Online Appendix

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

Xiaoyu Meng and Zhenlin Yang

In this Online Appendix, we provide proofs of Lemmas B.2 and B.3 in Appendix B of the main text, the detailed proofs of results in Sections 2 and 3, and the complete set of Monte Carlo results. Besides, an additional application using a simulated Boston housing price panel (obtained based on the popular Boston housing price data) is also given.

1. Proofs of Lemmas B.2 and B.3

Although $\mathbf{\Omega}_N(\delta)$ is defined differently in Sections 2 and 3, Lemmas B.2 and B.3 can be shown in a similar manner. In particular, Lemma B.2 (ii) is assumed to be true in Assumption F$'$ of Section 3. Therefore, the lemmas and following proofs are based on $\mathbf{\Omega}_N(\delta)$ in Section 2.

Proof of Lemma B.2:

Proof of (i)**.** Let $\mathbf{M}_{nT}(\delta) = \mathbf{A}'_{nT}(\lambda)\mathbf{B}'_{nT}(\rho)\mathbf{B}_{nT}(\rho)\mathbf{A}_{nT}(\lambda)$. As $\mathbf{\Omega}_N(\delta)$ is the principal submatrix of $\mathbf{M}^{-1}_{nT}(\delta)$, it is sufficient to show that $\mathbf{M}^{-1}_{nT}(\delta)$ is uniformly bounded in both row and column sums, uniformly in $\delta \in \Delta$, which is directly implied by Assumption E (i) and Lemma B.1. Similarly, we have

$\dot{\mathbf{\Omega}}_\lambda(\delta) = \mathcal{S}[(\frac{\partial}{\partial\lambda}\mathbf{A}^{-1}_{nT}(\lambda))\mathbf{B}^{-1}_{nT}(\rho)\mathbf{B}^{-1\prime}_{nT}(\rho)\mathbf{A}^{-1\prime}_{nT}(\lambda) + \mathbf{A}^{-1}_{nT}(\lambda)\mathbf{B}^{-1}_{nT}(\rho)\mathbf{B}^{-1\prime}_{nT}(\rho)(\frac{\partial}{\partial\lambda}\mathbf{A}^{-1\prime}_{nT}(\lambda))]\mathcal{S}'$, and

$\dot{\mathbf{\Omega}}_\rho(\delta) = \mathcal{S}\mathbf{A}^{-1}_{nT}(\lambda)[\frac{\partial}{\partial\rho}\mathbf{B}^{-1}_{nT}(\rho) + \frac{\partial}{\partial\rho}\mathbf{B}^{-1\prime}_{nT}(\rho)]\mathbf{A}^{-1\prime}_{nT}(\lambda)\mathcal{S}'$.

By Assumption E (i) and Lemma B.1, both $\frac{\partial}{\partial\lambda}\mathbf{A}^{-1}_{nT}(\lambda) = \mathbf{A}^{-1}_{nT}(\lambda)\mathbf{W}\mathbf{A}^{-1}_{nT}(\lambda)$ and $\frac{\partial}{\partial\rho}\mathbf{B}^{-1}_{nT}(\rho) = \mathbf{B}^{-1}_{nT}(\rho)\mathbf{M}\mathbf{B}^{-1}_{nT}(\rho)$ are bounded in both row and column sums, uniformly in $\delta \in \Delta$. Thus, $\dot{\mathbf{\Omega}}_\lambda(\delta)$ and $\dot{\mathbf{\Omega}}_\rho(\delta)$ are also bounded in row (column) sum, uniformly in $\delta \in \Delta$.

Note that $\mathbf{\Omega}^{-1}_N(\delta) = [\mathcal{S}\mathbf{M}^{-1}_{nT}(\delta)\mathcal{S}']^{-1}$. We first show $[S\mathbf{M}^{-1}_{nT}(\delta)S']^{-1}$, where S is a $(nT-1)\times nT$ selection matrix with a single observation omitted. Let r be the complement unit row such that $S'S + r'r = I_{nT}$. Thus, we have $\mathbf{M}_{nT}(\delta) = S'M_1(\delta)S + S'M_2(\delta)r + r'M_3(\delta)R + r'M_4(\delta)r$, where $M_1(\delta) = S\mathbf{M}_{nT}(\delta)S'$, $M_2(\delta) = S\mathbf{M}_{nT}(\delta)r'$, $M_3(\delta) = r\mathbf{M}_{nT}(\delta)S'$, and $M_4(\delta) = r\mathbf{M}_{nT}(\delta)r'$. As $\mathbf{M}_{nT}(\delta)$ is a positive definite (p.d.) matrix, all of its principal submatrices (including diagonal elements) are also p.d. matrices, uniformly in $\delta \in \Delta$. Thus, with the inverse formula of a partitioned matrix and $S'S + r'r = I_{nT}$, we have $\mathbf{M}^{-1}_{nT}(\delta) = S'[M_1(\delta) - M_2(\delta)M_4^{-1}(\delta)M_3(\delta)]^{-1}S + S'\ldots r + r'\ldots R + r'\ldots r$. Using $SS' = I_{(n-1)T}$ and $rS' = 0$, we have $[S\mathbf{M}^{-1}_{nT}(\delta)S']^{-1} = M_1(\delta) - M_2(\delta)M_4^{-1}(\delta)M_3(\delta) = S\mathbf{M}_{nT}(\delta)S' - \frac{S\mathbf{M}_{nT}(\delta)r'r\mathbf{M}_{nT}(\delta)S'}{r\mathbf{M}_{nT}(\delta)r'}$. As S, r and

$\mathbf{M}_{nT}(\delta)$ are all uniformly bounded in both row and column sums and $r\mathbf{M}_{nT}(\delta)r'$ is positive, we have $[S\mathbf{M}_{nT}^{-1}(\delta)S']^{-1}$ is also bounded in both row and column sums, uniformly in $\delta \in \Delta$, by Lemma B.1. Note that the selection matrix $\mathcal{S}$ can be written as the product of $nT - N$ selection matrices with a single observation deleted. Therefore, we can repeat the above procedures and show that $\mathbf{\Omega}_N^{-1}(\delta)$ is also bounded in both row and column sums, uniformly in $\delta \in \Delta$.

Proof of (ii)**.** Proof is simpler using a $\mathbf{D}_\alpha^\star$ under the constraint $\alpha_1 = 0$. Recall $\mathbb{D}(\delta) = [\mathbb{D}_\mu(\delta), \mathbb{D}_\alpha(\delta)]$ with $\mathbb{D}_\mu(\delta) = \mathbf{C}(\delta)\mathbf{D}_\mu$, $\mathbb{D}_\alpha(\delta) = \mathbf{C}(\delta)\mathbf{D}_\alpha^\star$ and $\mathbf{C}(\delta) = \mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)\mathcal{S}\mathbf{A}_{nT}^{-1}(\lambda)$. Denote $\mathcal{D}_{11}(\delta) = \mathbb{D}_\mu'(\delta)\mathbb{D}_\mu(\delta)$, $\mathcal{D}_{12}(\delta) = \mathbb{D}_\mu'(\delta)\mathbb{D}_\alpha(\delta)$, $\mathcal{D}_{22}(\delta) = \mathbb{D}_\alpha'(\delta)\mathbb{D}_\alpha(\delta)$. Using the inverse formula of a partitioned matrix, we have

$$[\mathbb{D}'(\delta)\mathbb{D}(\delta)]^{-1} = \begin{bmatrix} \mathcal{F}^{-1}(\delta) & -\mathcal{F}^{-1}(\delta)\mathcal{D}_{12}(\delta)\mathcal{D}_{22}^{-1}(\delta) \\ -\mathcal{D}_{22}^{-1}(\delta)\mathcal{D}_{12}'(\delta)\mathcal{F}^{-1}(\delta) & \mathcal{D}_{22}^{-1}(\delta) + \mathcal{D}_{22}^{-1}(\delta)\mathcal{D}_{12}'(\delta)\mathcal{F}^{-1}(\delta)\mathcal{D}_{12}(\delta)\mathcal{D}_{22}^{-1}(\delta) \end{bmatrix},$$

where $\mathcal{F}(\delta) = \mathcal{D}_{11}(\delta) - \mathcal{D}_{12}(\delta)\mathcal{D}_{22}^{-1}(\delta)\mathcal{D}_{12}'(\delta)$. Plugging this into $\mathbb{Q}_{\mathbb{D}}(\delta)$, we obtain,

$$\mathbb{Q}_{\mathbb{D}}(\delta) = \mathbb{Q}_{\mathbb{D}_\alpha}(\delta) - \mathbb{Q}_{\mathbb{D}_\alpha}(\delta)\mathbb{D}_\mu(\delta)[\mathbb{D}_\mu'(\delta)\mathbb{Q}_{\mathbb{D}_\alpha}(\delta)\mathbb{D}_\mu(\delta)]^{-1}\mathbb{D}_\mu'(\delta)\mathbb{Q}_{\mathbb{D}_\alpha}(\delta).$$

Plugging this into $\Psi(\delta)$, we first show $\mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)\mathbb{Q}_{\mathbb{D}_\alpha}(\delta)\mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)$. Given the special structure of $\mathbb{D}_\alpha(\delta)$, one has $\mathbb{Q}_{\mathbb{D}_\alpha}(\delta) = \texttt{blkdiag}(Q_1(\delta), \ldots, Q_T(\delta))$, where $Q_1(\delta) = I_n$ and $Q_t(\delta) = I_n - \frac{1}{n}C_t(\delta)l_n[\frac{1}{n}l_n'C_t'(\delta)C_t(\delta)l_n]^{-1}l_n'C_t'(\delta)$ for $t = 2, \cdots, T$. Note that $\mathbf{\Omega}_N(\delta)$ is block diagonal and denote its tth block by $\Omega_t(\delta)$. Thus, we only need to show that $\Omega_t^{-\frac{1}{2}}(\delta)Q_t(\delta)\Omega_t^{-\frac{1}{2}}(\delta), t = 1, \cdots, T$ are uniformly bounded in both row and column sums, uniformly in $\delta \in \Delta$. It is trivial when $t = 1$. For $t = 2, \cdots, T$, by Assumption E and Lemma B.2(i), the limit of $\frac{1}{n}l_n'C_t'(\delta)C_t(\delta)l_n$ is bounded away from zero and the elements of $\Omega_t^{-\frac{1}{2}}(\delta)C_t(\delta)l_nl_n'C_t'(\delta)\Omega_t^{-\frac{1}{2}}(\delta)$ are uniformly bounded, uniformly in $\delta \in \Delta$. Therefore, $\Omega_t^{-\frac{1}{2}}(\delta)Q_t(\delta)\Omega_t^{-\frac{1}{2}}(\delta), t = 2, \cdots, T$ must be uniformly bounded in both row and column sums, uniformly in $\delta \in \Delta$.

We next consider $\mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)\mathbb{Q}_{\mathbb{D}_\alpha}(\delta)\mathbb{D}_\mu(\delta)[\mathbb{D}_\mu'(\delta)\mathbb{Q}_{\mathbb{D}_\alpha}(\delta)\mathbb{D}_\mu(\delta)]^{-1}\mathbb{D}_\mu'(\delta)\mathbb{Q}_{\mathbb{D}_\alpha}(\delta)\mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)$. Denote it as $\bar{\mathcal{Q}}(\delta)$, which can be partitioned into $T \times T$ blocks with (s,t)th block being

$$\bar{\mathcal{Q}}_{s,t}(\delta) = \tfrac{1}{T}\Omega_s^{-\frac{1}{2}}(\delta)Q_s(\delta)C_s(\delta)[\tfrac{1}{T}\textstyle\sum_{t=1}^T C_t'(\delta)Q_t(\delta)C_t(\delta)]^{-1}C_t'(\delta)Q_t(\delta)\Omega_t^{-\frac{1}{2}}(\delta).$$

By assuming $A_s^{-1}(\lambda)[\frac{1}{T}\sum_{t=1}^T C_t'(\delta)Q_t(\delta)C_t(\delta)]^{-1}A_t^{-1\prime}(\lambda)$ is uniformly bounded in both row and column sum norms, uniformly in $\delta \in \Delta$, for all s and t, we have that the row and column sums of each $\bar{\mathcal{Q}}_{s,t}(\delta)$ must have uniform order $O(1/T)$, uniformly in $\delta \in \Delta$. As there are T blocks in each row or in each column of $\bar{\mathcal{Q}}(\delta)$, we must have $\bar{\mathcal{Q}}(\delta)$ bounded in both row and column sum norms, uniformly in $\delta \in \Delta$. Consequently, $\Psi(\delta)$ is bounded in both row and column sum norms, uniformly in $\delta \in \Delta$.

