To appear in Spatial Economics Analysis
Accepted on 6 Nov 2026

Dynamic Spatial Panel Data Models with Fixed Effects: Alternative Approaches and Application

Liyao Li[a] , Xiaoyu Meng[b*] and Zhenlin Yang[c]

[a]*School of Statistics, KLATASDS-MOE, East China Normal University, Shanghai, China*

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

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

Emails: lyli@fem.ecnu.edu.cn; xiaoyumeng@nankai.edu.cn; zlyang@smu.edu.sg

September 23, 2026

Abstract

We propose an alternative approach for estimating dynamic spatial panel data models with fixed effects that concentrates out the fixed effects and applies bias correction to the conditional concentrated quasi-score functions. We introduce an alternative corrected plug-in method for inference. The proposed methods are robust to initial conditions, require no data transformation, and are applicable in a wide range of settings, including time-varying and non-normalized spatial weight matrices, higher-order spatial effects, and heteroskedastic errors. The theoretical properties of the methods are rigorously studied, and the finite sample performance of the methods is assessed through Monte Carlo simulation. An empirical application to stock-exchange data for 1179 Chinese firms from 2015 to 2018 provides strong evidence of both dynamic and spatial interaction effects.

Key Words: Adjusted quasi scores; Dynamic spatial panel; Initial conditions; Heteroskedasticity.

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

*Correspondence: Xiaoyu Meng, School of Economics, Nankai University, No.94 Weijin Road, Tianjin, China. E-mail: xiaoyumeng@nankai.edu.cn

1. Introduction

The widely-used dynamic spatial panel data (DSPD) model with individual and time-specific fixed effects (FE) is expressed in the following general form for an n-vector:

$$\begin{aligned} y_t &= \rho y_{t-1} + \lambda_1 W_{1t} y_t + \lambda_2 W_{2t} y_{t-1} + x_t \beta + \mu + \alpha_t 1_n + u_t, \\ u_t &= \lambda_3 W_{3t} u_t + v_t, \quad t = 1, 2, \ldots, T, \end{aligned} \tag{1.1}$$

where y_t is an $n \times 1$ vector of responses observed at time t, and y_{t-1} denotes its one-period lag. The matrix x_t is an $n \times k$ matrix of regressors, and μ denotes the vector of individual fixed effects. The term α_t captures time-specific fixed effects, with 1_n denoting an $n \times 1$ vector of ones. The v_t is an $n \times 1$ vector of idiosyncratic errors with mean zero and variance σ_v^2. The matrices W_{rt}, $r = 1, 2, 3$, are known $n \times n$ spatial weight matrices that may vary over time and are not required to be row-normalized. The model nests several important features commonly studied in the spatial econometrics literature, including dynamic effects (ρy_{t-1}), endogenous spatial spillover effects ($\lambda_1 W_{1t} y_t$), space-time diffusion effects ($\lambda_2 W_{2t} y_{t-1}$) and spatial error dependence ($\lambda_3 W_{3t} u_t$). Furthermore, the framework is readily extendable to accommodate spatial Durbin effects, higher-order spatial effects, alternative fixed effects specifications, and unknown heteroskedasticity.

The standard estimation method for Model (1.1) maximizes the conditional (given the initial values y_0) concentrated quasi log-likelihood (CCQL), in which the fixed effects are concentrated out to obtain the CCQL estimator of the common parameters $\psi = (\beta', \rho, \lambda', \sigma_v^2)$, where $\lambda = (\lambda_1, \lambda_2, \lambda_3)'$. Then a post-estimation bias-correction is required for valid inference (Yu et al., 2008). However, this approach is valid only under large-T asymptotics and assumes homoskedastic errors. When T is small and fixed, Yang (2018) propose an M-estimation approach based on a first-differenced model, in which first differencing is used to eliminate individual fixed effects. The method applies bias correction to the conditional quasi-scores, thereby yielding unbiased estimating functions that are robust to initial conditions, and develops outer-product-of-martingale-difference (OPMD) procedures for inference. Subsequently, Li and Yang (2020) extend the framework of Yang (2018) to allow for unknown cross-sectional heteroskedasticity. However, these methods require the spatial weight matrices to be time-invariant and are valid only under fixed-T asymptotics. In addition, Lee and Yu (2014) propose a GMM approach for DSPD models that is valid for both fixed T and large T, but it requires time-invariant

spatial weight matrices (though seemingly extendable) and homoskedastic disturbances.

In this paper, we propose an alternative estimation strategy that bias-corrects the conditional concentrated quasi score (CCQS) and then solves the resulting *adjusted quasi score* (AQS) equations. Our method shares a basic idea of Yang (2018) and Li and Yang (2020) but has major differences – it does not require a transformation and takes care of *initial values problem* and *incidental parameters problem* simultaneously. It differs from (Yu et al., 2008) in that it corrects the CCQS instead of the CCQL maximizer. For statistical inference, we adopt an alternative approach that corrects the plug-in estimate of the variance-covariance (VC) matrix of the AQS function, extending the idea of Meng and Yang (2025a,b), instead of a variant of the OPMD method adopted in Yang (2018) and Li and Yang (2020). The difference is that our corrected plug-in estimate of the VC matrix needs to handle the expected quadratic forms of the initial values. The proposed methods have additional merits: they impose no restriction on the temporal variation of spatial weight matrices, allow T to be fixed or large, and extend naturally to models with richer structures, including fixed effects with group structure and heteroskedastic errors.

In summary, we contribute to the DSPD literature by proposing an alternative estimation method that jointly corrects the initial value bias and the incidental parameter bias in the CCQS function, and an alternative inference method that bias-corrects the plug-in estimate of the VC matrix of the adjusted quasi scores. The proposed methods are quite flexible and can accommodate (*i*) unobserved spatiotemporal heterogeneity in the form of fixed effects, (*ii*) unknown spatiotemporal heteroskedasticity, (*iii*) time-varying spatial weight matrices, and (*iv*) higher-order spatial effects (see Section 6 for details). We carry out extensive Monte Carlo experiments to investigate the finite sample performance of the proposed methods and to compare with their main competitors. The results favor the proposed methods. Finally, we illustrate the practical usefulness of the proposed methods through an application to China's stock-exchange data.

The rest of the paper is organized as follows. Section 2 introduces the M-estimation and inference methods under homoskedasticity. Section 3 extends these methods to accommodate heteroskedastic errors. Section 4 presents Monte Carlo results. Section 5 provides an empirical application to Chinese firm-level data. Section 6 discusses possible extensions of the proposed methods. Section 7 concludes the paper.

Notations and conventions. $|\cdot|$, $\mathtt{tr}(\cdot)$, $'$, $\|\cdot\|$ and $\otimes$ are common notations for

the determinant, trace, transpose, matrix norm and Kronecker product. $\texttt{diag}(\cdot)$ 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 $\texttt{bdiag}(\cdot)$ forms a block-diagonal matrix using the given sub-matrices. $\gamma_{\max}(A)$ and $\gamma_{\min}(A)$ denote, respectively, the largest and smallest eigenvalues of a real symmetric matrix A. The expectation and variance operators, $\mathrm{E}(\cdot)$ and $\mathrm{Var}(\cdot)$, correspond to the true parameter values. Denote a parametric quantity at the true parameter values by dropping its argument(s), e.g., $\mathbf{B}_1 \equiv \mathbf{B}_1(\lambda_{10})$, $\mathbf{B}_2 \equiv \mathbf{B}_2(\rho_0, \lambda_{20})$ and $\mathbf{M} \equiv \mathbf{M}(\lambda_{30})$.

2. M-Estimation under Homoskedasticity

To fix ideas, we begin by presenting a complete analysis of the first-order model (1.1) under homoskedasticity. Define the dummy-variable matrices $\mathbf{D}_\mu = (1_T \otimes I_n)$ and $\mathbf{D}_\alpha = (I_T \otimes 1_n)$ corresponding to individual and time fixed effects, respectively. Stacking the observations over t, Model (1.1) can be written as

$$\begin{aligned} \mathbf{Y} &= \rho\mathbf{Y}_{-1} + \lambda_1\mathbf{W}_1\mathbf{Y} + \lambda_2\mathbf{W}_2\mathbf{Y}_{-1} + \mathbf{X}\beta + \mathbf{D}_\mu\mu + \mathbf{D}_\alpha\alpha + \mathbf{U}, \\ \mathbf{U} &= \lambda_3\mathbf{W}_3\mathbf{U} + \mathbf{V}, \end{aligned} \tag{2.1}$$

where $\mathbf{X}$ collects regressors that vary across both individuals and time, including spatial Durbin terms when applicable. Any time-invariant or space-invariant covariates, if present, are absorbed by the individual and time fixed effects μ and α, respectively. The matrices $\mathbf{W}_\nu = \texttt{bdiag}(W_{\nu 1}, \ldots, W_{\nu T})$, for $\nu = 1, 2, 3$, denote block-diagonal spatial weight matrices constructed from the period-specific weights.

In Model (2.1), μ and α are not separately identified since $\mu_i + \alpha_t = (\mu_i + c) + (\alpha_t - c)$ for any constant c. Therefore, we normalize by dropping the last individual effect of μ and defining $\mu^* = (\mu_1^*, \ldots, \mu_{n-1}^*)'$. Correspondingly, let $\mathbf{D}_\mu^*$ denote the $nT \times (n-1)$ matrix obtained by removing the last column of $\mathbf{D}_\mu$. Define $\phi = (\mu^{*\prime}, \alpha')'$ and $\mathbf{D} = (\mathbf{D}_\mu^*, \mathbf{D}_\alpha)$, so that the likelihood function of the common and FE parameters is based on the following model form:

$$\begin{aligned} \mathbf{Y} &= \rho\mathbf{Y}_{-1} + \lambda_1\mathbf{W}_1\mathbf{Y} + \lambda_2\mathbf{W}_2\mathbf{Y}_{-1} + \mathbf{X}\beta + \mathbf{D}\phi + \mathbf{U}, \\ \mathbf{U} &= \lambda_3\mathbf{W}_3\mathbf{U} + \mathbf{V}. \end{aligned} \tag{2.2}$$

Define $\mathbf{B}_r(\lambda_r) = \texttt{bdiag}\{B_{r,1}(\lambda_r), \cdots, B_{r,T}(\lambda_r)\}, r = 1, 3$, where $B_{r,t}(\lambda_r) = I_n - \lambda_r W_{r,t}$, and $\mathbf{B}_2(\rho, \lambda_2) = \texttt{bdiag}\{B_{2,1}(\rho, \lambda_2), \cdots, B_{2,T}(\rho, \lambda_2)\}$, where $B_{2,t}(\rho, \lambda_2) = \rho I_n + \lambda_2 W_{2,t}$.

Let $\psi = (\beta', \rho, \lambda', \sigma_v^2)'$ collect all the common parameters, where $\lambda = (\lambda_1, \lambda_2, \lambda_3)'$. The conditional quasi-Gaussian loglikelihood function given y_0, **as if** $\mathbf{V} \sim N(0, \sigma_{v0}^2 I_{nT})$, is

$$\begin{aligned}\ell_{nT}(\psi, \phi) = &-\tfrac{nT}{2}\log(2\pi\sigma_v^2) - \tfrac{1}{2}\log|\mathbf{\Omega}(\lambda_3)| + \log|\mathbf{B}_1(\lambda_1)| \\ &- \tfrac{1}{2\sigma_v^2}[\mathbf{Z}(\beta, \rho, \lambda_1, \lambda_2) - \mathbf{D}\phi]'\mathbf{\Omega}^{-1}(\lambda_3)[\mathbf{Z}(\beta, \rho, \lambda_1, \lambda_2) - \mathbf{D}\phi],\end{aligned} \tag{2.3}$$

where $\mathbf{Z}(\beta, \rho, \lambda_1, \lambda_2) = \mathbf{B}_1(\lambda_1)\mathbf{Y} - \mathbf{B}_2(\rho, \lambda_2)\mathbf{Y}_{-1} - \mathbf{X}\beta$, $\mathbf{\Omega}(\lambda_3) = \texttt{bdiag}(\Omega_1, \cdots, \Omega_T)$, and $\Omega_t(\lambda_3) = \sigma_v^{-2}\mathrm{E}(u_t u_t') = [B_{3,t}'(\lambda_3)B_{3,t}(\lambda_3)]^{-1}$. Solving the first order condition, $\frac{\partial}{\partial\phi}\ell_{nT}(\psi, \phi) = 0$, yields the constrained conditional quasi maximum likelihood estimator of ϕ as a function of $\theta = (\beta', \rho, \lambda_1, \lambda_2, \lambda_3)$:

$$\tilde{\phi}(\theta) = [\mathbf{D}'\mathbf{\Omega}^{-1}(\lambda_3)\mathbf{D}]^{-1}\mathbf{D}'\mathbf{\Omega}^{-1}(\lambda_3)\mathbf{Z}(\beta, \rho, \lambda_1, \lambda_2). \tag{2.4}$$

Plugging $\tilde{\phi}(\theta)$ into $\ell_{nT}(\psi, \phi)$ yields the conditional concentrated quasi loglikelihood (CCQL) function of ψ:

$$\begin{aligned}\ell_{nT}^c(\psi) = &-\tfrac{nT}{2}\log(2\pi\sigma_v^2) + \log|\mathbf{B}_3(\lambda_3)| + \log|\mathbf{B}_1(\lambda_1)| \\ &- \tfrac{1}{2\sigma_v^2}\mathbf{Z}^{*\prime}(\theta)\mathbf{M}(\lambda_3)\mathbf{Z}^*(\theta),\end{aligned} \tag{2.5}$$

where $\mathbf{Z}^*(\theta) = \mathbf{B}_3(\lambda_3)\mathbf{Z}(\beta, \rho, \lambda_1, \lambda_2)$ and $\mathbf{M}(\lambda_3) = I_{nT} - \mathbf{B}_3(\lambda_3)\mathbf{D}[\mathbf{D}'\mathbf{\Omega}^{-1}(\lambda_3)\mathbf{D}]^{-1}\mathbf{D}'\mathbf{B}_3'(\lambda_3)$. Maximizing CCQL $\ell_{nT}^c(\psi)$ gives the conditional quasi maximum likelihood (CQML) estimator $\hat{\psi}_{\texttt{CQML}}$ of ψ. However, this estimator is affected by two well-known sources of bias: (i) conditioning on the initial observations y_0, commonly referred to as the *initial-value problem* (Hsiao et al., 2002), and (ii) the concentration/estimation of FE parameters ϕ, which gives rise to the *incidental parameters problem* (Neyman and Scott, 1948). As a result, $\hat{\psi}_{\texttt{CQML}}$ is inconsistent or asymptotically biased.

To clarify, note that under mild conditions, maximizing (2.5) is equivalent to solving $S_{nT}^c(\psi) = 0$, where $S_{nT}^c(\psi) = \frac{\partial}{\partial\psi}\ell_{nT}^c(\psi)$ denotes the concentrated conditional quasi-score (CCQS), given explicitly in (B.1) in Appendix B. A necessary condition for the consistency of $\hat{\psi}_{\texttt{CQML}}$ is $\mathrm{plim}_{nT\to\infty}\frac{1}{nT}S_{nT}^c(\psi_0) = 0$. This condition fails when T is fixed, implying that $\hat{\psi}_{\texttt{CQML}}$ cannot be consistent in this case. When T goes large with n, the above condition holds and consistency can be achieved. Nevertheless, the limiting distribution of $\sqrt{nT}(\hat{\psi}_{\texttt{CQML}} - \psi_0)$ has a non-zero mean $\lim_{nT\to\infty}[-\frac{1}{nT}\mathrm{E}(\frac{\partial}{\partial\psi'}S_{nT}^c(\psi_0))]^{-1}\frac{1}{\sqrt{nT}}\mathrm{E}[S_{nT}^c(\psi_0)]$, giving rise to the so-called *asymptotic bias*. While this bias can be removed by directly bias-correcting $\hat{\psi}_{\texttt{CQML}}$ as in Yu et al. (2008), valid inference under that approach requires

the additional sample size condition: $\frac{T}{n^{1/3}} \to \infty$.

In general, when inference concerns a fixed-dimensional parameter vector in the presence of nuisance parameters whose dimension grows with the sample size, three main bias-correction strategies are available: (i) bias correction of the CCQL function, (ii) bias correction of the CQML estimator, and (iii) bias correction of the score functions. For a comprehensive survey in non-spatial settings, see Arellano and Hahn (2007). Among these three options, method (iii) is the most practical for our setting, as the adjustments involve only common parameters and spatial-weight matrices, rather than unknown initial values or fixed effects, making implementation straightforward. Moreover, these corrections directly remove both the initial-value and incidental-parameter biases in the estimating functions and, as shown in Section 3, can be extended to unknown heteroskedasticity. By contrast, method (i) does not appear to have a straightforward implementation for our concentrated DSPD likelihood, while method (ii) provides an estimator-level bias correction and does not appear to admit a simple adaptation to unknown heteroskedasticity. In this paper, we adopt the score-correction approach, as in Yang (2018) and Li and Yang (2020), by explicitly deriving $\mathrm{E}[S_{nT}^c(\psi_0)]$ and using it to recenter the CCQS functions. This leads to unbiased estimating equations and thus a consistent and asymptotically unbiased estimator of ψ, without imposing any constraint on n and T. See also Dhaene and Jochmans (2016) for a related approach in autoregressive panel models.

To derive $\mathrm{E}[S_{nT}^c(\psi_0)]$, we first show, using (B.2) and (B.3) in Appendix B, that the stochastic component of the CCQS can be expressed as linear combinations of three types of terms, $\Pi'\mathbf{V}$, $\mathbf{V}'\Phi\mathbf{V}$, and $\mathbf{V}'\Psi\mathbf{Y}_0$, corresponding, respectively, to linear terms in $\mathbf{V}$, quadratic terms in $\mathbf{V}$, and bilinear terms involving $\mathbf{V}$ and $\mathbf{Y}_0 \equiv 1_T \otimes y_0$. Here, Φ and Ψ are non-stochastic matrices that depend on parameters ψ_0, the spatial weight matrices $\mathbf{W}_r$ (for $r = 1, 2, 3$), and $\mathbf{D}$, while Π additionally depends on $\mathbf{X}$ and ϕ_0. Precise definitions of these quantities are provided below (B.3). Under the standard assumption that y_0 is independent of $\{v_t, t \geq 1\}$ and the time-varying regressors $\{x_t, t = 0, 1, \ldots, T\}$, the terms linear in $\mathbf{V}$ and bilinear in $(\mathbf{V}, \mathbf{Y}_0)$ have zero expectations. As a result, $\mathrm{E}[S_{nT}^c(\psi_0)]$ in (B.4) depends only on the expectations of the quadratic terms and the nonstochastic components in $S_{nT}^c(\psi_0)$. Since $\mathrm{E}(\mathbf{V}'\Phi\mathbf{V}) = \sigma_{v0}^2 \mathtt{tr}(\Phi)$ and the nonstochastic terms in $S_{nT}^c(\psi_0)$ involve only common parameters and $\mathbf{W}_r$ $(r = 1, 3)$, the expression of $\mathrm{E}[S_{nT}^c(\psi_0)]$ is free from both the initial conditions and the fixed effects. It depends on the matrices:

$\Phi_\rho = \mathbf{M}\mathbf{B}_3\mathbf{Q}_{-1}\mathbf{B}_3^{-1}$, $\Phi_{\lambda_1} = \mathbf{M}\mathbf{B}_3\mathbf{W}_1\mathbf{Q}\mathbf{B}_3^{-1}$, $\Phi_{\lambda_2} = \mathbf{M}\mathbf{B}_3\mathbf{W}_2\mathbf{Q}_{-1}\mathbf{B}_3^{-1}$, $\Phi_{\lambda_3} = \mathbf{M}\mathbf{W}_3\mathbf{B}_3^{-1}\mathbf{M}$, where the $nT \times nT$ block matrices $\mathbf{Q}$ and $\mathbf{Q}_{-1}$ are defined in terms of $n \times n$ blocks. Let $\mathcal{B}_t = B_{1t}^{-1}B_{2t}$. For $\mathbf{Q}$, the (t,s)-block equals $(\mathcal{B}_t \cdots \mathcal{B}_{s+1})B_{1s}^{-1}$ for $t > s$, B_{1s}^{-1} for $t = s$, and zero otherwise. For $\mathbf{Q}_{-1}$, the (t,s)-block equals $(\mathcal{B}_{t-1} \cdots \mathcal{B}_{s+1})B_{1s}^{-1}$ for $t > s+1$, B_{1s}^{-1} when $t = s+1$, and zero otherwise. This observation motivates re-centering the CCQS by subtracting these expectations, as $S_{nT}^*(\psi_0) = S_{nT}^c(\psi_0) - \mathrm{E}[S_{nT}^c(\psi_0)]$. This leads to the bias-corrected estimating functions, which take the form at the general parameter ψ:

$$S_{nT}^*(\psi) = \begin{cases} \frac{1}{\sigma_v^2}\mathbf{X}'\mathbf{B}_3'(\lambda_3)\mathbf{M}(\lambda_3)\mathbf{Z}^*(\theta), \\ \frac{1}{\sigma_v^2}\mathbf{Y}_{-1}'\mathbf{B}_3'(\lambda_3)\mathbf{M}(\lambda_3)\mathbf{Z}^*(\theta) - \mathtt{tr}[\Phi_\rho(\delta)], \\ \frac{1}{\sigma_v^2}\mathbf{Y}'\mathbf{W}_1'\mathbf{B}_3'(\lambda_3)\mathbf{M}(\lambda_3)\mathbf{Z}^*(\theta) - \mathtt{tr}[\Phi_{\lambda_1}(\delta)], \\ \frac{1}{\sigma_v^2}\mathbf{Y}_{-1}'\mathbf{W}_2'\mathbf{B}_3'(\lambda_3)\mathbf{M}(\lambda_3)\mathbf{Z}^*(\theta) - \mathtt{tr}[\Phi_{\lambda_2}(\delta)], \\ \frac{1}{\sigma_v^2}\mathbf{Z}^{*\prime}(\theta)\mathbf{M}(\lambda_3)\mathbf{W}_3\mathbf{B}_3^{-1}(\lambda_3)\mathbf{M}(\lambda_3)\mathbf{Z}^*(\theta) - \mathtt{tr}[\Phi_{\lambda_3}(\lambda_3)], \\ \frac{1}{2\sigma_v^4}\mathbf{Z}^{*\prime}(\theta)\mathbf{M}(\lambda_3)\mathbf{Z}^*(\theta) - \frac{N}{2\sigma_v^2}, \end{cases} \tag{2.6}$$

where $\delta = (\rho, \lambda')'$ collects parameters that appear non-linearly in the estimating functions, and $N = nT - (n + T - 1)$ is the effective sample size after taking into account the estimation of fixed effects. The matrices $\Phi_\rho(\delta)$, $\Phi_{\lambda_1}(\delta)$, $\Phi_{\lambda_2}(\delta)$, and $\Phi_{\lambda_3}(\lambda_3)$ are evaluated at the general parameter values, using the same definitions as above. Detailed derivations of (2.6) are provided in Appendix B, following equation (B.2). By construction, $\mathrm{E}[S_{nT}^*(\psi_0)] = 0$, and it can be shown that $\mathrm{plim}_{nT\to\infty}\frac{1}{nT}S_{nT}^*(\psi_0) = 0$, paving the way for consistent and asymptotically unbiased estimation of ψ. Our M-estimator, $\hat{\psi}_{\mathtt{M}}$, is then defined as the solution to $S_{nT}^*(\psi) = 0$. Under regularity conditions, the asymptotic properties of $\hat{\psi}_{\mathtt{M}}$ are as follows.

