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

Xiaoyu Meng[1] and Zhenlin Yang[2†]

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

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

July 7, 2025

Abstract

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

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

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

*Zhenlin Yang gratefully acknowledges the research support from Singapore Management University. Xiaoyu Meng gratefully acknowledges funding support from the National Natural Science Foundation of China (Grant No. 72403136).

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

1. Introduction

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

$$\begin{aligned} Y_t &= \lambda_0 W_t Y_t + X_{1t}\beta_{10} + W_{dt}X_{2t}\beta_{20} + Z\gamma_0 + (\mathcal{Z}_t\zeta_0)l_n + \mu_0 + \alpha_{t0}l_n + U_t, \\ U_t &= \rho_0 M_t U_t + V_t, \quad t = 1, \ldots, T, \end{aligned} \tag{1.1}$$

where Y_t is a vector of response values on n spatial units at time t, X_{1t} and X_{2t} (typically a submatrix of X_{1t}) are matrices of observations on time-varying regressors, Z a matrix of observations on time-invariant regressors, $\mathcal{Z}_t$ a row vector of values of space-invariant regressors, and $U_t = (u_{1t}, u_{2t}, \ldots, u_{n_t t})'$ and $V_t = (v_{1t}, v_{2t}, \ldots, v_{n_t t})'$ are $n \times 1$ vectors of disturbance and idiosyncratic errors, respectively. W_t, W_{dt}, and M_t are given $n \times n$ spatial weight matrices, which together with the "spatial coefficients" λ_0, β_{20} and ρ_0, characterize the spatial lag or endogenous social effects (Manski, 1993), spatial Durbin or contextual effects, and spatial error (SE) effects, respectively. β_{10}, γ_0 and ζ_0 are vectors of regression coefficients. μ_0 is an n-vector of unit-specific effects and $\{\alpha_t\}$ are time-specific effects, which can be fixed effects (FE), random effects (RE), or correlated random effects (CRE). l_n is an $n \times 1$ vector of ones. Model (1.1) has been extensively studied. See, among others, Lee and Yu (2010a,b, 2015), Yang et al. (2016), and Liu and Yang (2020).

In many panels, not all (n) spatial units appeared in every time period, or even if they all appeared in every time period, some spatial units in certain time periods were not fully observed. Kelejian and Prucha (2010) classify the spatial units in spatial data into three groups: (1) units with full observations on themselves and on their neighbors, (2) units with observations on their neighbors missing, and (3) units with their own observations missing. Meng and Yang (2021) studied SPD models where all units are of Type (1) but the number of them can change from time to time, referred to as the SPD models with *genuine unbalancedness* (GU). In this paper, we study the SPD models where all units are of Types (2) and (3) but missing occurs only on responses, referred to as the *incomplete* SPD model with *missing responses* (MR), to emphasize the fact that although the panel is incomplete, the spatial connectivity or network structure is completely observed.

MR issue has drawn much attention in regular panels, and researchers (e.g., Pacini and Windmeijer, 2015; Abrevaya, 2019) have found that incorporating covariates information from periods with missing outcomes can improve estimation efficiency. MR issues can also frequently occur in spatial panels. In housing price panels, regions with transactions in a certain period have mean/median prices recorded, but regions without transactions have response values missing although their characteristics and spatial connectivity are fully observed. Educational studies often find that some students do not have test scores or graduation status, but their demographic and initial performance data and their "peers" are known. Household income data may be missing for certain years, but information on household characteristics and their neighborhood structure is usually fully recorded. However, essential methods for analyzing these types of data are lacking.

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

$$\begin{aligned} \mathcal{S}_t Y_t &= \mathcal{S}_t A_t^{-1}(\lambda_0)(X_{1t}\beta_{10} + W_{dt}X_{2t}\beta_{20} + Z\gamma_0 + (\mathcal{Z}_t\zeta_0)l_n + \mu_0 + \alpha_{t0}l_n + U_t), \\ U_t &= \rho_0 M_t U_t + V_t, \quad t = 1, \dots, T. \end{aligned} \tag{1.2}$$

The model exploits the observed responses $\mathcal{S}_t Y_t$ while maintaining the full structure in the other parts of the model, including regressors, spatial connectivity, and heterogeneity. Wang and Lee (2013) studied a simpler model ($\rho_0 = 0$) under RE and CRE specifications. They pointed out the difficulty in estimating a general model under FE specification. Zhou et al. (2022) studied a model with response for each unit following a pure AR(1) process. Liu et al. (2023) studied a dynamic SPD-MR model without μ_0, α_0 and ρ_0.

In this paper, we focus on the FE specification of Model (1.2) to fill in a major gap in the SPD-MR literature. We contribute to the literature by introducing a general M-estimation framework for model estimation and a novel *corrected plug-in* method for model inference, both taking into account the estimation of fixed effects. The proposed methods are then extended to allow for serial correlation. Consistency and asymptotic normality of the proposed M-estimators are established, and consistency of the proposed corrected plug-in estimators is proved. Monte Carlo results show that the proposed methods perform very well in finite samples and that "discarding" the observations with missing responses can give misleading results. An empirical application of our methods to US tax competition data points to the existence of tax competition and path dependence in US state taxes. Our methods apply to matrix exponential spatial specification and can be extended to include higher-order spatial effects, etc.

Standard approaches in nonlinear panel data with fixed effects bias-correct (*i*) the estimator, (*ii*) the concentrated score, and (*iii*) the concentrated likelihood, as elaborated by Arellano and Hahn (2007). Our approach falls into (*ii*) but with major differences: it does not require data to be independent, it provides exact bias corrections, and it does not impose any conditions on n and T (see the end of Sec. 2.1 for more details).

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

Notations and conventions. First, $|\cdot|$, $\texttt{tr}(\cdot)$, $'$ and $\|A\|$ are the usual notations for determinant, trace, transpose and matrix norm. For a real matrix A of full rank, $\mathbb{P}_A = A(A'A)^{-1}A'$ and $\mathbb{Q}_A = I_n - \mathbb{P}_A$ are the projection matrices. $\texttt{diagv}(\cdot)$ forms a column vector by the diagonal elements of a square matrix; $\texttt{bdiag}(\cdots)$ a block diagonal matrix; $[\cdot\,,\cdot\,,\ldots\,,\cdot]$ a row vector; and $[\cdot\,;\cdot\,;\ldots\,;\cdot]$ a column vector.

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

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

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

Model (2.1) in fact allows the time-invariant and space-invariant covariates effects $(Z, \mathcal{Z}_t)$, such as gender and policy. Our view is that they are a part of the FEs and can be "decomposed" from $\mathbf{D}$ by adding further constraints on ϕ (see Appendix E, Sec. E.2).

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

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

where $\mathbb{Y} = \mathbf{\Omega}_N^{-\frac{1}{2}}\mathcal{S}\mathbf{Y}$, $\mathbb{X} = \mathbf{C}\mathbf{X}$, $\mathbb{D} = \mathbf{C}\mathbf{D}$, $\mathbb{V} = \mathbf{C}\mathbf{B}_{nT}^{-1}\mathbf{V}$, and $\mathbf{C} = \mathbf{\Omega}_N^{-\frac{1}{2}}\mathcal{S}\mathbf{A}_{nT}^{-1}$. It is easy to see that $\mathrm{Var}(\mathbb{V}) = \sigma_{v0}^2 I_N$, and thus $\mathbb{V} \sim N(0, \sigma_{v0}^2 I_N)$ **if** $\mathbf{V} \sim N(0, \sigma_{v0}^2 I_{nT})$.

2.1. The M-estimation

We seek the desired estimating functions for M-estimation by exploiting the concentrated quasi scores of $\theta = (\beta', \sigma_v^2, \delta')'$. The quasi Gaussian loglikelihood of (θ, ϕ) in terms of the observed $\mathcal{S}\mathbf{Y}$, given the **exogenous** $(\mathbf{X}, \mathcal{S})$ and **as if** $\mathbb{V} \sim N(0, \sigma_{v0}^2 I_N)$, is:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Solving $S_N^{*c}(\delta) = 0$ gives us the unconstrained M-estimator $\hat{\delta}_{\mathtt{M}}$ of δ, and the M-estimators of β and σ_v^2: $\hat{\beta}_{\mathtt{M}} \equiv \hat{\beta}_{\mathtt{M}}(\hat{\delta}_{\mathtt{M}})$ and $\hat{\sigma}_{v,\mathtt{M}}^2 \equiv \hat{\sigma}_{v,\mathtt{M}}^2(\hat{\delta}_{\mathtt{M}})$. The M-estimator of θ is $\hat{\theta}_{\mathtt{M}} = (\hat{\beta}_{\mathtt{M}}', \hat{\sigma}_{v,\mathtt{M}}^2, \hat{\delta}_{\mathtt{M}}')'$.