Proof of (iii)**.** Let $Z_N(\delta) = [\frac{1}{N}\tilde{\mathbb{X}}'(\delta)\tilde{\mathbb{X}}(\delta)]^{-1} = [\frac{1}{N}\mathbb{X}'(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)\mathbb{X}(\delta)]^{-1}$ with its (j,k)th element being denoted by $z_{jk}(\delta)$. From Assumption C, $Z_N(\delta)$ converges to a finite limit uniformly in $\delta \in \Delta$. Therefore, there exists a constant c_z such that $|z_{jk}(\delta)| \leq c_z$ uniformly in $\delta \in \Delta$ for large enough N. Note that $\mathbb{X}(\delta) = \mathbf{C}(\delta)\mathbf{X}$. As the elements of $\mathbf{X}$ are uniformly bounded (Assumption C), and $\mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)\mathbf{C}(\delta) \equiv \Psi(\delta)\mathcal{S}\mathbf{A}_{nT}^{-1}(\lambda)$ are bounded in both row and column sum norms, uniformly in $\delta \in \Delta$, the elements of $\mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)\mathbb{X}(\delta)$ are also uniformly bounded, uniformly in $\delta \in \Delta$. Hence, there exists a constant c_x such that $|x_{jk}(\delta)| \leq c_x$ uniformly in $\delta \in \Delta$, where $x_{jk}(\delta)$ is the (j,k)th element of $\mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)\mathbb{X}(\delta)$. Let $p_{jl}(\delta)$ be the (j,l)th element of $\mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)\mathbb{P}_{\tilde{\mathbb{X}}}(\delta)\mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)$. It follows that uniformly in $\delta \in \Delta$, $\sum_{j=1}^{N}|p_{jl}(\delta)| \leq \frac{1}{N}\sum_{j=1}^{N}\sum_{r=1}^{k}\sum_{s=1}^{k}|z_{rs}(\delta)x_{jr}(\delta)x_{ls}(\delta)| \leq k^2 c_z c_x^2$ for all $l = 1,2,\ldots,N$. Similarly, uniformly in $\delta \in \Delta$, we have $\sum_{l=1}^{nT}|p_{jl}(\delta)| \leq \frac{1}{N}\sum_{l=1}^{nT}\sum_{r=1}^{k}\sum_{s=1}^{k}|z_{rs}(\delta)x_{jr}(\delta)x_{ls}(\delta)| \leq k^2 c_z c_x^2$ for all $j = 1,2,\ldots,N$. That is, both $\|\mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)\mathbb{P}_{\tilde{\mathbb{X}}}(\delta)\mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)\|_1$ and $\|\mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)\mathbb{P}_{\tilde{\mathbb{X}}}(\delta)\mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)\|_\infty$ are bounded, uniformly in $\delta \in \Delta$. ■

Proof of Lemma B.3: From the proof of Lemma B.2, the elements of $[\frac{1}{N}\mathbb{X}'(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)\mathbb{X}(\delta)]^{-1}$ are uniformly bounded, uniformly in $\delta \in \Delta$. If A_N and B_N are bounded in row (column) sum norm, then $A_N B_N$ is also bounded in row (column) sum norm. Thus, Lemma A.6 of Lee (2004) implies that the elements of $\frac{1}{N}\mathbf{X}'A_N B_N\mathbf{X}$ are uniformly bounded. It follows $\mathtt{tr}[A_N\mathbf{X}[\mathbb{X}'(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)\mathbb{X}(\delta)]^{-1}\mathbf{X}'B_N] = \mathtt{tr}[(\frac{1}{N}\mathbb{X}'(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)\mathbb{X}(\delta))^{-1}\frac{1}{N}\mathbf{X}'(\delta)A_N B_N\mathbf{X}(\delta)] = O(1)$, uniformly in $\delta \in \Delta$ because the number of regressors k is fixed. ■

2. Proofs for Section 2

Proofs use the following facts: (i) the eigenvalues of a projection matrix are either 0 or 1; (ii) the eigenvalues of a positive definite (p.d.) matrix are strictly positive; (iii) $\gamma_{min}(A)\mathtt{tr}(B) \leq \mathtt{tr}(AB) \leq \gamma_{max}(A)\mathtt{tr}(B)$ for symmetric matrix A and positive semi-definite (p.s.d.) matrix B; (iv) $\gamma_{max}(A+B) \leq \gamma_{max}(A) + \gamma_{max}(B)$ for symmetric matrices A and B; (v) $\gamma_{max}(AB) \leq \gamma_{max}(A)\gamma_{max}(B)$ for p.s.d. matrices A and B; and (vi) Let $\gamma_k(\cdot)$ denote the k-th smallest eigenvalue of a matrix. Then $\gamma_k(A) \leq \gamma_k(A_r) \leq \gamma_{k+n-r}(A)$, $1 \leq k \leq r$, for symmetric matrix A and its r-by-r principal submatrix A_r.

Proof of Theorem 2.1: By theorem 5.9 of Van der Vaart (1998), we only need to show $\sup_{\delta\in\delta}\frac{1}{N_1}\left\|S_N^{*c}(\delta) - \bar{S}_N^{*c}(\delta)\right\| \xrightarrow{p} 0$ under the assumptions in Theorem 2.1. From (2.9) and (C.2), the consistency of $\hat{\delta}_{\mathtt{M}}$ follows from:

(a) $\inf_{\delta\in\Delta}\bar{\sigma}_{v,\mathtt{M}}^2(\delta)$ 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 $\bar{\sigma}^2_{v,\mathtt{M}}(\delta)=\frac{1}{N_1}\eta'\boldsymbol{\Omega}_N^{-\frac12}(\delta)\mathbf{Q}(\delta)\boldsymbol{\Omega}_N^{-\frac12}(\delta)\eta+\frac{\sigma^2_{v0}}{N_1}\mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\delta)\mathcal{O}_N(\delta)]$. 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,

$$\frac{\sigma^2_{v0}}{N_1}\mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\delta)\mathcal{O}_N(\delta)]\geq\frac{\sigma^2_{v0}}{N_1}\gamma_{\min}[\mathcal{O}_N(\delta)]\mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\delta)]\geq\sigma^2_{v0}\gamma_{\max}(\boldsymbol{\Omega}_N)^{-1}\gamma_{\min}[\boldsymbol{\Omega}_N(\delta)]$$
$$\geq\sigma^2_{v0}\gamma_{\max}(\mathbf{A}_N'\mathbf{A}_N)^{-1}\gamma_{\max}(\mathbf{B}_N'\mathbf{B}_N)^{-1}\gamma_{\min}[\mathbf{A}_N'(\lambda)\mathbf{A}_N(\lambda)]\gamma_{\min}[\mathbf{B}_N'(\rho)\mathbf{B}_N(\rho)]>0,$$

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 $\bar{\mathbf{Q}}(\delta)=\boldsymbol{\Omega}_N^{-\frac12}(\delta)\mathbf{Q}(\delta)\boldsymbol{\Omega}_N^{-\frac12}(\delta)$ and using $\mathcal{S}\mathbf{Y}=\eta+\mathcal{S}\mathbf{A}_{nT}^{-1}\mathbf{B}_{nT}^{-1}\mathbf{V}$, we have

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

Note that $\bar{\mathbf{Q}}(\delta)\equiv\Psi(\delta)-\boldsymbol{\Omega}_N^{-\frac12}(\delta)\mathbb{P}_{\tilde{\mathbb{X}}}(\delta)\boldsymbol{\Omega}_N^{-\frac12}(\delta)$, bounded in both row and column sum norms, uniformly in $\delta\in\Delta$, by Lemma B.2. Then, by Assumption E and Lemma B.1, $\bar{\mathbf{Q}}(\delta)\mathcal{S}\mathbf{A}_{nT}^{-1}\mathbf{B}_{nT}^{-1}$ and $\mathbf{B}_{nT}^{-1\prime}\mathbf{A}_{nT}^{-1\prime}\mathcal{S}\bar{\mathbf{Q}}(\delta)\mathcal{S}\mathbf{A}_{nT}^{-1}\mathbf{B}_{nT}^{-1}$ are also bounded in both row and column sum norms, uniformly in $\delta\in\Delta$. Further, the elements of η are uniformly bounded. Thus, the pointwise convergence of the first term follows from Lemma B.4 (v), and that 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):

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

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 if $\sup_{\delta\in\Delta}\frac{1}{N_1}\mathbf{Y}'\mathcal{S}'[\frac{\partial}{\partial\varpi}\bar{\mathbf{Q}}(\delta)]\mathcal{S}\mathbf{Y}=O_p(1)$, $\varpi=\lambda,\rho$. We only show when $\varpi=\lambda$ as the proof of the other case is similar and simpler. To derive the expression of the partial derivative $\frac{\partial}{\partial\lambda}\bar{\mathbf{Q}}(\delta)$, write $\bar{\mathbf{Q}}(\delta)\equiv\Psi(\delta)-\Psi(\delta)\mathcal{X}(\lambda)[\mathcal{X}'(\lambda)\Psi(\delta)\mathcal{X}(\lambda)]^{-1}\mathcal{X}'(\lambda)\Psi(\delta)$, where $\mathcal{X}(\lambda)=\mathcal{S}\mathbf{A}_{nT}^{-1}(\lambda)\mathbf{X}$. For a conformable vector a and using (A.2) and $\mathbb{H}_\lambda(\delta) = \mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)[\frac{\partial}{\partial\lambda}\mathbf{\Omega}_N(\delta)]\mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)$, we have,

$$\begin{aligned} a'\tfrac{\partial}{\partial\lambda}\bar{\mathbf{Q}}(\delta)a = & -a'\bar{\mathbf{Q}}(\delta)[\tfrac{\partial}{\partial\lambda}\mathbf{\Omega}_N(\delta)]\bar{\mathbf{Q}}(\delta)a - 2a'\bar{\mathbf{Q}}(\delta)\mathbb{K}(\delta)a \\ & - 2a'\bar{\mathbf{Q}}(\delta)\mathcal{S}[\tfrac{\partial}{\partial\lambda}\mathbf{A}_{nT}^{-1}(\lambda)]\mathbf{X}[\mathcal{X}'(\lambda)\Psi(\delta)\mathcal{X}(\lambda)]^{-1}\mathcal{X}'(\lambda)\Psi(\delta)a. \end{aligned}$$

Again, Lemma B.2 implies $\bar{\mathbf{Q}}(\delta)$ is bounded in both row and column sum norms, uniformly in $\delta \in \Delta$. In addition, following exactly the same way of proving Lemma B.2(ii) and (iii), we show that $\mathbb{K}(\delta)$ and $\mathbf{X}[\mathcal{X}'(\lambda)\Psi(\delta)\mathcal{X}(\lambda)]^{-1}\mathcal{X}'(\lambda)$ are also bounded in both row and column sum norms, uniformly in $\delta \in \Delta$. For ease of presentation, we let $\bar{\mathbf{Q}}_\lambda^\dagger(\delta) = \bar{\mathbf{Q}}(\delta)[\frac{\partial}{\partial\lambda}\mathbf{\Omega}_N(\delta)]\bar{\mathbf{Q}}(\delta) + 2\bar{\mathbf{Q}}(\delta)\mathbb{K}(\delta) + 2\bar{\mathbf{Q}}(\delta)\mathcal{S}[\frac{\partial}{\partial\lambda}\mathbf{A}_{nT}^{-1}(\lambda)]\mathbf{X}[\mathcal{X}'(\lambda)\Psi(\delta)\mathcal{X}(\lambda)]^{-1}\mathcal{X}'(\lambda)\Psi(\delta)$ and then $a'\frac{\partial}{\partial\lambda}\bar{\mathbf{Q}}(\delta)a \equiv -a'\bar{\mathbf{Q}}_\lambda^\dagger(\delta)a$. With these, Lemma B.1 implies that $\|\bar{\mathbf{Q}}_\lambda^\dagger(\delta)\|_1$ and $\|\bar{\mathbf{Q}}_\lambda^\dagger(\delta)\|_\infty$ are bounded uniformly in $\delta \in \Delta$. Thus, Lemma B.4 implies

$$\begin{aligned} &\tfrac{1}{N_1}\mathbf{Y}'\mathcal{S}'[\tfrac{\partial}{\partial\lambda}\bar{\mathbf{Q}}(\delta)]\mathcal{S}\mathbf{Y} = -\tfrac{1}{N_1}\mathbf{Y}'\mathcal{S}'\bar{\mathbf{Q}}_\lambda^\dagger(\delta)\mathcal{S}\mathbf{Y} = -\tfrac{1}{N_1}(\eta + \mathcal{S}\mathbf{A}_{nT}^{-1}\mathbf{B}_{nT}^{-1}\mathbf{V})'\bar{\mathbf{Q}}_\lambda^\dagger(\delta)(\eta + \mathcal{S}\mathbf{A}_{nT}^{-1}\mathbf{B}_{nT}^{-1}\mathbf{V}) \\ &= -\tfrac{1}{N_1}\eta'\bar{\mathbf{Q}}_\lambda^\dagger(\delta)\eta - \tfrac{2}{N1}\eta'\bar{\mathbf{Q}}_\lambda^\dagger(\delta)\mathcal{S}\mathbf{A}_{nT}^{-1}\mathbf{B}_{nT}^{-1}\mathbf{V} - \tfrac{1}{N_1}\mathbf{V}'\mathbf{B}_{nT}^{-1\prime}\mathbf{A}_{nT}^{-1\prime}\mathcal{S}'\bar{\mathbf{Q}}_\lambda^\dagger(\delta)\mathcal{S}\mathbf{A}_{nT}^{-1}\mathbf{B}_{nT}^{-1}\mathbf{V} = O_p(1), \end{aligned}$$

uniformly in $\delta \in \Delta$. Thus, $\sup_{\delta\in\Delta}\frac{1}{N_1}\mathbf{Y}'\mathcal{S}'[\frac{\partial}{\partial\lambda}\bar{\mathbf{Q}}(\delta)]\mathcal{S}\mathbf{Y} = O_p(1)$. Following a similar analysis, $\sup_{\delta\in\Delta}\frac{1}{N_1}\mathbf{Y}'\mathcal{S}'[\frac{\partial}{\partial\rho}\bar{\mathbf{Q}}(\delta)]\mathcal{S}\mathbf{Y} = O_p(1)$. 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 $\bar{\mathbf{P}}(\delta) = \mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)\mathbb{P}_{\tilde{\mathbb{X}}}(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)\mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)$. We see that 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 similarly. Furthermore, by Lemma B.3, it is easy to see 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 $\mathbb{M}(\delta) = [\mathcal{S}(\frac{\partial}{\partial\lambda}\mathbf{A}_{nT}^{-1}(\lambda))\mathbf{X} - \mathbb{K}(\delta)\mathcal{X}(\lambda)][\mathcal{X}'(\lambda)\Psi(\delta)\mathcal{X}(\lambda)]^{-1}\mathcal{X}'(\lambda)\Psi(\delta)$. 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 easily 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

