Fixed Effects Estimation of Spatial Panel Model with Missing Responses: An Application to US State Tax Competition*

Xiaoyu Meng[1] and Zhenlin Yang[2†]

[1]*School of Economics, Nankai University, Tianjin, China*

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

December 23, 2023

Abstract

We consider estimation and inferences for general spatial panel data models with randomly missing observations on responses, allowing unobserved spatiotemporal heterogeneity, time-varying endogenous and contextual spatial interactions, time-varying cross-sectional error dependence, and serial correlation. A general M-estimation method is proposed for model estimation and a novel corrected plug-in method is proposed for model inference, taking into account the estimation of fixed effects. Asymptotic properties of the proposed methods are studied and finite sample properties are investigated. An empirical application is given using US state tax competition data. The proposed methods apply to matrix exponential spatial specification and can be further extended to include higher order spatial effects.

Keywords: Adjusted quasi score; Fixed effects; Missing responses; Spatial interactions; Time-varying spatial weights; Serial correlation.

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

*Zhenlin Yang gratefully acknowledges the research support from Singapore Management University.

†**Correspondence:** Zhenlin Yang, School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903. Email: zlyang@smu.edu.sg

1. Introduction

The classical spatial panel data (SPD) model takes the following vector form:

$$
\begin{aligned}
&Y_t = \lambda_0 W_t Y_t + X_{1t}\beta_{10} + W_{dt}X_{2t}\beta_{20} + Z\gamma_0 + \mu_0 + \alpha_{t0} l_n + U_t, \\
&U_t = \rho_0 M_t U_t + V_t, \quad t = 1, \ldots, T,
\end{aligned}
\tag{1.1}
$$

where Y_t is a vector of observations on n spatial units at time t, X_t is an $n \times k$ matrix containing values of k time-varying exogenous regressors, Z is an $n \times p$ matrix containing values of time-invariant 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 $n \times 1$ vectors of disturbance and idiosyncratic errors, respectively. W_t, W_{dt}, and M_t are given $n \times n$ spatial weight matrices, which together with the "spatial coefficients" λ_0, β_{20} and ρ_0, characterize the spatial lag or endogenous social effects, spatial Durbin or contextual effects, and spatial error (SE) effects, respectively.[1] β_{10} and γ_0 are vectors of regression coefficients, μ_0 is an $n \times 1$ vector of unit-specific effects, and $\alpha_0 = \{\alpha_{t0}\}_{t=1}^{T}$ a $T \times 1$ vector of time-specific effects. l_n is an $n \times 1$ vector of ones. The μ_0 and α_0 can be fixed effects (FE), or random effects (RE), or correlated random effects (CRE). Model (1.1) has been extensively studied. See, among others, Lee and Yu (2010a,b, 2015), Yang et al. (2016), Lee and Yu (2010b, 2015), and Liu and Yang (2020).

In many panels, not all (n) spatial units appeared in every time period, or even if they all appeared in every time period some spatial units at certain time periods were not fully observed. Kelejian and Prucha (2010) classify the spatial units in spatial data into three groups: (1) units with full observations on themselves and on their neighbors, (2) units with observations on their neighbors missing, and (3) units with their own observations missing. Meng and Yang (2021) studied SPD models where all units are of Type (1) but

[1] X_{2t} is typically a submatrix of X_{1t}. In the context of social networks, the three terms correspond to endogenous social effects, contextual effects, and correlated effects (Manski, 1993).

number of them can change from time to time. They refer to these as SPD models with *genuine unbalancedness.* In this paper, we study SPD models where all units are of Types (2) and (3) but missing occurs only on responses, referred to in this paper as *incomplete* SPD model with *missing responses* (MR) to emphasize on the fact that although panel is incomplete the spatial connectivity or network structure is completely observed.

Let $\mathcal{S}_t$ be an $n_t \times n$ selection matrix that selects the observed part of the $n \times 1$ vector of responses Y_t. Define $A_t(\lambda) = I_n - \lambda W_t$. If $A_t^{-1}(\lambda_0)$ exists, the SPD model with randomly missing responses has the following reduced-form representation:

$$\begin{aligned} \mathcal{S}_t Y_t &= \mathcal{S}_t A_t^{-1}(\lambda_0)(X_{1t}\beta_{10} + W_{dt}X_{2t}\beta_{20} + Z\gamma_0 + \mu_0 + \alpha_{t0} l_n + U_t), \\ U_t &= \rho_0 M_t U_t + V_t, \quad t = 1, \ldots, T. \end{aligned} \tag{1.2}$$

The model exploits the observed responses $\mathcal{S}_t Y_t$ while keeping the full structure on the other parts of the model, including regressors, spatial connectivity, and heterogeneity.

In this paper, we focus on the FE specification of Model (1.2) to give full control of unobserved unit-time heterogeneity. We propose a general M-estimation framework for model estimation and a novel *corrected plug-in* method for model inference, which take into account the estimation of the fixed effects.[2] The proposed methods are then extended to allow for serial correlation. Consistency and asymptotic normality of the proposed M-estimators are established, and consistency of the proposed corrected plug-in estimators of the variance-covariance matrices of M-estimators is proved. Monte Carlo results show that the proposed methods perform very well in finite samples and that "discarding" the spatial units with missing responses can give misleading results. An empirical illustration is given using US state tax competition data. Our methods apply to matrix exponential spatial specification, and can be further extended to include higher order spatial effects.

[2]Simpler versions of Model (1.2) are studied by Wang and Lee (2013) and Zhou et al. (2022) under RE or CRE specifications. Wang and Lee (2013, Appendix D) discussed the FE estimation of a much simpler model and pointed out the difficulty in handling the FEs for a general model.

Section 2 presents methods under iid errors. Section 3 extends the methods to allow for serial correlation. Section 4 presents some Monte Carlo results. Section 5 presents an empirical application. Section 6 concludes the paper and discusses some important extensions. Necessary results facilitating statistical inference are given in Appendix A. Technical lemmas and short proofs of the theories are presented in Appendices B-D. Detailed proofs and complete Monte Carlo results are given in `Online Appendix`.

Notations and conventions. First, $|\cdot|$, $\mathtt{tr}(\cdot)$, $'$ and $\|A\|$ are the usual notations for determinant, trace, transpose and matrix norm. For a real symmetric matrix, $\gamma_{\min}(\cdot)$ and $\gamma_{\max}(\cdot)$ denote its smallest and largest eigenvalues. For a real matrix A, $\|A\|_1$ and $\|A\|_\infty$ are the maximum absolute column and row sum norms, and $A^\circ = A + A'$. For a real matrix A of full rank, $\mathbb{P}_A = A(A'A)^{-1}A'$ and $\mathbb{Q}_A = I_n - \mathbb{P}_A$ are the projection matrices. $\mathtt{diagv}(\cdot)$ forms a column vector by the diagonal elements of a square matrix; $\mathtt{bdiag}(\cdots)$ a block-diagonal matrix; $[\cdot\ ,\cdot\ ,\ldots\ ,\cdot]$ a row vector; and $[\cdot\ ;\cdot\ ;\ldots\ ;\cdot]$ a column vector.

2. M-Estimation of Fixed Effects SPD-MR Model

Consider Model (1.2) with FE specification. For ease of exposition, the Z variables are dropped (see comments below (2.1)). Define $X_t = (X_{1t}, W_{dt}X_2)$, $\beta = (\beta_1', \beta_2')'$, and $k = \dim(\beta)$. Let $\mathbf{Y}$, $\mathbf{X}$, $\mathbf{U}$, and $\mathbf{V}$ be the stacked Y_t, X_t, U_t and V_t. Define $\mathbf{W} = \mathtt{bdiag}(W_1, \ldots, W_T)$, $\mathbf{M} = \mathtt{bdiag}(M_1, \ldots, M_T)$, $\mathbf{A}_{nT}(\lambda) = I_{nT} - \lambda\mathbf{W}$, and $\mathbf{B}_{nT}(\rho) = I_{nT} - \rho\mathbf{M}$, where I_m is an $m \times m$ identify matrix. To identify the FE parameters, a zero-sum constraint is imposed on $\{\alpha_t\}$. Define $\mathbf{D}_\mu = l_T \otimes I_n$, $\mathbf{D}_\alpha^\star = [-l_n l_{T-1}';\ I_{T-1} \otimes l_n]$, and $\mathbf{D} = [\mathbf{D}_\mu,\ \mathbf{D}_\alpha^\star]$. Let $\mathcal{S} = \mathtt{bdiag}(\mathcal{S}_1, \ldots, \mathcal{S}_T)$ be the $N \times nT$ selection matrix with $N = \sum_{t=1}^T n_t$. Denote $\phi = (\mu', \alpha_2, \ldots, \alpha_T)'$. Model (1.2) is written in matrix form:

$$\mathcal{S}\mathbf{Y} = \mathcal{S}\mathbf{A}_{nT}^{-1}(\lambda_0)[\mathbf{X}\beta_0 + \mathbf{D}\phi_0 + \mathbf{B}_{nT}^{-1}(\rho_0)\mathbf{V}]. \tag{2.1}$$

It is important to note that Model (2.1) allows time-invariant covariate effects, such as gender, to be "decomposed" from $\mathbf{D}$ by adding relevant constraints on it.

Let $\mathbf{\Omega}_N(\delta) = \mathcal{S}\mathbf{A}_{nT}^{-1}(\lambda)\mathbf{B}_{nT}^{-1}(\rho)\mathbf{B}_{nT}^{-1\prime}(\rho)\mathbf{A}_{nT}^{-1\prime}(\lambda)\mathcal{S}'$ and $\mathbf{\Omega}_N^{\frac{1}{2}}(\delta)$ be its square root matrix, where $\delta = (\lambda, \rho)'$. To simplify the presentation, denote a parametric quantity at the true parameter values by dropping its argument(s), e.g., $\mathbf{A} \equiv \mathbf{A}_{nT}(\lambda_0)$, $\mathbf{B} \equiv \mathbf{B}_{nT}(\rho_0)$, $\mathbf{\Omega}_N \equiv \mathbf{\Omega}_N(\delta_0)$. Pre-multiplying $\mathbf{\Omega}_N^{-\frac{1}{2}} \equiv \mathbf{\Omega}_N^{-\frac{1}{2}}(\delta_0)$, Model (2.1) is transformed to:

$$\mathbb{Y} = \mathbb{X}\beta_0 + \mathbb{D}\phi_0 + \mathbb{V}, \tag{2.2}$$

where $\mathbb{Y} = \mathbf{\Omega}_N^{-\frac{1}{2}}\mathcal{S}\mathbf{Y}$, $\mathbb{X} = \mathbf{C}\mathbf{X}$, $\mathbb{D} = \mathbf{C}\mathbf{D}$, and $\mathbb{V} = \mathbf{C}\mathbf{B}_{nT}^{-1}\mathbf{V}$, and $\mathbf{C} = \mathbf{\Omega}_N^{-\frac{1}{2}}\mathcal{S}\mathbf{A}_{nT}^{-1}$.

2.1. The M-estimation

If the elements of $\mathbf{V}$ are iid$(0, \sigma_{v0}^2)$, then $\mathrm{Var}(\mathbb{V}) = \sigma_{v0}^2 I_N$. If further, the time-varying regressors are exogenous, then the quasi Gaussian loglikelihood of $\theta = (\beta', \sigma_v^2, \delta')'$ and ϕ, given $\mathbf{X}$, in terms of the observed $\mathcal{S}\mathbf{Y}$ takes the form:

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

where $\mathbb{V}(\beta, \delta, \phi) = \mathbb{Y}(\delta) - \mathbb{X}(\delta)\beta - \mathbb{D}(\delta)\phi$, with $\mathbb{Y}(\delta)$, $\mathbb{X}(\delta)$ and $\mathbb{D}(\delta)$ being $\mathbb{Y}$, $\mathbb{X}$ and $\mathbb{D}$ at the general δ value. $\ell_N(\theta, \phi)$ is partially maximized at:

$$\hat{\phi}(\beta, \delta) = [\mathbb{D}'(\delta)\mathbb{D}(\delta)]^{-1}\mathbb{D}'(\delta)[\mathbb{Y}(\delta) - \mathbb{X}(\delta)\beta], \tag{2.4}$$

which is simply an OLS estimate of ϕ (given β and δ) from regressing $\mathbb{Y}(\delta) - \mathbb{X}(\delta)\beta$ on $\mathbb{D}(\delta)$. Therefore, the concentrated quasi Gaussian loglikelihood function of θ is:

$$\ell_N^{\mathrm{c}}(\theta) = -\tfrac{N}{2}\ln 2\pi - \tfrac{N}{2}\ln\sigma_v^2 - \tfrac{1}{2}\ln|\mathbf{\Omega}_N(\delta)| - \tfrac{1}{2\sigma_v^2}\tilde{\mathbb{V}}'(\beta, \delta)\tilde{\mathbb{V}}(\beta, \delta), \tag{2.5}$$

where $\tilde{\mathbb{V}}(\beta, \delta) = \mathbb{Q}_{\mathbb{D}}(\delta)[\mathbb{Y}(\delta) - \mathbb{X}(\delta)\beta]$, and $\mathbb{Q}_{\mathbb{D}}(\delta)$ is the projection matrix based on $\mathbb{D}(\delta)$. The QML estimator (QMLE) $\hat{\theta}_{\texttt{QML}}$ of θ maximizes $\ell_N^{\mathrm{c}}(\theta)$, which can be inconsistent or asymptotically biased due to the ignorance of the effect of estimating the incidental ϕ.

To rectify these problems, we adjust the concentrated quasi score (CQS) function $S_N^c(\theta) = \frac{\partial}{\partial\theta}\ell_N^c(\theta)$ to remove the effect of estimating ϕ, where,

$$S_N^c(\theta) = \begin{cases} \frac{1}{\sigma_v^2}\mathbb{X}'(\delta)\tilde{\mathbb{V}}(\beta,\delta), \\ \frac{1}{2\sigma_v^4}[\tilde{\mathbb{V}}'(\beta,\delta)\tilde{\mathbb{V}}(\beta,\delta) - N\sigma_v^2], \\ \frac{1}{2\sigma_v^2}\tilde{\mathbb{V}}'(\beta,\delta)\mathbb{H}_\lambda(\delta)\tilde{\mathbb{V}}(\beta,\delta) + \frac{1}{\sigma_v^2}\tilde{\mathbb{V}}'(\beta,\delta)\mathbb{J}(\delta)\boldsymbol{\varepsilon}(\beta,\delta) - \frac{1}{2}\mathtt{tr}[\mathbb{H}_\lambda(\delta)], \\ \frac{1}{2\sigma_v^2}\tilde{\mathbb{V}}'(\beta,\delta)\mathbb{H}_\rho(\delta)\tilde{\mathbb{V}}(\beta,\delta) - \frac{1}{2}\mathtt{tr}[\mathbb{H}_\rho(\delta)], \end{cases} \tag{2.6}$$

where $\mathbb{H}_\omega(\delta) = \boldsymbol{\Omega}_N^{-\frac{1}{2}}(\delta)[\frac{\partial}{\partial\omega}\boldsymbol{\Omega}_N(\delta)]\boldsymbol{\Omega}_N^{-\frac{1}{2}}(\delta), \omega = \lambda,\rho$, $\mathbb{J}(\delta) = \boldsymbol{\Omega}_N^{-\frac{1}{2}}(\delta)\mathcal{S}[\frac{\partial}{\partial\lambda}\mathbf{A}_{nT}^{-1}(\lambda)]$, and $\boldsymbol{\varepsilon}(\beta,\delta) = \mathbf{X}\beta + \mathbf{D}\hat{\phi}(\beta,\delta)$. Under mild conditions, $\hat{\theta}_{\mathtt{QML}} = \arg\{S_N^c(\theta) = 0\}$.

At the true θ_0, $\tilde{\mathbb{V}} = \mathbb{Q}_{\mathbb{D}}\mathbb{V}$ and $\boldsymbol{\varepsilon} = \mathbf{X}\beta_0 + \mathbf{D}\phi_0 + \mathbf{D}(\mathbb{D}'\mathbb{D})^{-1}\mathbb{D}'\mathbb{V}$. We have, $\mathrm{E}(\mathbb{X}'\tilde{\mathbb{V}}) = 0$, $\mathrm{E}(\tilde{\mathbb{V}}'\tilde{\mathbb{V}}) = (N-n-T+1)\sigma_{v0}^2$, $\mathrm{E}(\tilde{\mathbb{V}}'\mathbb{J}\boldsymbol{\varepsilon}) = 0$, and $\mathrm{E}(\tilde{\mathbb{V}}'\mathbb{H}_\omega\tilde{\mathbb{V}}) = \sigma_{v0}^2\mathtt{tr}(\mathbb{H}_\omega\mathbb{Q}_{\mathbb{D}})$, $\omega = \lambda,\rho$. Thus, $\frac{1}{N}\mathrm{E}[S_N^c(\theta_0)] = \frac{1}{N}\{0_k', -n-T+1, -\frac{1}{2}\mathtt{tr}(\mathbb{H}_\lambda\mathbb{P}_{\mathbb{D}}), -\frac{1}{2}\mathtt{tr}(\mathbb{H}_\rho\mathbb{P}_{\mathbb{D}})\}' \neq 0$, which may not even converge to 0 when either n or T is fixed. This is the root cause of inconsistency or asymptotic bias of the QMLE $\hat{\theta}_{\mathtt{QML}}$. Therefore, removing the bias in $S_N^c(\theta_0)$ due to the estimation of ϕ_0 may lead to a way for consistent and asymptotically unbiased estimation of θ. The adjusted quasi score (AQS), or *estimating function*, takes the general form:

$$S_N^*(\theta) = \begin{cases} \frac{1}{\sigma_v^2}\mathbb{X}'(\delta)\tilde{\mathbb{V}}(\beta,\delta), \\ \frac{1}{2\sigma_v^4}[\tilde{\mathbb{V}}'(\beta,\delta)\tilde{\mathbb{V}}(\beta,\delta) - N_1\sigma_v^2], \\ \frac{1}{2\sigma_v^2}\tilde{\mathbb{V}}'(\beta,\delta)\mathbb{H}_\lambda(\delta)\tilde{\mathbb{V}}(\beta,\delta) + \frac{1}{\sigma_v^2}\tilde{\mathbb{V}}'(\beta,\delta)\mathbb{J}(\delta)\boldsymbol{\varepsilon}(\beta,\delta) - \frac{1}{2}\mathtt{tr}[\mathbb{H}_\lambda(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)], \\ \frac{1}{2\sigma_v^2}\tilde{\mathbb{V}}'(\beta,\delta)\mathbb{H}_\rho(\delta)\tilde{\mathbb{V}}(\beta,\delta) - \frac{1}{2}\mathtt{tr}[\mathbb{H}_\rho(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)], \end{cases} \tag{2.7}$$

where $N_1 = N - n - T + 1$. Solving $S_N^*(\theta) = 0$ gives the M-estimator $\hat{\theta}_{\mathtt{M}}$ of θ.