As discussed in the Introduction, the standard methods in dealing with the fixed effects problem in nonlinear panels, or in general the *incidental parameters problem* of Neyman and Scott (1948), are (in our context) to bias-correct $\hat{\theta}_{\mathtt{QML}}$, or $S_N^c(\theta)$, or $\ell_N^c(\theta)$ (Arellano and Hahn, 2007). These methods require responses to be independent and T to increase with n as the corrections are derived under large-T approximations. Our method falls into the second category but it does not impose these conditions and provides an exact bias correction on $S_N^c(\theta_0)$. In addition, our method allows for the estimation of time- or unit-invariant covariates effects as indicated below (2.1). With our method, further bias corrections on $S_N^{*c}(\delta_0)$ can be made to correct the effect of estimating β and σ_v^2 on the estimation of δ, in light of Yang (2015) and Yang et al. (2016). This is particularly meaningful when β is of a large dimension and spatial dependence is heavy. Finally, our methods can be extended to a GMM framework by adding extra moments.

Lee and Yu (2010a) bias-correct the QMLE of θ for a complete spatial panel with FE, which requires $\frac{n}{T^3} \to 0$ and $\frac{T}{n^3} \to 0$ for valid inference (see Lee, 2023, p.326). The second method can be traced back to Neyman and Scott (1948, Sec. 5) but has the smallest literature. The third method may not apply to the type of model we consider.

2.2. Asymptotic properties of M-estimator

To study the asymptotic properties of the proposed M-estimator, it is necessary that the errors, regressors, selection matrix, and spatial weight matrices satisfy certain basic conditions. Let Δ_ϖ be the parameter space for $\varpi = \lambda, \rho$ and $\Delta = \Delta_\lambda \times \Delta_\rho$. For a real symmetric matrix, $\gamma_{\min}(\cdot)$ and $\gamma_{\max}(\cdot)$ denote its smallest and largest eigenvalues. For a real matrix A, $\|A\|_1$ and $\|A\|_\infty$ are the maximum absolute column and row sum norms.

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

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

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

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

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

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

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

Assumptions A-E are standard in spatial econometrics or missing-data literature (Lee, 2004; Abrevaya, 2019). For technical convenience, $\mathbf{X}$ and $\mathcal{S}$ are treated as non-stochastic. They can instead be stochastic but strictly exogenous (w.r.t. $\mathbf{V}$). The analyses are then interpreted conditionally on $\mathbf{X}$ and $\mathcal{S}$ (White, 2001, p.6). The strict exogeneity of $\mathcal{S}$ is in line with Little and Rubin (2019); see Appendix E, Sec. E.2. Assumption E ensures that $\mathbf{\Omega}_N(\delta)$, its partial derivatives, and its inverse are uniformly bounded in both row and column sum norms, uniformly in $\delta \in \Delta$ (see Lemma B.2(i)).

Some additional technical assumptions are required. Note that $\mathbf{A}_{nT}(\lambda)$ and $\mathbf{C}(\delta)$ are both block diagonal. Denote their tth blocks by $A_t(\lambda)$ and $C_t(\delta)$, respectively.

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

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

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

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

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

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

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

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

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

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

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

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

2.3. Estimation of the VC matrix

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

$\Gamma_N^*(\theta_0)$ contains the common parameters θ_0, the fixed effects ϕ_0 embedded in Π_2, and the skewness κ_3 and excess kurtosis κ_4 of the idiosyncratic errors. The common *plug-in* method may not be valid due to the involvement of incidental parameters ϕ_0. A *corrected plug-in* method is proposed. Let $\Gamma_N^*(\hat{\theta}_{\mathtt{M}}) = \Gamma_N^*(\theta)|_{(\theta=\hat{\theta}_{\mathtt{M}},\phi=\hat{\phi}_{\mathtt{M}},\kappa_3=\hat{\kappa}_{3,N},\kappa_4=\hat{\kappa}_{4,N})}$ be the plug-in estimator, where $\hat{\phi}_{\mathtt{M}}$ is the M-estimator of ϕ (or a GLS estimator by regressing $\mathcal{S}[\mathbf{Y}-\mathbf{A}_{nT}^{-1}(\hat{\lambda}_{\mathtt{M}})\mathbf{X}\hat{\beta}_{\mathtt{M}}]$ on $\mathcal{S}\mathbf{A}_{nT}^{-1}(\hat{\lambda}_{\mathtt{M}})\mathbf{D}$ with weight $\boldsymbol{\Omega}_N(\hat{\delta}_{\mathtt{M}})$), and $\hat{\kappa}_{3,N}$ and $\hat{\kappa}_{4,N}$ are consistent estimators of κ_3 and κ_4. When both n and T are large, $\Gamma_N^*(\hat{\theta}_{\mathtt{M}})$ would be consistent as $\hat{\phi}_{\mathtt{M}}$ is. However, when either n or T is fixed, $\hat{\phi}_{\mathtt{M}}$ is not consistent and a bias correction is necessary after plugging $\hat{\phi}_{\mathtt{M}}$ into $\Gamma_N^*(\theta)$. We show that the only term that cannot be consistently estimated is the one quadratic in ϕ_0, embedded in $\Pi_2'\Pi_2$.

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

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

where $\mathrm{Bias}^*(\delta_0)$ *has a single nonzero element on the diagonal corresponding to the* λ-λ *entry, given by* $\frac{1}{N_1}\mathtt{tr}[(\mathbb{D}'\mathbb{D})^{-1}\mathbf{D}'\mathbb{J}'\mathbb{Q}_{\mathbb{D}}\mathbb{J}\mathbf{D}]$.

See the proof of Corollary 2.1 in Appendix E (Sec. E.3) for details. Corollary 2.1 leads immediately to a general consistent estimator of $\Gamma^*_N(\theta_0)$:

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

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

Finally, we provide consistent estimators for κ_3 and κ_4. As $\mathbf{V}$ is infeasible for estimation due to the incidental parameters problem and incompleteness, we start from $\mathbf{\Omega}_N^{-\frac{1}{2}}\tilde{\mathbb{V}} = \mathbf{\Omega}_N^{-\frac{1}{2}}\mathbb{Q}_{\mathbb{D}}\mathbf{\Gamma}\mathbf{V}$, which can be "consistently" estimated by $\mathbf{\Omega}_N^{-\frac{1}{2}}(\hat{\delta}_{\mathtt{M}})\hat{\mathbb{V}}(\hat{\beta}_{\mathtt{M}}, \hat{\delta}_{\mathtt{M}}) = \mathbf{\Omega}_N^{-\frac{1}{2}}(\hat{\delta}_{\mathtt{M}})\mathbb{Q}_{\mathbb{D}}(\hat{\delta}_{\mathtt{M}})\mathbf{\Omega}_N^{-\frac{1}{2}}(\hat{\delta}_{\mathtt{M}})\mathcal{S}[\mathbf{Y}-\mathbf{A}_{nT}^{-1}(\hat{\lambda}_{\mathtt{M}})\mathbf{X}\hat{\beta}_{\mathtt{M}}]$. Let q_{jk} be the (j,k)th element of $N\times nT$ matrix $\bar{\mathbb{Q}}_{\mathbb{D}} \equiv \mathbf{\Omega}_N^{-\frac{1}{2}}\mathbb{Q}_{\mathbb{D}}\mathbf{\Gamma}$. Denote the elements of $\mathbf{V}$ by $v_l, l = 1, \dots, nT$, and the elements of $\bar{\mathbb{Q}}_{\mathbb{D}}\mathbf{V}$ by $\tilde{v}_j, j = 1, \dots, N$, where l and j are the combined index of cross-sectional and time dimensions. Then, $\tilde{v}_j = \sum_{k=1}^{nT} q_{jk}v_k$, and thus $\text{E}(\tilde{v}_j^3) = \sum_{k=1}^{nT} q_{jk}^3\text{E}(v_k^3) = \sigma_{v0}^3\kappa_3\sum_{k=1}^{nT} q_{jk}^3$. Summing $\text{E}(\tilde{v}_j^3)$ over j gives $\kappa_3 = (\sum_{j=1}^{N}\text{E}(\tilde{v}_j^3))(\sigma_{v0}^3\sum_{j=1}^{N}\sum_{k=1}^{nT} q_{jk}^3)^{-1}$. Its sample analog:

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

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

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

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

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

Corollary 2.2. *Under the assumptions of Theorem 2.1, we have, as $N \to \infty$, (i) $\hat{\kappa}_{3,N} \xrightarrow{p} \kappa_{3,0}$ and $\hat{\kappa}_{4,N} \xrightarrow{p} \kappa_{4,0}$; (ii) $\widehat{\Sigma}^*_N - \Sigma^*_N(\theta_0) \xrightarrow{p} 0$ and $\widehat{\Gamma}^*_N - \Gamma^*_N(\theta_0) \xrightarrow{p} 0$; and therefore $\widehat{\Sigma}^{*-1}_N\widehat{\Gamma}^*_N\widehat{\Sigma}^{*-1\prime}_N - \Sigma^{*-1}_N(\theta_0)\Gamma^*_N(\theta_0)\Sigma^{*-1\prime}_N(\theta_0) \xrightarrow{p} 0$.*

3. M-Estimation with Serial Correlation

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

Assumption A′: The innovations follow an MA process, $v_{it} = e_{it} + \tau e_{i,t-1}$, for all i and t with $|\tau| < 1$, $e_{it} \sim \text{iid}(0, \sigma_e^2)$, and $E|e_{it}|^{4+\epsilon_0} < \infty$ for some $\epsilon_0 > 0$.

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

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

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

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

where $\mathbb{H}_\lambda(\delta)$, $\mathbb{H}_\rho(\delta)$, $\mathbb{J}(\delta)$, $\tilde{\mathbb{V}}(\beta,\delta)$, and $\boldsymbol{\varepsilon}(\beta,\delta)$ in (2.6) are redefined to accommodate the extra τ, and $\mathbb{H}_\tau(\delta)$ is defined as $\mathbb{H}_\rho(\delta)$. Solving $S_N^{\diamond}(\theta) = 0$ gives the M-estimator $\hat{\theta}_{\mathtt{M}}^{\diamond}$ of θ.

The asymptotic properties of $\hat{\theta}_{\mathtt{M}}^{\diamond}$ can be established in a similar way as for $\hat{\theta}_{\mathtt{M}}$ in Sec. 2, based on a similar set of assumptions (A$'$ given above and B$'$−G$'$ in Appendix D).

Theorem 3.1. *Under Assumptions A$'$−G$'$, as $N \to \infty$, if $\frac{n_t}{n} \to c_t$ and $\frac{T_i}{T} \to d_i$, where $c_t, d_i \in (0, 1]$, and $\min(T_i) \geq 2$ and $\min(n_t) \geq 2$, then $\hat{\theta}_{\mathtt{M}}^{\diamond} \xrightarrow{p} \theta_0$.*

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

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

which is linear-quadratic in $\mathcal{E}$ with iid elements. Again, the importance of this representation is two-fold: it allows the application of CLT for LQ forms of Kelejian and Prucha (2001) and Wold device to establish the asymptotic normality of $\frac{1}{\sqrt{N_1}} S_N^{\diamond}(\theta_0)$ (thus the asymptotic normality $\hat{\theta}_{\mathtt{M}}^{\diamond}$) and an easy derivation of $\text{Var}[S_N^{\diamond}(\theta_0)]$ as seen in Appendix A.

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

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

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

For statistical inference, $\Sigma_N^{\diamond}(\theta_0)$ is estimated by $\widehat{\Sigma}_N^{\diamond} = -\frac{1}{N_1} \frac{\partial}{\partial \theta'} S_N^{\diamond}(\theta)|_{\theta = \hat{\theta}_{\mathtt{M}}^{\diamond}}$. The analytical expressions of $\frac{\partial}{\partial \theta'} S_N^{\diamond}(\theta)$ and $\Gamma_N^{\diamond}(\theta_0)$ are given in Appendix A. Similar to $\Gamma_N^{*}(\theta_0)$ in

Sec. 2, $\Gamma_N^\diamond(\theta_0)$ contains the common parameters θ_0, the incidental parameters ϕ_0, and the skewness κ_3^e and excess kurtosis κ_4^e of idiosyncratic errors $\{e_{it}\}$. Again, the usual plug-in estimator would not lead to a consistent estimate of $\Gamma_N^\diamond(\theta_0)$.

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

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

where $\text{Bias}^\diamond(\delta_0)$ *has a single nonzero element on the diagonal corresponding to the* λ-λ *entry, given by* $\frac{1}{N_1}\mathtt{tr}[(\mathbb{D}'\mathbb{D})^{-1}\mathbf{D}'\mathbb{J}'\mathbb{Q}_{\mathbb{D}}\mathbb{J}\mathbf{D}]$.

Thus, a corrected plug-in estimator (corrected in $\Pi_2'\Pi_2$) is developed:

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

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

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

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

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

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

4. Monte Carlo Results

Extensive Monte Carlo experiments are conducted to investigate (*i*) the finite sample performance of the proposed M-estimator and the corresponding corrected plug-in estimator of the VC matrix, (*ii*) the consequence of discarding observations with missing responses, (*iii*) the effect of ignoring the estimation of fixed effects, and (*iv*) the performance of some related estimators. The following data-generating process is used:

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

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

The spatial weight matrices can be `Group` interaction or `Queen` contiguity. To generate W_t under `Queen`, 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 $W_{t,ij} = 1$ if square j shares a common boundary or vertex with square i and 0 otherwise. To generate W_t under `Group`, let $K(n) = Round(n^{0.5})$ be the number of groups and then generate $K(n)$ group sizes according to a uniform distribution. The distribution of the idiosyncratic errors can be (1) `normal`, (2) standardized `normal mixture` (10% $N(0, 4^2)$ and 90% $N(0, 1)$), or (3) standardized `chi-square` with 3 degrees of freedom. See Yang (2015) for details. Both the case of iid errors and the case of serially correlated errors ($\tau = 0.5$) are considered.

The selection matrices $\mathcal{S}_t$ are generated according to two mechanisms: (i) MAR (missing at random) or (ii) MCAR (missing completely at random). The former depends on X_t and ϕ, but the latter simply on the outcomes of independent Bernoulli trials with the probability of missing p_t for period t. We design a MAR mechanism such that the missing percentage is about 25% (see Appendix E, Sec. E.5 for detail), and choose $p_t = 0.1$ or 0.3 for MCAR mechanism to see the effect of the degree of missingness.

Our Monte Carlo experiments involve nine estimators, but the main ones are `ME-MR` (the proposed M-estimator), `ME-GU` (M-estimator assuming *genuine unbalancedness* (GU) after deleting observations with missingness, considered in Meng and Yang (2021)), and

QMLE-MR (the QML estimator ignoring the effect of estimating the FEs). With these, the issues (*i*)-(*iii*) are addressed. The remaining six estimators relate to the "existing" methods, in particular the imputation methods, which address the issue (*iv*).

Table 1 contains partial Monte Carlo results on the three main estimators for the case of iid errors and MCAR. The results show an excellent performance of the proposed M-estimation and inference methods, irrespective of the error distributions, the spatial layouts, parameter values, as well as the missing percentage. In contrast, the QMLE-MRs (the closest to ME-MRs) of spatial parameters do not perform as well as the ME-MRs. This shows the consequence of ignoring the effects of estimating the FE parameters. By comparing ME-GU with ME-MR, we can see the consequences of treating MR mechanism as GU mechanism: ME-GUs of the spatial parameters perform poorly even when the sample size is fairly large. When the missing percentage is higher, ME-GUs become more biased. This is consistent with our expectation: treating MR as GU ignores spatial effects from the deleted units and the larger the missing percentage, the more serious the consequence.

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

Due to space constraints, we report the Monte Carlo results under MAR mechanism in online Appendix E (Table 9). Again, the results show that the proposed M-estimation and inference methods perform excellently in finite sample, and that their performance is not affected by allowing missingness to depend on the regressors and fixed effects.

