Identification and Estimation of Spatial Autoregressive Models with Common Factors*

Cynthia Fan Yang†

May 15, 2016

Abstract

This paper considers panel data models with both spatial autocorrelation and unobserved common factors. The conditions for model identification conditional on the factors are first derived, and then several estimation approaches that employ cross-sectional averages as factor proxies are suggested, including the 2SLS, B2SLS, and GMM estimation. We derive the asymptotic distribution of these estimators and compare their efficiency properties. Through extensive Monte Carlo experiments, we show that the proposed estimators are robust to unknown heteroskedasticity and serial correlation in the disturbances and have satisfactory finite sample performance when N is large and T is relatively modest, which are consistent with our theoretical findings. Further evidence supporting the identification conditions is also provided by simulations.

Keywords: Spatial panel data models, common factors, identification, two-stage least-squares (2SLS), Generalized method of moment (GMM)

1 Introduction

The past decade has seen a growing interest in the panel data models with cross sectional dependence, which, if ignored, could lead to inconsistent estimates and misleading inferences. Two approaches have been developed to address this issue, namely spatial models and common factor models. In the former, the dependence is represented by weight matrices typically based on physical, economic or social distance, while in the latter it is captured by a number of observable or latent factors (common shocks).

The two strands of literature have been growing rapidly and separately. As it is difficult to tell the true nature of dependence in economic and social activities, efforts are called for to investigate the connections and differences between the two types of models. However, not until recently did this subject receive attention from a few researchers. Among them, Pesaran and Tosetti (2011) consider a panel regression model with cross-sectionally dependent errors arising from spatial correlation and/or common shocks. They show that the Common Correlated Effects (CCE) estimator, developed by

*I am very grateful to M. Hashem Pesaran for his invaluable guidance and advice. I would also like to thank Hyungsik Roger Moon and Cheng Hsiao for helpful comments. I acknowledge the IAAE travel grant to attend the conference.

†Department of Economics, University of Southern California. Email: yang488@usc.edu.

Pesaran (2006), yields good estimates for the slope parameters. In contrast, Bai and Li (2014) specify the spatial autocorrelation on the dependent variable while assuming the presence of unobserved common shocks. To overcome the endogeneity problem, they advocate a pseudo maximum likelihood (ML) method that simultaneously estimate a large group of parameters including the heterogeneous factor loadings and heterogeneous variances of the disturbances. Besides computational complexities, there are two major drawbacks of the ML method. First, it is not robust to the presence of serial correlation in the errors. Second, the number of latent factors has to be estimated and this is well known to be a challenging task. In stead of estimating the two effects jointly, Bailey et al. (2014) propose a two-stage approach that extracts the common factors by cross-unit averages (or principle components) in the first stage and then estimates the spatial connections in the second stage. They demonstrate their modeling strategy with an application on US house price changes. Nonetheless, a formal distribution theory that takes into account the first-stage sampling errors is not yet available. Apart from estimation and inference, Chudik et al. (2011) introduce the concepts of weak and strong cross-sectional dependence and relate them to the properties of common factor models. Bailey et al. (2013) focus on how to measure the degree of cross-sectional dependence. They represent it by an exponent of dependence, which captures how fast the variance of the cross-sectional average declines with the cross-section dimension size. A spatial model, from this perspective, is a special form of weakly dependent process with weak factors. Using the exponent of cross-sectional dependence, **?** further develops a CD test of weak cross-sectional dependence in large panels.

This paper aims to join in the above efforts of bringing together spatial and factor models for a better understanding of the cross-sectional dependence. We set out to investigate the identification conditions for a joint model of spatial autocorrelation and common shocks, and then suggest several estimation methods. The identification of spatial models has been discussed in two recent studies: Lee and Yu (2015) consider the spatial Durbin dynamic models for finite samples, and Aquaro et al. (2015) examine the spatial autoregressive models with heterogeneous spatial coefficients. Our research sheds new light on the identification of spatial models with factors, and demonstrates that the conditions in Lee and Yu (2015) cannot be applied to large samples. With regard to estimation, the maximum likelihood method advocated by Bai and Li (2014) is computationally intensive and not robust enough, and this motives us to propose estimating the joint model by instrumental variables and/or other moment conditions. The main idea is to approximate the unobserved factors with cross-sectional averages and then estimate the model by two-stage least-squares (2SLS) or generalized methods of moments (GMM). Specifically, our study is built on two areas of works. On the one hand, the method of substituting cross-sectional averages for unknown factors in multi-factor panel data models originated from Pesaran (2006), and was later extended by Pesaran and Tosetti (2011) to allow more general forms of error dependence. Kapetanios et al. (2011) proceeded to show that the method continues to work even if the factors follow unit root processes. On the other hand, the existing literature on the IV/GMM estimation of spatial models also lays foundation for this paper. Among others, some notable contributions include **?**, **?**, Lee (2003), Lee (2007), Lin and Lee (2010), and Lee and Yu (2014). In particular, our construction of the moment conditions is closely related to Lee (2007) and Lee and

Yu (2014), and the idea of the best 2SLS (B2SLS) estimator is in the spirit of Lee (2003) and Lee (2007). We show that the proposed estimators, including the 2SLS, B2SLS and GMM estimators, are asymptotically normal and free of nuisance parameters so long as the time series dimension (T) is relatively small to the cross-section dimension (N), as both T and N tend towards infinity jointly. We find these estimation schemes appealing because they allow for heteroskedasticity and serial correlation in the error term, do not require estimating the number of factors, and are simple to implement.

The rest of this paper is organized as follows. Section 2 specifies the model and describes the idea of approximating the unobserved factors with cross-sectional averages. Section 3 details the assumptions underlying the model. Section 4 investigates the identification problem. The consistency and asymptotic distribution of the 2SLS, B2SLS and GMM estimators are established in Section 5, 6 and 7 respectively. Section 8 reports the Monte Carlo results for both the estimation and identification experiments, and finally the last section concludes. Proofs of main theorems and technical details are relegated to the Appendices.

Notation: Let $||\mathbf{A}|| = \sqrt{tr(\mathbf{AA}')}$, $||\mathbf{A}||_\infty = \max\limits_{1\leq i\leq N} \sum_{j=1}^{N} |a_{ij,N}|$ and $||\mathbf{A}||_1 = \max\limits_{1\leq j\leq N} \sum_{i=1}^{N} |a_{ij}|$ denote the Frobenius norm, the maximum row sum norm and maximum column sum norm of a matrix respectively. $\mathbb{M}^{n\times n}$ is the space of $n \times n$ real matrices. K is used generically for a finite positive constant. We say the row (column) sums of a (sequence of) matrix $\mathbf{A} \in \mathbb{M}^{n\times n}$ are uniformly bounded in absolute value, or equivalently we say $\mathbf{A}$ has bounded row (column) norm, if there exists a constant K, such that $||\mathbf{A}||_\infty < K < \infty$ ($||\mathbf{A}||_1 < K < \infty$) for all $n \in \mathbb{N}$ holds. $vec(\mathbf{A})$ is the column vector obtained by stacking the columns of $\mathbf{A}$. $diag(a_1, a_{2,...,}a_N)$ represents a diagonal matrix with diagonal entries $\{a_1, a_{2,...,}a_N\}$, and $diag(\mathbf{A}) = diag(a_{11}, a_{22}, ..., a_{NN})$. $\lambda_{\max}(\mathbf{A})$ ($\lambda_{\min}(\mathbf{A})$) is the largest (smallest) eigenvalue of matrix $\mathbf{A}$, $tr(\mathbf{A})$ denotes the trace, and $\det(\mathbf{A})$ denotes the determinant. $\boldsymbol{\tau}_N$ is an $N \times 1$ vector of ones. $\odot$ stands for the Hadamard products, and $\otimes$ is the Kronecker products. $(N,T) \overset{j}{\to} \infty$ denotes joint convergence of N and T.

Some frequently used notations are also summarized in Appendix A for easy reference.

2 The model

Consider the following spatial autoregressive model (SAR) with common factors

$$y_{it} = \rho \sum_{j=1}^{N} w_{ij} y_{jt} + \boldsymbol{\beta}'\mathbf{x}_{it} + \boldsymbol{\gamma}_i'\mathbf{f}_t + e_{it}, \tag{1}$$

$$\mathbf{x}_{it} = \mathbf{A}_i'\mathbf{f}_t + \mathbf{v}_{it}, \tag{2}$$

for $t = 1, 2, ..., T$, $i = 1, 2, ..., N$, where $\mathbf{x}_{it} = (x_{it,1}, x_{it,2}, ..., x_{it,k})'$ is a $k \times 1$ vector of explanatory variables, $\mathbf{W} = (w_{ij})_{N\times N}$ is a specified spatial weights matrix, and $\mathbf{f}_t$ is an $m \times 1$ vector of unobserved common effects.[1] In what follows, we also use y_{it}^* for brevity to denote $y_{it}^* = \mathbf{w}_i'\mathbf{y}_t$, where $\mathbf{w}_i'$ is the

[1]The model can be easily extended by adding observed factors, such as intercepts, seasonal dummies and deterministic trends, without additional complication.

ith row spatial weights matrix $\mathbf{W}$. For generality, we allow for the possibility that all variables and parameters depend on cross sectional sample size N, that is, they form triangular arrays, although we suppress the subscript N for simplicity of notation.

In stacked form for each time period,

$$\mathbf{y}_{.t} = \rho\mathbf{W}\mathbf{y}_{.t} + \mathbf{X}_{.t}\boldsymbol{\beta} + \boldsymbol{\Gamma}\mathbf{f}_t + \mathbf{e}_{.t}, \;\; t = 1, 2, ..., T, \tag{3}$$

where $\mathbf{y}_{.t} = (y_{1t}, y_{2t}, ..., y_{NT})'$, and $\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, ..., \boldsymbol{\gamma}_N)'$ is an $N \times m$ matrix of factor loadings. Let $\mathbf{S}(\rho) = \mathbf{I}_N - \rho\mathbf{W}$, and if $\mathbf{S}^{-1}(\rho)$ exists, then

$$\mathbf{y}_{.t} = \mathbf{S}^{-1}(\rho)(\mathbf{X}_{.t}\boldsymbol{\beta} + \boldsymbol{\Gamma}\mathbf{f}_t + \mathbf{e}_{.t}). \tag{4}$$

We rewrite the model as

$$\begin{bmatrix} y_{it} - \rho\sum_{j=1}^{N} w_{ij}y_{jt} - \boldsymbol{\beta}'\mathbf{x}_{it} \\ \mathbf{x}_{it} \end{bmatrix} = \boldsymbol{\Phi}_i'\mathbf{f}_t + \mathbf{u}_{it}, \tag{5}$$

where $\boldsymbol{\Phi}_i = [\boldsymbol{\gamma}_i, \mathbf{A}_i]$, $\mathbf{u}_{it} = (e_{it}, \mathbf{v}_{it}')'$. Let $\boldsymbol{\Delta}$ be a square matrix of dimension $N(k+1)$, of which the (i,j)th subblock is given by

$$\boldsymbol{\Delta}_{ij}(\rho, \boldsymbol{\beta}) = \begin{cases} \begin{bmatrix} 1 & -\boldsymbol{\beta}' \\ 0 & \mathbf{I}_k \end{bmatrix} & if \;\; i = j \\ \begin{bmatrix} -\rho w_{ij} & 0 \\ 0 & 0 \end{bmatrix} & if \;\; i \neq j \end{cases} . \tag{6}$$

Then the model can be rewritten more compactly as

$$\boldsymbol{\Delta}(\rho, \boldsymbol{\beta})\mathbf{z}_{.t} = \boldsymbol{\Phi}\mathbf{f}_t + \mathbf{u}_{.t}, \tag{7}$$

where $\mathbf{z}_{.t} = (\mathbf{z}_{1t}, \mathbf{z}_{2t}, ..., \mathbf{z}_{Nt})'$ with $\mathbf{z}_{it} = (y_{it}, \mathbf{x}_{it}')'$, $\boldsymbol{\Phi} = (\boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2, ..., \boldsymbol{\Phi}_N)'$ and $\mathbf{u}_{.t} = (\mathbf{u}_{1t}', \mathbf{u}_{2t}', ..., \mathbf{u}_{Nt}')'$. This way of stacking the equations was motivated by Bai and Li (2014), who have shown that $\boldsymbol{\Delta}^{-1}$ exists,[2] and its (i,j)th subblock is given by

$$\boldsymbol{\Delta}_{ij}^{-1}(\rho, \boldsymbol{\beta}) = \begin{cases} \begin{bmatrix} s_{ii}^{-1} & s_{ii}^{-1}\boldsymbol{\beta}' \\ \mathbf{0} & \mathbf{I}_k \end{bmatrix} & if \;\; i = j \\ \begin{bmatrix} s_{ij}^{-1} & s_{ij}^{-1}\boldsymbol{\beta}' \\ \mathbf{0} & \mathbf{0} \end{bmatrix} & if \;\; i \neq j \end{cases} , \tag{8}$$

where s_{ij}^{-1} denotes the (i,j)th element of $\mathbf{S}^{-1}(\rho) = (\mathbf{I} - \rho\mathbf{W})^{-1}$. Hence, the model can be recast into the following form

[2] See Lemma A.1 of Bai and Li (2014) for a proof.

$$\mathbf{z}_{.t} \quad = \quad \boldsymbol{\Delta}^{-1}(\boldsymbol{\Phi}\mathbf{f}_t + \mathbf{u}_{.t}) = \mathbf{C}'\mathbf{f}_t + \boldsymbol{\epsilon}_{.t}, \tag{9}$$

where $\mathbf{C} = (\boldsymbol{\Delta}^{-1}\boldsymbol{\Phi})'$ and $\boldsymbol{\epsilon}_{.t} = \boldsymbol{\Delta}^{-1}\mathbf{u}_{.t}$.

