HIGH-DIMENSIONAL-RESPONSES-ASSISTED HETEROGENEOUS NODAL INFLUENCE ANALYSIS

Dongxue Zhang, Wei Lan, Danyang Huang and Huazhen Lin

Southwestern University of Finance and Economics, Renmin University of China

Supplementary Material

This supplementary material includes six sections. Section S1 discusses the invalidity of ML estimation method. Section S2 gives the detailed interpretation of Condition (C4). Section S3 presents the proofs of Theorems 1, 2, and 3. Section S4 introduces a novel link function test to check the adequacy of the pre-specified link function. Section S5 provides additional simulation studies and empirical studies. Section S6 illustrates the details of variance designs V1 and V2.

S1 The Invalidity of ML Estimation Method

In this section, we discuss the general invalidity of traditional ML estimation method in the presence of unknown heteroscedastic errors. For Model (2.3), the log-likelihood function under the assumption of homogeneous error variances, σ^2, can be written as

$$\ell(\delta) = -\frac{mn}{2}\log(2\pi) - \frac{mn}{2}\log\sigma^2 - \frac{1}{2\sigma^2}e(\theta)^\top e(\theta) + \log\left|\det\left\{S(\gamma)\right\}\right|,$$

where $\delta = (\gamma^\top, \beta^\top, \sigma^2)^\top \in \mathbb{R}^{d+p+1}$, $\theta = (\gamma^\top, \beta^\top)^\top \in \mathbb{R}^{d+p}$, $p = p_1 + p_2$, $S(\gamma) = I_{mn} - \{V^\top \otimes \bar{U}(\gamma)\}$, and $e(\theta) = S(\gamma)y - x\beta$. Given γ, the maximum likelihood estimator (MLE) of β and σ^2 are $\widehat{\beta}_{ml}(\gamma) = (x^\top x)^{-1}x^\top S(\gamma)y$ and $\widehat{\sigma}^2_{ml}(\gamma) = \frac{1}{mn}\{S(\gamma)y - x\widehat{\beta}_{ml}(\gamma)\}^\top\{S(\gamma)y - x\widehat{\beta}_{ml}(\gamma)\}$, respectively. Subsequently, we can get the concentrated log-likelihood function of γ as

$$\ell_c(\gamma) = -\frac{mn}{2}\log(2\pi) - \frac{mn}{2}\log\widehat{\sigma}^2_{ml}(\gamma) - \frac{mn}{2} + \log\left|\det\{S(\gamma)\}\right|.$$

Let $\Lambda_{\gamma_k} := \partial\Lambda(\gamma)/\partial\gamma_k = \text{diag}\{z_{1k}F'(Z_1^\top\gamma), \cdots, z_{mk}F'(Z_m^\top\gamma)\}$, $S_{\gamma_k}(\gamma) := \partial S(\gamma)/\partial\gamma_k = -V^\top \otimes \bar{U}_{\gamma_k}$, where $F'(\cdot)$ is the first order derivative of F, and $\bar{U}_{\gamma_k} = U\Lambda_{\gamma_k}$. Accordingly, the first order condition for the concentrated log-likelihood function is

$$\frac{\partial\ell_c(\gamma)}{\partial\gamma_k} = -\frac{1}{\widehat{\sigma}^2_{ml}(\gamma)}y^\top S_{\gamma_k}(\gamma)^\top MS(\gamma)y + \text{tr}\left\{S(\gamma)^{-1}S_{\gamma_k}(\gamma)\right\},$$

where $M = I_{mn} - x(x^\top x)^{-1}x^\top$ and $k = 1, \cdots, d$.

Recall that $\theta_0 = (\gamma_0^\top, \beta_0^\top)^\top$ is the true parameter, and $e = S(\gamma_0)y - x\beta_0$. Under the assumption of heterogeneous error variances, we have,

$$\widehat{\sigma}^2_{ml}(\gamma_0) \widehat{=} \frac{1}{mn}\left\{S(\gamma_0)y - x\widehat{\beta}_{ml}(\gamma_0)\right\}^\top\left\{S(\gamma_0)y - x\widehat{\beta}_{ml}(\gamma_0)\right\} = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\sigma_{ij}^2 + o_p(1).$$

Define $G_{\gamma_k}(\gamma) = S_{\gamma_k}(\gamma)S(\gamma)^{-1}$, and denote $G_{\gamma_k} = G_{\gamma_k}(\gamma_0)$ and $S_{\gamma_k} =$

$S_{\gamma_k}(\gamma_0)$ for simplicity. Then we have,

$$\begin{aligned}\frac{1}{mn}\mathrm{E}\left\{\frac{\partial \ell_c(\gamma_0)}{\partial \gamma_k}\right\} &= -\frac{1}{\sum_{i=1}^m \sum_{j=1}^n \sigma_{ij}^2}\mathrm{E}\left\{(x\beta_0+e)^\top G_{\gamma_k}^\top M(x\beta_0+e)\right\} + \frac{1}{mn}\mathrm{tr}\left(G_{\gamma_k}\right) \\ &= \frac{1}{mn\bar{\sigma}^2}\sum_{i=1}^m \sum_{j=1}^n \left(G_{\gamma_k,m(j-1)+i,m(j-1)+i} - \bar{G}_{\gamma_k}\right)\left(\sigma_{ij}^2 - \bar{\sigma}^2\right) \\ &= \frac{1}{\bar{\sigma}^2}\mathrm{Cov}\left(G_{\gamma_k,m(j-1)+i,m(j-1)+i}, \sigma_{ij}^2\right),\end{aligned}$$

where $\bar{G}_{\gamma_k} = \frac{1}{mn}\mathrm{tr}(G_{\gamma_k}) = \frac{1}{mn}\sum_{i=1}^m \sum_{j=1}^n G_{\gamma_k,m(j-1)+i,m(j-1)+i}$, and $\bar{\sigma}^2 = \frac{1}{mn}\sum_{i=1}^m \sum_{j=1}^n \sigma_{ij}^2$.

The consistency of the MLE $\widehat{\gamma}_{k,ml}$ requires that $\frac{1}{mn}\mathrm{E}\{\partial \ell_c(\gamma_0)/\partial \gamma_k\}$ converges to zero as $mn \to \infty$. This holds if and only if the covariance between the diagonal elements of matrix G_{γ_k}, $G_{\gamma_k,m(j-1)+i,m(j-1)+i}$, and the error variances σ_{ij}^2 is zero as $mn \to \infty$. Hence, apart from the computational burden it imposes, the MLE for the MNIM with unknown heteroscedastic errors is inconsistent when the diagonal elements of matrix G_{γ_k} are not all equal.

S2 A Detailed Interpretation of Condition (C4)

Condition (C4) ensures the identification of θ_0 from the moment equations $\mathrm{E}\{g(\theta)\} = 0$ for a sufficiently large mn. For any possible value θ, the moment conditions take the following form:

$$E\{g(\theta)\} = \begin{pmatrix} d(\theta)^\top P_1 d(\theta) + \mathrm{tr}\left\{\Sigma S^{-1\top}(\mathcal{K}_0 - \mathcal{K})^\top W^\top P_1^s\right\} + \mathrm{tr}\left\{\Sigma S^{-1\top}(\mathcal{K}_0 - \mathcal{K})^\top W^\top P_1 W(\mathcal{K}_0 - \mathcal{K})S^{-1}\right\} \\ \vdots \\ d(\theta)^\top P_L d(\theta) + \mathrm{tr}\left\{\Sigma S^{-1\top}(\mathcal{K}_0 - \mathcal{K})^\top W^\top P_L^s\right\} + \mathrm{tr}\left\{\Sigma S^{-1\top}(\mathcal{K}_0 - \mathcal{K})^\top W^\top P_L W(\mathcal{K}_0 - \mathcal{K})S^{-1}\right\} \\ Q^\top d(\theta) \end{pmatrix}, \tag{S2.1}$$

where $d(\theta) = W(\mathcal{K}_0 - \mathcal{K})S^{-1}x\beta_0 + x(\beta_0 - \beta)$, $\mathcal{K} = I_n \otimes \Lambda(\gamma)$, $\mathcal{K}_0 = I_n \otimes \Lambda(\gamma_0)$, $W = V^\top \otimes U$, $S = S(\gamma_0)$. To ensure identification, the limiting moment equations must have a unique solution at θ_0, i.e.,

$$\lim_{mn\to\infty} \frac{1}{mn} \mathrm{E}\{g(\theta)\} = 0 \text{ if and only if } \theta = \theta_0.$$

We next elaborate on the two alternative conditions in C4.2, respectively, each of which is sufficient on its own for identification.