The root-finding process can be simplified by first solving the equations for β and σ_v^2:

$$\hat{\beta}_{\mathtt{M}}(\delta) = [\mathbb{X}'(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)\mathbb{X}(\delta)]^{-1}\mathbb{X}'(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)\mathbb{Y}(\delta) \quad \text{and} \quad \hat{\sigma}_{v,\mathtt{M}}^2(\delta) = \tfrac{1}{N_1}\hat{\mathbb{V}}'(\delta)\hat{\mathbb{V}}(\delta), \tag{2.8}$$

where $\hat{\mathbb{V}}(\delta) = \tilde{\mathbb{V}}(\hat{\beta}_{\mathtt{M}}(\delta), \delta)$. Then, plugging $\hat{\beta}_{\mathtt{M}}(\delta)$ and $\hat{\sigma}^2_{v,\mathtt{M}}(\delta)$ back into the δ-component of (2.7) gives the concentrated AQS (estimating) function of δ:

$$S_N^{*c}(\delta) = \begin{cases} \dfrac{\hat{\mathbb{V}}'(\delta)\mathbb{H}_\lambda(\delta)\hat{\mathbb{V}}(\delta)}{2\hat{\mathbb{V}}'(\delta)\hat{\mathbb{V}}(\delta)/N_1} + \dfrac{\hat{\mathbb{V}}'(\delta)\mathbb{J}(\delta)\boldsymbol{\varepsilon}(\hat{\beta}_{\mathtt{M}}(\delta), \delta)}{\hat{\mathbb{V}}'(\delta)\hat{\mathbb{V}}(\delta)/N_1} - \dfrac{1}{2}\mathtt{tr}[\mathbb{H}_\lambda(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)], \\ \dfrac{\hat{\mathbb{V}}'(\delta)\mathbb{H}_\rho(\delta)\hat{\mathbb{V}}(\delta)}{2\hat{\mathbb{V}}'(\delta)\hat{\mathbb{V}}(\delta)/N_1} - \dfrac{1}{2}\mathtt{tr}[\mathbb{H}_\rho(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)], \end{cases} \tag{2.9}$$

Solving $S_N^{*c}(\delta) = 0$ gives us the unconstrained M-estimator $\hat{\delta}_{\mathtt{M}}$ of δ, and the M-estimators of β and σ_v^2: $\hat{\beta}_{\mathtt{M}} \equiv \hat{\beta}_{\mathtt{M}}(\hat{\delta}_{\mathtt{M}})$ and $\hat{\sigma}^2_{v,\mathtt{M}} \equiv \hat{\sigma}^2_{v,\mathtt{M}}(\hat{\delta}_{\mathtt{M}})$. The M-estimator of θ is thus $\hat{\theta}_{\mathtt{M}} = (\hat{\beta}'_{\mathtt{M}}, \hat{\sigma}^2_{v,\mathtt{M}}, \hat{\delta}'_{\mathtt{M}})'$. As discussed in the introduction, this estimation method does not suffer from the problems associated with incidental parameters. It provides a consistent and asymptotically unbiased estimation of all parameters as long as N is large.

2.2. Asymptotic properties of M-estimator

To study the asymptotic properties of the proposed M-estimator, it is necessary that the errors, regressors, and spatial weight matrices satisfy certain basic conditions. Let Δ_ϖ be the parameter space for $\varpi = \lambda, \rho$ and $\Delta = \Delta_\lambda \times \Delta_\rho$.

Assumption A. The elements v_{it} of $\mathbf{V}$ are iid for all i and t with mean zero, variance σ^2_{v0}, and $E|v_{it}|^{4+\epsilon_0} < \infty$ for some $\epsilon_0 > 0$.

Assumption B. The space Δ of δ is compact with the true δ_0 in its interior.

Assumption C. The elements of $\mathbf{X}$ are non-stochastic and bounded uniformly in i and t. $\lim_{N\to\infty} \frac{1}{N}\mathbb{X}'(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)\mathbb{X}(\delta)$ exists and is non-singular, uniformly in $\delta \in \Delta$.

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

Assumption E. Denoting by $\mathbb{A}(\varpi)$ either $\mathbf{A}_N(\lambda)$ or $\mathbf{B}_N(\rho)$, where $\varpi = \lambda, \rho$,

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

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

Under Assumption E, $\mathbf{\Omega}_N(\delta)$, its partial derivatives, and its inverse are all uniformly bounded in both row and column sum norms, uniformly in $\delta \in \Delta$, as shown in Lemma B.2(i). Some additional technical assumptions are required. Note that $\mathbf{A}_{nT}(\lambda)$ and $\mathbf{C}(\delta)$ are both block diagonal. Denote their tth blocks by $A_t(\lambda)$ and $C_t(\delta)$, respectively.

Assumption F: $A_s^{-1}(\lambda)[\frac{1}{T}\sum_{t=1}^{T} C_t'(\delta)Q_t(\delta)C_t(\delta)]^{-1}A_t^{-1\prime}(\lambda)$ is bounded in both row and column sum norms, uniformly in $\delta \in \Delta$ for all s and t, where $Q_1(\delta) = I_{n_1}$ and $Q_t(\delta) = I_{n_t} - C_t(\delta)l_n[l_n'C_t'(\delta)C_t(\delta)l_n]^{-1}l_n'C_t'(\delta), t = 2, \ldots, T$.

Under Assumption F, $\mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)\mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)$ is bounded in both row and column sum norms uniformly in $\delta \in \Delta$, as shown in Lemma B.2(ii). This facilitates our asymptotic analysis. As in the GMM estimation, a high-level assumption, *identification uniqueness*, on the population object function $\bar{S}_N^{*c}(\delta)$ is imposed, where $\bar{S}_N^{*c}(\delta)$ is the concentrated $\mathrm{E}[S_N^*(\theta)]$ with β and σ_v^2 being concentrated out (see Appendix C).

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.*

More primitive conditions under which Assumption G is satisfied are discussed in Appendix C. Finally, to cater to various asymptotic scenarios, the missingness cannot be "too heavy", and in the case of a fixed T or n, the number of observed responses is at least 2 to ensure the spatial structure is complete after ϕ is concentrated out. Let T_i be the number of times that the unit-i response is observed.

Theorem 2.1. *Under Assumptions A-G, as $N \to \infty$, if $\frac{n_t}{n} \to c_t$ and $\frac{T_i}{T} \to d_i$, where $c_t, d_i \in (0, 1]$, and $\min(T_i \geq 2)$ and $\min(n_t) \geq 2$, then we have $\hat{\theta}_{\mathtt{M}} \xrightarrow{p} \theta_0$.*

The asymptotic distribution of $\hat{\theta}_{\mathtt{M}}$ can be derived by applying the mean value theorem: $0 = S_N^*(\hat{\theta}_{\mathtt{M}}) = S_N^*(\theta_0) + \frac{\partial}{\partial\theta'}S_N^*(\bar{\theta})(\hat{\theta}_{\mathtt{M}} - \theta_0)$, where $\bar{\theta}$ lies between $\hat{\theta}_{\mathtt{M}}$ and θ_0, and its value varies over the rows of $\frac{\partial}{\partial\theta'}S_N^*(\bar{\theta})$. The key result is the asymptotic normality of $\frac{1}{\sqrt{N_1}}S_N^*(\theta_0)$. Recall $\tilde{\mathbb{V}} = \mathbb{Q}_{\mathbb{D}}\mathbb{V}$, $\boldsymbol{\varepsilon} = \mathbf{X}\beta_0 + \mathbf{D}\phi_0 + \mathbf{D}(\mathbb{D}'\mathbb{D})^{-1}\mathbb{D}'\mathbb{V}$, and $\mathbb{V} = \boldsymbol{\Gamma}\mathbf{V}$, where $\boldsymbol{\Gamma} = \mathbf{C}\mathbf{B}_{nT}^{-1}$. Then, $S_N^*(\theta_0)$ can be written in linear-quadratic (LQ) forms in $\mathbf{V}$:

$$S_N^*(\theta_0) = \begin{cases} \frac{1}{\sigma_{v0}^2}\Pi_1'\mathbf{V}, \\ \frac{1}{2\sigma_{v0}^4}\mathbf{V}'\Phi_1\mathbf{V} - \frac{N_1}{2\sigma_{v0}^2}, \\ \frac{1}{2\sigma_{v0}^2}\mathbf{V}'\Phi_2\mathbf{V} + \frac{1}{\sigma_{v0}^2}\Pi_2'\mathbf{V} - \frac{1}{2}\mathtt{tr}(\mathbb{H}_\lambda\mathbb{Q}_{\mathbb{D}}), \\ \frac{1}{2\sigma_{v0}^2}\mathbf{V}'\Phi_3\mathbf{V} - \frac{1}{2}\mathtt{tr}(\mathbb{H}_\rho\mathbb{Q}_{\mathbb{D}}), \end{cases} \tag{2.10}$$

where $\Pi_1 = \boldsymbol{\Gamma}'\mathbb{Q}_{\mathbb{D}}\mathbb{X}$, $\Pi_2 = \boldsymbol{\Gamma}'\mathbb{Q}_{\mathbb{D}}\mathbb{J}(\mathbf{X}\beta_0 + \mathbf{D}\phi_0)$, $\Phi_1 = \boldsymbol{\Gamma}'\mathbb{Q}_{\mathbb{D}}\boldsymbol{\Gamma}$, $\Phi_2 = \boldsymbol{\Gamma}'\mathbb{Q}_{\mathbb{D}}[\mathbb{H}_\lambda\mathbb{Q}_{\mathbb{D}} + 2\mathbb{J}\mathbf{D}(\mathbb{D}\mathbb{D})^{-1}\mathbb{D}']\boldsymbol{\Gamma}$, and $\Phi_3 = \boldsymbol{\Gamma}'\mathbb{Q}_{\mathbb{D}}\mathbb{H}_\rho\mathbb{Q}_{\mathbb{D}}\boldsymbol{\Gamma}$.

The representation (2.10) allows the application of the central limit theorem (CLT) for LQ forms of Kelejian and Prucha (2001) and the Wold device to give $\frac{1}{\sqrt{N_1}}S_N^*(\theta_0) \xrightarrow{D} N(0, \lim_{N\to\infty}\Gamma_N^*(\theta_0))$, an important step toward the establishment of the asymptotic normality of $\hat{\theta}_{\mathtt{M}}$. It also allows an easy derivation of $\mathrm{E}[\frac{\partial}{\partial\theta'}S_N^*(\theta_0)]$ as seen in Appendix A. The consistency of $\hat{\theta}_{\mathtt{M}}$ leads to $\frac{1}{N_1}[\frac{\partial}{\partial\theta'}S_N^*(\bar{\theta}) - \mathrm{E}[\frac{\partial}{\partial\theta'}S_N^*(\theta_0)]] = o_p(1)$.

Theorem 2.2. *Under the assumptions of Theorem 2.1, we have, as $N \to \infty$,*

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

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.

2.3. Estimation of the VC matrix

Inferences for θ require a consistent estimator of the asymptotic variance-covariance (VC) matrix $\Sigma_N^{*-1}(\theta_0)\Gamma_N^*(\theta_0)\Sigma_N^{*-1\prime}(\theta_0)$. The analytical expressions of $\frac{\partial}{\partial\theta'}S_N^*(\theta)$ and $\Gamma_N^*(\theta_0)$ are given in Appendix A. First, it is easy to show that $\widehat{\Sigma}_N^* = -\frac{1}{N_1}\frac{\partial}{\partial\theta'}S_N^*(\theta)|_{\theta=\hat{\theta}_{\mathtt{M}}}$ consistently estimates $\Sigma_N^*(\theta_0)$, i.e., $\widehat{\Sigma}_N^* - \Sigma_N^*(\theta_0) = o_p(1)$.

$\Gamma_N^*(\theta_0)$ contains the common parameters θ_0, the fixed effects ϕ_0 embedded in Π_2, and the skewness κ_3 and excess kurtosis κ_4 of the idiosyncratic errors. The common *plug-in* method may not be valid due to the involvement of the incidental parameters ϕ_0. A *corrected plug-in* method is proposed. Let $\Gamma_N^*(\hat{\theta}_{\mathtt{M}}) = \Gamma_N^*(\theta)|_{(\theta=\hat{\theta}_{\mathtt{M}}, \phi=\hat{\phi}_{\mathtt{M}}, \kappa_3=\hat{\kappa}_{3,N}, \kappa_4=\hat{\kappa}_{4,N})}$ be the plug-in estimator, where $\hat{\phi}_{\mathtt{M}}$ is the M-estimator of ϕ,[3] and $\hat{\kappa}_{3,N}$ and $\hat{\kappa}_{4,N}$ are consistent estimators of κ_3 and κ_4. When both n and T are large, $\Gamma_N^*(\hat{\theta}_{\mathtt{M}})$ would be consistent as $\hat{\phi}_{\mathtt{M}}$ is. However, when either n or T is fixed, $\hat{\phi}_{\mathtt{M}}$ is not consistent and a bias correction is necessary after plugging $\hat{\phi}_{\mathtt{M}}$ into $\Gamma_N^*(\theta)$. We show that the only term that cannot be consistently estimated is the one quadratic in ϕ_0, embedded in $\Pi_2'\Pi_2$.

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

$$\Gamma_N^*(\hat{\theta}_{\mathtt{M}}) = \Gamma_N^*(\theta_0) + \text{Bias}^*(\delta_0) + o_p(1),$$

where $\text{Bias}^*(\delta_0)$ *has a solo non zero* λ-λ *entry* $\frac{1}{N_1}\mathtt{tr}[(\mathbb{D}'\mathbb{D})^{-1}\mathbf{D}'\mathbb{J}'\mathbb{Q}_{\mathbb{D}}\mathbb{J}\mathbf{D}]$.

See the proof of Corollary 2.1 in `Online Appendix` for details on the above discussions. Corollary 2.1 leads immediately to a general consistent estimator of $\Gamma_N^*(\theta_0)$:

$$\widehat{\Gamma}_N^* = \Gamma_N^*(\hat{\theta}_{\mathtt{M}}) - \text{Bias}^*(\hat{\delta}_{\mathtt{M}}),$$

referred to in this paper as the *corrected plug-in* estimator.

Finally, we provide consistent estimators for κ_3 and κ_4. As $\mathbf{V}$ is infeasible for estimation due to the incidental parameters problem and incompleteness, we start from $\mathbf{\Omega}_N^{-\frac{1}{2}}\tilde{\mathbb{V}} = \mathbf{\Omega}_N^{-\frac{1}{2}}\mathbb{Q}_{\mathbb{D}}\mathbf{\Gamma}\mathbf{V}$, which can be "consistently" estimated by $\mathbf{\Omega}_N^{-\frac{1}{2}}(\hat{\delta}_{\mathtt{M}})\hat{\mathbb{V}}(\hat{\beta}_{\mathtt{M}}, \hat{\delta}_{\mathtt{M}}) = \mathbf{\Omega}_N^{-\frac{1}{2}}(\hat{\delta}_{\mathtt{M}})\mathbb{Q}_{\mathbb{D}}(\hat{\delta}_{\mathtt{M}})\mathbf{\Omega}_N^{-\frac{1}{2}}(\hat{\delta}_{\mathtt{M}})\mathcal{S}[\mathbf{Y} - \mathbf{A}_{nT}^{-1}(\hat{\lambda}_{\mathtt{M}})\mathbf{X}\hat{\beta}_{\mathtt{M}}]$. Let q_{jk} be the (j,k)th element of $N \times nT$ matrix $\bar{\mathbb{Q}}_{\mathbb{D}} \equiv \mathbf{\Omega}_N^{-\frac{1}{2}}\mathbb{Q}_{\mathbb{D}}\mathbf{\Gamma}$. Denote the elements of $\mathbf{V}$ by $v_l, l = 1, \ldots, nT$, and the elements of

[3]Or a GLS estimator by regressing $\mathcal{S}[\mathbf{Y} - \mathbf{A}_{nT}^{-1}(\hat{\lambda}_{\mathtt{M}})\mathbf{X}\hat{\beta}_{\mathtt{M}}]$ on $\mathcal{S}\mathbf{A}_{nT}^{-1}(\hat{\lambda}_{\mathtt{M}})\mathbf{D}$ with VC matrix $\mathbf{\Omega}_N(\hat{\delta}_{\mathtt{M}})$.