Theorem 2.1. *Suppose Assumptions A - H in Appendix C hold. Assume further that* (i) $\gamma_{\max}[\mathrm{Var}(\mathbf{Y})]$ *and* $\gamma_{\max}[\mathrm{Var}(\mathbf{Y}_{-1})]$ *are bounded and* (ii) $\inf_{\delta\in\mathbf{\Delta}}\gamma_{\min}\big[\mathrm{Var}(\mathbf{B}_1(\lambda_1)\mathbf{Y} - \mathbf{B}_2(\rho,\lambda_2)\mathbf{Y}_{-1})\big] \geq \underline{c}_y > 0$. *We have as* $nT \to \infty$,

(i) $\hat{\psi}_{\mathtt{M}} \xrightarrow{p} \psi_0$,

(ii) $\sqrt{N}\left(\hat{\psi}_{\mathtt{M}} - \psi_0\right) \xrightarrow{D} N\left(0, \lim_{nT\to\infty}\Gamma_{nT}^{-1}(\psi_0)\Sigma_{nT}(\psi_0)\Gamma_{nT}'^{-1}(\psi_0)\right)$,

where $\Gamma_{nT}(\psi_0) = -\frac{1}{N}\mathrm{E}[\frac{\partial}{\partial\psi'}S_{nT}^*(\psi_0)]$ *and* $\Sigma_{nT}(\psi_0) = \frac{1}{N}\mathrm{Var}[S_{nT}^*(\psi_0)]$, *both assumed to exist and* $\Gamma_{nT}(\psi_0)$ *to be positive definite, for sufficiently large* nT.

Theorem 2.1 presents the large-sample foundations for inference in the DSPD model. However, its empirical implementation requires consistent estimators of $\Gamma_{nT}(\psi_0)$ and $\Sigma_{nT}(\psi_0)$. From the analytical expression for the "Hessian" matrix $\frac{\partial}{\partial\psi'}S^*_{nT}(\psi)$, provided in Appendix B, we see that it depends only on the common parameters. Thus, a consistent estimator of $\Gamma_{nT}(\psi_0)$ is given by its sample analogue, $\widehat{\Gamma}_{nT} = -\frac{1}{N}\frac{\partial}{\partial\psi'}S^*_{nT}(\hat{\psi}_{\mathtt{M}})$.

Complications arise in estimating the variance–covariance (VC) matrix of the estimating functions, $\Sigma_{nT}(\psi_0)$. As shown in (B.5) in Appendix B, $S^*_{nT}(\psi_0)$ can be expressed as linear combinations of three types of terms, namely $\Pi'\mathbf{V}$, $\mathbf{V}'\Psi\mathbf{Y}_0$, and $\mathbf{V}'\Phi\mathbf{V}$, where these conformable matrices Π, Ψ, and Φ are defined in (B.5). The elements of $\Sigma_{nT}(\psi_0)$ therefore depend on the variances and covariances of these three types of terms. The complete analytical expression for $\Sigma_{nT}(\psi_0)$ is provided in Appendix B. From this expression, we see that the entries of $\Sigma_{nT}(\psi_0)$ involve, in addition to known matrices and common parameters: (i) expectations of linear-quadratic (LQ) forms of the initial condition y_0; (ii) third and fourth moments of v_{it}; and (iii) LQ forms in FE parameters ϕ_0.

For component (i), we employ the sample analogue approach by directly removing the expectation sign in the expectations of the LQ forms in the initial values y_0. This simple idea not only addresses the initial value issue faced by the large-T approach of Yu et al. (2008) but also avoids the complications of the small-T approach of Yang (2018). For the estimation of the third and fourth moments in (ii), we adopt the approach of Meng and Yang (2025b). To estimate the quantities in (iii), we adopt a *corrected plug-in* approach. Since the matrix Π depends on the incidental parameter ϕ_0, for which a consistent estimator is unavailable under either fixed-T or fixed-n asymptotics, a naive plug-in may introduce bias. We propose to first substitute the M-estimator $\hat{\phi}_{\mathtt{M}} = \tilde{\phi}(\hat{\theta}_{\mathtt{M}})$ through (2.4) into the expressions involving Π, and then analytically derive the bias induced by this substitution. We find that terms linear in ϕ_0 do not generate bias under this plug-in, whereas terms that are quadratic in ϕ_0 do.

Let the $(k+5)\times(k+5)$ matrix $\Upsilon(\boldsymbol{\psi}_0)$ denote the resulting plug-in bias, whose non-zero elements are given by: $\Upsilon_{\omega,\varpi}(\psi_0) = \frac{1}{N}\mathtt{tr}(\Phi'_\omega\Phi_\varpi\mathbf{P})$, $\omega,\varpi = \rho,\lambda_1,\lambda_2$ where Φ_ρ, Φ_{λ_1}, and Φ_{λ_2} are defined below (B.3), and $\mathbf{P} = \mathbf{B}_3\mathbf{D}(\mathbf{D}'\boldsymbol{\Omega}^{-1}\mathbf{D})^{-1}\mathbf{D}'\mathbf{B}'_3$ is the projection matrix onto the column space of $\mathbf{B}_3\mathbf{D}$. Importantly, the bias terms depend only on common parameters and can therefore be consistently estimated. Moreover, the nonzero elements of $\Upsilon(\psi_0)$ are of order $O(n^{-1}) + O(T^{-1})$[1], so this bias vanishes when both n and T diverge, but can

remain non-negligible when one dimension is fixed. To correct for this bias, we subtract an estimate of $\Upsilon(\psi_0)$ from the plug-in estimator. Formally, letting $\Sigma_{nT}(\hat{\psi})$ denote the simple plug-in estimator, a consistent estimator of $\Sigma_{nT}(\psi_0)$ is given by: $\hat{\Sigma}_{nT} = \Sigma_{nT}(\hat{\psi}) - \Upsilon(\hat{\psi})$.

The consistency of this proposed estimator for the VC matrix of the M-estimator is established in the following theorem.

Theorem 2.2. *Under the assumptions of Theorem 2.1, as* $nT \to \infty$,

$$\widehat{\Gamma}_{nT} - \Gamma_{nT}(\psi_0) \xrightarrow{p} 0 \text{ and } \widehat{\Sigma}_{nT} - \Sigma_{nT}(\psi_0) \xrightarrow{p} 0,$$

and hence $\widehat{\Gamma}_{nT}^{-1}\widehat{\Sigma}_{nT}\widehat{\Gamma}_{nT}^{-1} - \Gamma_{nT}^{-1}(\psi_0)\Sigma_{nT}(\psi_0)\Gamma_{nT}^{-1}(\psi_0) \xrightarrow{p} 0$.

3. M-Estimation under Heteroskedasticity

Now suppose the idiosyncratic errors v_{it} are heteroskedastic in **both** the cross-sectional and time dimensions, such that $\text{Var}(v_{it}) = \sigma_{it}^2$. Let the $nT \times 1$ vector $\boldsymbol{\sigma}_{nT}$ collect all the error variances: $\boldsymbol{\sigma}_{nT} = (\sigma_{n,1}', \ldots, \sigma_{n,T}')'$, where $\sigma_{n,t} = (\sigma_{1t}^2, \ldots, \sigma_{nt}^2)'$.Under this setup, the expectations of the quadratic terms in (B.3) become:

$$\text{E}(\mathbf{V}'\Phi_r\mathbf{V}) = \texttt{tr}(\mathbf{H}\Phi_r), \quad r = \rho, \lambda_1, \lambda_2, \lambda_3, \tag{3.1}$$

where $\mathbf{H} \equiv \texttt{diag}(\boldsymbol{\sigma}_{nT}')$. These expectations now depend not only on the matrices Φ_r but also on the unknown heteroskedasticity matrix $\mathbf{H}$. As a result, the direct re-centering approach used to derive (2.6) is no longer valid under unknown heteroskedasticity.

To address this, we follow the idea in Li and Yang (2020) and construct alternative quadratic forms in $\mathbf{Z}^*$, whose expectations align with those in (3.1):

$$\text{E}(\mathbf{Z}^{*\prime}\mathbf{M}\mathbf{R}_r\mathbf{Z}^*) = \texttt{tr}(\mathbf{H}\Phi_r), \quad r = \rho, \lambda_1, \lambda_2, \lambda_3, \tag{3.2}$$

where $\mathbf{R}_r = \texttt{diag}(\Phi_r)\texttt{diag}(\mathbf{M})^{-1}, r = \rho, \lambda_1, \lambda_2, \lambda_3$. We can use this set of quadratic terms to recenter the CCQS and derive a set of unbiased estimating functions for ρ and

λ, robust against unknown $\mathbf{H}$:

$$S_{nT}^r(\theta) = \begin{cases} \mathbf{X}'\mathbf{B}_3'(\lambda_3)\mathbf{M}(\lambda_3)\mathbf{Z}^*(\theta), \\ \mathbf{Y}_{-1}'\mathbf{B}_3'(\lambda_3)\mathbf{M}(\lambda_3)\mathbf{Z}^*(\theta) - \mathbf{Z}^{*\prime}(\theta)\mathbf{M}(\lambda_3)\mathbf{R}_\rho(\delta)\mathbf{Z}^*(\theta), \\ \mathbf{Y}'\mathbf{W}_1'\mathbf{B}_3'(\lambda_3)\mathbf{M}(\lambda_3)\mathbf{Z}^*(\theta) - \mathbf{Z}^{*\prime}(\theta)\mathbf{M}(\lambda_3)\mathbf{R}_{\lambda_1}(\delta)\mathbf{Z}^*(\theta), \\ \mathbf{Y}_{-1}'\mathbf{W}_2'\mathbf{B}_3'(\lambda_3)\mathbf{M}(\lambda_3)\mathbf{Z}^*(\theta) - \mathbf{Z}^{*\prime}(\theta)\mathbf{M}(\lambda_3)\mathbf{R}_{\lambda_2}(\delta)\mathbf{Z}^*(\theta), \\ \mathbf{Z}^{*\prime}(\theta)[\mathbf{M}(\lambda_3)\mathbf{W}_3\mathbf{B}_3^{-1}(\lambda_3)\mathbf{M}(\lambda_3) - \mathbf{M}(\lambda_3)\mathbf{R}_{\lambda_3}(\delta)]\mathbf{Z}^*(\theta). \end{cases} \tag{3.3}$$

Note that σ_v^2 is removed from the estimating functions, as it is no longer relevant under unknown heteroskedasticity. By construction, we have $\mathrm{E}[S_{nT}^r(\theta_0)] = 0$, and it can further be shown that $\mathrm{plim}_{nT\to\infty}\frac{1}{N}S_{nT}^r(\theta_0) = 0$. Consequently, solving $S_{nT}^r(\theta) = 0$ yields consistent M-estimator, denoted by $\hat{\theta}_{\texttt{RM}}$, for θ, which is robust to unknown heteroskedasticity $\mathbf{H}$. In addition, as shown in (B.6) in Appendix B, the estimating function $S_{nT}^r(\theta_0)$ can be expressed as linear combinations of terms that are linear and quadratic in $\mathbf{V}$, as well as bilinear in $\mathbf{V}$ and $\mathbf{Y}_0$. As a result, the remainder of the M-estimation procedure parallels that of the homoskedastic case. Consistency and asymptotic normality follow by analogous arguments, as stated in the following corollary.

Corollary 3.1. *Suppose Assumptions A*, B-G and H* in Appendix C hold. Assume further (i)* $\gamma_{\max}[\mathrm{Var}(\mathbf{Y})]$ *and* $\gamma_{\max}[\mathrm{Var}(\mathbf{Y}_{-1})]$ *are bounded and (ii)* $\inf_{\delta\in\mathbf{\Delta}}\gamma_{\min}\big[\mathrm{Var}(\mathbf{B}_1\mathbf{Y} - \mathbf{B}_2\mathbf{Y}_{-1})\big] \geq \underline{c}_y > 0$. *We have as* $nT\to\infty$,

(*i*) $\hat{\theta}_{\texttt{RM}} \xrightarrow{p} \theta_0$,

(*ii*) $\sqrt{N}\left(\hat{\theta}_{\texttt{RM}} - \theta_0\right) \xrightarrow{D} N\left(0, \lim_{nT\to\infty}\Gamma^{r}{}_{nT}^{-1}(\theta_0)\Sigma_{nT}^r(\theta_0)\Gamma^{r\prime}{}_{nT}^{-1}(\psi_0)\right)$,

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

To conduct inference on θ, we require consistent estimators of $\Gamma_{nT}^r(\theta_0)$ and $\Sigma_{nT}^r(\theta_0)$, whose analytical expressions are provided in Appendix B. As before, $\Gamma_{nT}^r(\theta_0)$ can be consistently estimated using a plug-in approach, $\widehat{\Gamma}_{nT}^r = -\frac{1}{N}\frac{\partial}{\partial\theta'}S_{nT}^r(\hat{\theta}_{\texttt{RM}})$, since it depends only on common parameters that have already been consistently estimated.

Compared with the homoskedastic case, the elements of $\Sigma_{nT}^r(\theta_0)$ now depend not only on the initial condition y_0 and fixed-effects parameters ϕ_0, but also on the heteroskedasticity matrix $\mathbf{H}$. This substantially complicates the estimation of $\Sigma_{nT}^r(\theta_0)$, as the dimensions

of both ϕ and $\mathbf{H}$ grow with the sample size, leading to a more severe incidental-parameters problem. Another key feature of the analytical expression for $\Sigma^r_{nT}(\theta_0)$ is that it does not involve third- or fourth moments of the errors. This arises because the middle matrices in the quadratic terms $\mathbf{V}'\Phi^r_\omega\mathbf{V}$ for $\omega = \rho, \lambda_1, \lambda_2, \lambda_3$ in (B.6) of Appendix B, have zero diagonals by construction through (3.2). The expectation terms associated with the initial condition y_0 are likewise estimated using the sample analogue by directly plugging in the observed initial values y_0. To address the complications arising from ϕ_0 and $\mathbf{H}$, we propose a *corrected plug-in* approach to estimate $\Sigma^r_{nT}(\theta_0)$.

To make this dependence explicit, we write $\Sigma^r_{nT}(\theta_0)$ as $\Sigma^r_{nT}(\theta_0, \phi_0, \mathbf{H})$. Let $\hat{\phi}_{\texttt{RM}} = \tilde{\phi}(\hat{\theta}_{\texttt{RM}})$ be the RM-estimator for ϕ_0. For a given $\mathbf{H}$, a natural plug-in estimator is then $\Sigma^r_{nT}(\hat{\theta}_{\texttt{RM}}, \hat{\phi}_{\texttt{RM}}, \mathbf{H})$. However, as in the homoskedastic case, such a plug-in introduces non-negligible bias in terms that are quadratic in ϕ_0. In particular, the affected components are $\Sigma^r_{\omega\varpi}(\theta_0, \phi_0, \mathbf{H})$, for $\omega, \varpi = \rho, \lambda_1, \lambda_2, \lambda_3$, with bias terms taking the form $\frac{1}{N}\texttt{tr}(\mathbf{H}\mathbf{P}\Phi^{r\prime}_\omega\mathbf{H}\Phi^r_\varpi\mathbf{P})$, respectively. Regarding the estimation of $\mathbf{H}$, we note that $\mathrm{E}(\mathbf{M}\mathbf{Z}^* \odot \mathbf{M}\mathbf{Z}^*) = (\mathbf{M} \odot \mathbf{M})\boldsymbol{\sigma}_{nT}$, where $\odot$ denotes the Hadamard product. An estimator of $\mathbf{H}$ can therefore be constructed as $\hat{\mathbf{H}} = \texttt{diag}(\hat{\boldsymbol{\sigma}}_{nT})$, with

$$\hat{\boldsymbol{\sigma}}_{nT} = [\mathbf{M}(\hat{\lambda}_3) \odot \mathbf{M}(\hat{\lambda}_3)]^-[\mathbf{M}(\hat{\lambda}_3)\mathbf{Z}^*(\hat{\theta}) \odot \mathbf{M}(\hat{\lambda}_3)\mathbf{Z}^*(\hat{\theta})], \tag{3.4}$$

where $(\cdot)^-$ denotes a generalized inverse. However, directly substituting $\hat{\mathbf{H}}$ may introduce additional bias. As shown in Appendix B, the elements of $\Sigma^r_{nT}(\theta_0, \phi_0, \mathbf{H})$ take two generic forms: $\texttt{tr}(\mathbf{H}C_{nT})$ or $\texttt{tr}(\mathbf{H}A_{nT}\mathbf{H}B_{nT})$. In addition, the bias term caused by plug-in of $\hat{\phi}_{\texttt{RM}}$ is also of the second form. Therefore, it is important to assess the impact of replacing $\mathbf{H}$ with $\hat{\mathbf{H}}$ in these expressions. We show that such substitution has a non-negligible effect only for terms of the second form.

Let the $(k+4) \times (k+4)$ matrix $\Upsilon^r(\theta_0, \mathbf{H})$ denote the bias in the plug-in estimator of $\Sigma^r_{nT}(\theta_0, \phi_0, \mathbf{H})$, jointly accounting for the effects of substituting $\hat{\phi}$ and $\hat{\mathbf{H}}$. The nonzero components of $\Upsilon^r(\theta_0, \mathbf{H})$ are given by:

$$\Upsilon^r_{\omega,\varpi}(\theta_0, \mathbf{H}) = \tfrac{1}{N}\texttt{tr}(\mathbf{H}\mathbf{P}\Phi^{r\prime}_\omega\mathbf{H}\Phi^r_\varpi\mathbf{P}) + \tfrac{2}{N}\texttt{tr}(\mathbf{K}_{\omega,\varpi}\Xi), \quad \omega, \varpi = \rho, \lambda_1, \lambda_2, \lambda_3 \tag{3.5}$$

where $\mathbf{K}_{\omega,\varpi} = \Phi^r_\omega \odot (\Phi^r_\varpi + \Phi^{r\prime}_\varpi) - \mathbf{P}\Phi^{r\prime}_\omega \odot \Phi^r_\varpi\mathbf{P}$, $\Xi = (\mathbf{M} \odot \mathbf{M})^-[(\mathbf{M}\mathbf{H}\mathbf{M}) \odot (\mathbf{M}\mathbf{H}\mathbf{M})](\mathbf{M} \odot \mathbf{M})^-$. The first term in (3.5) captures the bias induced by substituting the fixed-effects estimator $\hat{\phi}_{\texttt{M}}$, which is typically of order $O(n^{-1}) + O(T^{-1})$. The second term reflects the bias arising from substituting $\hat{\mathbf{H}}$ (both in the original terms of $\Sigma^r_{nT}(\theta_0, \phi_0, \mathbf{H})$ and

in the first term of (3.5)) and is of order $O(1)$. Consequently, under heteroskedasticity, the plug-in estimator of $\Sigma^r_{nT}(\theta_0, \phi_0, \mathbf{H})$ is inconsistent even when both n and T diverge, and bias correction is therefore essential. Thus, a consistent estimator of $\Sigma^r_{nT}(\theta_0, \phi_0, \mathbf{H})$ is given by $\hat{\Sigma}^r_{nT} = \Sigma^r_{nT}(\hat{\theta}_{\texttt{RM}}, \hat{\phi}_{\texttt{RM}}, \hat{\mathbf{H}}) - \Upsilon^r(\hat{\theta}, \hat{\mathbf{H}})$. The consistency of the estimator of the VC matrix of the robust M-estimator is established in the following corollary.

Corollary 3.2. *Under the assumptions of Corollary 3.1, as* $nT \to \infty$,

$$\widehat{\Gamma}^r_{nT} - \Gamma^r_{nT}(\theta_0) \xrightarrow{p} 0 \text{ and } \widehat{\Sigma}^r_{nT} - \Sigma^r_{nT}(\theta_0) \xrightarrow{p} 0,$$

and hence $\widehat{\Gamma}^{r-1}_{nT}\widehat{\Sigma}^r_{nT}\widehat{\Gamma}^{r\prime-1}_{nT} - \Gamma^{r-1}_{nT}(\theta_0)\Sigma^r_{nT}(\theta_0)\Gamma^{r-1}_{nT}(\theta_0) \xrightarrow{p} 0$.

4. Monte Carlo Results

The Monte Carlo experiments in this section are designed to assess the finite sample performance of the proposed M-estimator and robust M-estimator, and their corresponding standard error estimators. The data generating process (DGP) used in these simulations is specified as follows:

$$\begin{aligned} y_t &= \rho y_{t-1} + \lambda_1 W_t y_t + \lambda_2 W_{t-1} y_{t-1} + x_t\beta + \mu + \alpha_t 1_n + u_t, \\ u_t &= \lambda_3 W_t u_t + v_t, \quad t = 1, 2, \ldots, T. \end{aligned} \tag{4.1}$$

where the elements of X_t are generated as $X_t = \mu_x + 0.01 \times t \times 1_n + \zeta_t$, where $(1 - 0.5L)\zeta_t = \varepsilon_t + 0.5\varepsilon_{t-1}$, $\varepsilon_t \sim N(0, I_n)$, $\mu_x = e_x + \frac{1}{T+m+1}\sum_{t=-m}^{T}\varepsilon_t$, and $e_x \sim N(0, 2^2 I_n)$. The two-way fixed effects are generated as follows: $\mu_i = \frac{1}{T}\Sigma_{t=1}^{T} X_{i,t} + e_i$, where $e_i \sim N(0,1)$, and $\alpha_t \sim N(0,1)$. The initial observation vector y_0 is obtained from a six-period pre-sample simulation. Starting from $\mu + \epsilon$, where $\epsilon \sim N(0, I_n)$, the same spatial-dynamic DGP is recursively applied, and the final pre-sample realization is taken as y_0. To cover both small and large T cases, we consider different combinations of n and T, including $n \in (50, 100, 200)$ and $T \in (5, 10, 20, 40)$. The distribution of the idiosyncratic errors $\{v_{it}\}$ can take one of two forms: (1) `normal` and (2) standardized `normal mixture` (10% $N(0, 4^2)$ and 90% $N(0,1)$). See Yang (2015) for details. The spatial weight matrices for the spatial lag effect, space–time effect, and spatial error effect are set to be identical, as described below. We set $\rho = 0.3$, $\beta = 1$, $\lambda_1 = 0.2$, $\lambda_2 = 0.2$ and $\lambda_3 = 0.2$. The number of Monte Carlo runs is 1000. Appendix F reports additional simulation results, including experiments with larger samples (up to $(n, T) = (200, 40)$), stronger spatial dependence,

the spatial-cointegration design of Yu et al. (2012), and comparisons with the estimators of Lee and Yu (2014), Yang (2018), and Li and Yang (2020) under time-invariant spatial weight matrices. Appendix G provides a computational note on implementing the proposed methods for large panels.