Firstly, C4.2(a) ensures the identification through the linear moment conditions corresponding to Q in (S2.1). Specifically, $\lim_{mn\to\infty} \frac{1}{mn} Q^\top d(\theta) = \lim_{mn\to\infty} \frac{1}{mn}(Q^\top W(\mathcal{K}_0 - \mathcal{K})S^{-1}x\beta_0 + Q^\top x(\beta_0 - \beta)) = 0$. They will uniquely identify θ_0 if the Jacobian $\partial\{Q^\top d(\theta)\}/\partial\theta^\top$ has full column rank $(d+p)$, for large enough mn. This rank condition implies the necessary condition that the matrix $(G_{\gamma_1}x\beta_0, \cdots, G_{\gamma_d}x\beta_0, x)$ has full column rank $(d+p)$ and that Q has at least rank $(d+p)$, for large enough mn. Under these circumstances, the parameter θ_0 can be identified from the linear moment equations alone.

Additionally, C4.2(b) ensures the identification through variation in nonlinear components. This condition addresses situations where the linear moment equations may not be informative enough–for example, if $\beta_0 = 0$, then the terms $G_{\gamma_k} x\beta_0$ vanish, and the Jacobian from C4.2(a) may lose rank. More generally, if $(G_{\gamma_1} x\beta_0, \cdots, G_{\gamma_d} x\beta_0, x)$ is not full rank, say it only has rank p, then γ_0 cannot be identified from the linear moments alone. In such cases, the identification of γ_0 must rely on the remaining quadratic moment equations in (S2.1): $\lim_{mn\to\infty} \frac{1}{mn}\Big[\mathrm{tr}\left\{\Sigma S^{-1\top}(\mathcal{K}_0 - \mathcal{K})^\top W^\top P_l^s\right\} + \mathrm{tr}\left\{\Sigma S^{-1\top}(\mathcal{K}_0 - \mathcal{K})^\top W^\top P_l W(\mathcal{K}_0 - \mathcal{K}) S^{-1}\right\}\Big] = 0$ for $l = 1, \cdots, L$. For identification, it is sufficient that the Jacobian of these quadratic terms with respect to γ has full column rank d for at least one l. This is implied that, for some l, the vector $(\mathrm{tr}\{\Sigma P_l^s G_{\gamma_1}\}, \cdots, \mathrm{tr}\{\Sigma P_l^s G_{\gamma_d}\})$ has rank d.

S3 Proofs of Theorems

S3.1 Proof of Theorem 1

In this section, we prove Theorem 1 in two steps: (1) demonstrating the consistency of $\widehat{\theta}$, as $mn \to \infty$ and (2) verifying that $\widehat{\theta}$ is asymptotically normal.

STEP I. To prove the consistency of an extremum estimate, we follow the techniques of Lin and Lee (2010) to show that $\frac{1}{mn}ag(\theta)$ converges in probability to $\frac{1}{mn}a\mathrm{E}\{g(\theta)\}$ uniformly in $\theta \in \Theta$ as $mn \to \infty$. Let $a = (a_1, \cdots, a_L, a_x)$, where a_j is the j-th column of the matrix and a_x is a submatrix. Let $a^{(i)} = (a_1^{(i)}, \cdots, a_L^{(i)}, a_x^{(i)})$ be the i-th row of the matrix, where $a_1^{(i)}, \cdots, a_L^{(i)}$ are scalars and $a_x^{(i)}$ is a row subvector with dimension $k^\star$. It is sufficient to consider the uniform convergence of $a^{(i)}g(\theta)$ for each i, where $a^{(i)}g(\theta) = e(\theta)^\top(\sum_{l=1}^L a_l^{(i)}P_l)e(\theta) + a_x^{(i)}Q^\top e(\theta)$.

We start with the uniform convergence of the first term $e(\theta)^\top(\sum_{l=1}^L a_l^{(i)}P_l)e(\theta)$. We denote $\mathcal{K} = I_n \otimes \Lambda(\gamma)$, $\mathcal{K}_0 = I_n \otimes \Lambda(\gamma_0)$, $W = V^\top \otimes U$ and $S = S(\gamma_0)$, yielding $S(\gamma) = S + W(\mathcal{K}_0 - \mathcal{K})$. By expansion, we have $e(\theta) = d(\theta) + e + W(\mathcal{K}_0 - \mathcal{K})S^{-1}e$, where $d(\theta) = W(\mathcal{K}_0 - \mathcal{K})S^{-1}x\beta_0 + x(\beta_0 - \beta)$. It follows that $e(\theta)^\top(\sum_{l=1}^L a_l^{(i)}P_l)e(\theta) = d(\theta)^\top(\sum_{l=1}^L a_l^{(i)}P_l)d(\theta) + l(\theta) + q(\theta)$, where $l(\theta) = d(\theta)^\top(\sum_{l=1}^L a_l^{(i)}P_l^s)\{e + W(\mathcal{K}_0 - \mathcal{K})S^{-1}e\}$ and $q(\theta) = \{e^\top + e^\top S^{-1\top}(\mathcal{K}_0 - \mathcal{K})^\top W^\top\}(\sum_{l=1}^L a_l^{(i)}P_l)\{e + W(\mathcal{K}_0 - \mathcal{K})S^{-1}e\}$. For the term $l(\theta)$, we have that

$$\begin{aligned}\frac{1}{mn}l(\theta) =& \frac{1}{mn}(x\beta_0)^\top S^{-1\top}(\mathcal{K}_0 - \mathcal{K})^\top W^\top(\sum_{l=1}^L a_l^{(i)}P_l^s)e + \frac{1}{mn}(\beta_0 - \beta)^\top x^\top(\sum_{l=1}^L a_l^{(i)}P_l^s)e \\ &+ \frac{1}{mn}(x\beta_0)^\top S^{-1\top}(\mathcal{K}_0 - \mathcal{K})^\top W^\top(\sum_{l=1}^L a_l^{(i)}P_l^s)W(\mathcal{K}_0 - \mathcal{K})S^{-1}e \\ &+ \frac{1}{mn}(\beta_0 - \beta)^\top x^\top(\sum_{l=1}^L a_l^{(i)}P_l^s)W(\mathcal{K}_0 - \mathcal{K})S^{-1}e = o_p(1),\end{aligned}$$

uniformly in $\theta \in \Theta$. Similarly,

$$\begin{aligned}\frac{1}{mn}q(\theta) =& \frac{1}{mn}e^\top\left(\sum_{l=1}^L a_l^{(i)}P_l\right)e + \frac{1}{mn}e^\top S^{-1\top}(\mathcal{K}_0-\mathcal{K})^\top W^\top\left(\sum_{l=1}^L a_l^{(i)}P_l^s\right)e\\ &+\frac{1}{mn}e^\top S^{-1\top}(\mathcal{K}_0-\mathcal{K})^\top W^\top\left(\sum_{l=1}^L a_l^{(i)}P_l\right)W(\mathcal{K}_0-\mathcal{K})S^{-1}e\\ =&\frac{1}{mn}\sum_{l=1}^L a_l^{(i)}\mathrm{tr}\left\{\Sigma S^{-1\top}(\mathcal{K}_0-\mathcal{K})^\top W^\top P_l^s\right\}\\ &+\frac{1}{mn}\sum_{l=1}^L a_l^{(i)}\mathrm{tr}\left\{\Sigma S^{-1\top}(\mathcal{K}_0-\mathcal{K})^\top W^\top P_l W(\mathcal{K}_0-\mathcal{K})S^{-1}\right\}\\ &+o_p(1),\end{aligned}$$

uniformly in $\theta \in \Theta$, and $\mathrm{E}(e^\top P_l e) = \mathrm{tr}\left\{\Sigma P_l\right\} = \mathrm{tr}\left\{\Sigma \mathrm{Diag}(P_l)\right\} = 0$ for all $l = 1, \cdots, L$. Consequently,

$$\begin{aligned}\frac{1}{mn}e(\theta)^\top\left(\sum_{l=1}^L a_l^{(i)}P_l\right)e(\theta) = &\frac{1}{mn}d(\theta)^\top\left(\sum_{l=1}^L a_l^{(i)}P_l\right)d(\theta)\\ &+\frac{1}{mn}\sum_{l=1}^L a_l^{(i)}\mathrm{tr}\left\{\Sigma S^{-1\top}(\mathcal{K}_0-\mathcal{K})^\top W^\top P_l^s\right\}\\ &+\frac{1}{mn}\sum_{l=1}^L a_l^{(i)}\mathrm{tr}\left\{\Sigma S^{-1\top}(\mathcal{K}_0-\mathcal{K})^\top W^\top P_l W(\mathcal{K}_0-\mathcal{K})S^{-1}\right\}\\ &+o_p(1),\end{aligned}$$

uniformly in $\theta \in \Theta$.