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

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 $-\Psi(\delta)\dot{\boldsymbol{\Omega}}_\lambda(\delta)\Psi(\delta) - \Psi(\delta)\mathbb{K}(\delta) - \mathbb{K}'(\delta)\Psi(\delta)$ and $-\Psi(\delta)\dot{\boldsymbol{\Omega}}_\rho(\delta)\Psi(\delta)$, 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 $\ddot{\boldsymbol{\Omega}}_{\omega\varpi}(\delta)$, $\omega, \varpi = \lambda, \rho$ is bounded in row and column sum norms, uniformly in $\delta \in \Delta$. Thus, by Lemma B.1, we have $\dot{\boldsymbol{\Psi}}_\omega(\delta)$ and $\ddot{\boldsymbol{\Psi}}_{\omega\varpi}(\delta)$, $\omega, \varpi = \lambda, \rho$ are all bounded in row and column sum norms, uniformly in $\delta \in \Delta$. With these, $\tilde{\mathbb{V}}(\beta_0, \delta_0) = \mathbb{Q}_{\mathbb{D}}\boldsymbol{\Gamma}\mathbf{V}$ 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

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

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 $\bar\delta$, we can write $\frac{1}{N_1}[H_N^{*\mathtt{S}}(\bar\beta,\bar\delta,\sigma_{v0}^2) - H_N^{*\mathtt{S}}(\theta_0)]$ as

$$\tfrac{1}{N_1}[\tfrac{\partial}{\partial\delta'}H_N^{*\mathtt{S}}(\bar\beta,\dot\delta,\sigma_{v0}^2)](\bar\delta-\delta_0) + \tfrac{1}{N_1}[H_N^{*\mathtt{S}}(\bar\beta,\delta_0,\sigma_{v0}^2) - H_N^{*\mathtt{S}}(\theta_0)],$$

where for simplicity, $\frac{\partial}{\partial\delta'}H_N^{*\mathtt{S}}(\bar\beta,\dot\delta,\sigma_{v0}^2)$ is used to denote $\frac{\partial}{\partial\delta'}H_N^{*\mathtt{S}}(\beta,\delta,\sigma_{v0}^2)\big|_{\beta=\bar\beta_s,\delta=\dot\delta_s \text{ in } s\text{th row}}$ and $\{\dot\delta_s\}$ are on the line segment between $\bar\delta$ and δ_0. Therefore, it suffices to show

(i) $\frac{1}{N_1}\frac{\partial}{\partial\delta'}H_N^{*\mathtt{S}}(\bar\beta,\dot\delta,\sigma_{v0}^2) = O_p(1)$ and (ii) $\frac{1}{N_1}[H_N^{*\mathtt{S}}(\bar\beta,\delta_0,\sigma_{v0}^2) - H_N^{*\mathtt{S}}(\theta_0)] = o_p(1)$.

We do so for the most complicated term, $H_{\lambda\lambda}^{*\mathtt{S}}(\theta)$. As $\frac{\partial}{\partial\lambda}H_{\lambda\lambda}^{*\mathtt{S}}(\bar\beta,\dot\delta,\sigma_{v0}^2)$ and $\frac{\partial}{\partial\rho}H_{\lambda\lambda}^{*\mathtt{S}}(\bar\beta,\dot\delta,\sigma_{v0}^2)$ can be analyzed in a similar manner, we show the later case for instance. We have,

$$\begin{aligned}\tfrac{1}{N_1}\tfrac{\partial}{\partial\rho}H_{\lambda\lambda}^{*\mathtt{S}}(\bar\beta,\dot\delta,\sigma_{v0}^2) =&\tfrac{2}{N_1\sigma_{v0}^2}\mathcal{V}'(\bar\beta,\dot\lambda)\ddot\Psi_{\lambda\rho}(\dot\delta)\mathcal{S}[\tfrac{\partial}{\partial\lambda}\mathbf{A}_{nT}^{-1}(\dot\lambda)]\mathbf{X}\bar\beta\\ &+\tfrac{2}{N_1\sigma_{v0}^2}\mathcal{V}'(\bar\beta,\dot\lambda)\dot\Psi_{\rho}(\dot\delta)[\tfrac{\partial}{\partial\lambda}\mathbf{A}_{nT}^{-1}(\dot\lambda)]\mathbf{W}_{nT}\mathbf{A}_{nT}^{-1}(\dot\lambda)\mathbf{X}\bar\beta\\ &-\tfrac{1}{N_1\sigma_{v0}^2}\bar\beta'\mathbf{X}'[\tfrac{\partial}{\partial\lambda}\mathbf{A}_{nT}^{-1}(\dot\lambda)]'\dot\Psi_{\rho}(\dot\delta)[\tfrac{\partial}{\partial\lambda}\mathbf{A}_{nT}^{-1}(\dot\lambda)]\mathbf{X}\bar\beta\\ &-\tfrac{1}{2N_1\sigma_{v0}^2}\mathcal{V}'(\bar\beta,\dot\lambda)[\tfrac{\partial}{\partial\rho}\ddot\Psi_{\lambda\lambda}(\dot\delta)]\mathcal{V}(\bar\beta,\dot\lambda)\end{aligned}$$

From (A.2), we note that the expression of $\ddot\Psi_{\lambda\lambda}(\delta)$ involves only $\ddot{\mathbf{\Omega}}_{\lambda\lambda}(\delta)$, $\dot\Psi_\lambda(\delta)$ and $\frac{\partial}{\partial\lambda}\mathbb{K}(\delta)$. The partial derivatives of these components w.r.t ρ are easily shown to be bounded in both row and column sums, uniformly in $\delta\in\Delta$. It follows by Lemmas B.1 and B.4 that the above equation is $O_p(N)$, and then the result (i) follows.

To prove (ii), we note that all the terms in $H_{\lambda\lambda}^{*\mathtt{S}}(\theta)$ are quadratic in β and therefore,

$$\tfrac{1}{N_1}[H_{\lambda\lambda}^{*\mathtt{S}}(\bar\beta,\delta_0,\sigma_{v0}^2) - H_{\lambda\lambda}^{*\mathtt{S}}(\theta_0)] = \tfrac{1}{N_1\sigma_{v0}^2}(\bar\beta+\beta_0)'\mathcal{H}(\delta_0)(\bar\beta-\beta_0),$$

where $\mathcal{H}(\delta) = 2\mathbf{X}'\mathbf{A}_{nT}^{-1\prime}(\lambda)\dot\Psi_\lambda(\delta)\mathcal{S}[\frac{\partial}{\partial\lambda}\mathbf{A}_{nT}^{-1}(\lambda)]\mathbf{X} + 2\mathbb{X}'(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)\mathbb{J}(\delta)\mathbf{W}_{nT}\mathbf{A}_{nT}^{-1}(\lambda)\mathbf{X} - \mathbf{X}'\mathbb{J}'(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)\mathbb{J}(\delta)\mathbf{X} - \frac{1}{2}\mathbf{X}'\mathbf{A}_{nT}^{-1\prime}(\lambda)\ddot\Psi_{\lambda\lambda}(\delta)\mathbf{A}_{nT}^{-1}(\lambda)\mathbf{X}$. By Lemmas B.1 and B.2, it is easy to show that $\mathcal{H}(\delta_0)$ are bounded in both row and column sums. Therefore, (ii) holds as $\bar\beta-\beta_0 = o_p(1)$.

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

$$\begin{aligned}&\tfrac{1}{N_1}[H_{\lambda\lambda}^{*\mathtt{NS}}(\bar\delta) - H_{\lambda\lambda}^{*\mathtt{NS}}(\delta_0)]\\ =&-(\bar\lambda-\lambda_0)\tfrac{1}{2N_1}\mathtt{tr}[\ddot{\mathbf{\Omega}}_{\lambda\lambda}(\dot\delta)\dot\Psi_\lambda(\dot\delta) + \dot{\mathbf{\Omega}}_\lambda(\dot\delta)\ddot\Psi_{\lambda\lambda}(\dot\delta) + [\tfrac{\partial}{\partial\lambda}\ddot{\mathbf{\Omega}}_{\lambda\lambda}(\dot\delta)]\Psi(\dot\delta) + \ddot{\mathbf{\Omega}}_{\lambda\lambda}(\dot\delta)\dot\Psi_\lambda(\dot\delta)]\\ &-(\bar\rho-\rho_0)\tfrac{1}{2N_1}\mathtt{tr}[\ddot{\mathbf{\Omega}}_{\lambda\rho}(\dot\delta)\dot\Psi_\lambda(\dot\delta) + \dot{\mathbf{\Omega}}_\lambda(\dot\delta)\ddot\Psi_{\lambda\rho}(\dot\delta) + [\tfrac{\partial}{\partial\rho}\ddot{\mathbf{\Omega}}_{\lambda\lambda}(\dot\delta)]\Psi(\dot\delta) + \ddot{\mathbf{\Omega}}_{\lambda\lambda}(\dot\delta)\dot\Psi_\rho(\dot\delta)],\end{aligned}$$

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

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 $\mathbf{V}$, and constants. The constant terms are canceled out. Other terms are shown to be $o_p(1)$ based on Lemma B.4. For example,

$$
\begin{aligned}
&\tfrac{1}{N_1}[H^*_{\rho\rho}(\rho_0) - \mathrm{E}(H^*_{\rho\rho}(\rho_0))] \\
&= \tfrac{1}{N_1\sigma_{v0}^2}[\mathbf{V}'\mathbf{B}_{nT}^{-1\prime}\mathbf{A}_{nT}^{-1\prime}\mathcal{S}'\ddot{\Psi}_{\rho\rho}(\delta_0)\mathcal{S}\mathbf{A}_{nT}^{-1}\mathbf{B}_{nT}^{-1}\mathbf{V} - \mathrm{E}(\mathbf{V}'\mathbf{B}_{nT}^{-1\prime}\mathbf{A}_{nT}^{-1\prime}\mathcal{S}'\ddot{\Psi}_{\rho\rho}(\delta_0)\mathcal{S}\mathbf{A}_{nT}^{-1}\mathbf{B}_{nT}^{-1}\mathbf{V})] = o_p(1). \ \blacksquare
\end{aligned}
$$

Proof of Corollary 2.1: Note that $\Gamma^*_N(\hat{\theta}_{\mathtt{M}}) = \Gamma^*_N(\theta)|_{(\theta=\hat{\theta}_{\mathtt{M}}, \phi=\hat{\phi}_{\mathtt{M}}, \kappa_3=\hat{\kappa}_{3,N}, \kappa_4=\hat{\kappa}_{4,N})}$. As $\hat{\theta}_{\mathtt{M}}$, $\hat{\kappa}_{3,N}$ and $\hat{\kappa}_{4,N}$ are consistent estimators for θ_0, κ_3 and κ_4, plugging these estimators into $\Gamma^*_N(\theta)$ will not bring additional bias to the estimation of $\Gamma^*_N(\theta_0)$. However, due to incidental parameters problem, the $\hat{\mu}_{\mathtt{M}}$ component of $\hat{\phi}_{\mathtt{M}}$ is not consistent for the estimation of μ_0 when T is fixed. To estimate the bias caused by replacing ϕ_0 by $\hat{\phi}_{\mathtt{M}}$, rewrite (2.4),

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

Thus, the unconstrained estimate of ϕ_0 is just $\hat{\phi}_{\mathtt{M}} = \hat{\phi}(\hat{\beta}_{\mathtt{M}}, \hat{\delta}_{\mathtt{M}})$. From the expression of $\Gamma^*_{\lambda\lambda}(\theta_0)$, we see that ϕ_0 is embedded in $\mathbf{\Gamma}'\mathbb{Q}_{\mathbb{D}}\mathbb{J}\mathbf{D}\phi_0$ from Π_2, where we recall $\mathbf{\Gamma}(\delta) = \mathbf{C}(\delta)\mathbf{B}_{nT}^{-1}(\rho)$, $\mathbf{C}(\delta) = \mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)\mathcal{S}\mathbf{A}_{nT}^{-1}(\lambda)$ and $\mathbb{J}(\delta) = \mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)\mathcal{S}[\frac{\partial}{\partial\lambda}\mathbf{A}_{nT}^{-1}(\lambda)]$. Thus, we have $\mathbf{\Gamma}'(\hat{\delta}_{\mathtt{M}})\mathbb{Q}_{\mathbb{D}}(\hat{\delta}_{\mathtt{M}})\mathbb{J}(\hat{\delta}_{\mathtt{M}})\mathbf{D}\hat{\phi}_{\mathtt{M}} = \mathbb{M}(\hat{\delta}_{\mathtt{M}})[\mathbf{A}_N(\hat{\lambda}_{\mathtt{M}})\mathbf{Y} - \mathbf{X}\hat{\beta}_{\mathtt{M}}]$, where $\mathbb{M}(\hat{\delta}_{\mathtt{M}}) = \mathbf{\Gamma}'(\hat{\delta}_{\mathtt{M}})\mathbb{Q}_{\mathbb{D}}(\hat{\delta}_{\mathtt{M}})\mathbb{J}(\hat{\delta}_{\mathtt{M}})\mathbf{D}[\mathbb{D}'(\hat{\delta}_{\mathtt{M}})\mathbb{D}(\hat{\delta}_{\mathtt{M}})]^{-1}\mathbb{D}'(\hat{\delta}_{\mathtt{M}})\mathbf{C}(\hat{\delta}_{\mathtt{M}})$. Note $\mathbf{A}_N\mathbf{Y} - \mathbf{X}\hat{\beta}_{\mathtt{M}} = \mathbf{A}_N\mathbf{Y} - \mathbf{X}\beta_0 - \mathbf{X}(\hat{\beta}_{\mathtt{M}} - \beta_0)$. Applying the MVT on each row of $\mathbb{M}(\hat{\delta}_{\mathtt{M}})[\mathbf{A}_N(\hat{\lambda}_{\mathtt{M}})\mathbf{Y} - \mathbf{X}\hat{\beta}_{\mathtt{M}}]$ w.r.t δ, we have,