While strictly speaking there are no existing methods for use in estimating our models, some may relate to ours. These include the three imputation estimators, one nonlinear least square estimator (Wang and Lee, 2013), one naïve estimator, and a QMLE under GU. See Appendix E (Sec. E.5) for a detailed definition of these estimators. It is interesting to know how these estimators perform in estimating our model. We therefore included these six estimators in our Monte Carlo experiments. A much larger set of Monte Carlo results, including these reported in the main text, is given in Appendix E (Sec. E.5). The results show that none of these six estimators perform satisfactorily.

5. An Empirical Application

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

In this section, we reconsider this study under the missing-on-response-only (MR) mechanism since the explanatory variables can be fully observed over a chosen period.

We construct two panels based on 48 contiguous US states over 19 years (1978-1996), the tax rates on beer and the tax rates on gasoline. The numbers of observations for beer and gasoline tax rates are, respectively, 911 and 888. We define the spatial neighboring states as those that share a common border. The overall spatial weight matrix W is row-normalized. The explanatory variables we use are state size (*Size*, measured by population density), spatially weighted size (*WSize*), dependency ratio (*DR*), government ideological orientation (*GIO*), lagged sales tax rate (*LSTR*), gross state product (*GSP*, in trillion), and public expenditure (*PE*, in billion). With these, we write the model as

$$\begin{aligned}\mathcal{S}_t Tax_t = \mathcal{S}_t A^{-1}(\lambda)\big(&Size_t\beta_1 + W \times Size_t\beta_2 + DR_t\beta_3 + GIO_t\beta_4 + LSTR_t\beta_5 \\ &+GSP_t\beta_6 + PE_t\beta_7 + \mu + \alpha_t l_n + U_t\big), \quad U_t = \rho W U_t + V_t, \;\; t = 1, \ldots, 19.\end{aligned}$$

Among the various model parameters, λ and τ are of particular interest as they quantify the intensity of tax competition and the path dependence in setting state tax rates.

Table 3 gives a descriptive summary of the data. Tables 4 and 5 summarize the empirical results for the beer tax rates and the gasoline tax rates, respectively. Besides the five estimators involved in the above Monte Carlo study, two additional M-estimators, `ME-IMR` and `ME-IMRSC`, based on *imputed data* (Honaker and King, 2010) under iid errors and serially correlated errors, respectively, are also included.

Our analyses lead to a deeper understanding of the mechanism of tax competition and offer more insight into the nature of spatial interactions. Both analyses based on the proposed methods point to the existence of strong and positive endogenous spatial spillover effects and strong and positive serial correlation. These imply that states mimic neighbors' tax moves (tax competition) and competition persists over time (a point not considered by Egger et al., 2005). They help mitigate revenue erosion and underscore the importance of multi-year fiscal planning for both temporary and permanent tax reforms.

From Table 4, `ME-MR` shows that the `SL` effect is significant and positive at 10% level, indicating the presence of beer tax competition. In contrast, none of the remaining four estimates reveals this, highlighting the limitations of QML, GU-based and imputation-based methods. QML method ignores the effect of estimating fixed effects; GU method ignores the spatial effects of the units with missing responses; and imputation methods do not account for spatial dependence during imputation. Interestingly, although the beer tax rates data have only one missing response, ignoring the spatial effects from this observation either through `ME-GU` or `ME-IMR` completely changes the conclusion on tax competition. All estimates show that the `SE` effect is negative but insignificant, consistent with findings of Egger et al. (2005). However, our proposed methods are able to tell that the `SE` effect is insignificant. Furthermore, the three MRSC-based estimates reveal that the serial correlation is positive and significant, suggesting the presence of path dependence in state beer tax rate decisions. `ME-MRSC` successfully identifies significant tax competition at 5% level, whereas `QMLE-MRSC` does not. After imputing the single missing response, `ME-IMRSC` shows the `SL` effect becomes less significant compared to `ME-MRSC`.

From Table 5, `ME-IMR`, `QMLE-MR` and `ME-MR` all provide significant evidence for a positive `SL` effect, indicating the existence of gasoline tax competition. However, this effect is not captured by GU-based estimates. Both `QMLE-MR` and `ME-IMR` appear to underestimate the competition effects compared to the proposed `ME-MR`. The underestimation by `QMLE-MR` may result from the incidental parameters problem, while that by `ME-IMR` may stem from neglecting the spatial dependence during imputation. Most estimates of the `SE` coefficient show insignificant `SE` effect. Furthermore, `ME-IMRSC` of the `SL` parameter is smaller than the proposed `ME-MRSC`, again showing an underestimation of competition effects. The `QMLE-MRSC` shows the `SL` effect is insignificant, and thus fails to capture the competition effects. Lastly, all estimates show serial correlation is significant and positive, further supporting the presence of path dependence in tax-setting decisions.

6. Conclusions and Discussions

We consider fixed effects estimation of spatial panel data models with missing responses. It allows for unobserved spatiotemporal heterogeneity, time-varying endogenous and contextual spatial interactions, time-varying cross-sectional error dependence, and serial correlation. We propose an M-estimation method for model estimation and a corrected plug-in method for model inference, both taking into account the effects of estimating the fixed effects. We study the asymptotic and finite sample properties of the proposed methods. We apply our methods to US state tax competition data, leading to a much deeper understanding of the tax competition mechanism. Our methods allow for the estimation of time or unit invariant covariates effects, such as *gender* and *policy*, by imposing relevant constraints on the FE parameters ϕ and the $\mathbf{D}$ matrix in Model (2.1).

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

Acknowledgments

We are grateful to Editor, Ivan Canay, an Associate Editor and two referees for their constructive comments that have led to significant improvement of the paper. We thank the participants of AMES 2021, Curtin University, Malaysia, and the seminar participants at WISE, Xiamen University, China, for their helpful comments.

Disclosure Statement

The authors report there are no competing interests to declare.

Appendix A: AQS, Hessian, and Variance of AQS

A.1. Derivation for Section 2

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

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

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

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

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

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

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

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

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

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

$$H_{\beta\rho}^*(\theta) = \tfrac{1}{\sigma_v^2}\mathbf{X}'\mathbf{A}_{nT}^{-1\prime}(\lambda)\mathcal{S}'\dot{\Psi}_\rho(\delta)\mathcal{V}(\beta,\lambda), \quad H_{\sigma_v^2\sigma_v^2}^*(\theta) = -\tfrac{1}{\sigma_v^6}\tilde{\mathbb{V}}'(\beta,\delta)\tilde{\mathbb{V}}(\beta,\delta) + \tfrac{1}{2\sigma_v^4}N_1,$$

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

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

$$\begin{aligned} H_{\lambda\lambda}^*(\theta) = {} & \tfrac{2}{\sigma_v^2}\mathcal{V}'(\beta,\lambda)\dot{\Psi}_\lambda(\delta)\mathcal{S}[\tfrac{\partial}{\partial\lambda}\mathbf{A}_{nT}^{-1}(\lambda)]\mathbf{X}\beta + \tfrac{2}{\sigma_v^2}\tilde{\mathbb{V}}'(\beta,\delta)\mathbb{J}(\delta)\mathbf{W}_{nT}\mathbf{A}_{nT}^{-1}(\lambda)\mathbf{X}\beta \\ & -\tfrac{1}{\sigma_v^2}\beta'\mathbf{X}'\mathbb{J}'(\delta)\mathbb{Q}_{\mathbb{D}}(\delta)\mathbb{J}(\delta)\mathbf{X}\beta - \tfrac{1}{2\sigma_v^2}\mathcal{V}'(\beta,\lambda)\ddot{\Psi}_{\lambda\lambda}(\delta)\mathcal{V}(\beta,\lambda) \\ & -\tfrac{1}{2}\mathtt{tr}[\dot{\mathbf{\Omega}}_\lambda(\delta)\dot{\Psi}_\lambda(\delta) + \ddot{\mathbf{\Omega}}_{\lambda\lambda}(\delta)\Psi(\delta)], \end{aligned}$$

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

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

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

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

The VC matrix. For stochastic terms of the forms in (2.10), we show that, for $r, s = 1,2,3$, (i) $\text{Cov}(\Pi_r'\mathbf{V},\ \Pi_s'\mathbf{V}) = \sigma_{v0}^2\Pi_r'\Pi_s$; (ii) $\text{Cov}(\mathbf{V}'\Phi_r\mathbf{V},\ \Pi_s'\mathbf{V}) = \sigma_{v0}^3\kappa_3\varphi_r'\Pi_s$; and (iii) $\text{Cov}(\mathbf{V}'\Phi_r\mathbf{V},\ \mathbf{V}'\Phi_s\mathbf{V}) = \sigma_{v0}^4\kappa_4\varphi_r'\phi_s + \sigma_{v0}^4\mathtt{tr}(\Phi_r\Phi_s^\circ)$, where $\varphi_r = \mathtt{diagv}(\Phi_r)$ and $\Phi_s^\circ = \Phi_s + \Phi_s'$. Apply these results to (2.10), we obtain,

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

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

A.2. Derivation for Section 3.