In addition, the second term $\frac{1}{mn}a_x^{(i)}Q^\top e(\theta) = \frac{1}{mn}a_x^{(i)}Q^\top d(\theta) + o_p(1)$ uniformly in $\theta \in \Theta$. As $g(\theta)$ is a quadratic function of θ and Θ is compact, $\frac{1}{mn}a\mathrm{E}\left\{g(\theta)\right\}$ is uniformly equicontinuous on Θ. The identification condition and the uniform equicontinuity of $\frac{1}{mn}a\mathrm{E}\left\{g(\theta)\right\}$ imply that the identification uniqueness condition for $\frac{1}{m^2n^2}\mathrm{E}\left\{g(\theta)\right\}^\top a^\top a\mathrm{E}\left\{g(\theta)\right\}$ must be satisfied. Thus, the consistency of $\widehat{\theta}$ follows from this uniform convergence and the identification uniqueness condition (White, 1994).

STEP II. For the asymptotic distribution of $\widehat{\theta}$, applying $\{\partial g(\widehat{\theta})/\partial\theta^\top\}a^\top ag(\widehat{\theta}) = 0$ and Taylor expansion of $g(\widehat{\theta})$ at θ_0, yields

$$\sqrt{mn}(\widehat{\theta}-\theta_0) = -\left(\frac{1}{mn}\frac{\partial g(\widehat{\theta})}{\partial\theta^\top}a^\top a\frac{1}{mn}\frac{\partial g(\bar{\theta})}{\partial\theta^\top}\right)^{-1}\frac{1}{mn}\frac{\partial g(\widehat{\theta})}{\partial\theta^\top}a^\top a\frac{1}{\sqrt{mn}}g(\theta_0),$$

where $\bar{\theta}$ lies between θ_0 and $\widehat{\theta}$.

Next, we verify $\frac{1}{mn}\{\partial g(\theta)/\partial\theta^\top\} = -\frac{1}{mn}D + o_p(1)$, where

$$\frac{\partial g(\theta)}{\partial\theta^\top} = \Big(P_1^s e(\theta),\cdots,P_L^s e(\theta),Q\Big)^\top\Big(S'_{\gamma_1}y,\cdots,S'_{\gamma_d}y,x\Big),$$

$S'_{\gamma_a} = -V^\top\otimes\bar{U}_{\gamma_a}$ and $\bar{U}_{\gamma_a} = U\Lambda_{\gamma_a}$. Explicitly, $\frac{1}{mn}e(\theta)^\top P_l^s S'_{\gamma_a}y = \frac{1}{mn}e(\theta)^\top P_l^s S'_{\gamma_a}S^{-1}x\beta_0 + \frac{1}{mn}e(\theta)^\top P_l^s S'_{\gamma_a}S^{-1}e$. Since

$$\begin{aligned}\frac{1}{mn}e(\theta)^\top P_l^s S'_{\gamma_a}S^{-1}x\beta_0 =& \frac{1}{mn}d(\theta)^\top P_l^s S'_{\gamma_a}S^{-1}x\beta_0 + \frac{1}{mn}e^\top P_l^s S'_{\gamma_a}S^{-1}x\beta_0\\ &+\frac{1}{mn}e^\top S^{-1\top}(\mathcal{K}_0-\mathcal{K})^\top W^\top P_l^s S'_{\gamma_a}S^{-1}x\beta_0\\ =&\frac{1}{mn}d(\theta)^\top P_l^s S'_{\gamma_a}S^{-1}x\beta_0 + o_p(1),\end{aligned}$$

and

$$\begin{aligned}\frac{1}{mn}e(\theta)^\top P_l^s S'_{\gamma_a}S^{-1}e =& \frac{1}{mn}d(\theta)^\top P_l^s S'_{\gamma_a}S^{-1}e + \frac{1}{mn}e^\top P_l^s S'_{\gamma_a}S^{-1}e\\ &+\frac{1}{mn}e^\top S^{-1\top}(\mathcal{K}_0-\mathcal{K})^\top W^\top P_l^s S'_{\gamma_a}S^{-1}e\\ =&\frac{1}{mn}\mathrm{tr}\left\{\Sigma P_l^s S'_{\gamma_a}S^{-1}\right\}\\ &+\frac{1}{mn}\mathrm{tr}\left\{\Sigma S^{-1\top}(\mathcal{K}_0-\mathcal{K})^\top W^\top P_l^s S'_{\gamma_a}S^{-1}\right\} + o_p(1),\end{aligned}$$

uniformly in $\theta \in \Theta$, we have

$$\frac{1}{mn} e(\theta)^\top P_l^s S'_{\gamma_a} y = \frac{1}{mn} d(\theta)^\top P_l^s S'_{\gamma_a} S^{-1} x \beta_0 + \frac{1}{mn} \mathrm{tr}\left\{\Sigma P_l^s S'_{\gamma_a} S^{-1}\right\} + \frac{1}{mn} \mathrm{tr}\left\{\Sigma S^{-1\top} (\mathcal{K}_0 - \mathcal{K})^\top W^\top P_l^s S'_{\gamma_a} S^{-1}\right\} + o_p(1),$$

uniformly in $\theta \in \Theta$. At θ_0, $d(\theta_0) = 0$, $S'_{\gamma_a} S^{-1} = G_{\gamma_a}$. Hence, $\frac{1}{mn} e(\theta)^\top P_l^s S'_{\gamma_a} y = \frac{1}{mn} \mathrm{tr}\{\Sigma P_l^s G_{\gamma_a}\} + o_p(1)$. In addition, $\frac{1}{mn} e(\theta)^\top P_l^s x = o_p(1)$ and $\frac{1}{mn} Q^\top S'_{\gamma_a} y = \frac{1}{mn} Q^\top G_{\gamma_a} x \beta_0 + o_p(1)$. In conclusion,

$$\frac{1}{mn} \frac{\partial g(\theta)}{\partial \theta^\top} = -\frac{1}{mn} D + o_p(1).$$

Since $\frac{1}{\sqrt{mn}} a g(\theta_0) \xrightarrow{d} N(0, \lim_{mn\to\infty} \frac{1}{mn} a \Omega a^\top)$, the asymptotic distribution of $\sqrt{mn}(\widehat{\theta} - \theta_0)$ then follows. This completes the entire proof.

S3.2 Proof of Theorem 2

According to the generalized Schwartz inequality, the optimal weighting matrix in Theorem 1 is $\left(\frac{1}{mn}\Omega\right)^{-1}$.