$\bar{\mathbb{Q}}_{\mathbb{D}}\mathbf{V}$ by $\tilde{v}_j, j = 1, \ldots, N$, where l and j are the combined index of cross-sectional and time dimensions. Then, $\tilde{v}_j = \sum_{k=1}^{nT} q_{jk}v_k$, and thus $\mathrm{E}(\tilde{v}_j^3) = \sum_{k=1}^{nT} q_{jk}^3\mathrm{E}(v_k^3) = \sigma_{v0}^3\kappa_3 \sum_{k=1}^{nT} q_{jk}^3$. Summing $\mathrm{E}(\tilde{v}_j^3)$ over j gives $\kappa_3 = (\sum_{j=1}^N \mathrm{E}(\tilde{v}_j^3))(\sigma_{v0}^3 \sum_{j=1}^N \sum_{k=1}^{nT} q_{jk}^3)^{-1}$. Its sample analog:

$$\hat{\kappa}_{3,N} = \frac{\sum_{j=1}^N \hat{v}_j^3}{\hat{\sigma}_{v,\mathtt{M}}^3 \sum_{j=1}^N \sum_{k=1}^{nT} \hat{q}_{jk}^3} \tag{2.11}$$

gives a consistent estimator of κ_3, where $\hat{v}_j$ is the jth element of $\mathbf{\Omega}_N^{-\frac{1}{2}}(\hat{\delta}_{\mathtt{M}})\hat{\mathbb{V}}(\hat{\beta}_{\mathtt{M}}, \hat{\delta}_{\mathtt{M}})$, and $\hat{q}_{jk}$ is the (j,k)th element of $\bar{\mathbb{Q}}_{\mathbb{D}}(\hat{\delta}_{\mathtt{M}})$. Similarly, to estimate κ_4, we have,

$$\begin{aligned}
\mathrm{E}(\tilde{v}_j^4) &= \textstyle\sum_{k=1}^{nT} q_{jk}^4\mathrm{E}(v_k^4) + 3\sigma_{v0}^4 \sum_{k=1}^{nT}\sum_{l=1}^{nT} q_{jk}^2 q_{jl}^2 - 3\sigma_{v0}^4 \sum_{k=1}^{nT} q_{jk}^4 \\
&= \textstyle\sum_{k=1}^{nT} q_{jk}^4\kappa_4\sigma_{v0}^4 + 3\sigma_{v0}^4 \sum_{k=1}^{nT}\sum_{l=1}^{nT} q_{jk}^2 q_{jl}^2,\ j = 1, \ldots, N,
\end{aligned}$$

which gives $\kappa_4 = \big(\sum_{j=1}^N \mathrm{E}(\tilde{v}_j^4) - 3\sigma_{v0}^4 \sum_{j=1}^N\sum_{k=1}^{nT}\sum_{l=1}^{nT} q_{jk}^2 q_{jl}^2\big)\big(\sigma_{v0}^4 \sum_{j=1}^N \sum_{k=1}^{nT} q_{jk}^4\big)^{-1}$, by summing $\mathrm{E}(\tilde{v}_j^4)$ over j. Hence, a consistent estimator for κ_4 is

$$\hat{\kappa}_{4,N} = \frac{\sum_{j=1}^N \hat{v}_j^4 - 3\hat{\sigma}_{v,\mathtt{M}}^4 \sum_{j=1}^N\sum_{k=1}^{nT}\sum_{l=1}^{nT} \hat{q}_{jk}^2\hat{q}_{jl}^2}{\hat{\sigma}_{v,\mathtt{M}}^4 \sum_{j=1}^N \sum_{k=1}^{nT} \hat{q}_{jk}^4}. \tag{2.12}$$

Corollary 2.2. *Under the assumptions of Theorem 2.1, we have, as $N \to \infty$,*

(i) $\hat{\kappa}_{3,N} \xrightarrow{p} \kappa_{3,0}$ *and* $\hat{\kappa}_{4,N} \xrightarrow{p} \kappa_{4,0}$; *(ii)* $\widehat{\Sigma}_N^* - \Sigma_N^*(\theta_0) \xrightarrow{p} 0$ *and* $\widehat{\Gamma}_N^* - \Gamma_N^*(\theta_0) \xrightarrow{p} 0$; *and therefore* $\widehat{\Sigma}_N^{*-1}\widehat{\Gamma}_N^*\widehat{\Sigma}_N^{*-1\prime} - \Sigma_N^{*-1}(\theta_0)\Gamma_N^*(\theta_0)\Sigma_N^{*-1\prime}(\theta_0) \xrightarrow{p} 0.$

See the proof of Corollary 2.2 in `Online Appendix` for details on the above discussions.

3. M-Estimation with Serial Correlation

In this section, we show that our M-estimation and inference methods introduced in Section 2 can be extended to allow the errors to be serially correlated.

Assumption A$'$: The innovations follow an MA process, $v_{it} = e_{it} + \tau e_{i,t-1}$, for all i and t with $|\tau| < 1$, $e_{it} \sim \mathrm{iid}(0, \sigma_e^2)$, and $E|e_{it}|^{4+\epsilon_0} < \infty$ for some $\epsilon_0 > 0$.

To conserve space, we use the same set of notations of Section 2, with relevant quantities being redefined to cater to the extra parameter τ. Let now $\delta = (\lambda, \rho, \tau)'$, $\theta = (\beta', \sigma_e^2, \delta')'$ and $\mathbf{\Omega}_N(\delta) \equiv \mathcal{S}\mathbf{A}_{nT}^{-1}(\lambda)\mathbf{B}_{nT}^{-1}(\rho)[\Upsilon(\tau)\Upsilon'(\tau) \otimes I_n]\mathbf{B}_{nT}^{-1\prime}(\rho)\mathbf{A}_{nT}^{-1\prime}(\lambda)\mathcal{S}'$, where $\Upsilon(\tau)$ is $T \times (T+1)$ with rows: $(\tau, 1, 0, \ldots, 0)$, $(0, \tau, 1, \ldots, 0)$, $\cdots$, $(0, 0, \ldots, \tau, 1)$.

With the redefined δ, θ and $\mathbf{\Omega}_N(\delta)$, update $\mathbb{Y}$, $\mathbb{X}$, $\mathbb{D}$, and $\mathbb{V}$ in (2.2). The transformed model remains in the form as (2.2) except that now $\mathrm{Var}(\mathbb{V}) = \sigma_{e0}^2 I_N$. The loglikelihood function of (θ, ϕ) remains in the same form as (2.3) with σ_{v0}^2 being replaced by σ_{e0}^2. The constrained QMLE of ϕ remains in the same form as (2.4). Updating $\mathbb{Q}_{\mathbb{D}}(\delta)$ with the updated $\mathbb{D}(\delta)$ and thus $\tilde{\mathbb{V}}(\beta, \delta)$, we then see that the concentrated quasi Gaussian loglikelihood of θ has the same form as (2.5), which leads to the direct QMLE of θ.

The CQS function of θ is obtained and its expectation at the true θ_0 is found in a similar way as that in Section 2. The desired AQS function of θ is obtained:

$$S_N^{\diamond}(\theta) = \begin{cases} \frac{1}{\sigma_e^2}\mathbb{X}'(\delta)\tilde{\mathbb{V}}(\beta, \delta), \\ \frac{1}{2\sigma_e^4}[\tilde{\mathbb{V}}'(\beta, \delta)\tilde{\mathbb{V}}(\beta, \delta) - N_1\sigma_e^2], \\ \frac{1}{2\sigma_e^2}\tilde{\mathbb{V}}'(\beta, \delta)\mathbb{H}_\lambda(\delta)\tilde{\mathbb{V}}(\beta, \delta) + \frac{1}{\sigma_e^2}\tilde{\mathbb{V}}'(\beta, \delta)\mathbb{J}(\delta)\boldsymbol{\varepsilon}(\beta, \delta) - \frac{1}{2}\mathtt{tr}[\mathbb{H}_\lambda(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)], \\ \frac{1}{2\sigma_e^2}\tilde{\mathbb{V}}'(\beta, \delta)\mathbb{H}_\rho(\delta)\tilde{\mathbb{V}}(\beta, \delta) - \frac{1}{2}\mathtt{tr}[\mathbb{H}_\rho(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)], \\ \frac{1}{2\sigma_e^2}\tilde{\mathbb{V}}'(\beta, \delta)\mathbb{H}_\tau(\delta)\tilde{\mathbb{V}}(\beta, \delta) - \frac{1}{2}\mathtt{tr}[\mathbb{H}_\tau(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)], \end{cases} \tag{3.1}$$

where $\mathbb{H}_\omega(\delta)$, $\mathbb{J}(\delta)$, and $\boldsymbol{\varepsilon}(\beta, \delta)$ are defined in (2.6), and now are extended/redefined to cater the extra parameter ($\omega = \lambda, \rho, \tau$). Solving $S_N^{\diamond}(\theta) = 0$ gives the M-estimator $\hat{\theta}_{\mathtt{M}}^{\diamond}$ of θ.

The asymptotic properties of $\hat{\theta}_{\mathtt{M}}^{\diamond}$ can be established in a similar way as for $\hat{\theta}_{\mathtt{M}}$ in Section 2, based on a similar sequence of assumptions (A′−G′). Assumption A′ can be extended to a higher order MA process. Assumptions B′ and C′ take the same form as Assumptions B and C but with relevant quantities being redefined, and D′ and E′ are the same as D and E. As $\mathbf{\Omega}_N(\delta)$ is no longer block diagonal, Assumption F needs to be modified.

Assumption F′: $\|\mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)\mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)\|_1$ *and* $\|\mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)\mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)\|_\infty$ *are bounded uniformly in* $\delta \in \Delta$.

The key steps for asymptotic analysis are: concentrating β and σ_e^2 out from (3.1) to give the concentrated AQS functions of δ, $S_N^{\diamond c}(\delta)$; concentrating β and σ_e^2 out from $\bar{S}_N^{\diamond}(\theta) = \mathrm{E}[S_N^{\diamond}(\theta)]$ to give the population counterpart $\bar{S}_N^{\diamond c}(\delta)$ of $S_N^{\diamond c}(\delta)$, and then showing that $\sup_{\delta\in\Delta}\frac{1}{N_1}\left\|S_N^{\diamond c}(\delta) - \bar{S}_N^{\diamond c}(\delta)\right\| \xrightarrow{p} 0$. This and the identification condition given below lead to the consistency of $\hat{\delta}_{\mathtt{M}}^{\diamond}$ for δ_0, and thus the consistency of $\hat{\theta}_{\mathtt{M}}^{\diamond}$.

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.

Theorem 3.1. *Under Assumptions A′-G′, as* $N \to \infty$*, if* $\frac{n_t}{n} \to c_t$ *and* $\frac{T_i}{T} \to d_i$*, where* $c_t, d_i \in (0,1]$*, and* $\min(T_i \geq 2)$ *and* $\min(n_t) \geq 2$*, then* $\hat{\theta}_{\mathtt{M}}^{\diamond} \xrightarrow{p} \theta_0$.

To derive the asymptotic distribution of $\hat{\theta}_{\mathtt{M}}^{\diamond}$, note that the AQS functions at the true θ_0, expressed in $\mathbf{V}$, take similar forms as (2.10), with an extra τ-component. In (2.10), replace $\mathbf{V}$ by $(\Upsilon \otimes I_n)\mathcal{E}$ and σ_{v0}^2 by σ_{e0}^2, where $\mathcal{E} = (\mathcal{E}_0', \mathcal{E}_1', \ldots, \mathcal{E}_T')'$, and $\mathcal{E}_t = (e_{1t}, e_{2t}, \ldots, e_{nt})'$; redefine $\mathbf{\Gamma}$ as $\mathbf{\Omega}_N^{-\frac{1}{2}}\mathcal{S}\mathbf{A}_{nT}^{-1}\mathbf{B}_{nT}^{-1}(\Upsilon \otimes I_n)$ and update Π_r and Φ_s accordingly, $r = 1, 2, s = 1, 2, 3$; and introduce new Φ_4 (defined as Φ_3) and $\mathbb{H}_\tau$ (defined as $\mathbb{H}_\rho$). We have,

$$
S_N^{\diamond}(\theta_0) = \begin{cases}
\frac{1}{\sigma_{e0}^2}\Pi_1'\mathcal{E}, \\
\frac{1}{2\sigma_{e0}^4}\mathcal{E}'\Phi_1\mathcal{E} - \frac{N_1}{2\sigma_{v0}^2}, \\
\frac{1}{2\sigma_{e0}^2}\mathcal{E}'\Phi_2\mathcal{E} + \frac{1}{\sigma_{v0}^2}\Pi_2'\mathbf{V} - \frac{1}{2}\mathtt{tr}(\mathbb{H}_\lambda\mathbb{Q}_{\mathbb{D}}), \\
\frac{1}{2\sigma_{e0}^2}\mathcal{E}'\Phi_3\mathcal{E} - \frac{1}{2}\mathtt{tr}(\mathbb{H}_\rho\mathbb{Q}_{\mathbb{D}}), \\
\frac{1}{2\sigma_{e0}^2}\mathcal{E}'\Phi_4\mathcal{E} - \frac{1}{2}\mathtt{tr}(\mathbb{H}_\tau\mathbb{Q}_{\mathbb{D}}),
\end{cases} \tag{3.2}
$$

which is linear-quadratic in $\mathcal{E}$ with iid elements. Again, the importance of this representation is two fold: it allows the application of CLT for LQ forms of Kelejian and Prucha

(2001) and the Wold device to establish the asymptotic normality of $\frac{1}{\sqrt{N_1}}S_N^\diamond(\theta_0)$ (thus the asymptotic normality $\hat{\theta}_{\mathtt{M}}^\diamond$) and an easy derivation of $\mathrm{E}[S_N^\diamond(\theta_0)]$ as seen in Appendix A.

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

$$\sqrt{N_1}(\hat{\theta}_{\mathtt{M}}^\diamond - \theta_0) \xrightarrow{D} N\Big(0, \lim_{N\to\infty} \Sigma_N^{\diamond-1}(\theta_0)\Gamma_N^\diamond(\theta_0)\Sigma_N^{\diamond-1\prime}(\theta_0)\Big),$$

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

For statistical inference, $\Sigma_N^\diamond(\theta_0)$ is estimated by $\widehat{\Sigma}_N^\diamond = -\frac{1}{N_1}\frac{\partial}{\partial\theta'}S_N^\diamond(\theta)|_{\theta=\hat{\theta}_{\mathtt{M}}^\diamond}$. The analytical expressions of $\frac{\partial}{\partial\theta'}S_N^\diamond(\theta)$ and $\Gamma_N^\diamond(\theta_0)$ are given in Appendix A. Similar to $\Gamma_N^*(\theta_0)$ in Section 2, $\Gamma_N^\diamond(\theta_0)$ contains the common parameters θ_0, the incidental parameters ϕ_0, and the skewness κ_3^e and the excess kurtosis κ_4^e of the idiosyncratic errors $\{e_{it}\}$. Again, the usual plug-in estimator would not lead to a consistent estimation of $\Gamma_N^\diamond(\theta_0)$, and a corrected plug-in estimator (correction on $\Pi_2'\Pi_2$) is developed:

$$\widehat{\Gamma}_N^\diamond = \Gamma_N^\diamond(\hat{\theta}_{\mathtt{M}}^\diamond) - \mathrm{Bias}^\diamond(\hat{\delta}_{\mathtt{M}}^\diamond),$$

where $\mathrm{Bias}^\diamond(\delta_0)$ has a sole non-zero λ-λ element $\frac{1}{N_1}\mathtt{tr}[(\mathbb{D}'\mathbb{D})^{-1}\mathbf{D}'\mathbb{J}'\mathbb{Q}_{\mathbb{D}}\mathbb{J}\mathbf{D}]$.

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

$$\Gamma_N^\diamond(\hat{\theta}_{\mathtt{M}}^\diamond) = \Gamma_N^\diamond(\theta_0) + \mathrm{Bias}^\diamond(\delta_0) + o_p(1),$$

where $\mathrm{Bias}^\diamond(\delta_0)$ *has a solo non zero λ-λ entry* $\frac{1}{N_1}\mathtt{tr}[(\mathbb{D}'\mathbb{D})^{-1}\mathbf{D}'\mathbb{J}'\mathbb{Q}_{\mathbb{D}}\mathbb{J}\mathbf{D}]$.

Finally, we note that $\mathbf{\Omega}_N^{-\frac{1}{2}}\tilde{\mathbb{V}} = \bar{\mathbb{Q}}_{\mathbb{D}}(\Upsilon \otimes I_n)\mathcal{E}$ can be "consistently" estimated by $\mathbf{\Omega}_N^{-\frac{1}{2}}(\hat{\delta}_{\mathtt{M}}^\diamond)\hat{\mathbb{V}}(\hat{\beta}_{\mathtt{M}}^\diamond, \hat{\delta}_{\mathtt{M}}^\diamond) = \mathbf{\Omega}_N^{-\frac{1}{2}}(\hat{\delta}_{\mathtt{M}}^\diamond)\mathbb{Q}_{\mathbb{D}}(\hat{\delta}_{\mathtt{M}}^\diamond)\mathbf{\Omega}_N^{-\frac{1}{2}}(\hat{\delta}_{\mathtt{M}}^\diamond)\mathcal{S}[\mathbf{Y} - \mathbf{A}_{nT}^{-1}(\hat{\lambda}_{\mathtt{M}}^\diamond)\mathbf{X}\hat{\beta}_{\mathtt{M}}^\diamond]$. We follow the idea of Corollary 2.2 and develop a pair of consistent estimators for κ_3^e and κ_4^e as follows:

$$\hat{\kappa}_{3,N}^e = \frac{\sum_{j=1}^N \hat{v}_j^3}{\hat{\sigma}_{e,\mathtt{M}}^{\diamond 3}\sum_{j=1}^N\sum_{k=1}^{n(T+1)}\hat{q}_{jk}^3} \quad \text{and} \quad \hat{\kappa}_{4,N}^e = \frac{\sum_{j=1}^N \hat{v}_j^4 - 3\hat{\sigma}_{e,\mathtt{M}}^{\diamond 4}\sum_{j=1}^N\sum_{k=1}^{n(T+1)}\sum_{l=1}^{nT}\hat{q}_{jk}^2\hat{q}_{jl}^2}{\hat{\sigma}_{e,\mathtt{M}}^{\diamond 4}\sum_{j=1}^N\sum_{k=1}^{n(T+1)}\hat{q}_{jk}^4}.$$

where $\hat{q}_{jk}$ is the (j,k)th element of $N \times n(T+1)$ matrix $\mathbf{\Omega}_N^{-\frac{1}{2}}(\hat{\delta}^{\diamond}_{\mathtt{M}})\bar{\mathbb{Q}}_{\mathbb{D}}(\hat{\delta}^{\diamond}_{\mathtt{M}})(\Upsilon(\hat{\tau}^{\diamond}_{\mathtt{M}}) \otimes I_n)$ and $\hat{v}_j$ the jth element of $\mathbf{\Omega}_N^{-\frac{1}{2}}(\hat{\delta}^{\diamond}_{\mathtt{M}})\hat{\mathbb{V}}(\hat{\beta}^{\diamond}_{\mathtt{M}}, \hat{\delta}^{\diamond}_{\mathtt{M}})$.