The Hessian matrix. With redefined $\mathbf{\Omega}_N(\delta)$, the non-$\tau$-block of $H_N^\diamond(\theta) = \frac{\partial}{\partial\theta'}S_N^\diamond(\theta)$ has the same form as $H_N^*(\theta)$ in Sec. 2. Extending the notations, $\dot{\mathbf{\Omega}}_\omega(\delta)$, $\ddot{\mathbf{\Omega}}_{\omega\varpi}(\delta)$, $\dot{\Psi}_\omega(\delta)$, and $\ddot{\Psi}_{\omega\varpi}(\delta)$ of Sec. 2 to $\omega,\varpi = \lambda,\rho,\tau$, we obtain the τ-components of $H_N^\diamond(\theta)$:

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

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

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

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

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

The VC matrix. Applying the results leading to $\text{Var}[S_N^*(\theta_0)]$ on (3.2), we obtain,

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

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

Appendix B: Some Basic Lemmas

The following lemmas are essential to the proofs of the main results in Sections 2 and 3. Lemmas B.2 and B.3 are new and their proofs are given in Appendix E (Sec. E.1).

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

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

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

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

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

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

Lemma B.4. *(Lee, 2004): Let A_N be an $N \times N$ matrix bounded in both row and column sum norms, with elements of uniform order $O(h_N^{-1})$, c_N be an $N \times 1$ vector with elements of uniform order $O(h_n^{-1/2})$, and $\mathbf{V} = (v_1, \ldots, v_N)'$ with $v_j \sim iid(0, \sigma^2)$. Then,*

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

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

(v) $c_N'A_N\mathbf{V} = O_p((\frac{N}{h_n})^{\frac{1}{2}})$.

Appendix C: Proofs for Section 2

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

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

where $\bar{\mathbb{V}}(\delta) = \tilde{\mathbb{V}}(\bar{\beta}_{\mathtt{M}}(\delta), \delta)$, obtained by first solving $\bar{S}_N^*(\theta) = \mathrm{E}[S_N^*(\theta)] = 0$ for β and σ^2:

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

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

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

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

$$\bar{\sigma}^2_{v,\mathtt{M}}(\delta) = \tfrac{1}{N_1}\mathrm{E}[\mathbb{Y}'(\delta)\mathbf{Q}(\delta)\mathbb{Y}(\delta)] + \tfrac{1}{N_1}\mathrm{E}\big\{[\mathbb{Y}(\delta) - \mathrm{E}(\mathbb{Y}(\delta))]'\mathbf{P}(\delta)[\mathbb{Y}(\delta) - \mathrm{E}(\mathbb{Y}(\delta))]\big\} \quad \text{(C.3)}$$

$$= \tfrac{1}{N_1}\eta'\mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)\mathbf{Q}(\delta)\mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)\eta + \tfrac{\sigma^2_{v0}}{N_1}\mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\delta)\mathcal{O}_N(\delta)], \quad \text{(C.4)}$$

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

The first term of (C.4) 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,

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

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

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

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

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

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

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

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

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

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

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

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

where $\bar{\delta}$ lies between δ_1 and δ_2. It follows that $\frac{1}{N_1}\mathbb{Y}'(\delta)\mathbf{Q}(\delta)\mathbb{Y}(\delta)$ is stochastically equicontinuous as $\sup_{\delta\in\Delta}\frac{1}{N_1}\mathbf{Y}'\mathcal{S}'[\frac{\partial}{\partial\varpi}\bar{\mathbf{Q}}(\delta)]\mathcal{S}\mathbf{Y} = O_p(1)$, $\varpi = \lambda, \rho$ (See Appendix E, Sec. E.3 for details). With the pointwise convergence of $\frac{1}{N_1}[\mathbb{Y}'(\delta)\mathbf{Q}(\delta)\mathbb{Y}(\delta) - \mathrm{E}(\mathbb{Y}'(\delta)\mathbf{Q}(\delta)\mathbb{Y}(\delta))]$ to zero for each $\delta \in \Delta$ and the stochastic equicontinuity of $\frac{1}{N_1}\mathbb{Y}'(\delta)\mathbf{Q}(\delta)\mathbb{Y}(\delta)$, the uniform convergence result, $\sup_{\delta\in\Delta}|\frac{1}{N_1}[\mathbb{Y}'(\delta)\mathbf{Q}(\delta)\mathbb{Y}(\delta) - \mathrm{E}(\mathbb{Y}'(\delta)\mathbf{Q}(\delta)\mathbb{Y}(\delta))]| = o_p(1)$, follows (Andrews, 1992). Thus, the result (b) is proved.

Proof of (c). We show only $\sup_{\delta\in\Delta}\frac{1}{N_1}\big|\hat{\mathbb{V}}'(\delta)\mathbb{H}_\lambda(\delta)\hat{\mathbb{V}}(\delta) - \mathrm{E}[\bar{\mathbb{V}}'(\delta)\mathbb{H}_\lambda(\delta)\bar{\mathbb{V}}(\delta)]\big| = o_p(1)$, as he other part is similar. By $\mathbb{H}_\lambda(\delta)$, $\hat{\mathbb{V}}(\delta)$ and $\bar{\mathbb{V}}(\delta)$ given below (2.6) and in the proofs of (a) and (b) above, we can write $\frac{1}{N_1}\hat{\mathbb{V}}'(\delta)\mathbb{H}_\lambda(\delta)\hat{\mathbb{V}}(\delta) - \frac{1}{N_1}\mathrm{E}[\bar{\mathbb{V}}'(\delta)\mathbb{H}_\lambda(\delta)\bar{\mathbb{V}}(\delta)]$ as

$$\frac{1}{N_1}[\mathbf{Y}'\mathcal{S}'\bar{\mathbf{Q}}(\delta)(\frac{\partial}{\partial\lambda}\mathbf{\Omega}_N(\delta))\bar{\mathbf{Q}}(\delta)\mathcal{S}\mathbf{Y} - \mathrm{E}(\mathbf{Y}'\mathcal{S}'\bar{\mathbf{Q}}(\delta)(\frac{\partial}{\partial\lambda}\mathbf{\Omega}_N(\delta))\bar{\mathbf{Q}}(\delta)\mathcal{S}\mathbf{Y})]$$
$$-\frac{\sigma_{v0}^2}{N_1}\mathtt{tr}[\bar{\mathbf{P}}(\delta)(\frac{\partial}{\partial\lambda}\mathbf{\Omega}_N(\delta))\bar{\mathbf{P}}(\delta)\mathbf{\Omega}_N],$$

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

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

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

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

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

$$0 = \tfrac{1}{\sqrt{N_1}} S_N^*(\hat{\theta}_{\mathtt{M}}) = \tfrac{1}{\sqrt{N_1}} S_N^*(\theta_0) + \big[\tfrac{1}{N_1}\tfrac{\partial}{\partial\theta'} S_N^*(\theta)\big|_{\theta=\bar{\theta}_r \text{ in } r\text{th row}}\big]\sqrt{N_1}(\hat{\theta}_{\mathtt{M}} - \theta_0), \qquad \text{(C.5)}$$