For consistency of $\widehat{\theta}_{opt}$, we consider

$$\frac{1}{mn} g(\theta)^\top \left(\frac{1}{mn}\widehat{\Omega}\right)^{-1} \frac{1}{mn} g(\theta) = \frac{1}{mn} g(\theta)^\top \Omega^{-1} g(\theta) + \frac{1}{mn} g(\theta)^\top \left(\widehat{\Omega}^{-1} - \Omega^{-1}\right) g(\theta). \tag{S3.1}$$

With $a = (\frac{1}{mn}\Omega)^{-1/2}$ in the proof of Theorem 1, the first term in (S3.1) converges in probability to a well-defined limit uniformly in $\theta \in \Theta$. For the second term, we have

$$\left\|\frac{1}{mn}g(\theta)^\top\left(\widehat{\Omega}^{-1}-\Omega^{-1}\right)g(\theta)\right\|\leq\left(\frac{1}{mn}\|g(\theta)\|\right)^2\left\|\left(\frac{1}{mn}\widehat{\Omega}\right)^{-1}-\left(\frac{1}{mn}\Omega\right)^{-1}\right\|, \tag{S3.2}$$

where $\|(\frac{1}{mn}\widehat{\Omega})^{-1}-(\frac{1}{mn}\Omega)^{-1}\|=o_p(1)$ uniformly in $\theta\in\Theta$. Next, to verify that (S3.2) converges in probability to zero uniformly in $\theta\in\Theta$, it is sufficient to show that $\frac{1}{mn}\|g(\theta)\|$ is $O_p(1)$ uniformly in $\theta\in\Theta$. From the proof of Theorem 1, we have $\frac{1}{mn}[g(\theta)-\mathrm{E}\{g(\theta)\}]=o_p(1)$ uniformly in $\theta\in\Theta$. In addition, we can calculate that $\frac{1}{mn}\mathrm{E}\{g(\theta)\}=O(1)$ uniformly in $\theta\in\Theta$, where

$$\begin{aligned}\frac{1}{mn}\mathrm{E}\left\{e(\theta)^\top P_l e(\theta)\right\}=&\frac{1}{mn}d(\theta)^\top P_l d(\theta)+\frac{1}{mn}\mathrm{tr}\left\{\Sigma S^{-1\top}(\mathcal{K}_0-\mathcal{K})^\top W^\top P_l^s\right\}\\&+\frac{1}{mn}\mathrm{tr}\left\{\Sigma S^{-1\top}(\mathcal{K}_0-\mathcal{K})^\top W^\top P_l W(\mathcal{K}_0-\mathcal{K})S^{-1}\right\}\\=&O(1)\end{aligned}$$

uniformly in $\theta\in\Theta$, and $\frac{1}{mn}\mathrm{E}\{Q^\top e(\theta)\}=\frac{1}{mn}Q^\top d(\theta)=O(1)$ uniformly in $\theta\in\Theta$. This, together with the Markov inequality, imply that $\frac{1}{mn}\|g(\theta)\|=O_p(1)$ uniformly in $\theta\in\Theta$. Therefore, the second term in (S3.1) is $o_p(1)$ uniformly in $\theta\in\Theta$. Thus, the consistency of $\widehat{\theta}_{opt}$ follows.

For the asymptotic distribution of $\widehat{\theta}_{opt}$, we employ similar techniques to those used in the proof of Theorem 1 and obtain

$$\begin{aligned}
&\sqrt{mn}(\widehat{\theta}_{opt}-\theta_0)\\
&\quad=-\left\{\frac{1}{mn}\frac{\partial g(\widehat{\theta})}{\partial\theta^\top}\left(\frac{1}{mn}\widehat{\Omega}\right)^{-1}\frac{1}{mn}\frac{\partial g(\bar{\theta})}{\partial\theta^\top}\right\}^{-1}\frac{1}{mn}\frac{\partial g(\widehat{\theta})}{\partial\theta^\top}\left(\frac{1}{mn}\widehat{\Omega}\right)^{-1}\frac{1}{\sqrt{mn}}g(\theta_0)\\
&\quad=\left\{\frac{1}{mn}D^\top\left(\frac{1}{mn}\Omega\right)^{-1}\frac{1}{mn}D\right\}^{-1}\frac{1}{mn}D^\top\left(\frac{1}{mn}\Omega\right)^{-1}\frac{1}{\sqrt{mn}}g(\theta_0)+o_p(1),
\end{aligned} \tag{S3.3}$$

where $\bar{\theta}$ lies between θ_0 and $\widehat{\theta}_{opt}$. Thus, the asymptotic distribution of $\sqrt{mn}(\widehat{\theta}_{opt}-\theta_0)$ follows, which completes the entire proof.

S3.3 Proof of Theorem 3

In Section 2.3 of the paper, we set $\theta=(\theta_1^\top,\theta_2^\top)^\top$, where $\theta_1=(\gamma_1,\beta^\top)^\top$ and $\theta_2=(\gamma_2,\cdots,\gamma_d)^\top$. The following notations and functions are based on this new setting. Under the null hypothesis $H_{0,\gamma}:\theta_2=(0,\cdots,0)^\top\in\mathbb{R}^{(d-1)\times 1}$, the true parameter is $\theta_0=(\theta_1^\top,0^\top)^\top$. Let $\widehat{\theta}^{(r)}$ be the resulting constrained feasible "optimal" GMME of θ; that is, $\widehat{\theta}^{(r)}$ is obtained by minimizing the function $g(\theta)^\top\widehat{\Omega}^{-1}g(\theta)$ with the constraint that $\theta_2=0$. Furthermore, we define $D=(D_1,D_2)$ and $D_\gamma=(D_1,0_{(L+k^\star)\times(d-1)})$, where $D_1=-\partial\mathrm{E}\left\{g(\theta_0)\right\}/\partial\theta_1^\top$ and $D_2=-\partial\mathrm{E}\left\{g(\theta_0)\right\}/\partial\theta_2^\top$.

Under the null hypothesis $H_{0,\gamma}$, we employ similar techniques in proving

(S3.3) to obtain that

$$\begin{aligned}
&\sqrt{mn}(\widehat{\theta}^{(r)} - \theta_0) \\
&\quad = \left\{ \frac{1}{mn} D_\gamma^\top \left(\frac{1}{mn} \Omega \right)^{-1} \frac{1}{mn} D_\gamma \right\}^{-1} \frac{1}{mn} D_\gamma^\top \left(\frac{1}{mn} \Omega \right)^{-1} \frac{1}{\sqrt{mn}} g(\theta_0) + o_p(1),
\end{aligned}$$

which implies that $\widehat{\theta}^{(r)}$ is a $\sqrt{mn}$-consistent estimator. Next, by applying Taylor expansion, we obtain

$$\frac{1}{\sqrt{mn}} g(\widehat{\theta}^{(r)}) = \frac{1}{\sqrt{mn}} g(\theta_0) + \frac{1}{mn} \frac{\partial g(\bar{\theta})}{\partial \theta^\top} \sqrt{mn}(\widehat{\theta}^{(r)} - \theta_0),$$

where $\bar{\theta}$ lies between $\widehat{\theta}^{(r)}$ and θ_0. In addition, applying Conditions (C1)–(C4) and the similar techniques in the proof of Theorem 1, yields $\frac{1}{mn}\{\partial g(\bar{\theta})/\partial \theta^\top\}$ $\xrightarrow{P} -\frac{1}{mn} D$. Accordingly,

$$\begin{aligned}
\frac{1}{\sqrt{mn}} g(\widehat{\theta}^{(r)}) &= \frac{1}{\sqrt{mn}} g(\theta_0) - \frac{1}{mn} D \sqrt{mn}(\widehat{\theta}^{(r)} - \theta_0) + o_p(1) \\
&= \frac{1}{\sqrt{mn}} \mathcal{H} g(\theta_0) + o_p(1),
\end{aligned}$$

where $\mathcal{H} = I_{L+k^\star} - \frac{1}{mn} D_1 \left\{ \frac{1}{mn} D_1^\top \left(\frac{1}{mn} \Omega \right)^{-1} \frac{1}{mn} D_1 \right\}^{-1} \frac{1}{mn} D_1^\top \left(\frac{1}{mn} \Omega \right)^{-1}$.