We follow the idea first proposed by Pesaran (2006) to approximate the unobserved factors with cross-sectional averages. To see why this strategy can be applied to our model with spatial autoregressive effects, let us consider the cross-sectional averages constructed using equal weights, namely $\bar{\mathbf{z}}_{.t} = \boldsymbol{\Theta}_a\mathbf{z}_{.t} = (\bar{y}_t, \bar{\mathbf{x}}'_{.t})'$, where $\boldsymbol{\Theta}_a \equiv \frac{1}{N}\boldsymbol{\tau}'_N \otimes \mathbf{I}_{k+1}$ and $\boldsymbol{\tau}_N$ is the $N \times 1$ vector of ones.[3] Now premultiplying both sides of (9) with $\boldsymbol{\Theta}_a$ yields

$$\bar{\mathbf{z}}_{.t} = \bar{\mathbf{C}}'\mathbf{f}_t + \bar{\boldsymbol{\epsilon}}_{.t}, \tag{10}$$

where

$$\bar{\mathbf{C}} \equiv (\boldsymbol{\Theta}_a\mathbf{C}')' = \frac{1}{N}\left[\sum_{i=1}^{N}\sum_{j=1}^{N} s_{ij}^{-1}(\boldsymbol{\gamma}_j + \mathbf{A}_j\boldsymbol{\beta}),\ \sum_{j=1}^{N}\mathbf{A}_j\right], \tag{11}$$

and $\bar{\boldsymbol{\epsilon}}_{.t} = \boldsymbol{\Theta}_a\boldsymbol{\epsilon}_{.t}$. Assuming that $\bar{\mathbf{C}}$ has full row rank, namely,

$$\text{Rank}(\bar{\mathbf{C}}) = m \leq k+1 \ \text{ for all } N \text{ including } N \to \infty, \tag{12}$$

we then have

$$\mathbf{f}_t = (\bar{\mathbf{C}}\bar{\mathbf{C}}')^{-1}\bar{\mathbf{C}}(\bar{\mathbf{z}}_{.t} - \bar{\boldsymbol{\epsilon}}_{.t}). \tag{13}$$

Importantly, we prove in Lemma 2 that for any t, $\bar{\boldsymbol{\epsilon}}_{.t}$ converges to zero in quadratic mean as $N \to \infty$, which leads to

$$\mathbf{f}_t \xrightarrow{p} (\mathbf{C}^*\mathbf{C}^{*'})^{-1}\mathbf{C}^*\bar{\mathbf{z}}_{.t} \ \text{ as } N \to \infty, \tag{14}$$

where

$$\mathbf{C}^* \equiv \lim_{N\to\infty}\bar{\mathbf{C}} = [E(\boldsymbol{\gamma}_i), E(\mathbf{A}_i)]\begin{bmatrix} s^* & \mathbf{0} \\ s^*\boldsymbol{\beta} & \mathbf{I}_k \end{bmatrix}, \tag{15}$$

and $s^* = \sum_{i=1}^{N}\sum_{j=1}^{N} s_{ij}^{-1}$.[4] Hence, this indicates that $\bar{\mathbf{z}}_{.t}$ can serve fairly well as a proxy for $\mathbf{f}_t$ for N sufficiently large.[5]

Finally, before moving on to the formal analysis, we intend to familiarize readers with the following definitions and notations. Let $\bar{\mathbf{Z}} = [\bar{\mathbf{z}}_{.1}, \bar{\mathbf{z}}_{.2}..., \bar{\mathbf{z}}_{.T}]'$ be the $T \times (k+1)$ matrix of observations on the cross-sectional averages, and $\mathbf{F} = (\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_T)'$ be the $T \times m$ matrices of unobserved common factors. In what follows, we define the "de-factoring" matrix as

[3] The equal weights assumption is nonessential to the analysis, which can be readily carried through with other weighting schemes satisfying the granularity conditions as in Pesaran (2006).

[4] $\mathbf{C}^*$ is nonsingular if the rank of$[E(\boldsymbol{\gamma}_i), E(\mathbf{A}_i)]$ is m.

[5] One may use only $\bar{y}_t$ as a proxy, but including $\bar{\mathbf{x}}_{.t}$ could improve efficiency (if there exist such exogenous variables that are also affected by the factors).

$$\mathbf{M}_f = \mathbf{I}_T - \mathbf{F}(\mathbf{F}'\mathbf{F})^{-}\mathbf{F}', \tag{16}$$

and its observable cross-sectional counterpart as

$$\bar{\mathbf{M}} = \mathbf{I}_T - \bar{\mathbf{Z}}(\bar{\mathbf{Z}}'\bar{\mathbf{Z}})^{-}\bar{\mathbf{Z}}', \tag{17}$$

where $\bar{\mathbf{Z}} = \mathbf{F}\bar{\mathbf{C}} + \bar{\boldsymbol{\epsilon}}$ with $\bar{\boldsymbol{\epsilon}} = (\bar{\boldsymbol{\epsilon}}_{.1}, ..., \bar{\boldsymbol{\epsilon}}_{.T})'$, and $(\mathbf{F}'\mathbf{F})^{-}$ and $(\bar{\mathbf{Z}}'\bar{\mathbf{Z}})^{-}$ denote the generalized inverses. It is convenient to further define $\mathbf{M}_f^b = \mathbf{M}_f \otimes \mathbf{I}_N$, and $\mathbf{M}^b = \bar{\mathbf{M}} \otimes \mathbf{I}_N$ that operate on the stacked vectors $\mathbf{Y} = (\mathbf{y}_{.1}^{'}, \mathbf{y}_{.2}^{'}, ..., \mathbf{y}_{.T}^{'})'$ and $\mathbf{X} = (\mathbf{X}_{.1}^{'}, \mathbf{X}_{.2}^{'}, ..., \mathbf{X}_{.T}^{'})'$.

3 Assumptions

To analyze our model, we make the following assumptions.

Assumption 1. *The unobserved common factors $\mathbf{f}_t$ are covariance stationary with absolutely summable autocovariances, and are distributed independently of $e_{it'}$ and $\mathbf{v}_{it'}$ for all i, t, t'.*

Assumption 2. *The idiosyncratic errors $\mathbf{u}_{it} = (e_{it}, \mathbf{v}_{it}')'$ are such that*

1. *For each i, e_{it} and $\mathbf{v}_{it}$ follow linear stationary processes with absolutely summable autocovariances: $e_{it} = \sum_{l=0}^{\infty} a_{il}\zeta_{i,t-l}$ and $\mathbf{v}_{it} = \sum_{l=0}^{\infty} \boldsymbol{\Xi}_{il}\boldsymbol{\varsigma}_{i,t-l}$, where $(\zeta_{it}, \boldsymbol{\varsigma}_{it}')' \sim IID(0_{k+1}, \mathbf{I}_{k+1})$ with finite fourth-order moments. For the GMM estimation below, it is assumed that e_{it} has absolutely summable cumulants up to fourth order.*

2. *e_{it} is independent of $\mathbf{v}_{jt'}$ for all i, j, t, t'.*

In addition, $Var(e_{it}) = \sum_{l=0}^{\infty} a_{il}^2 = \sigma_i^2 < K < \infty$, $Var(\mathbf{v}_{it}) = \sum_{l=0}^{\infty} \boldsymbol{\Xi}_{il}\boldsymbol{\Xi}_{il}^{'} = \boldsymbol{\Sigma}_{v,i} < K < \infty$, where $\sigma_i^2 > 0$ and $\boldsymbol{\Sigma}_{v,i}$ is positive definite. Accordingly, $Var(\mathbf{u}_{it}) = \boldsymbol{\Sigma}_{u,i} = diag(\sigma_i^2, \boldsymbol{\Sigma}_{v,i})$, and $Var(\mathbf{u}_{.t}) = \boldsymbol{\Sigma}_u = diag(\boldsymbol{\Sigma}_{u,1}, \boldsymbol{\Sigma}_{u,2}, ..., \boldsymbol{\Sigma}_{u,N})$. Both are block-diagonal matrices. For later use, we use $\boldsymbol{\Omega}_{e,i}$ to denote the autocovariance matrix of $\mathbf{e}_{i.}$, i.e., $\boldsymbol{\Omega}_{e,i} = E(\mathbf{e}_{i.}\mathbf{e}_{i.}^{'})$.

Assumption 3. *(i) The slope coefficients $\boldsymbol{\beta}$ are bounded, i.e. $||\boldsymbol{\beta}|| < K$. (ii) The factor loadings $\boldsymbol{\gamma}_i$ and $\mathbf{A}_i$ are bounded uniformly over i, i.e. $||\boldsymbol{\gamma}_i|| < K$ and $||\mathbf{A}_i|| < K$, for all i. They are independently and identically distributed over i with fixed mean and finite variances, and independent of e_{jt}, $\mathbf{v}_{jt}$, $\mathbf{f}_t$ for all i, j, and t. In particular, we assume $\boldsymbol{\gamma}_i = \boldsymbol{\gamma} + \boldsymbol{\eta}_i$, $\boldsymbol{\eta}_i \sim IID(\mathbf{0}, \boldsymbol{\Omega}_\eta)$, for all i, where $\boldsymbol{\Omega}_\eta$ is symmetric non-negative definite.*

Assumption 4. *The matrix $\bar{\mathbf{C}}$, given by (11), has full row rank m ($m \leq k+1$) for all N including $N \to \infty$.*

Assumption 5. *The spatial weights matrix $\mathbf{W}$ has bounded row and column norms, and*

$$|\rho| < max\{1/||\mathbf{W}||_1, 1/||\mathbf{W}||_\infty\} \tag{18}$$

for all $\rho \in \mathbb{P}$, where the parameter space $\mathbb{P}$ is compact. In addition, $\mathbf{W} \neq \mathbf{0}$ and all diagonal elements of $\mathbf{W}$ are zeros, i.e. $w_{ii} = 0$, for all $i = 1, 2, ..., N$.

