1686(.035)	.1888(.034)[.047]	.1981(.036)[.037]
ρ	.0425(.118)	.1324(.105)[.119]	.1636(.145)[.137]	.1284(.092)	.1995(.090)[.090]	.1980(.089)[.087]
σ_v^2	.8715(.175)	1.0019(.201)[.192]	–	.8744(.173)	1.0037(.198)[.193]	–
β_1	1.0011(.025)	1.0009(.025)[.025]	1.0006(.025)[.024]	1.0021(.025)	1.0004(.025)[.026]	.9995(.025)[.025]
λ	.1898(.039)	.1952(.037)[.054]	.1987(.042)[.042]	.1677(.035)	.1881(.035)[.047]	.1974(.036)[.037]
ρ	.0461(.115)	.1355(.102)[.117]	.1682(.139)[.138]	.1280(.092)	.1992(.090)[.090]	.1981(.090)[.088]
σ_v^2	.8646(.125)	.9940(.144)[.141]	–	.8751(.123)	1.0047(.141)[.144]	–
$n = 100$; error = 1, 2, 3, for the three panels below						
β_1	.9994(.018)	.9995(.018)[.018]	.9996(.018)[.018]	.9996(.018)	.9994(.018)[.018]	.9992(.018)[.018]
λ	.1901(.034)	.1936(.034)[.043]	.1980(.040)[.038]	.1760(.034)	.1889(.033)[.038]	.2000(.035)[.035]
ρ	.1077(.082)	.1470(.077)[.080]	.1809(.106)[.101]	.1627(.066)	.1979(.065)[.064]	.1979(.065)[.064]
σ_v^2	.8663(.053)	.9860(.060)[.067]	–	.8717(.054)	.9931(.062)[.067]	–
β_1	.9991(.018)	.9992(.018)[.018]	.9992(.018)[.018]	1.0003(.019)	1.0001(.019)[.018]	.9999(.019)[.018]
λ	.1902(.033)	.1937(.032)[.043]	.1978(.038)[.038]	.1746(.034)	.1875(.033)[.038]	.1985(.035)[.035]
ρ	.1106(.078)	.1497(.073)[.080]	.1848(.101)[.099]	.1605(.066)	.1957(.065)[.064]	.1957(.065)[.063]
σ_v^2	.8659(.124)	.9855(.141)[.139]	–	.8736(.125)	.9953(.143)[.139]	–
β_1	.9999(.018)	1.0000(.018)[.018]	1.0000(.018)[.018]	.9997(.019)	.9995(.019)[.018]	.9993(.019)[.018]
λ	.1922(.033)	.1957(.032)[.043]	.2005(.037)[.038]	.1738(.033)	.1867(.033)[.038]	.1977(.035)[.035]
ρ	.1077(.079)	.1470(.074)[.080]	.1808(.102)[.100]	.1607(.066)	.1958(.065)[.064]	.1958(.064)[.063]
σ_v^2	.8630(.088)	.9822(.100)[.102]	–	.8760(.091)	.9981(.104)[.103]	–
$n = 200$; error = 1, 2, 3, for the three panels below						
β_1	.9999(.013)	1.0000(.013)[.013]	1.0000(.013)[.013]	1.0004(.013)	1.0001(.013)[.013]	.9997(.013)[.013]
λ	.1938(.026)	.1951(.026)[.030]	.1976(.030)[.029]	.1833(.022)	.1892(.022)[.026]	.1997(.023)[.023]
ρ	.1420(.057)	.1601(.055)[.055]	.1955(.071)[.069]	.1830(.045)	.1995(.044)[.044]	.1999(.044)[.044]
σ_v^2	.8932(.038)	1.0097(.043)[.048]	–	.8922(.036)	1.0082(.041)[.048]	–