$$
\begin{aligned}
&\mathbb{M}(\hat{\delta}_{\mathtt{M}})[\mathbf{A}_N(\hat{\lambda}_{\mathtt{M}})\mathbf{Y} - \mathbf{X}\hat{\beta}_{\mathtt{M}}] \\
&= \mathbb{M}[\mathbf{A}_N\mathbf{Y} - \mathbf{X}\beta_0 - \mathbf{X}(\hat{\beta}_{\mathtt{M}} - \beta_0)] + [\tfrac{\partial}{\partial\rho}\mathbb{M}(\bar{\delta})][\mathbf{A}_N(\bar{\lambda})\mathbf{Y} - \mathbf{X}\hat{\beta}_{\mathtt{M}}](\hat{\rho}_{\mathtt{M}} - \rho_0) \\
&\quad + \big\{[\tfrac{\partial}{\partial\lambda}\mathbb{M}(\bar{\delta})][\mathbf{A}_N(\bar{\lambda})\mathbf{Y} - \mathbf{X}\hat{\beta}_{\mathtt{M}}] - \mathbb{M}(\bar{\delta})\mathbf{W}\mathbf{Y}\big\}(\hat{\lambda}_{\mathtt{M}} - \lambda_0) \\
&= \mathbf{\Gamma}'\mathbb{Q}_{\mathbb{D}}\mathbb{J}\mathbf{D}\phi_0 + \mathbb{M}\mathbf{B}_N^{-1}\mathbf{V} - \mathbb{M}\mathbf{X}(\hat{\beta}_{\mathtt{M}} - \beta_0) + [\tfrac{\partial}{\partial\rho}\mathbb{M}(\bar{\delta})][\mathbf{A}_N(\bar{\lambda})\mathbf{Y} - \mathbf{X}\hat{\beta}_{\mathtt{M}}](\hat{\rho}_{\mathtt{M}} - \rho_0) \\
&\quad + \big\{[\tfrac{\partial}{\partial\lambda}\mathbb{M}(\bar{\delta})][\mathbf{A}_N(\bar{\lambda})\mathbf{Y} - \mathbf{X}\hat{\beta}_{\mathtt{M}}] - \mathbb{M}(\bar{\delta})\mathbf{W}\mathbf{Y}\big\}(\hat{\lambda}_{\mathtt{M}} - \lambda_0), \qquad \text{(O.2)}
\end{aligned}
$$

where $\bar{\delta}$ lies between $\hat{\rho}_{\mathtt{M}}$ and ρ_0 and changes over the rows of $[\frac{\partial}{\partial\rho}\mathbb{M}(\bar{\delta})][\mathbf{A}_N(\bar{\lambda})\mathbf{Y} - \mathbf{X}\hat{\beta}_{\mathtt{M}}]$ and $[\frac{\partial}{\partial\lambda}\mathbb{M}(\bar{\delta})][\mathbf{A}_N(\bar{\lambda})\mathbf{Y} - \mathbf{X}\hat{\beta}_{\mathtt{M}}] - \mathbb{M}(\bar{\delta})\mathbf{W}\mathbf{Y}$. Note that $\mathbb{M}(\delta) \equiv \mathbf{B}_{nT}^{-1\prime}(\rho)\mathbf{A}_{nT}^{-1\prime}(\lambda)\mathcal{S}'\Psi(\delta)\mathbb{K}(\delta)\mathcal{S}\mathbf{A}_{nT}^{-1}(\lambda)$, and thus it is easy to see that $\mathbb{M}(\delta)$ and $\frac{\partial}{\partial\omega}\mathbb{M}(\bar{\delta})$, $\omega = \lambda, \rho$, are uniformly bounded in both row and column sums, uniformly in $\delta \in \Delta$, by Lemma B.1. From the expression of $\Gamma^*_N(\hat{\theta}_{\mathtt{M}})$, it has components linear or quadratic in $\mathbf{\Gamma}'(\hat{\delta}_{\mathtt{M}})\mathbb{Q}_{\mathbb{D}}(\hat{\delta}_{\mathtt{M}})\mathbb{J}(\hat{\delta}_{\mathtt{M}})\mathbf{D}\hat{\phi}_{\mathtt{M}}$. Let d_N be a non-stochastic N-vector with elements being of uniform order $O(1)$ or $O(h_n^{-1})$. Using (O.2), the terms of $\Gamma_N^*(\hat\theta_{\mathtt{M}})$ linear in $\mathbf{\Gamma}'(\hat\delta_{\mathtt{M}})\mathbb{Q}_{\mathbb{D}}(\hat\delta_{\mathtt{M}})\mathbb{J}(\hat\delta_{\mathtt{M}})\mathbf{D}\hat\phi_{\mathtt{M}}$ are represented as

$$\begin{aligned}
&\tfrac{1}{N_1}d_N'\mathbf{\Gamma}'(\hat\delta_{\mathtt{M}})\mathbb{Q}_{\mathbb{D}}(\hat\delta_{\mathtt{M}})\mathbb{J}(\hat\delta_{\mathtt{M}})\mathbf{D}\hat\phi_{\mathtt{M}}\\
=\ &\tfrac{1}{N_1}d_N'\mathbf{\Gamma}'\mathbb{Q}_{\mathbb{D}}\mathbb{J}\mathbf{D}\phi_0+\tfrac{1}{N_1}d_N'\mathbb{M}\mathbf{B}_N^{-1}\mathbf{V}-\tfrac{1}{N_1}d_N'\mathbb{M}\mathbf{X}(\hat\beta_{\mathtt{M}}-\beta_0)\\
&+\tfrac{1}{N_1}d_N'[\tfrac{\partial}{\partial\rho}\mathbb{M}(\bar\delta)][\mathbf{A}_N(\bar\lambda)\mathbf{Y}-\mathbf{X}\hat\beta_{\mathtt{M}}](\hat\rho_{\mathtt{M}}-\rho_0)\\
&+\tfrac{1}{N_1}d_N'\big\{[\tfrac{\partial}{\partial\lambda}\mathbb{M}(\bar\delta)][\mathbf{A}_N(\bar\lambda)\mathbf{Y}-\mathbf{X}\hat\beta_{\mathtt{M}}]-\mathbb{M}(\bar\delta)\mathbf{W}\mathbf{Y}\big\}(\hat\lambda_{\mathtt{M}}-\lambda_0)\\
=&\tfrac{1}{N_1}d_N'\mathbf{\Gamma}'\mathbb{Q}_{\mathbb{D}}\mathbb{J}\mathbf{D}\phi_0+o_p(1),
\end{aligned}$$

where the last equation holds because of the consistency of $\hat\theta_{\mathtt{M}}$ and Lemma B.4, using $\mathcal{S}\mathbf{Y}=\eta+\mathcal{S}\mathbf{A}_{nT}^{-1}\mathbf{B}_{nT}^{-1}\mathbf{V}$. Hence, we can conclude that the terms of $\Gamma_N^*(\theta_0)$ linear in ϕ_0 can be consistently estimated by simply replacing ϕ_0 with $\hat\phi_{\mathtt{M}}$.

Note that $\mathbf{\Gamma}\mathbf{\Gamma}'=I_N$. Hence, the only term that is quadratic in ϕ_0 is contained in $\Gamma_{\lambda\lambda}^*(\theta_0)$, $\frac{1}{N_1\sigma_{v0}^2}\phi_0'\mathbf{D}'\mathbb{J}'\mathbb{Q}_{\mathbb{D}}\mathbb{J}\mathbf{D}\phi_0$. The plug-in estimator, $\frac{1}{N_1\hat\sigma_{v,\mathtt{M}}^2}\hat\phi_{\mathtt{M}}'\mathbf{D}'\mathbb{J}'(\hat\delta_{\mathtt{M}})\mathbb{Q}_{\mathbb{D}}(\hat\delta_{\mathtt{M}})\mathbb{J}(\hat\delta_{\mathtt{M}})\mathbf{D}\hat\phi_{\mathtt{M}}$, estimates this term. Using (O.2), $\hat\theta_N^*-\theta_0=o_p(1)$ and Lemma B.4, we show that this estimator is biased/inconsistent:

$$\begin{aligned}
&\tfrac{1}{N_1\hat\sigma_{v,\mathtt{M}}^2}\hat\phi_{\mathtt{M}}'\mathbf{D}'\mathbb{J}'(\hat\delta_{\mathtt{M}})\mathbb{Q}_{\mathbb{D}}(\hat\delta_{\mathtt{M}})\mathbb{J}(\hat\delta_{\mathtt{M}})\mathbf{D}\hat\phi_{\mathtt{M}}\\
=\ &\tfrac{1}{N_1\sigma_{v0}^2}\phi_0'\mathbf{D}'\mathbb{J}'\mathbb{Q}_{\mathbb{D}}\mathbb{J}\mathbf{D}\phi_0+\tfrac{1}{N_1\sigma_{v0}^2}\mathbf{V}'\mathbf{B}_{nT}^{-1\prime}\mathbb{M}'\mathbb{M}\mathbf{B}_{nT}^{-1}\mathbf{V}+o_p(1)\\
=\ &\tfrac{1}{N_1\sigma_{v0}^2}\phi_0'\mathbf{D}'\mathbb{J}'\mathbb{Q}_{\mathbb{D}}\mathbb{J}\mathbf{D}\phi_0+\tfrac{1}{N_1}\mathtt{tr}[(\mathbb{D}'\mathbb{D})^{-1}\mathbf{D}'\mathbb{J}'\mathbb{Q}_{\mathbb{D}}\mathbb{J}\mathbf{D}]+o_p(1).
\end{aligned}$$

We see that the bias term, $\frac{1}{N_1}\mathtt{tr}[(\mathbb{D}'\mathbb{D})^{-1}\mathbf{D}'\mathbb{J}'\mathbb{Q}_{\mathbb{D}}\mathbb{J}\mathbf{D}]$, involves only the common parameters that can be consistently estimated. Thus, a bias correction can easily be made. Define

$$\mathrm{Bias}_{\lambda\lambda}^*(\delta)=\tfrac{1}{N_1}\mathtt{tr}[(\mathbb{D}'(\delta)\mathbb{D}(\delta))^{-1}\mathbf{D}'\mathbb{J}'(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)\mathbb{J}(\delta)\mathbf{D}]. \tag{O.3}$$

This gives the bias matrix of $\Gamma_N^*(\hat\theta_{\mathtt{M}})$, which is a matrix of the same dimension as $\Gamma_N^*(\hat\theta_{\mathtt{M}})$, and has the sole non-zero element $\mathrm{Bias}_{\lambda\lambda}^*(\delta_0)$ corresponding to the $\Gamma_{\lambda\lambda}^*(\hat\theta_{\mathtt{M}})$ component. ∎

Proof of Corollary 2.2.

Proof of (i). Recall $\bar{\mathbb{Q}}_{\mathbb{D}}\equiv\mathbf{\Omega}_N^{-\frac{1}{2}}\mathbb{Q}_{\mathbb{D}}\mathbf{\Gamma}$. Three vectors $\mathbf{V}$, $\mathbf{\Omega}_N^{-\frac{1}{2}}\tilde{\mathbb{V}}=\bar{\mathbb{Q}}_{\mathbb{D}}\mathbf{V}$ and $\mathbf{\Omega}_N^{-\frac{1}{2}}(\hat\delta_{\mathtt{M}})\hat{\mathbb{V}}(\hat\beta_{\mathtt{M}},\hat\delta_{\mathtt{M}})=\mathbf{\Omega}_N^{-\frac{1}{2}}(\hat\delta_{\mathtt{M}})\mathbb{Q}_{\mathbb{D}}(\hat\delta_{\mathtt{M}})\mathbf{\Omega}_N^{-\frac{1}{2}}(\hat\delta_{\mathtt{M}})\mathcal{S}[\mathbf{Y}-\mathbf{A}_{nT}^{-1}(\hat\lambda_{\mathtt{M}})\mathbf{X}\hat\beta_{\mathtt{M}}]$ are with respective elements $\{v_l\},\{\tilde v_j\}$ and $\{\hat v_j\}$, and $\bar{\mathbb{Q}}_{\mathbb{D}}$ has elements $\{q_{jl}\}$, $l=1,\dots,nT$, $j=1,\dots,N$, where j and l are the combined indices of both cross-sectional and time dimensions.

Consistency of $\hat\kappa_{3,N}$. As $\hat\sigma_{v,\mathtt{M}}-\sigma_{v0}=o_p(1)$ and $\hat\delta_N^*-\delta_0=o_p(1)$, the denominators of $\hat\kappa_{3,N}$ and κ_3 agree asymptotically. Thus, $\hat\kappa_{3,N}$ is consistent if $\frac{1}{N}\sum_{j=1}^N[\hat v_j^3-\mathrm{E}(\tilde v_j^3)]\xrightarrow{p}0$, or (a) $\frac{1}{N}\sum_{j=1}^N[\tilde v_j^3-\mathrm{E}(\tilde v_j^3)]\xrightarrow{p}0$ and (b) $\frac{1}{N}\sum_{j=1}^N(\hat v_j^3-\tilde v_j^3)\xrightarrow{p}0$.