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. Rewrite $\dot{\Psi}_\lambda(\delta)$ in (A.2) as $-\Psi(\delta)\dot{\mathbf{\Omega}}_\lambda(\delta)\Psi(\delta) - \Psi(\delta)\mathbb{K}(\delta) - \mathbb{K}'(\delta)\Psi(\delta)$ and $\dot{\Psi}_\rho(\delta)$ in (A.3) as $-\Psi(\delta)\dot{\mathbf{\Omega}}_\rho(\delta)\Psi(\delta)$. Following exactly the same way of proving Lemma B.2(*ii*), we show that both $\mathbb{K}(\delta)$ (defined below (A.3)) and $\frac{\partial}{\partial\omega}\mathbb{K}(\delta)$, $\omega = \lambda, \rho$ are uniformly bounded in both row and column sums, uniformly in $\delta \in \Delta$. In addition, the proof of Lemma B.2(*i*) also implies $\ddot{\mathbf{\Omega}}_{\omega\varpi}(\delta)$, $\omega, \varpi = \lambda, \rho$ is bounded in row and column sum norms, uniformly in $\delta \in \Delta$. Thus, by Lemma B.1, we have $\dot{\Psi}_\omega(\delta)$ and $\ddot{\Psi}_{\omega\varpi}(\delta)$, $\omega, \varpi = \lambda, \rho$ are all bounded in row and column sum norms, uniformly in $\delta \in \Delta$. With these, $\tilde{\mathbb{V}}(\beta_0, \delta_0) = \mathbb{Q}_{\mathbb{D}}\mathbf{\Gamma}\mathbf{V}$ and $\mathcal{V}(\beta_0, \lambda_0) = \boldsymbol{\mathcal{S}}\mathbf{A}_{nT}^{-1}[\mathbf{D}\phi_0 + \mathbf{B}_{nT}^{-1}\mathbf{V}]$, 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 $\bar{\delta}$, the result follows. For the non-stochastic part, the results can also be shown using the MVT (See Appendix E, Sec. E.3, for details).

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

Proof of Corollary 2.1. See Appendix E, Sec. E.3.

Proof of Corollary 2.2. See Appendix E, Sec. E.3.

Appendix D: Proofs for Section 3

Let now $\Delta = \Delta_\lambda \times \Delta_\rho \times \Delta_\tau$ be the parameter space for $\delta = (\lambda, \rho, \tau)'$, where Δ_ϖ is the parameter space for $\varpi = \lambda, \rho, \tau$. Let $\mathbf{C}(\delta) = \mathbf{\Omega}_N^{-\frac{1}{2}}(\delta)\mathcal{S}\mathbf{A}_{nT}^{-1}(\lambda)$, $\mathbb{X}(\delta) = \mathbf{C}(\delta)\mathbf{X}$, $\mathbb{D}(\delta) = \mathbf{C}(\delta)\mathbf{D}$, and $\mathbb{Q}_{\mathbb{D}}(\delta)$ be the projection matrix based on $\mathbb{D}(\delta)$, where $\mathbf{\Omega}_N(\delta) \equiv \mathcal{S}\mathbf{A}_{nT}^{-1}(\lambda)\mathbf{B}_{nT}^{-1}(\rho)[\Upsilon(\tau)\Upsilon'(\tau) \otimes I_n]\mathbf{B}_{nT}^{-1\prime}(\rho)\mathbf{A}_{nT}^{-1\prime}(\lambda)\mathcal{S}'$. The additional assumptions are:

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

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

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

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

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

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

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

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

Assumptions B′-E′ are either similar to or the same as Assumptions B-E. Assumption F′ extends Assumption F as $\mathbf{\Omega}_N(\delta)$ is no longer block diagonal. Assumption G′ extends Assumption G, and a more primitive version of it is given in Appendix E (Sec. E.2). Proofs of the results in Sec. 3 extend those in Sec. 2 (see `Appendix E`, Sec. E.4).

Appendix E: Online Supplementary Material

The `Supplementary Material` contains proofs of the two new lemmas in Appendix B, details on some important issues (literature, time/space invariant effects, computing) and technical assumptions, detailed proofs of the theories in the main text, a complete set of Monte Carlo results, and an additional application.

References

Abrevaya, J. (2019). "Missing dependent variables in fixed-effects models." *Journal of Econometrics*, 211(1), 151–165.

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

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

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

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

Honaker, J. and King, G. (2010). "What to do about missing values in time-series cross-section data." *American Journal of Political Science*, 54(2), 561–581.

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

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

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

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

Lee, L.-F. (2004). "Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models." *Econometrica*, 72, 1899–1925.

Lee, L.-F. (2023). *Spatial Econometrics: Spatial Autoregressive Models*. World Scientific.

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

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

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

Little, R. J. (1988). "A test of missing completely at random for multivariate data with missing values." *Journal of the American Statistical Association*, 83(404), 1198–1202.

Little, R. J. and Rubin, D. B. (2019). *Statistical Analysis with Missing Data*, Vol. 793. John Wiley & Sons.

Liu, J., Zhou, J., Lan, W., and Wang, H. (2023). "Spatial dynamic panel models with missing data." *Stat*, 12(1), e585.

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

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

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

Neyman, J. and Scott, E. L. (1948). "Consistent Estimates Based on Partially Consistent Observations." *Econometrica*, 16, 1–32.

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

Pacini, D. and Windmeijer, F. A. (2015). "Moment conditions for AR(1) panel data models with missing outcomes." *Discussion Papers 15/660*, University of Bristol, UK.

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

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

White, H. (2001). *Asymptotic Theory for Econometricians.* Emerald Group Publishing Limited.

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

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

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

Table 1: Empirical bias(sd)[$\hat{se}$] of estimators, MR model with iid errors.
Missing percentage=10%, $(\beta, \lambda, \rho, \sigma_v^2) = (1, 0.2, 0.2, 1)$, and `W = Group` and `M = Queen`.