We consider both homoskedastic and heteroskedastic error structures. Under homoskedasticity, the error variance is fixed at $\sigma_v^2 = 1$, while under heteroskedasticity, the average error variance is normalized to 1. The spatial weight matrices $\{W_t\}$ are generated from time-varying group interaction structures, with different group-formation mechanisms under homoskedasticity and heteroskedasticity. Under homoskedasticity, the n units are randomly repartitioned in each period into $K(n) = \text{Round}(n^{0.5})$ non-overlapping groups of heterogeneous sizes; we refer to this design as Group Interaction I (GI-I). Under heteroskedasticity, six group sizes are randomly generated in each period, subject to the constraint that they sum to 50. This group-size configuration is then replicated as needed to match the cross-sectional dimension n; we refer to this design as Group Interaction II (GI-II). In the heteroskedasticity case, the group-size variation does not shrink as n increases, which results in the M-estimation being inconsistent under heteroskedasticity, as shown in Liu and Yang (2015, 2020). Heteroskedasticity is generated by scaling the variance of each group according to its size as in Lin and Lee (2010): if the size of a group exceeds the mean group size, its variance is set equal to the group size; otherwise, the variance is set equal to the squared reciprocal of the group size.

Monte Carlo (empirical) bias and standard deviations (sd, in parentheses) are presented for the `CQMLE` and our proposed M-estimator (`M-Est`) and RM-estimator (`RM-Est`). The (empirical) coverage probabilities of 95% confidence intervals (CP95, in square brackets) are also reported for `M-Est` and `RM-Est` based on the results of Sections 2 and 3.

Tables 1 and 2 jointly report the Monte Carlo results for various sample sizes under both homoskedasticity and heteroskedasticity. Under homoskedasticity, `M-Est` and `RM-Est` exhibit very similar and excellent finite-sample performance and uniformly outperform the `CQMLE`, particularly in the estimation of the spatial and variance parameters, regardless of sample size or error distribution. Moreover, the coverage probabilities of confidence intervals based on both `M-Est` and `RM-Est` are close to the nominal level when the errors are homoskedastic, which validates the theoretical results. Under heteroskedasticity, only `RM-Est` remains valid, and the results confirm its excellent finite-sample per-

formance in terms of both point estimation and 95% confidence interval estimation. By contrast, the `CQMLE` and `M-Est` generally deliver severely distorted estimates for the spatial parameters, and their inconsistency persists even as the sample size increases, clearly demonstrating their failure in the presence of unknown heteroskedasticity.

5. Empirical Application

To illustrate the proposed alternative approach, we apply it to corporate data and study how firm policies evolve over time and respond to peer behavior. We estimate and conduct inference under unknown heteroskedasticity and two-way fixed effects to study how firm policies evolve over time and respond to peers. Our sample includes 1,179 firms listed on Chinese stock exchanges from 2015 to 2018. We focus on firm outcomes where both dynamics and peer effects are likely to play an important role, including investment, financing, and performance.

A large theoretical and empirical literature shows that firms' decisions are interdependent rather than isolated, as they respond to fundamentals, aggregate conditions, and peer behavior. Strategic complementarities or substitutabilities may arise through product-market competition, financial frictions, information diffusion, managerial incentives, and benchmarking (Bloom et al., 2013; Leary and Roberts, 2014; Bustamante and Frésard, 2021). Firms therefore internalize both their own trade-offs and the anticipated responses of peers, resulting equilibrium outcomes shaped by network interactions (Jaffe et al., 1993; Dougal et al., 2015; Grieser et al., 2022a).

Using a static spatial panel model with time fixed effects, Grieser et al. (2022a) document economically meaningful strategic interaction in capital structure decisions, and Grieser et al. (2022b) show similarly important peer effects in investment, financial policies, and performance. These interactions are unlikely to be purely contemporaneous. Adjustment frictions, target adjustment, habit formation, and learning imply dependence on firms' own past choices and peers' lagged decisions, motivating a more general dynamic framework capable of capturing both policy persistence and propagation of peer effects over time. Moreover, persistent unobserved firm traits, such as managerial quality or corporate culture, can systematically shape corporate policies, making firm fixed effects essential. Allowing for unknown heteroskedasticity accommodates time-varying risk

profiles and cross-firm volatility heterogeneity, features that are salient in the presence of macro-financial and industry shocks (e.g., uncertainty shocks in Bloom (2009)).

In addition, traditional industry networks rely on coarse classifications that treat firms within an industry as equally related and firms outside it as unrelated, a simplification that ignores heterogeneity in competitive overlap. Following Grieser et al. (2022a), we instead adopt the text-based similarity measure of Hoberg and Phillips (2016), which uses business descriptions of firms to generate continuous pairwise product-market similarity scores (ranging from 0 to 1), allowing a much more precise mapping of peer networks.

We estimate the DSPD model with two-way FEs and unknown heteroskedasticity:

$$\mathbf{Y} = \rho\mathbf{Y}_{-1} + \lambda_1\mathbf{W}_1\mathbf{Y} + \lambda_2\mathbf{W}_2\mathbf{Y}_{-1} + \mathbf{X}\beta + \mathbf{D}\phi + \mathbf{U}, \tag{5.1}$$

where spatial error dependence is excluded to focus on economically interpretable endogenous peer effects. Following Grieser and Hadlock (2019) and Grieser et al. (2022b), we analyze five outcomes, including *R&D*, *Leverage*, *ROA* (return on assets), CF (cash flow), and *Tobin's Q*, with controls for *Sales*, *Z-score*, *Age*, *Cash* (cash holdings), and *Tangibility* using CSMAR data (variable definitions in Appendix E). The time-invariant spatial weight matrices $W_1 = W_2$ are constructed from the similarity scores. For each firm, we retain the 100 closest rivals (about 8.5% of the sample) and set all other entries to zero, an economically reasonable cutoff since firms are unlikely to track every competitor. The results are robust to alternative cutoffs from 1% to 100% of firms. Relative to Grieser et al. (2022b), who keep only the top 10 rivals (0.016%) for computational reasons, this approach produces a much richer network. Similarity scores are taken from 2014 to mitigate endogenous peer selection, and the resulting matrices are row-normalized with zero diagonals, producing firm-specific competitor sets weighted by product-market similarity.

The main results are presented in Table 3. Consistent with Grieser et al. (2022b), we find positive and significant contemporaneous spillovers (λ_1) in firms' R&D, leverage and ROA. The spillover is highest for R&D, with an estimated coefficient of 0.350, indicating that innovation choices are more interdependent among product-market peers than other financial policies. We also find a significant positive spillover in ROA ($\lambda_1 = 0.268$), which implies that the profitability of a firm moves strongly with that of its peers. For leverage, the spillover is moderate ($\lambda_1 = 0.130$), similar to Grieser et al. (2022a).

In contrast, we do not find statistically significant spillover in CF and Tobin's Q. Given

our richer specification, this result is not unsurprising. Persistent firms' characteristics, correlated within peer groups, such as business models, accounting practices, or long-run profitability, are absorbed by firm fixed effects. Dynamic terms further capture serial dependence in these variables. Moreover, our robust treatment of spatiotemporal heteroskedasticity yields more reliable standard errors, reducing spurious significance that may arise under more restrictive error assumptions.

Turning to lagged peer effects, we find significant effects for R&D and leverage. The positive and relatively large lagged spillover in R&D ($\lambda_2 = 0.181$) indicates that firms increase current R&D in response to peers' past R&D. This delayed reaction may arise from learning and knowledge diffusion, which take time, or from adjustment frictions such as budgeting, approvals, and hiring that cause responses to materialize over multiple periods. Regardless of the channel, the results suggest that an R&D shock to some firms propagates through the network both immediately and with a delay, amplifying and prolonging its impact, an effect not documented in Grieser et al. (2022b).

For firms' capital structure, the negative and moderate lagged spillover in leverage ($\lambda_2 = -0.165$) suggests subsequent rebalancing or correction following an initial co-movement. This pattern is consistent with learning, whereby firms infer optimal choices by observing the decisions and outcomes of industry peers, a process that is often dynamic. As emphasized by Grieser et al. (2022a), widespread effects of industry in capital-structure policy suggest that the parameters that guide optimal choices have an industry-dependent element. Therefore, an optimal learning rule can incorporate a dynamic adjustment of firm decisions to its peers' choices.

Finally, we find significant dynamic effects in R&D and Tobin's Q. The dynamic effect of R&D is negative but very small ($\rho = -0.006$). This could reflect budget smoothing with a slight correction after a temporary spike. The contrast between weak internal persistence and strong peer spillovers indicates that persistence in R&D mainly arises at the network level rather than within firms. For Tobin's Q, the positive and moderate dynamic effect ($\rho = 0.047$) implies a gradual update of investors' beliefs about firm growth opportunities and risk. In the absence of significant peer spillovers, valuation dynamics appears to be largely driven by firm-specific information rather than network amplification. The dynamic effects for leverage, ROA, and cash flow are small and statistically insignificant. The coefficients on the control variables are broadly consistent with

previous findings in Grieser et al. (2022b).

6. Discussion

Our M-estimation and inference framework extends readily to alternative fixed-effect structures and higher-order spatial specifications, highlighting its flexibility.

Other forms of fixed-effects. Fixed effects may vary with the empirical context. Beyond individual effects, group-level effects—such as regional, sectoral, or institutional components—are often appropriate for hierarchical data (e.g., firms within industries or households within regions). In such cases, the dummy-variable matrix $\mathbf{D}$ can be redefined to include the relevant group indicators, with all subsequent steps of estimation and inference unchanged. The framework also accommodates richer structures such as multi-dimensional and interactive fixed effects. For instance, region-specific trends or sector–time interactions can be incorporated by augmenting $\mathbf{D}$ with the corresponding interaction terms.

Higher-order spatial effects. The framework also generalizes to models with multiple spatial and temporal lags without requiring row-normalization or time-invariance of spatial weight matrices. A specification with higher-order spatial lags can be of the form:

$$\begin{aligned} y_t &= \rho y_{t-1} + \sum_{\ell=1}^{q_1} \lambda_{1\ell} W_{1\ell t} y_t + \sum_{\ell=1}^{q_2} \lambda_{2\ell} W_{2\ell t} y_{t-1} + x_t\beta + \mu + \alpha_t 1_n + u_t, \\ u_t &= \sum_{\ell=1}^{q_3} \lambda_{3\ell} W_{3\ell t} u_t + v_t, \quad t = 1, 2, \ldots, T. \end{aligned} \tag{6.1}$$

The M-estimation and inference procedures extend directly: the estimating equations simply incorporate the additional lag components, and $\widehat{\Gamma}_{nT}$ and $\widehat{\Sigma}_{nT}$ expand correspondingly. The extra terms enter the gradient, Hessian, and variance–covariance matrices in a systematic manner and are handled using the same principles as in the baseline case. Details on these are available from the authors upon request.

7. Conclusion

This paper develops alternative estimation and inference approaches for dynamic spatial panel data (DSPD) models with fixed effects. The proposed approaches address

several fundamental challenges in the literature, including initial value bias, incidental parameter bias, spatiotemporal heteroskedasticity, time-varying spatial weight matrices, and higher-order spatial interactions. By bias-correcting the conditional concentrated quasi-scores, the method delivers consistent and asymptotically normal estimators without restrictions on the relative rates of growth of the cross-sectional and time dimensions. Furthermore, we propose a simple and novel corrected plug-in procedure that enables valid statistical inference in these settings, irrespective of the relative magnitudes of n and T.

The practical utility of our approaches is highlighted through an application to Chinese firm-level data in investigating dynamic and spatial interactions in corporate decision-making. Our analysis uncovers significant contemporaneous and lagged spillover effects in critical firm policies, including R&D, leverage, and profitability. R&D exhibits both immediate and delayed peer effects, indicative of knowledge diffusion and adjustment frictions, while leverage demonstrates dynamic responses to peer behavior. These findings highlight the importance of jointly modeling spatial dependence and temporal dynamics to better understand firm behavior and the propagation of shocks across networks.

The proposed framework is highly flexible. It accommodates a wide range of sub-models, alternative fixed-effect structures, and higher-order spatial effects without sacrificing theoretical rigor or computational tractability. As such, it provides a powerful and versatile tool for empirical researchers studying complex spatiotemporal interactions in economics and related fields. By offering a unified and robust methodology for DSPD models, this paper lays a solid foundation for future research using high-dimensional panel data with rich network structures.

Acknowledgments

The authors are listed alphabetically and contributed equally to this work. Part of this work was completed during Liyao Li's visit to the School of Economics at Singapore Management University, and she sincerely appreciates their hospitality. Any remaining errors are the author's own responsibility.

Declaration of Interest Statement

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

Data Availability Statement

The data that support the findings of this study are available in China Stock Market & Accounting Research Database (CSMAR) at https://data.csmar.com.

Funding

This work was supported by the National Natural Science Foundation of China under Grant Nos. 72673052 and 72403136.

References

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

Bao, Y. and Yu, J. (2026). A recentered method of moments estimator for spatial dynamic panels. *Journal of Business & Economic Statistics*. online.

Bloom, N. (2009). The impact of uncertainty shocks. *Econometrica*, 77:623–685.

Bloom, N., Schankerman, M., and Van Reenen, J. (2013). Identifying technology spillovers and product market rivalry. *Econometrica*, 81:1347–1393.

Bustamante, M. C. and Frésard, L. (2021). Does firm investment respond to peers' investment? *Management Science*, 67:4703–4724.

Dhaene, G. and Jochmans, K. (2016). Likelihood inference in an autoregression with fixed effects. *Econometric Theory*, 32(5):1178–1215.

Dougal, C., Parsons, C. A., and Titman, S. (2015). Urban vibrancy and corporate growth. *The Journal of Finance*, 70:163–210.

Grieser, W., Hadlock, C., LeSage, J., and Zekhnini, M. (2022a). Network effects in corporate financial policies. *Journal of Financial Economics*, 144:247–272.

Grieser, W., LeSage, J., and Zekhnini, M. (2022b). Industry networks and the geography of firm behavior. *Management Science*, 68:6163–6183.

Grieser, W. D. and Hadlock, C. J. (2019). Panel-data estimation in finance: Testable

assumptions and parameter (in) consistency. *Journal of Financial and Quantitative Analysis*, 54:1–29.

Hoberg, G. and Phillips, G. (2016). Text-based network industries and endogenous product differentiation. *Journal of Political Economy*, 124:1423–1465.

Hsiao, C., Pesaran, M. H., and Tahmiscioglu, A. K. (2002). Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. *Journal of Econometrics*, 109:107–150.

Jaffe, A. B., Trajtenberg, M., and Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *Quarterly Journal of Economics*, 108:577–598.

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

Leary, M. T. and Roberts, M. R. (2014). Do peer firms affect corporate financial policy? *The Journal of Finance*, 69:139–178.

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

Lee, L.-f. and Yu, J. (2014). Efficient GMM estimation of spatial dynamic panel data models with fixed effects. *Journal of Econometrics*, 180:174–197.

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

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

Liu, S. F. and Yang, Z. (2015). Modified QML estimation of spatial autoregressive models with unknown heteroskedasticity and nonnormality. *Regional Science and Urban Economics*, 52:50–70.

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

Meng, X. and Yang, Z. (2025a). Genuinely unbalanced spatial panel data models with fixed effects: M-estimation and inference with an application to fdi inflows. *Econometric Reviews*, pages 1–27.

Meng, X. and Yang, Z. (2025b). Fixed effects estimation of spatial panel model with missing responses: an application to US state tax competition. *Journal of Business & Economic Statistics*, pages 1–15.

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

Van der Vaart, A. W. (2000). *Asymptotic Statistics*, volume 3. Cambridge university press.

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

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

Yu, J., De Jong, R., and Lee, L.-f. (2008). Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large. *Journal of Econometrics*, 146:118–134.

Yu, J., De Jong, R., and Lee, L.-f. (2012). Estimation for spatial dynamic panel data with fixed effects: The case of spatial cointegration. *Journal of Econometrics*, 167(1):16–37.

Notes

[1]Because $\mathbf{P}$ is a rank $r = n + T - 1$ orthogonal projection, write $\mathbf{P} = \sum_{j=1}^{r} u_j u_j'$ with orthonormal $\{u_j\}_{j=1}^{r}$. Under Assumptions E–G, the spectral norm $\|\Phi_\omega' \Phi_\varpi\|_2$ is bounded. Hence $|\Upsilon_{\omega,\varpi}(\psi_0)| \leq N^{-1} \sum_{j=1}^{r} |u_j' \Phi_\omega' \Phi_\varpi u_j| \leq N^{-1} r \|\Phi_\omega' \Phi_\varpi\|_2 = O\{(n+T-1)/[(n-1)(T-1)]\} = O(n^{-1}) + O(T^{-1})$.

Table 1: Empirical bias(sd)[CP95] of `CQMLE`, `M-Est` and `RM-Est` based on time-varying weight matrices, and $(\beta, \rho, \lambda_1, \lambda_2, \lambda_3, \sigma_v^2) = (1, 0.3, 0.2, 0.2, 0.2, 1)$.

	Homoskedasticity			Heteroskedasticity		
	CQMLE	M-Est	RM-Est	CQMLE	M-Est	RM-Est
	$T = 5$, $n = 50$; error = 1, 2, for the two panels below					
β	.0063(.031)	-.0035(.031)[.958]	-.0036(.031)[.954]	.0101(.032)	.0005(.032)[.954]	.0001(.032)[.947]
ρ	-.0292(.024)	.0021(.025)[.966]	.0024(.025)[.960]	-.0253(.023)	.0007(.024)[.947]	.0011(.025)[.949]
λ_1	-.0130(.066)	-.0088(.069)[.944]	-.0092(.070)[.934]	-.0075(.048)	-.0041(.046)[.975]	-.0003(.052)[.943]
λ_2	.0090(.081)	.0106(.085)[.946]	.0112(.085)[.942]	.0051(.059)	.0036(.056)[.979]	-.0008(.062)[.935]
λ_3	-.0910(.181)	-.0369(.159)[.944]	-.0370(.163)[.940]	-.2097(.176)	-.1124(.139)[.989]	-.0881(.246)[.932]
σ_v^2	-.2398(.077)	-.0235(.099)[.920]	–	-.2437(.096)	-.0269(.124)[.945]	–
β	.0070(.031)	-.0028(.031)[.950]	-.0029(.031)[.946]	.0095(.032)	.0000(.032)[.972]	-.0004(.032)[.964]
ρ	-.0290(.026)	.0023(.026)[.938]	.0024(.027)[.938]	-.0267(.024)	-.0006(.024)[.936]	-.0006(.025)[.949]
λ_1	-.0063(.069)	-.0016(.071)[.922]	-.0018(.071)[.914]	-.0094(.049)	-.0059(.046)[.979]	-.0021(.051)[.945]
λ_2	.0011(.082)	.0021(.085)[.934]	.0024(.086)[.912]	.0075(.059)	.0060(.056)[.975]	.0016(.061)[.955]
λ_3	-.0972(.173)	-.0430(.152)[.958]	-.0424(.156)[.942]	-.2063(.167)	-.1097(.131)[.989]	-.0812(.240)[.947]
σ_v^2	-.2359(.174)	-.0181(.225)[.882]	–	-.2401(.217)	-.0220(.281)[.845]	–
	$T = 10$, $n = 50$; error = 1, 2, for the two panels below					
β	.0062(.020)	-.0002(.020)[.954]	-.0002(.020)[.950]	.0089(.021)	.0022(.021)[.918]	.0018(.021)[.924]
ρ	-.0102(.013)	.0010(.014)[.942]	.0010(.014)[.944]	-.0118(.013)	-.0009(.013)[.966]	-.0008(.013)[.958]
λ_1	.0002(.037)	.0008(.037)[.952]	.0008(.037)[.950]	-.0091(.027)	-.0040(.027)[.988]	-.0007(.028)[.932]
λ_2	-.0005(.041)	-.0022(.042)[.944]	-.0021(.042)[.940]	.0111(.031)	.0039(.030)[.992]	.0009(.031)[.942]
λ_3	-.0726(.092)	-.0188(.085)[.962]	-.0186(.085)[.960]	-.1785(.110)	-.0945(.094)[.940]	-.0482(.151)[.956]
σ_v^2	-.1284(.060)	-.0122(.068)[.918]	–	-.1287(.068)	-.0105(.077)[.962]	–
β	.0075(.019)	.0010(.019)[.954]	.0010(.019)[.960]	.0087(.021)	.0021(.021)[.922]	.0016(.021)[.932]
ρ	-.0116(.013)	-.0005(.014)[.960]	-.0005(.013)[.960]	-.0111(.013)	-.0003(.013)[.950]	-.0002(.013)[.948]
λ_1	-.0009(.039)	-.0002(.040)[.932]	-.0002(.040)[.936]	-.0104(.027)	-.0053(.026)[.984]	-.0019(.028)[.958]
λ_2	.0018(.044)	.0000(.045)[.938]	.0000(.045)[.932]	.0107(.031)	.0035(.030)[.988]	.0004(.031)[.962]
λ_3	-.0719(.093)	-.0181(.085)[.960]	-.0176(.086)[.948]	-.1820(.107)	-.0975(.091)[.944]	-.0510(.149)[.940]
σ_v^2	-.1315(.145)	-.0157(.164)[.894]	–	-.1321(.166)	-.0143(.189)[.882]	–
	$T = 20$, $n = 50$; error = 1, 2, for the two panels below					
β	.0036(.013)	-.0010(.013)[.946]	-.0010(.013)[.940]	.0053(.013)	.0000(.013)[.960]	.0001(.014)[.966]
ρ	-.0056(.009)	.0003(.009)[.960]	.0004(.009)[.958]	-.0066(.010)	-.0001(.010)[.948]	-.0004(.011)[.946]
λ_1	-.0005(.031)	-.0025(.031)[.956]	-.0026(.031)[.954]	-.0085(.017)	-.0039(.017)[.996]	.0020(.018)[.952]
λ_2	.0005(.035)	.0034(.036)[.944]	.0035(.036)[.944]	.0087(.020)	.0040(.019)[.996]	-.0021(.020)[.962]
λ_3	-.0615(.066)	-.0046(.062)[.948]	-.0046(.062)[.948]	-.1692(.071)	-.0834(.062)[.888]	-.0234(.099)[.950]
σ_v^2	-.0752(.042)	-.0085(.045)[.940]	–	-.0742(.054)	-.0054(.058)[.946]	–
β	.0041(.012)	-.0004(.012)[.964]	-.0004(.012)[.960]	.0064(.014)	.0011(.014)[.948]	.0012(.014)[.956]
ρ	-.0059(.010)	.0001(.010)[.950]	.0001(.010)[.946]	-.0071(.010)	-.0006(.010)[.934]	-.0008(.010)[.934]
λ_1	.0019(.031)	-.0001(.031)[.942]	-.0001(.031)[.942]	-.0090(.018)	-.0045(.017)[1.000]	.0009(.018)[.954]
λ_2	-.0039(.035)	-.0009(.036)[.946]	-.0009(.036)[.942]	.0088(.021)	.0042(.020)[.994]	-.0013(.021)[.960]
λ_3	-.0645(.067)	-.0072(.063)[.940]	-.0068(.063)[.926]	-.1653(.072)	-.0799(.063)[.908]	-.0165(.099)[.946]
σ_v^2	-.0761(.101)	-.0094(.108)[.934]	–	-.0816(.127)	-.0134(.137)[.896]	–
	$T = 40$, $n = 50$; error = 1, 2, for the two panels below					
β	.0028(.009)	.0002(.009)[.970]	.0002(.009)[.968]	.0027(.009)	-.0002(.009)[.960]	-.0003(.009)[.958]
ρ	-.0037(.007)	-.0003(.007)[.946]	-.0003(.007)[.946]	-.0033(.007)	.0001(.007)[.958]	-.0001(.007)[.960]
λ_1	.0008(.023)	.0007(.023)[.938]	.0007(.023)[.940]	-.0099(.015)	-.0050(.014)[.992]	.0004(.015)[.954]
λ_2	.0004(.026)	.0004(.026)[.940]	.0004(.026)[.940]	.0105(.018)	.0052(.017)[.992]	-.0002(.018)[.958]
λ_3	-.0630(.046)	-.0074(.043)[.944]	-.0073(.043)[.944]	-.1690(.049)	-.0824(.044)[.736]	-.0186(.068)[.938]
σ_v^2	-.0451(.031)	-.0027(.033)[.948]	–	-.0457(.038)	-.0015(.040)[.986]	–
β	.0025(.009)	.0000(.009)[.960]	.0000(.009)[.966]	.0036(.009)	.0007(.009)[.960]	.0006(.009)[.954]
ρ	-.0035(.007)	-.0001(.007)[.950]	-.0001(.007)[.952]	-.0037(.007)	-.0003(.007)[.940]	-.0005(.007)[.944]
λ_1	.0009(.022)	.0008(.022)[.938]	.0008(.022)[.942]	-.0095(.016)	-.0046(.015)[.986]	.0007(.016)[.942]
λ_2	-.0012(.024)	-.0013(.025)[.938]	-.0013(.025)[.946]	.0094(.019)	.0042(.018)[.988]	-.0011(.018)[.936]
λ_3	-.0604(.046)	-.0051(.043)[.944]	-.0051(.043)[.948]	-.1663(.048)	-.0801(.043)[.738]	-.0150(.066)[.950]
σ_v^2	-.0482(.077)	-.0060(.080)[.920]	–	-.0529(.093)	-.0091(.097)[.926]	–