To prove (a), noting that $\tilde{v}_j = \sum_{h=1}^{nT} q_{jh} v_h$, we have,

$$\frac{1}{N}\sum_{j=1}^{N}[\tilde{v}_j^3 - \mathrm{E}(\tilde{v}_j^3)] = \frac{1}{N}\sum_{j=1}^{N}\sum_{h=1}^{nT} q_{jh}^3[v_h^3 - \mathrm{E}(v_h^3)] + \frac{3}{N}\sum_{j=1}^{N}\sum_{l=1}^{nT}\sum_{\substack{m\neq l\\ m=1}}^{nT} q_{jl}^2 q_{jm} v_l^2 v_m$$
$$+\frac{6}{N}\sum_{j=1}^{N}\sum_{m=1}^{nT}\sum_{\substack{l\neq m\\ l=1}}^{nT}\sum_{\substack{h\neq m,l\\ h=1}}^{nT} q_{jm}q_{jl}q_{jh}v_m v_l v_h \equiv K_1 + K_2 + K_3.$$

First, consider the K_1 term. By Lemmas B.1 and B.2, $\bar{\mathbb{Q}}_{\mathbb{D}}$ is uniformly bounded in both row and column sums. This implies that the elements of $\bar{\mathbb{Q}}_{\mathbb{D}}$ are uniformly bounded. Therefore, there exists a constant $\bar{q}$ such that $|q_{jh}| \leq \bar{q}$ for all j and h. Given these, we have $\sum_{j=1}^{N} q_{jh}^3 \leq \sum_{j=1}^{N} |q_{jh}|^3 \leq \bar{q}^2 \sum_{j=1}^{N} |q_{jh}| < \infty$. Also, note $\{v_i\}$ are iid by Assumption A. Thus, Khinchine's weak law of large number (WLLN) (Feller, 1967, pp. 243-244) implies that K_1 converges to zero in probability as the sample size increases.

For the other two terms, we have by switching the order of summations when needed,

$$K_2 = \frac{3}{N}\sum_{j=1}^{N}\sum_{l=1}^{nT}\sum_{\substack{m\neq l\\ m=1}}^{nT} q_{jl}^2 q_{jm}(v_l^2 - \sigma_v^2)v_m + \frac{3}{N}\sum_{j=1}^{N}\sum_{l=1}^{nT}\sum_{\substack{m\neq l\\ m=1}}^{nT} q_{jl}^2 q_{jm}\sigma_v^2 v_m$$
$$= \frac{3}{N}\sum_{m=1}^{nT}(v_m^2 - \sigma_v^2)(\sum_{j=1}^{N}\sum_{l=1}^{m-1} q_{jm}^2 q_{jl} v_l) + \frac{3}{N}\sum_{m=1}^{nT} v_m[\sum_{j=1}^{N}\sum_{l=1}^{m-1} q_{jl}^2 q_{jm}(v_l^2 - \sigma_v^2)]$$
$$+\frac{3}{N}\sum_{m=1}^{nT}\sum_{j=1}^{N}\sum_{\substack{l\neq m\\ l=1}}^{N} q_{jl}^2 q_{jm}\sigma_v^2 v_m \equiv \frac{1}{N}\sum_{m=1}^{nT}(g_{1,m} + g_{2,m} + g_{3,m}), \text{ and}$$

$$K_3 = \frac{18}{N}\sum_{m=1}^{nT} v_m(\sum_{j=1}^{N}\sum_{l=1}^{m-1}\sum_{\substack{h\neq l\\ h=1}}^{m-1} q_{jm}q_{jl}q_{jh}v_l v_h) \equiv \frac{1}{N}\sum_{m=1}^{nT} g_{4,m},$$

where $g_{1,m} = 3(v_m^2 - \sigma_v^2)\sum_{j=1}^{N}\sum_{l=1}^{m-1} q_{jm}^2 q_{jl} v_l$, $g_{2,m} = 3v_m \sum_{j=1}^{N}\sum_{l=1}^{m-1} q_{jl}^2 q_{jm}(v_l^2 - \sigma_v^2)$, $g_{3,m} = 3\sum_{j=1}^{N}\sum_{\substack{l\neq m\\ l=1}}^{nT} q_{jl}^2 q_{jm}\sigma_v^2 v_m$, and $g_{4,m} = v_m \sum_{j=1}^{N}\sum_{l=1}^{m-1}\sum_{\substack{h\neq l\\ h=1}}^{m-1} q_{jm}q_{jl}q_{jh}v_l v_h$.

Let $\{\mathcal{G}_m\}$ be the increasing sequence of σ-fields generated by $(v_1, \cdots, v_j, j = 1, \cdots, m)$, $m = 1, \cdots, nT$. Then, $\mathrm{E}[(g_{1,m}, g_{2,m}, g_{3,m}, g_{4,m})|\mathcal{G}_{m-1}] = 0$; hence, $\{(g_{1,m}, g_{2,m}, g_{3,m}, g_{4,m})', \mathcal{G}_m\}$ form a vector martingale difference (M.D.) sequence. As $\bar{\mathbb{Q}}_{\mathbb{D}}$ is bounded in row and column sum norms, by Assumption A, it is easy to see that $\mathrm{E}|g_{s,m}|^{1+\epsilon} < \infty$, for $s = 1, 2, 3, 4$ and $\epsilon > 0$. Hence, $\{g_{1,m}\}$, $\{g_{2,m}\}$, $\{g_{3,m}\}$ and $\{g_{4,m}\}$ are uniformly integrable, and the WLLN of Davidson (1994, Theorem 19.7) applies to give $K_2 \xrightarrow{p} 0$ and $K_3 \xrightarrow{p} 0$.

To prove (b), let $\tilde{v}_j(\xi)$ be the jth element of $\mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)\tilde{\mathbb{V}}(\xi) = \Psi(\delta)\mathcal{S}[\mathbf{Y} - \mathbf{A}_{nT}^{-1}(\lambda)\mathbf{X}\beta]$, where $\xi = (\beta', \delta')'$. Thus, $\tilde{v}_j$ and $\hat{v}_j$ are just $\tilde{v}_j(\xi_0)$ and $\tilde{v}_j(\hat{\xi}_{\mathtt{M}})$, respectively. Let $\mathbf{S}(\xi) = \frac{\partial}{\partial \xi'}[\mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)\tilde{\mathbb{V}}(\xi)]$, which has components $\mathbf{S}_\beta(\xi) = -\Psi(\delta)\mathcal{S}\mathbf{A}_{nT}^{-1}(\lambda)\mathbf{X}$, $\mathbf{S}_\lambda(\xi) = \dot{\Psi}_\lambda(\delta)\mathcal{S}[\mathbf{Y} - \mathbf{A}_{nT}^{-1}(\lambda)\mathbf{X}\beta] - \Psi(\delta)\mathcal{S}[\frac{\partial}{\partial\lambda}\mathbf{A}_{nT}^{-1}(\lambda)]\mathbf{X}\beta$, and $\mathbf{S}_\rho(\xi) = \dot{\Psi}_\rho(\delta)\mathcal{S}[\mathbf{Y} - \mathbf{A}_{nT}^{-1}(\lambda)\mathbf{X}\beta]$. Let $s_j'(\xi)$ be the jth row of $\mathbf{S}(\xi)$. We have by the MVT, for each $j = 1, 2, \ldots, N$,

$$\hat{v}_j \equiv \tilde{v}_j(\hat{\xi}_{\mathtt{M}}) = \tilde{v}_j(\xi_0) + s_j'(\bar{\xi})(\hat{\xi}_{\mathtt{M}} - \xi_0) = \tilde{v}_j + \psi_j'(\hat{\xi}_{\mathtt{M}} - \xi_0) + o_p(\|\hat{\xi}_{\mathtt{M}} - \xi_0\|), \tag{O.4}$$

where $\bar{\xi}$ lies between $\hat{\xi}_{\mathtt{M}}$ and ξ_0, and $\psi_j' = \mathrm{plim}_{N\to\infty} s_j'(\bar{\xi})$, which is easily shown to be $O_p(1)$ as follows. Consider the first k (the number of regressors) elements of ψ_j' first. They are the limits of the jth row of $-\mathbb{X}(\bar{\rho})$, which are just the jth row of $-\Psi(\bar{\delta})\mathcal{S}\mathbf{A}_{nT}^{-1}(\bar{\lambda})\mathbf{X}$ because $\bar{\delta} \xrightarrow{p} \delta_0$, implied by $\hat{\delta}_N^* - \delta_0 = o_p(1)$. Hence, we conclude that the first k elements of ψ_j' are $O(1)$, for each $j = 1, 2, \ldots, N$. For the remaining two elements in each ψ_j', they are the limits of elements from the last two columns of $\mathbf{S}(\bar{\xi})$. It is easy to see the limits of the last two columns of $\mathbf{S}(\bar{\xi})$ are just $\mathbf{S}_\lambda(\xi_0)$ and $\mathbf{S}_\rho(\xi_0)$. Using $\mathcal{S}[\mathbf{Y} - \mathbf{A}_{nT}^{-1}\mathbf{X}\beta_0] = \mathcal{S}\mathbf{A}_{nT}^{-1}[\mathbf{D}\phi_0 + \mathbf{B}_{nT}^{-1}\mathbf{V}]$, we can easily see that each element of $\mathbf{S}_\lambda(\xi_0)$ and $\mathbf{S}_\rho(\xi_0)$ are $O_p(1)$, i.e., the last two elements in ψ_j' are also $O_p(1)$, for each $j = 1, 2, \ldots, N$.

As $\tilde{v}_j = O_p(1)$, $\psi_j' = O_p(1)$ and $\hat{\xi}_{\mathtt{M}} - \xi_0 = O_p(\frac{1}{\sqrt{N_1}})$, we have by (O.4), $\hat{v}_j^3 = \tilde{v}_j^3 + 3\tilde{v}_j^2\psi_j'(\hat{\xi}_{\mathtt{M}} - \xi_0) + o_p(\|\hat{\xi}_{\mathtt{M}} - \xi_0\|)$. It follows that

$$\begin{aligned}\tfrac{1}{N}\textstyle\sum_{j=1}^N(\hat{v}_j^3 - \tilde{v}_j^3) &= \tfrac{3}{N}\textstyle\sum_{j=1}^N \tilde{v}_j^2\psi_j'(\hat{\xi}_{\mathtt{M}} - \xi_0) + o_p(\|\hat{\xi}_{\mathtt{M}} - \xi_0\|)\\ &= \tfrac{3\sigma_v^2}{N}\textstyle\sum_{j=1}^N(\sum_{k=1}^{nT} q_{jk}^2\psi_j')(\hat{\xi}_{\mathtt{M}} - \xi_0) + o_p(\|\hat{\xi}_{\mathtt{M}} - \xi_0\|) = o_p(1),\end{aligned}$$

as $\frac{1}{N}\sum_{j=1}^N(\sum_{k=1}^{nT} q_{jk}^2\psi_j') = (\sum_{k=1}^{nT} q_{jk}^2)\frac{1}{N}(\sum_{j=1}^N \psi_j') = O(1)$.

Consistency of $\hat{\kappa}_{4,N}$. As $\hat{\sigma}_{v,\mathtt{M}} - \sigma_{v0} = o_p(1)$ and $\hat{\delta}_N^* - \delta_0 = o_p(1)$, the result follows if $\frac{1}{N}\sum_{j=1}^N[\hat{v}_j^4 - \mathrm{E}(\tilde{v}_j^4)] \xrightarrow{p} 0$. This shows that

$$(c)\ \tfrac{1}{N}\textstyle\sum_{j=1}^N[\tilde{v}_j^4 - \mathrm{E}(\tilde{v}_j^4)] \xrightarrow{p} 0 \ \text{ and } \ (d)\ \tfrac{1}{N}\sum_{j=1}^N(\hat{v}_j^4 - \tilde{v}_j^4) \xrightarrow{p} 0.$$

To prove (c), we have

$$\begin{aligned}&\tfrac{1}{N}\textstyle\sum_{j=1}^N \tilde{v}_j^4 - \tfrac{1}{N}\sum_{j=1}^N \mathrm{E}(\tilde{v}_j^4)\\ &= \tfrac{1}{N}\textstyle\sum_{j=1}^N\sum_{h=1}^{nT} q_{jh}^4[v_h^4 - \mathrm{E}(v_h^4)] + \tfrac{3}{N}\sum_{j=1}^N\sum_{l=1}^{nT}\sum_{\substack{m\neq l\\ m=1}}^{nT} q_{jl}^2 q_{jm}^2(v_l^2 v_m^2 - \sigma_v^4)\\ &\quad + \tfrac{4}{N}\textstyle\sum_{j=1}^N\sum_{l=1}^{nT}\sum_{\substack{m\neq l\\ m=1}}^{nT} q_{jl}^3 q_{jm} v_l^3 v_m + \tfrac{6}{N}\sum_{j=1}^N\sum_{l=1}^{nT}\sum_{\substack{m\neq l\\ m=1}}^{nT}\sum_{\substack{h\neq m,l\\ h=1}}^{nT} q_{jl}^2 q_{jm} q_{jh} v_l^2 v_m v_h\\ &\quad + \tfrac{1}{N}\textstyle\sum_{j=1}^N\sum_{l=1}^{nT}\sum_{\substack{m\neq l\\ m=1}}^{nT}\sum_{\substack{h\neq m,l\\ h=1}}^{nT}\sum_{\substack{p\neq m,l,h\\ p=1}}^{nT} q_{jl}q_{jm}q_{jh}q_{jp}v_l v_m v_h v_p \equiv \sum_{r=1}^5 R_r.\end{aligned}$$

By using WLLN of Davidson (1994, Theorem 19.7) for M.D. arrays as in the proof of (a), we have $R_r = o_p(1)$ for $r = 1, 3, 4, 5$. For R_2, we have

$$\begin{aligned}R_2 &= \tfrac{6}{N}\textstyle\sum_{l=1}^{nT}(v_l^2 - \sigma_v^2)[\sum_{j=1}^N\sum_{m=1}^{l-1} q_{jl}^2 q_{jm}^2(v_m^2 - \sigma_v^2)]\\ &\quad + \tfrac{6}{N}\textstyle\sum_{l=1}^{nT}[\sum_{j=1}^N\sum_{\substack{m\neq l\\ m=1}}^{nT} q_{jl}^2 q_{jm}^2\sigma_v^2(v_l^2 - \sigma_v^2)] \equiv \tfrac{6}{N}\sum_{l=1}^{nT}(f_{1,l} + f_{2,l}),\end{aligned}$$

noting that $v_l^2 v_m^2 - \sigma_v^4 = (v_l^2 - \sigma_v^2)(v_m^2 - \sigma_v^2) + \sigma_v^2(v_m^2 - \sigma_v^2) + \sigma_v^2(v_l^2 - \sigma_v^2)$. Since $\mathrm{E}[f_{1,l}|\mathcal{G}_{l-1}] = 0$ and $\{f_{2,l}\}$ are independent, both $\{f_l\}$ and $\{f_{2,l}\}$ form M.D. sequences. It is easy to see that $\mathrm{E}|f_{s,l}|^{1+\epsilon} < \infty$, for $s = 1, 2$ and $\epsilon > 0$, so that $\{f_{1,l}\}$ and $\{f_{2,l}\}$ are uniformly integrable. Therefore, the WLLN of Davidson (1994, Theorem 19.7) implies that $\frac{6}{N}\sum_{l=1}^{nT} f_{1,l} = o_p(1)$ and $\frac{6}{N}\sum_{l=1}^{nT} f_{2,l} = o_p(1)$.