Assumption 6. *The $N \times q$ instrument matrix $\mathbf{Q}_{.t}$ has full column rank $q \geq k + 1$, and its column dimension q is fixed for all N and t. $\mathbf{Q}_{.t}$ is composed of a subset of linearly independent columns of $(\mathbf{X}_{.t}, \mathbf{W}\mathbf{X}_{.t}, ..., \mathbf{W}^s\mathbf{X}_{.t})$, $s = 1, 2, 3, ...$, and $\mathbf{Q} = (\mathbf{Q}'_{.1}, \mathbf{Q}'_{.2}, ..., \mathbf{Q}'_{.T})'$.*[6]

Assumption 7. *(i) There exists N_0 and T_0, such that for all $N > N_0$ and $T > T_0$, $\frac{1}{NT}\mathbf{Q}'\mathbf{M}^b\mathbf{Q}$ and $\frac{1}{NT}\mathbf{Q}'\mathbf{M}_f^b\mathbf{Q}$ exist and are nonsingular. (ii) $plim_{N\to\infty}\frac{1}{NT}\mathbf{Q}'\mathbf{M}^b\left[(\mathbf{I}_T \otimes \mathbf{W})\mathbf{Y}, \mathbf{X}\right]$ has full column rank $k+1$. (iii) $E|x_{it,p}|^{2+\delta} < K < \infty$ for some $\delta > 0$, all $i = 1, ..., N$, $t = 1, ..., T$, and $p = 1, ..., k$.*

An attractive feature of our model and also advancement over previous studies is that we allow in Assumption 2 the presence of both heteroskedasticity and serial correlation in the disturbance process.[7] The large sample analysis is conducted under this general configuration and we also provide Monte Carlo evidence to show that our proposed estimators are robust to the unknown non-spherical errors for small sample sizes.

Assumption 5 ensures that matrix $\mathbf{S}(\rho) = \mathbf{I}_N - \rho\mathbf{W}$ is nonsingular for all $\rho \in \mathbb{P}$. To see this, note that $\mathbf{S}(\rho)$ is invertible if $|\lambda_{\max}(\rho\mathbf{W})| < 1$. Since $\lambda_{\max}(\rho\mathbf{W}) < |\rho|||\mathbf{W}||_1$ and $\lambda_{\max}(\rho\mathbf{W}) < |\rho|||\mathbf{W}||_\infty$, therefore $\mathbf{S}(\rho)$ is invertible if $|\rho| < \max\{1/||\mathbf{W}||_1, 1/||\mathbf{W}||_\infty\}$. Assumption 5 also implies that $\mathbf{S}^{-1}(\rho)$ is uniformly bounded in row and column sums in absolute value for all $\rho \in \mathbb{P}$, because

$$\begin{aligned} ||\mathbf{S}^{-1}||_1 &= ||\mathbf{I}_N + \rho\mathbf{W} + \rho^2\mathbf{W}^2 + ...||_1 \\ &\leq 1 + |\rho|||\mathbf{W}||_1 + |\rho|^2||\mathbf{W}||_1^2 + ... = \frac{1}{1 - |\rho|||\mathbf{W}||_1} < K < \infty. \end{aligned}$$

Similarly, we can show that $||\mathbf{S}^{-1}||_\infty < K < \infty$.

Assumption 7(i)(ii) are the standard rank conditions for the 2SLS estimator (and the GMM estimator) given below to be well defined asymptotically. The existence of higher-than-second moments in Assumption 7(iii) is required for the GMM estimation to apply a central limit theorem (CLT) for the linear and quadratic form (an extension of **?**). For the other estimation methods, the existence of the second moments would be sufficient.

4 Identification

To begin with, it is important to derive the identification conditions for a joint model containing both spatial interactions and common factors. As we have seen in (14), the unobservable factors $\mathbf{f}_t$ can be well approximated by cross-sectional averages $\bar{\mathbf{z}}_t$ for all values of ρ and $\boldsymbol{\beta}$ satisfying the assumptions set above, with an error being of order $O_p(N^{-1/2})$. Therefore, we will just focus on the identification problem conditional on the factors being observable.

Let us first consider the following SAR model without exogenous explanatory variables $\mathbf{x}_{it}$:

[6] The selection of instruments in Assumption 6 are proposed by **?**.

[7] In future work, we will also add spatial correlations in the error process as an extension.

$$y_{it} = \rho \sum_{j=1}^{N} w_{ij} y_{jt} + \boldsymbol{\gamma}_i^{'} \mathbf{f}_t + e_{it},\ t = 1, 2, ..., T,\ i = 1, 2, ..., N, \tag{19}$$

where $\mathbf{f}_t$ is an $m \times 1$ vector of observable factors, and the errors e_{it} are assumed to be independently and normally distributed with zero means and constant variances for all i and t, i.e. $e_{it} \sim IIDN(0, \sigma^2)$, where $0 < \sigma^2 < K < \infty$. Writing (19) in stacked form, we have

$$\mathbf{y}_{.t} = \rho \mathbf{W} \mathbf{y}_{.t} + \boldsymbol{\Gamma} \mathbf{f}_t + \mathbf{e}_{.t},\ t = 1, 2, ..., T, \tag{20}$$

where $\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, ..., \boldsymbol{\gamma}_N)^{'}$ is an $N \times m$ matrix of factor loadings. Define $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^{'}, \boldsymbol{\gamma}_2^{'}, ..., \boldsymbol{\gamma}_N^{'})^{'}$, and let $\boldsymbol{\varphi}_0 = (\rho_0, \boldsymbol{\gamma}_0^{'}, \sigma_0^2)^{'}$denote the true value of $\boldsymbol{\varphi} = (\rho, \boldsymbol{\gamma}^{'}, \sigma^2)^{'}$. Then the (quasi) log-likelihood function is given by

$$l(\boldsymbol{\varphi}) = -\frac{NT}{2} ln(2\pi) - \frac{NT}{2} ln\sigma^2 + T ln|\mathbf{S}(\rho)| - \frac{1}{2\sigma^2} \sum_{t=1}^{T} \left(\mathbf{S}(\rho)\mathbf{y}_{.t} - \boldsymbol{\Gamma}\mathbf{f}_t\right)^{'} \left(\mathbf{S}(\rho)\mathbf{y}_{.t} - \boldsymbol{\Gamma}\mathbf{f}_t\right), \tag{21}$$

and it follows that

$$\begin{aligned}
\frac{1}{NT} E_0 l(\boldsymbol{\varphi}) &= -\frac{1}{2} ln(2\pi) - \frac{1}{2} ln\sigma^2 + \frac{1}{N} ln|\mathbf{S}(\rho)| \\
&\quad -\frac{1}{2\sigma^2} \left\{ \left(\rho - \rho_0, (\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)^{'}\right) \mathbf{H}_f(\rho_0, \boldsymbol{\gamma}_0^{'}) \left(\rho - \rho_0, (\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)^{'}\right)^{'} + \frac{\sigma_0^2}{N} tr\left[\mathbf{S}_0^{-1} \mathbf{S}(\rho) \mathbf{S}^{'}(\rho) \mathbf{S}_0^{-1'}\right] \right\}, \\
\frac{1}{NT} E_0 l(\boldsymbol{\varphi}_0) &= -\frac{1}{2} \left[ln(2\pi) + 1\right] - \frac{1}{2} ln\sigma_0^2 + \frac{1}{N} ln|\mathbf{S}_0|,
\end{aligned}$$

where

$$\mathbf{H}_f(\rho_0, \boldsymbol{\gamma}_0^{'}) = \frac{1}{NT} E_0 \sum_{t=1}^{T} \left\{ \begin{bmatrix} (\mathbf{G}_0 \boldsymbol{\Gamma}_0 \mathbf{f}_t)^{'} \\ \mathbf{F}_t^{'} \end{bmatrix} \begin{bmatrix} \mathbf{G}_0 \boldsymbol{\Gamma}_0 \mathbf{f}_t, & \mathbf{F}_t \end{bmatrix} \right\}, \tag{22}$$

$\mathbf{G}(\rho) = \mathbf{W}\mathbf{S}^{-1}(\rho)$, $\mathbf{G}_0 = \mathbf{G}(\rho_0) = \mathbf{W}\mathbf{S}_0^{-1}$, $\mathbf{F}_t = \mathbf{I}_N \otimes \mathbf{f}_t^{'}$, and for the discussion of identification we use E_0 to emphasize that the expectation is taken based on the true values of parameters. Letting $Q_{NT}(\boldsymbol{\psi}) = \frac{1}{NT} E_0[l(\boldsymbol{\varphi}_0) - l(\boldsymbol{\varphi})]$, where $\boldsymbol{\psi} = (d, \boldsymbol{\zeta}^{'}, \vartheta)^{'}$, $d = \rho - \rho_0$, $\boldsymbol{\zeta} = \boldsymbol{\gamma} - \boldsymbol{\gamma}_0$, and $\vartheta = (\sigma^2 - \sigma_0^2)/\sigma^2 < 1$, we obtain

$$\begin{aligned}
Q_{NT}(\boldsymbol{\psi}) &= -\frac{1}{2} \left[ln(1 - \vartheta) + \vartheta\right] - \frac{1}{N} ln|\mathbf{I}_N - d\mathbf{G}_0| - \frac{1}{N}(1 - \vartheta) d\, tr(\mathbf{G}_0) + \frac{1}{2}(1 - \vartheta) d^2 \frac{tr(\mathbf{G}_0 \mathbf{G}_0^{'})}{N} \\
&\quad + \frac{1}{2} \sigma_0^2 (1 - \vartheta) \left(d, \boldsymbol{\zeta}^{'}\right) \mathbf{H}_f(\rho_0, \boldsymbol{\gamma}_0^{'}) \left(d, \boldsymbol{\zeta}^{'}\right)^{'}.
\end{aligned} \tag{23}$$

Then, by mean-value expansion and noting that $\partial Q_{NT}(\mathbf{0})/\partial \boldsymbol{\psi} = \mathbf{0}$, we have

$$Q_N(\boldsymbol{\psi}) = \frac{1}{2}\boldsymbol{\psi}' \boldsymbol{\Lambda}_{f,NT}(\bar{\boldsymbol{\psi}})\boldsymbol{\psi}, \tag{24}$$

where $\boldsymbol{\Lambda}_{f,NT}(\boldsymbol{\psi}) \equiv \frac{\partial^2 Q_{NT}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}\boldsymbol{\psi}'}$, a detailed expression of which is given by (C.1) in Appendix C, and $\bar{\boldsymbol{\psi}} = (\bar{d}, \bar{\boldsymbol{\zeta}}', \bar{\vartheta})' = (\bar{\rho} - \rho_0, \bar{\boldsymbol{\gamma}}' - \boldsymbol{\gamma}_0', (\bar{\sigma}^2 - \sigma_0^2)/\bar{\sigma}^2)'$ where $\bar{\rho}$, $\bar{\boldsymbol{\gamma}}$, $\bar{\sigma}^2$ lie between 0 and ρ_0, $\boldsymbol{\gamma}_0$, σ_0^2 respectively. It follows immediately that for all N (including $N \to \infty$) and all T , $\boldsymbol{\psi}_0$ are locally identified if and only if $\lambda_{\min}[\boldsymbol{\Lambda}_{f,NT}(\mathbf{0})] > 0$, where $\boldsymbol{\Lambda}_{f,NT}(\mathbf{0})$ is given by (C.2), and this condition can be further simplified after some algebra (see Appendix C for details). We formalize the identification results in the following theorem.

Theorem 1. *For all N (including $N \to \infty$) and all T, the true parameter values ρ_0, $\boldsymbol{\gamma}_0$ and σ_0^2 of model (19) are locally identified if and only if $\frac{tr(\mathbf{G}_0^2 + \mathbf{G}_0\mathbf{G}_0')}{N} > \frac{2[tr(\mathbf{G}_0)]^2}{N^2}$ and $\frac{1}{T}E_0\left(\mathbf{f}_t\mathbf{f}_t'\right)$ is positive definite.*

Notice that model (19) reduces to a pure spatial autoregressive model if there are no common factors, and the identification condition would simply be

$$\frac{tr(\mathbf{G}_0^2 + \mathbf{G}_0\mathbf{G}_0')}{N} > \frac{2\left[tr(\mathbf{G}_0)\right]^2}{N^2} \text{ for all } N \text{ including } N \to \infty. \tag{25}$$

This condition is in line with the findings in a recent study by Aquaro et al. (2015), who investigated the identification of a spatial model with heterogeneous spatial coefficients without factors. By replacing the heterogeneous coefficients in their identification condition with homogeneous ρ, one would arrive at the same inequality (25). To provide some intuition behind (25), we make the following four observations.