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

(i) $\hat{\kappa}^e_{3,N} \xrightarrow{p} \kappa^e_{3,0}$ *and* $\hat{\kappa}^e_{4,N} \xrightarrow{p} \kappa^e_{4,0}$; *(ii)* $\widehat{\Sigma}^{\diamond}_N - \Sigma^{\diamond}_N(\theta_0) \xrightarrow{p} 0$ *and* $\widehat{\Gamma}^{\diamond}_N - \Gamma^{\diamond}_N(\theta_0) \xrightarrow{p} 0$; *and therefore* $\widehat{\Sigma}^{\diamond-1}_N \widehat{\Gamma}^{\diamond}_N \widehat{\Sigma}^{\diamond-1\prime}_N - \Sigma^{\diamond-1}_N(\theta_0)\Gamma^{\diamond}_N(\theta_0)\Sigma^{\diamond-1\prime}_N(\theta_0) \xrightarrow{p} 0$.

4. Monte Carlo Results

Monte Carlo experiments are conducted to investigate (i) the finite sample performance of the proposed M-estimators of the common parameters and the corrected plug-in estimator of the VC matrix, and (ii) the consequence of discarding the units with missing responses. The following data generating process is used:

$$\mathcal{S}_t Y_t = \mathcal{S}_t A_t^{-1}(\lambda)(X_t\beta + \mu + \alpha_t l_n + U_t), \quad U_t = \rho M_t U_t + V_t, \ \ t = 1, \dots, T,$$

The parameters values are set at $\beta = 1$, $\lambda = 0.2$, $\rho = 0.2$, and $\sigma_v^2 = 1$. The $X_t's$ are generated independently from $N(0, 2^2 I_n)$, the individual fixed effects from $\mu = \frac{1}{T}\Sigma_{t=1}^T X_t + e$, where $e \sim N(0, I_n)$, and the time fixed effects α from $N(0, I_T)$. The sample sizes are based on $n \in (50, 100, 200, 400)$ and $T \in (5, 10)$. For each Monte Carlo experiment, the number of Monte Carlo runs is set to 1000.

The spatial weight matrices can be `Rook` contiguity or `Queen` contiguity. To generate W_{nt} under `Rook`, randomly permute the indices $\{1, 2, \dots, n\}$ for n spatial units and then allocate them into a lattice of $k \times m$ squares. Let $W_{nt,ij} = 1$ if the index j is in a square that is immediately left or right, above, or below the square that contains the index i. Similarly, W_{nt} under `Queen` is generated with additional neighbors sharing a common vertex with the unit i. The distribution of the idiosyncratic errors can be (i) `normal`, (ii) standardized `normal mixture` (10% $N(0,4^2)$ and 90% $N(0,1)$), or (iii) standardized `chi-square` with 3 degrees of freedom. See Yang (2015) for details.

The selection matrices $\mathcal{S}_t$ are generated as follows: for each t, associate with each row of I_n a uniform (0, 1) random number, and the rows with random numbers smaller than $p_t \in (0,1)$ are deleted. This gives $100p_t\%$ missing on the responses. We consider a case with iid errors, and a case with serially correlated errors with $\tau = 0.5$. We consider two randomly missing percentages, 10% and 30%, to see the effect of the degree of missingness.

The Monte Carlo experiments involve five estimators: `naïve`, `QMLE-GU` and `M-Est-GU`, and `QMLE-MR` and `M-Est-MR`. The `naïve` estimator is the M-estimator based on a balanced panel formed by deleting entirely the spatial units with missing responses; `QMLE-GU` and `M-Est-GU` are the estimators developed by Meng and Yang (2021) for genuinely unbalanced spatial panels; and `QMLE-MR` and `M-Est-MR` are our proposed estimators. Clearly, only the estimator `M-Est-MR` is valid when data is generated according to `FE-SPD-MR`. These Monte Carlo experiments allow us not only to see the finite sample performance of the proposed estimation and inference methods but also the consequence of a wrong choice of estimator, and a wrong choice of modeling mechanism. To conserve space, only partial Monte Carlo results are reported. The full set of results is in `Online Appendix`.

Table 1 contains partial results on `QMLE-MR`, `M-Est-GU` and `M-Est-MR` for the case of iid errors. The results (reported and unreported) show an excellent performance of the proposed M-estimation and inference methods, irrespective of the error distributions, the spatial layouts, parameter values, as well as the missing percentage. In contrast, the `QMLE-MR` (the closest estimator to `M-Est-MR`) of σ^2 is inconsistent, and the `QMLE-MRs` of spatial parameters do not perform as well as the `M-Est-MRs`. By comparing `M-Est-GU` with M-Est-MR, we can see the consequences of treating an MR mechanism as a GU mechanism: M-Est-GUs of the spatial parameters perform poorly even when the sample size is fairly large. When the missing percentage is higher, M-Est-GU estimator becomes more biased. This is consistent with our expectation as treating FE-SPD-MR as FE-SPD-GU ignores the spatial effects from the units with missing responses. The larger the missing percentage is, the more serious the consequence. Furthermore, the very poor performance of naïve and QMLE-GU is also clearly demonstrated by the Monte Carlo results (see Table 1c in Online Appendix). Therefore, in spatial panel frameworks, a wrong choice of estimator, i.e., näive estimator and QMLEs, and a wrong choice of mechanism, i.e., treating MR as GU, can have a very serious consequence.

Table 2 contains partial results on QMLE-MR and M-Est-MR for the case of serially correlated errors, as M-Est-GU is unavailable. The proposed M-Est-MRs of all the parameters have a very good finite sample performance. Their corresponding standard error estimates are also close to Monte Carlo standard deviations. In contrast, the QMLE-MR typically provides much worse estimates for error variance parameter σ^2 and serial correlation parameter τ, showing that the incidental parameters problem is more serious to the estimation of the parameters in the error term.

5. An Empirical Application

In this section, we present an empirical study to analyze horizontal competition in excise taxes on beer and gasoline among US states. The theoretical models set up in Kanbur and Keen (1993) and Nielsen (2001) imply that independent jurisdictions have incentives to engage in commodity tax competition in order to attract cross-border shoppers and thus maximize their tax revenue. Therefore, the tax rates of neighboring states are likely to play a role in the determination of the state's own tax policy. Egger et al. (2005) and Devereux et al. (2007) find empirical evidence for positive spillover effects. Egger et al. (2005) estimate the `SE` parameter using GMM and the `SL` parameter by 2SLS. Devereux et al. (2007) does not include the `SE` effect in the model. They treat the data as genuinely unbalanced (GU) panel data (some states in certain periods had observations on the response and covariate missing) in the sense of Meng and Yang (2021), thereby spillover effects to/from these ignored units with missing tax rates were not captured.

In this section, we reconsider this study under the missing-on-response-only (MR) mechanism since the explanatory variables can be fully observed over a chosen period. We construct two panels based on 48 contiguous US states over 19 years (1978-1996), the tax rates on beer and the tax rates on gasoline. The numbers of observations for beer and gasoline tax rates are, respectively, 911 and 888. We define the spatial neighboring states as those that share a common border. The overall spatial weight matrix W is row-normalized. The explanatory variables we use are state size (population density), spatially weighted size (WSize), dependency ratio (DR), government ideological orientation (GIO), lagged sales tax rate (LSTR), gross state product (GSP), and public expenditure (PE).[4]

Table A1 summarizes the estimation results based on tax rates on beer and tax rates on gasoline. For each set of data, we report three types of estimation results: M-estimation based on GU (Meng and Yang, 2021), M-estimation based on MR, and M-estimation based on MR with serial correlation (MRSC). Our analyses lead to a deeper understanding on the mechanism of tax competition, and offer more insights on the nature of spatial interactions. Both analyses point to the existence of strong and positive endogenous

[4]**Data sources.** Tax rates and PE: World Tax Database (`https://www.bus.umich.edu/otpr/otpr/default.asp`); GSP: US Bureau of Economic Analysis (`https://www.bea.gov/data/gdp/gdp-state`); other control variables: the data sources described in Egger et al. (2005); and the missing values on PE are recovered from United States Census Bureau (`https://www.census.gov/programs-surveys/state/data/historical_data.html`).

spatial spillover effects (tax competition), and strong and positive serial correlation.

Table A1: Spatial missing response panel analyses of US state tax rates

Variables	Beer			Gasoline		
	`GU`	`MR`	`MRSC`	`GU`	`MR`	`MRSC`
State Size	0.160***	0.159***	0.147***	0.038	0.053*	0.048
	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)
WSize	−0.113*	−0.124**	−0.155***	−0.129***	−0.119***	−0.114**
	(0.07)	(0.06)	(0.07)	(0.05)	(0.05)	(0.05)
DR	0.194*	0.175**	0.116	0.003	0.014	0.034
	(0.10)	(0.09)	(0.08)	(0.08)	(0.07)	(0.07)
GIO	−0.034**	−0.036***	−0.008	0.014	0.014	0.016
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
LSTR	0.273	0.270	−0.085	−0.044	0.007	−0.086
	(0.23)	(0.23)	(0.23)	(0.18)	(0.19)	(0.18)
GSP	−0.783***	−0.758***	−0.586***	−0.108	−0.173*	−0.146
	(0.09)	(0.10)	(0.10)	(0.09)	(0.10)	(0.10)
PE	0.007***	0.007***	0.005***	0.000	0.001	0.000
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
SL(λ)	0.197	0.316*	0.370**	0.081	0.329***	0.262**
	(0.24)	(0.18)	(0.18)	(0.11)	(0.12)	(0.13)
SE(ρ)	−0.026	−0.165	−0.245	0.270**	0.010	0.038
	(0.28)	(0.23)	(0.22)	(0.13)	(0.16)	(0.16)
SC(τ)			0.699***			0.692***
			(0.02)			(0.02)
Pseudo R^2	96.56%	96.64%	98.26%	82.31%	82.52%	90.59%
States FE	48	48	48	48	48	48
Years FE	19	19	19	19	19	19
Observed Responses	911	911	911	888	888	888

Significance levels: *:10%, **:5%, and ***: 1%. Standard error are in parentheses.

The first estimation serves to show the effects of ignoring the spatial effects from units with missing responses. For both data sets, the M-estimates of `SL` parameter based on MR or MRSC are all significantly positive, suggesting the existence of tax competition. In contrast, these based on GU are not. It is interesting to note that beer data have only one response value gone missing, but taking it into account by MR scheme completely changes the conclusion on tax competition. The estimates of `SE` coefficient exhibit negative but insignificant values for beer data, which is in line with Egger et al. (2005). However, our approach has an advantage in that we can make statistical inferences on the `SE` effect. The MR and MRSC estimates of `SE` coefficient for gasoline data are positive but insignificant. Lastly, the MRSC estimates of serial correlation are significantly positive for both data sets, suggesting the path dependence in setting state tax rates.

6. Conclusions and Discussions

We consider fixed effects estimation of spatial panel data models with missing responses, allowing unobserved spatiotemporal heterogeneity, time-varying endogenous and contextual spatial interactions, time-varying cross-sectional error dependence, and serial correlation. We propose an M-estimation method for model estimation and a corrected plug-in method for model inference, both taking into account the effects of estimating the fixed effects. We study the asymptotic and finite sample properties of the proposed methods. We apply our methods to US state tax competition data, which leads to a much deeper understanding on tax competition mechanism. An important feature of the proposed method is that it allows the estimation of time-invariant covariate effects, such as gender, by imposing relevant constraints on the $\mathbf{D}$ matrix in Model (2.1).

The proposed methods apply to matrix exponential spatial specification (MESS) by replacing, in Model (1.2), $I_n - \lambda W_t$ by $\exp(\lambda W_t) = \sum_{i=0}^{\infty}(\lambda W_t)^i/i!$ and $I_n - \rho M_t$ by $\exp(\rho M_t) = \sum_{i=0}^{\infty}(\rho M_t)^i/i!$, and can be easily extended to allow for a high-order MA process for serial correlation. They can be further extended to allow for high-order spatial effects by replacing $I_n - \lambda W_t$ by $I_n - \sum_{l=1}^{p} \lambda_l W_{lt}$ and $I_n - \rho M_t$ by $I_n - \sum_{e=1}^{p} \rho_e M_{et}$. Details on these are available from the authors upon request. Extending MESS to high order runs into a computational issue as the partial derivatives do not possess analytical form.

Acknowledgments

We thank the participants of AMES 2021, Curtin University, Malaysia, and the seminar participants at WISE, Xiamen University, China, for their helpful comments. Zhenlin Yang gratefully acknowledges the research support from Singapore Management University.

Appendix A: AQS, Hessian, and Variance of AQS

A.1. Derivation for Section 2

AQS function. Write $\tilde{\mathbb{V}}'(\beta,\delta)\tilde{\mathbb{V}}(\beta,\delta) = \mathcal{V}'(\beta,\lambda)\Psi(\delta)\mathcal{V}(\beta,\lambda)$, where $\mathcal{V}(\beta,\lambda) = \mathcal{S}[\mathbf{Y} - \mathbf{A}_{nT}^{-1}(\lambda)\mathbf{X}\beta]$ and $\Psi(\delta) = \mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)\mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)$. Letting $\mathcal{D}(\lambda) = \mathcal{S}\mathbf{A}_{nT}^{-1}(\lambda)\mathbf{D}$, then,

$$\Psi(\delta) = \mathbf{\Omega}_N^{-1}(\delta) - \mathbf{\Omega}_N^{-1}(\delta)\mathcal{D}(\lambda)[\mathcal{D}'(\lambda)\mathbf{\Omega}_N^{-1}(\delta)\mathcal{D}(\lambda)]^{-1}\mathcal{D}'(\lambda)\mathbf{\Omega}_N^{-1}(\delta), \tag{A.1}$$

which allows the use of the matrix result: $\frac{\partial}{\partial\omega}\mathbf{\Omega}_N^{-1}(\delta) = -\mathbf{\Omega}_N^{-1}(\delta)[\frac{\partial}{\partial\omega}\mathbf{\Omega}(\delta)]\mathbf{\Omega}_N^{-1}(\delta), \omega = \lambda, \rho$.

Denoting $\dot{\Psi}_\omega(\delta) \equiv \frac{\partial}{\partial\omega}\Psi(\delta), \omega = \lambda, \rho$, we obtain, after some tedious algebra:

$$\dot{\Psi}_\lambda(\delta) = -\mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)\mathbb{H}_\lambda(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)\mathbf{\Omega}_N^{-\frac{1}{2}}(\delta) - \Psi(\delta)\mathbb{K}(\delta) - \mathbb{K}'(\delta)\Psi(\delta), \tag{A.2}$$

$$\dot{\Psi}_\rho(\delta) = -\mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)\mathbb{H}_\rho(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)\mathbf{\Omega}_N^{-\frac{1}{2}}(\delta), \tag{A.3}$$

where $\mathbb{K}(\delta) = \mathcal{S}[\frac{\partial}{\partial\lambda}\mathbf{A}_{nT}^{-1}(\lambda)]\mathbf{D}[\mathbb{D}'(\delta)\mathbb{D}(\delta)]^{-1}\mathbb{D}'(\delta)\mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)$. These lead immediately to the CQS function (2.6) and thus the AQS function (2.7).

