Genuinely Unbalanced Spatial Panel Data Models with Fixed Effects: M-Estimation and Inference with an Application to FDI Inflows*

Xiaoyu Meng

School of Economics, Nankai University

Zhenlin Yang†

School of Economics, Singapore Management University

September 9, 2024

Abstract

We consider spatial panel data models with *genuine* unbalancedness arising from the non-presence of some spatial units in certain time periods. General M-estimation methods are proposed for model estimation, which take into account the estimation of the *incidental* fixed effects parameters, and allow for spatiotemporal heteroskedasticity and high-order time-varying spatial effects. Corrected plug-in methods are proposed for standard error estimation. The proposed estimation and inference methods are rigorously studied for their asymptotic properties and finite sample performance, and are empirically illustrated using China FDI inflow data.

Keywords: Adjusted quasi score; Fixed effects; Genuine unbalancedness; High-order spatial effects; Time-varying spatial weights; Spatiotemporal heteroskedasticity.

1. Introduction

The literature on spatial panel data (SPD) models has been fast-growing since Anselin (1988), due to the fact that the SPD models are able to take into account the spatial interaction effects and control for the unobservable heterogeneity. Most of the works on SPD models are based on balanced panels, i.e., a set of observations collected on n spatial units over the entire T periods in time (e.g., Baltagi et al., 2003; Lee and Yu, 2010; Baltagi and Yang, 2013a,b; Yang et al., 2016; Liu and Yang, 2020; Lu, 2023).

The literature on "unbalanced" SPD models is limited to only a few empirical works (Egger et al., 2005; Baltagi et al., 2007; and Baltagi et al., 2015), and a few theoretical works (Wang

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

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

and Lee, 2013b; Meng and Yang, 2021; Zhou et al., 2022; and Yang et al., 2023). This is in stark contrast to the sizable literature on usual unbalanced panels (e.g., Wansbeek and Kapteyn, 1989; Baltagi and Chang, 1994; Davis, 2002; Baltagi et al., 2001; Antweiler, 2001; Baltagi and Song, 2006; Wooldridge, 2019), textbook treatments (Baltagi, 2021; Hsiao, 2022; Greene, 2018), and software implementations (`STATA`, `SAS`, and `R`).

Unbalanced panels are more likely to be the norm in typical economic empirical settings (Baltagi and Song, 2006), so are the unbalanced spatial panels. Unbalancedness in regular panels is broadly viewed as due to either randomly or nonrandomly missing units/observations (Baltagi, 2021, Chap. 9), from a sampling perspective. In the case of the former, analyses are often done simply based on the available data as they still 'represent' the underlying population. However, this is not the case for unbalanced spatial panels.

A spatial autoregressive (`SAR`) model is a general equilibrium model and the number of units can be regarded as a 'population' (Lee, 2004; Wang and Lee, 2013a). In spatial panel context, the number of spatial units may change from one period to another due to the *non-presence* of some spatial units from time to time, giving rise to what we call in this paper the *genuinely unbalanced* (GU) SPD, to stress the fact that spatial units present at each period form a well defined `SAR` process with a complete connectivity structure. In other words, all spatial units have full observations of themselves and their neighbors in all periods, according to Kelejian and Prucha (2010). Examples of this include early dropouts and new entrants, mergers and acquisitions, splits, etc. Unbalanced spatial panels may be *not genuinely unbalanced* as they contain units with observations on their neighbors missing or units with their own observations missing (Kelejian and Prucha, 2010). Deleting these units ignores the impacts from or to the neighbors, rendering the subsequent analysis invalid.

Many important issues related to unbalanced spatial panels that are of wide practical interest have not been resolved. This paper focuses on (high-order) GU-SPD models. With unbalanced panels, the usual fixed effects estimation methods are no longer applicable. We introduce a general M-estimation method that allows for (i) unobserved spatiotemporal heterogeneity (the fixed effects), (ii) unknown spatiotemporal heteroskedasticity, (iii) high-order and time-varying spatial effects in responses, regressors and errors, and (iv) alternative forms of spatial specifications (see Section 2 for details). We propose a *corrected plug-in* method for standard error estimation. The proposed methods possess excellent finite sample property as seen through extensive Monte Carlo simulation. Their usefulness and empirical relevance are clearly demonstrated using China FDI inflow data.

Section 2 discusses model specifications. Section 3 presents results for first-order models under homoskedasticity. Section 4 extends the study to allow for spatiotemporal heteroskedasticity. Section 5 presents results for higher-order models. Section 6 presents partial Monte Carlo results. Section 7 presents an empirical application. Section 8 concludes. To save space, proofs and full MC results are relegated to the `Supplementary Material`.

2. Model Specification

Consider a study that lasts T periods and involves a total of n spatial units. At time t, only n_t of these n spatial units are present, which have full observations on themselves (responses and covariates) and on their neighbors, resulting in a set of GU-SPD, which is modeled with a `SAR` process on responses and a `SAR` process on errors, or `SARAR`:

$$\begin{aligned} Y_t &= \lambda_0 W_t Y_t + X_t\beta_0 + D_t Z\gamma_0 + D_t\mu_0 + \alpha_{t0} l_{n_t} + U_t, \\ U_t &= \rho_0 M_t U_t + V_t, \quad t = 1, \ldots, T,\ i = 1, \ldots, n_t, \end{aligned} \tag{2.1}$$

where Y_t is a vector of observations on n_t spatial units at time t, X_t is an $n_t \times k$ matrix containing values of k time-varying exogenous regressors, Z is an $n \times p$ matrix containing values of time-invariant regressors, and $U_t = (u_{1t}, u_{2t}, \ldots, u_{n_t t})'$ and $V_t = (v_{1t}, v_{2t}, \ldots, v_{n_t t})'$ are $n_t \times 1$ vectors of disturbance and idiosyncratic errors, respectively. W_t and M_t are given $n_t \times n_t$ spatial weight matrices, which together with the spatial coefficients λ_0 and ρ_0, characterize the spatial lag (`SL`) effects and spatial error (`SE`) effects, respectively. [1] β_0 and γ_0 are $k \times 1$ and $p \times 1$ vectors of regression coefficients, $\mu_0 = \{\mu_{i0}\}_{i=1}^n$ an $n \times 1$ vector of unit-specific effects, and $\alpha_0 = \{\alpha_{t0}\}_{t=1}^T$ a $T \times 1$ vector of time-specific effects. D_t is an $n_t \times n$ "selection" matrix obtained from the $n \times n$ identity matrix I_n by deleting its rows corresponding to the non-present units at time t, and l_{n_t} is an $n_t \times 1$ vector of ones.

When both μ_0 and α_0 are correlated with the time-varying regressors in an arbitrary manner, we have a fixed effects (FE) model; when they are uncorrelated with the regressors, we have a random effects (RE) model; and when they are correlated linearly with the regressors, we have a correlated random effects (CRE) model. The idiosyncratic errors $\{v_{it}\}$ can be iid (independent and identically distributed) across i and over t (homoskedasticity), or inid (independent but not identically distributed) along both i and t (heteroskedasticity).

The modeling strategy (2.1) allows full control of unobserved heterogeneity in all n units over entire T periods. Yang et al. (2023), following an early version of our paper (Meng and Yang, 2021), studied `MESS` version of Model (2.1), where `MESS` stands for matrix exponential

[1] Spatial Durbin (`SD`) terms or *contextual effects*, $\mathcal{W}_t X_t^*$, can be added as additional regressors, where X_t^* is a submatrix of X_t, and $\mathcal{W}_t$ is W_t or M_t or neither; see Anselin et al. (2008) and Lee and Yu (2016) for potential identification issues. All spatial weight matrices are time-varying due to the changes in the number of available units or the fundamental changes in connectivity, and are not necessarily row-normalized.

spatial specification (see below for details)[2].

Recently, theory and applications have advanced to SPD models with higher-order spatial effects for capturing various types of spatial interaction effects (see Kuersteiner and Prucha, 2020; Drukker et al., 2022). An SPD-GU model with a p-order SAR process on the responses and a q-order SAR process on the disturbances, or $\texttt{SARAR}(p,q)$, takes the form:

$$
\begin{aligned}
Y_t &= \textstyle\sum_{k=1}^{p} \lambda_{k0} W_{kt} Y_t + X_t\beta_0 + D_t Z\gamma_0 + D_t\mu_0 + \alpha_{t0} l_{n_t} + U_t, \\
U_t &= \textstyle\sum_{\ell=1}^{q} \rho_{\ell 0} M_{\ell t} U_t + V_t, \quad t = 1, \ldots, T,\ i = 1, \ldots, n_t.
\end{aligned} \tag{2.2}
$$

Our model can accommodate alternative spatial specifications. By replacing $I_{n_t} - \sum_{k=1}^{p}\lambda_k W_{kt}$ with $\exp(\sum_{k=1}^{p}\lambda_k W_{kt})$ and $I_{n_t} - \sum_{\ell=1}^{q}\rho_\ell M_{\ell t}$ with $\exp(\sum_{\ell=1}^{q}\rho_\ell M_{\ell t})$, we have a $\texttt{MESS}(p,q)$ form of the model, where $\exp(A) = \sum_{i=0}^{\infty} A^i/i!$ for a square matrix A and $A^0 = I_{\dim(A)}$.

$\texttt{MESS}(p,q)$ seems computationally simpler as $|\exp(\mathcal{W}_t)| = \exp(\texttt{tr}(\mathcal{W}_t)) = 1$, for $\mathcal{W}_t = \sum_{k=1}^{p}\lambda_k W_{kt}$, or $\sum_{\ell=1}^{q}\rho_\ell M_{\ell t}$. However, this is true only for first-order models, as the partial derivatives of $\exp(\mathcal{W}_t)$, required in the estimation and inference, do not have closed-form expressions unless $p = q = 1$. This issue does not apply to $I_{n_t} - \mathcal{W}_t$. Furthermore, the SAR-type specification is more widely adopted for modeling social interactions and network effects (Lee, 2007; Han et al., 2021), where the spatial lag and spatial Durbin effects correspond directly to what Manski (1993) called the endogenous social effects and contextual effects. Our paper provides a general framework for the estimation and inference applicable to the general $\texttt{SARAR}(p,q)$ specification. Section 8 discusses a possible way to tackle the partial derivative issue so as to accommodate the $\texttt{MESS}(p,q)$ specification.

Notations and conventions. First, $|\cdot|$, $\texttt{tr}(\cdot)$, $'$, and $\|\cdot\|$ are the common notations for determinant, trace, transpose, and matrix norm. For a real $n \times m$ matrix A, $\|A\|_1$ denotes the maximum column sum norm, $\|A\|_\infty$ maximum row sum norm, and $A^\circ = A + A'$. For a real $n \times m$ matrix A with a full column rank, $\mathbb{P}_A = A(A'A)^{-1}A'$ and $\mathbb{Q}_A = I_n - \mathbb{P}_A$ are the two orthogonal projection matrices. $\texttt{diag}(\cdot)$ forms a diagonal matrix by the diagonal elements of a square matrix or the elements of a given vector, $\texttt{diagv}(\cdot)$ forms a column vector by the diagonal elements of a square matrix, and $\texttt{blkdiag}(\cdots)$ forms a block-diagonal matrix. $\mathrm{E}(\cdot)$ and $\mathrm{Var}(\cdot)$ correspond to true parameter values.

3. M-Estimation under Homoskedasticity

To fix ideas, we first give a full treatment of the first-order model (2.1) under FE and GU specification, assuming that the errors $\{v_{it}\}$ are $\mathrm{iid}(0, \sigma_{v0}^2)$ across i and t. As our proposed

[2] A closely related model but with *missing on responses only* has been studied by Wang and Lee (2013b).

method starts with the joint quasi scores of both common and FE parameters, the joint quasi maximum likelihood (QML) estimation is discussed first.

3.1. Direct QML estimation with fixed effects

Let $\mathbf{Y} = (Y_1', \ldots, Y_T')'$, $\mathbf{X} = (X_1', \ldots, X_T')'$, $\mathbf{U} = (U_1', \ldots, U_T')'$, and $\mathbf{V} = (V_1', \ldots, V_T')'$. Define $\mathbf{W} = \texttt{blkdiag}(W_1, \ldots, W_T)$, $\mathbf{M} = \texttt{blkdiag}(M_1, \ldots, M_T)$, $\mathbf{D}_\mu = (D_1', \ldots, D_T')'$, and $\mathbf{D}_\alpha = \texttt{blkdiag}(l_{n_1}, \ldots, l_{n_T})$. Without the time-invariant regressors, model (2.1) is written in matrix form: $\mathbf{Y} = \lambda_0 \mathbf{WY} + \mathbf{X}\beta_0 + \mathbf{D}_\mu \mu_0 + \mathbf{D}_\alpha \alpha_0 + \mathbf{U}$ and $\mathbf{U} = \rho_0 \mathbf{MU} + \mathbf{V}$. Note that there are $n+T$ fixed effects parameters but only $n+T-1$ of them are identifiable. Therefore, a zero-sum constraint is put on the $\alpha_t's$, and the QML estimation of the common and FE parameters is based on the following model form:

$$\mathbf{Y} = \lambda_0 \mathbf{WY} + \mathbf{X}\beta_0 + \mathbf{D}_\mu \mu_0 + \mathbf{D}_\alpha^\star \alpha_0^\star + \mathbf{U}, \qquad \mathbf{U} = \rho_0 \mathbf{MU} + \mathbf{V}. \tag{3.1}$$

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

Denote the set of *common parameters* by $\theta = (\beta', \sigma_v^2, \delta)'$, where $\delta = (\lambda, \rho)'$, and the set of FE or *incidental parameters* by $\phi = (\mu', \alpha^{\star\prime})'$. Denote $\mathbf{A}_N(\lambda) = I_N - \lambda \mathbf{W}$, $\mathbf{B}_N(\rho) = I_N - \rho \mathbf{M}$, and $\mathbf{D} = [\mathbf{D}_\mu,\ \mathbf{D}_\alpha^\star]$, where $N = \sum_{t=1}^T n_t$ and I_N is the $N \times N$ identity matrix. We have the quasi Gaussian loglikelihood function:

$$\ell_N(\theta, \phi) = -\tfrac{N}{2} \ln 2\pi - \tfrac{N}{2} \ln \sigma_v^2 + \ln |\mathbf{A}_N(\lambda)| + \ln |\mathbf{B}_N(\rho)| - \tfrac{1}{2\sigma_v^2} \mathbf{V}'(\beta, \delta, \phi) \mathbf{V}(\beta, \delta, \phi), \tag{3.2}$$

where $\mathbf{V}(\beta, \delta, \phi) = \mathbf{B}_N(\rho)[\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\beta - \mathbf{D}\phi]$. $\ell_N(\theta, \phi)$ is partially maximized at

$$\hat{\phi}_N(\beta, \delta) = [\mathbb{D}'(\rho)\mathbb{D}(\rho)]^{-1} \mathbb{D}'(\rho) \mathbf{B}_N(\rho)[\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\beta], \tag{3.3}$$

where $\mathbb{D}(\rho) = \mathbf{B}_N(\rho)\mathbf{D}$. This leads to the concentrated quasi loglikelihood function of θ:

$$\ell_N^c(\theta) = -\tfrac{N}{2} \ln 2\pi - \tfrac{N}{2} \ln \sigma_v^2 + \ln |\mathbf{A}_N(\lambda)| + \ln |\mathbf{B}_N(\rho)| - \tfrac{1}{2\sigma_v^2} \tilde{\mathbf{V}}'(\beta, \delta) \tilde{\mathbf{V}}(\beta, \delta), \tag{3.4}$$

where $\tilde{\mathbf{V}}(\beta, \delta) = \mathbb{Q}_{\mathbb{D}}(\rho) \mathbf{B}_N(\rho)[\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\beta]$ and $\mathbb{Q}_{\mathbb{D}}(\rho)$ is the projection matrix based on $\mathbb{D}(\rho)$. The direct QML estimator (QMLE) $\hat{\theta}_{\texttt{QML}}$ of θ maximizes $\ell_N^c(\theta)$. However, such a direct estimation of the common parameters θ ignores the impact of estimating the fixed effects parameters ϕ. As a result, $\hat{\theta}_{\texttt{QML}}$ may be inconsistent or asymptotically biased, giving rise to the well-known *incidental parameters problem* of Neyman and Scott (1948). The transformation method of Lee and Yu (2010) works only for a balanced FE-SPD model with time-invariant and row-normalized spatial weight matrices. An alternative approach is to carry out a bias correction directly on $\hat{\theta}_{\texttt{QML}}$, which can be quite complicated and the resulting inference method can be valid only when T is also large.

3.2. M-estimation with fixed effects

We approach this problem by adjusting the concentrated quasi scores to remove the impact of estimating the incidental parameters, following Yang (2018b). The concentrated quasi score (CQS) vector, $S_N^c(\theta) = \partial\ell_N^c(\theta)/\partial\theta$, has the expression:

$$S_N^c(\theta) = \begin{cases} \frac{1}{\sigma_v^2}\mathbf{X}'\mathbf{B}_N'(\rho)\tilde{\mathbf{V}}(\beta,\delta), \\ \frac{1}{2\sigma_v^4}[\tilde{\mathbf{V}}'(\beta,\delta)\tilde{\mathbf{V}}(\beta,\delta) - N\sigma_v^2], \\ \frac{1}{\sigma_v^2}\mathbf{Y}'\mathbf{W}'\mathbf{B}_N'(\rho)\tilde{\mathbf{V}}(\beta,\delta) - \mathtt{tr}[\mathbf{F}_N(\lambda)], \\ \frac{1}{\sigma_v^2}\tilde{\mathbf{V}}'(\beta,\delta)\mathbf{G}_N(\rho)\tilde{\mathbf{V}}(\beta,\delta) - \mathtt{tr}[\mathbf{G}_N(\rho)], \end{cases} \tag{3.5}$$

where $\mathbf{F}_N(\lambda) = \mathbf{W}\mathbf{A}_N^{-1}(\lambda)$ and $\mathbf{G}_N(\rho) = \mathbf{M}\mathbf{B}_N^{-1}(\rho)$. See Appendix A for its derivation. For regular optimization problems, maximizing $\ell_N^c(\theta)$ is equivalent to solving $S_N^c(\theta) = 0$. Then, for the resulting root $\hat{\theta}_{\mathtt{QML}}$ to be consistent it is necessary that plim $_{N\to\infty}N^{-1}\partial\ell_N^c(\theta_0)/\partial\theta = 0$, where θ_0 denotes the true parameter vector. However,

$$\mathrm{E}[S_N^c(\theta_0)] = \begin{cases} 0_k, \\ -(n+T-1)/(2\sigma_{v0}^2), \\ \mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho_0)\mathbf{B}_N(\rho_0)\mathbf{F}_N(\lambda_0)\mathbf{B}_N^{-1}(\rho_0)] - \mathtt{tr}[\mathbf{F}_N(\lambda_0)], \\ \mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho_0)\mathbf{G}_N(\rho_0)] - \mathtt{tr}[\mathbf{G}_N(\rho_0)], \end{cases} \tag{3.6}$$

from which one sees that that $\lim_{N\to\infty}\mathrm{E}[S_N^c(\theta_0)]/N \neq 0$ when T is fixed. This suggests that $\mathrm{plim}_{N\to\infty}S_N^c(\theta_0)/N \neq 0$, and therefore $\hat{\theta}_{\mathtt{QML}}$ cannot be consistent. When T goes large with n, consistency can be achieved but the limiting distribution of $\sqrt{N}(\hat{\theta}_{\mathtt{QML}} - \theta_0)$ is not centered, suggesting that $\hat{\theta}_N$ is asymptotically biased.