Note: *error = 1(normal), 2(normal mixture). The spatial weight matrices are constructed under GI-I and GI-II for the homoskedastic and heteroskedastic designs, respectively.*

Table 2: Empirical bias(sd)[CP95] of `CQMLE`, `M-Est` and `RM-Est` based on time-varying weight matrices, and $(\beta, \rho, \lambda_1, \lambda_2, \lambda_3, \sigma_v^2) = (1, 0.3, 0.2, 0.2, 0.2, 1)$.

	Homoskedasticity			**Heteroskedasticity**		
	`CQMLE`	`M-Est`	`RM-Est`	`CQMLE`	`M-Est`	`RM-Est`
	$T = 5$, $n = 100$; error = 1, 2, for the two panels below					
β	.0090(.026)	.0006(.026)[.934]	.0006(.026)[.930]	.0075(.026)	.0015(.026)[.941]	.0012(.026)[.941]
ρ	-.0401(.019)	-.0018(.020)[.944]	-.0016(.020)[.944]	-.0318(.017)	-.0015(.018)[.948]	-.0010(.018)[.943]
λ_1	-.0001(.054)	.0017(.056)[.934]	.0016(.056)[.926]	-.0127(.024)	-.0037(.024)[.992]	-.0016(.026)[.937]
λ_2	-.0107(.064)	-.0029(.066)[.940]	-.0026(.067)[.928]	.0079(.025)	.0039(.024)[.998]	.0023(.025)[.968]
λ_3	-.0819(.131)	-.0415(.118)[.960]	-.0416(.120)[.954]	-.1134(.108)	-.0878(.092)[.948]	-.0303(.152)[.960]
σ_v^2	-.2327(.057)	-.0222(.073)[.912]	–	-.2186(.063)	-.0061(.081)[.974]	–
β	.0083(.024)	-.0003(.024)[.948]	-.0003(.024)[.952]	.0067(.026)	.0007(.026)[.945]	.0004(.026)[.943]
ρ	-.0404(.019)	-.0013(.020)[.956]	-.0013(.020)[.954]	-.0320(.018)	-.0016(.017)[.961]	-.0011(.018)[.957]
λ_1	-.0059(.054)	-.0048(.056)[.936]	-.0052(.057)[.938]	-.0132(.023)	-.0042(.023)[1.000]	-.0022(.026)[.961]
λ_2	-.0034(.064)	.0055(.068)[.938]	.0060(.068)[.936]	.0077(.024)	.0038(.024)[1.000]	.0024(.026)[.953]
λ_3	-.0664(.134)	-.0277(.122)[.968]	-.0268(.124)[.956]	-.1119(.103)	-.0864(.088)[.951]	-.0246(.146)[.953]
σ_v^2	-.2160(.116)	-.0008(.150)[.914]	–	-.2164(.157)	-.0030(.202)[.883]	–
	$T = 10$, $n = 100$; error = 1, 2, for the two panels below					
β	.0084(.015)	-.0001(.015)[.932]	-.0001(.015)[.934]	.0086(.014)	.0000(.014)[.940]	-.0002(.014)[.940]
ρ	-.0145(.011)	-.0003(.011)[.940]	-.0003(.011)[.928]	-.0142(.011)	-.0003(.011)[.944]	-.0003(.011)[.958]
λ_1	-.0024(.033)	-.0002(.033)[.956]	-.0002(.033)[.950]	-.0086(.020)	-.0046(.019)[.998]	-.0008(.020)[.942]
λ_2	.0001(.035)	-.0006(.036)[.950]	-.0005(.036)[.946]	.0096(.024)	.0044(.023)[.994]	.0004(.023)[.948]
λ_3	-.0430(.077)	-.0088(.072)[.952]	-.0087(.072)[.952]	-.1224(.073)	-.0876(.066)[.840]	-.0215(.108)[.954]
σ_v^2	-.1183(.040)	-.0097(.045)[.948]	–	-.1155(.052)	-.0062(.059)[.960]	–
β	.0091(.014)	.0006(.014)[.960]	.0006(.014)[.960]	.0098(.015)	.0012(.015)[.946]	.0010(.015)[.942]
ρ	-.0150(.011)	-.0010(.011)[.950]	-.0009(.011)[.956]	-.0145(.011)	-.0008(.011)[.948]	-.0007(.011)[.966]
λ_1	-.0035(.034)	-.0011(.035)[.932]	-.0011(.035)[.934]	-.0088(.019)	-.0048(.018)[.982]	-.0007(.019)[.956]
λ_2	.0015(.038)	.0006(.038)[.940]	.0006(.038)[.942]	.0112(.023)	.0060(.022)[.990]	.0018(.022)[.956]
λ_3	-.0482(.078)	-.0139(.073)[.944]	-.0137(.073)[.944]	-.1247(.072)	-.0897(.065)[.844]	-.0246(.108)[.954]
σ_v^2	-.1160(.097)	-.0071(.110)[.912]	–	-.1182(.123)	-.0093(.139)[.922]	–
	$T = 5$, $n = 200$; error = 1, 2, for the two panels below					
β	.0114(.016)	-.0014(.016)[.958]	-.0015(.016)[.952]	.0118(.016)	-.0003(.016)[.944]	-.0005(.016)[.944]
ρ	-.0325(.012)	-.0002(.012)[.956]	.0000(.012)[.970]	-.0292(.011)	.0001(.012)[.958]	.0002(.012)[.956]
λ_1	.0059(.050)	-.0038(.053)[.936]	-.0040(.053)[.934]	-.0045(.019)	-.0042(.019)[.992]	.0008(.021)[.964]
λ_2	-.0121(.062)	.0053(.067)[.944]	.0056(.067)[.936]	.0023(.025)	.0044(.025)[.988]	-.0008(.027)[.946]
λ_3	-.0511(.111)	-.0160(.103)[.942]	-.0155(.103)[.944]	-.0967(.074)	-.0880(.064)[.816]	-.0222(.111)[.954]
σ_v^2	-.2172(.041)	-.0088(.052)[.934]	–	-.2105(.049)	-.0005(.062)[.962]	–
β	.0136(.017)	.0006(.017)[.952]	.0006(.017)[.942]	.0113(.016)	-.0010(.017)[.944]	-.0011(.017)[.950]
ρ	-.0327(.013)	-.0003(.013)[.946]	-.0004(.013)[.954]	-.0293(.012)	.0002(.013)[.954]	.0003(.013)[.944]
λ_1	.0047(.050)	-.0051(.054)[.950]	-.0051(.054)[.944]	-.0053(.020)	-.0049(.020)[.996]	-.0007(.021)[.948]
λ_2	-.0117(.062)	.0061(.068)[.946]	.0060(.068)[.944]	.0037(.026)	.0058(.026)[.992]	.0014(.028)[.936]
λ_3	-.0519(.111)	-.0164(.103)[.960]	-.0163(.104)[.966]	-.0881(.077)	-.0807(.067)[.843]	-.0121(.114)[.956]
σ_v^2	-.2133(.081)	-.0038(.103)[.946]	–	-.2067(.112)	.0045(.143)[.918]	–
	$T = 10$, $n = 200$; error = 1, 2, for the two panels below					
β	.0081(.011)	-.0004(.011)[.922]	-.0004(.011)[.920]	.0086(.011)	.0004(.011)[.928]	.0004(.011)[.938]
ρ	-.0142(.008)	.0004(.008)[.958]	.0004(.008)[.962]	-.0133(.008)	.0000(.008)[.928]	.0000(.008)[.934]
λ_1	.0005(.030)	.0010(.030)[.954]	.0010(.030)[.954]	-.0040(.015)	-.0020(.015)[.980]	.0000(.016)[.954]
λ_2	-.0006(.035)	-.0006(.036)[.936]	-.0006(.036)[.940]	.0041(.018)	.0023(.018)[.982]	.0004(.018)[.958]
λ_3	-.0362(.066)	-.0120(.062)[.956]	-.0118(.062)[.954]	-.0684(.049)	-.0574(.046)[.830]	-.0077(.062)[.956]
σ_v^2	-.1069(.030)	-.0013(.034)[.946]	–	-.1064(.038)	-.0008(.043)[.956]	–
β	.0085(.010)	.0000(.010)[.940]	.0000(.010)[.948]	.0081(.011)	-.0001(.011)[.954]	-.0001(.011)[.950]
ρ	-.0147(.008)	-.0001(.008)[.928]	-.0001(.008)[.928]	-.0135(.008)	-.0001(.008)[.950]	-.0002(.008)[.946]
λ_1	.0005(.032)	.0010(.032)[.952]	.0010(.032)[.946]	-.0035(.016)	-.0014(.016)[.966]	.0005(.017)[.932]
λ_2	-.0003(.037)	-.0003(.037)[.946]	-.0003(.037)[.954]	.0026(.019)	.0008(.019)[.976]	-.0011(.019)[.922]
λ_3	-.0339(.065)	-.0100(.062)[.960]	-.0100(.062)[.956]	-.0675(.049)	-.0567(.046)[.826]	-.0069(.062)[.950]
σ_v^2	-.1081(.069)	-.0026(.078)[.934]	–	-.0979(.087)	.0088(.098)[.942]	–

Note: *error = 1(normal), 2(normal mixture). The spatial weight matrices are constructed under GI-I and GI-II for the homoskedastic and heteroskedastic designs, respectively.*

Table 3: Peer Effects in Firm Decisions

	R&D	*Leverage*	*ROA*	*CF*	*Tobin's Q*
Dynamic and spatial parameter estimates					
ρ	-0.006***	-0.004	0.011	0.000	0.047***
	(0.002)	(0.013)	(0.025)	(0.090)	(0.015)
λ_1	0.350***	0.130**	0.268***	0.023	0.045
	(0.082)	(0.056)	(0.104)	(0.150)	(0.036)
λ_2	0.181***	-0.165***	-0.115	-0.041	0.008
	(0.059)	(0.052)	(0.225)	(0.214)	(0.033)
Covariate parameter estimates					
Sales	0.002***	0.095***	0.017***	0.033***	-0.845***
	(0.000)	(0.003)	(0.002)	(0.004)	(0.033)
Age	-0.001***	0.007***	0.002	-0.001	-0.063**
	(0.000)	(0.002)	(0.001)	(0.002)	(0.026)
Cash	0.001***	-0.027***	0.012***	0.013***	0.260***
	(0.000)	(0.002)	(0.001)	(0.002)	(0.030)
Z-score	0.001*	0.000	0.000	0.000	-0.014
	(0.000)	(0.001)	(0.000)	(0.002)	(0.025)
ROA	0.000	-0.051***			0.304***
	(0.000)	(0.004)			(0.059)
Tobin's Q	0.005***	-0.018***	0.017***	0.031***	
	(0.001)	(0.003)	(0.002)	(0.002)	
Leverage	-0.002***				
	(0.000)				
Tangibility		-0.004	0.003**	0.012***	-0.014
		(0.002)	(0.001)	(0.001)	(0.027)
R&D			0.003**	-0.003	0.470***
			(0.002)	(0.003)	(0.039)
Observations	4716	4716	4716	4716	4716

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

Appendix A: Notations

The following list summarizes notations that are frequently used in the paper.

N	Effective sample size $N = nT - (n + T - 1)$.
μ, α_t, 1_n	Individual fixed effects, time fixed effect at time t, and $n \times 1$ vector of ones.
λ, δ, θ, ψ	$\lambda = (\lambda_1, \lambda_2, \lambda_3)'$, $\delta = (\rho, \lambda')'$, $\theta = (\beta', \rho, \lambda')'$, and $\psi = (\beta', \rho, \lambda', \sigma_v^2)'$.
$\mathbf{Y}$, $\mathbf{Y}_{-1}$, $\mathbf{Y}_0$	$\mathbf{Y} = (y_1', \ldots, y_T')'$, $\mathbf{Y}_{-1} = (y_0', \ldots, y_{T-1}')'$, and $\mathbf{Y}_0 = 1_T \otimes y_0$.
$\mathbf{X}$, $\mathbf{U}$, $\mathbf{V}$	$\mathbf{X} = (x_1', \ldots, x_T')'$, $\mathbf{U} = (u_1', \ldots, u_T')'$, and $\mathbf{V} = (v_1', \ldots, v_T')'$.
$\mathbf{W}_r$	$\mathbf{W}_r = \mathtt{bdiag}(W_{r1}, \ldots, W_{rT})$, $r = 1, 2, 3$.
$\mathbf{D}_\mu$, $\mathbf{D}_\alpha$, $\mathbf{D}$, ϕ	Dummy-variable matrices and normalized fixed-effect vector: $\mathbf{D}_\mu = 1_T \otimes I_n$, $\mathbf{D}_\alpha = I_T \otimes 1_n$, $\mathbf{D} = (\mathbf{D}_\mu^*, \mathbf{D}_\alpha)$, and $\phi = (\mu^{*'}, \alpha')'$.
B_{1t}, B_{2t}, B_{3t}	$B_{1t}(\lambda_1) = I_n - \lambda_1 W_{1t}$, $B_{2t}(\rho, \lambda_2) = \rho I_n + \lambda_2 W_{2t}$, and $B_{3t}(\lambda_3) = I_n - \lambda_3 W_{3t}$.
$\mathbf{B}_1$, $\mathbf{B}_2$, $\mathbf{B}_3$	$\mathbf{B}_r(\lambda_r) = \mathtt{bdiag}\{B_{r1}(\lambda_r), \ldots, B_{rT}(\lambda_r)\}$ for $r = 1, 3$, and $\mathbf{B}_2(\rho, \lambda_2) = \mathtt{bdiag}\{B_{21}(\rho, \lambda_2), \ldots, B_{2T}(\rho, \lambda_2)\}$.
$\mathbf{Z}(\theta)$, $\mathbf{Z}^*(\theta)$	$\mathbf{Z}(\theta) = \mathbf{B}_1\mathbf{Y} - \mathbf{B}_2\mathbf{Y}_{-1} - \mathbf{X}\beta$ and $\mathbf{Z}^*(\theta) = \mathbf{B}_3\mathbf{Z}(\theta)$.
$\Omega(\lambda_3)$, $\mathbf{M}(\lambda_3)$, $\mathbf{P}(\lambda_3)$	$\Omega(\lambda_3) = \mathtt{bdiag}(\Omega_1(\lambda_3), \ldots, \Omega_T(\lambda_3))$ with $\Omega_t(\lambda_3) = [B_{3t}'(\lambda_3)B_{3t}(\lambda_3)]^{-1}$; $\mathbf{P}(\lambda_3) = \mathbf{B}_3(\lambda_3)\mathbf{D}[\mathbf{D}'\Omega^{-1}(\lambda_3)\mathbf{D}]^{-1}\mathbf{D}'\mathbf{B}_3'(\lambda_3)$; and $\mathbf{M}(\lambda_3) = I_{nT} - \mathbf{P}(\lambda_3)$.
$\mathcal{B}_t$, $\mathbf{A}$, $\mathbf{A}_{-1}$	$\mathcal{B}_t = B_{1t}^{-1}B_{2t}$, $\mathbf{A} = \mathtt{bdiag}(\mathcal{B}_1, \mathcal{B}_2\mathcal{B}_1, \ldots, \mathcal{B}_T \cdots \mathcal{B}_1)$, and $\mathbf{A}_{-1} = \mathtt{bdiag}(I_n, \mathcal{B}_1, \ldots, \mathcal{B}_{T-1} \cdots \mathcal{B}_1)$.
$\mathbf{Q}$, $\mathbf{Q}_{-1}$	$(\mathbf{Q})_{ts} = (\mathcal{B}_t \cdots \mathcal{B}_{s+1})B_{1s}^{-1}$ for $t > s$ and $(\mathbf{Q})_{ss} = B_{1s}^{-1}$; $(\mathbf{Q}_{-1})_{ts} = (\mathcal{B}_{t-1} \cdots \mathcal{B}_{s+1})B_{1s}^{-1}$ for $t > s + 1$ and $(\mathbf{Q}_{-1})_{s+1,s} = B_{1s}^{-1}$; all other blocks are zero.
Π_β, Π_ρ, Π_{λ_1}, Π_{λ_2}	$\Pi_\beta = \mathbf{M}\mathbf{B}_3\mathbf{X}$, $\Pi_\rho = \mathbf{M}\mathbf{B}_3\eta_{-1} + \Phi_\rho\mathbf{B}_3\mathbf{D}\phi$, $\Pi_{\lambda_1} = \mathbf{M}\mathbf{B}_3\mathbf{W}_1\eta + \Phi_{\lambda_1}\mathbf{B}_3\mathbf{D}\phi$, and $\Pi_{\lambda_2} = \mathbf{M}\mathbf{B}_3\mathbf{W}_2\eta_{-1} + \Phi_{\lambda_2}\mathbf{B}_3\mathbf{D}\phi$, where $\eta = \mathbf{Q}\mathbf{X}\beta_0$ and $\eta_{-1} = \mathbf{Q}_{-1}\mathbf{X}\beta_0$.
Φ_ρ, Φ_{λ_1}, Φ_{λ_2}, Φ_{λ_3}	$\Phi_\rho = \mathbf{M}\mathbf{B}_3\mathbf{Q}_{-1}\mathbf{B}_3^{-1}$, $\Phi_{\lambda_1} = \mathbf{M}\mathbf{B}_3\mathbf{W}_1\mathbf{Q}\mathbf{B}_3^{-1}$, $\Phi_{\lambda_2} = \mathbf{M}\mathbf{B}_3\mathbf{W}_2\mathbf{Q}_{-1}\mathbf{B}_3^{-1}$, and $\Phi_{\lambda_3} = \mathbf{M}\mathbf{W}_3\mathbf{B}_3^{-1}\mathbf{M}$.
Ψ_ρ, Ψ_{λ_1}, Ψ_{λ_2}	$\Psi_\rho = \mathbf{M}\mathbf{B}_3\mathbf{A}_{-1}$, $\Psi_{\lambda_1} = \mathbf{M}\mathbf{B}_3\mathbf{W}_1\mathbf{A}$, and $\Psi_{\lambda_2} = \mathbf{M}\mathbf{B}_3\mathbf{W}_2\mathbf{A}_{-1}$.
Φ_ω^r, $\mathbf{R}_\omega$	$\mathbf{R}_\omega = \mathtt{diag}(\Phi_\omega)\mathtt{diag}(\mathbf{M})^{-1}$ and $\Phi_\omega^r = \Phi_\omega - \mathbf{M}\mathbf{R}_\omega$, $\omega \in \{\rho, \lambda_1, \lambda_2, \lambda_3\}$.

Appendix B: Detailed Derivations

B.1. Conditional Concentrated Quasi-Scores

Based on the conditional concentrated quasi log-likelihood (CCQL) function in (2.5), the corresponding conditional concentrated quasi-scores (CCQS) are given by

$$S_{nT}^{c}(\psi)=\begin{cases}\frac{1}{\sigma_v^2}\mathbf{X}'\mathbf{B}_3'(\lambda_3)\mathbf{M}(\lambda_3)\mathbf{Z}^*(\theta),\\ \frac{1}{\sigma_v^2}\mathbf{Y}_{-1}'\mathbf{B}_3'(\lambda_3)\mathbf{M}(\lambda_3)\mathbf{Z}^*(\theta),\\ \frac{1}{\sigma_v^2}\mathbf{Y}'\mathbf{W}_1'\mathbf{B}_3'(\lambda_3)\mathbf{M}(\lambda_3)\mathbf{Z}^*(\theta)-\mathtt{tr}[\mathbf{W}_1\mathbf{B}_1^{-1}(\lambda_1)],\\ \frac{1}{\sigma_v^2}\mathbf{Y}_{-1}'\mathbf{W}_2'\mathbf{B}_3'(\lambda_3)\mathbf{M}(\lambda_3)\mathbf{Z}^*(\theta),\\ \frac{1}{\sigma_v^2}\mathbf{Z}^{*\prime}(\theta)\mathbf{M}(\lambda_3)\mathbf{W}_3\mathbf{B}_3^{-1}(\lambda_3)\mathbf{M}(\lambda_3)\mathbf{Z}^*(\theta)-\mathtt{tr}[\mathbf{W}_3\mathbf{B}_3^{-1}(\lambda_3)],\\ \frac{1}{2\sigma_v^4}\mathbf{Z}^{*\prime}(\theta)\mathbf{M}(\lambda_3)\mathbf{Z}^*(\theta)-\frac{nT}{2\sigma_v^2}.\end{cases}\tag{B.1}$$

Under mild regularity conditions, maximizing the conditional quasi log-likelihood in (2.3) is equivalent to solving the estimating equations $S_{nT}^{c}(\psi)=0$.