Hessian matrix. To derive $H_N^*(\theta) = \frac{\partial}{\partial\theta'}S_N^*(\theta)$, let $\dot{\mathbf{\Omega}}_\omega(\delta)$ and $\ddot{\mathbf{\Omega}}_{\omega\varpi}(\delta)$ be the 1st- and 2nd-order partial derivatives of $\mathbf{\Omega}(\delta)$, $\omega, \varpi = \lambda, \rho$; similarly are $\dot{\mathbf{\Psi}}_\omega(\delta)$ and $\ddot{\mathbf{\Psi}}_{\omega\varpi}(\delta)$ defined. Denoting $\mathbb{J}(\delta) = \mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)\mathcal{S}[\frac{\partial}{\partial\lambda}\mathbf{A}_{nT}^{-1}(\lambda)]$, we obtain the components of $H_N^*(\theta)$:

$H_{\beta\beta}^*(\theta) = -\frac{1}{\sigma_v^2}\mathbb{X}'(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)\mathbb{X}(\delta)$, $\quad H_{\beta\sigma_v^2}^*(\theta) = -\frac{1}{\sigma_v^4}\mathbb{X}'(\rho)\tilde{\mathbb{V}}(\beta,\delta) = H_{\sigma_v^2\beta}^{*\prime}(\theta)$,

$H_{\beta\lambda}^*(\theta) = \frac{1}{\sigma_v^2}\mathbf{X}'\mathbb{J}'(\delta)\tilde{\mathbb{V}}(\beta,\delta) + \frac{1}{\sigma_v^2}\mathbf{X}'\mathbf{A}_{nT}^{-1\prime}(\lambda)\mathcal{S}'\dot{\Psi}_\lambda(\delta)\mathcal{V}(\beta,\lambda) - \frac{1}{\sigma_v^2}\mathbb{X}'(\rho)\mathbb{J}(\delta)\mathbf{X}\beta = H_{\lambda\beta}^{*\prime}(\theta)$,

$H_{\beta\rho}^*(\theta) = \frac{1}{\sigma_v^2}\mathbf{X}'\mathbf{A}_{nT}^{-1\prime}(\lambda)\mathcal{S}'\dot{\Psi}_\rho(\delta)\mathcal{V}(\beta,\lambda) = H_{\rho\beta}^{*\prime}(\theta)$,

$H_{\sigma_v^2\sigma_v^2}^*(\theta) = -\frac{1}{\sigma_v^6}\tilde{\mathbb{V}}'(\beta,\delta)\tilde{\mathbb{V}}(\beta,\delta) + \frac{1}{2\sigma_v^4}N_1$,

$H_{\sigma_v^2\lambda}^*(\theta) = \frac{1}{2\sigma_v^4}\mathcal{V}'(\beta,\lambda)\dot{\Psi}_\lambda(\delta)\mathcal{V}(\beta,\lambda) - \frac{1}{\sigma_v^4}\tilde{\mathbb{V}}'(\beta,\delta)\mathbb{J}(\delta)\mathbf{X}\beta = H_{\lambda\sigma_v^2}^{*\prime}(\theta)$,

$H_{\sigma_v^2\rho}^*(\theta) = \frac{1}{2\sigma_v^4}\mathcal{V}'(\beta,\lambda)\dot{\Psi}_\rho(\delta)\mathcal{V}(\beta,\lambda) = H_{\rho\sigma_v^2}^{*\prime}(\theta)$,

$H_{\lambda\lambda}^*(\theta) = \frac{2}{\sigma_v^2}\mathcal{V}'(\beta,\lambda)\dot{\Psi}_\lambda(\delta)\mathcal{S}[\frac{\partial}{\partial\lambda}\mathbf{A}_{nT}^{-1}(\lambda)]\mathbf{X}\beta + \frac{2}{\sigma_v^2}\tilde{\mathbb{V}}'(\beta,\delta)\mathbb{J}(\delta)\mathbf{W}_{nT}\mathbf{A}_{nT}^{-1}(\lambda)\mathbf{X}\beta$

$$-\tfrac{1}{\sigma_v^2}\beta'\mathbf{X}'\mathbb{J}'(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)\mathbb{J}(\delta)\mathbf{X}\beta - \tfrac{1}{2\sigma_v^2}\mathcal{V}'(\beta,\lambda)\ddot{\Psi}_{\lambda\lambda}(\delta)\mathcal{V}(\beta,\lambda)$$

$$-\tfrac{1}{2}\mathtt{tr}[\dot{\mathbf{\Omega}}_\lambda(\delta)\dot{\Psi}_\lambda(\delta) + \ddot{\mathbf{\Omega}}_{\lambda\lambda}(\delta)\Psi(\delta)],$$

$$H^*_{\lambda\rho}(\theta) = -\tfrac{1}{2\sigma_v^2}\mathcal{V}'(\beta,\lambda)\ddot{\Psi}_{\lambda\rho}(\delta)\mathcal{V}(\beta,\lambda) + \tfrac{1}{\sigma_v^2}\mathcal{V}'(\beta,\lambda)\dot{\Psi}_\rho(\delta)\mathcal{S}[\tfrac{\partial}{\partial\lambda}\mathbf{A}_{nT}^{-1}(\lambda)]\mathbf{X}\beta$$

$$-\tfrac{1}{2}\mathtt{tr}[\dot{\mathbf{\Omega}}_\lambda(\delta)\dot{\Psi}_\rho(\delta) + \ddot{\mathbf{\Omega}}_{\lambda\rho}(\delta)\Psi(\delta)],$$

$$H^*_{\rho\lambda}(\theta) = -\tfrac{1}{2\sigma_v^2}\mathcal{V}'(\beta,\lambda)\ddot{\Psi}_{\lambda\rho}(\delta)\mathcal{V}(\beta,\lambda) + \tfrac{1}{\sigma_v^2}\mathcal{V}'(\beta,\lambda)\dot{\Psi}_\rho(\delta)\mathcal{S}[\tfrac{\partial}{\partial\lambda}\mathbf{A}_{nT}^{-1}(\lambda)]\mathbf{X}\beta$$

$$-\tfrac{1}{2}\mathtt{tr}[\dot{\mathbf{\Omega}}_\rho(\delta)\dot{\Psi}_\lambda(\delta) + \ddot{\mathbf{\Omega}}_{\lambda\rho}(\delta)\Psi(\delta)],$$

$$H^*_{\rho\rho}(\theta) = -\tfrac{1}{2\sigma_v^2}\mathcal{V}'(\beta,\lambda)\ddot{\Psi}_{\rho\rho}(\delta)\mathcal{V}(\beta,\lambda) - \tfrac{1}{2}\mathtt{tr}[\dot{\mathbf{\Omega}}_\rho(\delta)\dot{\Psi}_\rho(\delta) + \ddot{\mathbf{\Omega}}_{\rho\rho}(\delta)\Psi(\delta)].$$

VC matrix. For stochastic terms of the forms in (2.10), one easily shows,

(i) $\mathrm{Cov}(\Pi_r'\mathbf{V},\ \Pi_s'\mathbf{V}) = \sigma_{v0}^2\Pi_r'\Pi_s, r,s = 1,2,\ldots;$

(ii) $\mathrm{Cov}(\mathbf{V}'\Phi_r\mathbf{V},\ \Pi_s'\mathbf{V}) = \sigma_{v0}^3\kappa_3\varphi_r'\Pi_s, r = 1,2,...,s = 1,2,\ldots;$ and

(iii) $\mathrm{Cov}(\mathbf{V}'\Phi_r\mathbf{V},\ \mathbf{V}'\Phi_s\mathbf{V}) = \sigma_{v0}^4\kappa_4\varphi_r'\phi_s + \sigma_{v0}^4\mathtt{tr}(\Phi_r\Phi_s^\circ), r,s = 1,2,\ldots,$

where $\varphi_r = \mathtt{diagv}(\Phi_r)$, for $r = 1,2,\ldots.$

Apply these results to (2.10), we obtain,

$$\mathrm{Var}[S_N^*(\theta_0)] = \frac{1}{\sigma_{v0}^2}\begin{pmatrix} \Pi_1'\Pi_1, & \frac{1}{\sigma_0}\kappa_3\Pi_1'\varphi_1, & \Pi_1'\Pi_2 + \sigma_0\kappa_3\Pi_1'\varphi_2, & \sigma_0\kappa_3\Pi_1'\varphi_3 \\ \sim, & \frac{1}{\sigma_0^2}\Xi_{11}, & \Xi_{12}, & \frac{1}{\sigma_0^2}\Xi_{13} \\ \sim, & \sim, & \Xi_{22} + \Pi_2'\Pi_2 + 2\sigma_0\kappa_3\Pi_2'\varphi_2, & \Xi_{23} + \sigma_0\kappa_3\Pi_2'\varphi_3 \\ \sim, & \sim, & \sim, & \Xi_{33} \end{pmatrix}$$

where $\Xi_{rs} = \mathtt{tr}(\Phi_r\Phi_s^\circ) + \kappa_4\varphi_r'\varphi_s,\ r,s = 1,2,3.$

A.2. Derivation for Section 3.

Hessian matrix. With $\mathbf{\Omega}_N(\delta) = \mathcal{S}\mathbf{A}_{nT}^{-1}(\lambda)\mathbf{B}_{nT}^{-1}(\rho)(\Upsilon'(\tau)\Upsilon(\tau)\otimes I_n)\mathbf{B}_{nT}^{-1\prime}(\rho)\mathbf{A}_{nT}^{-1\prime}(\lambda)\mathcal{S}'$, the non-$\tau$-components of $H_N^\diamond(\theta) = \frac{\partial}{\partial\theta'}S_N^*(\theta)$ remain in the same form as those of $H_N^*(\theta)$ in Section 2. Extending the notations, $\dot{\mathbf{\Omega}}_\omega(\delta)$, $\ddot{\mathbf{\Omega}}_{\omega\varpi}(\delta)$, $\dot{\mathbf{\Psi}}_\omega(\delta)$, and $\ddot{\mathbf{\Psi}}_{\omega\varpi}(\delta)$ of Section 2 to $\omega,\varpi = \lambda,\rho,\tau$, we obtain the τ-components of $H_N^\diamond(\theta)$:

$H^{\diamond}_{\beta\tau}(\theta) = \frac{1}{\sigma_e^2}\mathbf{X}'\mathbf{A}^{-1\prime}_{nT}(\lambda)\mathcal{S}'\dot{\Psi}_{\tau}(\delta)\mathcal{V}(\beta,\lambda) = H^{\diamond\prime}_{\tau\beta}(\theta),$

$H^{\diamond}_{\sigma_e^2\tau}(\theta) = \frac{1}{2\sigma_e^4}\mathcal{V}'(\beta,\lambda)\dot{\Psi}_{\tau}(\delta)\mathcal{V}(\beta,\lambda) = H^{\diamond\prime}_{\tau\sigma_e^2}(\theta),$

$$H^{\diamond}_{\lambda\tau}(\theta) = -\frac{1}{2\sigma_e^2}\mathcal{V}'(\beta,\lambda)\ddot{\Psi}_{\lambda\tau}(\delta)\mathcal{V}(\beta,\lambda) + \frac{1}{\sigma_e^2}\mathcal{V}'(\beta,\lambda)\dot{\Psi}_{\tau}(\delta)\mathcal{S}[\tfrac{\partial}{\partial\lambda}\mathbf{A}^{-1}_{nT}(\lambda)]\mathbf{X}\beta$$
$$-\frac{1}{2}\mathtt{tr}[\dot{\mathbf{\Omega}}_{\lambda}(\delta)\dot{\Psi}_{\tau}(\delta) + \ddot{\mathbf{\Omega}}_{\lambda\tau}(\delta)\Psi(\delta)],$$

$$H^{\diamond}_{\tau\lambda}(\theta) = -\frac{1}{2\sigma_e^2}\mathcal{V}'(\beta,\lambda)\ddot{\Psi}_{\lambda\tau}(\delta)\mathcal{V}(\beta,\lambda) + \frac{1}{\sigma_e^2}\mathcal{V}'(\beta,\lambda)\dot{\Psi}_{\tau}(\delta)\mathcal{S}[\tfrac{\partial}{\partial\lambda}\mathbf{A}^{-1}_{nT}(\lambda)]\mathbf{X}\beta$$
$$-\frac{1}{2}\mathtt{tr}[\dot{\mathbf{\Omega}}_{\tau}(\delta)\dot{\Psi}_{\lambda}(\delta) + \ddot{\mathbf{\Omega}}_{\lambda\tau}(\delta)\Psi(\delta)],$$

$$H^{\diamond}_{\rho\tau}(\theta) = -\frac{1}{2\sigma_e^2}\mathcal{V}'(\beta,\lambda)\ddot{\Psi}_{\rho\tau}(\delta)\mathcal{V}(\beta,\lambda) - \frac{1}{2}\mathtt{tr}[\mathbf{\Omega}_{\rho}(\delta)\dot{\Psi}_{\tau}(\delta) + \ddot{\mathbf{\Omega}}_{\rho\tau}(\delta)\Psi(\delta)],$$

$$H^{\diamond}_{\tau\rho}(\theta) = -\frac{1}{2\sigma_e^2}\mathcal{V}'(\beta,\lambda)\ddot{\Psi}_{\tau\rho}(\delta)\mathcal{V}(\beta,\lambda) - \frac{1}{2}\mathtt{tr}[\dot{\mathbf{\Omega}}_{\tau}(\delta)\dot{\Psi}_{\rho}(\delta) + \ddot{\mathbf{\Omega}}_{\tau\rho}(\delta)\Psi(\delta)],$$

$$H^{\diamond}_{\tau\tau}(\theta) = -\frac{1}{2\sigma_e^2}\mathcal{V}'(\beta,\lambda)\ddot{\Psi}_{\tau\tau}(\delta)\mathcal{V}(\beta,\lambda) - \frac{1}{2}\mathtt{tr}[\dot{\mathbf{\Omega}}_{\tau}(\delta)\dot{\Psi}_{\tau}(\delta) + \ddot{\mathbf{\Omega}}_{\tau\tau}(\delta)\Psi(\delta)].$$

VC matrix. Applying the results (i)-(iii) in Section A.1 to (3.2), we obtain,

$$\mathrm{Var}[S^{\diamond}_N(\theta_0)]$$
$$= \frac{1}{\sigma_{e0}^2}\begin{pmatrix} \Pi_1'\Pi_1, & \frac{1}{\sigma_{e0}}\kappa_3\Pi_1'\varphi_1, & \Pi_1'\Pi_2 + \sigma_{e0}\kappa_3\Pi_1'\varphi_2, & \sigma_{e0}\kappa_3\Pi_1'\varphi_3, & \sigma_{e0}\kappa_3\Pi_1'\varphi_4 \\ \sim, & \frac{1}{\sigma_{e0}^2}\Xi_{11}, & \Xi_{12}, & \frac{1}{\sigma_{e0}^2}\Xi_{13}, & \frac{1}{\sigma_{e0}^2}\Xi_{14} \\ \sim, & \sim, & \Xi_{22} + \Pi_2'\Pi_2 + 2\sigma_{e0}\kappa_3\Pi_2'\varphi_2, & \Xi_{23} + \sigma_{e0}\kappa_3\Pi_2'\varphi_3, & \Xi_{24} + \sigma_{e0}\kappa_3\Pi_2'\varphi_4 \\ \sim, & \sim, & \sim, & \Xi_{33}, & \Xi_{34} \\ \sim, & \sim, & \sim, & \sim, & \Xi_{44} \end{pmatrix}$$

where $\Xi_{rs} = \mathtt{tr}(\Phi_r\Phi_s^{\circ}) + \kappa_4\varphi_r'\varphi_s,\ r, s = 1, 2, 3, 4.$

Appendix B: Some Basic Lemmas

The following lemmas, existing or new, are essential to the proofs of main results in Sections 2 and 3. The proofs of the new lemmas are given in `Online Appendix`.

Lemma B.1. *(Kelejian and Prucha, 1999): Let $\{A_N\}$ and $\{B_N\}$ be two sequences of $N \times N$ matrices that are bounded in both r-norm and c-norm. 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 bounded in both row and column sum norms,*

(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 B.2. *Under the setup of Section 2 and Assumptions C-F, the following matrices are all bounded in both row and column sum norms, uniformly in* $\delta \in \Delta$*: (i)* $\mathbf{\Omega}_N(\delta)$, $\dot{\mathbf{\Omega}}_\omega(\delta) \equiv \frac{\partial}{\partial\omega}\mathbf{\Omega}_N(\delta)$, $\omega = \lambda, \rho$, $\mathbf{\Omega}_N^{-1}(\delta)$, *(ii)* $\mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)\mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)$, *and (iii)* $\mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)\mathbb{P}_{\tilde{\mathbb{X}}}(\delta)\mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)$, *where* $\mathbb{P}_{\tilde{\mathbb{X}}}(\delta)$ *is the projection matrix based on* $\tilde{\mathbb{X}}(\delta) = \mathbb{Q}_{\mathbb{D}}(\delta)\mathbb{X}(\delta)$.

Lemma B.3. *Under Assumptions C-E,* $\mathtt{tr}[A_N\mathbf{X}[\mathbb{X}'(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)\mathbb{X}(\delta)]^{-1}\mathbf{X}'B_N] = O(1)$, *uniformly in* $\delta \in \Delta$, *for* A_N *and* B_N *bounded in row or column sum norm.*

Lemma B.4. *Let* A_N *be an* $N \times N$ *matrix bounded in row and column sum norm, with elements being* $O(h_N^{-1})$ *uniformly in* i *and* j. *Let* $\mathbf{V} = (v_1, \cdots, v_N)'$ *and* $v_j \sim iid(0, \sigma^2)$. *Let* c_N *be an* $N \times 1$ *vector with elements of uniform order* $O(h_n^{-1/2})$. *Then,*

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

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

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

Appendix C: Proofs for Section 2

Population objective function. The population counterpart of $S_N^{*c}(\delta)$ is

$$\bar{S}_N^{*c}(\delta) = \begin{cases} \dfrac{\mathrm{E}[\bar{\mathbb{V}}'(\delta)\mathbb{H}_\lambda(\delta)\bar{\mathbb{V}}(\delta)]}{2\mathrm{E}[\bar{\mathbb{V}}'(\delta)\bar{\mathbb{V}}(\delta)]/N_1} + \dfrac{\mathrm{E}[\bar{\mathbb{V}}'(\delta)\mathbb{J}(\delta)\boldsymbol{\varepsilon}(\bar{\beta}_{\mathtt{M}}(\delta),\delta)]}{\mathrm{E}[\bar{\mathbb{V}}'(\delta)\bar{\mathbb{V}}(\delta)]/N_1} - \dfrac{1}{2}\mathtt{tr}[\mathbb{H}_\lambda(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)], \\ \dfrac{\mathrm{E}[\bar{\mathbb{V}}'(\delta)\mathbb{H}_\rho(\delta)\bar{\mathbb{V}}(\delta)]}{2\mathrm{E}[\bar{\mathbb{V}}'(\delta)\bar{\mathbb{V}}(\delta)]/N_1} - \dfrac{1}{2}\mathtt{tr}[\mathbb{H}_\rho(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)], \end{cases} \tag{C.1}$$

where $\bar{\mathbb{V}}(\delta) = \tilde{\mathbb{V}}(\bar{\beta}_{\mathtt{M}}(\delta), \delta)$, obtained by first solving $\bar{S}_N^*(\theta) = \mathrm{E}[S_N^*(\theta)]$ for β and σ^2:

$$\bar{\beta}_{\mathtt{M}}(\delta) = [\mathbb{X}'(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)\mathbb{X}(\delta)]^{-1}\mathbb{X}'(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)\mathrm{E}[\mathbb{Y}(\delta)] \quad \text{and} \quad \bar{\sigma}_{v,\mathtt{M}}^2(\delta) = \tfrac{1}{N_1}\mathrm{E}[\bar{\mathbb{V}}'(\delta)\bar{\mathbb{V}}(\delta)], \tag{C.2}$$

and then substituting $\bar{\beta}_{\mathtt{M}}(\delta)$ and $\bar{\sigma}^2_{v,\mathtt{M}}(\delta)$ back into the δ-component of $\bar{S}^*_N(\theta)$.