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

$$S_N^*(\theta) = \begin{cases} \frac{1}{\sigma_v^2}\mathbf{X}'\mathbf{B}_N'(\rho)\tilde{\mathbf{V}}(\beta,\delta), \\ \frac{1}{2\sigma_v^4}\left[\tilde{\mathbf{V}}'(\beta,\delta)\tilde{\mathbf{V}}(\beta,\delta) - (N-n-T+1)\sigma_v^2\right], \\ \frac{1}{\sigma_v^2}\mathbf{Y}'\mathbf{W}'\mathbf{B}_N'(\rho)\tilde{\mathbf{V}}(\beta,\delta) - \mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho)\mathbf{F}_N(\lambda)\mathbf{B}_N^{-1}(\rho)], \\ \frac{1}{\sigma_v^2}\tilde{\mathbf{V}}'(\beta,\delta)\mathbf{G}_N(\rho)\tilde{\mathbf{V}}(\beta,\delta) - \mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{G}_N(\rho)]. \end{cases} \tag{3.7}$$

Solving the AQS equations: $S_N^*(\theta) = 0$, gives the M-estimator of θ, i.e.,

$$\hat{\theta}_N^* = \arg\{S_N^*(\theta) = 0\}.$$

It is easy to verify that $\mathrm{E}[S_N^*(\theta_0)] = 0$ and plim $S_N^*(\theta_0)/N = 0$, making it possible for $\hat{\theta}_N^*$ to be $\sqrt{N_1}$-consistent with proper limiting distribution, where $N_1 = N - n - T + 1$, the *effective* sample size after taking into account the estimation of fixed effects.

The proposed approach applies to general T and general spatial weight matrices. It offers a feasible way to fully control the unobserved heterogeneity in all units and periods involved in the study even if the SPD is unbalanced. It can be applied to all the models discussed in Section 2 except $\texttt{MESS}(p,q)$ which has a computation issue. For a balanced spatial panel, it offers a more general method than Lee and Yu (2010). [3]

To simplify the root-finding process, given δ we solve $S_N^*(\theta) = 0$ at:

$$\hat{\beta}_N^*(\delta) = [\mathbb{X}'(\rho)\mathbb{X}(\rho)]^{-1}\mathbb{X}'(\rho)\mathbf{B}_N(\rho)\mathbf{A}_N(\lambda)\mathbf{Y} \quad \text{and} \quad \hat{\sigma}_{v,N}^{*2}(\delta) = \tfrac{1}{N_1}\hat{\mathbf{V}}'(\delta)\hat{\mathbf{V}}(\delta), \tag{3.8}$$

where $\mathbb{X}(\rho) = \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho)\mathbf{X}$ and $\hat{\mathbf{V}}(\delta) = \tilde{\mathbf{V}}(\hat{\beta}_N^*(\delta), \delta)$. Substituting $\hat{\beta}_N^*(\delta)$ and $\hat{\sigma}_{v,N}^{*2}(\delta)$ back into the third and fourth components of (3.7) gives the concentrated AQS functions of δ:

$$S_N^{*c}(\delta) = \begin{cases} \frac{1}{\hat{\sigma}_{v,N}^{*2}(\delta)}\mathbf{Y}'\mathbf{W}'\mathbf{B}_N'(\rho)\hat{\mathbf{V}}(\delta) - \mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho)\mathbf{F}_N(\lambda)\mathbf{B}_N^{-1}(\rho)], \\ \frac{1}{\hat{\sigma}_{v,N}^{*2}(\delta)}\hat{\mathbf{V}}'(\delta)\mathbf{G}_N(\rho)\hat{\mathbf{V}}(\delta) - \mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{G}_N(\rho)]. \end{cases} \tag{3.9}$$

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

3.3. Asymptotic properties of the M-estimator

We now study the asymptotic properties of the proposed M-estimator to provide a theoretical base for empirical applications. First, for $\hat{\theta}_N^*$ to be consistent, it is necessary that some basic conditions hold for the errors, regressors, and spatial weight matrices. Let Δ be the parameter space for δ, and Δ_λ and Δ_ρ be the sub-spaces for λ and ρ.

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

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

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

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 $h_n/n \to 0$, as $n \to \infty$; (ii) diagonal

[3] The AQS method of **?** is seen to be rather versatile in dealing with the *incidental parameters problem*, a problem raised by Neyman and Scott (1948) and solutions sought thereafter. See also Baltagi and Yang (2013a,b), Liu and Yang (2020), Li and Yang (2020, 2021), and Xu and Yang (2020).

elements are zero; and (iii) column and row sum norms are bounded.

Further, the two key matrices $\mathbf{A}_N(\lambda)$ and $\mathbf{B}_N(\rho)$, denoted by $\mathbb{A}(\varpi)$ with $\varpi = \lambda$ or ρ, need to be invertible with their inverses satisfy certain boundedness conditions.

Assumption E. (i) both $\|\mathbb{A}^{-1}(\varpi_0)\|_\infty$ and $\|\mathbb{A}^{-1}(\varpi_0)\|_1$ are bounded;
(ii) either $\|\mathbb{A}^{-1}(\varpi)\|_\infty$ or $\|\mathbb{A}^{-1}(\varpi)\|_1$ is bounded, uniformly in $\varpi \in \Delta_\varpi$;
(iii) $0 < \underline{c}_\varpi \leq \inf_{\varpi\in\Delta_\varpi} \gamma_{\min}[\mathbb{A}'(\varpi)\mathbb{A}(\varpi)] \leq \sup_{\varpi\in\Delta_\varpi} \gamma_{\max}[\mathbb{A}'(\varpi)\mathbb{A}(\varpi)] \leq \bar{c}_\varpi < \infty$.

As we are dealing with unbalanced spatial panels, the asymptotics depends on the growth of $N(= \sum_{t=1}^T n_t = \sum_{i=1}^n T_i)$ where unit i appears T_i times in T periods. While no restrictions are imposed on the relative magnitude of n and T, we require that n_t/n does not shrink when n increases and T_i/T does not shrink when T increases.

Assumption F. (i) As $n \to \infty$, $n_t/n \to c_t$, where $c_t \in (0, 1], \forall t$; (ii) as $T \to \infty$, $T_i/T \to d_i$, where $d_i \in (0, 1], \forall i$, and T_i is the number of times the ith unit shows up in the entire T periods; and (iii) $\min_i(T_i) \geq 2$ and $\min_t(n_t) \geq 2$.

Assumption F(iii) ensures the spatial structure is complete after μ and α are concentrated out and all parameters are identified. Clearly, the scenario under Assumption F(i) with $\min_i(T_i) \geq 2$ is of greater interest in spatial econometrics as it is when n is large that one needs to impose structures on the spatial connectivity matrices.

Like in GMM estimation, identification uniqueness is an important but difficult issue, and often a high-level assumption is given. For our M-estimation, let $\bar{S}_N^{*c}(\delta)$ be the population counterpart of $S_N^{*c}(\delta)$ obtained by concentrating $\bar{S}_N^*(\theta) = \mathrm{E}[S_N^*(\theta)]$:

$$\bar{S}_N^{*c}(\delta) = \begin{cases} \frac{1}{\bar{\sigma}^{*2}_{v,N}(\delta)} \mathrm{E}[\mathbf{Y}'\mathbf{W}'\mathbf{B}_N'(\rho)\bar{\mathbf{V}}(\delta)] - \mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho)\mathbf{F}_N(\lambda)\mathbf{B}_N^{-1}(\rho)], \\ \frac{1}{\bar{\sigma}^{*2}_{v,N}(\delta)} \mathrm{E}[\bar{\mathbf{V}}'(\delta)\mathbf{G}_N(\rho)\bar{\mathbf{V}}(\delta)] - \mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{G}_N(\rho)]. \end{cases} \tag{3.10}$$

where $\bar{\sigma}^{*2}_{v,N}(\delta) = \mathrm{E}[\bar{\mathbf{V}}'(\delta)\bar{\mathbf{V}}(\delta)]/N_1$, $\bar{\mathbf{V}}(\delta) = \tilde{\mathbf{V}}(\bar{\beta}_N^*(\delta), \delta) = \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho)[\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\bar{\beta}_N^*(\delta)]$, and $\bar{\beta}_N^*(\delta) = [\mathbb{X}'(\rho)\mathbb{X}(\rho)]^{-1}\mathbb{X}'(\rho)\mathbf{B}_N(\rho)\mathbf{A}_N(\lambda)\mathrm{E}(\mathbf{Y})$. Clearly, $\bar{S}_N^{*c}(\delta_0) = 0$ and $\bar{\sigma}^{*2}_{v,N}(\delta_0) = \sigma^2_{v0}$, and $S_N^{*c}(\hat{\delta}_N^*) = 0$. Thus, by Theorem 5.9 of Van der Vaart (1998), $\hat{\delta}_N^*$ is consistent if $\sup_{\delta\in\Delta} \left\|S_N^{*c}(\delta) - \bar{S}_N^{*c}(\delta)\right\| / N_1 \xrightarrow{p} 0$, and the following identification condition holds.

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

Assumption G is a high-level assumption being put up for simplicity of presentation. It can be shown to be true under low-level conditions (see Appendix C in `Supplementary Material`). Finally, a minor technical assumption is needed to ensure the uniform boundedness of $\|\mathbb{Q}_{\mathbb{D}}(\rho)\|_1$ and $\|\mathbb{Q}_{\mathbb{D}}(\rho)\|_\infty$. Let $B_t(\rho)$ be the tth diagonal block of $\mathbf{B}_N(\rho)$.

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

See for detail Lemma B.3 in `Supplementary Material`. Once consistency of $\hat{\delta}_N^*$ is established, consistency of $\hat{\beta}_N^*$ and $\hat{\sigma}_{v,N}^{*2}$ follows from (3.8) and Assumptions C-E, H.

Theorem 1. *Suppose Assumptions A-H hold. We have, as $N \to \infty$, $\hat{\theta}_N^* \xrightarrow{p} \theta_0$.*

To derive the asymptotic distribution of $\hat{\theta}_N^*$, we have by the mean value theorem,

$$0 = S_N^*(\hat{\theta}_N^*) = S_N^*(\theta_0) + \tfrac{\partial}{\partial\theta'}S_N^*(\bar{\theta})(\hat{\theta}_N^* - \theta_0),$$

with a different $\bar{\theta}$ (between $\hat{\theta}_N^*$ and θ_0) for each row of $\partial S_N^*(\bar{\theta})/\partial\theta'$. The asymptotic normality of $\sqrt{N_1}(\hat{\theta}_N^* - \theta_0)$ depends on that of $S_N^*(\theta_0)/\sqrt{N_1}$ and a proper behavior of $\partial S_N^*(\bar{\theta})/\partial\theta'$. Note

$$S_N^*(\theta_0) = \begin{cases} \frac{1}{\sigma_{v0}^2}\mathbb{X}'\mathbf{V}, \\ \frac{1}{2\sigma_{v0}^4}(\mathbf{V}'\mathbb{Q}_{\mathbb{D}}\mathbf{V} - N_1\sigma_{v0}^2), \\ \frac{1}{\sigma_{v0}^2}\mathbf{V}'\mathbb{Q}_{\mathbb{D}}\mathbf{B}_N\mathbf{F}_N\eta + \frac{1}{\sigma_{v0}^2}\mathbf{V}'\mathbb{Q}_{\mathbb{D}}\bar{\mathbf{F}}_N\mathbf{V} - \mathtt{tr}(\mathbb{Q}_{\mathbb{D}}\bar{\mathbf{F}}_N), \\ \frac{1}{\sigma_{v0}^2}\mathbf{V}'\mathbb{Q}_{\mathbb{D}}\mathbf{G}_N\mathbb{Q}_{\mathbb{D}}\mathbf{V} - \mathtt{tr}(\mathbb{Q}_{\mathbb{D}}\mathbf{G}_N), \end{cases} \tag{3.11}$$

by using $\tilde{\mathbf{V}}(\beta_0, \delta_0) = \mathbb{Q}_{\mathbb{D}}\mathbf{V}$ and $\mathbf{Y} = \mathbf{A}_N^{-1}(\eta + \mathbf{B}_N^{-1}\mathbf{V})$ with $\eta = \mathbf{X}\beta_0 + \mathbf{D}\phi_0$, and the shorthand notations $\mathbf{A}_N \equiv \mathbf{A}_N(\lambda_0)$, $\mathbf{B}_N \equiv \mathbf{B}_N(\rho_0)$, $\mathbb{Q}_{\mathbb{D}} \equiv \mathbb{Q}_{\mathbb{D}}(\rho_0)$, etc., and $\bar{\mathbf{F}}_N = \mathbf{B}_N\mathbf{F}_N\mathbf{B}_N^{-1}$. As $S_N^*(\theta_0)$ is linear-quadratic (LQ) in $\mathbf{V}$, the central limit theorem (CLT) for LQ forms of Kelejian and Prucha (2001) can be applied to show that $S_N^*(\theta_0)/\sqrt{N_1}$ is asymptotically normal with **zero mean**. This leads to the following theorem, showing the importance of adjusting $S_N^c(\theta_0)$.

Theorem 2. *Under Assumptions A-H, we have, as $N \to \infty$,*

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

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

3.4. Inference based on M-estimation

The "Hessian" matrix, $\partial S_N^*(\theta)/\partial\theta'$, is given in Appendix A from which $\Sigma_N^*(\theta_0)$ can easily be found. The analytical expression of $\Gamma_N^*(\theta_0)$ is also given there for ease of presentation. To conduct inferences for θ, consistent estimates of $\Sigma_N^*(\theta_0)$ and $\Gamma_N^*(\theta_0)$ are required. As $\Sigma_N^*(\theta)$ and $\partial S_N^*(\theta)/\partial\theta'$ depend only on the common parameters θ, either the plug-in estimator $\Sigma_N^*(\hat{\theta}_N^*)$ or the sample analogue, $\widehat{\Sigma}_N^* = -N_1^{-1}\partial S_N^*(\theta)/\partial\theta'|_{\theta=\hat{\theta}_N^*}$, can be used to estimate $\Sigma_N^*(\theta_0)$. Consistency of $\Sigma_N^*(\hat{\theta}_N^*)$ and $\widehat{\Sigma}_N^*$ is proved in the proof of Theorem 2.

However, the estimation of $\Gamma_N^*(\theta_0)$ is more complicated as it involves not only the common parameters θ, but also the fixed effects ϕ embedded in η, and the skewness κ_3 and excess kurtosis κ_4 of the idiosyncratic errors. Thus, the common plug-in approach may not provide a valid estimate as the estimator of ϕ may not be consistent.

Let $\Gamma_N^*(\hat\theta_N^*) = \Gamma_N^*(\theta)|_{(\theta=\hat\theta_N^*, \phi=\hat\phi_N^*, \kappa_3=\hat\kappa_{3,N}, \kappa_4=\hat\kappa_{4,N})}$ be the plug-in estimator, where $\hat\phi_N^*$ is the M-estimator of ϕ obtained through (3.3), i.e., $\hat\phi_N^* = \hat\phi_N(\hat\beta_N^*, \hat\delta_N^*)$, and $\hat\kappa_{3,N}$ and $\hat\kappa_{4,N}$ are the consistent estimators of κ_3 and κ_4 to be given later. When both n and T are large, $\Gamma_N^*(\hat\theta_N^*)$ would be consistent as $\hat\phi_N^*$ is. However, when either n or T is fixed, then $\hat\alpha_N^{**}$ or $\hat\mu_N^*$ is not consistent. Plugging $\hat\phi_N^*$ into $\Gamma_N^*(\theta)$ will induce a bias (inconsistency), and a bias correction is necessary. However, only the λ-components of $\Gamma_N^*(\theta_0)$ involve ϕ (linearly or quadratic). We show that the terms linear in ϕ can be consistently estimated by the plug-in method. Therefore, the only term that may not be consistently estimated by the plug-in method is $\eta'\mathbf{F}_N'\mathbf{B}_N'\mathbb{Q}_{\mathbb{D}}\mathbf{B}_N\mathbf{F}_N\eta/\sigma_{v0}^2$ associated with the λ-λ component of $\Gamma_N^*(\theta_0)$. A consistent estimator of $\Gamma_N^*(\theta_0)$ is thus derived:

$$\widehat\Gamma_N^* = \Gamma_N^*(\hat\theta_N^*) - \text{Bias}^*(\hat\delta_N^*), \tag{3.12}$$

referred to in this paper as the *corrected plug-in* estimator, where the matrix $\text{Bias}^*(\delta_0)$ has sole non-zero element $\mathtt{tr}(\bar{\mathbf{F}}_N'\mathbb{Q}_{\mathbb{D}}\bar{\mathbf{F}}_N\mathbb{P}_{\mathbb{D}})/N_1$ at the λ-λ entry.