First, it is worth pointing out that a necessary condition for (25) is that there exists an $\varepsilon > 0$ such that

$$\frac{tr(\mathbf{G}_0\mathbf{G}_0')}{N} > \varepsilon > 0, \text{ for all } N \text{ including } N \to \infty. \tag{26}$$

To see this, using Schur's inequality, $tr(\mathbf{G}_0^2)/N \leq tr(\mathbf{G}_0\mathbf{G}_0')/N$, (25) implies that

$$\frac{tr(\mathbf{G}_0\mathbf{G}_0')}{N} > \frac{\left[tr(\mathbf{G}_0)\right]^2}{N^2}. \tag{27}$$

Then, by the Cauchy-Schwarz inequality we have $tr(\mathbf{G}_0\mathbf{G}_0')/N \geq \left[tr(\mathbf{G}_0)\right]^2/N^2$. To exclude the equality (26) is needed, because $tr(\mathbf{G}_0\mathbf{G}_0')/N = 0$ implies $tr(\mathbf{G}_0)/N = 0$ for all N including $N \to \infty$. Also required for the strict inequality is that $\mathbf{G}_0$ cannot be proportional to $\mathbf{I}_N$, which is automatically satisfied under the assumption that all the diagonal elements of $\mathbf{W}$ are zeros, and therefore $\mathbf{G}_0 \neq c\mathbf{I}_N$ for all $c \neq 0$.

Second, under Assumption 5, a necessary and sufficient condition for (26) is that there exists an $\varepsilon > 0$ such that

$$\frac{tr(\mathbf{W}'\mathbf{W})}{N} > \varepsilon > 0, \text{ for all } N \text{ including } N \to \infty. \tag{28}$$

To see why, we note that $\lambda_{\min}\left\{\mathbf{S}^{'}(\rho)\mathbf{S}(\rho)\right\} > 0$, which immediately follows from the non-singularity of $\mathbf{S}(\rho)$, and also

$$\lambda_{\max}\left[\mathbf{S}^{'}(\rho)\mathbf{S}(\rho)\right] \leq ||\mathbf{S}(\rho)||_1||\mathbf{S}(\rho)||_\infty \leq (1+|\rho|||\mathbf{W}||_1)(1+|\rho|||\mathbf{W}||_\infty) < K < \infty. \tag{29}$$

Therefore, we have $\lambda_{\max}\left\{\left[\mathbf{S}^{'}(\rho)\mathbf{S}(\rho)\right]^{-1}\right\} < K < \infty$, and $\lambda_{\min}\left\{\left[\mathbf{S}^{'}(\rho)\mathbf{S}(\rho)\right]^{-1}\right\} > 0$. It then follows that[8]

$$\frac{tr(\mathbf{G}_0\mathbf{G}_0^{'})}{N} = \frac{tr\left[(\mathbf{S}_0^{'}\mathbf{S}_0)^{-1}\mathbf{W}'\mathbf{W}\right]}{N} \leq \lambda_{\max}\left\{(\mathbf{S}_0^{'}\mathbf{S}_0)^{-1}\right\}\frac{tr(\mathbf{W}'\mathbf{W})}{N} < K\frac{tr(\mathbf{W}'\mathbf{W})}{N}, \tag{30}$$

which establishes necessity, and

$$\frac{tr(\mathbf{G}_0\mathbf{G}_0^{'})}{N} = \frac{tr\left[(\mathbf{S}_0^{'}\mathbf{S}_0)^{-1}\mathbf{W}'\mathbf{W}\right]}{N} \geq \lambda_{\min}\left\{(\mathbf{S}_0^{'}\mathbf{S}_0)^{-1}\right\}\frac{tr(\mathbf{W}'\mathbf{W})}{N}, \tag{31}$$

which establishes sufficiency. As a simple necessary condition for identification, (28) does not depend on any unknown parameters and can be easily applied to check identification in practice.

Third, it is readily seen that (26) simplifies to $\mathbf{W} \neq \mathbf{0}$ when N is finite. Also of interest is the fact that $\rho_0 = 0$ is identified if $tr(\mathbf{W}^{'}\mathbf{W})/N > \varepsilon > 0$, which can be established by replacing $\mathbf{G}_0$ with $\mathbf{W}$ and $tr(\mathbf{G}_0) = 0$ in (25), so (26) is both necessary and sufficient for identification in this case.

Fourth, if there were no common factors, our model would become a special case of the spatial Durbin models considered by Lee and Yu (2015). They restrict their attention to the finite sample case and argue that $\boldsymbol{\varphi}_0$ are identified if $\mathbf{I}_N$, $\mathbf{W}+\mathbf{W}'$ and $\mathbf{W}^{'}\mathbf{W}$ are linearly independent. However, it is possible that the necessary condition (26) we found is violated while the independence condition by Lee and Yu (2015) still holds when N is large. As an illustration, we will construct such an example in Section 8 and provide Monte Carlo evidence to show that $\boldsymbol{\varphi}_0$ are not identified in this case.

We now proceed to add exogenous explanatory variables $\mathbf{x}_{it}$ to (19) and consider the following model

$$y_{it} = \rho\sum_{j=1}^{N} w_{ij}y_{jt} + \boldsymbol{\beta}'\mathbf{x}_{it} + \boldsymbol{\gamma}_i'\mathbf{f}_t + e_{it}. \tag{32}$$

In contrast with model (1)(2), here we assume that $\mathbf{x}_{it}$ are uncorrelated with $\mathbf{f}_t$ for all i and t, and $e_{it} \sim IIDN(0,\sigma^2)$. Although it seems that $\mathbf{f}_t$ are nothing special but simply a subgroup of $\mathbf{x}_{it}$, their cross-sectionally invariability in fact carry notable implications for the identification of ρ_0. With a slight abuse of notation, we use the same letter $\boldsymbol{\varphi}$ to denote the parameters of this model, $\boldsymbol{\varphi} = (\rho, \boldsymbol{\beta}^{'}, \boldsymbol{\gamma}^{'}, \sigma^2)'$, and its true value is denoted as $\boldsymbol{\varphi}_0 = (\rho_0, \boldsymbol{\beta}_0^{'}, \boldsymbol{\gamma}_0^{'}, \sigma_0^2)'$. By similar reasoning, we proclaim the following identification theorem, whose proof is provided in Appendix C.

Theorem 2. *Consider the model given by (32), where $\mathbf{x}_{it}$ are exogenous and uncorrelated with $\mathbf{f}_t$ for all i and t. For all N (including $N \to \infty$) and all T, the true parameter values ρ_0 and σ_0^2 are locally*

[8] For real symmetric matrix $\mathbf{A}$ and real positive-semidefinite matrix $\mathbf{B}$ of the same size, $\lambda_{\min}(\mathbf{A})tr(\mathbf{B}) \leq tr(\mathbf{AB}) \leq \lambda_{\max}(\mathbf{A})tr(\mathbf{B})$.

identified if $\frac{tr(\mathbf{G}_0^2+\mathbf{G}_0\mathbf{G}_0')}{N} > \frac{2[tr(\mathbf{G}_0)]^2}{N^2}$ *or/and if* $\mathbf{H}(\rho_0, \boldsymbol{\beta}_0')$ *is positive definite, where* $\mathbf{H}(\rho_0, \boldsymbol{\beta}_0')$ *is given by (C.5). Given* ρ_0*,* $\boldsymbol{\beta}_0$ *are identified if* $\frac{1}{NT}E_0(\mathbf{X}'\mathbf{X})$ *is positive definite.* $\boldsymbol{\gamma}_0$ *is identified if* $\frac{1}{T}E_0\left(\mathbf{f}_t\mathbf{f}_t'\right)$ *is positive definite.*

Remark 1. Note that if $\mathbf{H}(\rho_0, \boldsymbol{\beta}_0')$ is positive definite, then both ρ_0 and $\boldsymbol{\beta}_0$ are identified. Compared with the identification condition for the pure SAR model, including exogenous variables $\mathbf{x}_{it}$ that vary across individuals introduces an additional means to identify ρ_0, however, adding observed common factors does not help. This is not surprising because the latter does not introduce information regarding cross-sectional variation.

Remark 2. If there are no common factors, model (1) would reduce to a SAR model with exogenous regressors, and Theorem 2 provides the identification condition for ρ_0, $\boldsymbol{\beta}_0$ and σ_0^2 , which is valid even if $N \to \infty$ (cf. Lee and Yu, 2015 for finite N).

Finally, let us go back to our main model (1), where $\mathbf{x}_{it}$ may be correlated with $\mathbf{f}_t$ and follow the process given by (2). Suppose that we are only interested in identifying ρ_0 and $\boldsymbol{\beta}_0$, as is the case in the following analysis, we can partialling out the effect of $\mathbf{f}_t$ by premultiplying (3) with $\mathbf{M}_f$. By Frisch-Waugh-Lovell Theorem, the identification conditions can be readily established as a corollary to Theorem 2.

Corollary 1. *Consider model (1), where* $\mathbf{x}_{it}$ *is given by (2). For all* N *(including* $N \to \infty$*) and all* T*, the true parameter values* ρ_0 *are locally identified if* $\frac{tr(\mathbf{G}_0^2+\mathbf{G}_0\mathbf{G}_0')}{N} > \frac{2[tr(\mathbf{G}_0)]^2}{N^2}$ *or/and if* $\mathring{\mathbf{H}}(\rho_0, \boldsymbol{\beta}_0')$ *is positive definite, where* $\mathring{\mathbf{H}}(\rho_0, \boldsymbol{\beta}_0')$ *is given by*

$$\mathring{\mathbf{H}}(\rho_0, \boldsymbol{\beta}_0') = \frac{1}{NT}E_0\left\{\begin{bmatrix} (\mathbf{G}_0^b\mathbf{M}_f^b\mathbf{X}\boldsymbol{\beta}_0)' \\ (\mathbf{M}_f^b\mathbf{X})' \end{bmatrix}\begin{bmatrix} \mathbf{G}_0^b\mathbf{M}_f^b\mathbf{X}\boldsymbol{\beta}_0, & \mathbf{M}_f^b\mathbf{X} \end{bmatrix}\right\} \tag{33}$$

Given ρ_0*,* $\boldsymbol{\beta}_0$ *are identified if* $\frac{1}{NT}E_0(\mathbf{X}'\mathbf{M}_f^b\mathbf{X})$ *is positive definite, which is ensured if* $\mathring{\mathbf{H}}(\rho_0, \boldsymbol{\beta}_0')$ *is positive definite.*

5 2SLS Estimation

5.1 Consistency

In this section, we discuss the 2SLS estimation of model (1)(2) based on the IV matrix $\mathbf{Q}$ specified in Assumption 6. The (pooled) 2SLS estimator for $\boldsymbol{\delta}_0 = (\rho_0, \boldsymbol{\beta}_0')'$ is given by

$$\hat{\boldsymbol{\delta}}_{2sls} = \left[\mathbf{L}'\mathbf{P}_Q\mathbf{L}\right]^{-1}\left[\mathbf{L}'\mathbf{P}_Q\mathbf{Y}\right], \tag{34}$$

where $\mathbf{P}_Q = \mathbf{M}^b\mathbf{Q}(\mathbf{Q}'\mathbf{M}^b\mathbf{Q})^{-1}\mathbf{Q}'\mathbf{M}^b$, $\mathbf{L} = [\mathbf{Y}^*, \mathbf{X}]$ and $\mathbf{Y}^* = (\mathbf{I}_T \otimes \mathbf{W})\mathbf{Y}$. The 2SLS estimator can be interpreted as first "de-factoring" the data with cross-sectional averages, i.e. $\mathring{\mathbf{Y}} = \mathbf{M}^b\mathbf{Y}$, and $\mathring{\mathbf{L}} = \mathbf{M}^b\mathbf{L}$, and then applying the standard 2SLS estimation using $\mathring{\mathbf{Y}}$ and $\mathring{\mathbf{L}}$. Alternatively, one may think of $\mathbf{M}^b\mathbf{Q}$ directly as the instruments.