More on Assumption G. By (C.2), we have $\bar{\mathbb{V}}(\delta) = \mathbb{Q}_{\mathbb{D}}(\delta)\mathbb{Y}(\delta) - \mathbb{Q}_{\mathbb{D}}(\delta)\mathbb{X}(\delta)\bar{\beta}_{\mathtt{M}}(\delta) = \mathbb{Q}_{\tilde{\mathbb{X}}}(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)\mathbb{Y}(\delta) + \mathbb{P}_{\tilde{\mathbb{X}}}(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)[\mathbb{Y}(\delta) - \mathrm{E}(\mathbb{Y}(\delta))]$, where $\mathbb{P}_{\tilde{\mathbb{X}}}(\delta)$ and $\mathbb{Q}_{\tilde{\mathbb{X}}}(\delta)$ are the projection matrices based on $\tilde{\mathbb{X}}(\delta) = \mathbb{Q}_{\mathbb{D}}(\delta)\mathbb{X}(\delta)$. Let $\eta = \mathcal{S}\mathbf{A}_{nT}^{-1}(\mathbf{X}\beta_0 + \mathbf{D}\phi_0)$. As $\mathbb{Y}(\delta) = \mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)\eta + \mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)\mathcal{S}\mathbf{A}_{nT}^{-1}\mathbf{B}_{nT}^{-1}\mathbf{V}$, we have by orthogonality between $\mathbb{Q}_{\tilde{\mathbb{X}}}(\delta)$ and $\mathbb{P}_{\tilde{\mathbb{X}}}(\delta)$,

$$
\begin{aligned}
\bar{\sigma}^2_{v,\mathtt{M}}(\delta) &= \tfrac{1}{N_1}\mathrm{E}[\bar{\mathbb{V}}'(\delta)\bar{\mathbb{V}}(\delta)] \\
&= \tfrac{1}{N_1}\mathrm{E}[\mathbb{Y}'(\delta)\mathbf{Q}(\delta)\mathbb{Y}(\delta)] + \tfrac{1}{N_1}\mathrm{E}\big\{[\mathbb{Y}(\delta) - \mathrm{E}(\mathbb{Y}(\delta))]'\mathbf{P}(\delta)[\mathbb{Y}(\delta) - \mathrm{E}(\mathbb{Y}(\delta))]\big\} \qquad \text{(C.3)} \\
&= \tfrac{1}{N_1}\mathrm{E}(\mathbb{Y}(\delta))'\mathbf{Q}(\delta)\mathrm{E}(\mathbb{Y}(\delta)) + \tfrac{1}{N_1}\mathrm{E}\big\{[\mathbb{Y}(\delta) - \mathrm{E}(\mathbb{Y}(\delta))]'[\mathbf{Q}(\delta) + \mathbf{P}(\delta)][\mathbb{Y}(\delta) - \mathrm{E}(\mathbb{Y}(\delta))]\big\} \\
&= \tfrac{1}{N_1}\mathrm{E}(\mathbb{Y}(\delta))'\mathbf{Q}(\delta)\mathrm{E}(\mathbb{Y}(\delta)) + \tfrac{1}{N_1}\mathrm{E}\big\{[\mathbb{Y}(\delta) - \mathrm{E}(\mathbb{Y}(\delta))]'\mathbb{Q}_{\mathbb{D}}(\delta)[\mathbb{Y}(\delta) - \mathrm{E}(\mathbb{Y}(\delta))]\big\} \\
&= \tfrac{1}{N_1}\eta'\mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)\mathbf{Q}(\delta)\mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)\eta + \tfrac{\sigma^2_{v0}}{N_1}\mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\delta)\mathcal{O}_N(\delta)],
\end{aligned}
$$

where $\mathbf{Q}(\delta) = \mathbb{Q}_{\mathbb{D}}(\delta)\mathbb{Q}_{\tilde{\mathbb{X}}}(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)$, $\mathbf{P}(\delta) = \mathbb{Q}_{\mathbb{D}}(\delta)\mathbb{P}_{\tilde{\mathbb{X}}}(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)$ and $\mathcal{O}_N(\delta) = \mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)\mathbf{\Omega}_N\mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)$.

Denote $\mathcal{Q}_N(\delta) = \mathbb{Q}_{\tilde{\mathbb{X}}}(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)\mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)$ and $\mathbb{D}^\dagger(\delta) = \mathbf{D}[\mathbb{D}'(\delta)\mathbb{D}(\delta)]^{-1}\mathbb{D}'(\delta)$. A more primitive condition for Assumption G to hold is that either (a) or (b) holds for $\delta \neq \delta_0$:

(a) $\frac{1}{2\bar{\sigma}^2_{v,\mathtt{M}}(\delta)}\eta'\mathcal{Q}_N'(\delta)\mathbb{H}_\lambda(\delta)\mathcal{Q}_N(\delta)\eta + \frac{1}{\bar{\sigma}^2_{v,\mathtt{M}}(\delta)}\eta'\mathcal{Q}_N'(\delta)\mathbb{J}(\delta)[\mathbb{D}^\dagger(\delta)\mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)\eta - \mathbb{X}(\delta)\bar{\beta}_{\mathtt{M}}(\delta) + \mathbf{X}\bar{\beta}_{\mathtt{M}}(\delta)] + \frac{\sigma^2_{v0}}{\bar{\sigma}^2_{v,\mathtt{M}}(\delta)}\mathtt{tr}[\mathbb{J}(\delta)\mathbb{D}^\dagger(\delta)\mathcal{O}_N(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)] + \frac{1}{2}\mathtt{tr}[\frac{\sigma^2_{v0}}{\bar{\sigma}^2_{v,\mathtt{M}}(\delta)}\mathbb{H}_\lambda(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)\mathcal{O}_N(\delta)\mathbb{Q}_{\mathbb{D}}(\delta) - \mathbb{H}_\lambda(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)] \neq 0$,

(b) $\frac{1}{2\bar{\sigma}^2_{v,\mathtt{M}}(\delta)}\eta'\mathcal{Q}_N'(\delta)\mathbb{H}_\rho(\delta)\mathcal{Q}_N(\delta)\eta + \frac{1}{2}\mathtt{tr}[\frac{\sigma^2_{v0}}{\bar{\sigma}^2_{v,\mathtt{M}}(\delta)}\mathbb{H}_\rho(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)\mathcal{O}_N(\delta)\mathbb{Q}_{\mathbb{D}}(\delta) - \mathbb{H}_\rho(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)] \neq 0$.

As $\mathcal{Q}_N(\delta_0)\eta = 0$, $\mathcal{O}_N(\delta_0) = I_N$, $\bar{\beta}_{\mathtt{M}}(\delta_0) = \beta_0$ and $\bar{\sigma}^2_{v,\mathtt{M}}(\delta_0) = \sigma^2_{v0}$, the two quantities in (a) and (b) are both 0 when $\delta = \delta_0$.

Proof of Theorem 2.1: By theorem 5.9 of Van der Vaart (1998), we only need to show $\sup_{\delta\in\delta}\frac{1}{N_1}\left\|S^{*c}_N(\delta) - \bar{S}^{*c}_N(\delta)\right\| \xrightarrow{p} 0$ under the assumptions in Theorem 2.1. From (2.9) and (C.1), the consistency of $\hat{\delta}_{\mathtt{M}}$ follows from:

(a) $\inf_{\delta\in\Delta}\bar{\sigma}^2_{v,\mathtt{M}}(\delta)$ is bounded away from zero,

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

(c) $\sup_{\delta\in\Delta}\frac{1}{N_1}\left|\hat{\mathbb{V}}'(\delta)\mathbb{H}_\omega(\delta)\hat{\mathbb{V}}(\delta)-\mathrm{E}[\bar{\mathbb{V}}'(\delta)\mathbb{H}_\omega(\delta)\bar{\mathbb{V}}(\delta)]\right| = o_p(1)$, for $\omega=\lambda,\rho$,

(d) $\sup_{\delta\in\Delta}\frac{1}{N_1}\left|\hat{\mathbb{V}}'(\delta)\mathbb{J}(\delta)\boldsymbol{\varepsilon}(\hat{\beta}_{\mathtt{M}}(\delta),\delta)-\mathrm{E}[\bar{\mathbb{V}}'(\delta)\mathbb{J}(\delta)\boldsymbol{\varepsilon}(\bar{\beta}_{\mathtt{M}}(\delta),\delta)]\right| = o_p(1)$.

Proof of (a). Note that $\bar{\sigma}^2_{v,\mathtt{M}}(\delta) = \frac{1}{N_1}\eta'\mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)\mathbf{Q}(\delta)\mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)\eta + \frac{\sigma^2_{v0}}{N_1}\mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\delta)\mathcal{O}_N(\delta)]$. 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,

$$\frac{\sigma^2_{v0}}{N_1}\mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\delta)\mathcal{O}_N(\delta)] \geq \frac{\sigma^2_{v0}}{N_1}\gamma_{\min}[\mathcal{O}_N(\delta)]\mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\delta)] \geq \sigma^2_{v0}\gamma_{\max}(\mathbf{\Omega}_N)^{-1}\gamma_{\min}[\mathbf{\Omega}_N(\delta)]$$

$$\geq \sigma^2_{v0}\gamma_{\max}(\mathbf{A}'_N\mathbf{A}_N)^{-1}\gamma_{\max}(\mathbf{B}'_N\mathbf{B}_N)^{-1}\gamma_{\min}[\mathbf{A}'_N(\lambda)\mathbf{A}_N(\lambda)]\gamma_{\min}[\mathbf{B}'_N(\rho)\mathbf{B}_N(\rho)] > 0,$$

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

Proof of (b). From (2.8), $\hat{\mathbb{V}}(\delta)=\mathbb{Q}_{\mathbb{D}}(\delta)[\mathbb{Y}(\delta)-\mathbb{X}(\delta)\hat{\beta}_{\mathtt{M}}(\delta)] = \mathbb{Q}_{\tilde{\mathbb{X}}}(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)\mathbb{Y}(\delta)$ and $\hat{\sigma}^2_{v,\mathtt{M}}(\delta)=\frac{1}{N_1}\mathbb{Y}'(\delta)\mathbf{Q}(\delta)\mathbb{Y}(\delta)$. From (C.3), $\bar{\sigma}^2_{v,\mathtt{M}}(\delta)=\frac{1}{N_1}\mathrm{E}[\mathbb{Y}'(\delta)\mathbf{Q}(\delta)\mathbb{Y}(\delta)]+\frac{\sigma^2_{v0}}{N_1}\mathtt{tr}[\mathbf{P}(\delta)\mathcal{O}_N(\delta)]$. Thus, $\hat{\sigma}^2_{v,\mathtt{M}}(\delta)-\bar{\sigma}^2_{v,\mathtt{M}}(\delta)=\frac{1}{N_1}[\mathbb{Y}'(\delta)\mathbf{Q}(\delta)\mathbb{Y}(\delta)-\mathrm{E}(\mathbb{Y}'(\delta)\mathbf{Q}(\delta)\mathbb{Y}(\delta))]-\frac{\sigma^2_{v0}}{N_1}\mathtt{tr}[\mathbf{P}(\delta)\mathcal{O}_N(\delta)]$.

For the second term, $0\leq\frac{1}{N_1}\mathtt{tr}[\mathbf{P}(\delta)\mathcal{O}_N(\delta)]\leq\frac{1}{N_1}\gamma_{\max}[\mathcal{O}_N(\delta)]\gamma^2_{\max}[\mathbb{Q}_{\mathbb{D}}(\delta)]\mathtt{tr}[\mathbb{P}_{\tilde{\mathbb{X}}}(\delta)] = o(1)$, because $\mathtt{tr}[\mathbb{P}_{\tilde{\mathbb{X}}}(\delta)]=k$, $\gamma_{\max}[\mathbb{Q}_{\mathbb{D}}(\delta)]=1$ and, by Assumption E(iii),

$$\gamma_{\max}[\mathcal{O}_N(\delta)]\leq\gamma_{\min}(\mathbf{A}'_N\mathbf{A}_N)^{-1}\gamma_{\min}(\mathbf{B}'_N\mathbf{B}_N)^{-1}\gamma_{\max}[\mathbf{A}'_N(\lambda)\mathbf{A}_N(\lambda)]\gamma_{\max}[\mathbf{B}'_N(\rho)\mathbf{B}_N(\rho)]<\infty.$$

Therefore, one has $\sup_{\delta\in\Delta}|\frac{\sigma^2_{v0}}{N_1}\mathtt{tr}[\mathbf{P}(\delta)\mathcal{O}_N(\delta)]|=o(1)$. For the first term, letting $\bar{\mathbf{Q}}(\delta)=\mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)\mathbf{Q}(\delta)\mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)$ and using $\mathcal{S}\mathbf{Y}=\eta+\mathcal{S}\mathbf{A}^{-1}_{nT}\mathbf{B}^{-1}_{nT}\mathbf{V}$, we have

$$\frac{1}{N_1}[\mathbb{Y}'(\delta)\mathbf{Q}(\delta)\mathbb{Y}(\delta)-\mathrm{E}(\mathbb{Y}'(\delta)\mathbf{Q}(\delta)\mathbb{Y}(\delta))]=\frac{1}{N_1}[\mathbf{Y}'\mathcal{S}'\bar{\mathbf{Q}}(\delta)\mathcal{S}\mathbf{Y}-\mathrm{E}(\mathbf{Y}'\mathcal{S}'\bar{\mathbf{Q}}(\delta)\mathcal{S}\mathbf{Y})]$$

$$=\frac{2}{N_1}\eta'\bar{\mathbf{Q}}(\delta)\mathcal{S}\mathbf{A}^{-1}_{nT}\mathbf{B}^{-1}_{nT}\mathbf{V}+\frac{1}{N_1}[\mathbf{V}'\mathbf{B}^{-1\prime}_{nT}\mathbf{A}^{-1\prime}_{nT}\mathcal{S}\bar{\mathbf{Q}}(\delta)\mathcal{S}\mathbf{A}^{-1}_{nT}\mathbf{B}^{-1}_{nT}\mathbf{V}-\sigma^2_{v0}\mathtt{tr}(\bar{\mathbf{Q}}(\delta)\mathbf{\Omega}_N)].$$

Thus, the pointwise convergence of the first term follows from Lemma B.4(v), and the pointwise convergence of the second term follows from Lemma B.4(iv). Therefore, $\frac{1}{N_1}[\mathbb{Y}'(\delta)\mathbf{Q}(\delta)\mathbb{Y}(\delta)-\mathrm{E}(\mathbb{Y}'(\delta)\mathbf{Q}(\delta)\mathbb{Y}(\delta))]\xrightarrow{p}0$, for each $\delta\in\Delta$.

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

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

where $\bar{\delta}$ lies between δ_1 and δ_2. It follows that $\frac{1}{N_1}\mathbb{Y}'(\delta)\mathbf{Q}(\delta)\mathbb{Y}(\delta)$ is stochastically equicontinuous as $\sup_{\delta\in\Delta}\frac{1}{N_1}\mathbf{Y}'\mathcal{S}'[\frac{\partial}{\partial\varpi}\bar{\mathbf{Q}}(\delta)]\mathcal{S}\mathbf{Y} = O_p(1)$, $\varpi = \lambda, \rho$ (See `Online Appendix` for details). With the pointwise convergence of $\frac{1}{N_1}[\mathbb{Y}'(\delta)\mathbf{Q}(\delta)\mathbb{Y}(\delta) - \mathrm{E}(\mathbb{Y}'(\delta)\mathbf{Q}(\delta)\mathbb{Y}(\delta))]$ to zero for each $\delta\in\Delta$ and the stochastic equicontinuity of $\frac{1}{N_1}\mathbb{Y}'(\delta)\mathbf{Q}(\delta)\mathbb{Y}(\delta)$, the uniform convergence result, $\sup_{\delta\in\Delta}|\frac{1}{N_1}[\mathbb{Y}'(\delta)\mathbf{Q}(\delta)\mathbb{Y}(\delta) - \mathrm{E}(\mathbb{Y}'(\delta)\mathbf{Q}(\delta)\mathbb{Y}(\delta))]| = o_p(1)$, follows (Andrews, 1992). Thus, (b) is shown.

Proof of (c). As the two results can be shown in a similar manner, we only show $\sup_{\delta\in\Delta}\frac{1}{N_1}\big|\hat{\mathbb{V}}'(\delta)\mathbb{H}_\lambda(\delta)\hat{\mathbb{V}}(\delta) - \mathrm{E}[\bar{\mathbb{V}}'(\delta)\mathbb{H}_\lambda(\delta)\bar{\mathbb{V}}(\delta)]\big| = o_p(1)$. By the expressions of $\mathbb{H}_\lambda(\delta)$, $\hat{\mathbb{V}}(\delta)$ and $\bar{\mathbb{V}}(\delta)$ given above, we have

$$\begin{aligned}
&\tfrac{1}{N_1}\hat{\mathbb{V}}'(\delta)\mathbb{H}_\lambda(\delta)\hat{\mathbb{V}}(\delta) - \tfrac{1}{N_1}\mathrm{E}[\bar{\mathbb{V}}'(\delta)\mathbb{H}_\lambda(\delta)\bar{\mathbb{V}}(\delta)]\\
&= \tfrac{1}{N_1}[\mathbf{Y}'\mathcal{S}'\bar{\mathbf{Q}}(\delta)(\tfrac{\partial}{\partial\lambda}\mathbf{\Omega}_N(\delta))\bar{\mathbf{Q}}(\delta)\mathcal{S}\mathbf{Y} - \mathrm{E}(\mathbf{Y}'\mathcal{S}'\bar{\mathbf{Q}}(\delta)(\tfrac{\partial}{\partial\lambda}\mathbf{\Omega}_N(\delta))\bar{\mathbf{Q}}(\delta)\mathcal{S}\mathbf{Y})]\\
&\quad - \tfrac{\sigma_{v0}^2}{N_1}\mathtt{tr}[\bar{\mathbf{P}}(\delta)(\tfrac{\partial}{\partial\lambda}\mathbf{\Omega}_N(\delta))\bar{\mathbf{P}}(\delta)\mathbf{\Omega}_N],
\end{aligned}$$

where $\bar{\mathbf{P}}(\delta) = \mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)\mathbb{P}_{\tilde{\mathbb{X}}}(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)\mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)$. The first term is similar in form to $\frac{1}{N_1}[\mathbf{Y}'\mathcal{S}'\bar{\mathbf{Q}}(\delta)\mathcal{S}\mathbf{Y} - \mathrm{E}(\mathbf{Y}'\mathcal{S}'\bar{\mathbf{Q}}(\delta)\mathcal{S}\mathbf{Y})]$ from **(b)** and its uniform convergence is shown in a similar way. Furthermore, by Lemma B.3, the second term is $o(1)$ uniformly in $\delta\in\Delta$.