It is left to provide consistent estimators for κ_3 and κ_4. As $\mathbf{V} = \mathbf{B}_N(\mathbf{A}_N\mathbf{Y} - \eta)$ is infeasible due to the incidental parameters problem, we start from $\tilde{\mathbf{V}} = \mathbb{Q}_{\mathbb{D}}\mathbf{V}$, which can be "consistently" estimated by $\hat{\mathbf{V}} = \mathbb{Q}_{\mathbb{D}}(\hat\rho_N^*)\mathbf{B}_N(\hat\rho_N^*)[\mathbf{A}_N(\hat\lambda_N^*)\mathbf{Y} - \mathbf{X}\hat\beta_N^*]$. Let q_{jk} be the (j,k)th element of $\mathbb{Q}_{\mathbb{D}}$. Denote the elements of $\mathbf{V}$ by v_j, and the elements of $\tilde{\mathbf{V}}$ by $\tilde v_j, j = 1, \ldots, N$, where j is the combined index for $i = 1, \ldots, n_t$ and $t = 1, \ldots, T$. Then, $\tilde v_j = q_{j1}v_1 + q_{j2}v_2 + \cdots + q_{jN}v_N$, and,

$$\mathrm{E}(\tilde v_j^3) = \textstyle\sum_{k=1}^N q_{jk}^3 \mathrm{E}(v_k^3) = \sigma_{v0}^3\kappa_3 \sum_{k=1}^N q_{jk}^3, \ j = 1, \ldots, N.$$

Summing $\mathrm{E}(\tilde v_j^3)$ over j gives $\kappa_3 = \big(\sum_{j=1}^N \mathrm{E}(\tilde v_j^3)\big)\big(\sigma_{v0}^3 \sum_{j=1}^N \sum_{k=1}^N q_{jk}^3\big)^{-1}$. Its sample analogue using $\hat v_j$, the jth element of $\hat{\mathbf{V}}(\hat\beta_N^*, \hat\lambda_N^*)$, gives a consistent estimator of κ_3:

$$\hat\kappa_{3,N} = \frac{\sum_{j=1}^N \hat v_j^3}{\hat\sigma_{v,N}^{*3} \sum_{j=1}^N \sum_{k=1}^N \hat q_{jk}^3}, \tag{3.13}$$

where $\hat q_{jk}$ is the (j,k)th element of $\mathbb{Q}_{\mathbb{D}}(\hat\rho_N^*)$. Similarly, to estimate κ_4, we have,

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

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

$$\hat{\kappa}_{4,N} = \frac{\sum_{j=1}^N \hat{v}_j^4 - 3\hat{\sigma}_{v,N}^{*4} \sum_{j=1}^N \sum_{k=1}^N \sum_{l=1}^N \hat{q}_{jk}^2 \hat{q}_{jl}^2}{\hat{\sigma}_{v,N}^{*4} \sum_{j=1}^N \sum_{k=1}^N \hat{q}_{jk}^4}. \tag{3.14}$$

For consistency of the proposed estimators $\widehat{\Sigma}_N^*$, $\widehat{\Gamma}_N^*$, $\hat{\kappa}_{3,N}$ and $\hat{\kappa}_{4,N}$, see Corollaries C.1 and C.2 and their proofs given in `Supplementary Material`.

4. M-Estimation under Unknown Heteroskedasticity

Cross-sectional heteroskedasticity is rather common in spatial regression models due to misspecification, peer interaction, aggregation, clustering, etc. (Anselin, 1988). The same is true for (unbalanced) SPD models. Robust methods have been introduced for SPD models, but are limited to balanced panels with cross-sectional heteroskedasticity only (Moscone and Tosetti, 2011; Baltagi and Yang, 2013b; Badinger and Egger, 2015; Liu and Yang, 2020). Time-series heteroskedasticity is also important, in particular in short panels (Bai, 2013). Therefore, Assumption A is relaxed as follows.

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

4.1. Heteroskedasticity Robust M-Estimation

Denote $\mathbf{H} = \mathtt{diag}(\sigma_1^2, \sigma_2^2, \cdots, \sigma_N^2)$, and hence $\mathrm{Var}(\mathbf{V}) = \mathbf{H}$. Under this relaxed condition, the CQS function $S_N^c(\theta)$ given in (3.5) needs to be readjusted to be robust against unknown spatiotemporal heteroskedasticity $\mathbf{H}$. As in Liu and Yang (2020), we adjust the relevant components of $S_N^c(\theta)$, so that their expectations at θ_0 are zero under $\mathbf{H}$.

First, consider the stochastic element of the λ-component of $S_N^c(\theta)$ given in (3.5). Recall $\bar{\mathbf{F}}_N(\delta) = \mathbf{B}_N(\rho)\mathbf{F}_N(\lambda)\mathbf{B}_N^{-1}(\rho)$. Denote as usual $\bar{\mathbf{F}}_N = \bar{\mathbf{F}}_N(\delta_0)$. As $\tilde{\mathbf{V}}(\beta_0, \delta_0) = \mathbb{Q}_{\mathbb{D}}\mathbf{V}$, $\mathbf{B}_N\mathbf{W}\mathbf{Y} = \bar{\mathbf{F}}_N\mathbf{B}_N\mathbf{A}_N\mathbf{Y}$, $\mathbf{B}_N\mathbf{A}_N\mathbf{Y} = \mathbf{B}_N\eta + \mathbf{V}$, and $\eta = \mathbf{X}\beta_0 + \mathbf{D}\phi_0$, we have,

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

where $\bar{\mathbb{F}}_N' = \bar{\mathbb{F}}_N'(\delta_0)$ and $\bar{\mathbb{F}}_N'(\delta) = \mathtt{diag}[\bar{\mathbf{F}}_N'(\delta)\mathbb{Q}_{\mathbb{D}}(\rho)]\mathtt{diag}[\mathbb{Q}_{\mathbb{D}}(\rho)]^{-1}$. Taking the difference between the quantities inside the second expectation and the last expectation, we obtain an AQS function for λ, which is robust against unknown heteroskedasticity:

$$\mathbf{Y}'\mathbf{A}_N'(\lambda)\mathbf{B}_N'(\rho)[\bar{\mathbf{F}}_N'(\delta) - \bar{\mathbb{F}}_N'(\delta)]\tilde{\mathbf{V}}(\beta, \delta). \tag{4.1}$$

Now, consider the stochastic element of the ρ-component of $S_N^c(\theta)$. We have,

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

where $\bar{\mathbf{G}}_N(\rho) = \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{G}_N(\rho)$ and $\bar{\mathbb{G}}_N(\rho) = \mathtt{diag}[\bar{\mathbf{G}}_N(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)]\mathtt{diag}[\mathbb{Q}_{\mathbb{D}}(\rho)]^{-1}$. Replacing the $\mathbf{V}'$ by $[\mathbf{A}_N(\lambda_0)\mathbf{Y} - \mathbf{X}\beta_0]'\mathbf{B}_N'(\rho_0)$, and taking the difference between the two quantities inside the second and last expectations, we obtain a robust AQS function for ρ:

$$[\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\beta]'\mathbf{B}_N'(\rho)[\bar{\mathbf{G}}_N(\rho) - \bar{\mathbb{G}}_N(\rho)]\tilde{\mathbf{V}}(\beta, \delta). \tag{4.2}$$

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

$$S_N^{\diamond}(\beta, \delta) = \begin{cases} \mathbb{X}'(\rho)\tilde{\mathbf{V}}(\beta, \delta), \\ \mathbf{Y}'\mathbf{A}_N'(\lambda)\mathbf{B}_N'(\rho)[\bar{\mathbf{F}}_N'(\delta) - \bar{\mathbb{F}}_N'(\delta)]\tilde{\mathbf{V}}(\beta, \delta), \\ [\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\beta]'\mathbf{B}_N'(\rho)[\bar{\mathbf{G}}_N(\rho) - \bar{\mathbb{G}}_N(\rho)]\tilde{\mathbf{V}}(\beta, \delta). \end{cases} \tag{4.3}$$

Solving $S_N^{\diamond}(\beta, \delta) = 0$ gives the robust M-estimators (RM-estimators), $\hat{\beta}_N^{\diamond}$ and $\hat{\delta}_N^{\diamond}$, of β and δ, which is simplified by numerically solving for δ using the concentrated robust-AQS functions:

$$S_N^{\diamond c}(\delta) = \begin{cases} \mathbf{Y}'\mathbf{A}_N'(\lambda)\mathbf{B}_N'(\rho)[\bar{\mathbf{F}}_N'(\delta) - \bar{\mathbb{F}}_N'(\delta)]\hat{\mathbf{V}}(\delta), \\ [\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\hat{\beta}_N^{\diamond}(\delta)]'\mathbf{B}_N'(\rho)[\bar{\mathbf{G}}_N(\rho) - \bar{\mathbb{G}}_N(\rho)]\hat{\mathbf{V}}(\delta), \end{cases} \tag{4.4}$$

where $\hat{\beta}_N^{\diamond}(\delta) = \hat{\beta}_N^{*}(\delta)$ given in (3.8), and $\hat{\mathbf{V}}(\delta) = \tilde{\mathbf{V}}(\hat{\beta}_N^{\diamond}(\delta), \delta)$. Then, solving $S_N^{\diamond c}(\delta) = 0$, we obtain the RM-estimator $\hat{\delta}_N^{\diamond}$ of δ, and thus the RM-estimator $\hat{\beta}_N^{\diamond} \equiv \hat{\beta}_N^{\diamond}(\hat{\delta}_N^{\diamond})$ of β.

4.2. Asymptotic properties of the RM-estimator

Let $\bar{S}_N^{\diamond c}(\delta)$ be the concentrated $\mathrm{E}[S_N^{\diamond}(\theta)]$. Similar to Sec. 3, the key to the consistency of $\hat{\delta}_N^{\diamond}$ is the uniform convergence $\sup_{\delta\in\Delta} \left\| S_N^{\diamond c}(\delta) - \bar{S}_N^{\diamond c}(\delta) \right\| / N_1 \xrightarrow{p} 0$, and

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

Again, this is a high-level assumption put for simplicity, which holds under some low-level conditions. [4] Let $\xi = (\beta', \delta')'$ and $\hat{\xi}_N^{\diamond} = (\hat{\beta}_N^{\diamond\prime}, \hat{\delta}_N^{\diamond\prime})'$. We have the following theorem.

Theorem 3. *Under Assumptions A′, B-F and G′, we have, as $N \to \infty$, $\hat{\xi}_N^{\diamond} \xrightarrow{p} \xi_0$.*

Similarly, the asymptotic normality of $\hat{\xi}_N^{\diamond}$ can be established, by applying the mean value theorem to each element of $S_N^{\diamond}(\hat{\xi}_N^{\diamond}) = 0$ at ξ_0. The robust AQS function at ξ_0 is $S_N^{\diamond}(\xi_0) =$

[4] See Appendix D in `Supplementary Material` for details on $\bar{S}_N^{\diamond c}(\delta)$ and Assumption G′.

$[\mathbb{X}'\mathbf{V};\ \eta'\mathbf{B}_N'(\bar{\mathbf{F}}_N' - \bar{\mathbb{F}}_N')\mathbb{Q}_{\mathbb{D}}\mathbf{V} + \mathbf{V}'(\bar{\mathbf{F}}_N' - \bar{\mathbb{F}}_N')\mathbb{Q}_{\mathbb{D}}\mathbf{V};\ \phi_0'\mathbb{D}'(\bar{\mathbf{G}}_N - \bar{\mathbb{G}}_N)\mathbb{Q}_{\mathbb{D}}\mathbf{V} + \mathbf{V}'(\bar{\mathbf{G}}_N - \bar{\mathbb{G}}_N)\mathbb{Q}_{\mathbb{D}}\mathbf{V}]$, which is shown to be asymptotically normal by using the CLT for LQ forms of Kelejian and Prucha (2001). The adjusted Hessian $\partial S_N^{\diamond}(\bar{\xi})/\partial\xi'$ given in Appendix A, is shown to have a proper asymptotic behavior, for some $\bar{\xi}$ lying between $\hat{\xi}_N^{\diamond}$ and ξ_0 elementwise. Consequently, the asymptotic normality of $\hat{\xi}_N^{\diamond}$ is proved.

Theorem 4. *Under the assumptions of Theorem 3, we have, as $N \to \infty$,*

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

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

4.3. Heteroskedasticity robust inference

Robust inference for ξ_0 depends on the availability of consistent estimators of $\Sigma_N^{\diamond}(\xi_0)$ and $\Gamma_N^{\diamond}(\xi_0)$. As in the homoskedasticity case, $\Sigma_N^{\diamond}(\xi_0)$ can be estimated by its observed counterpart $\widehat{\Sigma}_N^{\diamond} = -N_1^{-1}\partial S_N^{\diamond}(\xi)/\partial\xi'|_{\xi=\hat{\xi}_N^{\diamond}}$, with detailed expression of $\partial S_N^{\diamond}(\xi)/\partial\xi'$ being given in Appendix A. The consistency of $\widehat{\Sigma}_N^{\diamond}$ is proved in the proof of Theorem 4.

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

$$\begin{aligned}
N_1\Gamma_{\beta\xi}^{\diamond} &= [\mathbb{X}'\mathbf{H}\mathbb{X},\ \ \mathbb{X}'\mathbf{H}\mathbb{L}_\lambda\mathbf{B}_N\eta,\ \ \mathbb{X}'\mathbf{H}\mathbb{L}_\rho\mathbb{D}\phi_0],\\
N_1\Gamma_{\lambda\lambda}^{\diamond} &= \eta'\mathbf{B}_N'\mathbb{L}_\lambda'\mathbf{H}\mathbb{L}_\lambda\mathbf{B}_N\eta + \mathtt{tr}(\mathbf{H}\mathbb{L}_\lambda\mathbf{H}\mathbb{L}_\lambda^{\circ}),\\
N_1\Gamma_{\lambda\rho}^{\diamond} &= \eta'\mathbf{B}_N'\mathbb{L}_\lambda'\mathbf{H}\mathbb{L}_\rho\mathbb{D}\phi_0 + \mathtt{tr}(\mathbf{H}\mathbb{L}_\lambda\mathbf{H}\mathbb{L}_\rho^{\circ}),\\
N_1\Gamma_{\rho\rho}^{\diamond} &= \phi_0'\mathbb{D}'\mathbb{L}_\rho'\mathbf{H}\mathbb{L}_\rho\mathbb{D}\phi_0 + \mathtt{tr}(\mathbf{H}\mathbb{L}_\rho\mathbf{H}\mathbb{L}_\rho^{\circ}),
\end{aligned} \tag{4.5}$$

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

Write $\Gamma_N^{\diamond}(\xi_0)$ as $\Gamma_N^{\diamond}(\xi_0, \phi, \mathbf{H})$. Let $\hat{\phi}_N^{\diamond}$ be the estimator of ϕ by plugging the RM-estimator $\hat{\xi}_N^{\diamond}$ in (3.3). Let $\Gamma_N^{\diamond}(\hat{\xi}_N^{\diamond}, \hat{\phi}_N^{\diamond}, \mathbf{H})$ be the plug-in estimator of $\Gamma_N^{\diamond}(\xi_0)$ for a given $\mathbf{H}$. We show that such a 'plug-in' results in a non-negligible bias: $\mathrm{Bias}_\phi^{\diamond}(\delta_0, \mathbf{H})$, with β-related entries being zero, and δ entries being $\mathtt{tr}(\mathbf{H}\mathbb{P}_{\mathbb{D}}\mathbb{L}_a'\mathbf{H}\mathbb{L}_b\mathbb{P}_{\mathbb{D}})/N_1, a, b = \lambda, \rho$.

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

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

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

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

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

From (4.5), we see that the elements of $\Gamma_N^\diamond(\xi_0, \phi, \mathbf{H})$ take two forms: $\mathtt{tr}(\mathbf{H}C_N)$ or $\mathtt{tr}(\mathbf{H}A_N\mathbf{H}B_N)$. Further, the bias term, $\mathrm{Bias}_\phi^\diamond(\delta_0, \mathbf{H})$, is also of the second form. It is important to know the effects of replacing $\mathbf{H}$ by $\hat{\mathbf{H}}$ in these forms. We show that the effect is non-negligible only for the second form. The resulting bias, denoted by $\mathrm{Bias}_{\mathbf{H}}^\diamond(\delta_0, \mathbf{H})$, has non-zero δ-entries, $2N_1^{-1}\mathtt{tr}((\mathbb{L}_a \odot \mathbb{L}_b^\circ - \mathbb{P}_{\mathbb{D}}\mathbb{L}_a' \odot \mathbb{L}_b\mathbb{P}_{\mathbb{D}})\Pi_N\Lambda(\mathbf{H})\Pi_N)$, $a, b = \lambda, \rho$; $\Pi_N(\rho) = [\mathbb{Q}_{\mathbb{D}}(\rho) \odot \mathbb{Q}_{\mathbb{D}}(\rho)]^-$; and $\Lambda(\mathbf{H}) = \{(q_j'\mathbf{H}q_k)^2\}_{j,k=1}^N$ with q_j' being the jth row of $\mathbb{Q}_{\mathbb{D}}$.

Combining the two results above, a consistent estimator of $\Gamma_N^\diamond(\xi_0)$ is given as follows:

$$\widehat{\Gamma}_N^\diamond = \Gamma_N^\diamond(\hat{\xi}_N^\diamond, \hat{\phi}_N^\diamond, \widehat{\mathbf{H}}) - \mathrm{Bias}_\phi^\diamond(\hat{\delta}_N^\diamond, \widehat{\mathbf{H}}) - \mathrm{Bias}_{\mathbf{H}}^\diamond(\hat{\delta}_N^\diamond, \widehat{\mathbf{H}}). \tag{4.6}$$

See `Supplementary Material` for the derivation of $\Gamma_N^\diamond(\xi_0)$ (Lemma B.5), the derivations of $\mathrm{Bias}_\phi^\diamond(\delta_0, \mathbf{H})$ and $\mathrm{Bias}_{\mathbf{H}}^\diamond(\delta_0, \mathbf{H})$ (Corollary D.1 and Lemma D.1), and the proofs of consistency of $\widehat{\Sigma}_N^\diamond$ and $\widehat{\Gamma}_N^\diamond$ (Corollary D.2).

5. M-Estimation of High-Order Models

Consider Model (2.2). Let $A_t(\boldsymbol{\lambda}) = I_{n_t} - \sum_{k=1}^p \lambda_k W_{kt}$ and $B_t(\boldsymbol{\rho}) = I_{n_t} - \sum_{\ell=1}^q \rho_\ell M_{\ell t}$; and $\mathbf{A}_N(\boldsymbol{\lambda}) = \mathtt{blkdiag}\{A_1(\boldsymbol{\lambda}), \dots, A_T(\boldsymbol{\lambda})\}$ and $\mathbf{B}_N(\boldsymbol{\rho}) = \mathtt{blkdiag}\{B_1(\boldsymbol{\rho}), \dots, B_T(\boldsymbol{\rho})\}$, where $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_p)'$ and $\boldsymbol{\rho} = (\rho_1, \dots, \rho_q)'$. Let $\boldsymbol{\delta} = (\boldsymbol{\lambda}', \boldsymbol{\rho}')'$, and $\boldsymbol{\theta} = (\beta', \sigma_v^2, \boldsymbol{\delta}')'$.