Denote a parametric quantity at the true parameters by dropping its argument(s), e.g., $B_{t,1}=B_{1,t}(\lambda_{10})$. Backward substitutions of the first equation in (1.1) give

$$\mathbf{Y}=\mathbf{A}\mathbf{Y}_0+\boldsymbol{\eta}+\mathbf{Q}\mathbf{Z}\quad\text{and}\quad\mathbf{Y}_{-1}=\mathbf{A}_{-1}\mathbf{Y}_0+\boldsymbol{\eta}_{-1}+\mathbf{Q}_{-1}\mathbf{Z},\tag{B.2}$$

where $\mathbf{Y}_0=1_T\otimes y_0$, 1_T is a $T\times 1$ vector of ones, $\mathbf{Z}=\mathbf{Z}(\beta_0,\rho_0,\lambda_{10},\lambda_{20})$, $\boldsymbol{\eta}=\mathbf{Q}\mathbf{X}\beta_0$, $\boldsymbol{\eta}_{-1}=\mathbf{Q}_{-1}\mathbf{X}\beta_0$, $\mathbf{A}=\mathtt{bdiag}(\mathcal{B}_1,\mathcal{B}_2\mathcal{B}_1,\ \ldots,\ \mathcal{B}_T\cdots\mathcal{B}_1)$, $\mathbf{A}_{-1}=\mathtt{bdiag}(I_n,\ \mathcal{B}_1,\ \ldots,\ \mathcal{B}_{T-1}\cdots\mathcal{B}_1)$. With these expressions and noting $\mathbf{M}\mathbf{Z}^*=\mathbf{M}\mathbf{V}$, we can write

$$S_{nT}^{c}(\psi_0)=\begin{cases}\frac{1}{\sigma_{v0}^2}\Pi_\beta'\mathbf{V}\\ \frac{1}{\sigma_{v0}^2}\mathbf{V}'\Psi_\rho\mathbf{Y}_0+\frac{1}{\sigma_{v0}^2}\mathbf{V}'\Phi_\rho\mathbf{V}+\frac{1}{\sigma_{v0}^2}\Pi_\rho'\mathbf{V},\\ \frac{1}{\sigma_{v0}^2}\mathbf{V}'\Psi_{\lambda_1}\mathbf{Y}_0+\frac{1}{\sigma_{v0}^2}\mathbf{V}'\Phi_{\lambda_1}\mathbf{V}+\frac{1}{\sigma_{v0}^2}\Pi_{\lambda_1}'\mathbf{V}-\mathtt{tr}(\mathbf{W}_1\mathbf{B}_1^{-1}),\\ \frac{1}{\sigma_{v0}^2}\mathbf{V}'\Psi_{\lambda_2}\mathbf{Y}_0+\frac{1}{\sigma_{v0}^2}\mathbf{V}'\Phi_{\lambda_2}\mathbf{V}+\frac{1}{\sigma_{v0}^2}\Pi_{\lambda_2}'\mathbf{V},\\ \frac{1}{\sigma_{v0}^2}\mathbf{V}'\Phi_{\lambda_3}\mathbf{V}-\mathtt{tr}(\mathbf{W}_3\mathbf{B}_3^{-1}),\\ \frac{1}{2\sigma_{v0}^4}\mathbf{V}'\mathbf{M}\mathbf{V}-\frac{nT}{2\sigma_{v0}^2},\end{cases}\tag{B.3}$$

which are linear combinations of terms that are linear or quadratic in $\mathbf{V}$, and bilinear in $\mathbf{V}$ and $\mathbf{Y}_0$, where $\Pi_\beta=\mathbf{M}\mathbf{B}_3\mathbf{X}$, $\Pi_\rho=\mathbf{M}\mathbf{B}_3\boldsymbol{\eta}_{-1}+\Phi_\rho\mathbf{B}_3\mathbf{D}\phi$, $\Pi_{\lambda_1}=\mathbf{M}\mathbf{B}_3\mathbf{W}_1\boldsymbol{\eta}+\Phi_{\lambda_1}\mathbf{B}_3\mathbf{D}\phi$,

and $\Pi_{\lambda_2} = \mathbf{M}\mathbf{B}_3\mathbf{W}_2\boldsymbol{\eta}_{-1} + \Phi_{\lambda_2}\mathbf{B}_3\mathbf{D}\phi$; $\Phi_\rho = \mathbf{M}\mathbf{B}_3\mathbf{Q}_{-1}\mathbf{B}_3^{-1}$, $\Phi_{\lambda_1} = \mathbf{M}\mathbf{B}_3\mathbf{W}_1\mathbf{Q}\mathbf{B}_3^{-1}$, $\Phi_{\lambda_2} = \mathbf{M}\mathbf{B}_3\mathbf{W}_2\mathbf{Q}_{-1}\mathbf{B}_3^{-1}$, and $\Phi_{\lambda_3} = \mathbf{M}\mathbf{W}_3\mathbf{B}_3^{-1}\mathbf{M}$; $\Psi_\rho = \mathbf{M}\mathbf{B}_3\mathbf{A}_{-1}$, $\Psi_{\lambda_1} = \mathbf{M}\mathbf{B}_3\mathbf{W}_1\mathbf{A}$, and $\Psi_{\lambda_2} = \mathbf{M}\mathbf{B}_3\mathbf{W}_2\mathbf{A}_{-1}$.

Using (B.3), the expectation of the CCQS evaluated at ψ_0 is given by:

$$\mathrm{E}[S_{nT}^c(\psi_0)] = \begin{cases} 0 \\ \mathtt{tr}(\Phi_\rho), \\ \mathtt{tr}(\Phi_{\lambda_1}) - \mathtt{tr}(\mathbf{W}_1\mathbf{B}_1^{-1}), \\ \mathtt{tr}(\Phi_{\lambda_2}), \\ \mathtt{tr}(\Phi_{\lambda_3}) - \mathtt{tr}(\mathbf{W}_3\mathbf{B}_3^{-1}), \\ \frac{1}{2\sigma_{v0}^2}\mathtt{tr}(\mathbf{M}) - \frac{nT}{2\sigma_{v0}^2}, \end{cases} \tag{B.4}$$

To illustrate the source of inconsistency, consider the σ_v^2 component of the CCQS. Its expectation, normalized by nT and evaluated at the true parameter values, is $-\frac{n+T-1}{2\sigma_{v0}^2 nT}$, which does not converge to zero when either n or T is fixed. This confirms that the CQML estimator fails to satisfy the necessary condition for consistency under such cases.

B.2. LQ Form Representation of Estimating Functions

The estimating functions in (2.6), evaluated at the true parameter values, also admit representations as linear combinations of terms that are linear or quadratic in $\mathbf{V}$, as well as bilinear in $(\mathbf{V}, \mathbf{Y}_0)$. These representations play two important roles: they facilitate the derivation of the asymptotic properties of $\hat{\psi}_{\mathtt{M}}$ and provide a convenient basis for constructing the variance–covariance (VC) matrix of the estimating functions.

Under homoskedasticity, the estimating functions at ψ_0 can be written as

$$S_{nT}^*(\psi_0) = \begin{cases} \frac{1}{\sigma_{v0}^2}\Pi_\beta'\mathbf{V} \\ \frac{1}{\sigma_{v0}^2}\mathbf{V}'\Psi_\rho\mathbf{Y}_0 + \frac{1}{\sigma_{v0}^2}\mathbf{V}'\Phi_\rho\mathbf{V} + \frac{1}{\sigma_{v0}^2}\Pi_\rho'\mathbf{V} - \mu_\rho, \\ \frac{1}{\sigma_{v0}^2}\mathbf{V}'\Psi_{\lambda_1}\mathbf{Y}_0 + \frac{1}{\sigma_{v0}^2}\mathbf{V}'\Phi_{\lambda_1}\mathbf{V} + \frac{1}{\sigma_{v0}^2}\Pi_{\lambda_1}'\mathbf{V} - \mu_{\lambda_1}, \\ \frac{1}{\sigma_{v0}^2}\mathbf{V}'\Psi_{\lambda_2}\mathbf{Y}_0 + \frac{1}{\sigma_{v0}^2}\mathbf{V}'\Phi_{\lambda_2}\mathbf{V} + \frac{1}{\sigma_{v0}^2}\Pi_{\lambda_2}'\mathbf{V} - \mu_{\lambda_2}, \\ \frac{1}{\sigma_{v0}^2}\mathbf{V}'\Phi_{\lambda_3}\mathbf{V} - \mu_{\lambda_3}, \\ \frac{1}{2\sigma_{v0}^4}\mathbf{V}'\mathbf{M}\mathbf{V} - \mu_\sigma, \end{cases} \tag{B.5}$$

where $\mu_r = \mathtt{tr}(\Phi_r), r = \rho, \lambda_1, \lambda_2, \lambda_3$, $\mu_\sigma = \frac{N}{2\sigma_{v0}^2}$, and other quantities are defined in (B.3).

Under heteroskedasticity, the robust estimating functions evaluated at θ_0 are

$$S_{nT}^{r}(\theta_0) = \begin{cases} \Pi_{\beta}'\mathbf{V} \\ \mathbf{V}'\Psi_{\rho}\mathbf{Y}_0 + \mathbf{V}'\Phi_{\rho}^{r}\mathbf{V} + \Pi_{\rho}^{r\prime}\mathbf{V}, \\ \mathbf{V}'\Psi_{\lambda_1}\mathbf{Y}_0 + \mathbf{V}'\Phi_{\lambda_1}^{r}\mathbf{V} + \Pi_{\lambda_1}^{r\prime}\mathbf{V}, \\ \mathbf{V}'\Psi_{\lambda_2}\mathbf{Y}_0 + \mathbf{V}'\Phi_{\lambda_2}^{r}\mathbf{V} + \Pi_{\lambda_2}^{r\prime}\mathbf{V}, \\ \mathbf{V}'\Phi_{\lambda_3}^{r}\mathbf{V} + \Pi_{\lambda_3}^{r\prime}\mathbf{V}, \end{cases} \tag{B.6}$$

where $\Pi_{\rho}^{r} = \mathbf{M}\mathbf{B}_3\boldsymbol{\eta}_{-1} + \Phi_{\rho}^{r}\mathbf{B}_3\mathbf{D}\phi$, $\Pi_{\lambda_1}^{r} = \mathbf{M}\mathbf{B}_3\mathbf{W}_1\boldsymbol{\eta} + \Phi_{\lambda_1}^{r}\mathbf{B}_3\mathbf{D}\phi$, $\Pi_{\lambda_2}^{r} = \mathbf{M}\mathbf{B}_3\mathbf{W}_2\boldsymbol{\eta}_{-1} + \Phi_{\lambda_2}^{r}\mathbf{B}_3\mathbf{D}\phi$, and $\Pi_{\lambda_3}^{r} = \Phi_{\lambda_3}^{r}\mathbf{B}_3\mathbf{D}\phi$; $\Phi_{\rho}^{r} = \Phi_{\rho} - \mathbf{M}\mathbf{R}_{\rho}$, $\Phi_{\lambda_1}^{r} = \Phi_{\lambda_1} - \mathbf{M}\mathbf{R}_{\lambda_1}$, $\Phi_{\lambda_2}^{r} = \Phi_{\lambda_2} - \mathbf{M}\mathbf{R}_{\lambda_2}$, and $\Phi_{\lambda_3}^{r} = \Phi_{\lambda_3} - \mathbf{M}\mathbf{R}_{\lambda_3}$; Π_{β}, Ψ_{ρ}, Ψ_{λ_1}, and Ψ_{λ_2} are defined in (B.3). A key feature of this construction is that the diagonal elements of Φ_{ν}^{r} are all zero as $\mathtt{diag}(\Phi_{\nu}^{r}) = \mathtt{diag}(\mathbf{M}\mathbf{R}_{\nu}), \nu = \rho, \lambda_1, \lambda_2, \lambda_3$. As a result, the variance–covariance matrix of $S_{nT}^{r}(\theta_0)$ is free of the third and fourth moments of the error terms.

B.3. Hessian Matrices

Let the $nT \times 1$ vector X_p, $p = 1, \cdots, k$, be the pth column of $\mathbf{X}$. Denote $N \times 1$ vectors, $X_{k+1} = Y_{-1}$, $X_{k+2} = \mathbf{W}_1 Y$, $X_{k+3} = \mathbf{W}_2 Y_{-1}$. Further, denote $\beta_{k+1} = \rho$, $\beta_{k+2} = \lambda_1$, and $\beta_{k+3} = \lambda_2$. The Hessian matrix, $H(\psi) = \frac{\partial}{\partial \psi'} S_{nT}^{*}(\psi)$, has the elements:

$$\begin{aligned}
H_{\beta_p\beta_q} &= -\tfrac{1}{\sigma_v^2} X_p'\mathbf{B}_3'\mathbf{M}\mathbf{B}_3 X_q - \dot{\mu}_{\beta_p,\beta_q}, & H_{\beta_p\lambda_3} &= H_{\lambda_3\beta_p} - \dot{\mu}_{\beta_p,\lambda_3}, \\
H_{\beta_p\sigma_v^2} &= -\tfrac{1}{\sigma_v^4} X_p'\mathbf{B}_3'\mathbf{M}\mathbf{Z}^{*}(\theta), & H_{\lambda_3\beta_p} &= -\tfrac{1}{\sigma_v^2}\mathbf{Z}^{*\prime}(\theta)\mathbf{M}(\mathbf{G}+\mathbf{G}')\mathbf{M}\mathbf{B}_3 X_p, \\
H_{\lambda_3\sigma_v^2} &= -\tfrac{1}{\sigma_v^4}\mathbf{Z}^{*\prime}(\theta)\mathbf{M}\mathbf{G}\mathbf{M}\mathbf{Z}^{*}(\theta), & H_{\sigma_v^2\beta_p} &= H_{\beta_p\sigma_v^2}, \\
H_{\sigma_v^2\sigma_v^2} &= -\tfrac{1}{\sigma_v^6}\mathbf{Z}^{*\prime}(\theta)\mathbf{M}\mathbf{Z}^{*}(\theta) + \tfrac{nT-n-T+1}{2\sigma_v^4}, & H_{\sigma_v^2\lambda_3} &= H_{\lambda_3\sigma_v^2}, \\
H_{\lambda_3\lambda_3} &= \tfrac{1}{\sigma_v^2}\mathbf{Z}^{*\prime}(\theta)\mathbf{M}(\mathbf{G}^{\circ}\mathbf{P}\mathbf{G}^{\circ} - \mathbf{G}'\mathbf{G})\mathbf{M}\mathbf{Z}^{*}(\theta) - \dot{\mu}_{\lambda_3,\lambda_3},
\end{aligned}$$

where $\mathbf{G} = \mathbf{W}_3\mathbf{B}_3^{-1}$, $\mathbf{G}^{\circ} = \mathbf{G} + \mathbf{G}'$ and $\dot{\mu}_{\omega,\varpi} = \frac{\partial}{\partial \varpi}\mu_{\omega}$ with $\omega, \varpi = (\beta_1, \ldots \beta_{k+3}, \lambda_3)$ and $\dot{\mu}_{\beta_p,\beta_q} = 0$ for $p \leq k$ or $q \leq k$.

For the heteroskedasticity case, the Hessian takes the same form as above with terms related to σ_v^2 deleted and the terms μ_{ω} redefined:

$$\mu_{\omega} = \mathbf{Z}^{*\prime}(\theta)\mathbf{M}(\lambda_3)\mathbf{R}_{\omega}(\delta)\mathbf{Z}^{*}(\theta), \omega = \rho, \lambda_1, \lambda_2, \lambda_3.$$

B.4. VC Matrices of Estimating Functions

To estimate the VC matrix of the estimating functions, we derive its analytical form using results in (B.5) and (B.6). Let γ and κ denote the skewness and the excess kurtosis of v_{it}. Under homoskedasticity, $nT\Sigma_{NT}(\psi_0) = \text{Var}[S^*_{nT}(\psi_0)]$ has the elements:

$$\Sigma_{\beta\beta} = \Pi'_\beta\Pi_\beta/\sigma^2_{v0}; \quad \Sigma_{\beta\rho} = \Pi'_\beta(\xi_\rho/\sigma^2_{v0} + \gamma\phi_\rho/\sigma_{v0}); \quad \Sigma_{\beta\lambda_1} = \Pi'_\beta(\xi_{\lambda_1}/\sigma^2_{v0} + \gamma\phi_{\lambda_1}/\sigma_{v0});$$

$$\Sigma_{\beta\lambda_3} = \gamma\Pi'_\beta\phi_{\lambda_3}/\sigma_{v0}; \quad \Sigma_{\beta\lambda_2} = \Pi'_\beta(\xi_{\lambda_2}/\sigma^2_{v0} + \gamma\phi_{\lambda_2}/\sigma_{v0}); \quad \Sigma_{\beta\sigma^2} = \gamma\Pi'_\beta\phi_\sigma/\sigma_{v0};$$

$$\Sigma_{\rho\rho} = \xi'_\rho\xi_\rho/\sigma^2_{v0} + \mathbf{G}_{\rho,\rho} + 2\gamma\xi'_\rho\phi_\rho/\sigma_{v0};$$

$$\Sigma_{\rho\lambda_1} = \xi'_\rho\xi_{\lambda_1}/\sigma^2_{v0} + \mathbf{G}_{\rho,\lambda_1} + \gamma\xi'_\rho\phi_{\lambda_1}/\sigma_{v0} + \gamma\xi'_{\lambda_1}\phi_\rho/\sigma_{v0};$$

$$\Sigma_{\rho\lambda_2} = \xi'_\rho\xi_{\lambda_2}/\sigma^2_{v0} + \mathbf{G}_{\rho,\lambda_2} + \gamma\xi'_\rho\phi_{\lambda_2}/\sigma_{v0} + \gamma\xi'_{\lambda_2}\phi_\rho/\sigma_{v0};$$

$$\Sigma_{\rho\lambda_3} = \mathbf{G}_{\rho,\lambda_3} + \gamma\xi'_\rho\phi_{\lambda_3}/\sigma_{v0}; \quad \Sigma_{\rho\sigma^2} = \mathbf{G}_{\rho,\sigma} + \gamma\xi'_\rho\phi_\sigma/\sigma_{v0};$$

$$\Sigma_{\lambda_1\lambda_1} = \xi'_{\lambda_1}\xi_{\lambda_1}/\sigma^2_{v0} + \mathbf{G}_{\lambda_1,\lambda_1} + 2\gamma\xi'_{\lambda_1}\phi_{\lambda_1}/\sigma_{v0};$$

$$\Sigma_{\lambda_1\lambda_2} = \xi'_{\lambda_1}\xi_{\lambda_2}/\sigma^2_{v0} + \mathbf{G}_{\lambda_1,\lambda_2} + \gamma\xi'_{\lambda_1}\phi_{\lambda_2}/\sigma_{v0} + \gamma\xi'_{\lambda_2}\phi_{\lambda_1}/\sigma_{v0};$$

$$\Sigma_{\lambda_1\lambda_3} = \mathbf{G}_{\lambda_1,\lambda_3} + \gamma\xi'_{\lambda_1}\phi_{\lambda_3}/\sigma_{v0}; \quad \Sigma_{\lambda_1\sigma^2} = \mathbf{G}_{\lambda_1,\sigma} + \gamma\xi'_{\lambda_1}\phi_\sigma/\sigma_{v0};$$

$$\Sigma_{\lambda_2\lambda_2} = \xi'_{\lambda_2}\xi_{\lambda_2}/\sigma^2_{v0} + \mathbf{G}_{\lambda_2,\lambda_2} + 2\gamma\xi'_{\lambda_2}\phi_{\lambda_2}/\sigma_{v0};$$

$$\Sigma_{\lambda_2\lambda_3} = \mathbf{G}_{\lambda_2,\lambda_3} + \gamma\xi'_{\lambda_2}\phi_{\lambda_3}/\sigma_{v0}; \quad \Sigma_{\lambda_2\sigma^2} = \mathbf{G}_{\lambda_2,\sigma} + \gamma\xi'_{\lambda_2}\phi_\sigma/\sigma_{v0};$$

$$\Sigma_{\lambda_3\lambda_3} = \mathbf{G}_{\lambda_3,\lambda_3}; \quad \Sigma_{\lambda_3\sigma^2} = \mathbf{G}_{\lambda_3,\sigma}; \quad \Sigma_{\sigma^2\sigma^2} = \mathbf{G}_{\sigma,\sigma};$$

where $\phi_\omega = \texttt{diagv}(\Phi_\omega)$, for $\omega = \rho, \lambda_1, \lambda_2, \lambda_3, \sigma$, $\Phi_\sigma = \mathbf{M}/(2\sigma^2_{v0})$, $\xi_\nu = \Pi_\nu + \Psi_\nu\text{E}(\mathbf{Y}_0)$, for $\nu = \rho, \lambda_1, \lambda_2$, and $\mathbf{G}_{\omega,\varpi} = \kappa\phi'_\omega\phi_\varpi + \text{tr}[\Phi_\omega(\Phi_\varpi + \Phi'_\varpi)]$ for $\omega, \varpi = \rho, \lambda_1, \lambda_2, \lambda_3, \sigma$.