Proof of (d). Again, using the expressions of $\hat{\beta}_{\mathtt{M}}(\delta)$, $\bar{\beta}_{\mathtt{M}}(\delta)$, $\hat{\mathbb{V}}(\delta)$ and $\bar{\mathbb{V}}(\delta)$, we have

$$\begin{aligned}
&\tfrac{1}{N_1}\hat{\mathbb{V}}'(\delta)\mathbb{J}(\delta)\boldsymbol{\varepsilon}(\hat{\beta}_{\mathtt{M}}(\delta),\delta) - \tfrac{1}{N_1}\mathrm{E}[\bar{\mathbb{V}}'(\delta)\mathbb{J}(\delta)\boldsymbol{\varepsilon}(\bar{\beta}_{\mathtt{M}}(\delta),\delta)]\\
&= \tfrac{1}{N_1}[\mathbf{Y}'\mathcal{S}'\bar{\mathbf{Q}}(\delta)(\mathbb{M}(\delta)+\mathbb{K}(\delta))\mathcal{S}\mathbf{Y} - \mathrm{E}(\mathbf{Y}'\mathcal{S}'\bar{\mathbf{Q}}(\delta)(\mathbb{M}(\delta)+\mathbb{K}(\delta))\mathcal{S}\mathbf{Y})]\\
&\quad - \tfrac{\sigma_{v0}^2}{N_1}\mathtt{tr}[\bar{\mathbf{P}}(\delta)\mathbb{K}(\delta)\mathbf{\Omega}_N] - \tfrac{\sigma_{v0}^2}{N_1}\mathtt{tr}[\bar{\mathbf{Q}}(\delta)\mathbb{M}(\delta)\mathbf{\Omega}_N],
\end{aligned}$$

where $\mathbb{M}(\delta) = [\mathcal{S}(\frac{\partial}{\partial\lambda}\mathbf{A}_{nT}^{-1}(\lambda))\mathbf{X} - \mathbb{K}(\delta)\mathcal{X}(\lambda)][\mathcal{X}'(\lambda)\Psi(\delta)\mathcal{X}(\lambda)]^{-1}\mathcal{X}'(\lambda)\Psi(\delta)$, and $\mathcal{X}(\lambda) =$

$\mathcal{S}\mathbf{A}_{nT}^{-1}(\lambda)\mathbf{X}$. Therefore, the uniform convergence of the first term can be shown in a similar way as we do for $\frac{1}{N_1}[\mathbf{Y}'\mathcal{S}'\bar{\mathbf{Q}}(\delta)\mathcal{S}\mathbf{Y} - \mathrm{E}(\mathbf{Y}'\mathcal{S}'\bar{\mathbf{Q}}(\delta)\mathcal{S}\mathbf{Y})]$ from **(b)** due to their similar forms. By Lemma B.3, the remaining two terms are shown to be $o(1)$, uniformly in $\delta \in \Delta$. ■

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

$$0 = \tfrac{1}{\sqrt{N_1}} S_N^*(\hat{\theta}_{\mathtt{M}}) = \tfrac{1}{\sqrt{N_1}} S_N^*(\theta_0) + \big[\tfrac{1}{N_1}\tfrac{\partial}{\partial\theta'} S_N^*(\theta)\big|_{\theta=\bar{\theta}_r \text{ in } r\text{th row}}\big]\sqrt{N_1}(\hat{\theta}_{\mathtt{M}} - \theta_0), \tag{C.4}$$

where $\{\bar{\theta}_r\}$ are on the line segment between $\hat{\theta}_{\mathtt{M}}$ and θ_0. The result 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). As seen from (2.10), the elements of $S_N^*(\theta_0)$ are linear-quadratic forms in $\mathbf{V}$. Thus, for every non-zero $(k+3)\times 1$ constant vector 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 B.1 and Lemma B.2) 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 Appendix A.

Proof of (b). The Hessian matrix $H_N^*(\theta) = \frac{\partial}{\partial\theta'} S_N^*(\theta)$ is given in Appendix A. Note that we can rewrite $\dot{\Psi}_\lambda(\delta)$ in (A.2) and $\dot{\Psi}_\rho(\delta)$ in (A.3) as $-\Psi(\delta)\dot{\mathbf{\Omega}}_\lambda(\delta)\Psi(\delta) - \Psi(\delta)\mathbb{K}(\delta) - \mathbb{K}'(\delta)\Psi(\delta)$ and $-\Psi(\delta)\dot{\mathbf{\Omega}}_\rho(\delta)\Psi(\delta)$, respectively. Following exactly the same way of proving Lemma B.2(ii), we show that both $\mathbb{K}(\delta)$ and $\frac{\partial}{\partial\omega}\mathbb{K}(\delta)$,$\omega = \lambda, \rho$ are uniformly bounded in both row and column sums, uniformly in $\delta \in \Delta$. In addition, the proof of Lemma B.2(i)

also implies $\ddot{\mathbf{\Omega}}_{\omega\varpi}(\delta)$, $\omega, \varpi = \lambda, \rho$ is bounded in row and column sum norms, uniformly in $\delta \in \Delta$. Thus, by Lemma B.1, we have $\dot{\mathbf{\Psi}}_{\omega}(\delta)$ and $\ddot{\mathbf{\Psi}}_{\omega\varpi}(\delta)$, $\omega, \varpi = \lambda, \rho$ are all bounded in row and column sum norms, uniformly in $\delta \in \Delta$. With these, $\tilde{\mathbb{V}}(\beta_0, \delta_0) = \mathbb{Q}_{\mathbb{D}}\mathbf{\Gamma}\mathbf{V}$ and $\mathcal{V}(\beta_0, \lambda_0) = \mathcal{S}\mathbf{A}_{nT}^{-1}[\mathbf{D}\phi_0 + \mathbf{B}_{nT}^{-1}\mathbf{V}]$, Lemma B.4 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 $\hat{\theta}_{\mathtt{M}} \xrightarrow{p} \theta_0$, where for simplicity, $H_N^*(\bar{\theta})$ 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, $\frac{1}{N_1}H_N^*(\bar{\theta}) = \frac{1}{N_1}H_N^*(\bar{\beta}, \bar{\delta}, \sigma_{v0}^2) + o_p(1)$. Thus, the proof of **(b)** is equivalent to the proof of

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

or the proofs of $\frac{1}{N_1}[H_N^{*\mathtt{S}}(\bar{\beta}, \bar{\delta}, \sigma_{v0}^2) - H_N^{*\mathtt{S}}(\theta_0)] \xrightarrow{p} 0$ 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 that all the components of $H_N^{*\mathtt{S}}(\beta, \delta, \sigma_{v0}^2)$ are linear or quadratic in β, but nonlinear in δ. Hence, with an application of the MVT on $H_N^{*\mathtt{S}}(\bar{\beta}, \bar{\delta}, \sigma_{v0}^2)$ w.r.t $\bar{\delta}$, the result follows. For the non-stochastic part, the results can also be shown using the MVT (See `Online Appendix` for details).

Proof of (c). Since $\tilde{\mathbb{V}}(\beta_0, \delta_0) = \mathbb{Q}_{\mathbb{D}}\mathbf{\Gamma}\mathbf{V}$ and $\mathcal{V}(\beta_0, \lambda_0) = \mathcal{S}\mathbf{A}_{nT}^{-1}[\mathbf{D}\phi_0 + \mathbf{B}_{nT}^{-1}\mathbf{V}]$, the Hessian matrix at true θ_0 are seen to be linear combinations of terms linear or quadratic in $\mathbf{V}$. We have, e.g., $\frac{1}{N_1}[H_{\rho\rho}^*(\rho_0) - \mathrm{E}(H_{\rho\rho}^*(\rho_0))] = \frac{1}{N_1\sigma_{v0}^2}[\mathbf{V}'\mathbf{B}_{nT}^{-1\prime}\mathbf{A}_{nT}^{-1\prime}\mathcal{S}'\ddot{\Psi}_{\rho\rho}(\delta_0)\mathcal{S}\mathbf{A}_{nT}^{-1}\mathbf{B}_{nT}^{-1}\mathbf{V} - \mathrm{E}(\mathbf{V}'\mathbf{B}_{nT}^{-1\prime}\mathbf{A}_{nT}^{-1\prime}\mathcal{S}'\ddot{\Psi}_{\rho\rho}(\delta_0)\mathcal{S}\mathbf{A}_{nT}^{-1}\mathbf{B}_{nT}^{-1}\mathbf{V})] = o_p(1)$. The other terms follow similarly. ■

Appendix D: Proofs for Section 3

More on Assumption G′. With the redefined δ and $\mathbf{\Omega}_N(\delta)$, update $\mathbb{Y}(\delta)$, $\mathbb{X}(\delta)$, $\mathbb{D}(\delta)$, and $\bar{\mathbb{V}}(\delta)$ in (C.2) and obtain $\bar{\beta}_{\mathtt{M}}^{\diamond}(\delta)$ and $\bar{\sigma}_{v,\mathtt{M}}^{\diamond 2}(\delta)$. Similarly, we can obtain the population counterpart $\bar{S}_N^{\diamond c}(\delta)$ of $S_N^{\diamond c}(\delta)$, corresponding to $\bar{S}_N^{*c}(\delta)$ in (C.1). With the updated $\bar{\sigma}_{v,\mathtt{M}}^{\diamond 2}(\delta)$, $\mathcal{Q}_N(\delta)$, $\mathbb{J}(\delta)$, $\mathbb{D}^{\dagger}(\delta)$ and $\mathbb{H}_\omega(\delta)$, $\omega = \lambda, \rho, \tau$, a more primitive condition for Assumption G to hold is that either (a), (b) or (c) holds for $\delta \neq \delta_0$, where

(a) $\frac{1}{2\bar{\sigma}_{v,\mathtt{M}}^{\diamond 2}(\delta)}\eta'\mathcal{Q}_N'(\delta)\mathbb{H}_\lambda(\delta)\mathcal{Q}_N(\delta)\eta + \frac{1}{\bar{\sigma}_{v,\mathtt{M}}^{\diamond 2}(\delta)}\eta'\mathcal{Q}_N'(\delta)\mathbb{J}(\delta)\big\{\mathbb{D}^{\dagger}(\delta)\mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)[\eta - \mathcal{S}\mathbf{A}_{nT}^{-1}(\lambda)\mathbf{X}\bar{\beta}_{\mathtt{M}}^{\diamond}(\delta)] + \mathbf{X}\bar{\beta}_{\mathtt{M}}^{\diamond}(\delta)\big\} + \frac{\sigma_{v0}^2}{\bar{\sigma}_{v,\mathtt{M}}^{\diamond 2}(\delta)}\mathtt{tr}[\mathbb{J}(\delta)\mathbb{D}^{\dagger}(\delta)\mathcal{O}_N(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)] + \frac{1}{2}\mathtt{tr}[\frac{\sigma_{v0}^2}{\bar{\sigma}_{v,\mathtt{M}}^{\diamond 2}(\delta)}\mathbb{H}_\lambda(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)\mathcal{O}_N(\delta)\mathbb{Q}_{\mathbb{D}}(\delta) - \mathbb{H}_\lambda(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)] \neq 0$,

(b) $\frac{1}{2\bar{\sigma}_{v,\mathtt{M}}^{\diamond 2}(\delta)}\eta'\mathcal{Q}_N'(\delta)\mathbb{H}_\rho(\delta)\mathcal{Q}_N(\delta)\eta + \frac{1}{2}\mathtt{tr}[\frac{\sigma_{v0}^2}{\bar{\sigma}_{v,\mathtt{M}}^{\diamond 2}(\delta)}\mathbb{H}_\rho(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)\mathcal{O}_N(\delta)\mathbb{Q}_{\mathbb{D}}(\delta) - \mathbb{H}_\rho(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)] \neq 0$,

(c) $\frac{1}{2\bar{\sigma}_{v,\mathtt{M}}^{\diamond 2}(\delta)}\eta'\mathcal{Q}_N'(\delta)\mathbb{H}_\tau(\delta)\mathcal{Q}_N(\delta)\eta + \frac{1}{2}\mathtt{tr}[\frac{\sigma_{v0}^2}{\bar{\sigma}_{v,\mathtt{M}}^{\diamond 2}(\delta)}\mathbb{H}_\tau(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)\mathcal{O}_N(\delta)\mathbb{Q}_{\mathbb{D}}(\delta) - \mathbb{H}_\tau(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)] \neq 0$.

Again, as $\mathcal{Q}_N(\delta_0)\eta = 0$, $\mathcal{O}_N(\delta_0) = I_N$, $\bar{\beta}_{\mathtt{M}}^{\diamond}(\delta_0) = \beta_0$ and $\bar{\sigma}_{v,\mathtt{M}}^{\diamond 2}(\delta_0) = \sigma_{v0}^2$, the quantities in (a), (b) and (b) are all 0 when $\delta = \delta_0$.

Proof of Theorem 3.1: Similar to the proof of Theorem 2.1 in Appendix C and with δ and $\Omega_N(\delta)$ redefined, the consistency of $\hat{\delta}_{\mathtt{M}}^{\diamond}$ follows if:

(a) $\inf_{\delta\in\Delta}\bar{\sigma}_{v,\mathtt{M}}^{\diamond 2}(\delta)$ is bounded away from zero,

(b) $\sup_{\delta\in\Delta}\big|\hat{\sigma}_{v,\mathtt{M}}^{\diamond 2}(\delta) - \bar{\sigma}_{v,\mathtt{M}}^{\diamond 2}(\delta)\big| = o_p(1)$,

(c) $\sup_{\delta\in\Delta}\frac{1}{N_1}\big|\hat{\mathbb{V}}'(\delta)\mathbb{H}_\omega(\delta)\hat{\mathbb{V}}(\delta) - \mathrm{E}[\bar{\mathbb{V}}'(\delta)\mathbb{H}_\omega(\delta)\bar{\mathbb{V}}(\delta)]\big| = o_p(1)$, for $\omega = \lambda, \rho, \tau$,

(d) $\sup_{\delta\in\Delta}\frac{1}{N_1}\big|\hat{\mathbb{V}}'(\delta)\mathbb{J}(\delta)\boldsymbol{\varepsilon}(\hat{\beta}_{\mathtt{M}}^{\diamond}(\delta),\delta) - \mathrm{E}[\bar{\mathbb{V}}'(\delta)\mathbb{J}(\delta)\boldsymbol{\varepsilon}(\bar{\beta}_{\mathtt{M}}^{\diamond}(\delta),\delta)]\big| = o_p(1)$.

Proof of (a). Note that $\bar{\sigma}_{v,\mathtt{M}}^{\diamond 2}(\delta) = \frac{1}{N_1}\eta'\mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)\mathbf{Q}(\delta)\mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)\eta + \frac{\sigma_{v0}^2}{N_1}\mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\delta)\mathcal{O}_N(\delta)]$. The first term is still non-negative as it can be written in the form of $a'(\delta)a(\delta)$ for an $N\times 1$ vector function of δ, uniformly in $\delta\in\Delta$. For the second term, as $0 < \underline{c}_\tau \leq \inf_{\tau\in\Delta_\tau}\gamma_{\min}[\Upsilon(\tau)\Upsilon'(\tau)\otimes I_n] \leq \sup_{\tau\in\Delta_\tau}\gamma_{\max}[\Upsilon(\tau)\Upsilon'(\tau)\otimes I_n] \leq \bar{c}_\tau < \infty$,

$$\frac{\sigma_{v0}^2}{N_1}\mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\delta)\mathcal{O}_N(\delta)] \geq \frac{\sigma_{v0}^2}{N_1}\gamma_{\min}[\mathcal{O}_N(\delta)]\mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\delta)] \geq \sigma_{v0}^2\gamma_{\max}[\mathbf{\Omega}_N(\delta)]^{-1}\gamma_{\min}(\mathbf{\Omega}_N)$$

$$\geq \frac{\underline{c}_\tau}{\bar{c}_\tau}\sigma_{v0}^2\gamma_{\max}(\mathbf{A}_N'\mathbf{A}_N)^{-1}\gamma_{\max}(\mathbf{B}_N'\mathbf{B}_N)^{-1}\gamma_{\min}[\mathbf{A}_N'(\lambda)\mathbf{A}_N(\lambda)]\gamma_{\min}[\mathbf{B}_N'(\rho)\mathbf{B}_N(\rho)] > 0,$$

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

Proofs of (b), (c), and (d) are similar to proofs of (b), (c), and (d) of Theorem 2.1 (the results in Lemma B.2 still hold with redefined $\Omega_N(\delta)$). Thus, they are omitted. ■

Proof of Theorem 3.2: Applying the MVT to each element of $S^\diamond_N(\hat{\theta}_{\mathtt{M}})$, we have

$$0 = \tfrac{1}{\sqrt{N_1}} S^\diamond_N(\hat{\theta}^\diamond_{\mathtt{M}}) = \tfrac{1}{\sqrt{N_1}} S^\diamond_N(\theta_0) + \big[\tfrac{1}{N_1}\tfrac{\partial}{\partial\theta'} S^\diamond_N(\theta)\big|_{\theta=\bar{\theta}_r \text{ in } r\text{th row}}\big]\sqrt{N_1}(\hat{\theta}^\diamond_{\mathtt{M}} - \theta_0), \tag{D.1}$$

where $\{\bar{\theta}_r\}$ are on the line segment between $\hat{\theta}^\diamond_{\mathtt{M}}$ and θ_0. The result follows if

(a) $\frac{1}{\sqrt{N_1}} S^\diamond_N(\theta_0) \xrightarrow{D} N[0, \lim_{N\to\infty}\Gamma^\diamond_N(\theta_0)]$,

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

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

Proof of (a). Again, from (3.2), the elements of $S^\diamond_N(\theta_0)$ are linear-quadratic forms in $\mathcal{E}$. Thus, for every non-zero $(k+3)\times 1$ constant vector a, $a' S^\diamond_N(\theta_0)$ is of the form:

$$a' S^\diamond_N(\theta_0) = b'_N\mathcal{E} + \mathcal{E}'\Phi_N\mathcal{E} - \sigma^2_v \mathtt{tr}(\Phi_N),$$

for suitably defined non-stochastic vector b_N and matrix Φ_N. Based on Assumptions A′-F′, it is easy to verify (by Lemma B.1 and Lemma B.2) 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 $\frac{1}{\sqrt{N_1}} a' S^\diamond_N(\theta_0)$ follows. By Cramér-Wold device, $\frac{1}{\sqrt{N_1}} S^\diamond_N(\theta_0) \xrightarrow{D} N[0, \lim_{N\to\infty}\Gamma^\diamond_N(\theta_0)]$, where elements of $\Gamma^\diamond_N(\theta_0)$ are given in Appendix A.