Homoskedasticity. With these extended notations, under Assumption A, the concentrated Gaussian loglikelihood function of $\boldsymbol{\theta}$ remains in the same form as (3.4). The AQS vector for Model (2.2) is derived along the same idea as that for (2.1):

$$S_N^*(\boldsymbol{\theta}) = \begin{cases} \frac{1}{\sigma_v^2}\mathbf{X}'\mathbf{B}_N'(\boldsymbol{\rho})\tilde{\mathbf{V}}(\beta, \boldsymbol{\delta}), \\ \frac{1}{2\sigma_v^4}\left[\tilde{\mathbf{V}}'(\beta, \boldsymbol{\delta})\tilde{\mathbf{V}}(\beta, \boldsymbol{\delta}) - N_1\sigma_v^2\right], \\ \frac{1}{\sigma_v^2}\tilde{\mathbf{V}}'(\beta, \boldsymbol{\delta})\mathbf{B}_N(\boldsymbol{\rho})\dot{\mathbf{A}}_{Nk}(\boldsymbol{\lambda})\mathbf{Y} - \mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\boldsymbol{\rho})\bar{\mathbf{F}}_{Nk}(\boldsymbol{\delta})],\ k = 1, \dots, p, \\ \frac{1}{\sigma_v^2}\tilde{\mathbf{V}}'(\beta, \boldsymbol{\delta})\mathbf{G}_{N\ell}(\boldsymbol{\rho})\tilde{\mathbf{V}}(\beta, \boldsymbol{\delta}) - \mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\boldsymbol{\rho})\mathbf{G}_{N\ell}(\boldsymbol{\rho})],\ \ell = 1, \dots, q, \end{cases} \tag{5.1}$$

where $\dot{\mathbf{A}}_{Nk}(\boldsymbol{\lambda}) = -\partial\mathbf{A}_N(\boldsymbol{\lambda})/\partial\lambda_k$ and $\mathbf{F}_{Nk}(\boldsymbol{\lambda}) = \dot{\mathbf{A}}_{Nk}(\boldsymbol{\lambda})\mathbf{A}_N^{-1}(\boldsymbol{\lambda})$; $\dot{\mathbf{B}}_{N\ell}(\boldsymbol{\rho}) = -\partial\mathbf{B}_N(\boldsymbol{\rho})/\partial\rho_\ell$; and $\mathbf{G}_{N\ell}(\boldsymbol{\rho}) = \dot{\mathbf{B}}_{N\ell}(\boldsymbol{\rho})\mathbf{B}_N^{-1}(\boldsymbol{\rho})$; and $\bar{\mathbf{F}}_{Nk}(\boldsymbol{\delta}) = \mathbf{B}_N(\boldsymbol{\rho})\mathbf{F}_{Nk}(\boldsymbol{\lambda})\mathbf{B}_N^{-1}(\boldsymbol{\rho})$.

The M-estimator $\hat{\boldsymbol{\theta}}_N^*$ of $\boldsymbol{\theta}_0$ solves $S_N^*(\boldsymbol{\theta}) = 0$. Consistency and asymptotic normality of $\hat{\boldsymbol{\theta}}_N^*$ can be established in a similar way as for the first-order model. In particular, under the extended Assumptions A-H, as $N_1 \to \infty$, $\hat{\boldsymbol{\theta}}_N^* \xrightarrow{p} \boldsymbol{\theta}_0$, and

$$\sqrt{N_1}(\hat{\boldsymbol{\theta}}_N^* - \boldsymbol{\theta}_0) \xrightarrow{D} N\big[0, \lim_{N\to\infty} \boldsymbol{\Sigma}_N^{*-1}(\boldsymbol{\theta}_0)\boldsymbol{\Gamma}_N^*(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}_N^{*-1\prime}(\boldsymbol{\theta}_0)\big], \tag{5.2}$$

where $\boldsymbol{\Sigma}_N^*(\boldsymbol{\theta}_0) = -N_1^{-1}\partial S_N^*(\boldsymbol{\theta})/\partial\boldsymbol{\theta}'$ and $\boldsymbol{\Gamma}_N^*(\boldsymbol{\theta}_0) = \mathrm{Var}[S_N^*(\boldsymbol{\theta}_0)]/N_1$. For practical applications, the analytical expressions for $\partial S_N^*(\boldsymbol{\theta})/\partial\boldsymbol{\theta}'$ and $\mathrm{Var}[S_N^*(\boldsymbol{\theta}_0)]$ are given in Appendix A, with which the corrected plug-in estimator of the VC matrix of $\hat{\boldsymbol{\theta}}_N^*$ is derived.

Heteroskedasticity. When errors are heteroskedastic as in Assumption A$'$, we need to find an alternative set of estimating functions robust against unknown $\mathbf{H}$. Following the same idea leading to (4.3), we obtain the robust AQS function of $\boldsymbol{\xi} = (\beta', \boldsymbol{\delta}')'$:

$$S_N^\diamond(\boldsymbol{\xi}) = \begin{cases} \mathbb{X}'(\boldsymbol{\rho})\tilde{\mathbf{V}}(\beta,\boldsymbol{\delta}), \\ \mathbf{Y}'\mathbf{A}_N'(\boldsymbol{\lambda})\mathbf{B}_N'(\boldsymbol{\rho})[\bar{\mathbf{F}}_{Nk}'(\boldsymbol{\delta}) - \bar{\mathbb{F}}_{Nk}'(\boldsymbol{\delta})]\tilde{\mathbf{V}}(\beta,\boldsymbol{\delta}),\ k = 1,\ldots,p, \\ [\mathbf{A}_N(\boldsymbol{\lambda})\mathbf{Y} - \mathbf{X}\beta]'\mathbf{B}_N'(\boldsymbol{\rho})[\bar{\mathbf{G}}_{N\ell}(\boldsymbol{\rho}) - \bar{\mathbb{G}}_{N\ell}(\boldsymbol{\rho})]\tilde{\mathbf{V}}(\beta,\boldsymbol{\delta}),\ \ell = 1,\ldots,q, \end{cases} \tag{5.3}$$

where $\bar{\mathbb{F}}_{Nk}'(\boldsymbol{\delta}) = \mathtt{diag}[\bar{\mathbf{F}}_{Nk}'(\boldsymbol{\delta})\mathbb{Q}_{\mathbb{D}}(\boldsymbol{\rho})]\mathtt{diag}[\mathbb{Q}_{\mathbb{D}}(\boldsymbol{\rho})]^{-1}$, $\bar{\mathbf{G}}_{N\ell}(\boldsymbol{\rho}) = \mathbb{Q}_{\mathbb{D}}(\boldsymbol{\rho})\mathbf{G}_{N\ell}(\boldsymbol{\rho})$, and $\bar{\mathbb{G}}_{N\ell}(\boldsymbol{\rho})$ $= \mathtt{diag}[\bar{\mathbf{G}}_{N\ell}(\boldsymbol{\rho})\mathbb{Q}_{\mathbb{D}}(\boldsymbol{\rho})]\mathtt{diag}[\mathbb{Q}_{\mathbb{D}}(\boldsymbol{\rho})]^{-1}$. Solving $S_N^\diamond(\boldsymbol{\xi}) = 0$ gives the RM-estimator $\hat{\boldsymbol{\xi}}_N^\diamond$ of $\boldsymbol{\xi}$. Consistency and asymptotic normality of $\hat{\boldsymbol{\xi}}_N^\diamond$ can be proved in a similar manner as for the first-order model in Section 4. For practical applications, the analytical expressions of $\partial S_N^\diamond(\boldsymbol{\xi})/\partial\boldsymbol{\xi}'$ and $\mathrm{Var}[S_N^\diamond(\boldsymbol{\xi}_0)]$ are given in Appendix A, with which the corrected plug-in estimator of the VC matrix of $\hat{\boldsymbol{\xi}}_N^\diamond$ is obtained.

6. Monte Carlo Results

Extensive Monte Carlo experiments are conducted to investigate the finite sample performance of the proposed M-estimators and the corresponding standard error estimators. We first consider a $\mathtt{SARAR}(1,1)$ data-generating process (DGP):

$$Y_t = \lambda W_t Y_t + X_t\beta + D_t\mu + \alpha_t l_{n_t} + U_t, \quad U_t = \rho M_t U_t + V_t, \ \ t = 1,\ldots,T.$$

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

The spatial weight matrices can be `Rook` contiguity, `Queen` contiguity, or `Group` interactions. The distribution of the idiosyncratic errors can be (i) `normal`, (ii) standardized `normal mixture` (10% $N(0,4^2)$ and 90% $N(0,1)$), or (iii) standardized `chi-square` with 3 degrees of freedom. The selection matrices D_t are generated as follows: for each t, associate with each row of I_n a uniform (0, 1) random number. Delete the rows if the corresponding random numbers are smaller than $p_t \in (0,1)$. This give $100p_t$% non-presence units in t-th period. To generate spatial panel data with GU, we first generate the full vectors/matrices $(V_t^*, \mu, X_t^*, W_t^*, M_t^*)$ for each t, then do deletions according to the generated D_t to give $V_t = D_t V_t^*$, $X_t = D_t X_t^*$, $W_t = D_t W_t^* D_t'$, and $M_t = D_t M_t^* D_t'$, and then generate Y_t according to the `DGP`. See `Supplementary Material` for details.

Monte Carlo (empirical) means and standard deviations (sd, shown in parentheses) are recorded for the `naïve` estimator,[5] `QMLE`, M-estimator (`M-Est`), and RM-estimator (`RM-Est`). The empirical averages of the standard error estimates ($\hat{se}$, shown in square brackets) are recorded for the `M-Est` and `RM-Est`, based on the methods introduced in Sections 3-5. Partial Monte Carlo results on `QMLE`, `M-Est` and `RM-Est` are reported in Table 1, with the full set being given in `Supplementary Material`.

Under homoskedasticity, both `M-Est` and `RM-Est` perform excellently in the finite sample and uniformly outperform the `QMLE`, in particular in the estimation of λ and ρ, irrespective of the values of n and T, spatial layouts, and error distributions. The proposed standard error estimators for the `M-Est` and `RM-Est` also perform excellently, with the estimates of standard errors $\hat{sd}$'s being on average very close to the corresponding Monte Carlo sd's. The $\sqrt{N_1}$-consistency of the `M-Est` and `RM-Est` is clearly demonstrated by the reduction of the Monte Carlo sds and the estimated sds as N increases.

Under heteroskedasticity with a `Group` scheme of a fixed set of group sizes (3, 5, 7, 9, 11, 15),[6] only `RM-Est` is valid and the Monte Carlo results (reported and unreported) confirm its excellent finite sample performance in terms of point estimation and standard error estimation. The $\sqrt{N_1}$-consistency of the `RM-Est` is also well demonstrated by the Monte Carlo results. In contrast, the QMLE, and `M-Est` generally provide very poor estimates for spatial parameters, and their inconsistency is clearly demonstrated.

[5]The `naïve` estimator is the M-estimator based on the balanced panel formed by including only the spatial units that are present in every period. This allows us to see the consequence of 'balancing by deletion' in a spatial context. See Monte Carlo results in `Supplementary Material`.

[6]Replicating the set increases n. Under this `Group` scheme, the variation in group sizes does not shrink to zero as n increases, giving a scenario where the M-estimators are inconsistent when heteroskedasticity depends on group sizes (Liu and Yang, 2020). We follow Lin and Lee (2010) to generate such an $\mathbf{H}$.

Table 1: Empirical mean(sd)[$\hat{se}$] of `QMLE`, `M-Est` and `RM-Est` of `SARAR(1,1)`: Unbalancedness percentage = 10% , and $(\beta, \lambda, \rho, \sigma_v^2) = (1, 0.2, 0.2, 1)$.

	T=5			T=10		
	`QMLE`	`M-Est`	`RM-Est`	`QMLE`	`M-Est`	`RM-Est`
Homoskedasticity, $n = 100$; error = 1, 2, 3, for the three panels below; `W = Rook` and `M = Queen`						
β	1.0010(.027)	1.0011(.026)[.027]	1.0011(.026)[.027]	1.0001(.018)	.9997(.018)[.018]	.9997(.018)[.018]
λ	.1922(.043)	.1993(.043)[.042]	.1994(.043)[.042]	.1924(.027)	.1993(.027)[.027]	.1993(.027)[.027]
ρ	.1565(.099)	.1906(.096)[.100]	.1906(.096)[.099]	.1600(.063)	.1952(.062)[.063]	.1952(.062)[.063]
σ_v^2	.7617(.060)	.9942(.078)[.076]	–	.8792(.044)	.9986(.050)[.050]	–
β	.9993(.028)	.9994(.028)[.027]	.9994(.028)[.027]	1.0005(.018)	1.0000(.018)[.018]	1.0000(.018)[.018]
λ	.1923(.042)	.1994(.042)[.042]	.1994(.042)[.042]	.1932(.027)	.2001(.027)[.027]	.2000(.027)[.027]
ρ	.1623(.102)	.1962(.099)[.099]	.1962(.099)[.096]	.1634(.062)	.1985(.061)[.063]	.1985(.061)[.062]
σ_v^2	.7624(.128)	.9951(.167)[.160]	–	.8773(.102)	.9964(.116)[.112]	–
β	.9983(.027)	.9984(.027)[.027]	.9984(.027)[.027]	1.0005(.018)	1.0001(.018)[.018]	1.0001(.018)[.018]
λ	.1937(.043)	.2009(.043)[.042]	.2009(.043)[.042]	.1923(.027)	.1992(.027)[.027]	.1992(.027)[.027]
ρ	.1621(.100)	.1961(.097)[.099]	.1961(.097)[.098]	.1609(.064)	.1961(.063)[.063]	.1961(.063)[.063]
σ_v^2	.7625(.092)	.9951(.120)[.118]	–	.8782(.073)	.9975(.083)[.082]	–
Homoskedasticity, $n = 400$; error = 1, 2, 3, for the three panels below; `W = Rook` and `M = Queen`						
β	1.0003(.014)	1.0003(.014)[.013]	1.0003(.014)[.013]	1.0004(.009)	1.0003(.009)[.009]	1.0003(.009)[.009]
λ	.1985(.019)	.2001(.019)[.019]	.2001(.019)[.019]	.1982(.014)	.1998(.014)[.014]	.1999(.014)[.014]
ρ	.1936(.049)	.1982(.048)[.047]	.1982(.048)[.047]	.1918(.033)	.1989(.032)[.031]	.1989(.032)[.031]
σ_v^2	.7738(.028)	.9966(.036)[.038]	–	.8854(.022)	.9982(.024)[.025]	–
β	1.0001(.013)	1.0000(.013)[.013]	1.0000(.013)[.013]	.9998(.009)	.9997(.009)[.009]	.9997(.009)[.009]
λ	.1985(.019)	.2001(.019)[.020]	.2001(.019)[.019]	.1983(.013)	.1999(.013)[.014]	.2000(.013)[.014]
ρ	.1937(.048)	.1983(.047)[.048]	.1983(.047)[.047]	.1931(.031)	.2001(.030)[.031]	.2001(.030)[.031]
σ_v^2	.7782(.063)	1.0023(.081)[.082]	–	.8847(.050)	.9974(.056)[.057]	–
β	1.0001(.013)	1.0001(.013)[.013]	1.0001(.013)[.013]	.9995(.009)	.9994(.009)[.009]	.9994(.009)[.009]
λ	.1972(.020)	.1988(.020)[.020]	.1987(.020)[.019]	.1978(.013)	.1994(.013)[.014]	.1996(.013)[.014]
ρ	.1944(.050)	.1990(.049)[.047]	.1990(.049)[.047]	.1931(.031)	.2002(.031)[.031]	.2002(.031)[.031]
σ_v^2	.7743(.049)	.9973(.063)[.060]	–	.8873(.038)	1.0004(.043)[.042]	–
Heteroskedasticity, $n = 100$; error = 1, 2, 3, for the three panels below; `W = M = Group`						
β	1.0005(.028)	1.0005(.028)[.028]	1.0002(.028)[.028]	.9994(.018)	.9995(.018)[.018]	.9996(.018)[.018]
λ	.1927(.049)	.1954(.048)[.064]	.1992(.056)[.054]	.1901(.034)	.1936(.034)[.043]	.1980(.040)[.038]
ρ	.0883(.131)	.1304(.120)[.127]	.1573(.167)[.160]	.1077(.082)	.1470(.077)[.080]	.1809(.106)[.101]
σ_v^2	.7572(.071)	.9925(.093)[.105]	–	.8663(.053)	.9860(.060)[.067]	–
β	1.0004(.029)	1.0004(.029)[.028]	1.0001(.029)[.028]	.9991(.018)	.9992(.018)[.018]	.9992(.018)[.018]
λ	.1921(.049)	.1948(.048)[.063]	.1985(.055)[.053]	.1902(.033)	.1937(.032)[.043]	.1978(.038)[.038]
ρ	.0884(.129)	.1305(.118)[.128]	.1578(.165)[.157]	.1106(.078)	.1497(.073)[.080]	.1848(.101)[.099]
σ_v^2	.7554(.155)	.9901(.203)[.199]	–	.8659(.124)	.9855(.141)[.139]	–
β	.9997(.029)	.9997(.029)[.028]	.9995(.029)[.028]	.9999(.018)	1.0000(.018)[.018]	1.0000(.018)[.018]
λ	.1914(.049)	.1941(.048)[.063]	.1979(.055)[.054]	.1922(.033)	.1957(.032)[.043]	.2005(.037)[.038]
ρ	.0877(.130)	.1299(.119)[.128]	.1566(.167)[.159]	.1077(.079)	.1470(.074)[.080]	.1808(.102)[.100]
σ_v^2	.7614(.115)	.9979(.150)[.152]	–	.8630(.088)	.9822(.100)[.102]	–
Heteroskedasticity, $n = 400$; error = 1, 2, 3, for the three panels below; `W = M = Group`						
β	.9999(.014)	.9999(.014)[.013]	.9999(.014)[.014]	1.0003(.009)	1.0009(.009)[.009]	1.0003(.009)[.009]
λ	.1966(.026)	.1970(.026)[.031]	.1998(.030)[.030]	.1974(.016)	.2490(.018)[.019]	.2010(.018)[.018]
ρ	.1491(.060)	.1550(.058)[.058]	.1892(.075)[.074]	.1533(.037)	.1210(.028)[.037]	.1959(.046)[.047]
σ_v^2	.7849(.034)	1.0110(.044)[.052]	–	.8923(.027)	1.0063(.030)[.034]	–
β	.9998(.014)	.9998(.014)[.013]	.9998(.014)[.014]	.9995(.009)	1.0001(.009)[.009]	.9995(.009)[.009]
λ	.1968(.027)	.1972(.027)[.031]	.1998(.031)[.030]	.1965(.017)	.2493(.019)[.019]	.1997(.019)[.018]
ρ	.1509(.061)	.1568(.059)[.058]	.1914(.075)[.074]	.1562(.038)	.1232(.029)[.037]	.1996(.048)[.047]
σ_v^2	.7878(.079)	1.0148(.102)[.103]	–	.8933(.061)	1.0075(.069)[.072]	–
β	1.0000(.013)	1.0000(.013)[.013]	1.0000(.014)[.014]	.9998(.009)	1.0004(.009)[.009]	.9998(.009)[.009]
λ	.1949(.027)	.1953(.027)[.031]	.1980(.031)[.030]	.1968(.017)	.2485(.018)[.019]	.2003(.019)[.018]
ρ	.1500(.059)	.1559(.057)[.058]	.1904(.073)[.074]	.1531(.038)	.1208(.029)[.037]	.1958(.047)[.047]
σ_v^2	.7869(.059)	1.0136(.076)[.078]	–	.8960(.045)	1.0105(.051)[.053]	–

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

Table 2: Empirical mean(sd)[$\hat{se}$] of `QMLE`, `M-Est` and `RM-Est` of `SARAR(2,2)`: Unbalancedness percentage = 10%, and $(\beta, \lambda_1, \lambda_2, \rho_1, \rho_2, \sigma_v^2) = (1, 0.2, 0.3, 0.2, 0.3, 1)$.