	T=5			**T=10**		
	`QMLE-MR`	`ME-GU`	`ME-MR`	`QMLE-MR`	`ME-GU`	`ME-MR`
$n = 100$; error = 1, 2, 3, for the three panels below; Missing percentage = 10%						
β	.0010(.027)	.0042(.027)[.027]	.0007(.027)[.027]	-.0005(.018)	.0025(.018)[.018]	-.0007(.018)[.018]
λ	-.0120(.052)	-.0438(.054)[.051]	-.0009(.051)[.051]	-.0122(.033)	-.0221(.034)[.034]	-.0016(.033)[.033]
ρ	-.0182(.133)	-.0147(.105)[.106]	-.0095(.114)[.103]	-.0271(.071)	-.0018(.065)[.066]	-.0044(.064)[.064]
σ_v^2	-.2408(.058)	.0068(.076)[.076]	-.0122(.075)[.075]	-.1238(.043)	.0083(.049)[.051]	-.0050(.049)[.050]
β	-.0003(.027)	.0028(.027)[.027]	-.0005(.027)[.027]	.0003(.019)	.0033(.019)[.018]	.0000(.019)[.018]
λ	-.0190(.055)	-.0496(.054)[.051]	-.0077(.054)[.052]	-.0111(.032)	-.0206(.034)[.034]	-.0006(.032)[.032]
ρ	-.0200(.131)	-.0158(.103)[.106]	-.0121(.108)[.103]	-.0278(.068)	-.0025(.062)[.066]	-.0051(.061)[.064]
σ_v^2	-.2354(.127)	.0130(.167)[.159]	-.0052(.165)[.158]	-.1295(.105)	.0013(.120)[.113]	-.0115(.119)[.112]
β	-.0002(.027)	.0030(.027)[.027]	-.0004(.027)[.027]	.0007(.018)	.0037(.018)[.018]	.0005(.018)[.018]
λ	-.0181(.053)	-.0471(.051)[.051]	-.0068(.052)[.052]	-.0148(.034)	-.0241(.035)[.034]	-.0041(.033)[.033]
ρ	-.0125(.133)	-.0084(.104)[.105]	-.0049(.114)[.102]	-.0296(.072)	-.0044(.066)[.066]	-.0066(.064)[.064]
σ_v^2	-.2433(.095)	.0023(.125)[.116]	-.0154(.124)[.115]	-.1201(.074)	.0118(.084)[.082]	-.0007(.084)[.082]
$n = 400$; error = 1, 2, 3, for the three panels below; Missing percentage = 10%						
β	.0004(.013)	.0008(.013)[.013]	.0003(.013)[.013]	.0000(.009)	-.0005(.009)[.009]	-.0001(.009)[.009]
λ	-.0083(.036)	-.0457(.036)[.035]	-.0033(.036)[.036]	-.0081(.024)	-.0487(.022)[.022]	-.0022(.024)[.024]
ρ	.0426(.064)	-.0005(.053)[.053]	-.0025(.052)[.051]	.0111(.035)	.0011(.033)[.032]	-.0006(.032)[.031]
σ_v^2	-.2286(.030)	.0092(.038)[.039]	-.0005(.038)[.039]	-.1152(.023)	.0067(.025)[.025]	-.0010(.025)[.025]
β	-.0002(.013)	.0002(.013)[.013]	-.0002(.013)[.013]	-.0004(.009)	-.0008(.009)[.009]	-.0005(.009)[.009]
λ	-.0092(.037)	-.0464(.037)[.035]	-.0041(.037)[.036]	-.0064(.025)	-.0472(.023)[.022]	-.0005(.025)[.024]
ρ	.0446(.061)	.0011(.051)[.053]	-.0012(.050)[.051]	.0114(.034)	.0009(.032)[.032]	-.0004(.031)[.031]
σ_v^2	-.2308(.063)	.0063(.082)[.083]	-.0034(.082)[.083]	-.1128(.051)	.0096(.058)[.058]	.0017(.058)[.057]
β	.0001(.014)	.0005(.014)[.013]	.0001(.014)[.013]	-.0001(.009)	-.0005(.009)[.009]	-.0002(.009)[.009]
λ	-.0081(.035)	-.0460(.035)[.035]	-.0030(.034)[.036]	-.0080(.025)	-.0486(.023)[.022]	-.0021(.025)[.024]
ρ	.0435(.063)	-.0003(.052)[.053]	-.0021(.051)[.051]	.0120(.034)	.0016(.032)[.032]	.0002(.031)[.031]
σ_v^2	-.2301(.049)	.0073(.063)[.061]	-.0025(.063)[.061]	-.1143(.038)	.0078(.043)[.042]	.0001(.043)[.041]
$n = 100$; error = 1, 2, 3, for the three panels below; Missing percentage = 30%						
β	-.0002(.035)	.0019(.035)[.035]	-.0003(.034)[.035]	.0014(.021)	.0045(.020)[.021]	-.0006(.021)[.021]
λ	-.0182(.064)	-.0994(.055)[.055]	-.0087(.064)[.061]	-.0117(.036)	-.0615(.035)[.039]	-.0020(.039)[.039]
ρ	-.0515(.246)	-.0098(.158)[.161]	-.0099(.145)[.145]	-.0408(.095)	.0015(.086)[.090]	-.0145(.089)[.087]
σ_v^2	-.3128(.073)	.0227(.098)[.093]	-.0202(.092)[.091]	-.1576(.050)	.0272(.058)[.060]	-.0046(.063)[.059]
β	-.0003(.034)	.0015(.034)[.035]	-.0008(.036)[.035]	.0014(.020)	.0045(.020)[.021]	-.0004(.021)[.021]
λ	-.0161(.064)	-.1003(.055)[.055]	-.0040(.064)[.061]	-.0156(.039)	-.0632(.040)[.039]	-.0042(.039)[.039]
ρ	-.0393(.259)	-.0011(.163)[.161]	-.0170(.144)[.147]	-.0475(.103)	-.0052(.093)[.091]	-.0057(.088)[.086]
σ_v^2	-.3205(.140)	.0153(.184)[.182]	-.0087(.200)[.182]	-.1473(.115)	.0381(.133)[.132]	-.0137(.125)[.128]
β	-.0033(.036)	-.0013(.035)[.035]	-.0017(.035)[.035]	-.0016(.022)	.0015(.022)[.021]	.0016(.020)[.021]
λ	-.0140(.066)	-.0989(.057)[.055]	-.0049(.061)[.062]	-.0159(.039)	-.0638(.041)[.039]	-.0018(.041)[.039]
ρ	-.0683(.246)	-.0224(.157)[.162]	-.0082(.144)[.148]	-.0350(.098)	.0060(.090)[.089]	-.0058(.082)[.086]
σ_v^2	-.3183(.102)	.0161(.138)[.133]	-.0174(.141)[.136]	-.1573(.088)	.0258(.100)[.095]	-.0103(.100)[.092]
$n = 400$; error = 1, 2, 3, for the three panels below; Missing percentage = 30%						
β	.0001(.016)	-.0014(.016)[.016]	-.0003(.016)[.016]	-.0001(.010)	-.0005(.010)[.010]	-.0002(.010)[.010]
λ	-.0087(.046)	-.0978(.039)[.037]	-.0050(.044)[.042]	-.0066(.030)	-.0989(.028)[.024]	-.0008(.027)[.027]
ρ	.0631(.091)	.0035(.073)[.072]	-.0009(.064)[.067]	.0127(.046)	.0033(.044)[.044]	-.0041(.041)[.040]
σ_v^2	-.2990(.036)	.0127(.047)[.045]	-.0037(.044)[.045]	-.1478(.025)	.0189(.029)[.030]	-.0016(.030)[.029]
β	-.0007(.016)	-.0022(.016)[.016]	.0001(.016)[.016]	-.0008(.010)	-.0012(.010)[.010]	-.0003(.010)[.010]
λ	-.0100(.049)	-.0975(.039)[.037]	-.0020(.041)[.042]	-.0074(.028)	-.0974(.025)[.024]	-.0014(.026)[.026]
ρ	.0649(.092)	.0049(.073)[.072]	-.0030(.065)[.067]	.0153(.044)	.0061(.043)[.044]	.0015(.041)[.040]
σ_v^2	-.2967(.071)	.0162(.093)[.095]	-.0096(.087)[.092]	-.1535(.059)	.0119(.067)[.065]	-.0027(.067)[.065]
β	.0008(.016)	-.0007(.016)[.016]	-.0012(.016)[.016]	.0007(.011)	.0003(.011)[.010]	.0004(.010)[.010]
λ	-.0100(.048)	-.0984(.040)[.037]	-.0061(.044)[.042]	-.0056(.028)	-.0968(.024)[.024]	-.0017(.026)[.026]
ρ	.0673(.088)	.0051(.071)[.072]	-.0045(.066)[.067]	.0116(.047)	.0023(.045)[.044]	.0005(.039)[.040]
σ_v^2	-.2992(.052)	.0129(.068)[.069]	-.0063(.070)[.069]	-.1483(.041)	.0184(.047)[.048]	-.0040(.046)[.047]

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

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

	T=5		T=10	
	`QMLE-MRSC`	`ME-MRSC`	`QMLE-MRSC`	`ME-MRSC`
	$n = 100$; error = 1, 2, 3, for the three panels below			
β	.0024(.028)	.0011(.024)[.024]	-.0002(.015)	-.0004(.015)[.016]
λ	-.0150(.052)	-.0038(.043)[.044]	-.0089(.030)	-.0001(.029)[.029]
ρ	-.0098(.147)	-.0039(.096)[.094]	-.0142(.065)	-.0051(.057)[.057]
τ	-.6263(.573)	.0183(.082)[.079]	-.0829(.045)	.0019(.038)[.040]
σ_v^2	-.3089(.156)	-.0195(.077)[.077]	-.1159(.044)	-.0080(.049)[.050]
β	.0005(.032)	.0007(.025)[.024]	.0006(.016)	.0003(.016)[.016]
λ	-.0123(.052)	-.0028(.044)[.044]	-.0097(.030)	-.0003(.029)[.029]
ρ	-.0054(.139)	-.0034(.094)[.094]	-.0158(.066)	-.0064(.057)[.057]
τ	-.6840(.594)	.0163(.095)[.096]	-.0785(.049)	.0061(.043)[.045]
σ_v^2	-.3171(.182)	-.0084(.153)[.157]	-.1186(.097)	-.0113(.109)[.108]
β	.0002(.029)	-.0008(.023)[.024]	.0005(.016)	.0001(.016)[.016]
λ	-.0154(.053)	-.0036(.045)[.044]	-.0117(.032)	-.0021(.032)[.029]
ρ	-.0173(.145)	-.0092(.095)[.095]	-.0145(.066)	-.0055(.058)[.057]
τ	-.6763(.590)	.0140(.080)[.085]	-.0803(.047)	.0042(.042)[.042]
σ_v^2	-.3189(.175)	-.0148(.123)[.114]	-.1185(.072)	-.0111(.081)[.079]
	$n = 400$; error = 1, 2, 3, for the three panels below			
β	-.0007(.014)	-.0008(.012)[.012]	.0000(.008)	-.0001(.008)[.008]
λ	-.0094(.036)	-.0035(.031)[.032]	-.0068(.024)	-.0005(.024)[.024]
ρ	.0478(.063)	-.0027(.046)[.046]	.0209(.031)	-.0001(.028)[.028]
τ	-.4706(.399)	.0029(.036)[.037]	-.0803(.022)	.0028(.020)[.020]
σ_v^2	-.2466(.106)	-.0070(.041)[.038]	-.1065(.023)	-.0033(.025)[.025]
β	-.0003(.014)	-.0002(.013)[.012]	.0000(.008)	.0000(.008)[.008]
λ	-.0070(.036)	.0002(.031)[.032]	-.0067(.025)	-.0007(.025)[.024]
ρ	.0469(.061)	-.0030(.045)[.045]	.0204(.031)	-.0002(.028)[.028]
τ	-.5007(.438)	.0052(.042)[.045]	-.0839(.026)	-.0002(.022)[.022]
σ_v^2	-.2566(.122)	-.0066(.078)[.080]	-.1087(.050)	-.0056(.056)[.055]
β	-.0001(.013)	-.0001(.012)[.012]	.0006(.008)	.0007(.008)[.008]
λ	-.0081(.038)	-.0014(.034)[.032]	-.0079(.023)	-.0016(.023)[.024]
ρ	.0505(.063)	.0005(.045)[.046]	.0172(.030)	-.0036(.026)[.028]
τ	-.4567(.395)	.0063(.041)[.040]	-.0807(.023)	.0029(.020)[.021]
σ_v^2	-.2429(.107)	-.0043(.059)[.059]	-.1051(.038)	-.0018(.042)[.041]