Proofs of (b) and (c) are similar to those of Theorem 2.2, and thus are omitted. ■

Online Appendix: Detailed Proofs and More Results

This `Online Appendix` contains detailed proofs of the lemmas in Appendix B and the theories introduced in the main text, the complete set of Monte Carlo results, and an additional application using a simulated Boston housing price panel.

References

Andrews, D. W. (1992). "Generic uniform convergence." *Econometric theory*, 8, 241–257.

Devereux, M. P., Lockwood, B., and Redoano, M. (2007). "Horizontal and vertical indirect tax competition: Theory and some evidence from the USA." *Journal of Public Economics*, 91, 451–479.

Egger, P., Pfaffermayr, M., and Winner, H. (2005). "An unbalanced spatial panel data approach to US state tax competition." *Economics Letters*, 88, 329–335.

Kanbur, R. and Keen, M. (1993). "Jeux sans frontieres: Tax competition and tax coordination when countries differ in size." *The American Economic Review*, 877–892.

Kelejian, H. H. and Prucha, I. R. (1999). "A generalized moments estimator for the autoregressive parameter in a spatial model." *International economic review*, 40, 509–533.

Kelejian, H. H. and Prucha, I. R. (2001). "On the asymptotic distribution of the Moran I test statistic with applications." *Journal of Econometrics*, 104, 219–257.

Kelejian, H. H. and Prucha, I. R. (2010). "Spatial models with spatially lagged dependent variables and incomplete data." *Journal of geographical systems*, 12, 241–257.

Lee, L.-F. and Yu, J. (2010a). "Estimation of spatial autoregressive panel data models with fixed effects." *Journal of econometrics*, 154, 165–185.

Lee, L.-F. and Yu, J. (2010b). "Some recent developments in spatial panel data models." *Regional Science and Urban Economics*, 40(5), 255–271.

Lee, L.-F. and Yu, J. (2015). "Spatial Panel Data Models." *The Oxford Handbook of Panel Data*, Oxford University Press.

Liu, S. F. and Yang, Z. (2020). "Robust estimation and inference of spatial panel data models with fixed effects." *Japanese Journal of Statistics and Data Science*, 3, 257–311.

Manski, C. F. (1993). "Identification of endogenous social effects: The reflection problem." *The review of economic studies*, 60(3), 531–542.

Meng, X. and Yang, Z. (2021). "Unbalanced spatial panel data models with fixed effects." *Working Paper, Singapore Management University.*

Nielsen, S. B. (2001). "A simple model of commodity taxation and cross-border shopping." *Scandinavian Journal of Economics*, 103, 599–623.

Van der Vaart, A. W. (1998). *Asymptotic Statistics.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Wang, W. and Lee, L.-F. (2013). "Estimation of spatial panel data models with randomly missing data in the dependent variable." *Regional Science and Urban Economics*, 43, 521–538.

Yang, Z. (2015). "A general method for third-order bias and variance corrections on a nonlinear estimator." *Journal of Econometrics*, 186, 178–200.

Yang, Z., Yu, J., and Liu, S. F. (2016). "Bias correction and refined inferences for fixed effects spatial panel data models." *Regional Science and Urban Economics*, 61, 52–72.

Zhou, J., Liu, J., Wang, F., and Wang, H. (2022). "Autoregressive model with spatial dependence and missing data." *Journal of Business & Economic Statistics*, 40, 28–34.

Table 1: Empirical mean$(sd)[\hat{se}]$ of estimators, MR model with iid errors. $(\beta, \lambda, \rho, \sigma_v^2) = (1, 0.2, 0.2, 1)$, and `W = Queen` and `M=Rook`.

	T=5			T=10		
	QMLE-MR	M-Est-GU	M-Est-MR	QMLE-MR	M-Est-GU	M-Est-MR
$n = 100$; error = 1, 2, 3, for the three panels below; Missing percentage = 10%						
β	1.0015(.030)	1.0039(.028)[.028]	1.0016(.030)[.029]	1.0008(.019)	.9970(.018)[.018]	1.0008(.019)[.018]
λ	.1913(.041)	.1666(.039)[.040]	.1990(.041)[.041]	.1905(.028)	.1607(.029)[.027]	.1998(.029)[.029]
ρ	.2228(.100)	.1983(.083)[.084]	.1956(.079)[.082]	.1972(.057)	.1991(.049)[.051]	.2000(.052)[.051]
σ_v^2	.7603(.064)	1.0228(.080)[.078]	.9937(.082)[.076]	.8710(.043)	1.0175(.052)[.051]	.9903(.048)[.050]
β	1.0008(.029)	1.0039(.029)[.028]	1.0008(.029)[.029]	.9993(.017)	.9967(.018)[.018]	.9993(.017)[.018]
λ	.1875(.040)	.1650(.040)[.040]	.1949(.040)[.041]	.1916(.031)	.1632(.027)[.027]	.2005(.031)[.029]
ρ	.2267(.108)	.2023(.081)[.084]	.1985(.085)[.083]	.1972(.054)	.1948(.050)[.051]	.2000(.049)[.051]
σ_v^2	.7614(.126)	1.0263(.170)[.162]	.9963(.164)[.161]	.8740(.104)	1.0121(.117)[.112]	.9937(.118)[.112]
β	1.0023(.029)	1.0028(.028)[.028]	1.0024(.029)[.029]	1.0007(.018)	.9980(.017)[.018]	1.0007(.018)[.018]
λ	.1874(.040)	.1669(.041)[.040]	.1957(.040)[.041]	.1908(.030)	.1629(.027)[.027]	.1997(.030)[.029]
ρ	.2285(.105)	.1916(.083)[.084]	.1999(.082)[.083]	.1945(.056)	.1981(.049)[.051]	.1976(.050)[.051]
σ_v^2	.7554(.093)	1.0238(.123)[.121]	.9886(.120)[.117]	.8743(.074)	1.0200(.084)[.082]	.9940(.084)[.081]
$n = 400$; error = 1, 2, 3, for the three panels below; Missing percentage = 10%						
β	1.0000(.013)	.9990(.014)[.014]	1.0000(.013)[.014]	1.0000(.009)	1.0009(.009)[.009]	.9999(.009)[.009]
λ	.1972(.019)	.1701(.020)[.020]	.1991(.020)[.020]	.1977(.014)	.1609(.014)[.014]	.1998(.014)[.014]
ρ	.2597(.051)	.1948(.041)[.041]	.2006(.040)[.040]	.2191(.028)	.1971(.024)[.026]	.2001(.026)[.025]
σ_v^2	.7655(.031)	1.0239(.039)[.039]	.9964(.039)[.038]	.8835(.024)	1.0208(.027)[.026]	.9994(.027)[.025]
β	.9993(.014)	1.0001(.013)[.014]	.9992(.014)[.014]	.9994(.009)	1.0028(.008)[.009]	.9992(.009)[.009]
λ	.1968(.020)	.1701(.020)[.020]	.1986(.020)[.020]	.1983(.014)	.1596(.014)[.014]	.2005(.014)[.014]
ρ	.2617(.050)	.1965(.042)[.041]	.2017(.039)[.041]	.2179(.030)	.1988(.026)[.026]	.1990(.027)[.025]
σ_v^2	.7641(.063)	1.0241(.078)[.082]	.9949(.081)[.082]	.8845(.049)	1.0216(.055)[.058]	1.0005(.055)[.058]
β	1.0000(.013)	.9996(.013)[.014]	1.0000(.014)[.014]	1.0002(.009)	1.0017(.009)[.009]	1.0001(.009)[.009]
λ	.1979(.020)	.1695(.021)[.020]	.1995(.020)[.019]	.1981(.016)	.1608(.014)[.014]	.2004(.016)[.014]
ρ	.2649(.051)	.1948(.042)[.041]	.2047(.041)[.040]	.2187(.027)	.1974(.025)[.026]	.1998(.025)[.025]
σ_v^2	.7636(.047)	1.0220(.057)[.060]	.9945(.061)[.060]	.8864(.036)	1.0211(.043)[.042]	1.0026(.040)[.042]
$n = 100$; error = 1, 2, 3, for the three panels below; Missing percentage = 30%						
β	.9992(.035)	.9980(.035)[.037]	.9988(.035)[.036]	1.0004(.021)	.9983(.021)[.021]	1.0002(.021)[.021]
λ	.1895(.046)	.0923(.042)[.044]	.1978(.048)[.049]	.1920(.032)	.1310(.031)[.032]	.2010(.034)[.033]
ρ	.2336(.189)	.1879(.129)[.126]	.1948(.127)[.121]	.1884(.073)	.1962(.064)[.067]	.1959(.064)[.065]
σ_v^2	.6662(.066)	1.0538(.091)[.098]	.9832(.090)[.095]	.8461(.049)	1.0409(.060)[.059]	.9941(.058)[.058]
β	1.0008(.037)	.9992(.037)[.037]	1.0003(.037)[.036]	.9993(.022)	.9972(.022)[.021]	.9990(.022)[.021]
λ	.1941(.050)	.0979(.045)[.044]	.2024(.051)[.049]	.1878(.031)	.1281(.030)[.031]	.1971(.032)[.033]
ρ	.2428(.185)	.1890(.125)[.126]	.2007(.122)[.124]	.1938(.071)	.2005(.065)[.067]	.2005(.063)[.065]
σ_v^2	.6643(.131)	1.0541(.194)[.188]	.9813(.190)[.183]	.8389(.110)	1.0306(.131)[.127]	.9856(.129)[.125]
β	1.0009(.036)	.9994(.036)[.037]	1.0003(.036)[.036]	1.0007(.021)	.9987(.021)[.021]	1.0005(.021)[.021]
λ	.1910(.050)	.0945(.045)[.044]	.1988(.052)[.049]	.1883(.033)	.1278(.032)[.032]	.1980(.033)[.033]
ρ	.2455(.176)	.1941(.122)[.126]	.2016(.117)[.121]	.1885(.073)	.1970(.066)[.067]	.1959(.065)[.065]
σ_v^2	.6625(.100)	1.0487(.149)[.141]	.9781(.143)[.138]	.8463(.083)	1.0395(.100)[.093]	.9942(.098)[.091]
$n = 400$; error = 1, 2, 3, for the two panels below; Missing percentage = 30%						
β	1.0008(.015)	1.0061(.015)[.017]	1.0007(.015)[.016]	.9997(.011)	1.0054(.011)[.011]	.9996(.011)[.010]
λ	.1972(.024)	.1249(.022)[.022]	.1996(.024)[.024]	.1965(.015)	.1237(.015)[.016]	.1991(.016)[.017]
ρ	.2869(.075)	.1947(.055)[.056]	.1999(.054)[.055]	.2281(.037)	.1958(.032)[.034]	.2022(.032)[.034]
σ_v^2	.6963(.037)	1.0496(.050)[.047]	.9945(.049)[.046]	.8474(.027)	1.0490(.032)[.031]	.9986(.031)[.030]
β	1.0003(.016)	1.0053(.016)[.017]	.9999(.016)[.016]	1.0003(.011)	1.0059(.011)[.011]	1.0001(.011)[.010]
λ	.1950(.024)	.1228(.023)[.022]	.1977(.025)[.024]	.1972(.016)	.1249(.016)[.016]	.1998(.017)[.017]
ρ	.2859(.077)	.1930(.057)[.057]	.1989(.056)[.055]	.2259(.037)	.1958(.034)[.035]	.2004(.032)[.033]
σ_v^2	.6926(.070)	1.0438(.099)[.094]	.9891(.098)[.092]	.8443(.055)	1.0445(.066)[.066]	.9948(.064)[.065]
β	.9997(.017)	1.0051(.017)[.017]	.9995(.016)[.016]	.9997(.011)	1.0052(.011)[.011]	.9995(.011)[.010]
λ	.1976(.025)	.1244(.023)[.022]	.2006(.025)[.024]	.1964(.016)	.1241(.016)[.016]	.1990(.016)[.017]
ρ	.2912(.073)	.1975(.055)[.056]	.2030(.054)[.055]	.2230(.036)	.1919(.032)[.035]	.1977(.031)[.034]
σ_v^2	.6962(.053)	1.0514(.075)[.070]	.9948(.073)[.069]	.8484(.040)	1.0491(.048)[.049]	.9996(.047)[.048]

Note: error = 1(`normal`), 2(`normal mixture`), 3(`chi-square`).

Table 2: Empirical mean(sd)[$\hat{se}$] of estimators, MR model with serially correlated errors. Missing percentage=10%, $(\beta, \lambda, \rho, \tau, \sigma_e^2) = (1, 0.2, 0.2, 0.5, 1)$, and `W = Queen` and `M = Rook`.

	T=5		T=10	
	QMLE-MR	M-Est-MR	QMLE-MR	M-Est-MR
	$n = 100$; error = 1, 2, 3, for the two panels below			
β_1	1.0020(.027)	1.0018(.026)[.025]	1.0003(.016)	1.0000(.015)[.016]
λ	.1894(.035)	.1952(.037)[.035]	.1922(.024)	.1993(.025)[.025]
ρ	.2325(.101)	.1908(.073)[.076]	.2048(.049)	.1953(.043)[.045]
τ	.2325(.096)	.5291(.066)[.074]	.4245(.042)	.5041(.038)[.039]
σ_v^2	.7779(.063)	.9823(.080)[.076]	.8824(.045)	.9930(.051)[.050]
β_1	1.0023(.026)	1.0020(.025)[.025]	1.0012(.015)	1.0008(.015)[.016]
λ	.1881(.038)	.1971(.039)[.035]	.1923(.026)	.2001(.028)[.025]
ρ	.2370(.105)	.1913(.075)[.076]	.2124(.052)	.2022(.045)[.045]
τ	.2576(.118)	.5471(.098)[.091]	.4283(.051)	.5075(.045)[.043]
σ_v^2	.7742(.124)	.9735(.160)[.154]	.8826(.092)	.9933(.104)[.109]
β_1	1.0002(.026)	.9998(.025)[.025]	1.0001(.016)	1.0000(.016)[.016]
λ	.1948(.038)	.2023(.038)[.035]	.1913(.025)	.1987(.026)[.025]
ρ	.2425(.102)	.1972(.074)[.075]	.2132(.050)	.2025(.044)[.045]
τ	.2447(.106)	.5387(.076)[.082]	.4233(.048)	.5031(.042)[.041]
σ_v^2	.7769(.095)	.9795(.121)[.115]	.8852(.078)	.9964(.088)[.080]
	$n = 400$; error = 1, 2, 3, for the two panels below			
β	1.0010(.012)	1.0009(.012)[.012]	.9989(.009)	.9989(.009)[.008]
λ	.1962(.017)	.1985(.017)[.017]	.1942(.011)	.1959(.012)[.013]
ρ	.2714(.050)	.2009(.037)[.037]	.2233(.019)	.1950(.016)[.016]
τ	.2040(.057)	.5072(.032)[.036]	.4129(.021)	.4953(.017)[.018]
σ_v^2	.7856(.029)	.9919(.036)[.038]	.8918(.017)	.9986(.019)[.020]
β	.9990(.013)	.9992(.012)[.012]	.9970(.008)	.9969(.008)[.008]
λ	.1966(.020)	.1991(.020)[.017]	.1983(.012)	.2019(.013)[.013]
ρ	.2686(.049)	.1974(.035)[.037]	.2329(.020)	.2043(.018)[.019]
τ	.2071(.069)	.5118(.037)[.042]	.4168(.028)	.4977(.025)[.024]
σ_v^2	.7820(.062)	.9863(.078)[.079]	.8807(.040)	.9864(.044)[.043]
β	.9988(.014)	.9985(.014)[.012]	1.0002(.009)	1.0002(.009)[.008]
λ	.1985(.019)	.2010(.017)[.017]	.1970(.014)	.1975(.014)[.013]
ρ	.2709(.050)	.1998(.036)[.037]	.2366(.025)	.2065(.022)[.023]
τ	.2091(.061)	.5093(.038)[.038]	.4282(.021)	.5072(.018)[.019]
σ_v^2	.7871(.047)	.9934(.058)[.058]	.8914(.035)	.9986(.039)[.040]

Note: error = 1(`normal`), 2(`normal mixture`), 3(`chi-square`).