β_1	.9999(.013)	1.0000(.013)[.013]	1.0000(.013)[.013]	1.0007(.012)	1.0005(.012)[.013]	1.0000(.012)[.013]
λ	.1958(.026)	.1972(.025)[.030]	.2000(.029)[.029]	.1821(.022)	.1880(.022)[.026]	.1985(.023)[.023]
ρ	.1399(.056)	.1581(.054)[.055]	.1926(.070)[.069]	.1798(.045)	.1964(.045)[.044]	.1968(.045)[.044]
σ_v^2	.8939(.089)	1.0105(.101)[.101]	–	.8912(.090)	1.0070(.102)[.100]	–
β_1	.9993(.013)	.9993(.013)[.013]	.9993(.013)[.013]	1.0006(.013)	1.0003(.013)[.013]	.9999(.013)[.013]
λ	.1955(.026)	.1969(.026)[.030]	.1998(.029)[.029]	.1830(.023)	.1890(.023)[.026]	.1996(.024)[.023]
ρ	.1379(.057)	.1561(.055)[.055]	.1903(.071)[.070]	.1784(.043)	.1949(.043)[.044]	.1953(.043)[.044]
σ_v^2	.8916(.063)	1.0080(.072)[.074]	–	.8956(.065)	1.0121(.074)[.074]	–
$n = 400$; error = 1, 2, 3, for the three panels below						
β_1	1.0003(.009)	1.0009(.009)[.009]	1.0003(.009)[.009]	.9997(.009)	.9997(.009)[.009]	.9994(.009)[.009]
λ	.1974(.016)	.2490(.018)[.019]	.2010(.018)[.018]	.1860(.016)	.1888(.016)[.018]	.1991(.016)[.017]
ρ	.1533(.037)	.1210(.028)[.037]	.1959(.046)[.047]	.1924(.032)	.1995(.031)[.031]	.2001(.031)[.032]
σ_v^2	.8923(.027)	1.0063(.030)[.034]	–	.8913(.028)	1.0049(.031)[.034]	–
β_1	.9995(.009)	1.0001(.009)[.009]	.9995(.009)[.009]	1.0003(.009)	1.0003(.009)[.009]	1.0000(.009)[.009]
λ	.1965(.017)	.2493(.019)[.019]	.1997(.019)[.018]	.1871(.016)	.1900(.016)[.018]	.2004(.017)[.017]
ρ	.1562(.038)	.1232(.029)[.037]	.1996(.048)[.047]	.1914(.031)	.1985(.030)[.031]	.1991(.030)[.031]
σ_v^2	.8933(.061)	1.0075(.069)[.072]	–	.8923(.063)	1.0060(.071)[.071]	–
β_1	.9998(.009)	1.0004(.009)[.009]	.9998(.009)[.009]	1.0003(.009)	1.0002(.009)[.009]	1.0000(.009)[.009]
λ	.1968(.017)	.2485(.018)[.019]	.2003(.019)[.018]	.1864(.016)	.1892(.016)[.018]	.1996(.016)[.017]
ρ	.1531(.038)	.1208(.029)[.037]	.1958(.047)[.047]	.1922(.031)	.1993(.031)[.031]	.1999(.031)[.032]
σ_v^2	.8960(.045)	1.0105(.051)[.053]	–	.8950(.047)	1.0090(.053)[.053]	–

Note: error = 1(normal), 2(normal mixture), 3(chi-square); X_t values are generated from $N(0, 2^2)$.

Table 4. Empirical mean$(sd)[\hat{sd}]$ of estimators for FE-SPD model with `SL-SD` effects 10% random missing, **heteroskedasticity**, $(\beta_1, \beta_2, \lambda, \sigma_v^2) = (1, 0.2, 0.2, 1)$, `W=Group-II`.