First, we show that the 2SLS estimator $\hat{\boldsymbol{\delta}}_{2sls}$ given by (34) is consistent as $N \to \infty$, for T fixed or $T \to \infty$. Note that

$$\hat{\boldsymbol{\delta}}_{2sls} - \boldsymbol{\delta}_0 \quad = \quad \left[\mathbf{L}^{'}\mathbf{P}_Q\mathbf{L}\right]^{-1}\mathbf{L}^{'}\mathbf{P}_Q\left[(\mathbf{I}_T \otimes \boldsymbol{\Gamma}_0)\mathbf{f} + \mathbf{e}\right],$$

and then

$$\begin{aligned}\sqrt{NT}(\hat{\boldsymbol{\delta}}_{2sls} - \boldsymbol{\delta}_0) \quad &= \quad \left[\frac{1}{NT}\mathbf{L}'\mathbf{M}^b\mathbf{Q}\left(\frac{1}{NT}\mathbf{Q}'\mathbf{M}^b\mathbf{Q}\right)^{-1}\frac{1}{NT}\mathbf{Q}'\mathbf{M}^b\mathbf{L}\right]^{-1} \\ &\times\left[\frac{1}{NT}\mathbf{L}'\mathbf{M}^b\mathbf{Q}\left(\frac{1}{NT}\mathbf{Q}'\mathbf{M}^b\mathbf{Q}\right)^{-1}\frac{1}{\sqrt{NT}}\mathbf{Q}'\mathbf{M}^b\left[(\mathbf{I}_T \otimes \boldsymbol{\Gamma}_0)\mathbf{f} + \mathbf{e}\right]\right].\end{aligned}$$

By Lemma 6, we have

$$\boldsymbol{\Psi}_{QMQ} \equiv \underset{N\to\infty}{plim}\frac{1}{NT}\mathbf{Q}'\mathbf{M}^b\mathbf{Q} = \underset{N\to\infty}{plim}\frac{1}{NT}\mathbf{Q}'\mathbf{M}_f^b\mathbf{Q}, \tag{35}$$

$$\boldsymbol{\Psi}_{QML} = \underset{N\to\infty}{plim}\frac{1}{NT}\mathbf{Q}'\mathbf{M}^b\mathbf{L} = \underset{N\to\infty}{plim}\frac{1}{NT}\mathbf{Q}'\mathbf{M}_f^b\mathbf{L}_0, \tag{36}$$

where

$$\mathbf{L}_0 = [(\mathbf{I}_T \otimes \mathbf{G}_0)\mathbf{X}\boldsymbol{\beta}_0, \mathbf{X}], \tag{37}$$

and it follows that

$$\boldsymbol{\Psi}_{LPL} \equiv \underset{N\to\infty}{plim}\frac{1}{NT}\mathbf{L}'\mathbf{P}_Q\mathbf{L} = \underset{N\to\infty}{plim}\frac{1}{NT}\mathbf{L}_0'\mathbf{P}_{Q,f}\mathbf{L}_0 \tag{38}$$

where $\mathbf{P}_{Q,f} = \mathbf{M}_f^b\mathbf{Q}(\mathbf{Q}'\mathbf{M}_f^b\mathbf{Q})^{-1}\mathbf{Q}'\mathbf{M}_f^b$. Under Assumption 7 and the identification theorem, the limit in (38) exists and is nonsingular. Furthermore, we have shown in Lemma 6 that $\underset{N\to\infty}{plim}\frac{1}{NT}\mathbf{Q}'\mathbf{M}_f^b[(\mathbf{I}_T \otimes \boldsymbol{\Gamma}_0)\mathbf{f} + \mathbf{e}] = 0$. Hence, $\hat{\boldsymbol{\delta}}_{2sls}$ is consistent as $N \to \infty$.

5.2 Asymptotic Distribution

The following theorem presents the asymptotic distribution of $\hat{\boldsymbol{\delta}}_{2sls}$. The proof is provided in the appendix.

Theorem 3. *Consider model (1)(2). Suppose that Assumptions 1-7 hold, the 2SLS estimator* $\hat{\boldsymbol{\delta}}_{2sls}$ *defined by (34) is consistent for* $\boldsymbol{\delta}_0$ *as* $N \to \infty$, *for* T *fixed or* $T \to \infty$. *Moreover, as* $(N,T) \overset{j}{\to} \infty$ *and* $T/N \to 0$, *we have*

$$\sqrt{NT}(\hat{\boldsymbol{\delta}}_{2sls} - \boldsymbol{\delta}_0) \overset{d}{\to} N(\mathbf{0}, \boldsymbol{\Sigma}_{2sls}), \tag{39}$$

where

$$\boldsymbol{\Sigma}_{2sls} = \boldsymbol{\Psi}_{LPL}^{-1}\boldsymbol{\Omega}_{LPe}\boldsymbol{\Psi}_{LPL}^{-1}, \tag{40}$$

$$\mathbf{\Omega}_{LPe} = \mathbf{\Psi}_{QML}^{'} \mathbf{\Psi}_{QMQ}^{-1} \mathbf{\Omega}_{QMe} \mathbf{\Psi}_{QMQ}^{-1} \mathbf{\Psi}_{QML}, \tag{41}$$

$\mathbf{\Psi}_{QMQ}$, $\mathbf{\Psi}_{QML}$, $\mathbf{\Psi}_{LPL}$ are defined in (35) (36) (38) respectively, and

$$\mathbf{\Omega}_{QMe} = \lim_{N \to \infty} \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{\Omega}_{iQMe} \right), \tag{42}$$

$$\mathbf{\Omega}_{iQMe} = \operatorname*{plim}_{T \to \infty} \left[\frac{1}{T} \mathbf{Q}_{i.}^{'} \mathbf{M}_f \mathbf{\Omega}_{e,i} \mathbf{M}_f \mathbf{Q}_{i.} \right], \tag{43}$$

where $\mathbf{Q}_{i.} = (\mathbf{Q}_{i1}, \mathbf{Q}_{i2}, ..., \mathbf{Q}_{iT})'$.

A consistent estimator for the asymptotic variance matrix $\mathbf{\Sigma}_{2sls}$ is given by

$$\hat{\mathbf{\Sigma}}_{2sls} = \left(\frac{1}{NT} \mathbf{L} \mathbf{P}_Q \mathbf{L} \right)^{-1} \hat{\mathbf{\Omega}}_{LPe} \left(\frac{1}{NT} \mathbf{L} \mathbf{P}_Q \mathbf{L} \right)^{-1}, \tag{44}$$

where $\hat{\mathbf{\Omega}}_{LPe}$ can be obtained by following a Newey-West type procedure as follows

$$\hat{\mathbf{\Omega}}_{LPe} = \frac{1}{N} \sum_{i=1}^{N} \hat{\mathbf{\Omega}}_{iLPe}, \tag{45}$$

$$\hat{\mathbf{\Omega}}_{iLPe} = \hat{\mathbf{\Omega}}_{iLPe,0} + \sum_{h=1}^{M_l} \left(1 - \frac{h}{M_l + 1} \right) \left(\hat{\mathbf{\Omega}}_{iLPe,h} + \hat{\mathbf{\Omega}}_{iLPe,h}^{'} \right), \tag{46}$$

$$\hat{\mathbf{\Omega}}_{iLPe,h} = \frac{1}{T} \sum_{t=h+1}^{T} \hat{e}_{it} \hat{e}_{i,t-h} \hat{\mathbf{L}}_{it} \hat{\mathbf{L}}_{i,t-h}^{'}, \tag{47}$$

where M_l is the maximum lag length, $\hat{e}_{it}$ are the elements of $\hat{\mathbf{e}} = \mathbf{M}^b \left(\mathbf{Y} - \mathbf{L} \hat{\boldsymbol{\delta}}_{2sls} \right)$, and $\hat{\mathbf{L}}_{it}$ are the elements of $\hat{\mathbf{L}} = \mathbf{P}_Q \mathbf{L}$.

6 B2SLS Estimation

Now a question naturally arises whether best instruments are available for the model, or how many variables should be included from the list $\{\mathbf{X}_{.t}, \mathbf{W}\mathbf{X}_{.t}, \mathbf{W}^2\mathbf{X}_{.t}, ...\}$. Lee (2003) suggests a best generalized 2SLS estimator for a cross-sectional spatial model that constructs the best IVs based on consistent initial estimates of the parameters and then re-estimate the model. He shows that the estimator is the "best" in the sense that it has the smallest asymptotic variance among all the IV estimators of the model. In light of this idea, we investigate if using

$$\hat{\mathbf{Q}}^* = \mathbf{M}^b \left[\left(\mathbf{I}_T \otimes \mathbf{G}(\hat{\rho}_{2sls}) \right) \mathbf{X} \hat{\boldsymbol{\beta}}_{2sls}, \mathbf{X} \right] \tag{48}$$

as an IV matrix, where $\hat{\rho}_{2sls}$ and $\hat{\boldsymbol{\beta}}_{2sls}$ are the first-step 2SLS estimates (or maybe some other consistent estimates), to obtain the best 2SLS (B2SLS) estimator,

$$\hat{\boldsymbol{\delta}}_{b2sls} = \left[\hat{\mathbf{Q}}^{*\prime}\mathbf{L}\right]^{-1}\left[\hat{\mathbf{Q}}^{*\prime}\mathbf{Y}\right], \tag{49}$$

can achieve the least asymptotic variance for our model. The intuition behind the best IV $\hat{\mathbf{Q}}^*$ is straightforward, namely, to exploit the part of $\mathbf{Y}^*$ that is uncorrelated with the errors. The following theorem gives the asymptotic properties of $\hat{\boldsymbol{\delta}}_{b2sls}$.

Theorem 4. *Consider model (1)(2). Under the same assumptions as in Theorem 3, the B2SLS estimator $\hat{\boldsymbol{\delta}}_{b2sls}$ defined by (49) is consistent for $\boldsymbol{\delta}_0$ as $N \to \infty$, for T fixed or $T \to \infty$; and as $(N,T) \xrightarrow{j} \infty$ and $T/N \to 0$, it has the following distribution*

$$\sqrt{NT}(\hat{\boldsymbol{\delta}}_{b2sls} - \boldsymbol{\delta}_0) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}_{b2sls}), \tag{50}$$

where

$$\boldsymbol{\Sigma}_{b2sls} = \boldsymbol{\Psi}_{LML}^{-1}\boldsymbol{\Omega}_{LMe}\boldsymbol{\Psi}_{LML}^{-1}, \tag{51}$$

$$\boldsymbol{\Psi}_{LML} \equiv \underset{N\to\infty}{plim}\frac{1}{NT}\mathbf{L}_0'\mathbf{M}_f^b\mathbf{L}_0, \tag{52}$$

$$\boldsymbol{\Omega}_{LMe} = \lim_{N\to\infty}\left(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\Omega}_{iLMe}\right), \tag{53}$$

$$\boldsymbol{\Omega}_{iLMe} = \underset{T\to\infty}{plim}\left[\frac{1}{T}\mathbf{L}_{0,i.}^{\prime}\mathbf{M}_f\boldsymbol{\Omega}_{e,i}\mathbf{M}_f\mathbf{L}_{0,i.}\right], \tag{54}$$

and $\mathbf{L}_0$ is defined in (37). $\hat{\boldsymbol{\delta}}_{b2sls}$ is the best IV estimator if the disturbances $\{e_{it}\}$ of model (1) are independently and identically distributed with mean zero and variance σ_e^2.

Note that under Assumption 5, $(\mathbf{I}_N - \rho\mathbf{W})^{-1}\mathbf{X}_{.t}\boldsymbol{\beta} = \sum_{i=1}^{N}\rho^i\mathbf{W}^i\mathbf{X}_{.t}\boldsymbol{\beta}$. Hence, the best IV $\mathbf{Q}^*$ can be approximated by linear combinations of $\mathbf{M}^b\mathbf{Q} = \left(\mathbf{X}, (\mathbf{I}_T\otimes\mathbf{W})\mathbf{X}, (\mathbf{I}_T\otimes\mathbf{W}^2)\mathbf{X}, ...\right)$. Clearly, the higher power of $\mathbf{W}$ included, the better the approximation. However, in practice the efficiency gain by including more instruments may not be significant and one might as well use $\hat{\boldsymbol{\delta}}_{2sls}$ instead of $\hat{\boldsymbol{\delta}}_{b2sls}$. In the Monte Carlo experiments, we will compare the finite sample performance of $\hat{\boldsymbol{\delta}}_{2sls}$ based on instruments$(\mathbf{X}_{.t}, \mathbf{W}\mathbf{X}_{.t})$ with that of $\hat{\boldsymbol{\delta}}_{b2sls}$ based on $\hat{\mathbf{Q}}_{t.}^*$.