	T=5			T=10		
	QMLE	M-Est	RM-Est	QMLE	M-Est	RM-Est
	Homoskedasticity, $n = 100$; error = 1, 2, for the two panels below; $\texttt{W}_1$ = $\texttt{M}_1$ = Queen, $\texttt{W}_2$ = $\texttt{M}_2$ = Rook					
β	.9979(.027)	.9977(.027)[.028]	.9977(.027)[.028]	.9995(.018)	.9995(.018)[.018]	.9995(.018)[.018]
λ_1	.1998(.053)	.1953(.055)[.052]	.1954(.055)[.052]	.1981(.036)	.1967(.037)[.037]	.1966(.037)[.036]
λ_2	.2970(.044)	.2982(.043)[.042]	.2983(.044)[.042]	.2996(.029)	.3002(.030)[.029]	.3003(.030)[.029]
ρ_1	.1573(.156)	.1988(.131)[.122]	.1987(.131)[.121]	.1568(.081)	.1969(.076)[.076]	.1972(.076)[.075]
ρ_2	.3417(.119)	.2968(.100)[.096]	.2972(.100)[.095]	.3034(.064)	.2997(.059)[.059]	.2998(.059)[.058]
σ_v^2	.7451(.062)	.9794(.080)[.077]	–	.8734(.045)	.9907(.051)[.050]	–
β	.9992(.029)	.9990(.029)[.028]	.9990(.029)[.028]	.9980(.018)	.9980(.018)[.018]	.9980(.018)[.017]
λ_1	.1989(.052)	.1946(.054)[.052]	.1943(.054)[.051]	.2006(.035)	.1994(.036)[.036]	.1994(.036)[.036]
λ_2	.2976(.041)	.2990(.040)[.042]	.2992(.040)[.041]	.2990(.028)	.2999(.029)[.029]	.2999(.029)[.029]
ρ_1	.1438(.157)	.1877(.129)[.123]	.1881(.130)[.122]	.1535(.082)	.1937(.076)[.076]	.1938(.076)[.075]
ρ_2	.3457(.114)	.2988(.096)[.095]	.2989(.096)[.093]	.3011(.065)	.2973(.060)[.059]	.2976(.060)[.058]
σ_v^2	.7443(.124)	.9787(.163)[.158]	–	.8723(.102)	.9895(.116)[.112]	–
	Homoskedasticity, $n = 200$; error = 1, 2, for the two panels below; $\texttt{W}_1$ = $\texttt{M}_1$ = Queen, $\texttt{W}_2$ = $\texttt{M}_2$ = Rook					
β	.9991(.020)	.9991(.020)[.019]	.9991(.020)[.019]	.9994(.013)	.9994(.013)[.012]	.9994(.013)[.012]
λ_1	.2013(.034)	.1993(.035)[.034]	.1993(.035)[.033]	.1995(.025)	.1986(.026)[.026]	.1986(.026)[.025]
λ_2	.2971(.026)	.2992(.026)[.027]	.2992(.026)[.027]	.2993(.018)	.2999(.019)[.019]	.2999(.019)[.019]
ρ_1	.1980(.103)	.1969(.093)[.083]	.1970(.094)[.083]	.1907(.056)	.2014(.052)[.053]	.2014(.052)[.053]
ρ_2	.3660(.075)	.2988(.068)[.065]	.2991(.068)[.065]	.3163(.045)	.2974(.042)[.040]	.2975(.042)[.041]
σ_v^2	.7528(.044)	.9902(.057)[.055]	–	.8763(.033)	.9942(.037)[.036]	–
β	.9998(.019)	.9997(.019)[.019]	.9997(.019)[.019]	.9996(.012)	.9996(.012)[.012]	.9996(.012)[.012]
λ_1	.1999(.033)	.1978(.035)[.033]	.1977(.035)[.033]	.2002(.025)	.1993(.025)[.025]	.1993(.025)[.025]
λ_2	.2984(.026)	.3003(.026)[.027]	.3003(.026)[.027]	.2987(.019)	.2992(.019)[.019]	.2993(.019)[.019]
ρ_1	.2026(.098)	.2013(.091)[.082]	.2016(.091)[.081]	.1867(.056)	.1978(.052)[.053]	.1980(.052)[.053]
ρ_2	.3694(.077)	.3025(.070)[.065]	.3027(.070)[.064]	.3186(.045)	.2994(.041)[.041]	.2994(.041)[.040]
σ_v^2	.7504(.091)	.9873(.119)[.114]	–	.8766(.069)	.9946(.078)[.080]	–
	Heteroskedasticity, $n = 100$; error = 1, 2, for the two panels below; $\texttt{W}_1$ = $\texttt{M}_1$ = Group, $\texttt{W}_2$ = $\texttt{M}_2$ = Rook					
β	.9980(.029)	.9972(.029)[.028]	.9983(.030)[.029]	.9992(.018)	1.0000(.018)[.018]	.9999(.018)[.018]
λ_1	.1922(.044)	.1856(.046)[.061]	.1963(.054)[.052]	.1942(.027)	.1950(.022)[.032]	.2003(.025)[.025]
λ_2	.2966(.038)	.2952(.041)[.039]	.2964(.039)[.040]	.2974(.028)	.2995(.028)[.028]	.2993(.028)[.029]
ρ_1	.0382(.191)	.1294(.130)[.111]	.1754(.193)[.178]	.0534(.177)	.1284(.071)[.076]	.1843(.109)[.104]
ρ_2	.3479(.109)	.3069(.109)[.091]	.3036(.100)[.092]	.3080(.066)	.2983(.057)[.059]	.2977(.057)[.059]
σ_v^2	.7486(.076)	.9790(.100)[.102]	–	.8728(.055)	.9914(.061)[.067]	–
β	.9986(.028)	.9976(.029)[.028]	.9987(.030)[.029]	.9991(.018)	.9996(.018)[.018]	.9995(.018)[.018]
λ_1	.1893(.042)	.1832(.047)[.061]	.1945(.054)[.053]	.1926(.025)	.1940(.022)[.032]	.1994(.025)[.025]
λ_2	.2964(.038)	.2945(.041)[.038]	.2955(.040)[.039]	.2977(.029)	.2997(.029)[.028]	.2995(.029)[.029]
ρ_1	.0434(.179)	.1324(.135)[.114]	.1797(.202)[.177]	.0577(.148)	.1251(.070)[.077]	.1801(.107)[.103]
ρ_2	.3518(.105)	.3131(.110)[.092]	.3088(.099)[.090]	.3121(.064)	.3020(.059)[.059]	.3015(.059)[.057]
σ_v^2	.7517(.161)	.9832(.211)[.205]	–	.8678(.128)	.9865(.145)[.142]	–
	Heteroskedasticity, $n = 200$; error = 1, 2, for the two panels below; $\texttt{W}_1$ = $\texttt{M}_1$ = Group, $\texttt{W}_2$ = $\texttt{M}_2$ = Rook					
β	1.0009(.019)	1.0012(.019)[.020]	1.0011(.019)[.019]	1.0002(.012)	1.0003(.012)[.012]	1.0001(.012)[.012]
λ_1	.1982(.029)	.1970(.027)[.033]	.2013(.030)[.031]	.1926(.016)	.1937(.016)[.022]	.1993(.018)[.019]
λ_2	.2965(.028)	.3004(.028)[.028]	.3001(.028)[.028]	.2987(.019)	.3001(.019)[.020]	.3000(.019)[.019]
ρ_1	.1005(.144)	.1290(.069)[.074]	.1809(.103)[.104]	.1014(.072)	.1268(.050)[.054]	.1885(.076)[.075]
ρ_2	.3721(.076)	.2970(.064)[.065]	.2981(.065)[.064]	.3236(.044)	.3010(.041)[.041]	.2999(.041)[.041]
σ_v^2	.7581(.048)	.9924(.061)[.066]	–	.8739(.039)	.9914(.044)[.048]	–
β	.9977(.020)	.9982(.019)[.020]	.9981(.019)[.019]	.9997(.013)	.9998(.012)[.012]	.9996(.012)[.012]
λ_1	.1976(.031)	.1964(.028)[.033]	.2007(.031)[.031]	.1931(.017)	.1942(.017)[.022]	.1997(.019)[.019]
λ_2	.2952(.028)	.2994(.028)[.028]	.2990(.029)[.028]	.2994(.018)	.3007(.018)[.020]	.3007(.018)[.019]
ρ_1	.0956(.143)	.1252(.069)[.076]	.1756(.102)[.103]	.1036(.063)	.1278(.049)[.055]	.1900(.074)[.074]
ρ_2	.3733(.073)	.2976(.061)[.065]	.2993(.062)[.064]	.3218(.044)	.2994(.041)[.041]	.2985(.041)[.040]
σ_v^2	.7607(.105)	.9956(.137)[.137]	–	.8711(.093)	.9882(.105)[.103]	–

Note: error = 1(normal), 2(normal mixture).

We then extend the Monte Carlo experiments by using a SARAR$(2, 2)$ DGP with an additional set of spatial weight matrices, to demonstrate the finite sample performance of the proposed set of estimation and inference methods for a higher-order GU-SPD model. Partial Monte Carlo results on `QMLE`, `M-Est` and `RM-Est` are reported in Table 2, with the full set being given in `Supplementary Material`. General conclusions remain.

In summary, Monte Carlo results provide clear and strong support to our theoretical predictions, suggesting that the proposed methods are very reliable. They further suggest that in real applications when homoskedasticity holds either M-Estimator or RM-Estimator can be used but when it is in doubt one should use the RM-Estimator.

7. An Empirical Illustration

The spillover effect as a determinant of inward or outward foreign direct investment (FDI) has been extensively studied (see, among others, Coughlin and Segev, 2000; Baltagi et al., 2007, 2008; Blonigen et al., 2007). Coughlin and Segev (2000) studied the total FDI inflows (1990-1997) of 29 administrative *divisions* of China mainland (excluding Tibet), based on a spatial cross-sectional model and maximum likelihood estimation. The result exhibits a significant spatial error dependence. However, their study ignores the panel structure of the data. Practically important issues, such as the heterogeneity across 'provinces' and over time, spatiotemporal heteroskedasticity, and the effects of splits of the two provinces, Guangdong and Sichuan, were not investigated.

In this section, an extended study is given on the issue of spatial spillovers of FDI inflows (denoted by F_t) across Chinese administrative divisions, based on our GU-SDP model. First, our panel data covers a total of $T = 16$ years (1985 to 2000), including the year 1988 when Hainan became an individual province split from Guangdong, and the year 1997 when Chongqing became a municipality separated from Sichuan. With Sichuan and Guangdong provinces before and after the splits being treated as 'different' spatial units, we have a total of $n = 33$ spatial units in the study, with 29 in the first 3 years, 30 in the next 9 years, and 31 in last 4 years (including 22 provinces, 5 autonomous regions, and 4 municipalities), giving rise to a genuinely unbalanced spatial panel data. See Table 3 for details[7].

[7] Sources: Data from 1985 to 1991 are drawn from the *China Foreign Economic Statistics 1979–1991*. Data from 1992 to 1995 are sourced from the *China Foreign Economic Statistical Yearbook 1996*. Data from 1996 to 2000 are provided by the *State Statistical Bureau Yearbook 1997–2001*.

Table 3: China FDI inflows by administrative divisions (US$ Million), 1980 Constant Prices

Year	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
Beijing	68	105	77	350	213	176	148	205	380	762	584	816	818	1096	977	806
Tianjin	34	32	97	43	21	23	80	63	299	564	822	1054	1289	1068	872	558
Hebei	3	5	7	13	29	28	34	66	226	291	296	434	565	722	515	325
Shanxi	0.3	0.1	4	5	7	2	2	32	49	18	35	72	137	124	194	108
InnerMongolia	4	1.0	4	4	3	7	1	3	49	22	31	38	38	46	32	51
Liaoning	12	25	66	91	84	162	219	303	729	800	770	913	1132	1107	525	978
Jilin	2	2	5	7	7	11	19	44	157	134	221	237	207	207	149	161
Heilongjiang	2	13	10	48	38	18	13	42	132	193	279	288	377	266	157	144
Shanghai	48	73	155	162	280	110	88	290	1801	1374	1564	2070	2169	1821	1403	1512
Jiangsu	9	14	63	87	84	84	133	859	1621	2091	2806	2736	2790	3352	3006	3075
Zhejiang	13	14	26	30	36	31	56	141	588	639	680	799	772	666	610	772
Anhui	2	6	2	19	6	9	6	32	147	206	261	266	223	140	129	152
Fujian	90	46	40	101	231	202	285	836	1635	2063	2186	2145	2154	2129	1990	1642
Jiangxi	4	3	4	6	6	5	12	59	119	145	156	158	245	235	159	109
Shandong	4	15	47	62	109	117	131	589	1068	1418	1454	1361	1280	1114	1117	1422
Henan	4	5	10	45	31	7	23	31	174	215	259	275	355	312	258	270
Hubei	4	44	19	16	19	20	28	119	308	334	338	357	406	492	453	452
Hunan	13	7	2	9	15	9	15	78	249	184	274	369	471	414	323	325
Guangdong	436	425	578	-	-	-	-	-	-	-	-	-	-	-	-	-
Guangdong*	-	-	-	871	879	998	1175	2173	4308	5257	5546	6105	6012	6076	5766	5398
Guangxi	10	28	33	15	35	22	19	107	504	465	364	345	452	448	314	251
Hainan	-	-	-	82	71	65	107	266	403	510	574	414	362	363	240	206
Chongqing	-	-	-	-	-	-	-	-	-	-	-	-	199	218	118	117
Sichuan	5	17	18	28	9	15	49	66	326	512	293	223	-	-	-	-
Sichuan*	-	-	-	-	-	-	-	-	-	-	-	-	127	188	169	209
Guizhou	1.1	0.6	1	7	8	7	9	12	24	35	31	16	26	23	20	12
Yunnan	1.2	3	5	6	5	5	2	17	55	36	53	34	85	74	76	61
Tibet	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0
Shaanxi	6	7	53	78	65	30	19	27	134	133	175	171	322	152	120	138
Gansu	2	0.2	0.2	2	0	1	3	0.2	7	49	35	47	21	20	20	30
Qinghai	0.8	0.02	0	2	0	0	0	0.4	2	1.3	0.9	0.5	1.3	0	2	0
Ningxia	0.2	1.3	0.02	0.2	0.7	0.2	0.1	2	7	4	2.1	2.9	3.4	9	25	8
Xinjiang	1.2	11	13	4	0.6	3	0.1	0	30	27	30	34	13	11	12	9
Total	441	442	1337	2193	2292	2167	2677	6463	15533	18482	20116	21780	23050	22893	19751	19301

Note: Guangdong and Sichuan Provinces and Guangdong* and Sichuan* Provinces represent the provinces before and after the splits, respectively.

We extend the model considered in Coughlin and Segev (2000) by including all three types of spillovers, spatial lag, spatial error, and spatial Durbin, controlling for two-way fixed effects, and allowing for unknown spatiotemporal heteroskedasticity. A similar set of exogenous variables as in Coughlin and Segev (2000) is used: `GPP` (Gross Provincial Product in billion yuan), `WAGE` (average annual wage of staff and workers in yuan), `PROD` (overall labor productivity of industrial enterprises in yuan), `AIR` (number of staff and workers in state-owned units of airway transportation per ten thousand population), `HIWAY` (length of the paved highway (km) divided by area (1000 km^2)), and `UGRD` (number of undergraduate graduates per ten thousand population)[8]. Spatial weight matrices are constructed as in Coughlin and Segev (2000) with regions sharing a common border being treated as neighbors, which are row-normalized for each period. The selection matrix D_t and the number of units in each year n_t are defined based on

[8]Sources: All the explanatory variables are from *China Statistical Yearbook 1986-2001*, National Bureau of Statistics of China, https://www.stats.gov.cn/english/.