Using Conditions (C1)–(C5), we obtain $(\frac{1}{mn}\widehat{\Omega}^{(r)})^{-1} \xrightarrow{P} (\frac{1}{mn}\Omega)^{-1}$. Then, we have

$$\begin{aligned}
T_\gamma =& g\big(\widehat{\theta}^{(r)}\big)^\top \widehat{\Omega}^{(r)-1} g\big(\widehat{\theta}^{(r)}\big) \\
=& \frac{1}{\sqrt{mn}} g\big(\theta_0\big)^\top \mathcal{H}^\top \left(\frac{1}{mn} \Omega \right)^{-1} \mathcal{H} \frac{1}{\sqrt{mn}} g\big(\theta_0\big) + o_p(1) \\
=& \left\{ \left(\frac{1}{mn} \Omega \right)^{-1/2} \frac{1}{\sqrt{mn}} g(\theta_0) \right\}^\top (I_{L+k^\star} - \mathcal{H}_1) \left\{ \left(\frac{1}{mn} \Omega \right)^{-1/2} \frac{1}{\sqrt{mn}} g(\theta_0) \right\} + o_p(1),
\end{aligned}$$

where $\mathcal{H}_1 = \left(\frac{1}{mn}\Omega\right)^{-1/2} \frac{1}{mn} D_1 \left\{\frac{1}{mn} D_1^\top \left(\frac{1}{mn}\Omega\right)^{-1} \frac{1}{mn} D_1\right\}^{-1} \frac{1}{mn} D_1^\top \left(\frac{1}{mn}\Omega\right)^{-1/2}$,

and

$$\begin{aligned}\operatorname{tr}(\mathcal{H}_1) =& \operatorname{tr}\left[\frac{1}{mn} D_1^\top \left(\frac{1}{mn}\Omega\right)^{-1} \frac{1}{mn} D_1 \left\{\frac{1}{mn} D_1^\top \left(\frac{1}{mn}\Omega\right)^{-1} \frac{1}{mn} D_1\right\}^{-1}\right] \\ =& \operatorname{tr}(I_{p+1}) = p + 1.\end{aligned}$$

In addition, by Lemma A.5 in Lin and Lee (2010), we have $\left(\frac{1}{mn}\Omega\right)^{-1/2} \frac{1}{\sqrt{mn}} g(\theta_0)$ $\xrightarrow{d} N(0, I_{L+k^\star})$. The above results imply that $T_\gamma \xrightarrow{d} \chi^2(L + k^\star - p - 1)$, which completes the entire proof.

S4 A Link Function Test

To perform GMME, we need specify the link function $F(\cdot)$. To assess the adequacy of the link function, we develop a testing procedure in this. Any two specific link functions, F_1 and F_2, are not necessarily nested in the sense that one is not obtained as a special case of the other. Hence, we adapt the method of Rivers and Vuong (2002) and Hall and Pelletier (2011) to assess whether F_1 and F_2 are equally close to F_0. First, for $F \in \{F_1, F_2\}$, we denote the population analogue of the GMM minimands as

$$\mathcal{Q}(\theta^{(F)}) = \mathrm{E}\left\{g(\theta^{(F)})\right\}^\top \mathrm{E}\left\{g(\theta^{(F)})\right\},$$

and $\widetilde{\theta}^{(F)} = \operatorname{argmin}_{\theta^{(F)} \in \Theta} \mathcal{Q}(\theta^{(F)})$. Next, to measure the closeness of the two link functions F and F_0, we employ the difference between the minimands as DIFF $:= \mathcal{Q}(\widetilde{\theta}^{(F)}) - \mathcal{Q}(\theta^{(F_0)})$, where $\theta^{(F_0)} = \theta_0$ is the true parameter, as defined in Section 2.2 of the paper. In addition, for $F \in \{F_1, F_2\}$, we denote the GMME as $\widehat{\theta}^{(F)} = \mathrm{argmin}_{\theta^{(F)} \in \Theta} \widehat{\mathcal{Q}}(\theta^{(F)})$, where

$$\widehat{\mathcal{Q}}(\theta^{(F)}) = \left\{g(\theta^{(F)})\right\}^\top \left\{g(\theta^{(F)})\right\}.$$

Notably, $\widehat{\theta}^{(F)}$ is the empirical version of $\widetilde{\theta}^{(F)}$. Analogous to Theorem 1, $\widehat{\theta}^{(F)}$ is also a consistent estimator of $\widetilde{\theta}^{(F)}$.

Given any two specific link functions, F_1 and F_2, it is natural to select the model that is closest to the true link function, F_0. To this end, we consider the following hypotheses:

$$H_{0,LF} : \mathcal{Q}(\widetilde{\theta}^{(F_1)}) - \mathcal{Q}(\widetilde{\theta}^{(F_2)}) = 0 \quad \text{vs.} \quad H_{1,LF} : \quad \mathcal{Q}(\widetilde{\theta}^{(F_1)}) - \mathcal{Q}(\widetilde{\theta}^{(F_2)}) \neq 0.$$

Under the null hypothesis of $H_{0,LF}$, the two models with link functions F_1 and F_2 are equivalent. Under the alternative hypothesis of $H_{1,LF}$, one is better than the other. As we can estimate $\mathrm{E}\{g(\widetilde{\theta}^{(F)})\}$ by $g(\widehat{\theta}^{(F)})$ for $F \in \{F_1, F_2\}$, the test statistic can be proposed as

$$T_{LF} = \widehat{\mathcal{Q}}(\widehat{\theta}^{(F_1)}) - \widehat{\mathcal{Q}}(\widehat{\theta}^{(F_2)}).$$

Moreover, we need additional notations for the theoretical analysis. For convenience, we combine the parameters and moment functions from both models into one vector, which leads to $\eta = (\theta^{(F_1)\top}, \theta^{(F_2)\top})^\top$, $h(\eta) = \left(g(\theta^{(F_1)})^\top, g(\theta^{(F_2)})^\top\right)^\top$. Without ambiguity, we denote $\widetilde{\eta} = (\widetilde{\theta}^{(F_1)\top}, \widetilde{\theta}^{(F_2)\top})^\top$,

$\widehat{\eta} = (\widehat{\theta}^{(F_1)\top}, \widehat{\theta}^{(F_2)\top})^\top$. Further, we define $\zeta(\eta) = h(\eta) - \mathrm{E}\left\{h(\eta)\right\}$, $R(\eta) = \left(2\mathrm{E}\left\{g(\theta^{(F_1)})\right\}^\top, -2\mathrm{E}\left\{g(\theta^{(F_2)})\right\}^\top\right)^\top$. For $F \in \{F_1, F_2\}$, we denote $\mathcal{G}(\theta^{(F)}) = \partial \mathrm{E}\left\{g(\theta^{(F)})\right\}/\partial\theta^{(F)\top}$, $\widehat{\mathcal{G}}(\theta^{(F)}) = \partial g(\theta^{(F)})/\partial\theta^{(F)\top}$. To investigate the asymptotic property of T_{LF}, we impose the following set of assumptions.

(S.C1) For $F \in \{F_1, F_2\}$, $\mathrm{E}\left\{g(\theta^{(F)})\right\}$ exists and is finite for all $\theta^{(F)} \in \Theta$, and assume that $\mathrm{E}\left\{\sup_{\theta^{(F)}\in\Theta}\left\|g(\theta^{(F)})\right\|\right\} < \infty$.

(S.C2) For $F \in \{F_1, F_2\}$, $\widetilde{\theta}^{(F)}$ is an interior point of Θ, and that there exists $\widetilde{\theta}^{(F)} \in \Theta$ such that $\mathcal{Q}(\widetilde{\theta}^{(F)}) < \mathcal{Q}(\theta)$ for all $\theta \in \Theta \setminus \{\widetilde{\theta}^{(F)}\}$.

(S.C3) For $F \in \{F_1, F_2\}$, $\mathcal{G}(\widetilde{\theta}^{(F)})$ exists, is finite, and has the full rank $d + p$, and assume that $\mathcal{G}(\theta)$ is continuous on some ϵ-neighborhood N_ϵ of $\widetilde{\theta}^{(F)}$, and $\sup_{\theta\in N_\epsilon}\left\|\widehat{\mathcal{G}}(\theta) - \mathcal{G}(\theta)\right\| \xrightarrow{P} 0$.

(S.C4) For $F \in \{F_1, F_2\}$, $\left\|\mathrm{E}\left\{g\left(\theta^{(F)}\right)\right\}\right\| \neq 0$ for all $\theta^{(F)} \in \Theta$.

(S.C5) $(mn)^{-1/2}\zeta(\widetilde{\eta}) \xrightarrow{d} N(0, \mathcal{J})$, where $\mathcal{J} = \lim_{mn\to\infty}(mn)^{-1}\mathrm{Var}\left\{\zeta(\widetilde{\eta})\right\}$ is a positive definite matrix of finite constants.