To prove (d), $\hat{v}_j^4 = \tilde{v}_j^4 + 4\tilde{v}_j^3\psi_j'(\hat{\xi}_{\mathtt{M}} - \xi_0) + o_p(\|\hat{\xi}_{\mathtt{M}} - \xi_0\|)$ by (O.4). It follows that

$$\begin{aligned}\tfrac{1}{N}\textstyle\sum_{j=1}^N(\hat{v}_j^4 - \tilde{v}_j^4) &= \tfrac{4}{N}\textstyle\sum_{j=1}^N \tilde{v}_j^3\psi_j'(\hat{\xi}_{\mathtt{M}} - \xi_0) + o_p(\|\hat{\xi}_{\mathtt{M}} - \xi_0\|)\\ &= \tfrac{4\sigma_v^3\kappa_3}{N}\textstyle\sum_{j=1}^N(\sum_{k=1}^{nT} q_{jk}^3\psi_j')(\hat{\xi}_{\mathtt{M}} - \xi_0) + o_p(\|\hat{\xi}_{\mathtt{M}} - \xi_0\|) = o_p(1).\end{aligned}$$

Proof of (ii). The consistency of $\widehat{\Sigma}^*_N$ to $\Sigma^*_N(\theta_0)$ can be shown similarly as what we do in the proof of Theorem 2.2 for results (b) and (c). For $\widehat{\Gamma}^*_N - \Gamma^*_N(\theta_0) \xrightarrow{p} 0$, we only need to show that $\text{Bias}^*(\hat{\delta}^*_N) - \text{Bias}^*(\delta_0) = o_p(1)$, based on Corollary 2.1. That is to show

$$\tfrac{1}{N_1}\{\mathtt{tr}[(\mathbb{D}'(\hat{\delta}_{\mathtt{M}})\mathbb{D}(\hat{\delta}_{\mathtt{M}}))^{-1}\mathbf{D}'\mathbb{J}'(\hat{\delta}_{\mathtt{M}})\mathbb{Q}_{\mathbb{D}}(\hat{\delta}_{\mathtt{M}})\mathbb{J}(\hat{\delta}_{\mathtt{M}})\mathbf{D}] - \mathtt{tr}[(\mathbb{D}'\mathbb{D})^{-1}\mathbf{D}'\mathbb{J}'\mathbb{Q}_{\mathbb{D}}\mathbb{J}\mathbf{D}]\} = o_p(1),$$

which can be proved as that for $\frac{1}{N_1}[H^{*\mathtt{NS}}_{\lambda\lambda}(\bar{\delta}) - H^{*\mathtt{NS}}_{\lambda\lambda}(\delta_0)]$ in the proof of Theorem 2.2 (b). ∎

3. Proofs for Section 3

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

- (a) $\inf_{\delta\in\Delta}\bar{\sigma}^{\diamond 2}_{v,\mathtt{M}}(\delta)$ is bounded away from zero,
- (b) $\sup_{\delta\in\Delta}\left|\hat{\sigma}^{\diamond 2}_{v,\mathtt{M}}(\delta) - \bar{\sigma}^{\diamond 2}_{v,\mathtt{M}}(\delta)\right| = o_p(1)$,
- (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 $\omega = \lambda, \rho, \tau$,
- (d) $\sup_{\delta\in\Delta}\frac{1}{N_1}\left|\hat{\mathbb{V}}'(\delta)\mathbb{J}(\delta)\boldsymbol{\varepsilon}(\hat{\beta}^{\diamond}_{\mathtt{M}}(\delta),\delta) - \mathrm{E}[\bar{\mathbb{V}}'(\delta)\mathbb{J}(\delta)\boldsymbol{\varepsilon}(\bar{\beta}^{\diamond}_{\mathtt{M}}(\delta),\delta)]\right| = o_p(1)$.

Proof of (a). Note that $\bar{\sigma}^{\diamond 2}_{v,\mathtt{M}}(\delta) = \frac{1}{N_1}\eta'\mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)\mathbf{Q}(\delta)\mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)\eta + \frac{\sigma^2_{v0}}{N_1}\mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\delta)\mathcal{O}_N(\delta)]$. The first term 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 $0 < \underline{c}_{\tau} \le \inf_{\tau\in\Delta_{\tau}}\gamma_{\min}[\Upsilon(\tau)\Upsilon'(\tau)\otimes I_n] \le \sup_{\tau\in\Delta_{\tau}}\gamma_{\max}[\Upsilon(\tau)\Upsilon'(\tau)\otimes I_n] \le \bar{c}_{\tau} < \infty$,

$$\begin{aligned}\tfrac{\sigma^2_{v0}}{N_1}\mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\delta)\mathcal{O}_N(\delta)] &\ge \tfrac{\sigma^2_{v0}}{N_1}\gamma_{\min}[\mathcal{O}_N(\delta)]\mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\delta)] \ge \sigma^2_{v0}\gamma_{\max}[\mathbf{\Omega}_N(\delta)]^{-1}\gamma_{\min}(\mathbf{\Omega}_N)\\ &\ge \frac{\underline{c}_{\tau}}{\bar{c}_{\tau}}\sigma^2_{v0}\gamma_{\max}(\mathbf{A}'_N\mathbf{A}_N)^{-1}\gamma_{\max}(\mathbf{B}'_N\mathbf{B}_N)^{-1}\gamma_{\min}[\mathbf{A}'_N(\lambda)\mathbf{A}_N(\lambda)]\gamma_{\min}[\mathbf{B}'_N(\rho)\mathbf{B}_N(\rho)] > 0,\end{aligned}$$

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 quite similar to the proofs of (b), (c) and (d) of Theorem 2.1 (the results of Lemma B.2 still hold with the 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

$$0 = \tfrac{1}{\sqrt{N_1}}S^{\diamond}_N(\hat{\theta}^{\diamond}_{\mathtt{M}}) = \tfrac{1}{\sqrt{N_1}}S^{\diamond}_N(\theta_0) + \left[\tfrac{1}{N_1}\tfrac{\partial}{\partial\theta'}S^{\diamond}_N(\theta)\big|_{\theta=\bar{\theta}_r \text{ in } r\text{th row}}\right]\sqrt{N_1}(\hat{\theta}^{\diamond}_{\mathtt{M}} - \theta_0), \tag{O.5}$$

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, $a'S_N^{\diamond}(\theta_0)$ is of the form:

$$a'S_N^{\diamond}(\theta_0) = b_N'\mathcal{E} + \mathcal{E}'\Phi_N\mathcal{E} - \sigma_v^2\mathtt{tr}(\Phi_N),$$

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 $\frac{1}{\sqrt{N_1}}a'S_N^{\diamond}(\theta_0)$ follows. By Cramér-Wold device, $\frac{1}{\sqrt{N_1}}S_N^{\diamond}(\theta_0) \xrightarrow{D} N[0, \lim_{N\to\infty}\Gamma_N^{\diamond}(\theta_0)]$, where elements of $\Gamma_N^{\diamond}(\theta_0)$ are given in Appendix A.

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

Proofs of the results in Corollaries 3.1 and 3.2 are similar to those of Corollaries 2.1 and 2.2 and thus are omitted to conserve space. They are available from the authors upon request.

4. Full Monte Carlo Results

In this section, we design a Monte Carlo experiment for the FE-ISPD-MR model with iid errors or serially correlated errors. For the model with serial correlation, we assume $v_{it} = \tau e_{it} + e_{i,t-1}$ and $\{e_{it}\}$ are iid$(0, \sigma_e^2)$ across i and t. Let $\epsilon_t = (e_{1t}, e_{2t}, \ldots, e_{nt})'$. Thus, two types of DGPs employed in this Monte Carlo study are

$$\mathtt{DGP\ 1}: \ \mathcal{S}_tY_t = \mathcal{S}_tA_t^{-1}(\lambda)(X_t\beta + \mu + \alpha_t l_n + U_t), \ \ U_t = \rho M_tU_t + V_t,$$

$$\mathtt{DGP\ 2}: \ \mathcal{S}_tY_t = \mathcal{S}_tA_t^{-1}(\lambda)(X_t\beta + \mu + \alpha_t l_n + U_t), \ \ U_t = \rho M_tU_t + V_t, \ \ V_t = \tau V_{t-1} + \epsilon_t,$$

for $t = 1, \ldots, T$. We choose $n = 50, 100, 200, 400$, and $T = 5, 10$. The parameters values are set at $\beta = 1$, $\lambda = 0.2$, $\rho = 0.2$ and $\sigma_v^2 = 1$ for `DGP 1`, and $\beta = 1$, $\lambda = 0.2$, $\rho = 0.2$, $\tau = 0.5$ and $\sigma_e^2 = 1$ for `DGP 2`. $X_t's$ are generated independently from $N(0, 2^2I_n)$, and individual effects are set to be $\mu = \frac{1}{T}\Sigma_{t=1}^T X_t + e$, where $e \sim N(0, I_n)$. The time fixed effects α are generated from $N(0, I_T)$. The number of Monte Carlo runs is 1000.

The spatial weight matrices can be `Rook` contiguity and `Queen` contiguity. To generate W_t under `Rook`, randomly permute the indices $\{1, 2, \ldots, n\}$ 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 $\{v_{it}\}$ 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 two randomly missing percentages, at 10% and 30%, which allow us to see the effect of the degree of missingness on the estimation. To generate the MR data, a full sample of size n is first generated for each period, and then the "observed" responses are selected based on the generated selection matrix.

The Monte Carlo experiments involve five estimators: `naïve` estimator, `QMLE-GU`, `QMLE-MR`, `M-Est-GU`, and `M-Est-MR`. The `naïve` estimator (naïve to missingness) is the M-estimator based on a balanced panel formed by deleting the spatial units with missing responses; `QMLE-GU` and `QMLE-MR` are the QMLEs assuming GU and MR, respectively; and `M-Est-GU` and `M-Est-MR` are the M-estimators assuming GU and MR, respectively. Clearly, only the estimator `M-Est-MR` is valid as the data is generated according to `FE-ISPD-MR` and the estimation is done accordingly. 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 or model specification. Monte Carlo (empirical) means and standard deviations (*sd*, shown in parentheses) are recorded for the `naïve` estimator, `QMLE`, M-estimator (`M-Est`), and RM-estimator (`RM-Est`). The empirical averages of the standard error estimates ($\hat{se}$, shown in square brackets) are also recorded for the `naïve` estimator, `M-Est` and `RM-Est`, based on the VC matrix estimates in Sections 2 and 3.

Tables 1a and 1b present Monte Carlo results for the `FE-ISPD-MR` model with missing percentages 10% and 30%, respectively, when $\{W_t\}$ are `Rook` and $\{M_t\}$ are `Queen`. As expected, the results show an excellent performance of `M-Est-MR` and its inference methods. `M-Est-MR` performs well even when the sample size is quite small, and shows convergence to their true values as the sample size increases. Their corresponding standard error estimates are also close to Monte Carlo standard deviations. In contrast, the `QMLE` of σ^2 is inconsistent and the finite sample performance of the `QMLE` of the spatial estimates is not as good as that of the proposed M-estimation. By comparing `M-Est-GU` and `M-Est-MR`, we can see the consequences of treating MR models as GU models in that `M-Est-GU` cannot provide consistent estimation for spatial parameters even when the sample size is large enough. When the missing percentage is higher, `M-Est-GU` becomes more biased. This is consistent with our expectation as treating an `FE-SPD-MR` mechanism as `FE-SPD-GU` will ignore the spatial effects from the missing units. The larger the missing percentage is, the more serious the consequence is.