Under heteroskedasticity, the elements of $nT\Sigma^r_{NT}(\psi_0) = \text{Var}[S^r_{nT}(\psi_0)]$ based on the robust estimating functions are:

$$\Sigma^r_{\beta\beta} = \Pi'_\beta\mathbf{H}\Pi_\beta; \quad \Sigma^r_{\beta\rho} = \Pi'_\beta\mathbf{H}\Pi^r_\rho + \Pi'_\beta\mathbf{H}\Psi_\rho\text{E}(\mathbf{Y}_0); \quad \Sigma^r_{\beta\lambda_1} = \Pi'_\beta\mathbf{H}\Pi^r_{\lambda_1} + \Pi'_\beta\mathbf{H}\Psi_{\lambda_1}\text{E}(\mathbf{Y}_0);$$

$$\Sigma^r_{\beta\lambda_3} = \Pi'_\beta\mathbf{H}\Pi^r_{\lambda_3}; \quad \Sigma^r_{\beta\lambda_2} = \Pi'_\beta\mathbf{H}\Pi^r_{\lambda_2} + \Pi'_\beta\mathbf{H}\Psi_{\lambda_2}\text{E}(\mathbf{Y}_0);$$

$$\Sigma^r_{\rho\rho} = \Pi^{r\prime}_\rho\mathbf{H}\Pi^r_\rho + \mathbf{G}^r_{\rho,\rho} + 2\Pi^{r\prime}_\rho\mathbf{H}\Psi_\rho\text{E}(\mathbf{Y}_0) + \text{E}(\mathbf{Y}'_0\Psi'_\rho\mathbf{H}\Psi_\rho\mathbf{Y}_0);$$

$$\Sigma^r_{\rho\lambda_1} = \Pi^{r\prime}_\rho\mathbf{H}\Pi^r_{\lambda_1} + \mathbf{G}^r_{\rho,\lambda_1} + (\Pi^{r\prime}_\rho\mathbf{H}\Psi_{\lambda_1} + \Pi^{r\prime}_{\lambda_1}\mathbf{H}\Psi_\rho)\text{E}(\mathbf{Y}_0) + \text{E}(\mathbf{Y}'_0\Psi'_\rho\mathbf{H}\Psi_{\lambda_1}\mathbf{Y}_0);$$

$$\Sigma^r_{\rho\lambda_2} = \Pi^{r\prime}_\rho\mathbf{H}\Pi^r_{\lambda_2} + \mathbf{G}^r_{\rho,\lambda_2} + (\Pi^{r\prime}_\rho\mathbf{H}\Psi_{\lambda_2} + \Pi^{r\prime}_{\lambda_2}\mathbf{H}\Psi_\rho)\text{E}(\mathbf{Y}_0) + \text{E}(\mathbf{Y}'_0\Psi'_\rho\mathbf{H}\Psi_{\lambda_2}\mathbf{Y}_0);$$

$$\Sigma^r_{\rho\lambda_3} = \Pi^{r\prime}_\rho\mathbf{H}\Pi^r_{\lambda_3} + \mathbf{G}^r_{\rho,\lambda_3} + \Pi^{r\prime}_{\lambda_3}\mathbf{H}\Psi_\rho\text{E}(\mathbf{Y}_0);$$

$$\Sigma^r_{\lambda_1\lambda_1} = \Pi^{r\prime}_{\lambda_1}\mathbf{H}\Pi^r_{\lambda_1} + \mathbf{G}^r_{\lambda_1,\lambda_1} + 2\Pi^{r\prime}_{\lambda_1}\mathbf{H}\Psi_{\lambda_1}\text{E}(\mathbf{Y}_0) + \text{E}(\mathbf{Y}'_0\Psi'_{\lambda_1}\mathbf{H}\Psi_{\lambda_1}\mathbf{Y}_0);$$

$$\Sigma^r_{\lambda_1\lambda_2} = \Pi^{r\prime}_{\lambda_1}\mathbf{H}\Pi^r_{\lambda_2} + \mathbf{G}^r_{\lambda_1,\lambda_2} + (\Pi^{r\prime}_{\lambda_1}\mathbf{H}\Psi_{\lambda_2} + \Pi^{r\prime}_{\lambda_2}\mathbf{H}\Psi_{\lambda_1})\text{E}(\mathbf{Y}_0) + \text{E}(\mathbf{Y}'_0\Psi'_{\lambda_1}\mathbf{H}\Psi_{\lambda_2}\mathbf{Y}_0);$$

$$\Sigma^r_{\lambda_1\lambda_3} = \Pi^{r\prime}_{\lambda_1}\mathbf{H}\Pi^r_{\lambda_3} + \mathbf{G}^r_{\lambda_1,\lambda_3} + \Pi^{r\prime}_{\lambda_3}\mathbf{H}\Psi_{\lambda_1}\mathrm{E}(\mathbf{Y}_0);$$

$$\Sigma^r_{\lambda_2\lambda_2} = \Pi^{r\prime}_{\lambda_2}\mathbf{H}\Pi^r_{\lambda_2} + \mathbf{G}^r_{\lambda_2,\lambda_2} + 2\Pi^{r\prime}_{\lambda_2}\mathbf{H}\Psi_{\lambda_2}\mathrm{E}(\mathbf{Y}_0) + \mathrm{E}(\mathbf{Y}_0'\Psi'_{\lambda_2}\mathbf{H}\Psi_{\lambda_2}\mathbf{Y}_0);$$

$$\Sigma^r_{\lambda_2\lambda_3} = \Pi^{r\prime}_{\lambda_2}\mathbf{H}\Pi^r_{\lambda_3} + \mathbf{G}^r_{\lambda_2,\lambda_3} + \Pi^{r\prime}_{\lambda_3}\mathbf{H}\Psi_{\lambda_2}\mathrm{E}(\mathbf{Y}_0);$$

$$\Sigma^r_{\lambda_3\lambda_3} = \Pi^{r\prime}_{\lambda_3}\mathbf{H}\Pi^r_{\lambda_3} + \mathbf{G}^r_{\lambda_3,\lambda_3},$$

where $\mathbf{G}^r_{\omega,\varpi} = \mathtt{tr}[\mathbf{H}\Phi^r_\omega\mathbf{H}(\Phi^r_\varpi + \Phi^{r\prime}_\varpi)]$, $\omega, \varpi = \rho, \lambda_1, \lambda_2, \lambda_3$.

Appendix C: Assumptions

For a real $n \times m$ matrix A, $\|A\|_1$ denotes the maximum column sum norm, $\|A\|_\infty$ maximum row sum norm. Rigorous studies on the asymptotic properties of the proposed M-estimator require the following basic regularity conditions. We define $\bar{S}^{*c}_{nT}(\delta)$ as the population counterpart of the concentrated estimating function $S^{*c}_{nT}(\delta)$, obtained by concentrating out β and σ^2_v from $\mathrm{E}[S^*_{nT}(\psi)]$. Similarly, $\bar{S}^{rc}_{nT}(\delta)$ denotes the population counterpart of $S^{rc}_{nT}(\delta)$, obtained by concentrating out β from $\mathrm{E}[S^r_{nT}(\theta)]$.

Assumption A. *The innovations v_{it} are iid for all i and t with $E(v_{it}) = 0$, $Var(v_{it}) = \sigma^2_{v0}$, and $E|v_{it}|^{4+\epsilon_0} < \infty$ for some $\epsilon_0 > 0$.*

Assumption A*. *The innovations v_{it} are inid across i and t with $E(v_{it}) = 0$, $Var(v_{it}) = \sigma^2_{it}$, and $\sup_{i,t}\mathrm{E}|v_{it}|^4 < \infty$.*

Assumption B. *(i) Process started at $t = -m$ ($m \geq 0$) and data collection started at $t = 0$; (ii) y_0 is independent of $\{v_t, t = 1, \ldots, T\}$.*

Assumption C. *The parameter space $\boldsymbol{\Delta}$ of δ is compact, and the true parameter vector δ_0 lies in its interior*

Assumption D. *(i) The elements of the time-varying regressors $\{x_t, t = 1, \ldots, T\}$, the individual effects $\{\mu_i\}$ and time effects $\{\alpha_t\}$ are nonstochastic and are uniformly bounded in i and t, and (ii) the limit $\lim_{nT\to\infty}\frac{1}{nT}\mathbf{X}'\mathbf{M}(\lambda_3)\mathbf{X}$ exists and is nonsingular, uniformly in λ_3 in a compact parameter space $\boldsymbol{\Lambda}_3$.*

Assumption E. *(i) For $\nu = 1, 2, 3$, the elements $w_{\nu,ijt}$ of $W_{\nu t}$ are at most of order h_n^{-1}, uniformly in all i,j and t, and $w_{\nu,iit} = 0$ for all i and t; (ii) $h_n/n \to 0$ as $n \to \infty$; (iii) $\{W_{\nu t}, \nu = 1, 2, 3\}$ and $\{B^{-1}_{\nu t0}, \nu = 1, 3\}$ are uniformly bounded in both row and column sum norms; (iv) For $B_{\nu t} = B_{\nu t}(\lambda_\nu)$ with $\nu = 1, 3$, either $\|B^{-1}_{\nu t}\|_\infty$ or $\|B^{-1}_{\nu t}\|_1$ is*

bounded, uniformly in λ_ν *in a compact parameter space* $\mathbf{\Lambda}_\nu$ *uniformly in* t*, and* $0 < \underline{c}_\nu \leq \inf_{t,\lambda_\nu \in \mathbf{\Lambda}_\nu} \gamma_{\min}(B_\nu' B_\nu) \leq \sup_{t,\lambda_\nu \in \mathbf{\Lambda}_\nu} \gamma_{\max}(B_\nu' B_\nu) \leq \bar{c}_\nu < \infty$.

Assumption F. *For an* $n \times n$ *matrix* Φ *uniformly bounded in either row or column sums, with elements of uniform order* h_n^{-1}*, and an* $n \times 1$ *vector* b *with elements of uniform order* $h_n^{-1/2}$*, (i) (1)* $\frac{h_n}{n} y_0' \Phi y_0 = O_p(1)$*, (2)* $\frac{h_n}{n}[y_0 - \mathrm{E}(y_0)]'b = o_p(1)$*, (3)* $\frac{h_n}{n}[y_0' \Phi y_0 - \mathrm{E}(y_0' \Phi y_0)] = o_p(1)$*; (ii)* $\sup_t \max\{\|\mathcal{B}_t\|_1, \|\mathcal{B}_t\|_\infty\} < 1$.

Assumption G. *Both* $\|\mathbf{M}(\lambda_3)\|_\infty$ *and* $\|\mathbf{M}(\lambda_3)\|_1$ *are bounded, uniformly in* λ_3 *in a compact parameter space* $\mathbf{\Lambda}_3$.

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

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

Assumptions A and A* correspond to the homoskedastic and heteroskedastic error settings, respectively. They require the idiosyncratic errors v_{it} to be independent across both the cross-sectional and time dimensions. Cross-sectional and serial correlations are not a primary concern in the present framework, as such dependence is explicitly modeled through the spatial lag, time lag, space–time lag, and spatial error components of the DSPD model. In the robust case, the zero diagonals of Φ_ω^r remove the diagonal fourth-cumulant terms, so the LQ-form CLT for the robust score only requires a finite $2 + \epsilon_0$ moment. Assumption A* retains a uniformly bounded fourth moment for the corrected plug-in VC estimator, where it bounds the stochastic remainders from the plug-in bias corrections: the centered quadratic terms created by the trace corrections and the squared-innovation products that appear when $\mathbf{H}$ is replaced by $\widehat{\mathbf{H}}$. Assumptions B-D are standard in the literature. Assumption D (ii) ensures the consistency of $\hat{\beta}_{\mathtt{M}}$ and $\hat{\sigma}_{\mathtt{M}}^2$ follows from that of $\hat{\delta}_{\mathtt{M}}$. Assumption E imposes regularity conditions on the spatial weight matrices, paralleling Assumption E in Yang (2018) and closely related to the conditions in Lee (2004). Allowing h_n to grow with n but at a slower rate is useful as it corresponds to an important spatial layout where the degree of spatial dependence increases with n, see Lee (2004) and Yang (2015) for related discussions. Assumption F(i) ensures that the initial observations have a proper stochastic behavior. It is satisfied if the process starts from the infinite past and has evolved according to (1.1), provided

that Assumption F(ii) holds. Assumption F(ii) is a high-level condition that restricts temporal and cross-sectional dependence and ensures the stationarity of y_t when T is allowed to grow. Throughout the paper, whenever T diverges, we restrict attention to stationary dynamic settings. Similar assumptions are imposed in Yu et al. (2008) and Lee and Yu (2014). Assumption G facilitates our asymptotic analysis to establish the consistency of our M-estimators, related to Assumption F in Meng and Yang (2025b).

Assumption H provides the identification condition under homoskedasticity, while Assumption H* gives the corresponding condition under heteroskedasticity. Both require that the population concentrated corrected scores have no common zero other than δ_0, which is the standard identification requirement in extremum and M-estimation; see, for example, Theorem 5.9 of Van der Vaart (2000). Similar high-level identification conditions are used in spatial and dynamic panel models, including Lee and Yu (2014), Yang (2018), and Kuersteiner and Prucha (2020). Because the AQS equations in (2.6) and (3.3) are highly nonlinear in the spatial parameters, dynamic propagation matrices, and fixed-effect projection, primitive conditions are difficult to state. As suggested by Bao and Yu (2026) for nonlinear moment conditions in spatial dynamic panels, a useful practical diagnostic is to simulate the population analogue of the moment equations over the admissible parameter space and inspect the surface for flat regions or additional zeros.

Kuersteiner and Prucha (2020) (Section 2, Lemma EX1) provide explicit population-rank sufficient conditions for an example without temporal, space-time lags, or spatial error term. Under their rank conditions, the true parameter vector is uniquely determined by the linear moment restrictions alone or, when these are insufficient, by imposing the linear and quadratic moment restrictions jointly. Our population concentrated corrected scores likewise combine expectations of linear and recentered quadratic forms in the error. However, this shared structure does not by itself establish identification. The linear and corrected quadratic contributions enter our estimating equations jointly and may cancel at false parameter values. Moreover, the temporal and space–time lags and spatial errors in our model further complicate the extension of the identification argument developed for KP's example. We therefore retain Assumption H; see also Bao and Yu (2026) (Remark 1 and Supplement, Section S.4.1) for closely related discussions.

Appendix D: Proof of Theorems

Proof of Theorem 2.1 (i): This proof closely follows the Proof of Theorem 2.1 of Li and Yang (2021) and the Proof of Theorem 3.1 of Yang (2018). As a result, some details are omitted here for brevity. Given the nonlinear parameters δ, solving the estimating equations for β and σ^2 from (2.6), we obtain the constrained M-estimators $\hat{\beta}(\delta)$ and $\hat{\sigma}^2(\delta)$. Substituting $\hat{\beta}(\delta)$ and $\hat{\sigma}^2(\delta)$ back into the δ-component of $S^*_{nT}(\psi)$ gives the concentrated estimating function $S^{*c}_{nT}(\delta)$. Similarly, $\bar{S}^{*c}_{nT}(\delta)$ is the population counterpart of the concentrated estimating function. It is easy to see that $S^{*c}_{nT}(\hat{\delta}_{\texttt{M}}) = \mathbf{0}$, and $\bar{S}^{*c}_{nT}(\delta_0) = \mathbf{0}$. By Theorem 5.9 of Van der Vaart (2000), $\hat{\delta}_{\texttt{M}}$ will be consistent for δ_0 if $\sup_{\delta\in\boldsymbol{\Delta}} \frac{1}{N}\big\|S^{*c}_{nT}(\delta) - \bar{S}^{*c}_{nT}(\delta)\big\| \xrightarrow{p} 0$, and the identification condition holds. Using expressions in (B.2), both components of the $S^*_{nT}(\psi)$ can be expressed as linear, quadratic terms in $\mathbf{V}$, and bilinear terms in $\mathbf{V}$ and $\mathbf{Y}_0$. Using the uniform boundedness of row or column sums of the relevant matrices and the assumption on a compact parameter space, the point-wise convergence of the quadratic terms is established by Lemma A.4 of Yang (2018); the point-wise convergence of the linear terms is established by Chebyshev's inequality; and the point-wise convergence of the bilinear terms is established by Assumption F (i). With stochastically equicontinuity of $\bar{S}^{*c}_{nT}(\delta)$, we establish the uniform convergence $\sup_{\delta\in\boldsymbol{\Delta}} \frac{1}{N}\big\|S^{*c}_{nT}(\delta) - \bar{S}^{*c}_{nT}(\delta)\big\| \xrightarrow{p} 0$, and with Assumption H, the consistency of the M-estimator follows. The result can be established in a similar manner for the heteroskedasticity case. ∎

Proof of Theorem 2.1 (ii): By applying the mean value theorem (henceforth MVT) to each element of $S^*_{nT}(\hat{\psi}_{\texttt{M}})$, we have,

$$\tfrac{1}{N}S^*_{nT}(\hat{\psi}_{\texttt{M}}) = \tfrac{1}{N}S^*_{NT}(\psi_0) + \big[\tfrac{1}{N}\tfrac{\partial}{\partial\psi'}S^*_N(\psi)|_{\psi=\bar{\psi}_r \text{ in } r\text{th row}}\big](\hat{\psi}_{\texttt{M}} - \psi_0) = 0, \tag{D.1}$$

where $\{\bar{\psi}_r\}$ are between $\hat{\psi}$ and ψ_0 elementwise. The result of the theorem follows if

(a) $\frac{1}{\sqrt{N}}S^*_{nT}(\psi_0) \xrightarrow{D} N\big[0, \lim_{nT\to\infty}\Sigma_{nT}(\psi_0)\big]$,

(b) $\frac{1}{N}\big[\frac{\partial}{\partial\psi'}S^*_{nT}(\psi|_{\psi=\bar{\psi}_r \text{ in } r\text{th row}}) - \frac{\partial}{\partial\psi'}S^*_{nT}(\psi_0)\big] \xrightarrow{p} 0$, and

(c) $\frac{1}{N}\big[\frac{\partial}{\partial\psi'}S^*_{nT}(\psi_0) - \mathrm{E}\big(\frac{\partial}{\partial\psi'}S^*_{nT}(\psi_0)\big)\big] \xrightarrow{p} 0$.

This theorem is proved in a manner similar to Theorem 2.2 in Li and Yang (2021) and Theorem 3.2 in Yang (2018). Therefore, the details are omitted for brevity. For (a),

using the representation in equation (B.5) and the uniform boundedness of the row or column sums of the relevant matrices, the joint asymptotic normality of $\frac{1}{\sqrt{N}}S^*_{nT}(\psi_0)$ is established by Lemma A.5 of Yang (2018) and the Cramér-Wold device.

For (b), the detailed expression of the Hessian matrix $H(\psi) = \frac{\partial}{\partial \psi'}S^*_{nT}(\psi)$ is given in Appendix A. The convergence can be shown by using the Mean Value Theorem under the uniform boundedness of row or column sums of the relevant matrices, Assumptions C and F, and the consistency of the estimator.

For (c), using the representation of $\mathbf{Y}$ and $\mathbf{Y}_{-1}$ in equation (B.2), the elements of the Hessian matrix can be written as linear combinations of quadratic and linear terms of $\mathbf{V}$, quadratic and linear terms of $\mathbf{Y}_0$, and bilinear terms of $\mathbf{V}$ and $\mathbf{Y}_0$. Thus, the results follow by repeatedly applying Lemma A.4 of Yang (2018) under the uniform boundedness of the relevant matrices and Assumption F. ■

Proof of Theorem 2.2: First, the result $\Gamma_{NT}(\hat{\psi}_{\mathtt{M}}) - \Gamma_{NT}(\psi_0) \xrightarrow{p} 0$ is implied by result **(b)** in the proof of Theorem 2.1 (ii). Next, to show $\hat{\Sigma}_{nT} - \Sigma_{nT}(\psi_0) \xrightarrow{p} 0$, we only need to show that $\Upsilon(\hat{\psi}_{\mathtt{M}}) - \Upsilon(\psi_0) = o_p(1)$. This can be shown by repeatedly applying MVT under the uniform boundedness of the relevant matrices in Assumptions C and F, and the consistency of the estimator. ■

Proof s of Corollary 3.1 follows closely the arguments used in the proof of Theorem 2.1 and is therefore omitted. ■

Proof of Corollary 3.2 follows closely the Proof of Corollary D.1 and D.2 of Meng and Yang (2025a). ■

Appendix E: Variable Definitions

Table D.1: Variable Constructions of Empirical Study

Variables	Descriptions
Sales	Log of sales
Age	Years since establishment
Cash	Cash holdings
Z-score	$6.56 \cdot \frac{wcapt}{at} + 3.26 \cdot \frac{re}{at} + 6.72 \cdot \frac{ebit}{at} + 1.05 \cdot \frac{ceq}{lt}$
ROA	Return on assets
Tobin's Q	(Assets + market equity - book equity)/Assets
Leverage	Total book debt divided by total book assets
Tangibility	Net property, plant, and equipment scaled by total assets
R&D	Research and development expenditures divided by sales

Note: Variable construction follow Grieser et al. (2022b). All variables are standardized.

Appendix F: Additional Monte Carlo Results

This appendix reports additional Monte Carlo experiments designed to examine the finite-sample performance of the proposed estimators under larger sample sizes, stronger spatial dependence, and spatial cointegration, as well as to compare them with existing methods under time-invariant spatial weight matrices.

We consider models with both time-invariant and time-varying spatial weight matrices. For the time-varying cases, we adopt the same interaction structures as in the main text, namely GI-I under homoskedasticity and GI-II under heteroskedasticity. For the time-invariant cases, we consider two additional interaction structures. Under homoskedasticity, Group Interaction III (GI-III) randomly partitions the n units into $K(n) = \text{Round}(n^{0.5})$ non-overlapping groups of heterogeneous sizes. Under heteroskedasticity, Group Interaction IV (GI-IV) employs the fixed group-size pattern $(3, 5, 7, 9, 11, 15)$, replicated to match the cross-sectional dimension. Heteroskedasticity under GI-IV is generated as in the main text following Lin and Lee (2010).

Tables F.1–F.3 report Monte Carlo results for models with time-varying spatial weight matrices under both homoskedasticity and heteroskedasticity, comparing the CQMLE, M-Est, and RM-Est. Table F.1 considers the baseline design in the main text with larger sample sizes, $(N, T) = (100, 20), (100, 40), (200, 20)$, and $(200, 40)$. Table F.2 con-

siders a high-spatial-dependence design with $(\beta, \rho, \lambda_1, \lambda_2, \lambda_3, \sigma_v^2) = (1, 0.2, 0.6, 0.1, 0.2, 1)$, whereas Table F.3 examines a spatial-cointegration design with $(\beta, \rho, \lambda_1, \lambda_2, \lambda_3, \sigma_v^2) = (1, 0.4, 0.4, 0.2, 0.2, 1)$ following Yu et al. (2012). Across all designs, the qualitative patterns are broadly consistent with those documented in the main text. However, establishing a formal asymptotic theory for spatially cointegrated models is beyond the scope of the present paper and is left for future research.

Tables F.4 and F.5 examine models with time-constant spatial weight matrices. Table F.4 considers a homoskedastic model without a spatial-error effect and compares the CQMLE, the M-estimator of Yang (2018) (M-Yang), the GMM estimator of Lee and Yu (2014), M-Est, and RM-Est. Table F.5 considers a heteroskedastic model and compares the CQMLE, the heteroskedasticity-robust estimator of Li and Yang (2020) (RM-LY), M-Est, and RM-Est. Across all scenarios, the proposed estimators perform well and consistently outperform the benchmark estimators in terms of bias and efficiency, while the proposed inference procedures maintain reliable coverage probabilities.