Notably, Conditions (S.C1)–(S.C3) involve some regularity and identification conditions on $g(\theta^{(F)})$, $\widetilde{\theta}^{(F)}$, and $\partial g(\theta^{(F)})/\partial\theta^{(F)\top}$, which are carefully studied in Hall (2005). Condition (S.C4) indicates that the model setting we consider here is a case of nonlocal misspecification (see e.g., Hall and Inoue (2003); Hall and Pelletier (2011)). Condition (S.C5) applies the law of large numbers, which can be also found in Hall and Pelletier (2011). Based on these conditions, the limiting distribution of T_{LF} is given in the following theorem.

Theorem S.1. *Under Conditions (S.C1)–(S.C5) and the null hypothesis of $H_{0,LF}$, as $mn \to \infty$, we have*

$$T_{LF} \xrightarrow{d} N(0, \sigma_{LF}^2),$$

where $\sigma_{LF}^2 = R(\widetilde{\eta})^\top \mathrm{Var}\left\{\zeta(\widetilde{\eta})\right\} R(\widetilde{\eta})$.

Ultimately, Theorem S.1 indicates that under the null hypothesis of $H_{0,LF}$, T_{LF} asymptotically follows a normal distribution with a mean of 0 and a variance of σ_{LF}^2. In practice, σ_{LF}^2 is unknown and can be estimated by the sample variance of $R(\widehat{\eta})^\top \zeta(\widehat{\eta})$. At the given significance level α, we reject $H_{0,LF}$ and select link function F_1 if $T_{LF}/\widehat{\sigma}_{LF} < -z_{\alpha/2}$, where $z_{\alpha/2}$ is the $\alpha/2$-th upper quantile of the standard normal distribution. In contrast, we reject $H_{0,LF}$ and select link function F_2 if $T_{LF}/\widehat{\sigma}_{LF} > z_{\alpha/2}$. Otherwise, F_1 and F_2 are considered to be equivalent, and we could select the link function with smaller estimated objective function in empirical studies. In addition, Sections S4.1 and S4.2 provide detailed simulation and empirical results for this theorem, respectively.

Proof of Theorem S.1. For $F \in \{F_1, F_2\}$, analogous to the proof of

Theorem 1, we can easily obtain $(mn)^{1/2}\left(\widehat{\theta}^{(F)}-\widetilde{\theta}^{(F)}\right) = O_p(1)$. Then applying Taylor expansion to $\widehat{\mathcal{Q}}(\theta^{(F)})$ around $\theta^{(F)}=\widetilde{\theta}^{(F)}$, we have

$$\widehat{\mathcal{Q}}(\widehat{\theta}^{(F)}) = \widehat{\mathcal{Q}}(\widetilde{\theta}^{(F)}) + \frac{\partial \widehat{\mathcal{Q}}(\widetilde{\theta}^{(F)})}{\partial \theta^{(F)\top}}\left(\widehat{\theta}^{(F)}-\widetilde{\theta}^{(F)}\right) + o_p\left((mn)^{-1/2}\right). \qquad \text{(S4.1)}$$

Define $\psi(\theta^{(F)}) = 2\mathrm{E}\left\{g(\theta^{(F)})\right\}^\top \mathcal{G}(\theta^{(F)})$ and $\widehat{\psi}(\theta^{(F)}) = 2g(\theta^{(F)})^\top \widehat{\mathcal{G}}(\theta^{(F)})$. By the law of large numbers and the continuous mapping theorem, we can show that $\widehat{\psi}(\widetilde{\theta}^{(F)}) \xrightarrow{P} \psi(\widetilde{\theta}^{(F)})$. It therefore follows from (S4.1) that

$$\begin{aligned}\widehat{\mathcal{Q}}(\widehat{\theta}^{(F_1)}) - \widehat{\mathcal{Q}}(\widehat{\theta}^{(F_2)}) = \widehat{\mathcal{Q}}(\widetilde{\theta}^{(F_1)}) - \widehat{\mathcal{Q}}(\widetilde{\theta}^{(F_2)}) + \psi(\widetilde{\theta}^{(F_1)})\left(\widehat{\theta}^{(F_1)}-\widetilde{\theta}^{(F_1)}\right)& \\ - \psi(\widetilde{\theta}^{(F_2)})\left(\widehat{\theta}^{(F_2)}-\widetilde{\theta}^{(F_2)}\right) + o_p\left((mn)^{-1/2}\right).& \end{aligned} \qquad \text{(S4.2)}$$

Moreover, under conditions (C8)–(C12), the GMME $\widehat{\theta}^{(F)}$ can be obtained by solving the first-order conditions associated with $\mathrm{argmin}_{\theta^{(F)}\in\Theta}\widehat{\mathcal{Q}}(\theta^{(F)})$; that is, $\widehat{\psi}(\widehat{\theta}^{(F)})=0$. The probability limits $\widetilde{\theta}^{(F)}$ must satisfy the analogous population moment condition $\psi(\widetilde{\theta}^{(F)})=0$. Therefore, we write (S4.2) as

$$\widehat{\mathcal{Q}}(\widehat{\theta}^{(F_1)}) - \widehat{\mathcal{Q}}(\widehat{\theta}^{(F_2)}) = \widehat{\mathcal{Q}}(\widetilde{\theta}^{(F_1)}) - \widehat{\mathcal{Q}}(\widetilde{\theta}^{(F_2)}) + o_p\left((mn)^{-1/2}\right). \qquad \text{(S4.3)}$$

Under $H_{0,LF}$, we have $\mathcal{Q}(\widetilde{\theta}^{(F_1)}) = \mathcal{Q}(\widetilde{\theta}^{(F_2)})$; therefore, we can write (S4.3) as

$$\widehat{\mathcal{Q}}(\widehat{\theta}^{(F_1)}) - \widehat{\mathcal{Q}}(\widehat{\theta}^{(F_2)}) = \widehat{\mathcal{Q}}(\widetilde{\theta}^{(F_1)}) - \mathcal{Q}(\widetilde{\theta}^{(F_1)}) - \widehat{\mathcal{Q}}(\widetilde{\theta}^{(F_2)}) - \mathcal{Q}(\widetilde{\theta}^{(F_2)}) + o_p\left((mn)^{-1/2}\right). \qquad \text{(S4.4)}$$

In addition, notice that for $F \in \{F_1, F_2\}$,

$$\begin{aligned}\widehat{\mathcal{Q}}(\widetilde{\theta}^{(F)}) - \mathcal{Q}(\widetilde{\theta}^{(F)}) &= g(\widetilde{\theta}^{(F)})^\top g(\widetilde{\theta}^{(F)}) - \mathrm{E}\left\{g(\widetilde{\theta}^{(F)})\right\}^\top \mathrm{E}\left\{g(\widetilde{\theta}^{(F)})\right\} \\ &= g(\widetilde{\theta}^{(F)})^\top \left[g(\widetilde{\theta}^{(F)}) - \mathrm{E}\left\{g(\widetilde{\theta}^{(F)})\right\}\right] \\ &+ \mathrm{E}\left\{g(\widetilde{\theta}^{(F)})\right\}^\top \left[g(\widetilde{\theta}^{(F)}) - \mathrm{E}\left\{g(\widetilde{\theta}^{(F)})\right\}\right].\end{aligned} \tag{S4.5}$$

Therefore, by (S4.4) and (S4.5), we have

$$\begin{aligned}\widehat{\mathcal{Q}}(\widehat{\theta}^{(F_1)}) - \widehat{\mathcal{Q}}(\widehat{\theta}^{(F_2)}) &= 2\mathrm{E}\left\{g(\widetilde{\theta}^{(F_1)})\right\}^\top \left[g(\widetilde{\theta}^{(F_1)}) - \mathrm{E}\left\{g(\widetilde{\theta}^{(F_1)})\right\}\right] \\ &- 2\mathrm{E}\left\{g(\widetilde{\theta}^{(F_2)})\right\}^\top \left[g(\widetilde{\theta}^{(F_2)}) - \mathrm{E}\left\{g(\widetilde{\theta}^{(F_2)})\right\}\right] + o_p\left((mn)^{-1/2}\right) \\ &= R(\widetilde{\eta})^\top \zeta(\widetilde{\eta}) + o_p\left((mn)^{-1/2}\right).\end{aligned}$$

This completes the entire proof.