Table 1c reports `QMLE-GU` and the `naïve` estimator for the `FE-ISPD-MR` model with iid errors and missing percentages 10% , when $\{W_t\}$ are `Queen` contiguity and $\{M_t\}$ are `Rook` contiguity. We can see that `QMLE-GU` cannot provide unbiased estimation for spatial parameters even when the sample size is large enough. It can even give the opposite sign for spatial parameter estimation when the sample size is small. This is due to the fact that treating an `FE-ISPD-MR model` as an `FE-USPD-GU` model will ignore the spatial effects of the missing units on their neighbors. Similarly, the `naïve` estimation cannot give us unbiased estimation for spatial parameters as well, as it completely ignores the spatial effects of deleted units on remaining units. These together show again the serious consequences when models are misspecified and wrong estimation methods are applied.

Table 2 reports Monte Carlo results for the `FE-ISPD-MR` model with serial correlation. We only report `QMLE-MR` and `M-Est-MR`, as `QMLE-GU` and `M-Est-GU` are unavailable and `naïve` estimator is expected to have a very poor performance. The `M-Est-MR` of all the parameters has a 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, spatial error parameter ρ, and serial correlation parameter τ.

5. Analyses of Simulated Housing Price Panels

Lastly, we generate an *incomplete* housing price panel by mimicking the housing price environment using the popular *Boston housing price data*. The data is given by Harrison Jr and Rubinfeld (1978) and is corrected and augmented with longitude and latitude by Gilley and Pace (1996). It is cross-sectional data with the median housing price (for each of the 506 census tracts in the Boston metropolitan statistical area) as the response. The explanatory variables include per capita crime rate by town (`crime`), proportion of residential land zoned for lots over 25,000 square feet (`zoning`), proportion of non-retail business acres per town (`industry`), tract bounding river Charles River dummy (`charlesr`), nitric oxide concentration (`noxsq`), average number of rooms per dwelling (`rooms`), proportion of owner-occupied units built prior to 1940 (`houseage`), weighted distances to five Boston employment centers (`distance`), index of accessibility to radial highways (`access`), full-value property-tax rate per 10,000 (`taxrate`), pupil-teacher ratio by town (`ptratio`), $1000(Bk-0.63)$ where Bk is the proportion of blacks by town (`blackpop`), and lower status of the population proportion (`lowclass`).

Housing price panels are usually formed through aggregation (e.g., median, mean) and may be incomplete (with missing observations on response) when researchers aggregate the data to the least possible level (e.g., the census tracts as in the *Boston housing price data*) to keep as much information as possible. However, the characteristics (values of the explanatory variables) of the aggregated spatial units are usually completely available, giving rise to incomplete spatial panel data with missing responses (MR). The spatial weight matrix is constructed using the Euclidean distance with longitude and latitude.

We first estimate a spatial cross-sectional model with spatial lag and spatial error based on the original data and using QML method. Then, we use the estimated model as the true model to generate data for "future" periods. For the time-invariant variables (`charlesr` and `distance`), we can simply repeat them for each period. For the time-varying variables, we perturb them to give their observations for future periods. With error distribution assumed to be log-normally distributed, we generate values of the dependent variable using the generated panel of explanatory variables, and individual and time-fixed effects in a similar way as in the Monte Carlo experiments. We generate four more periods of data to give a 506×5 panel (with the original data in the first period). Clearly, the characteristics of the census tracts are always available but the median price may not be available at every census tract and in every time period, if the time periods are short, e.g., month, or quarter.

For the missing percentages, Ross and Zhou (2021), Shen and Ross (2021) and Nowak and Smith (2020) point out that housing transaction frequencies have a cyclical nature, with higher volumes during economic booms and lower volumes during recessions. In addition, high-quality houses are more likely to sell than low-quality houses, so the frequency of transactions in high-quality census tracts would be higher. Moreover, high-quality houses recover faster than low-quality houses in the economic boom period in terms of transactions, so the average house quality of each census tract also influences the change in the frequency of transactions with the cycle. To capture these features, we first divide the units into high-quality and low-quality groups, depending on whether or not their first-period dependent variable values are higher than the average of the first-period dependent variable values. With these considerations, the randomly missing percentages of the dependent variable are chosen to be [15%, 25%, 15%, 5%, 15%] in the five periods for the high-quality group, and [25%, 40%, 35%, 25%, 40%] in the five periods for the low-quality group. The number of Monte Carlo runs is also set to 1000.

Table 3 presents Monte Carlo results for this experiment, in which three types of estimations are reported: the `naïve` estimation, `M-Est-GU` and `M-Est-MR`. The `naïve` estimation result shows a large bias for spatial parameters as a lot of units are deleted whose spatial effects on remaining units are totally ignored. This leads to a highly biased estimation for variance parameter and thus standard error estimates. The `M-Est-GU` cannot give us unbiased estimation for spatial parameters either, as treating an `FE-ISPD-MR` model as an `FE-USPD-GU` model completely ignores the spatial interaction effects from those spatial units with MR. In contrast, we can see much more reasonable results from the `M-Est-MR`. The standard error estimates are also close to the empirical standard errors.

References

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

Davidson, J. (1994). *Stochastic limit theory: An introduction for econometricians*. OUP Oxford.

Feller, W. (1967). "An introduction to probability theory and its applications." *1957*.

Gilley, O. W. and Pace, R. K. (1996). "On the harrison and rubinfeld data." *Journal of Environmental Economics and Management*, 31(3), 403–405.

Harrison Jr, D. and Rubinfeld, D. L. (1978). "Hedonic housing prices and the demand for clean air." *Journal of environmental economics and management*, 5(1), 81–102.

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.

Lee, L.-F. (2004). "A supplement to 'asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models'." *https://www.asc.ohio-state.edu/lee.1777/wp/sar-qml-r-appen-04feb.pdf*.

Nowak, A. D. and Smith, P. S. (2020). "Quality-adjusted house price indexes." *American Economic Review: Insights*, 2(3), 339–356.

Ross, S. L. and Zhou, T. (2021). "Loss aversion in housing sales prices: Evidence from focal point bias." *Report No. w28796*, National Bureau of Economic Research.

Shen, L. and Ross, S. (2021). "Information value of property description: A machine learning approach." *Journal of Urban Economics*, 121, 103299.

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

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

Table 1a: Empirical mean(sd)[$\hat{se}$] of QMLE, M-Est-GU and M-Est-MR: **MR model** with **iid errors**, Missing percentage = 10%, $(\beta, \lambda, \rho, \sigma_v^2) = (1, 0.2, 0.2, 1)$, and `W = Queen` and `M=Rook`.

	T=5			**T=10**		
	QMLE	M-Est-GU	M-Est-MR	QMLE	M-Est-GU	M-Est-MR
			$n = 50$; error = 1, 2, 3, for the three panels below			
β	.9982(.041)	1.0066(.038)[.038]	.9983(.040)[.040]	1.0012(.026)	1.0015(.024)[.024]	1.0013(.026)[.026]
λ	.1811(.060)	.1348(.059)[.057]	.1980(.061)[.062]	.1839(.040)	.1731(.037)[.036]	.2018(.040)[.042]
ρ	.1849(.155)	.2005(.119)[.125]	.1978(.122)[.119]	.1667(.078)	.1925(.072)[.074]	.1987(.071)[.073]
σ_v^2	.7471(.089)	1.0013(.111)[.110]	.9843(.115)[.107]	.8618(.063)	1.0161(.077)[.073]	.9875(.072)[.071]
β	.9977(.040)	1.0030(.038)[.038]	.9978(.039)[.041]	.9993(.026)	1.0016(.022)[.024]	.9994(.026)[.025]
λ	.1818(.064)	.1357(.057)[.057]	.1993(.065)[.062]	.1772(.041)	.1712(.036)[.036]	.1952(.042)[.042]
ρ	.1861(.146)	.2030(.126)[.125]	.1983(.115)[.119]	.1687(.080)	.1996(.073)[.074]	.2006(.073)[.074]
σ_v^2	.7572(.177)	.9911(.226)[.217]	.9966(.231)[.219]	.8603(.138)	1.0073(.168)[.155]	.9856(.158)[.154]
β	.9999(.039)	1.0070(.039)[.038]	1.0001(.039)[.040]	.9985(.027)	1.0000(.023)[.024]	.9985(.027)[.026]
λ	.1805(.060)	.1392(.057)[.058]	.1976(.061)[.061]	.1812(.042)	.1709(.036)[.036]	.1990(.042)[.042]
ρ	.1840(.148)	.2011(.126)[.125]	.1972(.116)[.119]	.1671(.079)	.1989(.072)[.074]	.1991(.072)[.073]
σ_v^2	.7401(.126)	1.0060(.175)[.162]	.9750(.165)[.158]	.8669(.104)	1.0070(.120)[.113]	.9933(.119)[.115]
			$n = 100$; error = 1, 2, 3, for the three panels below			
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]
			$n = 200$; error = 1, 2, 3, for the three panels below			
β	.9995(.020)	.9981(.020)[.019]	.9994(.020)[.020]	1.0015(.012)	1.0028(.012)[.012]	1.0016(.012)[.013]
λ	.1985(.029)	.1688(.028)[.027]	.2024(.029)[.029]	.1927(.020)	.1733(.020)[.020]	.1973(.020)[.021]
ρ	.2479(.073)	.1957(.060)[.059]	.1989(.057)[.058]	.2089(.038)	.1958(.036)[.036]	.1974(.034)[.036]
σ_v^2	.7632(.043)	1.0208(.055)[.056]	.9960(.054)[.054]	.8825(.031)	1.0166(.036)[.036]	.9992(.035)[.036]
β	1.0000(.020)	.9996(.020)[.020]	.9999(.020)[.020]	1.0008(.012)	1.0016(.012)[.012]	1.0008(.012)[.013]
λ	.1982(.026)	.1674(.027)[.027]	.2020(.027)[.028]	.1957(.021)	.1735(.020)[.020]	.2000(.021)[.021]
ρ	.2450(.071)	.1966(.057)[.059]	.1968(.056)[.058]	.2138(.041)	.1952(.037)[.036]	.2019(.037)[.036]
σ_v^2	.7616(.087)	1.0288(.120)[.117]	.9935(.112)[.114]	.8803(.072)	1.0203(.085)[.081]	.9968(.082)[.081]
β	.9997(.020)	1.0002(.020)[.019]	.9996(.020)[.020]	.9991(.013)	1.0023(.012)[.012]	.9991(.013)[.013]
λ	.1954(.028)	.1671(.028)[.027]	.1993(.028)[.028]	.1947(.020)	.1720(.020)[.020]	.1993(.020)[.021]
ρ	.2482(.077)	.1975(.056)[.059]	.1995(.061)[.058]	.2122(.039)	.1961(.036)[.036]	.2004(.035)[.036]
σ_v^2	.7599(.067)	1.0263(.089)[.086]	.9918(.087)[.084]	.8752(.053)	1.0178(.057)[.059]	.9911(.060)[.058]
			$n = 400$; error = 1, 2, 3, for the three panels below			
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]

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

Table 1b: Empirical mean(sd)[$\hat{se}$] of QMLE, M-Est-GU and M-Est-MR: **MR model** with **iid errors**, Missing percentage = 30%, $(\beta, \lambda, \rho, \sigma_v^2) = (1, 0.2, 0.2, 1)$, `W = Queen` and `M = Rook`.

	T=5			**T=10**		
	QMLE	M-Est-GU	M-Est-MR	QMLE	M-Est-GU	M-Est-MR
$n = 50$; error = 1, 2, 3, for the three panels below						
β	.9983(.047)	1.0034(.047)[.047]	.9977(.047)[.046]	1.0003(.027)	1.0017(.027)[.028]	.9994(.027)[.027]
λ	.1790(.075)	.1029(.064)[.063]	.1982(.079)[.072]	.1827(.041)	.1290(.040)[.041]	.1990(.043)[.044]
ρ	.1498(.234)	.1844(.169)[.172]	.1887(.162)[.166]	.1447(.111)	.1944(.099)[.104]	.1984(.098)[.100]
σ_v^2	.6621(.090)	1.0179(.132)[.129]	.9601(.127)[.126]	.8252(.067)	1.0371(.083)[.086]	.9851(.080)[.084]
β	1.0017(.044)	1.0067(.044)[.047]	1.0011(.044)[.046]	1.0013(.027)	1.0029(.027)[.028]	1.0004(.027)[.027]
λ	.1749(.074)	.0991(.062)[.063]	.1935(.076)[.072]	.1844(.042)	.1286(.042)[.041]	.2005(.043)[.044]
ρ	.1719(.230)	.1947(.158)[.175]	.2038(.158)[.170]	.1583(.115)	.2097(.104)[.104]	.2107(.102)[.101]
σ_v^2	.6663(.182)	1.0265(.265)[.245]	.9666(.261)[.240]	.8264(.150)	1.0411(.184)[.180]	.9861(.178)[.176]
β	.9986(.047)	1.0037(.046)[.047]	.9979(.046)[.046]	.9978(.028)	.9996(.028)[.028]	.9972(.028)[.027]
λ	.1812(.074)	.1052(.067)[.064]	.2013(.080)[.073]	.1857(.044)	.1311(.042)[.041]	.2016(.045)[.044]
ρ	.1680(.225)	.1903(.160)[.174]	.2008(.157)[.165]	.1475(.104)	.1994(.095)[.103]	.2010(.092)[.101]
σ_v^2	.6696(.137)	1.0283(.201)[.188]	.9700(.196)[.185]	.8307(.121)	1.0444(.148)[.136]	.9914(.144)[.132]
$n = 100$; error = 1, 2, 3, for the three panels below						
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]
$n = 200$; error = 1, 2, 3, for the three panels below						
β	.9994(.023)	1.0039(.023)[.024]	.9992(.023)[.024]	1.0005(.015)	1.0027(.015)[.014]	1.0003(.015)[.014]
λ	.1960(.031)	.1195(.030)[.031]	.1998(.032)[.032]	.1943(.023)	.1361(.024)[.024]	.1986(.023)[.023]
ρ	.2675(.107)	.1921(.079)[.083]	.1980(.077)[.080]	.2119(.053)	.1917(.048)[.049]	.1977(.046)[.047]
σ_v^2	.6857(.049)	1.0501(.068)[.067]	.9878(.065)[.065]	.8433(.036)	1.0443(.043)[.043]	.9945(.042)[.042]
β	1.0010(.023)	1.0053(.023)[.024]	1.0007(.023)[.024]	1.0005(.014)	1.0029(.014)[.015]	1.0004(.014)[.014]
λ	.1961(.032)	.1206(.031)[.031]	.1998(.033)[.032]	.1956(.023)	.1355(.025)[.024]	.2004(.023)[.023]
ρ	.2685(.113)	.1916(.080)[.083]	.1983(.080)[.081]	.2164(.056)	.1940(.051)[.049]	.2016(.049)[.048]
σ_v^2	.6864(.094)	1.0516(.133)[.135]	.9895(.132)[.132]	.8505(.074)	1.0550(.089)[.095]	1.0031(.087)[.093]
β	.9986(.023)	1.0030(.023)[.024]	.9982(.023)[.024]	1.0003(.014)	1.0026(.014)[.015]	1.0001(.014)[.014]
λ	.1979(.032)	.1197(.031)[.031]	.2023(.033)[.032]	.1937(.022)	.1350(.023)[.024]	.1987(.023)[.023]
ρ	.2662(.110)	.1903(.080)[.083]	.1969(.078)[.081]	.2174(.053)	.1962(.047)[.049]	.2025(.046)[.047]
σ_v^2	.6886(.075)	1.0559(.109)[.101]	.9920(.104)[.098]	.8474(.058)	1.0500(.068)[.069]	.9993(.068)[.067]
$n = 400$; error = 1, 2, 3, for the three panels below						
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 1c: Empirical mean(sd)[$\hat{se}$] of QMLE based on GU and the `naïve` estimator: **MR model** with **iid errors**, Missing percentage = 10%, $(\beta, \lambda, \rho, \sigma_v^2) = (1, 0.2, 0.2, 1)$, `W = Queen` and `M = Rook`.