Table F.1: Empirical bias(sd)[CP95] of `CQMLE`, `M-Est`, and `RM-Est` based on time-varying weight matrices; $(\beta, \rho, \lambda_1, \lambda_2, \lambda_3, \sigma_v^2) = (1, 0.3, 0.2, 0.2, 0.2, 1)$.

	Homoskedasticity			**Heteroskedasticity**		
	`CQMLE`	`M-Est`	`RM-Est`	`CQMLE`	`M-Est`	`RM-Est`
			$T = 20$, $n = 100$; error = 1, 2, for the two panels below			
β	.0056(.009)	.0009(.009)[.950]	.0009(.009)[.950]	.0045(.010)	-.0001(.010)[.948]	-.0002(.010)[.950]
ρ	-.0075(.007)	-.0002(.007)[.956]	-.0002(.007)[.956]	-.0065(.007)	.0002(.007)[.950]	.0002(.007)[.950]
λ_1	-.0012(.026)	-.0007(.026)[.950]	-.0006(.026)[.946]	-.0051(.014)	-.0026(.014)[.992]	.0009(.015)[.954]
λ_2	.0026(.030)	.0013(.030)[.942]	.0013(.030)[.946]	.0058(.017)	.0027(.017)[.988]	-.0009(.017)[.966]
λ_3	-.0438(.054)	-.0066(.051)[.946]	-.0066(.051)[.948]	-.1050(.046)	-.0695(.043)[.760]	-.0126(.062)[.972]
σ_v^2	-.0614(.030)	-.0028(.032)[.954]	–	-.0587(.039)	.0002(.041)[.960]	–
β	.0048(.009)	.0001(.009)[.956]	.0001(.009)[.950]	.0048(.009)	.0003(.009)[.948]	.0002(.009)[.956]
ρ	-.0071(.008)	.0002(.008)[.964]	.0002(.008)[.956]	-.0069(.007)	-.0002(.007)[.964]	-.0002(.007)[.962]
λ_1	.0001(.025)	.0006(.025)[.956]	.0006(.025)[.946]	-.0056(.014)	-.0031(.014)[.972]	.0001(.015)[.950]
λ_2	.0007(.028)	-.0008(.029)[.942]	-.0007(.029)[.946]	.0070(.018)	.0040(.017)[.978]	.0007(.018)[.944]
λ_3	-.0465(.056)	-.0092(.054)[.928]	-.0091(.054)[.924]	-.1016(.049)	-.0663(.045)[.782]	-.0084(.065)[.940]
σ_v^2	-.0647(.072)	-.0063(.076)[.944]	–	-.0656(.093)	-.0071(.098)[.936]	–
			$T = 40$, $n = 100$; error = 1, 2, for the two panels below			
β	.0025(.007)	-.0002(.007)[.946]	-.0002(.007)[.948]	.0034(.007)	.0006(.007)[.940]	.0006(.007)[.936]
ρ	-.0037(.005)	-.0002(.005)[.934]	-.0002(.005)[.934]	-.0037(.005)	-.0003(.005)[.938]	-.0004(.005)[.938]
λ_1	.0003(.019)	.0001(.019)[.942]	.0001(.019)[.940]	-.0063(.009)	-.0038(.009)[.996]	-.0001(.009)[.952]
λ_2	-.0004(.021)	-.0006(.022)[.936]	-.0006(.022)[.938]	.0067(.011)	.0037(.011)[.994]	.0001(.011)[.946]
λ_3	-.0421(.038)	-.0035(.036)[.952]	-.0035(.036)[.950]	-.0990(.033)	-.0621(.031)[.566]	-.0022(.043)[.966]
σ_v^2	-.0356(.022)	-.0022(.023)[.954]	–	-.0329(.028)	.0008(.029)[.950]	–
β	.0030(.007)	.0003(.007)[.956]	.0003(.007)[.958]	.0029(.007)	.0001(.007)[.952]	.0001(.007)[.954]
ρ	-.0036(.005)	-.0002(.005)[.950]	-.0002(.005)[.952]	-.0031(.005)	.0002(.005)[.954]	.0002(.005)[.952]
λ_1	.0020(.018)	.0018(.018)[.954]	.0018(.018)[.952]	-.0068(.009)	-.0043(.009)[.992]	-.0004(.009)[.964]
λ_2	-.0027(.020)	-.0029(.021)[.936]	-.0029(.021)[.934]	.0071(.010)	.0040(.010)[.996]	.0003(.010)[.962]
λ_3	-.0446(.035)	-.0058(.033)[.970]	-.0058(.033)[.966]	-.1024(.033)	-.0654(.031)[.582]	-.0069(.044)[.966]
σ_v^2	-.0363(.050)	-.0028(.052)[.956]	–	-.0316(.065)	.0022(.067)[.948]	–
			$T = 20$, $n = 200$; error = 1, 2, for the two panels below			
β	.0046(.007)	-.0003(.007)[.958]	-.0003(.007)[.960]	.0050(.006)	.0004(.006)[.960]	.0004(.006)[.960]
ρ	-.0066(.005)	.0002(.005)[.958]	.0002(.005)[.958]	-.0063(.005)	-.0001(.005)[.946]	-.0001(.005)[.946]
λ_1	.0020(.023)	.0022(.023)[.942]	.0022(.023)[.942]	-.0038(.010)	-.0025(.009)[.984]	.0002(.010)[.954]
λ_2	-.0018(.025)	-.0021(.025)[.942]	-.0021(.025)[.940]	.0042(.011)	.0024(.011)[.988]	-.0003(.012)[.952]
λ_3	-.0342(.046)	-.0081(.044)[.938]	-.0080(.044)[.938]	-.0734(.033)	-.0588(.031)[.610]	-.0035(.043)[.958]
σ_v^2	-.0557(.021)	-.0012(.022)[.952]	–	-.0536(.027)	.0010(.028)[.968]	–
β	.0051(.007)	.0002(.007)[.932]	.0002(.007)[.938]	.0047(.006)	.0002(.006)[.976]	.0002(.006)[.974]
ρ	-.0068(.005)	.0000(.005)[.950]	-.0001(.005)[.954]	-.0062(.005)	.0001(.005)[.948]	.0000(.005)[.952]
λ_1	-.0024(.022)	-.0023(.022)[.954]	-.0023(.022)[.960]	-.0036(.010)	-.0024(.009)[.994]	.0003(.010)[.942]
λ_2	.0023(.024)	.0021(.024)[.966]	.0021(.024)[.962]	.0040(.011)	.0021(.011)[.988]	-.0005(.012)[.954]
λ_3	-.0252(.043)	.0005(.042)[.948]	.0005(.042)[.948]	-.0738(.034)	-.0591(.032)[.570]	-.0044(.044)[.954]
σ_v^2	-.0534(.054)	.0012(.057)[.936]	–	-.0540(.067)	.0006(.071)[.938]	–
			$T = 40$, $n = 200$; error = 1, 2, for the two panels below			
β	.0028(.005)	.0001(.005)[.956]	.0001(.005)[.956]	.0027(.005)	.0002(.005)[.954]	.0001(.005)[.952]
ρ	-.0033(.004)	.0001(.004)[.958]	.0001(.004)[.958]	-.0031(.003)	.0002(.003)[.942]	.0001(.003)[.946]
λ_1	-.0006(.014)	-.0004(.014)[.950]	-.0004(.014)[.948]	-.0052(.006)	-.0040(.006)[.996]	-.0004(.006)[.962]
λ_2	.0013(.016)	.0005(.016)[.962]	.0005(.016)[.964]	.0053(.007)	.0037(.007)[.986]	.0003(.007)[.956]
λ_3	-.0274(.029)	-.0005(.028)[.944]	-.0006(.028)[.946]	-.0811(.022)	-.0645(.021)[.170]	-.0022(.030)[.944]
σ_v^2	-.0308(.017)	-.0016(.017)[.938]	–	-.0293(.018)	.0000(.019)[.972]	–
β	.0024(.005)	-.0002(.005)[.954]	-.0002(.005)[.956]	.0023(.005)	-.0002(.005)[.958]	-.0003(.005)[.956]
ρ	-.0033(.004)	.0001(.004)[.952]	.0001(.004)[.954]	-.0031(.004)	.0001(.004)[.932]	.0000(.004)[.938]
λ_1	-.0010(.015)	-.0007(.015)[.946]	-.0007(.015)[.946]	-.0041(.006)	-.0029(.006)[.990]	.0006(.006)[.956]
λ_2	.0018(.017)	.0009(.017)[.942]	.0009(.017)[.942]	.0043(.007)	.0027(.007)[.988]	-.0007(.007)[.958]
λ_3	-.0271(.030)	-.0003(.029)[.950]	-.0004(.029)[.952]	-.0803(.023)	-.0637(.022)[.218]	-.0009(.031)[.934]
σ_v^2	-.0292(.039)	.0001(.040)[.936]	–	-.0277(.044)	.0016(.045)[.956]	–

Note: *error = 1(normal), 2(normal mixture). The spatial weight matrices are constructed under GI-I and GI-II for the homoskedastic and heteroskedastic designs, respectively.*

Table F.2: Empirical bias(sd)[CP95] of `CQMLE`, `M-Est`, and `RM-Est` based on time-varying weight matrices; $(\beta, \rho, \lambda_1, \lambda_2, \lambda_3, \sigma_v^2) = (1, 0.2, 0.6, 0.1, 0.2, 1)$.

	Homoskedasticity			Heteroskedasticity		
	CQMLE	M-Est	RM-Est	CQMLE	M-Est	RM-Est
	$T = 5$, $n = 50$; error = 1, 2, for the two panels below					
β	.0020(.030)	-.0007(.030)[.962]	-.0007(.030)[.964]	.0044(.030)	.0015(.030)[.965]	.0006(.030)[.954]
ρ	-.0123(.014)	-.0008(.014)[.956]	-.0008(.014)[.958]	-.0066(.011)	.0008(.011)[.950]	.0004(.011)[.944]
λ_1	-.0064(.027)	-.0012(.027)[.940]	-.0013(.027)[.932]	-.0113(.027)	-.0107(.027)[.933]	-.0031(.027)[.914]
λ_2	.0037(.032)	.0008(.033)[.926]	.0009(.033)[.928]	.0136(.034)	.0131(.033)[.937]	.0042(.034)[.919]
λ_3	-.0897(.156)	-.0406(.134)[.968]	-.0398(.136)[.966]	-.1899(.169)	-.0947(.133)[.988]	-.0717(.226)[.960]
σ_v^2	-.2398(.079)	-.0298(.101)[.920]	–	-.2385(.098)	-.0256(.125)[.927]	–
β	.0011(.032)	-.0016(.032)[.936]	-.0016(.032)[.932]	.0034(.032)	.0003(.032)[.952]	-.0008(.032)[.940]
ρ	-.0114(.015)	.0000(.015)[.936]	.0002(.015)[.944]	-.0063(.011)	.0010(.011)[.954]	.0008(.011)[.956]
λ_1	-.0069(.027)	-.0017(.027)[.938]	-.0016(.027)[.932]	-.0103(.026)	-.0097(.025)[.954]	-.0015(.026)[.929]
λ_2	.0050(.032)	.0021(.033)[.946]	.0021(.033)[.938]	.0112(.032)	.0106(.032)[.952]	.0011(.033)[.917]
λ_3	-.0922(.160)	-.0431(.139)[.954]	-.0426(.141)[.946]	-.2056(.168)	-.1080(.132)[.998]	-.1040(.247)[.954]
σ_v^2	-.2416(.168)	-.0320(.214)[.874]	–	-.2289(.226)	-.0127(.289)[.856]	–
	$T = 5$, $n = 100$; error = 1, 2, for the two panels below					
β	.0089(.024)	-.0005(.024)[.930]	-.0005(.024)[.926]	.0095(.023)	.0023(.023)[.958]	.0012(.023)[.942]
ρ	-.0186(.013)	.0007(.014)[.954]	.0008(.014)[.958]	-.0197(.014)	.0004(.014)[.954]	-.0001(.014)[.956]
λ_1	-.0084(.023)	-.0025(.023)[.944]	-.0025(.023)[.944]	-.0124(.016)	-.0097(.016)[.954]	-.0002(.017)[.952]
λ_2	.0056(.027)	.0022(.027)[.948]	.0022(.027)[.938]	.0146(.023)	.0109(.023)[.958]	.0008(.024)[.958]
λ_3	-.0509(.118)	-.0245(.105)[.972]	-.0242(.106)[.968]	-.1254(.107)	-.0906(.089)[.962]	-.0545(.161)[.964]
σ_v^2	-.2177(.054)	-.0091(.069)[.952]	–	-.2160(.067)	-.0057(.085)[.948]	–
β	.0089(.024)	-.0004(.024)[.952]	-.0004(.024)[.958]	.0072(.021)	.0001(.021)[.971]	-.0011(.022)[.943]
ρ	-.0188(.014)	.0004(.014)[.938]	.0004(.014)[.932]	-.0197(.014)	.0003(.014)[.933]	.0001(.014)[.929]
λ_1	-.0078(.023)	-.0019(.022)[.946]	-.0019(.022)[.956]	-.0124(.017)	-.0097(.017)[.951]	.0001(.017)[.959]
λ_2	.0055(.026)	.0021(.027)[.958]	.0021(.027)[.960]	.0130(.024)	.0092(.024)[.951]	-.0011(.025)[.955]
λ_3	-.0567(.124)	-.0295(.109)[.962]	-.0290(.109)[.950]	-.1306(.113)	-.0944(.095)[.949]	-.0639(.171)[.953]
σ_v^2	-.2222(.119)	-.0147(.151)[.912]	–	-.2180(.164)	-.0079(.209)[.882]	–
	$T = 10$, $n = 100$; error = 1, 2, for the two panels below					
β	.0043(.014)	.0006(.014)[.950]	.0006(.014)[.950]	.0060(.014)	.0020(.014)[.940]	.0001(.014)[.936]
ρ	-.0073(.007)	.0000(.008)[.948]	.0000(.008)[.950]	-.0062(.007)	-.0005(.007)[.952]	-.0001(.007)[.954]
λ_1	-.0030(.016)	-.0017(.016)[.942]	-.0017(.016)[.932]	-.0115(.009)	-.0095(.009)[.968]	.0003(.010)[.958]
λ_2	.0030(.020)	.0020(.020)[.946]	.0021(.020)[.946]	.0130(.012)	.0105(.012)[.954]	.0000(.012)[.956]
λ_3	-.0426(.071)	-.0082(.067)[.962]	-.0082(.067)[.960]	-.1056(.073)	-.0722(.066)[.894]	-.0297(.100)[.952]
σ_v^2	-.1158(.043)	-.0081(.048)[.944]	–	-.1126(.050)	-.0047(.056)[.964]	–
β	.0040(.014)	.0003(.014)[.944]	.0003(.014)[.944]	.0058(.014)	.0017(.014)[.958]	-.0001(.014)[.950]
ρ	-.0072(.007)	.0001(.007)[.970]	.0001(.007)[.970]	-.0063(.007)	-.0005(.007)[.956]	-.0001(.007)[.956]
λ_1	-.0013(.016)	.0000(.016)[.956]	.0001(.016)[.956]	-.0118(.010)	-.0098(.010)[.928]	.0000(.010)[.934]
λ_2	.0008(.020)	-.0002(.020)[.944]	-.0002(.020)[.946]	.0127(.013)	.0101(.013)[.942]	-.0002(.013)[.948]
λ_3	-.0461(.076)	-.0117(.071)[.958]	-.0115(.071)[.944]	-.1037(.069)	-.0704(.063)[.922]	-.0263(.094)[.946]
σ_v^2	-.1177(.095)	-.0102(.106)[.928]	–	-.1090(.116)	-.0007(.130)[.950]	–
	$T = 10$, $n = 200$; error = 1, 2, for the two panels below					
β	.0060(.011)	.0008(.011)[.930]	.0008(.011)[.932]	.0057(.010)	.0018(.010)[.940]	.0003(.010)[.940]
ρ	-.0087(.006)	.0002(.006)[.962]	.0002(.006)[.962]	-.0075(.005)	-.0010(.005)[.926]	-.0004(.005)[.940]
λ_1	-.0022(.015)	-.0005(.015)[.952]	-.0005(.015)[.954]	-.0107(.007)	-.0085(.007)[.890]	-.0003(.008)[.948]
λ_2	.0025(.018)	.0003(.018)[.948]	.0003(.018)[.948]	.0123(.010)	.0092(.010)[.914]	.0004(.010)[.946]
λ_3	-.0284(.065)	-.0066(.061)[.942]	-.0066(.061)[.946]	-.0590(.050)	-.0504(.047)[.864]	-.0072(.063)[.952]
σ_v^2	-.1081(.029)	-.0037(.033)[.956]	–	-.1044(.038)	-.0001(.042)[.962]	–
β	.0053(.010)	.0001(.010)[.936]	.0001(.010)[.934]	.0055(.009)	.0015(.009)[.960]	.0001(.009)[.980]
ρ	-.0090(.006)	.0000(.006)[.960]	.0000(.006)[.956]	-.0069(.005)	-.0004(.005)[.944]	.0001(.005)[.960]
λ_1	-.0018(.014)	-.0001(.014)[.942]	-.0001(.014)[.946]	-.0108(.007)	-.0086(.007)[.888]	-.0004(.007)[.962]
λ_2	.0026(.018)	.0004(.018)[.942]	.0004(.018)[.938]	.0124(.010)	.0094(.010)[.904]	.0005(.010)[.940]
λ_3	-.0306(.063)	-.0088(.059)[.944]	-.0088(.059)[.942]	-.0596(.048)	-.0509(.045)[.864]	-.0075(.062)[.952]
σ_v^2	-.1058(.074)	-.0011(.082)[.928]	–	-.1032(.087)	.0013(.097)[.938]	–

Note: *error = 1(normal), 2(normal mixture). The spatial weight matrices are constructed under GI-I and GI-II for the homoskedastic and heteroskedastic designs, respectively.*

Table F.3: Empirical bias(sd)[CP95] of `CQMLE`, `M-Est`, and `RM-Est` based on time-varying weight matrices; $(\beta, \rho, \lambda_1, \lambda_2, \lambda_3, \sigma_v^2) = (1, 0.4, 0.4, 0.2, 0.2, 1)$.

	Homoskedasticity			Heteroskedasticity		
	CQMLE	M-Est	RM-Est	CQMLE	M-Est	RM-Est
	$T = 5$, $n = 50$; error = 1, 2, for the two panels below					
β	.0099(.030)	.0006(.030)[.940]	.0005(.030)[.938]	.0085(.028)	.0019(.027)[.987]	.0003(.027)[.959]
ρ	-.0257(.022)	.0004(.023)[.932]	.0007(.023)[.930]	-.0159(.015)	.0003(.015)[.976]	.0004(.016)[.955]
λ_1	-.0086(.050)	-.0048(.051)[.944]	-.0049(.051)[.936]	-.0145(.026)	-.0088(.026)[.972]	-.0028(.026)[.950]
λ_2	.0050(.059)	.0050(.060)[.948]	.0051(.060)[.932]	.0152(.037)	.0118(.036)[.957]	.0044(.037)[.957]
λ_3	-.1044(.175)	-.0491(.151)[.956]	-.0491(.154)[.954]	-.2397(.175)	-.1351(.136)[.989]	-.1144(.274)[.920]
σ_v^2	-.2426(.076)	-.0269(.098)[.918]	–	-.2349(.104)	-.0170(.134)[.937]	–
β	.0105(.030)	.0012(.031)[.948]	.0012(.030)[.934]	.0061(.029)	-.0003(.029)[.972]	-.0018(.029)[.955]
ρ	-.0268(.024)	-.0007(.024)[.926]	-.0008(.024)[.922]	-.0156(.016)	.0006(.016)[.957]	.0005(.016)[.942]
λ_1	-.0049(.053)	-.0014(.054)[.918]	-.0013(.054)[.904]	-.0099(.027)	-.0046(.027)[.976]	.0009(.028)[.944]
λ_2	.0009(.061)	.0011(.063)[.914]	.0009(.063)[.898]	.0078(.038)	.0049(.037)[.970]	-.0020(.038)[.953]
λ_3	-.1176(.189)	-.0599(.162)[.944]	-.0601(.163)[.932]	-.2132(.158)	-.1152(.124)[.998]	-.0780(.243)[.942]
σ_v^2	-.2328(.173)	-.0137(.224)[.896]	–	-.2442(.228)	-.0296(.295)[.827]	–
	$T = 5$, $n = 100$; error = 1, 2, for the two panels below					
β	.0046(.022)	.0003(.022)[.938]	.0002(.022)[.936]	.0091(.025)	.0019(.025)[.932]	.0014(.025)[.928]
ρ	-.0202(.013)	-.0002(.014)[.952]	-.0002(.014)[.940]	-.0182(.013)	.0004(.013)[.944]	.0001(.013)[.944]
λ_1	.0055(.042)	.0007(.044)[.940]	.0006(.044)[.938]	-.0138(.021)	-.0071(.020)[.988]	-.0010(.022)[.956]
λ_2	-.0112(.052)	-.0003(.054)[.944]	-.0003(.055)[.940]	.0146(.026)	.0081(.026)[.990]	.0009(.027)[.942]
λ_3	-.0916(.136)	-.0422(.121)[.950]	-.0423(.122)[.948]	-.1049(.116)	-.0789(.099)[.954]	-.0404(.146)[.952]
σ_v^2	-.2239(.055)	-.0142(.071)[.928]	–	-.2178(.072)	-.0077(.092)[.950]	–
β	.0042(.023)	-.0003(.023)[.940]	-.0003(.023)[.946]	.0074(.021)	.0002(.021)[.958]	-.0003(.021)[.962]
ρ	-.0200(.013)	.0003(.014)[.940]	.0002(.014)[.948]	-.0177(.012)	.0008(.012)[.966]	.0005(.012)[.976]
λ_1	.0016(.041)	-.0036(.044)[.950]	-.0039(.044)[.938]	-.0138(.020)	-.0074(.020)[.992]	-.0012(.021)[.954]
λ_2	-.0075(.052)	.0040(.056)[.938]	.0043(.056)[.946]	.0147(.025)	.0085(.025)[.982]	.0013(.027)[.960]
λ_3	-.0811(.149)	-.0325(.135)[.922]	-.0317(.136)[.918]	-.1017(.118)	-.0762(.101)[.950]	-.0359(.154)[.946]
σ_v^2	-.2253(.124)	-.0159(.159)[.876]	–	-.2180(.151)	-.0078(.192)[.914]	–
	$T = 10$, $n = 100$; error = 1, 2, for the two panels below					
β	.0025(.013)	-.0005(.013)[.944]	-.0005(.013)[.944]	.0050(.014)	.0007(.014)[.934]	-.0001(.014)[.928]
ρ	-.0046(.006)	.0004(.006)[.950]	.0004(.006)[.948]	-.0068(.007)	-.0004(.007)[.946]	-.0002(.007)[.952]
λ_1	-.0024(.026)	-.0008(.026)[.926]	-.0008(.026)[.922]	-.0110(.014)	-.0063(.014)[.976]	.0005(.014)[.940]
λ_2	.0027(.029)	.0007(.030)[.926]	.0007(.030)[.916]	.0135(.018)	.0068(.018)[.970]	-.0008(.018)[.936]
λ_3	-.0510(.079)	-.0163(.074)[.952]	-.0164(.074)[.952]	-.0987(.074)	-.0694(.067)[.874]	-.0267(.095)[.946]
σ_v^2	-.1122(.042)	-.0040(.047)[.946]	–	-.1130(.051)	-.0048(.057)[.952]	–
β	.0035(.013)	.0005(.013)[.932]	.0005(.013)[.934]	.0052(.013)	.0009(.013)[.954]	.0001(.013)[.962]
ρ	-.0050(.006)	-.0001(.006)[.950]	-.0001(.006)[.960]	-.0070(.007)	-.0005(.007)[.924]	-.0003(.007)[.952]
λ_1	-.0018(.024)	-.0003(.024)[.954]	-.0003(.024)[.962]	-.0117(.013)	-.0069(.013)[.984]	-.0002(.014)[.956]
λ_2	.0027(.027)	.0007(.027)[.960]	.0007(.027)[.962]	.0147(.017)	.0079(.017)[.976]	.0004(.017)[.956]
λ_3	-.0498(.076)	-.0150(.071)[.958]	-.0146(.071)[.956]	-.0963(.075)	-.0673(.068)[.890]	-.0233(.096)[.930]
σ_v^2	-.1154(.096)	-.0076(.108)[.926]	–	-.1157(.116)	-.0078(.131)[.924]	–
	$T = 10$, $n = 200$; error = 1, 2, for the two panels below					
β	.0043(.010)	.0005(.010)[.936]	.0005(.010)[.934]	.0044(.009)	.0006(.009)[.938]	-.0001(.009)[.938]
ρ	-.0067(.005)	-.0002(.005)[.956]	-.0002(.005)[.956]	-.0067(.005)	-.0006(.005)[.928]	-.0004(.005)[.942]
λ_1	-.0046(.018)	-.0004(.018)[.938]	-.0004(.018)[.936]	-.0107(.008)	-.0070(.008)[.966]	.0000(.008)[.952]
λ_2	.0070(.021)	.0008(.021)[.936]	.0008(.021)[.932]	.0132(.011)	.0085(.011)[.962]	.0002(.011)[.954]
λ_3	-.0246(.061)	-.0063(.057)[.946]	-.0062(.057)[.952]	-.0840(.049)	-.0748(.045)[.724]	-.0098(.069)[.946]
σ_v^2	-.1081(.031)	-.0038(.034)[.932]	–	-.1053(.038)	-.0006(.042)[.958]	–
β	.0040(.009)	.0002(.009)[.952]	.0002(.009)[.946]	.0051(.009)	.0013(.009)[.934]	.0006(.009)[.930]
ρ	-.0067(.005)	-.0002(.005)[.948]	-.0002(.005)[.934]	-.0067(.004)	-.0006(.004)[.952]	-.0003(.004)[.956]
λ_1	-.0029(.018)	.0013(.018)[.912]	.0013(.018)[.912]	-.0107(.009)	-.0071(.009)[.964]	.0001(.009)[.948]
λ_2	.0052(.021)	-.0011(.021)[.920]	-.0011(.021)[.924]	.0132(.012)	.0086(.012)[.960]	.0001(.012)[.936]
λ_3	-.0300(.061)	-.0116(.058)[.956]	-.0118(.058)[.954]	-.0861(.050)	-.0766(.046)[.694]	-.0126(.071)[.946]
σ_v^2	-.1064(.070)	-.0019(.078)[.936]	–	-.1004(.091)	.0049(.101)[.930]	–