S5 Additional Simulation and Empirical Results

S5.1 Simulation results for the link function test

Theorem S.1 establishes that the link function test statistic T_{LF} has an asymptotic normal distribution under the null hypothesis. To evaluate the performance of T_{LF}, we examine the percentage of the true link function selection (TS) and false link function selection (FS). In this study, we assume that the true link function is LINK I (logistic). Next, we compare the following two pairs of link functions: (i) LINK I (logistic) vs. LINK II (inverse of log-log) and (ii) LINK I (logistic) vs. LINK III (inverse of probit). Thus, the percentages of TS and FS can be computed based on the selection results of the K replications. Table S.1 reports the TS and FS results for two pairs of link functions under variance designs V1 and V2 with a significance level of 0.05. The table reveals that TS increases and tends to 100% given an increase in mn, while FS is almost 0. As expected, the selection results are almost correct and at least are not false. These findings demonstrate the validity of our proposed testing procedure.

Table S.1: The selection results of the link function test T_{LF} for two pairs of link functions with a significance level of 0.05.

		LINK I vs. LINK II		LINK I vs. LINK III	
	(m, n)	TS	FS	TS	FS
Design V1	(50,70)	65.5%	0%	59%	0%
	(50,90)	68%	0%	65%	0%
	(100,90)	71.5%	0%	66.5%	0%
Design V2	(50,70)	64.5%	0%	59.5%	0%
	(50,90)	67.5%	0%	63.5%	0%
	(100,90)	71%	0%	65.5%	0%

S5.2 Simulation Results for Other Sparse Adjacency Matrix

To explore the simulation performance under various sparse patterns in the adjacency matrix, we examine three other representative forms of sparsity:

(i) Dyadic Sparsity: A small set of node pairs (dyads) that are highly connected, while other connections are sparse—reflecting scenarios where strong interactions are limited to a few node pairs.

(ii) Block Sparsity: Nodes are grouped into blocks with dense intra-block and sparse inter-block connections, capturing community structures often seen in social and collaborative networks.

(iii) Power-law Sparsity: The degree distribution follows a power law, resulting in a few highly connected nodes (hubs) and many sparsely connected ones. This type of sparsity is typical in scale-free networks like social media or citation networks.

We conduct simulations under each of these sparse structures, and the results for LINK I are reported in Table S.2. The results show that the model performs well under all three patterns, demonstrating the robustness and accuracy of the proposed method across different sparsity scenarios.

Table S.2: The performance of the GMMEs of parameters θ under variance designs V1 and V2 and three sparse patterns, respectively, for LINK I. The BIAS, SE and tSE values ($\times 10^2$) are reported for θ estimates.

			Design V1							Design V2						
	(m, n)	Measures	$\widehat{\gamma}_1$	$\widehat{\gamma}_2$	$\widehat{\gamma}_3$	$\widehat{\beta}_{1,1}$	$\widehat{\beta}_{1,2}$	$\widehat{\beta}_{2,1}$	$\widehat{\beta}_{2,2}$	$\widehat{\gamma}_1$	$\widehat{\gamma}_2$	$\widehat{\gamma}_3$	$\widehat{\beta}_{1,1}$	$\widehat{\beta}_{1,2}$	$\widehat{\beta}_{2,1}$	$\widehat{\beta}_{2,2}$
Dyad	(50,70)	BIAS	0.71	-0.01	0.38	-0.04	-0.03	0.00	0.00	0.10	0.02	0.03	-0.01	0.00	0.00	0.00
		SE	3.77	26.10	4.75	0.17	0.12	0.11	0.08	1.35	10.30	1.77	0.06	0.04	0.02	0.02
		tSE	5.63	44.15	7.92	0.22	0.17	0.10	0.09	1.81	14.09	2.39	0.07	0.05	0.03	0.03
	(50,90)	BIAS	0.32	-0.68	-0.06	-0.01	0.01	-0.01	0.00	0.05	-0.09	0.00	0.00	0.00	0.00	0.00
		SE	2.70	10.88	3.13	0.14	0.07	0.07	0.06	0.98	4.04	1.19	0.05	0.02	0.02	0.02
		tSE	2.95	11.97	3.24	0.15	0.07	0.06	0.06	0.99	4.04	1.15	0.05	0.02	0.02	0.02
	(100,90)	BIAS	-0.01	0.03	-0.07	0.00	0.00	0.00	0.00	0.01	0.07	0.04	0.00	0.00	0.00	0.00
		SE	0.25	2.73	1.37	0.03	0.03	0.03	0.02	0.13	1.55	0.71	0.02	0.02	0.02	0.01
		tSE	0.24	2.64	1.31	0.03	0.03	0.03	0.02	0.14	1.57	0.72	0.01	0.02	0.02	0.01
Block	(50,70)	BIAS	0.46	2.90	2.84	-0.01	-0.01	0.00	0.01	0.02	-0.11	0.01	0.00	0.00	0.00	0.00
		SE	2.18	21.87	19.39	0.09	0.20	0.11	0.09	0.49	5.29	4.62	0.02	0.06	0.02	0.02
		tSE	2.07	19.76	16.74	0.08	0.20	0.10	0.09	0.47	4.87	4.26	0.02	0.05	0.02	0.02
	(50,90)	BIAS	0.29	-0.33	0.81	0.00	0.00	-0.01	0.00	0.03	0.06	0.07	0.00	0.00	0.00	0.00
		SE	2.12	8.02	7.26	0.09	0.07	0.07	0.06	0.70	2.71	2.36	0.03	0.02	0.02	0.02
		tSE	2.06	7.86	6.95	0.08	0.07	0.06	0.06	0.71	2.75	2.31	0.03	0.02	0.02	0.02
	(100,90)	BIAS	-0.03	0.01	-0.29	0.00	0.00	0.00	0.00	0.00	-0.01	-0.03	0.00	0.00	0.00	0.00
		SE	0.52	1.59	2.72	0.03	0.03	0.03	0.03	0.30	0.86	1.44	0.02	0.02	0.02	0.01
		tSE	0.50	1.45	2.57	0.03	0.03	0.03	0.02	0.29	0.87	1.49	0.02	0.01	0.02	0.01
Power-law	(50,70)	BIAS	0.01	0.26	0.21	-0.01	-0.01	0.00	0.00	-0.01	0.06	0.02	0.00	0.00	0.00	0.00
		SE	1.07	4.73	1.95	0.09	0.10	0.10	0.08	0.16	0.91	0.38	0.02	0.03	0.02	0.02
		tSE	0.96	4.63	1.90	0.08	0.10	0.09	0.08	0.18	0.90	0.37	0.02	0.02	0.02	0.02
	(50,90)	BIAS	-0.05	0.01	-0.14	0.00	0.00	-0.01	0.00	0.00	-0.01	0.01	0.00	0.00	0.00	0.00
		SE	0.63	1.38	1.53	0.06	0.07	0.07	0.06	0.19	0.37	0.43	0.02	0.02	0.02	0.02
		tSE	0.61	1.37	1.44	0.06	0.07	0.06	0.06	0.17	0.40	0.39	0.02	0.02	0.02	0.02
	(100,90)	BIAS	0.01	0.02	-0.06	0.00	0.00	0.00	0.00	0.01	-0.01	-0.03	0.00	0.00	0.00	0.00
		SE	0.30	1.35	0.98	0.03	0.03	0.03	0.03	0.18	0.78	0.59	0.02	0.02	0.02	0.01
		tSE	0.29	1.24	0.93	0.03	0.03	0.03	0.02	0.17	0.73	0.57	0.02	0.01	0.02	0.01

S5.3 Empirical Results for the Link Function Test

To assess the adequacy of the link function and select the optimal one, we first apply the link function test to each pair of three link functions in Section 2.1; that is, LINK I (logistic) vs. LINK II (inverse of log-log), LINK I (logistic) vs. LINK III (inverse of probit), and LINK II (inverse of log-log) vs. LINK III (inverse of probit). For each test setting, we use the previous quarterly dataset from the second quarter of 2022 to select an optimal link function. We cannot reject any of the null hypotheses H_0 at a significance level of 0.05, which indicates that LINKs I, II and III are equivalent for this dataset. Finally, we apply LINK I with the smallest estimated objective function (e.g., see Rivers and Vuong (2002)) to analyze the recent dataset from the third quarter of 2022. We further conduct empirical studies by applying LINKs II and III. Tables S.3 and S.4 report the regression results of MNIM for LINKs II and III, respectively, which yield similar findings to those obtained from LINK I in Table 2 of the paper. This further supports that the results are indeed similar for three link functions.