Table 3. These give us the final econometric model:

$$
\begin{aligned}
F_t = {} & \lambda W_t F_t + \beta_1 \texttt{GPP}_t + \beta_2 \texttt{WAGE}_t + \beta_3 \texttt{PROD}_t + \beta_4 \texttt{AIR}_t + \beta_5 \texttt{HIWAY}_t + \beta_6 \texttt{UGRD}_t \\
& + \beta_7 W_t \times \texttt{GPP}_t + \beta_8 W_t \times \texttt{WAGE}_t + \beta_9 W_t \times \texttt{PROD}_t + \beta_{10} W_t \times \texttt{AIR}_t + \beta_{11} W_t \times \texttt{HIWAY}_t \\
& + \beta_{12} W_t \times \texttt{UGRD}_t + D_t \mu + \alpha_t l_{n_t} + U_t, \qquad U_t = \rho W_t U_t + V_t.
\end{aligned} \tag{7.1}
$$

Table 4 reports the estimation results for (7.1) based on our M-estimation and RM-estimation under the GU specification in the columns below `M-Est` and `RM-Est`, respectively. As a comparison, we also report results under a balanced panel setup (`M-Est-Ba` and `RM-Est-Ba`) that ignore the two splits. We can see the significance levels and values of spatial parameter estimates under GU and balanced specification are totally different. Both M- and RM-estimators for λ under balanced setup are statistically insignificant, while RM-estimator under GU specification is statistically significantly positive, suggesting the existence of spillover effects among China FDI inflows. The latter is in line with the theoretical prediction in Coughlin and Segev (2000) that FDI agglomeration may lead to higher FDI levels in neighboring provinces to the extent that its beneficial effects spill over province borders. Both GU estimators for ρ are not statistically significant, while `RM-Est-Ba` for ρ in the third column exhibits statistically significant results. We also see the estimate values are different based on M- and RM-estimation. As administrative divisions differ substantially in FDI inflows and economic sizes, heteroskedasticity likely exists, and therefore the results based on RM-estimation are more reliable. In addition, all the estimators report statistically significantly positive estimates for `GPP` and `WAGE`, and both GU estimators also find significantly positive effects of `HIGHWAY`.

The spatial interdependence effects are also captured by the spatially weighted regressors. Overall, we can see that more spatial Durbin (or contextual) effects are found under GU specification than under the balanced panel setup. RM-estimator under GU specification reports negative and statistically significant contextual effects of `GPP` and `WAGE`, indicating that the factors attracting FDI in one place have negative impacts on FDI in neighboring places. Finally, both GU estimations report statistically significantly positive estimates for the spatial lag of `PROD`. Higher productivity in neighboring provinces may make the return on investment in that province more profitable and therefore attract more FDI.

Table 4: Empirical results for the China FDI inflows

Variables	M-Est-Ba	RM-Est-Ba	M-Est	RM-Est
SL(λ)	0.069	0.044	−0.228	0.375***
	(0.170)	(0.083)	(0.582)	(0.115)
SE(ρ)	0.268	0.284***	0.449	−0.182
	(0.199)	(0.096)	(0.505)	(0.152)
GPP	0.401***	0.401***	0.383***	0.400***
	(0.024)	(0.043)	(0.029)	(0.046)
WAGE	0.152***	0.151***	0.099***	0.113***
	(0.030)	(0.036)	(0.029)	(0.025)
PROD	0.009	0.009	0.009	0.005
	(0.006)	(0.006)	(0.007)	(0.005)
AIR	9.384	9.277	4.860	6.226
	(9.996)	(6.943)	(8.663)	(5.345)
HIGHWAY	5.375	5.440	11.087*	15.635**
	(6.057)	(7.267)	(6.359)	(7.399)
UGRD	−17.107	−17.012	−10.229	−8.735
	(18.484)	(14.987)	(17.784)	(15.046)
W × GPP	−0.097	−0.087	−0.041	−0.283***
	(0.080)	(0.071)	(0.254)	(0.083)
W × WAGE	−0.085	−0.082	−0.015	−0.064**
	(0.057)	(0.053)	(0.069)	(0.030)
W × PROD	0.034**	0.035***	0.036***	0.030***
	(0.015)	(0.015)	(0.012)	(0.013)
W × AIR	−20.903	−20.687	−6.837	−14.933
	(20.088)	(19.413)	(20.214)	(14.530)
W × HIGHWAY	4.700	4.948	−2.694	−7.444
	(14.031)	(13.691)	(11.187)	(8.218)
W × UGRD	14.623	14.755	8.421	−12.110
	(44.959)	(43.242)	(39.464)	(40.496)
Pseudo R^2	86.93%	86.96%	89.56%	89.73%
Observations	464	464	481	481

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

8. Conclusions and Discussions

We propose an M-estimation method for estimating an unbalanced spatial panel data (SPD) model with fixed effects, where the unbalancedness is of the genuine type due to the non-presence of spatial units. The method allows for the presence of high-order and time-varying spatial effects in response, regressors and errors, as well as unknown spatiotemporal heteroskedasticity. For statistical inference, we propose a simple *corrected plug-in* method that corrects the effect of plugging in the estimates of fixed effects parameters, and/or the estimates of the unknown spatiotemporal heteroskedasticity parameters. The proposed estimation and inference methods are seen to be simple and reliable and thus can be trustfully implemented by practitioners in an easy manner. New research can be generated as well.

The proposed methods are potentially applicable to many other scenarios, such as the SPD-GU model with (correlated) random effects and heteroskedasticity and SPD-GU model with multi-level effects and heteroskedasticity. The latter is particularly relevant to the social interaction and network models where time-varying group effects are of interest. Furthermore, the proposed methods are also applicable to the alternative $\texttt{MESS}(p, q)$ specification suggested by Yang (2018a). Another issue of immediate interest is when data contain both genuine non-presence and random missing, and in this case, imputation may be necessary (Loh et al., 2020; Sun and Wang, 2020).

Declaration of Interest Statement

The authors do not have any relevant financial or non-financial competing interests.

Acknowledgements

An early version of the paper entitled: *Unbalanced Spatial Panel Data Models with Fixed effects* and is labeled as Meng and Yang (2021), was presented at the Asia Meeting of the Econometric Society (AMES) 2019, Xiamen, China; XIII Conference of Spatial Econometrics Association 2019, Pittsburgh, US; Shanghai Workshop of Econometrics 2019, SUFE, Shanghai, China; AMES 2021, Curtin University, Malaysia; and Seminar at WISE, Xiamen University, China, 2021. The authors are grateful to Lung-Fei Lee, Ingmar Prucha, Xi Qu, and the conference and seminar participants for their helpful comments.

Data Availability Statement

The data that support the findings of this study are openly available in 4TU.ResearchData at https://data.4tu.nl/private_datasets/ztD3LvCNWEeHXXwAVpnWw2bOEq07rf8vdxuehk_SzJ8.

References

Anselin, L. (1988). *Spatial Econometrics: Methods and Models.* Springer Science & Business Media.

Anselin, L., J. L. Gallo, and H. Jayet (2008). Spatial Panel Econometrics. In L. Mátyás and P. Sevestre (Eds.), *The Econometrics of Panel Data: Fundamentals and Recent Developments* *in Theory and Practice*, Advanced Studies in Theoretical and Applied Econometrics, pp. 625–660. Berlin, Heidelberg: Springer.

Antweiler, W. (2001). Nested random effects estimation in unbalanced panel data. *Journal of Econometrics 101*, 295–313.

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

Bai, J. (2013). Likelihood approach to dynamic panel models with interactive effects. *Available at SSRN 2332992*.

Baltagi, B. H. (2021). *Econometric Analysis of Panel Data* (6th ed.). Springer.

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

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

Baltagi, B. H., P. Egger, and M. Pfaffermayr (2007). Estimating models of complex FDI: Are there third-country effects? *Journal of Econometrics 140*, 260–281.

Baltagi, B. H., P. Egger, and M. Pfaffermayr (2008). Estimating regional trade agreement effects on FDI in an interdependent world. *Journal of Econometrics 145*(1-2), 194–208.

Baltagi, B. H. and S. H. Song (2006). Unbalanced panel data: A survey. *Statistical Papers 47*, 493–523.

Baltagi, B. H., S. H. Song, and B. C. Jung (2001). The unbalanced nested error component regression model. *Journal of Econometrics 101*, 357–381.

Baltagi, B. H., S. H. Song, and W. Koh (2003). Testing panel data regression models with spatial error correlation. *Journal of Econometrics 117*, 123–150.

Baltagi, B. H. and Z. Yang (2013a). Heteroskedasticity and non-normality robust LM tests for spatial dependence. *Regional Science and Urban Economics 43*, 725–739.

Baltagi, B. H. and Z. Yang (2013b). Standardized LM tests for spatial error dependence in linear or panel regressions. *The Econometrics Journal 16*, 103–134.

Blonigen, B. A., R. B. Davies, G. R. Waddell, and H. T. Naughton (2007). FDI in space: Spatial autoregressive relationships in foreign direct investment. *European economic review 51*(5), 1303–1325.

Coughlin, C. C. and E. Segev (2000). Foreign direct investment in china: a spatial econometric study. *The world economy 51*, 1–23.

Davis, P. (2002). Estimating multi-way error components models with unbalanced data structures. *Journal of Econometrics 106*, 67–95.

Drukker, D. M., P. H. Egger, and I. R. Prucha (2022). Simultaneous equations models with higher-order spatial or social network interactions. *Econometric Theory*, 1–48.

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

Greene, W. H. (2018). *Econometric analysis* (8th ed.). Pearson.

Han, X., C.-S. Hsieh, and S. I. Ko (2021). Spatial modeling approach for dynamic network formation and interactions. *Journal of Business & Economic Statistics 39*(1), 120–135.

Hsiao, C. (2022). *Analysis of panel data* (8th ed.). Cambridge university press.

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

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

Kuersteiner, G. M. and I. R. Prucha (2020). Dynamic spatial panel models: Networks, common shocks, and sequential exogeneity. *Econometrica 88*(5), 2109–2146.

Lee, L.-F. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica 72*(6), 1899–1925.

Lee, L.-F. (2007). Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. *Journal of Econometrics 140*(2), 333–374.

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

Lee, L.-F. and J. Yu (2016). Identification of spatial Durbin panel models. *Journal of Applied Econometrics 31*, 133–162.

Li, L. and Z. Yang (2020). Estimation of fixed effects spatial dynamic panel data models with small T and unknown heteroskedasticity. *Regional Science and Urban Economics 81*, 103520.

Li, L. and Z. Yang (2021). Spatial dynamic panel data models with correlated random effects. *Journal of Econometrics 221*, 424–454.

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

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

Loh, W.-Y., Q. Zhang, W. Zhang, and P. Zhou (2020). Missing data, imputation and regression trees. *Statistica Sinica 30*(4), 1697–1722.

Lu, L. (2023). Simultaneous spatial panel data models with common shocks. *Journal of Business & Economic Statistics 41*(2), 608–623.

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

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

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

Neyman, J. and E. L. Scott (1948). Consistent estimates based on partially consistent observations. *Econometrica 16*, 1–32.

Sun, Z. and H. Wang (2020). Network imputation for a spatial autoregression model with incomplete data. *Statistica Sinica 30*(3), 1419–1436.

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

Wang, W. and L.-F. Lee (2013a). Estimation of spatial autoregressive models with randomly missing data in the dependent variable. *The Econometrics Journal 16*, 73–102.

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

Wansbeek, T. and A. Kapteyn (1989). Estimation of the error-components model with incomplete panels. *Journal of Econometrics 41*, 341–361.

Wooldridge, J. M. (2019). Correlated random effects models with unbalanced panels. *Journal of Econometrics 211*, 137–150.

Xu, Y. and Z. Yang (2020). Specification tests for temporal heterogeneity in spatial panel data models with fixed effects. *Regional Science and Urban Economics 81*, 103488.

Yang, Y., O. Doğan, and S. Taşpinar (2023). Estimation of matrix exponential unbalanced panel data models with fixed effects: An application to us outward FDI stock. *Journal of Business & Economic Statistics*, 1–16.

Yang, Z. (2018a). Bootstrap LM tests for higher-order spatial effects in spatial linear regression models. *Empirical Economics 55*, 35–68.

Yang, Z. (2018b). Unified M-estimation of fixed-effects spatial dynamic models with short panels. *Journal of Econometrics 205*, 423–447.

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

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

Appendix A: AQS, Hessian, and Variance of AQS

For Section 3. Write the key quantity in the concentrated quasi loglikelihood function (3.4) as $\tilde{\mathbf{V}}'(\beta,\delta)\tilde{\mathbf{V}}(\beta,\delta) = [\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\beta]'\mathbf{B}'_N(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho)[\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\beta]$. The derivatives of $\Phi(\rho) \equiv \mathbf{B}'_N(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho)$, in deriving the ρ-components of $S^*_N(\theta)$ and $\partial S^*_N(\theta)/\partial\theta'$, are the most complicated. With $\partial\mathbb{D}(\rho)/\partial\rho = -\mathbf{MD} = -\mathbf{G}_N(\rho)\mathbb{D}(\rho)$, we obtain:

$$\dot{\mathbb{Q}}_{\mathbb{D}}(\rho) \equiv \tfrac{\partial}{\partial\rho}\mathbb{Q}_{\mathbb{D}}(\rho) = \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{G}_N(\rho)\mathbb{P}_{\mathbb{D}}(\rho) + \mathbb{P}_{\mathbb{D}}(\rho)\mathbf{G}'_N(\rho)\mathbb{Q}_{\mathbb{D}}(\rho), \tag{A.1}$$

$$\dot{\Phi}(\rho) \equiv -\tfrac{\partial}{\partial\rho}\Phi(\rho) = \mathbf{B}'_N(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{G}^\circ_N(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho), \tag{A.2}$$

$$S^*_{N,\rho}(\theta) = \tfrac{1}{2\sigma_v^2}\tilde{\mathbf{V}}'(\beta,\delta)\mathbf{G}^\circ_N(\rho)\tilde{\mathbf{V}}'(\beta,\delta) - \mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{G}_N(\rho)], \tag{A.3}$$

$$a'[\tfrac{\partial}{\partial\rho}\dot{\Phi}(\rho)]a = 2a'\mathbf{B}'_N(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)[\mathbf{G}^\circ_N(\rho)\mathbb{P}_{\mathbb{D}}(\rho)\mathbf{G}^\circ_N(\rho) - \mathbf{G}'_N(\rho)\mathbf{G}_N(\rho)]\mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{B}_N(\rho)a, \tag{A.4}$$

where a is a constant vector, and $S^*_{N,\rho}(\theta)$ denotes the ρ-component of $S^*_N(\theta)$.

By (A.1)-(A.4), $S^*_N(\theta)$ in (3.7), $\partial \ln|\mathbf{A}(\lambda)|/\partial\lambda = \mathtt{tr}[\mathbf{A}^{-1}(\lambda)\partial\mathbf{A}(\lambda)/\partial\lambda]$ and $\partial\mathbf{A}^{-1}(\lambda)/\partial\lambda = -\mathbf{A}^{-1}(\lambda)[\partial\mathbf{A}(\lambda)/\partial\lambda]\mathbf{A}^{-1}(\lambda)$, one has the components of $H^*_N(\theta) = \partial S^*_N(\theta)/\partial\theta'$:

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

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

Now, applying Lemma B.5 to (3.11), we obtain $\Gamma^*_N(\theta_0)$ having distinct elements:

$$
\begin{aligned}
&N_1\Gamma^*_{\beta\theta} = \big[\tfrac{1}{\sigma_{v0}^2}\mathbb{X}'\mathbb{X},\ \tfrac{\kappa_3}{2\sigma_{v0}^3}\mathbb{X}'q,\ \tfrac{\kappa_3}{\sigma_{v0}}\mathbb{X}'\bar{f} + \tfrac{1}{\sigma_{v0}^2}\mathbb{X}'\eta^*,\ \tfrac{\kappa_3}{\sigma_{v0}}\mathbb{X}'g\big],\\
&N_1\Gamma^*_{\sigma_v^2\sigma_v^2} = \tfrac{1}{4\sigma_{v0}^4}(2N_1 + \kappa_4 q'q), \qquad N_1\Gamma^*_{\sigma_v^2\lambda} = \tfrac{\kappa_3}{2\sigma_{v0}^3}q'\eta^* + \tfrac{1}{2\sigma_{v0}^2}[2\mathtt{tr}(\bar{\mathcal{F}}_N) + \kappa_4 q'\bar{f}],\\
&N_1\Gamma^*_{\sigma_v^2\rho} = \tfrac{1}{2\sigma_{v0}^2}[2\mathtt{tr}(\mathcal{G}) + \kappa_4 q'g], \quad N_1\Gamma^*_{\lambda\lambda} = \tfrac{1}{\sigma_{v0}^2}\eta^{*\prime}\eta^* + \tfrac{2\kappa_3}{\sigma_{v0}}\bar{f}'\eta^* + \mathtt{tr}(\bar{\mathcal{F}}\bar{\mathcal{F}}^\circ) + \kappa_4\bar{f}'\bar{f},\\
&N_1\Gamma^*_{\lambda\rho} = \mathtt{tr}(\mathcal{G}\bar{\mathcal{F}}^\circ) + \kappa_4\bar{f}'g + \tfrac{\kappa_3}{\sigma_{v0}}g'\eta^*, \quad N_1\Gamma^*_{\rho\rho} = \mathtt{tr}(\mathcal{G}\mathcal{G}^\circ) + \kappa_4 g'g,
\end{aligned}
$$

where $\eta^* = \mathbb{Q}_{\mathbb{D}}\mathbf{B}_N\mathbf{F}_N\eta$, $\bar{\mathcal{F}} = \mathbb{Q}_{\mathbb{D}}\bar{\mathbf{F}}_N$, $\mathcal{G} = \mathbb{Q}_{\mathbb{D}}\mathbf{G}_N\mathbb{Q}_{\mathbb{D}}$, and $\bar{f}, g$, and q are the vectors of diagonal elements of $\bar{\mathcal{F}}$, $\mathcal{G}$ and $\mathbb{Q}_{\mathbb{D}}$, respectively.