7 GMM Estimation

In this section, we consider GMM estimation of the model. We construct the moment conditions to include r quadratic moments and q linear moments, following Lee (2007) and subsequent studies (Lin and Lee, 2010; Lee and Yu, 2014):

$$\mathbf{g}_{NT}(\boldsymbol{\delta}) = \begin{bmatrix} \boldsymbol{\xi}'(\boldsymbol{\delta})\mathbf{M}^b\mathbf{P}_1^b\mathbf{M}^b\boldsymbol{\xi}(\boldsymbol{\delta}) \\ \vdots \\ \boldsymbol{\xi}'(\boldsymbol{\delta})\mathbf{M}^b\mathbf{P}_r^b\mathbf{M}^b\boldsymbol{\xi}(\boldsymbol{\delta}) \\ \mathbf{Q}'\mathbf{M}^b\boldsymbol{\xi}(\boldsymbol{\delta}) \end{bmatrix}, \tag{55}$$

where

$$\boldsymbol{\xi}(\boldsymbol{\delta}) = [\mathbf{I}_T \otimes \mathbf{S}(\rho)]\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}, \tag{56}$$

and $\boldsymbol{\delta} = (\rho, \boldsymbol{\beta}')'$ represent unknown parameters in the parameter space Δ_{sp} where the identification conditions for the model hold. This implies that $p\lim_{N\to\infty} \frac{1}{NT}\mathbf{g}_{NT}(\boldsymbol{\delta}) = \mathbf{0}$ has a unique solution at $\boldsymbol{\delta}_0$ in Δ_{sp}. For $l = 1, 2, ..., r$, we define $\mathbf{P}_l^b = \mathbf{I}_T \otimes \mathbf{P}_l$, where $\mathbf{P}_l$ is an $N \times N$ matrix satisfying $diag(\mathbf{P}_l) = \mathbf{0}$.

Essentially, the idea of the quadratic moments is to achieve zero expectations by introducing some matrix $\mathbf{P}_l$ such that it eliminates the correlations among the elements of $\mathbf{e}_{.t}$. As an illustrative example, imagine that if there are no factors, at $\boldsymbol{\delta}_0$ the (population) quadratic moments would be simplified to $E(\mathbf{e}'\mathbf{P}_l^b\mathbf{e}) = \sum_{i=1}^N \sum_{j=1}^N p_{l,ji} E(\mathbf{e}_{i.}'\mathbf{e}_{j.}) = \sum_{i=1}^N p_{l,ii} E(\mathbf{e}_{i.}'\mathbf{e}_{i.}) = 0$, where $p_{l,ji}$ is the (j,i)th element of matrix $\mathbf{P}_l$, and the last equality follows from the property $diag(\mathbf{P}_l) = \mathbf{0}$. As we can see, the moment conditions are built on the key assumption of the cross-sectional uncorrelatedness between $\mathbf{e}_{i.}$ and $\mathbf{e}_{j.}$ $(j \neq i)$, and since we allow unknown heteroskedasticity, we need all diagonal elements of $\mathbf{P}_l$ to be zeros to remove the variances of e_{it} from the moments. By contrast, imposing $tr(\mathbf{P}_l) = 0$ would be sufficient if e_{it} are homoskedastic (see, for example, Lee, 2007 and Lee and Yu, 2014). The utilization of the quadratic moments in addition to the 2SLS-type linear moments not only could improve efficiency, but also makes it possible to estimate the spatial coefficient when there are no exogenous regressors or they are barely relevant.

Note that we are only using the aggregated moment conditions over time instead of a moment condition for each period separately, because the latter approach induces the many-moment bias problem (Lee and Yu, 2014) and is left for future research.

Given the moments (55), the GMM estimator $\hat{\boldsymbol{\delta}}_{GMM}$ is defined as

$$\hat{\boldsymbol{\delta}}_{GMM} = \underset{\boldsymbol{\delta}\in\Delta_{sp}}{\operatorname{argmin}}\ \mathbf{g}_{NT}'(\boldsymbol{\delta})\mathbf{A}_{NT}^{w\prime}\mathbf{A}_{NT}^{w}\mathbf{g}_{NT}(\boldsymbol{\delta}), \tag{57}$$

where $\mathbf{A}_{NT}^w = (\mathbf{a}^{(1)}, \mathbf{a}^{(2)}, ..., \mathbf{a}^{(r)}, \mathbf{A}^{(Q)})$ is a $k_a \times (r+q)$ matrix $(k_a \geq k+1)$. Finally, we make the following additional assumption for the GMM estimation and then present the asymptotic results.

Assumption 8. *Given the moment conditions (55), we assume that $\mathbf{P}_l$ $(l = 1, 2, ..., r)$ is nonstochastic and has bounded row and column norms. $\mathbf{A}_{NT}^w$ is assumed to converge in probability to a constant full rank matrix $\mathbf{A}_0^w$.*

Theorem 5. *Consider model (1)(2). Suppose that Assumptions 1-8 hold, the GMM estimator $\hat{\boldsymbol{\delta}}_{GMM}$ defined by (57) is consistent for $\boldsymbol{\delta}_0$ as $N \to \infty$, for T fixed or $T \to \infty$. Furthermore, as $(N,T) \xrightarrow{j} \infty$ and $T/N \to 0$, we have*

$$\sqrt{NT}(\hat{\boldsymbol{\delta}}_{GMM}-\boldsymbol{\delta}_0)\xrightarrow{d}N(\mathbf{0},\boldsymbol{\Sigma}_{GMM}), \tag{58}$$

where

$$\boldsymbol{\Sigma}_{GMM} = \underset{N,T\to\infty}{plim}\left(\mathbf{D}'\mathbf{A}^{w'}_{NT}\mathbf{A}^{w}_{NT}\mathbf{D}\right)^{-1}\mathbf{D}'\mathbf{A}^{w'}_{NT}\mathbf{A}^{w}_{NT}\boldsymbol{\Sigma}_g\mathbf{A}^{w'}_{NT}\mathbf{A}^{w}_{NT}\mathbf{D}\left(\mathbf{D}'\mathbf{A}^{w'}_{NT}\mathbf{A}^{w}_{NT}\mathbf{D}\right)^{-1}, \tag{59}$$

$$\mathbf{D}=\frac{1}{NT}E\begin{bmatrix} T\sum_{i=1}^{N}\tilde{g}^s_{ii,1}\sigma_i^2 & ... & T\sum_{i=1}^{N}\tilde{g}^s_{ii,r}\sigma_i^2 & \{(\mathbf{M}_f\otimes\mathbf{G}_0)\mathbf{X}\boldsymbol{\beta}_0\}'\mathbf{Q} \\ 0 & ... & 0 & \mathbf{X}'\mathbf{M}^b_f\mathbf{Q}\end{bmatrix}', \tag{60}$$

and

$$\begin{aligned}\boldsymbol{\Sigma}_g &= Var\left(\frac{1}{\sqrt{NT}}\mathbf{g}_{NT}(\boldsymbol{\delta}_0)\right)\\ &= \frac{1}{NT}\begin{bmatrix} tr\left[(\mathbf{P}_1\odot\mathbf{P}^s_1)\boldsymbol{\Sigma}_{ee}\right] & tr\left[(\mathbf{P}_1\odot\mathbf{P}^s_2)\boldsymbol{\Sigma}_{ee}\right] & \cdots & 0\\ tr\left[(\mathbf{P}_2\odot\mathbf{P}^s_1)\boldsymbol{\Sigma}_{ee}\right] & tr\left[(\mathbf{P}_2\odot\mathbf{P}^s_2)\boldsymbol{\Sigma}_{ee}\right] & \cdots & 0\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & \boldsymbol{\Omega}_{QMe}\end{bmatrix}\\ &= \frac{1}{NT}\begin{bmatrix} \sum_{i=1}^{N}\sum_{j=1}^{N}p_{1,ji}(p_{1,ij}+p_{1,ji})tr[\boldsymbol{\Omega}_{e,i}\boldsymbol{\Omega}_{e,j}] & * & \cdots & 0\\ \sum_{i=1}^{N}\sum_{j=1}^{N}p_{2,ji}(p_{1,ij}+p_{1,ji})tr[\boldsymbol{\Omega}_{e,i}\boldsymbol{\Omega}_{e,j}] & * & \cdots & 0\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & \boldsymbol{\Omega}_{QMe}\end{bmatrix},\end{aligned} \tag{61}$$

where $\tilde{g}^s_{ii,l}$ *($l=1,...,r$) is the (i,i)th element of matrix* $\tilde{\mathbf{G}}_l(\rho_0)=\mathbf{P}^s_l\mathbf{G}_0$*,* $\mathbf{P}^s_l=\mathbf{P}_l+\mathbf{P}'_l$*,* $\boldsymbol{\Sigma}_{ee}$ *is an* $N\times N$ *matrix of which the (i,j)th element is* $tr[\boldsymbol{\Omega}_{e,i}\boldsymbol{\Omega}_{e,j}]$*, and* $\boldsymbol{\Omega}_{QMe}$ *is defined by (42).*

The next theorem states the asymptotic distribution of the efficient GMM estimator obtained using the optimal weighting matrix $\boldsymbol{\Sigma}_g^{-1}$.

Theorem 6. *Under the same assumptions as in Theorem 5, the efficient GMM estimator,* $\hat{\boldsymbol{\delta}}^*_{GMM}=\underset{\boldsymbol{\delta}}{argmin}\ \mathbf{g}'_{NT}(\boldsymbol{\delta})\boldsymbol{\Sigma}_g^{-1}\mathbf{g}_{NT}(\boldsymbol{\delta})$*, has the following asymptotic distribution as* $(N,T)\xrightarrow{j}\infty$ *and* $T/N\to 0$*,*

$$\sqrt{NT}(\hat{\boldsymbol{\delta}}^*_{GMM}-\boldsymbol{\delta}_0)\xrightarrow{d}N(\mathbf{0},\boldsymbol{\Sigma}^*_{GMM}), \tag{62}$$

where

$$\boldsymbol{\Sigma}^*_{GMM}=\underset{N,T\to\infty}{plim}\left(\mathbf{D}'\boldsymbol{\Sigma}_g^{-1}\mathbf{D}\right)^{-1}. \tag{63}$$

A consistent estimator of $\boldsymbol{\Sigma}_{GMM}$ can be obtained by replacing $\mathbf{D}$ and $\boldsymbol{\Sigma}_g$ in (59) with $\hat{\mathbf{D}}$ and $\hat{\boldsymbol{\Sigma}}_g$

respectively:

$$\hat{\mathbf{D}} = \frac{1}{NT}\begin{bmatrix} \sum_{i=1}^{N}\hat{\tilde{g}}_{ii,1}^{s}\hat{\mathbf{e}}_{i.}'\hat{\mathbf{e}}_{i.} & ... & \sum_{i=1}^{N}\hat{\tilde{g}}_{ii,r}^{s}\hat{\mathbf{e}}_{i.}'\hat{\mathbf{e}}_{i.} & \{(\mathbf{I}_T\otimes\mathbf{W})\mathbf{Y}\}'\mathbf{M}^b\mathbf{Q} \\ 0 & ... & 0 & \mathbf{X}'\mathbf{M}^b\mathbf{Q} \end{bmatrix}', \tag{64}$$

$$\hat{\mathbf{\Sigma}}_g = \frac{1}{NT}\begin{bmatrix} \sum_{i=1}^{N}\sum_{j=1}^{N}p_{1,ji}(p_{1,ij}+p_{1,ji})\hat{s}_{e,ij} & * & \cdots & 0 \\ \sum_{i=1}^{N}\sum_{j=1}^{N}p_{2,ji}(p_{1,ij}+p_{1,ji})\hat{s}_{e,ij} & * & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \hat{\mathbf{\Omega}}_{QMe} \end{bmatrix}, \tag{65}$$

where $\hat{\mathbf{e}} = \mathbf{M}^b\left(\mathbf{Y}-\mathbf{L}\hat{\boldsymbol{\delta}}_{GMM}\right)$, $\hat{\tilde{g}}_{ii,l}^{s}$ is the (i,i)th element of $\tilde{\mathbf{G}}_l(\hat{\rho})$, $\hat{s}_{e,ij} = T\hat{\gamma}_{e,i}(0)\hat{\gamma}_{e,j}(0)+2\sum_{h=1}^{M_l}(T-h)(1-\frac{h}{M_l+1})\hat{\gamma}_{e,i}(h)\hat{\gamma}_{e,j}(h)$, $\hat{\gamma}_{e,i}(h) = \frac{1}{T}\sum_{t=h+1}^{T}\hat{e}_{it}\hat{e}_{i,t-h}$ and M_l is the maximum lag. Similarly, we can estimate $\mathbf{\Sigma}_{GMM}^{*}$ by $\hat{\mathbf{\Sigma}}_{GMM}^{*} = \left(\hat{\mathbf{D}}^{*\prime}\hat{\mathbf{\Sigma}}_{g}^{*-1}\hat{\mathbf{D}}^{*}\right)^{-1}$, where $\hat{\mathbf{D}}^{*}$ and $\hat{\mathbf{\Sigma}}_{g}^{*}$ are computed using $\hat{\mathbf{e}}^{*} = \mathbf{M}^b\left(\mathbf{Y}-\mathbf{L}\hat{\boldsymbol{\delta}}_{GMM}^{*}\right)$.