	T=5			**T=10**		
	QMLE	AQSE	RAQSE	QMLE	AQSE	RAQSE
			$n = 50$; error = 1, 2, 3, for the three panels below			
β_1	1.0113(.045)	1.0054(.045)[.042]	1.0033(.045)[.044]	1.0134(.028)	1.0058(.028)[.027]	1.0024(.028)[.029]
β_2	.7304(.237)	.6103(.223)[.277]	.5672(.256)[.269]	.6962(.153)	.5798(.143)[.172]	.5278(.167)[.190]
λ	.0569(.146)	.1317(.137)[.152]	.1583(.159)[.165]	.0811(.089)	.1517(.084)[.096]	.1831(.099)[.114]
σ_v^2	.7704(.102)	1.0208(.135)[.148]	–	.8676(.074)	.9945(.085)[.093]	–
β_1	1.0124(.045)	1.0065(.044)[.042]	1.0045(.045)[.043]	1.0130(.028)	1.0054(.028)[.027]	1.0021(.028)[.028]
β_2	.7306(.230)	.6121(.216)[.276]	.5705(.247)[.261]	.6972(.151)	.5811(.141)[.172]	.5302(.166)[.187]
λ	.0572(.141)	.1310(.132)[.151]	.1566(.153)[.160]	.0805(.089)	.1508(.083)[.096]	.1816(.099)[.112]
σ_v^2	.7664(.224)	1.0156(.297)[.273]	–	.8662(.171)	.9930(.196)[.189]	–
β_1	1.0084(.044)	1.0025(.044)[.043]	1.0003(.044)[.043]	1.0127(.028)	1.0051(.028)[.027]	1.0017(.029)[.029]
β_2	.7193(.232)	.5998(.218)[.276]	.5557(.249)[.266]	.6969(.156)	.5804(.146)[.172]	.5285(.171)[.189]
λ	.0627(.143)	.1373(.134)[.152]	.1645(.155)[.164]	.0809(.092)	.1514(.086)[.096]	.1828(.102)[.113]
σ_v^2	.7771(.166)	1.0297(.219)[.211]	–	.8710(.122)	.9984(.140)[.142]	–
			$n = 100$; error = 1, 2, 3, for the three panels below			
β_1	1.0090(.032)	1.0056(.032)[.030]	1.0028(.032)[.032]	1.0080(.020)	1.0045(.020)[.019]	1.0011(.020)[.021]
β_2	.6379(.166)	.5817(.161)[.185]	.5351(.185)[.201]	.6174(.106)	.5644(.102)[.118]	.5130(.120)[.125]
λ	.1149(.095)	.1497(.092)[.103]	.1785(.109)[.119]	.1250(.063)	.1588(.061)[.066]	.1914(.073)[.077]
σ_v^2	.7609(.070)	.9942(.091)[.105]	–	.8659(.052)	.9843(.060)[.067]	–
β_1	1.0081(.031)	1.0047(.031)[.029]	1.0019(.031)[.032]	1.0080(.020)	1.0045(.020)[.019]	1.0012(.020)[.021]
β_2	.6336(.161)	.5778(.156)[.185]	.5313(.179)[.196]	.6179(.104)	.5650(.101)[.118]	.5141(.118)[.124]
λ	.1168(.092)	.1513(.089)[.102]	.1801(.105)[.116]	.1247(.063)	.1584(.061)[.067]	.1907(.072)[.076]
σ_v^2	.7653(.160)	.9998(.209)[.200]	–	.8667(.119)	.9853(.136)[.138]	–
β_1	1.0081(.031)	1.0048(.031)[.030]	1.0020(.031)[.032]	1.0077(.020)	1.0042(.020)[.019]	1.0009(.021)[.021]
β_2	.6370(.163)	.5810(.158)[.185]	.5348(.182)[.200]	.6186(.106)	.5656(.103)[.118]	.5150(.121)[.124]
λ	.1150(.095)	.1497(.092)[.103]	.1782(.108)[.118]	.1238(.064)	.1575(.062)[.067]	.1897(.075)[.076]
σ_v^2	.7629(.115)	.9968(.150)[.152]	–	.8691(.090)	.9880(.102)[.102]	–