For Section 4. Let $\mathbb{L}_\lambda(\delta) = \mathbb{Q}_{\mathbb{D}}(\rho)[\bar{\mathbf{F}}_N(\delta) - \bar{\mathbb{F}}'_N(\delta)]$ and $\mathbb{L}_\rho(\rho) = \mathbb{Q}_{\mathbb{D}}(\rho)[\bar{\mathbf{G}}'_N(\rho) - \bar{\mathbb{G}}_N(\rho)]$. For $S^\diamond_N(\xi)$ in (4.3), $H^\diamond_N(\xi) = \partial S^\diamond_N(\xi)/\partial\xi'$ is shown to have components:

$$
\begin{aligned}
&H^\diamond_{\beta\beta}(\xi) = -\mathbb{X}'(\rho)\mathbb{X}(\rho), && H^\diamond_{\beta\lambda}(\xi) = -\mathbb{X}'(\rho)\mathbb{Y}(\rho),\\
&H^\diamond_{\beta\rho}(\xi) = -\mathbb{X}'(\rho)\mathbf{G}^\circ_N(\rho)\tilde{\mathbf{V}}(\beta,\delta), && H^\diamond_{\lambda\beta}(\xi) = -\mathbf{Y}'\mathbf{C}'_N(\delta)\mathbb{L}'_\lambda(\delta)\mathbb{X}(\rho),\\
&H^\diamond_{\lambda\lambda}(\xi) = -\mathbf{Y}'\mathbf{W}'\mathbf{B}'_N(\rho)\mathbb{L}'_\lambda(\delta)\tilde{\mathbf{V}}(\beta,\delta) + \mathbf{Y}'\mathbf{C}'_N(\delta)\big[\mathbb{L}'_{\lambda\lambda}(\delta)\tilde{\mathbf{V}}(\beta,\delta) - \mathbb{L}'_\lambda(\delta)\mathbb{Y}(\rho)\big],\\
&H^\diamond_{\lambda\rho}(\xi) = \mathbf{Y}'\mathbf{C}'_N(\delta)\big[\mathbb{L}'_{\lambda\rho}(\delta) + \mathbb{L}'_\lambda(\delta)\mathbb{G}_N(\rho) - \mathbf{G}'_N(\rho)\mathbb{L}'_\lambda(\delta)\big]\tilde{\mathbf{V}}(\beta,\delta),\\
&H^\diamond_{\rho\beta}(\xi) = -\mathbf{V}'_0(\beta,\delta)\mathbb{L}'_\rho(\rho)\mathbb{X}(\rho) - \tilde{\mathbf{V}}'(\beta,\delta)\mathbb{L}_\rho(\rho)\mathbf{B}_N(\rho)\mathbf{X},\\
&H^\diamond_{\rho\lambda}(\xi) = -\mathbf{Y}'\mathbf{W}'\mathbf{B}'_N(\rho)\mathbb{L}'_\rho(\rho)\tilde{\mathbf{V}}(\beta,\delta) - \mathbf{V}'_0(\beta,\delta)\mathbb{L}'_\rho(\rho)\mathbb{Y}(\rho),\\
&H^\diamond_{\rho\rho}(\xi) = \mathbf{V}'_0(\beta,\delta)\big[\mathbb{L}_{\rho\rho}(\rho) + \mathbb{L}'_\rho(\rho)\mathbb{G}_N(\rho) - \mathbf{G}'_N(\delta)\mathbb{L}'_\rho(\rho)\big]\tilde{\mathbf{V}}(\beta,\delta),
\end{aligned}
$$

where $\mathbf{V}_0(\beta,\delta) = \mathbf{B}_N(\rho)[\mathbf{A}_N(\lambda)\mathbf{Y} - \mathbf{X}\beta]$, $\mathbb{G}_N(\rho) = \mathbb{P}_{\mathbb{D}}(\rho)\mathbf{G}'_N(\rho) - \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{G}_N(\rho)$,

$$
\begin{aligned}
&\mathbb{L}_{\lambda\lambda}(\delta) = \mathbb{Q}_{\mathbb{D}}(\rho)\big[\mathbf{B}_N(\rho)\mathbf{F}^2_N(\lambda)\mathbf{B}^{-1}_N(\rho) - \mathtt{diag}[\mathbf{B}^{-1\prime}_N(\rho)\mathbf{F}'^2_N(\lambda)\mathbf{B}'_N(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)]\mathtt{diag}[\mathbb{Q}_{\mathbb{D}}(\rho)]^{-1}\big],\\
&\mathbb{L}_{\lambda\rho}(\delta) = \tfrac{\partial}{\partial\rho}\mathbb{L}_\lambda(\delta) = \dot{\mathbb{Q}}_{\mathbb{D}}(\rho)[\bar{\mathbf{F}}_N(\delta) - \bar{\mathbb{F}}'_N(\delta)] + \mathbb{Q}_{\mathbb{D}}(\rho)[\bar{\mathbf{F}}_\rho(\delta) - \bar{\mathbb{F}}'_\rho(\delta)],\\
&\bar{\mathbf{F}}_\rho(\delta) = -\mathbf{M}\mathbf{F}_N(\lambda)\mathbf{B}^{-1}_N(\rho) + \mathbf{B}_N(\rho)\mathbf{F}_N(\lambda)\mathbf{B}^{-1}_N(\rho)\mathbf{G}_N(\rho),\\
&\bar{\mathbb{F}}'_\rho(\delta) = \mathtt{diag}[\bar{\mathbf{F}}'_\rho(\delta)\mathbb{Q}_{\mathbb{D}}(\rho) + \bar{\mathbf{F}}'_N(\delta)\dot{\mathbb{Q}}_{\mathbb{D}}(\rho)]\mathtt{diag}[\mathbb{Q}_{\mathbb{D}}(\rho)]^{-1} + \mathtt{diag}[\bar{\mathbf{F}}'_N(\delta)\mathbb{Q}_{\mathbb{D}}(\rho)]\bar{\mathbb{Q}}_{\mathbb{D}}(\rho),
\end{aligned}
$$

$$\bar{\mathbb{Q}}_{\mathbb{D}}(\rho) = -\mathtt{diag}[\mathbb{Q}_{\mathbb{D}}(\rho)]^{-1}\mathtt{diag}[\dot{\mathbb{Q}}_{\mathbb{D}}(\rho)]\mathtt{diag}[\mathbb{Q}_{\mathbb{D}}(\rho)]^{-1},$$
$$\mathbb{L}_{\rho\rho}(\rho) = \tfrac{\partial}{\partial\rho}\mathbb{L}'_{\rho}(\rho) = \dot{\mathbb{Q}}_{\mathbb{D}}(\rho)[\bar{\mathbf{G}}'_N(\rho)-\bar{\mathbb{G}}_N(\rho)] + \mathbb{Q}_{\mathbb{D}}(\rho)[\bar{\mathbf{G}}'_{\rho}(\rho)-\bar{\mathbb{G}}_{\rho}(\rho)],$$
$$\bar{\mathbf{G}}_{\rho}(\rho) = \dot{\mathbb{Q}}_{\mathbb{D}}(\rho)\mathbf{G}_N(\rho) + \mathbb{Q}_{\mathbb{D}}(\rho)\mathbf{G}_N^2(\rho), \text{ and}$$
$$\bar{\mathbb{G}}_{\rho}(\rho) = \mathtt{diag}[\bar{\mathbf{G}}_{\rho}(\rho)\mathbb{Q}_{\mathbb{D}}(\rho) + \bar{\mathbf{G}}_N(\rho)\dot{\mathbb{Q}}_{\mathbb{D}}(\rho)]\mathtt{diag}[\mathbb{Q}_{\mathbb{D}}(\rho)]^{-1} + \mathtt{diag}[\bar{\mathbf{G}}_N(\rho)\mathbb{Q}_{\mathbb{D}}(\rho)]\bar{\mathbb{Q}}_{\mathbb{D}}(\rho).$$

For Section 5 (Homoskedasticity). Again, the derivations of the $\boldsymbol{\rho}$-components of $S_N^*(\boldsymbol{\theta})$ give in (5.1) and the $\boldsymbol{\rho}$-components of $\partial S_N^*(\boldsymbol{\theta})/\partial\rho_\ell$ are the most complicated. Denote $\Phi(\boldsymbol{\rho}) = \mathbf{B}'_N(\boldsymbol{\rho})\mathbb{Q}_{\mathbb{D}}(\boldsymbol{\rho})\mathbf{B}_N(\boldsymbol{\rho})$. We have, similarly to (A.1)-(A.4),

$$\dot{\mathbb{Q}}_{\mathbb{D}\ell}(\boldsymbol{\rho}) \equiv \tfrac{\partial}{\partial\rho_\ell}\mathbb{Q}_{\mathbb{D}}(\boldsymbol{\rho}) = \mathbb{Q}_{\mathbb{D}}(\boldsymbol{\rho})\mathbf{G}_{N\ell}(\boldsymbol{\rho})\mathbb{P}_{\mathbb{D}}(\boldsymbol{\rho}) + \mathbb{P}_{\mathbb{D}}(\boldsymbol{\rho})\mathbf{G}'_{N\ell}(\boldsymbol{\rho})\mathbb{Q}_{\mathbb{D}}(\boldsymbol{\rho}),$$
$$\dot{\Phi}_\ell(\boldsymbol{\rho}) \equiv -\tfrac{\partial}{\partial\rho_\ell}\Phi(\boldsymbol{\rho}) = \mathbf{B}'_N(\boldsymbol{\rho})\mathbb{Q}_{\mathbb{D}}(\boldsymbol{\rho})\mathbf{G}^{\circ}_{N\ell}(\boldsymbol{\rho})\mathbb{Q}_{\mathbb{D}}(\boldsymbol{\rho})\mathbf{B}_N(\boldsymbol{\rho}),$$
$$S^*_{N,\rho_\ell}(\boldsymbol{\theta}) = \tfrac{1}{2\sigma_v^2}\tilde{\mathbf{V}}'(\beta,\boldsymbol{\delta})\mathbf{G}^{\circ}_{N\ell}(\boldsymbol{\rho})\tilde{\mathbf{V}}'(\beta,\boldsymbol{\delta}) - \mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\boldsymbol{\rho})\mathbf{G}_{N\ell}(\boldsymbol{\rho})],$$
$$a'[\tfrac{\partial}{\partial\rho_{\ell_2}}\dot{\Phi}_{\ell_1}(\boldsymbol{\rho})]a = 2a'\mathbf{B}'_N(\boldsymbol{\rho})\mathbb{Q}_{\mathbb{D}}(\boldsymbol{\rho})[\mathbf{G}^{\circ}_{N\ell_1}(\boldsymbol{\rho})\mathbb{P}_{\mathbb{D}}(\boldsymbol{\rho})\mathbf{G}^{\circ}_{N\ell_2}(\boldsymbol{\rho}) - \mathbf{G}'_{N\ell_1}(\boldsymbol{\rho})\mathbf{G}_{N\ell_2}(\boldsymbol{\rho})]\mathbb{Q}_{\mathbb{D}}(\boldsymbol{\rho})\mathbf{B}_N(\boldsymbol{\rho})a,$$

for $k, k_1, k_2 = 1, \ldots, p$ and $\ell, \ell_1, \ell_2 = 1, \ldots, q$, where a is a constant vector.

With the above results, we obtain $H_N^*(\boldsymbol{\theta}) = \partial S_N^*(\boldsymbol{\theta})/\partial\boldsymbol{\theta}'$ having components:

$$H^*_{\beta\beta}(\boldsymbol{\theta}) = -\tfrac{1}{\sigma_v^2}\mathbb{X}'(\boldsymbol{\rho})\mathbb{X}(\boldsymbol{\rho}), \qquad H^*_{\beta\sigma_v^2}(\boldsymbol{\theta}) = -\tfrac{1}{\sigma_v^4}\mathbb{X}'(\boldsymbol{\rho})\tilde{\mathbf{V}}(\beta,\boldsymbol{\delta}) = H^{*\prime}_{\sigma_v^2\beta}(\boldsymbol{\theta}),$$
$$H^*_{\beta\lambda_k}(\boldsymbol{\theta}) = -\tfrac{1}{\sigma_v^2}\mathbb{X}'(\boldsymbol{\rho})\mathbb{Y}_k(\boldsymbol{\rho}) = H^{*\prime}_{\lambda_k\beta}(\boldsymbol{\theta}), \quad H^*_{\beta\rho_\ell}(\boldsymbol{\theta}) = -\tfrac{1}{\sigma_v^2}\mathbb{X}'(\boldsymbol{\rho})\mathbf{G}^{\circ}_{N\ell}(\boldsymbol{\rho})\tilde{\mathbf{V}}(\beta,\boldsymbol{\delta}) = H^{*\prime}_{\rho_\ell\beta}(\boldsymbol{\theta}),$$
$$H^*_{\sigma_v^2\sigma_v^2}(\boldsymbol{\theta}) = -\tfrac{1}{\sigma_v^6}\tilde{\mathbf{V}}'(\beta,\boldsymbol{\delta})\tilde{\mathbf{V}}(\beta,\boldsymbol{\delta}) + \tfrac{1}{2\sigma_v^4}N_1, \quad H^*_{\sigma_v^2\lambda_k}(\boldsymbol{\theta}) = -\tfrac{1}{\sigma_v^4}\tilde{\mathbf{V}}'(\beta,\boldsymbol{\delta})\mathbb{Y}_k(\boldsymbol{\rho}) = H^{*\prime}_{\lambda_k\sigma_v^2}(\boldsymbol{\theta}),$$
$$H^*_{\sigma_v^2\rho_\ell}(\boldsymbol{\theta}) = -\tfrac{1}{2\sigma_v^4}\tilde{\mathbf{V}}'(\boldsymbol{\rho})\mathbf{G}^{\circ}_{N\ell}(\boldsymbol{\rho})\tilde{\mathbf{V}}(\beta,\boldsymbol{\delta}) = H^{*\prime}_{\rho_\ell\sigma_v^2}(\boldsymbol{\theta}), \quad H^*_{\rho_\ell\lambda_k}(\boldsymbol{\theta}) = -\tfrac{1}{\sigma_v^2}\tilde{\mathbf{V}}'(\boldsymbol{\rho})\mathbf{G}^{\circ}_{N\ell}(\boldsymbol{\rho})\mathbb{Y}_k(\boldsymbol{\rho})$$
$$H^*_{\lambda_{k_1}\lambda_{k_2}}(\boldsymbol{\theta}) = -\tfrac{1}{\sigma_v^2}\mathbb{Y}'_{k_1}(\boldsymbol{\rho})\mathbb{Y}_{k_2}(\boldsymbol{\rho}) - \mathtt{tr}[\mathbb{Q}_{\mathbb{D}}(\boldsymbol{\rho})\bar{\mathbf{F}}_{Nk_1k_2}(\boldsymbol{\delta})],$$
$$H^*_{\lambda_k\rho_\ell}(\boldsymbol{\theta}) = -\tfrac{1}{\sigma_v^2}\tilde{\mathbf{V}}'(\boldsymbol{\rho})\mathbf{G}^{\circ}_{N\ell}(\boldsymbol{\rho})\mathbb{Y}_k(\boldsymbol{\rho}) + \mathtt{tr}[\mathbf{F}_{Nk}(\boldsymbol{\lambda})\mathbb{R}_{N\ell}(\boldsymbol{\rho})], \, ,$$
$$H^*_{\rho_{\ell_1}\rho_{\ell_2}}(\boldsymbol{\theta}) = \tfrac{1}{\sigma_v^2}\tilde{\mathbf{V}}'(\beta,\boldsymbol{\delta})\mathcal{R}_{N\ell_1\ell_2}(\boldsymbol{\rho})\tilde{\mathbf{V}}(\beta,\boldsymbol{\delta}) - \mathtt{tr}[\dot{\mathbb{Q}}_{\mathbb{D}\ell_2}(\boldsymbol{\rho})\mathbf{G}_{N\ell_1}(\boldsymbol{\rho}) + \mathbb{Q}_{\mathbb{D}}(\boldsymbol{\rho})\mathbf{G}_{N\ell_1}(\boldsymbol{\rho})\mathbf{G}_{N\ell_2}(\boldsymbol{\rho})],$$

where $\mathbb{X}(\boldsymbol{\rho}) = \mathbb{Q}_{\mathbb{D}}(\boldsymbol{\rho})\mathbf{B}_N(\boldsymbol{\rho})\mathbf{X}$, $\mathbb{Y}_k(\boldsymbol{\rho}) = \mathbb{Q}_{\mathbb{D}}(\boldsymbol{\rho})\mathbf{B}_N(\boldsymbol{\rho})\dot{\mathbf{A}}_{Nk}(\boldsymbol{\lambda})\mathbf{Y}$, $\mathbf{F}_{Nk_1k_2}(\boldsymbol{\lambda}) = \mathbf{F}_{Nk_1}(\boldsymbol{\lambda})\mathbf{F}_{Nk_2}(\boldsymbol{\lambda})$, $\bar{\mathbf{F}}_{Nk_1k_2}(\boldsymbol{\delta}) = \mathbf{B}_N(\boldsymbol{\rho})\mathbf{F}_{Nk_1k_2}(\boldsymbol{\lambda})\mathbf{B}_N^{-1}(\boldsymbol{\rho})$, $\mathbb{R}_{N\ell}(\boldsymbol{\rho}) = \mathbf{B}_N^{-1}(\boldsymbol{\rho})\mathbb{P}_{\mathbb{D}}(\boldsymbol{\rho})\mathbf{G}^{\circ}_{N\ell}(\boldsymbol{\rho})\mathbb{Q}_{\mathbb{D}}(\boldsymbol{\rho})\mathbf{B}_N(\boldsymbol{\rho})$, and lastly $\mathcal{R}_{N\ell_1\ell_2}(\boldsymbol{\rho}) = \mathbf{G}^{\circ}_{N\ell_1}(\boldsymbol{\rho})\mathbb{P}_{\mathbb{D}}(\boldsymbol{\rho})\mathbf{G}^{\circ}_{N\ell_2}(\boldsymbol{\rho}) - \mathbf{G}'_{N\ell_1}(\boldsymbol{\rho})\mathbf{G}_{N\ell_2}(\boldsymbol{\rho})$.