Observe that the 2SLS estimator fits in the GMM framework by setting the weights as

$$\mathbf{A}_{NT}^{w\prime}\mathbf{A}_{NT}^{w} = \begin{bmatrix} \mathbf{0}_{r\times r} & \mathbf{0}_{r\times q} \\ \mathbf{0}_{q\times r} & (\mathbf{Q}'\mathbf{M}^b\mathbf{Q})^{-1} \end{bmatrix}, \tag{66}$$

which is clearly not the optimal GMM weighting matrix. Therefore, $\hat{\boldsymbol{\delta}}_{2sls}$ is asymptotically less efficient than $\hat{\boldsymbol{\delta}}_{GMM}$. We will compare their small sample performance through Monte Carlo simulations.

Turning to the choice of $\mathbf{P}_l$ for the quadratic moments, note that the precision matrix of the efficient GMM estimator is given by

$$\mathbf{D}'\mathbf{\Sigma}_g^{-1}\mathbf{D} = \frac{1}{NT}\begin{bmatrix} \mathbf{D}_p'\mathbf{\Sigma}_{g,p}^{-1}\mathbf{D}_p & \mathbf{0}_{1\times k} \\ \mathbf{0}_{k\times 1} & \mathbf{0}_{k\times k} \end{bmatrix} + \frac{1}{NT}\left(\mathbf{Q}'\mathbf{M}_f^b\mathbf{L}_0\right)'\mathbf{\Omega}_{QMe}\left(\mathbf{Q}'\mathbf{M}_f^b\mathbf{L}_0\right). \tag{67}$$

where $\mathbf{D}_p = \left[T\sum_{i=1}^{N}\tilde{g}_{ii,1}^{s}\sigma_i^2, ..., T\sum_{i=1}^{N}\tilde{g}_{ii,r}^{s}\sigma_i^2\right]'$ and

$$\mathbf{\Sigma}_{g,p}^{-1} = \begin{bmatrix} tr\left[(\mathbf{P}_1\odot\mathbf{P}_1^s)\mathbf{\Sigma}_{ee}\right] & \cdots & tr\left[(\mathbf{P}_1\odot\mathbf{P}_r^s)\mathbf{\Sigma}_{ee}\right] \\ \vdots & & \vdots \\ tr\left[(\mathbf{P}_r\odot\mathbf{P}_1^s)\mathbf{\Sigma}_{ee}\right] & \cdots & tr\left[(\mathbf{P}_r\odot\mathbf{P}_r^s)\mathbf{\Sigma}_{ee}\right] \end{bmatrix}. \tag{68}$$

Ideally, we should choose $\mathbf{P}_l$ $(l = 1, ..., r)$ to maximize $\mathbf{D}_p'\mathbf{\Sigma}_{g,p}^{-1}\mathbf{D}_p$, which, however, depends on the unknown variance structure of the disturbances. If the disturbances are i.i.d., it is known in the spatial literature that the best $\mathbf{P}_l$ within the class of matrices of which the diagonal elements are all zeros is given by $\mathbf{P}^* = [\mathbf{G}_0 - diag(\mathbf{G}_0)]$ (Lee, 2007, Lee and Yu, 2014). Following the same arguments, the results can be readily extended to our model with common factors. To put it more clearly, provided that the disturbances are i.i.d., a best GMM (BGMM) estimator is available and it can be obtained by minimizing the optimally weighted moments (55), where $\mathbf{P}^b$ and $\mathbf{Q}$ are set to be

$\hat{\mathbf{P}}^* = [\mathbf{G}(\hat{\rho}) - diag(\mathbf{G}(\hat{\rho}))]$ and $\hat{\mathbf{Q}}^*$ given by (48) respectively. Nonetheless, in the presence of unknown heteroskedasticity and serial correlations, the BGMM estimator may not be the most efficient.

8 Monte Carlo Experiments

8.1 Monte Carlo Design

8.1.1 Estimation Experiments

Following closely the design of Pesaran (2006), the data generating process (DGP) is specified as follows.

$$
\begin{aligned}
y_{it} &= \rho \sum_{j=1}^{N} w_{ij} y_{jt} + \beta_1 x_{it1} + \beta_2 x_{it2} + \gamma_{y,i}' f_t + e_{it}, \\
x_{itp} &= \gamma_{x,ip}' f_t + \upsilon_{itp}, \quad p = 1, 2.
\end{aligned} \tag{69}
$$

The factors are generated by

$$
\begin{aligned}
f_{lt} &= \rho_{fl} f_{l,t-1} + \varsigma_{f_{lt}}, \ l = 1, ..., m, \ t = -49, ..., 0, 1, ..., T, \\
\varsigma_{f_{lt}} &\sim IIDN(0, 1 - \rho_{fl}^2), \quad \rho_{fl} = 0.5, \quad f_{l,-50} = 0,
\end{aligned}
$$

and the first 50 observations are discarded. The factor loadings are assumed to be $\gamma_{y,i1} \sim IIDN(1, 0.2)$, $\gamma_{y,i2} \sim IIDN(1, 0.2)$, and

$$
\begin{pmatrix} \gamma_{x,i11} & \gamma_{x,i12} \\ \gamma_{x,i21} & \gamma_{x,i22} \end{pmatrix} \sim IID \begin{pmatrix} N(0.5, 0.5) & N(0, 0.5) \\ N(0, 0.5) & N(0.5, 0.5) \end{pmatrix}. \tag{70}
$$

The idiosyncratic errors $(\upsilon_{it1}, \upsilon_{it2})'$ are generated as

$$
\begin{aligned}
\upsilon_{it,p} &= \rho_{\upsilon_{ip}} \upsilon_{it-1,p} + \vartheta_{it,p}, \quad t = -49, ..., 0, 1, ..., T, \\
\vartheta_{it,p} &\sim N(0, 1 - \rho_{\vartheta_{ip}}^2), \quad \upsilon_{ip,-50} = 0, \\
\rho_{\vartheta_{ip}} &\sim IIDU(0.05, 0.95), \quad p = 1, 2,
\end{aligned}
$$

and the first 50 observations are discarded.

We consider three different designs for the the idiosyncratic errors e_{it}:

- In the baseline case, e_{it} are generated from $IIDN(0, 1)$. The main goal of this simplest setup is to compare the efficiency properties of the competing estimators and examine if the theoretical conjecture on the efficiency relations can be supported.

- e_{it} are independent over time and heteroskedastic. Specifically,

$$\begin{aligned} e_{it} &= \sigma_i \zeta_{it},\ i = 1, ..., N,\ t = 1, ..., T, \\ \zeta_{it} &\sim IIDN(0,1),\ \sigma_i^2 \sim IIDU(0.5, 1.5). \end{aligned}$$

- e_{it} are serially correlated and heteroskedastic. In particular, they are specified as AR(1) processes for the first half of individual units, and as MA(1) processes for the remaining half.

$$\begin{aligned} e_{it} &= \rho_{ie} e_{i,t-1} + \sigma_i (1 - \rho_{ie}^2)^{1/2} \zeta_{it},\ i = 1, ..., \lfloor N/2 \rfloor, \\ e_{it} &= \sigma_i (1 + \theta_{ie}^2)^{-1/2} (\zeta_{it} + \theta_{ie} \zeta_{i,t-1}),\ i = \lfloor N/2 \rfloor + 1, ..., N, \\ \zeta_{it} &\sim IIDN(0,1),\ \sigma_i^2 \sim IIDU(0.5, 1.5), \\ \rho_{ie} &\sim IIDU(0.05, 0.95),\ e_{i,-50} = 0. \end{aligned}$$

To give the spatial weights matrix some practical meaning, we adopt the 1-ahead-and-1-behind circular neighbors spatial weights. An n-ahead-and-n-behind spatial matrix is generated by arranging all units in a circle and assigning equal weights to the n neighbors immediately ahead and behind a particular unit. For example, for the 2-ahead-and-2-behind spatial matrix, the ith row has non-zero elements in the positions $i-2, i-1, i+1, i+2$, and each weigh 1/4 due to row-normalization. Other specifications of the spatial weights matrices, including the 2, 3, 4- ahead-and-behind circular neighbors and unequally weighted matrices, have also been employed for robustness checks.[9]

In all experiments, the number of factors is set to $m = 2$ and the slope parameters are set to $\beta_1 = 1$ and $\beta_2 = 2$. We consider ρ =0.4 and 0.8, which represent low and high intensity of spatial dependence respectively, and combinations N, T =20,30,50,100,500,1000. Each experiment is repeated 2000 times.

8.1.2 Identification Experiments

For the identification experiments, it is sufficient to consider a simple SAR model without exogenous variables:

$$y_{it} = \rho \sum_{j=1}^{N} w_{ij} y_{jt} + e_{it}, \tag{71}$$

where $e_{it} \sim IIDN(0,1)$. To confirm that ρ is not identified when condition (28) is violated, we construct the weights matrix $\mathbf{W}$ as follows:

The first $N_1 = \lfloor N^\alpha \rfloor$ rows of $\mathbf{W}$ are nonzero, where $\alpha \in [0,1]$ and $\lfloor N^\alpha \rfloor$ is the integer part of N^α,

[9] These results are not presented to save space, but are available upon request. We find the proposed estimators exhibit robust performance across various spatial weights matrices.

and the rest $N_2 = N - N_1$ rows of $\mathbf{W}$ are all zeros. Notice that (28) fails to hold if $\alpha < 1$ because

$$\begin{aligned} \frac{tr(\mathbf{W}'\mathbf{W})}{N} &= \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}^2}{N} \\ &= \frac{\sum_{i=1}^{\lfloor N^\alpha \rfloor} K_i + \sum_{i=\lfloor N^\alpha \rfloor+1}^{N} 0}{N} \\ &= K_i \frac{\lfloor N^\alpha \rfloor}{N} = K_i N^{\alpha-1}, \end{aligned}$$

where in the second line we have used $\sum_{j=1}^{N} w_{ij}^2 = K_i < \infty$ for all i. Clearly, $tr(\mathbf{W}'\mathbf{W})/N$ goes to zero faster as $N \to \infty$ for smaller α.