S5.4 Empirical Results for the Grouped Regression

Inspired by Remark 2, we further explore covariate heterogeneity under a group specification. Using spectral clustering on the similarity matri-

Table S.3: The regression results of MNIM for LINK II with coefficient estimates, estimated standard errors, and p-values.

	Variables	Estimate	Standard Error	p-value
X_1	size	0.1501	0.0325	≪0.0001
	volatility	-0.2009	0.8697	0.8173
	return	0.4022	0.4804	0.4024
	age	-0.5581	0.0834	≪0.0001
X_2	size	0.0001	0.0000	≪0.0001
	volatility	0.0021	0.0004	≪0.0001
	return	0.2021	0.1903	0.2881
Z	intercept	0.2794	0.9382	0.7659
	size	0.4978	0.1184	≪0.0001
	volatility	0.4235	0.8948	0.6360
	degree	0.2643	0.0486	≪0.0001

Table S.4: The regression results of MNIM for LINK III with coefficient estimates, estimated standard errors, and p-values.

	Variables	Estimate	Standard Error	p-value
X_1	size	0.1465	0.0394	0.0002
	volatility	-0.2009	1.2433	0.8716
	return	0.4023	0.6521	0.5373
	age	-0.5596	0.0938	≪0.0001
X_2	size	0.0001	0.0000	≪0.0001
	volatility	0.0030	0.0004	≪0.0001
	return	0.2022	0.1992	0.3102
Z	intercept	0.2999	1.1128	0.7875
	size	0.4976	0.1215	≪0.0001
	volatility	0.4011	2.4366	0.8693
	degree	0.3017	0.0636	≪0.0001

ces—constructed from $\mathcal{X}_1$ and $\mathcal{X}_2$ via a Gaussian kernel—we identify three distinct groups among funds and a single unified group among stocks. Accordingly, we partition X_1 into group-specific components X_{1,g_1}, X_{1,g_2}, and X_{1,g_3}, while maintaining X_2 and Z common across all observations. The grouped regression results, shown in Table S.5, reveal substantial heterogeneity across groups, which is obscured under the pooled model. Specifically, for X_1, size shows significantly positive effects in groups g_2 and g_3, but is slightly negative in g_1, averaging to a moderate positive effect in the pooled model; return is positively associated in g_2 but negatively in g_1, resulting in an insignificant pooled effect; age consistently has a negative effect across groups, with the strongest impact observed in g_2. For X_2, allowing group-specific variation in X_1 leads to more precise estimates, with volatility and return becoming significant compared to the pooled model. Importantly, the estimates for Z remain stable, indicating that group heterogeneity primarily affects covariate-response relationships rather than the structural component. Overall, introducing group-specific effects uncovers richer covariate structures and improves model interpretability compared to the aggregated approach.

Table S.5: The grouped regression results of MNIM for LINK I with coefficient estimates, estimated standard errors, and p-values.

	Variables	Estimate	Standard Error	p-value
X_{1,g_1}	size	-0.0998	0.0602	0.0973
	volatility	0.0877	0.0441	0.0467
	return	-0.1128	0.0419	0.0072
	age	-0.0453	0.0433	0.2948
X_{1,g_2}	size	0.3645	0.0195	$\ll$0.0001
	volatility	0.0198	0.0664	0.7656
	return	0.3010	0.0299	$\ll$0.0001
	age	-0.5014	0.0004	$\ll$0.0001
X_{1,g_3}	size	0.0562	0.0273	0.0397
	volatility	0.0330	0.0405	0.4151
	return	0.0319	0.0347	0.3581
	age	-0.0408	0.0354	0.2484
X_2	size	-0.0148	0.0200	0.4953
	volatility	0.0959	0.0318	0.0026
	return	-0.0467	0.0206	0.0235
Z	intercept	0.3001	2.0282	0.8824
	size	0.4978	0.2217	0.0241
	volatility	0.4013	4.0782	0.9219
	degree	0.2990	0.1081	0.0056

S6 The Details of Variance Designs V1 and V2

To generate heteroscedastic disturbances, we consider group structures. That is, the error terms ϵ_{ij} are generated from the normal distribution with means of 0 and different variances across groups. First, we randomly divide the m individuals and n dimensions of vector response into $\mathcal{M} = m/10$ and $\mathcal{N} = n/10$ groups (round to the closest integer), respectively. This leads to a total $\mathcal{M} \times \mathcal{N}$ groups. Figure 1 presents the structure of the groups. For $h = 1, \cdots, \mathcal{M} - 1$, the individual group size r_h is determined by a uniform $U(3, 17)$ variable (rounded to the closest integer); therefore, the mean

	dimension group 1 (c_1)	dimension group 2 (c_2)	...	dimension group $\mathcal{N}$ ($c_{\mathcal{N}}$)
individual group 1 (r_1)	group 1 ($r_1 \times c_1$)	group 2 ($r_1 \times c_2$)	...	group $\mathcal{N}$ ($r_1 \times c_{\mathcal{N}}$)
individual group 2 (r_2)	group $\mathcal{N}+1$ ($r_2 \times c_1$)	group $\mathcal{N}+2$ ($r_2 \times c_2$)	...	group $2\times\mathcal{N}$ ($r_2 \times c_{\mathcal{N}}$)
...	...	...	...	...
individual group $\mathcal{M}$ ($r_{\mathcal{M}}$)	group $(\mathcal{M}-1)\times\mathcal{N}+1$ ($r_{\mathcal{M}} \times c_1$)	group $(\mathcal{M}-1)\times\mathcal{N}+2$ ($r_{\mathcal{M}} \times c_2$)	...	group $\mathcal{M}\times\mathcal{N}$ ($r_{\mathcal{M}} \times c_{\mathcal{N}}$)

Figure 1: The illustration of group structures. For $h = 1, \cdots, \mathcal{M}$ and $s = 1, \cdots, \mathcal{N}$, each group size is the product of the corresponding individual group size r_h and dimension group size c_s, which are listed in parentheses.

row group size is approximately 10. For $s = 1, \cdots, \mathcal{N} - 1$, we similarly determine the dimension group size c_s as the individual group size with a mean column group size of approximately 10. For completeness, we define $r_{\mathcal{M}} = m - \sum_{h=1}^{\mathcal{M}-1} r_h$ and $c_{\mathcal{N}} = m - \sum_{s=1}^{\mathcal{N}-1} c_s$. Accordingly, for $h = 1, \cdots, \mathcal{M}$ and $s = 1, \cdots, \mathcal{N}$, the unit group size is $r_h \times c_s$. Next, we consider two variance structures based on this particular design.

Design V1. For individual $i \in h$ and dimension $j \in s$, if the group size $r_h \times c_s$ is greater than 100, then the variance of ϵ_{ij} is $(r_h \times c_s/m/n)^2$; otherwise, the variance is $(r_h \times c_s)^{-4}$.

Design V2. For individual $i \in h$ and dimension $j \in s$, the variance of ϵ_{ij}

is $(r_h \times c_s)^{-2}$.

As the group size increases, the variance function of V1 decreases and then increases, while that of V2 decreases. Thus, under the two variance designs, we can generate the error term vector $e = (\epsilon_{11}, \cdots, \epsilon_{m1}, \cdots, \epsilon_{1n}, \cdots, \epsilon_{mn})$.

References

Hall. A. R. (2005), *Generalized Method of Moments.* Oxford University Press, London.

Hall, A. R. and Inoue, A. (2003). The large sample behavior of the generalized method of moments estimator in misspecified models. *Journal of Econometrics* **114**, 361–394.

Hall, A. R. and Pelletier, D. (2011). Nonnested testing in models estimated via generalized method of moments. *Econometric Theory* **27**, 443–456.

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

Rivers, D. and Vuong, Q. (2002). Model selection tests for nonlinear dynamic models. *Econometrics Journal* **5**, 1–39.

White, H. (1994). *Estimation, inference and specification analysis.* Cambridge University Press, New York.