Note: *error = 1(normal), 2(normal mixture). The spatial weight matrices are constructed under GI-I and GI-II for the homoskedastic and heteroskedastic designs, respectively.*

Table F.4: Empirical bias(sd)[CP95] of `CQMLE`, `M-Yang`, `GMM`, `M-Est`, and `RM-Est` based on time-constant weight matrix; $(\beta, \rho, \lambda_1, \lambda_2, \sigma_v^2) = (1, 0.3, 0.2, 0.2, 1)$; Homoskedasticity.

	CQMLE	M-Yang	GMM	M-Est	RM-Est
			$\boldsymbol{T=5}$, $\boldsymbol{n=50}$; **error = 1, 2, for the two panels below**		
β	.0166(.033)	.0152(.033)	.0056(.074)	.0016(.033)[.964]	.0016(.033)[.950]
ρ	-.0351(.027)	-.0355(.027)	-.0120(.157)	-.0010(.028)[.940]	-.0009(.029)[.944]
λ_1	-.0283(.057)	-.0055(.063)	-.0009(.083)	-.0102(.056)[.948]	-.0101(.056)[.948]
λ_2	.0350(.058)	.0184(.062)	-.0034(.216)	.0084(.059)[.962]	.0083(.059)[.956]
σ_v^2	-.2455(.077)	-.0577(.096)	–	-.0297(.100)[.914]	–
β	.0149(.034)	.0135(.034)	-.0046(.074)	-.0003(.034)[.946]	-.0003(.034)[.926]
ρ	-.0356(.029)	-.0360(.029)	.0063(.154)	-.0012(.029)[.932]	-.0013(.030)[.930]
λ_1	-.0272(.062)	-.0044(.069)	.0073(.088)	-.0091(.060)[.932]	-.0090(.060)[.938]
λ_2	.0356(.065)	.0189(.068)	-.0338(.218)	.0087(.065)[.910]	.0086(.065)[.926]
σ_v^2	-.2405(.170)	-.0515(.212)	–	-.0229(.221)[.854]	–
			$\boldsymbol{T=5}$, $\boldsymbol{n=100}$; **error = 1, 2, for the two panels below**		
β	.0098(.024)	.0107(.024)	-.0026(.035)	-.0005(.024)[.936]	-.0005(.024)[.934]
ρ	-.0364(.020)	-.0366(.020)	.0032(.078)	-.0006(.021)[.934]	-.0008(.021)[.940]
λ_1	-.0333(.058)	-.0002(.068)	.0122(.072)	-.0067(.058)[.954]	-.0066(.058)[.942]
λ_2	.0295(.050)	.0142(.051)	-.0251(.122)	.0024(.049)[.946]	.0024(.049)[.950]
σ_v^2	-.2229(.054)	-.0294(.067)	–	-.0103(.069)[.944]	–
β	.0087(.024)	.0095(.024)	-.0020(.035)	-.0016(.024)[.944]	-.0015(.024)[.944]
ρ	-.0347(.020)	-.0349(.020)	.0034(.082)	.0007(.020)[.930]	.0007(.020)[.924]
λ_1	-.0369(.060)	-.0047(.070)	.0118(.082)	-.0103(.059)[.946]	-.0103(.059)[.950]
λ_2	.0313(.049)	.0163(.051)	-.0067(.118)	.0043(.049)[.932]	.0043(.049)[.938]
σ_v^2	-.2292(.117)	-.0373(.146)	–	-.0184(.150)[.914]	–
			$\boldsymbol{T=10}$, $\boldsymbol{n=100}$; **error = 1, 2, for the two panels below**		
β	.0100(.016)	.0096(.016)	.0004(.028)	-.0001(.016)[.938]	-.0001(.016)[.940]
ρ	-.0170(.012)	-.0174(.012)	-.0025(.038)	-.0007(.012)[.944]	-.0008(.012)[.946]
λ_1	-.0124(.033)	.0158(.037)	.0292(.056)	-.0005(.033)[.944]	-.0005(.033)[.940]
λ_2	.0152(.037)	-.0076(.040)	-.0543(.107)	.0016(.037)[.926]	.0016(.037)[.930]
σ_v^2	-.1181(.043)	-.0209(.047)	–	-.0085(.048)[.942]	–
β	.0109(.015)	.0106(.015)	.0013(.026)	.0008(.015)[.954]	.0008(.015)[.956]
ρ	-.0169(.012)	-.0173(.012)	-.0023(.036)	-.0007(.012)[.952]	-.0007(.012)[.946]
λ_1	-.0163(.033)	.0115(.037)	.0242(.055)	-.0044(.033)[.960]	-.0043(.033)[.956]
λ_2	.0177(.035)	-.0047(.038)	-.0484(.106)	.0041(.035)[.948]	.0040(.035)[.952]
σ_v^2	-.1150(.094)	-.0174(.104)	–	-.0050(.105)[.930]	–
			$\boldsymbol{T=10}$, $\boldsymbol{n=200}$; **error = 1, 2, for the two panels below**		
β	.0093(.010)	.0084(.010)	-.0007(.024)	-.0003(.010)[.958]	-.0003(.010)[.954]
ρ	-.0153(.008)	-.0152(.008)	-.0002(.037)	.0000(.008)[.952]	.0000(.008)[.950]
λ_1	-.0095(.024)	.0092(.025)	.0201(.046)	-.0019(.023)[.956]	-.0019(.023)[.958]
λ_2	.0151(.026)	-.0009(.027)	-.0407(.094)	.0026(.026)[.938]	.0026(.026)[.934]
σ_v^2	-.1100(.029)	-.0114(.032)	–	-.0043(.033)[.952]	–
β	.0101(.010)	.0092(.010)	-.0001(.025)	.0005(.010)[.936]	.0005(.010)[.932]
ρ	-.0161(.008)	-.0160(.008)	-.0007(.037)	-.0009(.008)[.942]	-.0009(.008)[.944]
λ_1	-.0076(.023)	.0112(.025)	.0226(.044)	-.0001(.022)[.950]	-.0001(.022)[.950]
λ_2	.0130(.026)	-.0030(.027)	-.0445(.088)	.0006(.026)[.956]	.0006(.026)[.952]
σ_v^2	-.1118(.073)	-.0134(.081)	–	-.0063(.082)[.928]	–

Note: *error = 1(normal), 2(normal mixture). The spatial weight matrices are constructed under GI-III.* `M-Yang` *denotes the M-estimator of Yang (2018), and* `GMM` *denotes the GMM estimator of Lee and Yu (2014), which combines linear IV moments and quadratic moments and employs an estimated optimal weighting matrix.*

Table F.5: Empirical bias(sd)[CP95] of `CQMLE`, `RM-LY`, `M-Est`, and `RM-Est` based on time-constant weight matrix; $(\beta, \rho, \lambda_1, \lambda_2, \lambda_3, \sigma_v^2) = (1, 0.3, 0.2, 0.2, 0.2, 1)$; Heteroskedasticity.

	CQMLE	RM-LY	M-Est	RM-Est
	$T = 5$, $n = 50$; error = 1, 2, for the two panels below			
β	.0117(.031)	-.0022(.031)	-.0011(.031)[.968]	-.0019(.031)[.950]
ρ	-.0351(.028)	.0017(.033)	.0007(.028)[.942]	.0011(.029)[.932]
λ_1	-.0162(.067)	-.0057(.078)	-.0078(.065)[.970]	-.0055(.075)[.898]
λ_2	.0200(.076)	.0028(.088)	.0077(.076)[.938]	.0026(.081)[.912]
λ_3	-.2219(.178)	-.2630(.268)	-.1336(.162)[.978]	-.1376(.238)[.926]
σ_v^2	-.2553(.092)	-.0612(.119)	-.0398(.120)[.916]	–
β	.0130(.031)	-.0018(.035)	.0000(.031)[.968]	-.0005(.031)[.952]
ρ	-.0365(.029)	.0026(.049)	-.0004(.029)[.946]	-.0006(.029)[.936]
λ_1	-.0080(.067)	.0028(.081)	.0006(.065)[.972]	.0042(.075)[.900]
λ_2	.0112(.077)	-.0044(.095)	-.0014(.077)[.942]	-.0075(.081)[.924]
λ_3	-.2316(.174)	-.2783(.283)	-.1429(.158)[.982]	-.1513(.244)[.916]
σ_v^2	-.2402(.205)	-.0353(.271)	-.0196(.267)[.868]	–
	$T = 5$, $n = 100$; error = 1, 2, for the two panels below			
β	.0103(.024)	.0000(.025)	.0001(.024)[.946]	-.0004(.024)[.936]
ρ	-.0356(.019)	-.0010(.023)	.0002(.020)[.956]	.0004(.020)[.944]
λ_1	-.0092(.042)	-.0018(.049)	-.0002(.041)[.996]	-.0017(.048)[.946]
λ_2	.0156(.039)	.0020(.045)	.0020(.040)[.986]	.0000(.042)[.934]
λ_3	-.1023(.109)	-.0906(.148)	-.0686(.105)[.944]	-.0405(.140)[.930]
σ_v^2	-.2258(.063)	-.0241(.081)	-.0145(.081)[.970]	–
β	.0112(.024)	.0009(.026)	.0010(.024)[.942]	.0005(.024)[.942]
ρ	-.0364(.020)	-.0015(.034)	-.0009(.020)[.948]	-.0003(.019)[.946]
λ_1	-.0092(.040)	-.0005(.046)	-.0002(.040)[.996]	-.0009(.045)[.944]
λ_2	.0140(.037)	-.0004(.045)	.0001(.036)[.986]	-.0026(.038)[.960]
λ_3	-.1077(.108)	-.0995(.148)	-.0738(.103)[.952]	-.0471(.138)[.944]
σ_v^2	-.2276(.139)	-.0227(.179)	-.0165(.179)[.908]	–
	$T = 10$, $n = 100$; error = 1, 2, for the two panels below			
β	.0106(.015)	.0009(.017)	.0007(.015)[.952]	.0005(.015)[.938]
ρ	-.0160(.011)	-.0004(.017)	.0000(.011)[.956]	.0000(.011)[.958]
λ_1	-.0027(.027)	.0008(.032)	-.0001(.027)[.992]	.0017(.030)[.944]
λ_2	.0050(.028)	.0006(.038)	.0012(.028)[.980]	-.0018(.030)[.942]
λ_3	-.0885(.067)	-.0652(.090)	-.0510(.064)[.942]	-.0170(.083)[.960]
σ_v^2	-.1166(.051)	-.0158(.057)	-.0075(.058)[.956]	–
β	.0104(.016)	.0008(.021)	.0004(.016)[.932]	.0003(.016)[.932]
ρ	-.0166(.012)	-.0013(.025)	-.0005(.012)[.946]	-.0005(.012)[.932]
λ_1	-.0048(.028)	-.0009(.033)	-.0022(.028)[.978]	-.0006(.031)[.942]
λ_2	.0062(.029)	.0006(.039)	.0024(.029)[.976]	-.0004(.031)[.940]
λ_3	-.0874(.068)	-.0653(.091)	-.0500(.065)[.938]	-.0158(.084)[.958]
σ_v^2	-.1140(.123)	-.0105(.139)	-.0045(.138)[.924]	–
	$T = 10$, $n = 200$; error = 1, 2, for the two panels below			
β	.0092(.010)	-.0002(.012)	-.0001(.010)[.940]	-.0002(.010)[.942]
ρ	-.0153(.008)	-.0001(.012)	-.0002(.008)[.928]	-.0002(.008)[.932]
λ_1	-.0029(.017)	.0002(.019)	-.0010(.017)[.996]	.0000(.019)[.944]
λ_2	.0051(.018)	-.0015(.023)	.0011(.018)[.984]	-.0011(.019)[.950]
λ_3	-.0600(.046)	-.0303(.058)	-.0424(.045)[.890]	-.0067(.056)[.970]
σ_v^2	-.1077(.034)	-.0071(.038)	-.0023(.038)[.978]	–
β	.0098(.010)	.0008(.014)	.0005(.010)[.944]	.0004(.010)[.946]
ρ	-.0159(.008)	-.0015(.018)	-.0009(.008)[.952]	-.0009(.008)[.966]
λ_1	-.0034(.018)	-.0005(.019)	-.0015(.018)[.984]	-.0005(.019)[.946]
λ_2	.0070(.019)	.0010(.024)	.0029(.019)[.984]	.0007(.020)[.950]
λ_3	-.0602(.047)	-.0309(.060)	-.0426(.046)[.892]	-.0070(.058)[.960]
σ_v^2	-.1073(.088)	-.0054(.099)	-.0018(.099)[.932]	–

Note: *error = 1(normal), 2(normal mixture). The spatial weight matrices are constructed under GI-IV.* `RM-LY` *denotes the heteroskedasticity robust estimator of Li and Yang (2020).*

Appendix G: A Computational Note

The formal expressions of the main results involve several $nT \times nT$ matrices. In case of a large panel, direct calculations would hinder the applicability of the proposed methods. This note explains how computations involving large matrices are avoided.

First, note that our methods do not involve the computation of the determinant of a matrix, and that the matrices $\mathbf{B}_1$ and $\mathbf{B}_3$ are block diagonal, and thus their inverse calculations reduce to those of $n \times n$ matrices. The remaining computations involving the $nT \times nT$ matrices are performed through operator-based, recursive, and block-wise computations, and thus explicit construction and storage of large dense matrices are avoided. These computational tools are implemented in Matlab and the codes for Monte Carlo simulation are made available online at http://www.mysmu.edu/faculty/zlyang/SubPages/publication.htm.

G.1 Computation of Matrix Products and Diagonal Terms

We first describe how the matrix products and diagonal terms appearing in the estimating equations (2.6) and (3.3) are computed without explicitly constructing $\mathbf{M}$ or other dense $nT \times nT$ matrices. The key idea is to treat $\mathbf{M}$ as a linear operator. For any vector or thin matrix Z, $\mathbf{M}Z = Z - \mathbf{B}_3\mathbf{D}\big[\big(\mathbf{D}'\mathbf{B}_3'\mathbf{B}_3\mathbf{D}\big)^{-1}\mathbf{D}'\mathbf{B}_3'Z\big]$. Thus, the residualized quantity $\mathbf{M}Z$ is computed using only matrix-vector or thin-matrix products together with the inverse of $\mathbf{D}'\mathbf{B}_3'\mathbf{B}_3\mathbf{D}$, whose dimension is $(n+T-1) \times (n+T-1)$. The direct inversion is not needed as the desired computation is of the form $x = A^{-1}B$, which can be handled efficiently by solving the linear systems $Ax = B$; see Section G.2 for details. Full matrix multiplications are also avoided due to the block-diagonal structure of $\mathbf{B}_3$.

The robust estimating equations in (3.3) additionally involve the calculation of $\mathbf{R}_r\mathbf{Z}^*$, for $r = \rho, \lambda_1, \lambda_2, \lambda_3$, where $\mathbf{R}_r = \mathtt{diag}(\Phi_r)\mathtt{diag}(\mathbf{M})^{-1}$. To obtain these diagonal entries, the identity matrix is partitioned into blocks of canonical basis vectors. The matrix $\mathbf{M}$ and the matrix operations defining Φ_r are then applied sequentially to each block, taking advantage of the block-diagonal structure of $\mathbf{B}_3$ and $\mathbf{W}_s$, $s = 1, 2, 3$, together with the lower block-triangular structure of $\mathbf{Q}$ and $\mathbf{Q}_{-1}$. The corresponding diagonal elements are extracted and accumulated across blocks. Repeating this procedure over all blocks produces all diagonal entries without constructing any dense $nT \times nT$ matrix.

G.2 Computation of Generalized Inverse

Next, we describe how the generalized inverse of $\mathbf{M} \odot \mathbf{M}$ involved in (3.4) is avoided. Instead of explicitly applying $(\mathbf{M} \odot \mathbf{M})^-$ to $(\mathbf{MZ}^*) \odot (\mathbf{MZ}^*)$, we solve the linear system

$$(\mathbf{M} \odot \mathbf{M})x = \mathbf{MZ}^* \odot \mathbf{MZ}^*.$$

The system is solved by an iterative Krylov-subspace method implemented using `Matlab`'s sparse iterative solvers. In addition, we have $(\mathbf{M}\odot\mathbf{M})x = (1-2p)\odot x + \mathtt{diag}\big[\mathbf{P}\mathtt{diag}(x)\mathbf{P}\big]$, where $p = \mathtt{diag}(\mathbf{P})$ and $\mathbf{P} = \mathbf{B}_3\mathbf{D}\big(\mathbf{D}'\mathbf{B}_3'\mathbf{B}_3\mathbf{D}\big)^{-1}\mathbf{D}'\mathbf{B}_3'$. The vector p and the diagonal term $\mathtt{diag}\big[\mathbf{P}\mathtt{diag}(x)\mathbf{P}\big]$ are then calculated directly from the thin matrix $\mathbf{B}_3\mathbf{D}$ and the reduced-dimensional matrix $\mathbf{D}'\mathbf{B}_3'\mathbf{B}_3\mathbf{D}$. Consequently, each iteration of the iterative solver only involves operations on thin matrices and vectors.

G.3 Computation of Trace Terms

We first consider trace terms involving $\mathbf{M}$. As $\mathbf{M} = I_{nT} - \mathbf{P}$, we have, for any conformable matrix $\mathbf{C}$, $\mathtt{tr}(\mathbf{MC}) = \mathtt{tr}(\mathbf{C}) - \mathtt{tr}(\mathbf{PC})$. The second term can be reduced by cyclic permutation: $\mathtt{tr}(\mathbf{PC}) = \mathtt{tr}\Big[\big(\mathbf{D}'\mathbf{B}_3'\mathbf{B}_3\mathbf{D}\big)^{-1}\mathbf{D}'\mathbf{B}_3'\mathbf{C}\mathbf{B}_3\mathbf{D}\Big]$, which only involves matrices of dimension $(n+T-1)\times(n+T-1)$ after applying $\mathbf{C}$ to the thin matrix $\mathbf{B}_3\mathbf{D}$. The remaining computational issue is the evaluation of $\mathtt{tr}(\mathbf{C})$, which is obtained by accumulating the diagonals of $\mathbf{C}$ using the diagonal-extraction procedure in G.1.

Moreover, the robust variance correction terms in (3.5) involving Ξ are evaluated using repeated applications of the iterative procedure described in G.2, so that the generalized inverse $(\mathbf{M}\odot\mathbf{M})^-$ is not explicitly computed. The associated trace terms are subsequently obtained using the diagonal-extraction procedure described in G.1.

These implementation strategies substantially reduce memory usage and computational cost, making the proposed methods feasible for large-scale applications. To illustrate the computational burden, we report the elapsed time for one replication of the largest time-varying design, $(n, T) = (200, 40)$, on a personal computer equipped with a 13th-generation Intel Core i7-13700H processor, 64 GB of RAM, and MATLAB R2025b. The estimation and inference steps require approximately 26 and 32 seconds, respectively, for `M-Est`, and 67 and 90 seconds, respectively, for `RM-Est`.