To derive $\mathrm{Var}[S_N^*(\boldsymbol{\theta}_0)]$, we first derive $S_N^*(\boldsymbol{\theta}_0)$ as done in (3.11) for the first-order model. Then, denote $\eta_k^* = \mathbb{Q}_{\mathbb{D}}\mathbf{B}_N\mathbf{F}_{Nk}\eta$, $\bar{\mathcal{F}}_k = \mathbb{Q}_{\mathbb{D}}\bar{\mathbf{F}}_{Nk}$, and $\mathcal{G}_\ell = \mathbb{Q}_{\mathbb{D}}\mathbf{G}_{N\ell}\mathbb{Q}_{\mathbb{D}}$. By applying Lemma B.5 to $S_N^*(\boldsymbol{\theta}_0)$, we obtain $\mathrm{Var}[S_N^*(\boldsymbol{\theta}_0)]$, which has distinct elements:

$$N_1\Gamma^*_{\beta[\beta,\sigma_v^2,\lambda_k,\rho_\ell]} = \big[\tfrac{1}{\sigma_{v0}^2}\mathbb{X}'\mathbb{X},\ \ \tfrac{\kappa_3}{2\sigma_{v0}^3}\mathbb{X}'q,\ \ \tfrac{\kappa_3}{\sigma_{v0}}\mathbb{X}'\bar{f}_k + \tfrac{1}{\sigma_{v0}^2}\mathbb{X}'\eta_k^*,\ \ \tfrac{\kappa_3}{\sigma_{v0}}\mathbb{X}'g_\ell\big],$$
$$N_1\Gamma^*_{\sigma_v^2\sigma_v^2} = \tfrac{1}{4\sigma_{v0}^4}(2N_1 + \kappa_4 q'q), \qquad N_1\Gamma^*_{\sigma_v^2\lambda_k} = \tfrac{\kappa_3}{2\sigma_{v0}^3}q'\eta_k^* + \tfrac{1}{2\sigma_{v0}^2}[2\mathtt{tr}(\bar{\mathcal{F}}_k) + \kappa_4 q'\bar{f}_k],$$
$$N_1\Gamma^*_{\sigma_v^2\rho_\ell} = \tfrac{1}{2\sigma_{v0}^2}[2\mathtt{tr}(\mathcal{G}_\ell) + \kappa_4 q'g_\ell],$$
$$N_1\Gamma^*_{\lambda_{k_1}\lambda_{k_2}} = \tfrac{1}{\sigma_{v0}^2}\eta_{k_1}^{*\prime}\eta_{k_2}^* + \tfrac{2\kappa_3}{\sigma_{v0}}\bar{f}'_{k_1}\eta_{k_2}^* + \mathtt{tr}(\bar{\mathcal{F}}_{k_1}\bar{\mathcal{F}}^{\circ}_{k_2}) + \kappa_4\bar{f}'_{k_1}\bar{f}_{k_2},$$
$$N_1\Gamma^*_{\lambda_k\rho_\ell} = \mathtt{tr}(\mathcal{G}_\ell\bar{\mathcal{F}}^{\circ}_k) + \kappa_4\bar{f}'_k g_\ell + \tfrac{\kappa_3}{\sigma_{v0}}g'_\ell\eta_k^*, \qquad N_1\Gamma^*_{\rho_{\ell_1}\rho_{\ell_2}} = \mathtt{tr}(\mathcal{G}_{\ell_1}\mathcal{G}^{\circ}_{\ell_2}) + \kappa_4 g'_{\ell_1}g_{\ell_2},$$

where $\bar{f}_k = \mathtt{diagv}(\bar{\mathcal{F}}_k)$ and $g_\ell = \mathtt{diagv}(\mathcal{G}_\ell)$. A consistent estimator of $\Gamma_N^*(\theta_0)$ is $\Gamma_N^*(\hat{\theta}_N^*) - \text{Bias}^*(\hat{\delta}_N^*)$, where $\text{Bias}^*(\delta_0)$ has non-zero λ_{k_1}-λ_{k_2} entries: $N_1^{-1}\mathtt{tr}(\bar{\mathbf{F}}'_{Nk_1}\mathbb{Q}_\mathbb{D}\bar{\mathbf{F}}_{Nk_2}\mathbb{P}_\mathbb{D})$.

For Section 5 (Heteroskedasticity). Let $\boldsymbol{\xi} = (\beta', \boldsymbol{\delta}')'$, $\mathbb{L}_{\lambda_k}(\boldsymbol{\delta}) = \mathbb{Q}_\mathbb{D}(\boldsymbol{\rho})[\bar{\mathbf{F}}_{Nk}(\boldsymbol{\delta}) - \bar{\mathbb{F}}'_{Nk}(\boldsymbol{\delta})]$, and $\mathbb{L}_{\rho_\ell}(\boldsymbol{\rho}) = \mathbb{Q}_\mathbb{D}(\boldsymbol{\rho})[\bar{\mathbf{G}}'_{N\ell}(\boldsymbol{\rho}) - \bar{\mathbb{G}}_{N\ell}(\boldsymbol{\rho})]$. $H_N^\diamond(\boldsymbol{\xi}) = \partial S_N^\diamond(\boldsymbol{\xi})/\partial\boldsymbol{\xi}'$ has components:

$$H_{\beta\beta}^\diamond(\boldsymbol{\xi}) = -\mathbb{X}'(\boldsymbol{\rho})\mathbb{X}(\boldsymbol{\rho}), \qquad H_{\beta\lambda_k}^\diamond(\boldsymbol{\xi}) = -\mathbb{X}'(\boldsymbol{\rho})\mathbb{Y}_k(\boldsymbol{\rho}),$$
$$H_{\beta\rho_\ell}^\diamond(\boldsymbol{\xi}) = -\mathbb{X}'(\boldsymbol{\rho})\mathbf{G}_{N\ell}^\circ(\boldsymbol{\rho})\tilde{\mathbf{V}}(\beta,\boldsymbol{\delta}), \qquad H_{\lambda_k\beta}^\diamond(\boldsymbol{\xi}) = -\mathbf{Y}'\mathbf{C}'_N(\boldsymbol{\delta})\mathbb{L}'_{\lambda_k}(\boldsymbol{\delta})\mathbb{X}(\boldsymbol{\rho}),$$
$$H_{\lambda_{k_1}\lambda_{k_2}}^\diamond(\boldsymbol{\xi}) = -\mathbf{Y}'\dot{\mathbf{A}}'_{Nk_2}(\boldsymbol{\lambda})\mathbf{B}'_N(\boldsymbol{\rho})\mathbb{L}'_{\lambda_{k_1}}(\boldsymbol{\delta})\tilde{\mathbf{V}}(\beta,\boldsymbol{\delta}) + \mathbf{Y}'\mathbf{C}'_N(\boldsymbol{\delta})\big[\mathbb{L}'_{\lambda_{k_1}\lambda_{k_2}}(\boldsymbol{\delta})\tilde{\mathbf{V}}(\beta,\boldsymbol{\delta}) - \mathbb{L}'_{\lambda_{k_1}}(\boldsymbol{\delta})\mathbb{Y}_{k_2}(\boldsymbol{\rho})\big],$$
$$H_{\lambda_k\rho_\ell}^\diamond(\boldsymbol{\xi}) = \mathbf{Y}'\mathbf{C}'_N(\boldsymbol{\delta})\big[\mathbb{L}'_{\lambda_k\rho_\ell}(\boldsymbol{\delta}) + \mathbb{L}'_{\lambda_k}(\boldsymbol{\delta})\mathbb{G}_{N\ell}(\boldsymbol{\rho}) - \mathbf{G}'_{N\ell}(\boldsymbol{\rho})\mathbb{L}'_{\lambda_k}(\boldsymbol{\delta})\big]\tilde{\mathbf{V}}(\beta,\boldsymbol{\delta}),$$
$$H_{\rho_\ell\beta}^\diamond(\boldsymbol{\xi}) = -\mathbf{V}'_0(\beta,\boldsymbol{\delta})\mathbb{L}'_{\rho_\ell}(\boldsymbol{\rho})\mathbb{X}(\boldsymbol{\rho}) - \tilde{\mathbf{V}}'(\beta,\boldsymbol{\delta})\mathbb{L}_{\rho_\ell}(\boldsymbol{\rho})\mathbf{B}_N(\boldsymbol{\rho})\mathbf{X},$$
$$H_{\rho_\ell\lambda_k}^\diamond(\boldsymbol{\xi}) = -\mathbf{Y}'\dot{\mathbf{A}}'_{Nk}(\boldsymbol{\lambda})\mathbf{B}'_N(\boldsymbol{\rho})\mathbb{L}'_{\rho_\ell}(\boldsymbol{\rho})\tilde{\mathbf{V}}(\beta,\boldsymbol{\delta}) - \mathbf{V}'_0(\beta,\boldsymbol{\delta})\mathbb{L}'_{\rho_\ell}(\boldsymbol{\rho})\mathbb{Y}_{\lambda_k}(\boldsymbol{\rho}),$$
$$H_{\rho_{\ell_1}\rho_{\ell_2}}^\diamond(\boldsymbol{\xi}) = \mathbf{V}'_0(\beta,\boldsymbol{\delta})\big[\mathbb{L}_{\rho_{\ell_1}\rho_{\ell_2}}(\boldsymbol{\rho}) + \mathbb{L}'_{\rho_{\ell_1}}(\boldsymbol{\rho})\mathbb{G}_{N\ell_2}(\boldsymbol{\rho}) - \mathbf{G}'_{N\ell_2}(\boldsymbol{\delta})\mathbb{L}'_{\rho_{\ell_1}}(\boldsymbol{\rho})\big]\tilde{\mathbf{V}}(\beta,\boldsymbol{\delta}),$$

where $\mathbf{V}_0(\beta,\boldsymbol{\delta}) = \mathbf{B}_N(\boldsymbol{\rho})[\mathbf{A}_N(\boldsymbol{\lambda})\mathbf{Y} - \mathbf{X}\beta]$, $\mathbb{G}_{N\ell}(\boldsymbol{\rho}) = \mathbb{P}_\mathbb{D}(\boldsymbol{\rho})\mathbf{G}'_{N\ell}(\boldsymbol{\rho}) - \mathbb{Q}_\mathbb{D}(\boldsymbol{\rho})\mathbf{G}_{N\ell}(\boldsymbol{\rho})$,
$\mathbb{L}_{\lambda_{k_1}\lambda_{k_2}}(\boldsymbol{\delta}) = \mathbb{Q}_\mathbb{D}(\boldsymbol{\rho})\big[\mathbf{B}_N(\boldsymbol{\rho})\mathbf{F}_{Nk_1k_2}(\boldsymbol{\lambda})\mathbf{B}_N^{-1}(\boldsymbol{\rho}) - \mathtt{diag}[\mathbf{B}_N^{-1\prime}(\boldsymbol{\rho})\mathbf{F}'_{Nk_1k_2}(\boldsymbol{\lambda})\mathbf{B}'_N(\boldsymbol{\rho})\mathbb{Q}_\mathbb{D}(\boldsymbol{\rho})]\mathtt{diag}[\mathbb{Q}_\mathbb{D}(\boldsymbol{\rho})]^{-1}\big]$,
$\mathbb{L}_{\lambda_k\rho_\ell}(\boldsymbol{\delta}) = \dot{\mathbb{Q}}_{\mathbb{D}\ell}(\boldsymbol{\rho})[\bar{\mathbf{F}}_{Nk}(\boldsymbol{\delta}) - \bar{\mathbb{F}}'_{Nk}(\boldsymbol{\delta})] + \mathbb{Q}_\mathbb{D}(\boldsymbol{\rho})[\bar{\mathbf{F}}_{Nk\ell}(\boldsymbol{\delta}) - \bar{\mathbb{F}}'_{Nk\ell}(\boldsymbol{\delta})]$,
$\bar{\mathbf{F}}_{Nk\ell}(\boldsymbol{\delta}) = -\dot{\mathbf{B}}_{N\ell}(\boldsymbol{\rho})\mathbf{F}_{Nk}(\boldsymbol{\lambda})\mathbf{B}_N^{-1}(\boldsymbol{\rho}) + \mathbf{B}_N(\boldsymbol{\rho})\mathbf{F}_{Nk}(\boldsymbol{\lambda})\mathbf{B}_N^{-1}(\boldsymbol{\rho})\mathbf{G}_{N\ell}(\boldsymbol{\rho})$,
$\bar{\mathbb{F}}'_{Nk\ell}(\boldsymbol{\delta}) = \mathtt{diag}[\bar{\mathbf{F}}'_{Nk\ell}(\boldsymbol{\delta})\mathbb{Q}_\mathbb{D}(\boldsymbol{\rho}) + \bar{\mathbf{F}}'_{Nk}(\boldsymbol{\delta})\dot{\mathbb{Q}}_{\mathbb{D}\ell}(\boldsymbol{\rho})]\mathtt{diag}[\mathbb{Q}_\mathbb{D}(\boldsymbol{\rho})]^{-1} + \mathtt{diag}[\bar{\mathbf{F}}'_{Nk}(\boldsymbol{\delta})\mathbb{Q}_\mathbb{D}(\boldsymbol{\rho})]\bar{\mathbb{Q}}_{\mathbb{D}\ell}(\boldsymbol{\rho})$,
$\bar{\mathbb{Q}}_{\mathbb{D}\ell}(\boldsymbol{\rho}) = -\mathtt{diag}[\mathbb{Q}_\mathbb{D}(\boldsymbol{\rho})]^{-1}\mathtt{diag}[\dot{\mathbb{Q}}_{\mathbb{D}\ell}(\boldsymbol{\rho})]\mathtt{diag}[\mathbb{Q}_\mathbb{D}(\boldsymbol{\rho})]^{-1}$,
$\mathbb{L}_{\rho_{\ell_1}\rho_{\ell_2}}(\boldsymbol{\rho}) = \dot{\mathbb{Q}}_{\mathbb{D}\ell_2}(\boldsymbol{\rho})[\bar{\mathbf{G}}'_{N\ell_1}(\boldsymbol{\rho}) - \bar{\mathbb{G}}_{N\ell_1}(\boldsymbol{\rho})] + \mathbb{Q}_\mathbb{D}(\boldsymbol{\rho})[\bar{\mathbf{G}}'_{N\ell_1\ell_2}(\boldsymbol{\rho}) - \bar{\mathbb{G}}_{N\ell_1\ell_2}(\boldsymbol{\rho})]$,
$\bar{\mathbf{G}}_{N\ell_1\ell_2}(\boldsymbol{\rho}) = \dot{\mathbb{Q}}_{\mathbb{D}\ell_2}(\boldsymbol{\rho})\mathbf{G}_{N\ell_1}(\boldsymbol{\rho}) + \mathbb{Q}_\mathbb{D}(\boldsymbol{\rho})\mathbf{G}_{N\ell_1}(\boldsymbol{\rho})\mathbf{G}_{N\ell_2}(\boldsymbol{\rho})$, and
$\bar{\mathbb{G}}_{N\ell_1\ell_2}(\boldsymbol{\rho}) = \mathtt{diag}[\bar{\mathbf{G}}_{N\ell_1\ell_2}(\boldsymbol{\rho})\mathbb{Q}_\mathbb{D}(\boldsymbol{\rho}) + \bar{\mathbf{G}}_{N\ell_1}(\boldsymbol{\rho})\dot{\mathbb{Q}}_{\mathbb{D}\ell_2}(\boldsymbol{\rho})]\mathtt{diag}[\mathbb{Q}_\mathbb{D}(\boldsymbol{\rho})]^{-1} + \mathtt{diag}[\bar{\mathbf{G}}_{N\ell_1}(\boldsymbol{\rho})\mathbb{Q}_\mathbb{D}(\boldsymbol{\rho})]\bar{\mathbb{Q}}_{\mathbb{D}\ell_2}(\boldsymbol{\rho})$.

Similar to (4.5), the distinct elements of $\Gamma_N^\diamond(\boldsymbol{\xi}_0) \equiv \text{Var}\big[S_N^\diamond(\boldsymbol{\xi}_0)\big]/N_1$ are:

$$\begin{aligned} N_1\Gamma^\diamond_{\beta[\beta,\lambda_k,\rho_\ell]} &= [\mathbb{X}'\mathbf{H}\mathbb{X},\ \ \mathbb{X}'\mathbf{H}\mathbb{L}_{\lambda_k}\mathbf{B}_N\eta,\ \ \mathbb{X}'\mathbf{H}\mathbb{L}_{\rho_\ell}\mathbb{D}\phi_0],\\ N_1\Gamma^\diamond_{\lambda_{k_1}\lambda_{k_2}} &= \eta'\mathbf{B}'_N\mathbb{L}'_{\lambda_{k_1}}\mathbf{H}\mathbb{L}_{\lambda_{k_2}}\mathbf{B}_N\eta + \mathtt{tr}(\mathbf{H}\mathbb{L}_{\lambda_{k_1}}\mathbf{H}\mathbb{L}^\circ_{\lambda_{k_2}}),\\ N_1\Gamma^\diamond_{\lambda_k\rho_\ell} &= \eta'\mathbf{B}'_N\mathbb{L}'_{\lambda_k}\mathbf{H}\mathbb{L}_{\rho_\ell}\mathbb{D}\phi_0 + \mathtt{tr}(\mathbf{H}\mathbb{L}_{\lambda_k}\mathbf{H}\mathbb{L}^\circ_{\rho_\ell}),\\ N_1\Gamma^\diamond_{\rho_{\ell_1}\rho_{\ell_2}} &= \phi'_0\mathbb{D}'\mathbb{L}'_{\rho_{\ell_1}}\mathbf{H}\mathbb{L}_{\rho_{\ell_2}}\mathbb{D}\phi_0 + \mathtt{tr}(\mathbf{H}\mathbb{L}_{\rho_{\ell_1}}\mathbf{H}\mathbb{L}^\circ_{\rho_{\ell_2}}). \end{aligned} \tag{A.5}$$

A consistent estimator of $\Gamma_N^\diamond(\boldsymbol{\xi}_0)$ is $\widehat{\Gamma}_N^\diamond = \Gamma_N^\diamond(\hat{\boldsymbol{\xi}}_N^\diamond, \hat{\phi}_N^\diamond, \widehat{\mathbf{H}}) - \text{Bias}_\phi^\diamond(\hat{\boldsymbol{\delta}}_N^\diamond, \widehat{\mathbf{H}}) - \text{Bias}_\mathbf{H}^\diamond(\hat{\boldsymbol{\delta}}_N^\diamond, \widehat{\mathbf{H}})$, where $\text{Bias}_\phi^\diamond(\boldsymbol{\delta}_0, \mathbf{H})$ and $\text{Bias}_\mathbf{H}^\diamond(\boldsymbol{\delta}_0, \mathbf{H})$ have non-zero $\boldsymbol{\delta}$ entries: $\mathtt{tr}(\mathbf{H}\mathbb{P}_\mathbb{D}\mathbb{L}'_a\mathbf{H}\mathbb{L}_b\mathbb{P}_\mathbb{D})/N_1$ and $2\mathtt{tr}((\mathbb{L}_a \odot \mathbb{L}_b^\circ - \mathbb{P}_\mathbb{D}\mathbb{L}'_a \odot \mathbb{L}_b\mathbb{P}_\mathbb{D})\Pi_N\Lambda(\mathbf{H})\Pi_N)/N_1$, respectively, for $a, b = \lambda_k, \rho_\ell$.

Supplementary Material

The online `Supplementary Material` contains proofs of the theoretical results and a full set of the Monte Carlo results.