			$n = 200$; error = 1, 2, 3, for the three panels below			
β_1	1.0067(.021)	1.0050(.021)[.021]	1.0017(.022)[.023]	1.0053(.015)	1.0036(.015)[.014]	1.0005(.015)[.015]
β_2	.5960(.117)	.5690(.116)[.129]	.5193(.133)[.146]	.5831(.077)	.5563(.076)[.082]	.5074(.086)[.093]
λ	.1412(.067)	.1573(.066)[.068]	.1869(.077)[.085]	.1485(.044)	.1651(.044)[.046]	.1953(.051)[.055]
σ_v^2	.7670(.051)	.9909(.066)[.073]	–	.8830(.039)	.9979(.044)[.047]	–
β_1	1.0056(.021)	1.0038(.021)[.021]	1.0005(.022)[.023]	1.0051(.014)	1.0034(.014)[.014]	1.0003(.015)[.015]
β_2	.5952(.113)	.5683(.111)[.129]	.5179(.127)[.144]	.5807(.075)	.5540(.073)[.083]	.5048(.083)[.093]
λ	.1438(.064)	.1598(.063)[.068]	.1898(.074)[.084]	.1495(.043)	.1660(.043)[.046]	.1964(.050)[.055]
σ_v^2	.7684(.110)	.9926(.143)[.141]	–	.8814(.088)	.9962(.099)[.099]	–
β_1	1.0062(.022)	1.0044(.021)[.021]	1.0011(.022)[.023]	1.0049(.015)	1.0032(.015)[.014]	1.0001(.015)[.015]
β_2	.5952(.116)	.5682(.114)[.129]	.5180(.131)[.145]	.5839(.077)	.5572(.076)[.083]	.5081(.086)[.093]
λ	.1434(.066)	.1595(.065)[.068]	.1893(.076)[.084]	.1487(.044)	.1652(.043)[.046]	.1956(.051)[.055]
σ_v^2	.7646(.082)	.9878(.105)[.106]	–	.8848(.063)	1.0000(.071)[.074]	–
			$n = 400$; error = 1, 2, 3, for the three panels below			
β_1	1.0049(.015)	1.0040(.015)[.014]	1.0006(.015)[.016]	1.0039(.010)	1.0030(.010)[.010]	1.0000(.010)[.011]
β_2	.5716(.079)	.5582(.078)[.086]	.5083(.090)[.103]	.5631(.051)	.5498(.051)[.057]	.5031(.057)[.065]
λ	.1562(.046)	.1643(.046)[.048]	.1947(.054)[.062]	.1616(.030)	.1698(.030)[.032]	.1987(.034)[.039]
σ_v^2	.7778(.036)	.9995(.046)[.052]	–	.8933(.027)	1.0073(.031)[.034]	–
β_1	1.0048(.014)	1.0040(.014)[.014]	1.0006(.015)[.015]	1.0041(.010)	1.0032(.010)[.010]	1.0003(.010)[.010]
β_2	.5695(.080)	.5561(.079)[.086]	.5057(.091)[.103]	.5627(.051)	.5495(.051)[.057]	.5030(.057)[.065]
λ	.1579(.047)	.1660(.047)[.048]	.1968(.055)[.061]	.1609(.030)	.1691(.029)[.032]	.1979(.034)[.039]
σ_v^2	.7778(.078)	.9995(.100)[.102]	–	.8948(.064)	1.0089(.072)[.072]	–
β_1	1.0047(.014)	1.0038(.014)[.014]	1.0005(.014)[.015]	1.0038(.010)	1.0030(.010)[.010]	1.0000(.010)[.010]
β_2	.5714(.078)	.5580(.077)[.086]	.5084(.089)[.103]	.5615(.050)	.5483(.049)[.057]	.5017(.056)[.065]
λ	.1560(.046)	.1641(.046)[.048]	.1944(.054)[.062]	.1620(.029)	.1701(.029)[.032]	.1990(.033)[.039]
σ_v^2	.7777(.057)	.9993(.073)[.076]	–	.8905(.047)	1.0041(.053)[.053]	–

Note: error = 1(normal), 2(normal mixture), 3(chi-square); X_t values are generated from $N(0, 2^2)$.