	T=5		**T=10**	
	QMLE-GU	Naïve-Est	QMLE-GU	Naïve-Est
	***n* = 50; error = 1, 2, 3, for the three panels below**			
β	1.0002(.041)	.9972(.047)[.050]	.9999(.026)	.9996(.040)[.040]
λ	.1223(.067)	.0802(.053)[.060]	.1652(.042)	.0589(.035)[.037]
ρ	-.0395(.287)	.0292(.122)[.172]	-.0361(.132)	.0543(.089)[.111]
σ_v^2	.7688(.088)	1.0225(.152)[.144]	.8986(.069)	1.0694(.114)[.112]
β	1.0026(.038)	1.0000(.048)[.048]	.9992(.023)	1.0019(.038)[.040]
λ	.1151(.070)	.0770(.057)[.059]	.1620(.041)	.0572(.034)[.037]
ρ	-.0080(.269)	.0336(.127)[.170]	-.0341(.129)	.0594(.090)[.112]
σ_v^2	.7677(.182)	.9983(.321)[.273]	.8972(.146)	1.0695(.249)[.229]
β	1.0061(.041)	.9970(.050)[.049]	1.0007(.024)	.9979(.041)[.040]
λ	.1245(.066)	.0808(.057)[.060]	.1626(.041)	.0584(.034)[.037]
ρ	-.0296(.286)	.0309(.119)[.174]	-.0431(.124)	.0542(.093)[.112]
σ_v^2	.7748(.131)	1.0111(.219)[.205]	.8911(.104)	1.0603(.176)[.169]
	***n* = 100; error = 1, 2, 3, for the three panels below**			
β	1.0024(.028)	1.0013(.035)[.035]	.9978(.018)	1.0001(.026)[.026]
λ	.1486(.048)	.1154(.048)[.055]	.1466(.031)	.0541(.026)[.028]
ρ	.0278(.185)	.0286(.101)[.128]	.0312(.088)	.0527(.063)[.071]
σ_v^2	.8029(.064)	1.0511(.097)[.096]	.9179(.049)	1.0681(.076)[.075]
β	1.0040(.029)	1.0025(.034)[.035]	.9977(.017)	.9994(.024)[.026]
λ	.1522(.048)	.1127(.046)[.055]	.1469(.032)	.0539(.026)[.028]
ρ	.0197(.187)	.0393(.107)[.129]	.0330(.095)	.0500(.061)[.071]
σ_v^2	.8029(.130)	1.0582(.212)[.196]	.9156(.106)	1.0599(.166)[.162]
β	1.0050(.029)	1.0005(.035)[.035]	.9975(.017)	.9990(.026)[.026]
λ	.1516(.049)	.1109(.050)[.054]	.1494(.031)	.0562(.027)[.028]
ρ	.0338(.180)	.0299(.100)[.128]	.0328(.093)	.0494(.063)[.071]
σ_v^2	.8047(.099)	1.0508(.151)[.145]	.9224(.075)	1.0641(.129)[.117]
	***n* = 200; error = 1, 2, 3, for the three panels below**			
β	.9983(.020)	.9987(.023)[.025]	1.0020(.012)	.9999(.019)[.020]
λ	.1565(.032)	.0904(.025)[.028]	.1609(.023)	.0472(.018)[.020]
ρ	.0584(.127)	.0276(.063)[.077]	.0414(.066)	.0407(.042)[.051]
σ_v^2	.8115(.044)	1.0702(.076)[.073]	.9248(.033)	1.0780(.056)[.059]
β	.9990(.019)	.9995(.024)[.025]	1.0007(.012)	1.0003(.020)[.020]
λ	.1576(.031)	.0889(.025)[.028]	.1596(.024)	.0468(.019)[.020]
ρ	.0539(.125)	.0233(.061)[.078]	.0476(.062)	.0425(.042)[.051]
σ_v^2	.8127(.096)	1.0548(.155)[.149]	.9267(.079)	1.0913(.132)[.130]
β	.9987(.020)	.9991(.024)[.025]	1.0013(.012)	1.0008(.020)[.020]
λ	.1545(.031)	.0897(.027)[.028]	.1614(.024)	.0476(.018)[.020]
ρ	.0594(.129)	.0245(.061)[.077]	.0440(.067)	.0355(.043)[.051]
σ_v^2	.8162(.068)	1.0733(.117)[.114]	.9240(.055)	1.0806(.094)[.094]
	***n* = 400; error = 1, 2, 3, for the three panels below**			
β	.9987(.014)	1.0061(.018)[.018]	1.0007(.009)	1.0026(.015)[.016]
λ	.1571(.022)	.0953(.021)[.022]	.1451(.017)	.0440(.013)[.014]
ρ	.0726(.084)	.0298(.047)[.058]	.0755(.044)	.0354(.029)[.037]
σ_v^2	.8176(.031)	1.0668(.051)[.049]	.9296(.023)	1.0880(.046)[.045]
β	.9983(.013)	1.0063(.018)[.018]	1.0005(.009)	1.0037(.015)[.016]
λ	.1556(.024)	.0969(.021)[.022]	.1452(.016)	.0436(.013)[.014]
ρ	.0812(.089)	.0319(.047)[.058]	.0730(.045)	.0341(.030)[.037]
σ_v^2	.8168(.068)	1.0675(.102)[.102]	.9279(.053)	1.0842(.095)[.098]
β	.9978(.014)	1.0079(.019)[.018]	1.0014(.009)	1.0027(.015)[.016]
λ	.1569(.023)	.0963(.020)[.022]	.1458(.017)	.0433(.013)[.014]
ρ	.0773(.087)	.0300(.047)[.058]	.0710(.046)	.0380(.029)[.037]
σ_v^2	.8171(.049)	1.0607(.076)[.074]	.9297(.037)	1.0862(.072)[.072]

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

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

	T=5		**T=10**	
	QMLE	M-Est-MR	QMLE	M-Est-MR
	$n = 50$; error = 1, 2, 3, for the three panels below			
β_1	.9987(.037)	.9985(.036)[.034]	.9998(.020)	.9993(.020)[.021]
λ	.1828(.054)	.1979(.059)[.056]	.1880(.034)	.1995(.035)[.032]
ρ	.2070(.162)	.2032(.116)[.111]	.1870(.068)	.2009(.059)[.065]
τ	.2873(.124)	.5872(.114)[.130]	.4234(.067)	.5071(.057)[.058]
σ_v^2	.7504(.089)	.9394(.120)[.117]	.8700(.062)	.9877(.071)[.071]
β_1	.9987(.034)	.9987(.032)[.034]	1.0000(.021)	.9997(.020)[.021]
λ	.1838(.059)	.2010(.063)[.057]	.1883(.033)	.1990(.035)[.032]
ρ	.2074(.146)	.1973(.107)[.111]	.1777(.069)	.1926(.061)[.065]
τ	.3076(.157)	.6045(.139)[.150]	.4283(.066)	.5105(.058)[.062]
σ_v^2	.7460(.172)	.9298(.223)[.215]	.8744(.136)	.9927(.155)[.154]
β_1	.9993(.037)	1.0005(.036)[.033]	1.0001(.022)	.9996(.022)[.021]
λ	.1788(.060)	.1965(.061)[.056]	.1883(.032)	.2005(.034)[.032]
ρ	.2175(.147)	.2081(.107)[.110]	.1891(.071)	.2037(.063)[.065]
τ	.3000(.144)	.5901(.128)[.136]	.4239(.068)	.5073(.058)[.060]
σ_v^2	.7300(.136)	.9151(.177)[.164]	.8703(.106)	.9879(.120)[.111]
	$n = 100$; error = 1, 2, 3, for the three 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]
	$n = 200$; error = 1, 2, 3, for the three panels below			
β_1	1.0000(.019)	1.0001(.018)[.017]	1.0003(.011)	1.0002(.011)[.011]
λ	.1927(.026)	.1971(.025)[.025]	.1954(.017)	.1985(.018)[.018]
ρ	.2635(.067)	.2009(.050)[.052]	.2200(.036)	.1981(.031)[.032]
τ	.2321(.078)	.5200(.047)[.051]	.4209(.032)	.5010(.027)[.027]
σ_v^2	.7892(.042)	.9928(.052)[.055]	.8886(.031)	.9970(.035)[.036]
β_1	.9980(.020)	.9977(.019)[.017]	1.0008(.011)	1.0007(.011)[.011]
λ	.1950(.026)	.1998(.026)[.025]	.1960(.017)	.1991(.018)[.017]
ρ	.2628(.071)	.2004(.052)[.052]	.2202(.037)	.1982(.032)[.032]
τ	.2445(.079)	.5235(.054)[.060]	.4238(.035)	.5038(.031)[.030]
σ_v^2	.7934(.094)	.9974(.118)[.113]	.8818(.071)	.9892(.080)[.079]
β_1	.9989(.019)	.9991(.018)[.017]	1.0004(.012)	1.0002(.012)[.011]
λ	.1932(.026)	.1982(.027)[.025]	.1959(.018)	.1997(.018)[.017]
ρ	.2574(.070)	.1969(.052)[.052]	.2211(.037)	.1989(.033)[.032]
τ	.2342(.081)	.5242(.052)[.055]	.4216(.033)	.5016(.029)[.029]
σ_v^2	.7891(.063)	.9908(.079)[.082]	.8880(.054)	.9962(.061)[.057]
	$n = 400$; error = 1, 2, 3, for the three 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`).

Table 3: Empirical mean(sd)[$\hat{se}$] of Naïve-Est, M-Est-GU and M-Est-MR: Boston housing price DGP with MR, $n = 506$ and $T = 5$.

	QMLE-CS	Naïve-Est	M-Est-GU	M-Est-MR
Time-varying variables				
crime	-1.0982	-1.0945(.360)[.347]	-1.0893(.134)[.135]	-1.0950(.133)[.141]
zoning	.9164	.9405(.202)[.196]	.9152(.111)[.112]	.9193(.112)[.115]
industry	-.0143	-.0209(.225)[.230]	-.0274(.124)[.124]	-.0244(.124)[.127]
noxsq	-2.0339	-2.0458(.240)[.243]	-2.0196(.125)[.127]	-2.0273(.125)[.133]
rooms2	2.9906	3.0222(.199)[.209]	2.9747(.121)[.120]	2.9858(.119)[.127]
houseage	-.6105	-.6130(.208)[.219]	-.6152(.120)[.120]	-.6170(.120)[.123]
access	2.7855	2.7978(.230)[.243]	2.7643(.124)[.124]	2.7759(.123)[.132]
taxrate	-2.2076	-2.2131(.246)[.239]	-2.1823(.128)[.124]	-2.1952(.128)[.133]
ptratio	-1.3875	-1.3902(.209)[.212]	-1.3726(.118)[.118]	-1.3793(.117)[.122]
blackpop	.9256	.9213(.308)[.310]	.9220(.137)[.136]	.9231(.136)[.139]
lowclass	-3.0573	-3.0746(.243)[.243]	-3.0290(.130)[.125]	-3.0465(.128)[.141]
Spatial dependence				
SL(λ)	.0828	.0937(.022)[.064]	.0266(.044)[.041]	.0920(.062)[.067]
SE(ρ)	.6277	.1278(.030)[.088]	.5404(.043)[.041]	.6449(.062)[.063]
Variance parameter				
σ_v^2	15.3924	20.7510(1.890)[1.924]	16.3664(1.005)[.980]	15.8950(1.079)[1.075]

Missing percentages: [15%, 25%, 15%, 5%, 15%] for high-quality group; [25%, 40%, 35%, 25%, 40%] for low-quality group.