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

Table 3: Descriptive statistics for the data.

Variables	Obs	Mean	Std	Min	Max
Beer Tax Rates	911	0.193	0.152	0.017	0.768
Gasoline Tax Rates	888	0.137	0.052	0.040	0.380
Size	912	0.647	0.907	0.017	4.279
DR	912	0.540	0.055	0.430	0.720
GIO	912	0.523	0.108	0.213	0.728
LSTR	912	0.042	0.017	0.000	0.080
GSP	912	0.098	0.123	0.004	0.964
PE	912	9.391	12.756	0.448	109.000

Note. Tax rates and PE are from World Tax Database (`https://www.bus.umich.edu/otpr/otpr/default.asp`); GSP from US Bureau of Economic Analysis (`https://www.bea.gov/data/gdp/gdp-state`); other control variables from Egger et al. (2005); and the missing values on PE are recovered from United States Census Bureau (`https://www.census.gov/programs-surveys/state/data/historical_data.html`). Little's test of missing completely at random (Little, 1988) has a p-value of 0.9886, and thus is not rejected.

Table 4: Estimation results for beer tax rates using various methods.

Variables	QMLE-GU	ME-GU	ME-IMR	QMLE-MR	ME-MR	ME-IMRSC	QMLE-MRSC	ME-MRSC
Size	0.158***	0.160***	0.156***	0.158***	0.159***	0.148***	0.147***	0.147***
	(4.39)	(4.45)	(2.72)	(4.27)	(4.30)	(3.39)	(3.42)	(3.42)
WSize	−0.107	−0.113*	−0.087	−0.114*	−0.124**	−0.152**	−0.142**	−0.155***
	(−1.64)	(−1.73)	(−0.42)	(−1.91)	(−2.09)	(−2.25)	(−2.18)	(−2.37)
DR	0.193*	0.194*	0.232	0.185**	0.175**	0.125	0.133*	0.116
	(1.89)	(1.90)	(1.49)	(2.10)	(1.98)	(1.44)	(1.66)	(1.46)
GIO	−0.035**	−0.034**	−0.031	−0.036***	−0.036***	−0.009	−0.009	−0.008
	(−2.29)	(−2.24)	(−0.93)	(−2.44)	(−2.42)	(−0.64)	(−0.66)	(−0.60)
LSTR	0.268	0.273	0.282	0.268	0.270	−0.061	−0.075	−0.085
	(1.15)	(1.17)	(1.22)	(1.15)	(1.16)	(−0.25)	(−0.33)	(−0.37)
GSP	−0.786***	−0.783***	−0.762***	−0.775***	−0.758***	−0.603***	−0.613***	−0.586***
	(−8.95)	(−8.91)	(−2.79)	(−7.86)	(−7.69)	(−5.54)	(−6.12)	(−5.85)
PE	0.007***	0.007***	0.007***	0.007***	0.007***	0.005***	0.006***	0.005***
	(10.00)	(9.84)	(3.28)	(8.50)	(8.14)	(6.15)	(7.00)	(6.48)
SL(λ)	0.168	0.197	0.025	0.234	0.316*	0.346*	0.244	0.370**
	(0.71)	(0.83)	(0.02)	(1.28)	(1.74)	(1.68)	(1.39)	(2.10)
SE(ρ)	−0.036	−0.026	0.150	−0.111	−0.165	−0.222	−0.139	−0.245
	(−0.13)	(−0.09)	(0.11)	(−0.49)	(−0.72)	(−0.89)	(−0.63)	(−1.12)
SC(τ)						0.663***	0.688***	0.699***
						(8.89)	(38.44)	(38.96)
Pseudo R^2	96.55%	96.56%	96.51%	96.59%	96.64%	98.09%	98.21%	98.26%
States	48	48	48	48	48	48	48	48
Years	19	19	19	19	19	19	19	19
N	911	911	911	911	911	911	911	911

Significance levels: *:10%, **:5%, and ***: 1%; t-statistic values in parentheses.

Table 5: Estimation results for gasoline tax rates using various methods.

Variables	QMLE-GU	ME-GU	ME-IMR	QMLE-MR	ME-MR	ME-IMRSC	QMLE-MRSC	ME-MRSC
Size	0.041	0.038	0.033	0.051*	0.053*	0.029	0.046	0.048
	(1.40)	(1.31)	(1.01)	(1.66)	(1.73)	(0.74)	(1.37)	(1.41)
WSize	−0.134***	−0.129***	−0.131***	−0.127***	−0.119***	−0.133**	−0.120**	−0.114**
	(−2.65)	(−2.55)	(−2.48)	(−2.70)	(−2.52)	(−2.08)	(−2.28)	(−2.15)
DR	0.012	0.003	0.013	0.016	0.014	0.014	0.036	0.034
	(0.16)	(0.04)	(0.17)	(0.25)	(0.21)	(0.16)	(0.54)	(0.50)
GIO	0.014	0.014	0.006	0.014	0.014	0.002	0.016	0.016
	(1.21)	(1.16)	(0.48)	(1.22)	(1.19)	(0.13)	(1.38)	(1.38)
LSTR	−0.031	−0.044	0.194	0.005	0.007	0.178	−0.085	−0.086
	(−0.17)	(−0.24)	(0.93)	(0.03)	(0.04)	(0.76)	(−0.46)	(−0.47)
GSP	−0.136	−0.107	−0.195**	−0.179*	−0.173*	−0.206**	−0.153	−0.146
	(−1.45)	(−1.15)	(−2.10)	(−1.79)	(−1.72)	(−2.06)	(−1.53)	(−1.47)
PE	0.000	0.000	0.001	0.001	0.000	0.001	0.000	0.000
	(0.24)	(−0.07)	(0.98)	(0.65)	(0.58)	(1.03)	(0.44)	(0.37)
SL(λ)	0.100	0.081	0.251***	0.267**	0.329***	0.186***	0.208	0.262**
	(0.94)	(0.76)	(3.67)	(2.28)	(2.81)	(3.86)	(1.63)	(2.05)
SE(ρ)	0.197	0.270**	0.018	0.026	0.010	0.037	0.045	0.038
	(1.55)	(2.12)	(0.18)	(0.16)	(0.06)	(0.56)	(0.28)	(0.23)
SC(τ)						0.415***	0.682***	0.691***
						(18.85)	(41.53)	(42.11)
Pseudo R^2	82.19%	82.31%	77.62%	82.37%	82.52%	81.79%	90.52%	90.59%
States	48	48	48	48	48	48	48	48
Years	19	19	19	19	19	19	19	19
N	888	888	888	888	888	888	888	888

Significance levels: *:10%, **:5%, and ***; t-statistic values in parentheses.