Notice that the identification condition proposed by Lee and Yu (2015) is satisfied in this example. This can be verified by rewriting the condition that $\mathbf{I}_N$, $\mathbf{W}+\mathbf{W}'$ and $\mathbf{W}'\mathbf{W}$ being linearly independent as that there does not exist constants $c_1 > 0$ and $c_2 \neq 0$ such that

$$\begin{cases} \sum_{k=1}^{N} w_{ki}^2 = c_1 & \forall i \\ \sum_{k=1}^{N} w_{ki} w_{kj} = c_2(w_{ij} + w_{ji}) & \forall i, j \text{ and } i \neq j \\ w_{ii} = 0 & \forall i \end{cases} \tag{72}$$

Without loss of generality, we use the 5-ahead-and-5-behind circular spatial weights in the first N_1 rows, and set the rest N_2 rows equal to zeros in the experiments. We examine $\alpha = 1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}$, and combinations of $N = 20, 50, 100, 500, 1000$ and $T = 1, 20, 50, 100$. We can easily check (72) and confirm that no such c_1 and c_2 exist if $\alpha < 1$. In other words, the weight matrices satisfy the condition by Lee and Yu (2015) but violate our identification condition given in (28) under $\alpha < 1$.

Each experiment is repeated 2000 times and the true value of ρ is 0.2. The model is estimated by the maximum likelihood method (see, for example, **?**).

8.2 Monte Carlo Results

The model (69) is estimated by the following methods:

- The naive 2SLS estimator, which ignores the latent factors and apply a 2SLS estimation directly with IV matrix $(\mathbf{X}_{.t}, \mathbf{W}\mathbf{X}_{.t})$.
- The infeasible 2SLS estimator, which assumes the factors are known and uses IV matrix $(\mathbf{X}_{.t}, \mathbf{W}\mathbf{X}_{.t})$.
- The 2SLS estimator with IV matrix $(\mathbf{X}_{.t}, \mathbf{W}\mathbf{X}_{.t})$.
- The Best 2SLS (B2SLS) estimator with IV matrix $\hat{\mathbf{Q}}^* = \mathbf{M}^b \left[(\mathbf{I}_T \otimes \mathbf{G}(\hat{\rho}_{2sls}))\, \mathbf{X}\hat{\boldsymbol{\beta}}_{2sls}, \mathbf{X} \right]$.
- The efficient GMM estimator, which uses $\mathbf{P} = \mathbf{W}$ in the quadratic moments and $(\mathbf{X}_{.t}, \mathbf{W}\mathbf{X}_{.t})$ as IV in the linear moments, is obtained by a two-step procedure:
  - Step 1: Take the identity matrix as the weighting matrix, and compute a preliminary GMM estimate.

– Step 2: Use the inverse of the covariance of the moments as the weighting matrix and re-estimate the model.

- The maximum likelihood estimator (MLE) developed by Bai and Li (2014), which assumes that the disturbances of the model are independently distributed with heteroskedastic variances and explicitly estimate all the heteroskedasticity and factor loadings. It is important to note that the asymptotic distribution of the MLE was derived under the assumption that $N, T \to \infty$ and $\sqrt{N}/T \to 0$. The incidental parameters in the time dimension are avoided by estimating the sample variance of the factors rather than estimating individual factors.[10] With regards to computation, we follow the EM algorithm suggested by Bai and Li (2014). For simplicity, the number of factors is assumed known in the experiments.[11] The size and power properties of the MLE are not reported in their paper.

For the robust variance estimation of the above methods (excluding MLE), the Bartlett window size is chosen to be $\lfloor 2\sqrt{T} \rfloor$.[12]

Table 1 - 8 collect the results of the estimation experiments. Each table reports the estimates of bias, root mean squared error (RMSE), size and power for various estimators. We omit the results of β_2 to save space as they are similar to those of β_1. The results of the naive estimator are only presented in the first two tables since ignoring the factors produces disastrous biases and variances in all experiments, as expected.

We first observe that the 2SLS estimator exhibits very small biases and declining RMSEs as N and/or T increase. A comparison between the 2SLS and the infeasible 2SLS estimators suggests that the efficiency loss from using cross-sectional averages to approach the unobserved factors is quite small, and almost indiscernible when the sample size is large. The B2SLS estimator is only marginally more efficient than the 2SLS estimator for the spatial parameter ρ when N is small, and provides little or no improvement for the slope parameter $\boldsymbol{\beta}$. This implies that the IV matrix $(\mathbf{X}_{.t}, \mathbf{W}\mathbf{X}_{.t})$ approximates the best IV quite well for our experimental designs. The GMM estimator for ρ outperforms the 2SLS and B2SLS estimator in reducing the RMSEs, and even beats the infeasible 2SLS estimator for modest to large sample size ($N \geq 100$). Finally, the MLE by Bai and Li (2014) produces the smallest RMSEs among all estimation methods, and the improvement for ρ is especially notable. Nonetheless, its computation for large values of N and T is rather strenuous, and its performance could deteriorate if the number of factors is estimated.

Turning to the size and power properties, as anticipated by the theory, our proposed estimators have good power and empirical sizes that are close to the 5% nominal size for large N and small to modest T, irrespective of whether the errors are heteroskedastic and serially correlated. In cases of

[10] Bai and Li (2014) argue that if T is much smaller than N, one could estimate individual factors and the sample variance of factor loadings instead by switching the role of N and T. However, as they did not provide the asymptotic analysis for this scenario, we compute the MLE in the same way for all combinations of N and T.

[11] Bai and Li (2014) propose using an information criteria to estimate the number of factors and show in their Monte Carlo experiments that the estimates are near 100% correct.

[12] We also considered $\lfloor T^{1/3} \rfloor$ as the window size and found only marginal differences in the size and power properties. (Results are available upon request.)

small N and large T, the rejection frequencies under the null of the 2SLS and B2SLS estimators could reach as high as 30%, and the GMM estimators are even more severally over-sized especially for the spatial parameter ρ. In view of this, it is worthwhile to bear in mind that our variance estimators cannot be applied to the small N large T scenarios. In contrast, the MLE performs well when the errors are independent: it has higher power than the other estimators and proper sizes close to the 5% nominal level when N is not too large relative to T. However, as its theory does not permit the presence of serial correlation in the errors, the tests are significantly over-sized.

In sum, our proposed estimators exhibit robust performance to unknown heteroskedasticity and serial correlation in the errors. Furthermore, inspection of Table 7 and 8 reveals that they are also robust to different intensity of spatial dependence.

Finally, the last table reports the results of the identification experiments. As expected, the RMSEs are enormous when $\alpha < 1$, and the smaller α is, the greater the RMSEs become. This further corroborates our finding that $tr(\mathbf{W}'\mathbf{W})/N > \varepsilon > 0$ for all N including $N \to \infty$ is indeed necessary for identification.

9 Concluding Remarks

This paper is concerned with panel data models in the joint presence of two sources of cross-sectional dependence: spatial autoregressive effects and common shocks. It derives the identification conditions conditional on the factors and proposes several estimation methods via instrumental variables and moment conditions based on the variance-covariance properties of the disturbances. By replacing the unobserved factors with cross-sectional averages, these estimators are shown to be consistent and free of nuisance parameters asymptotically provided that T is of a smaller order of magnitude than N as $N, T \to \infty$ jointly. Importantly, unlike MLE, the number of latent factors need not be estimated and general forms of serial correlation in the disturbances are allowed for. A wide range of Monte Carlo exercises lend further support to our theoretical results.

We foresee many potential applications in different areas, especially in the fields of macroeconomics and finance, where the estimation results may not be reliable without accounting for the unobserved factors. In future work, we could extend and improve the current framework by incorporating spatial dependence in the error terms, and accommodating dynamics and more heterogeneity in the model. Furthermore, it may be worthwhile to relax the assumption that the number of instruments is fixed and the same across individuals.

References

Aquaro, M., Bailey, N., Pesaran, M. H., 2015. Quasi maximum likelihood estimation of spatial models with heterogeneous coefficients, working Paper.

Bai, J., Li, K., 2014. Spatial panel data models with common shocks, working Paper.

Bailey, N., Holly, S., Pesaran, M. H., 2014. A two stage approach to spatiotemporal analysis with strong and weak cross-sectional dependence, cESifo Working Paper.

Bailey, N., Kapetanios, G., Pesaran, H., 2013. Exponent of cross-sectional dependence: Estimation and inference, discussion Paper series, Forschungsinstitut zur Zukunft der Arbeit.

Chudik, A., Pesaran, M. H., Tosetti, E., 2011. Weak and strong cross-section dependence and estimation of large panels. The Econometrics Journal 14 (1), C45–C90.

Kapetanios, G., Pesaran, M. H., Yamagata, T., 2011. Panels with non-stationary multifactor error structures. Journal of Econometrics 160 (2), 326–348.

Lee, L.-f., 2003. Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances. Econometric Reviews 22 (4), 307–335.

Lee, L.-f., 2007. Gmm and 2sls estimation of mixed regressive, spatial autoregressive models. Journal of Econometrics 137 (2), 489–514.

Lee, L.-f., Yu, J., 2014. Efficient gmm estimation of spatial dynamic panel data models with fixed effects. Journal of Econometrics 180 (2), 174–197.

Lee, L.-f., Yu, J., 2015. Identification of spatial durbin panel models. Journal of Applied Econometrics.

Lin, X., Lee, L.-f., 2010. Gmm estimation of spatial autoregressive models with unknown heteroskedasticity. Journal of Econometrics 157 (1), 34–52.

Liu, X., Lee, L.-f., Bollinger, C. R., 2010. An efficient gmm estimator of spatial autoregressive models. Journal of Econometrics 159 (2), 303–319.

Pesaran, M. H., 2006. Estimation and inference in large heterogeneous panels with a multifactor error structure. Econometrica 74 (4), 967–1012.

Pesaran, M. H., Tosetti, E., 2011. Large panels with common factors and spatial correlation. Journal of Econometrics 161 (2), 182–202.

Rothenberg, T. J., 1971. Identification in parametric models. Econometrica: Journal of the Econometric Society, 577–591.

Yu, J., de Jong, R., 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 (1), 118–134.

Table 1: Small sample properties of estimators for spatial parameter ρ ($\rho = 0.4$, i.i.d. errors)

	Bias(×100)						RMSE(×100)						Size(×100)						Power(×100)					
N\T	20	30	50	100	500	1000	20	30	50	100	500	1000	20	30	50	100	500	1000	20	30	50	100	500	1000
	Naive 2SLS estimator (excluding factors)																							
20	13.88	14.01	14.06	14.14	14.26	14.27	14.60	14.63	14.53	14.55	14.59	14.61	94.25	97.65	98.80	99.35	99.65	99.85	97.10	98.45	99.35	99.55	99.95	99.95
30	13.81	13.94	14.07	14.14	14.24	14.27	14.36	14.37	14.41	14.41	14.45	14.48	98.10	99.15	99.75	99.90	100.00	100.00	99.30	99.75	99.95	100.00	100.00	100.00
50	13.84	14.01	14.11	14.15	14.25	14.27	14.25	14.32	14.34	14.33	14.38	14.40	99.75	99.90	100.00	100.00	100.00	100.00	99.85	100.00	100.00	100.00	100.00	100.00
100	13.85	14.06	14.12	14.16	14.28	14.30	14.18	14.28	14.28	14.27	14.35	14.36	99.95	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
500	13.81	13.97	14.09	14.13	14.26	14.29	14.06	14.14	14.19	14.19	14.28	14.30	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
1000	13.82	13.98	14.08	14.14	14.26	14.29	14.06	14.14	14.18	14.19	14.28	14.30	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	Infeasible 2SLS estimator (including factors)																							
20	-0.16	-0.10	-0.11	-0.06	0.00	-0.01	3.12	2.44	1.85	1.26	0.56	0.40	5.20	4.85	5.65	4.65	5.10	5.40	9.40	13.05	17.45	32.90	94.70	99.90
30	-0.11	-0.06	-0.01	0.00	0.01	0.00	2.45	1.96	1.50	1.01	0.45	0.32	4.00	4.35	5.00	4.10	4.80	4.70	10.30	16.00	27.00	48.90	99.65	100.00
50	0.01	0.03	-0.01	-0.01	0.01	0.00	1.89	1.51	1.14	0.79	0.35	0.25	4.85	5.00	4.85	4.35	5.00	4.80	16.90	25.25	39.65	69.90	100.00	100.00
100	0.00	0.01	0.01	0.01	0.00	0.00	1.33	1.05	0.80	0.56	0.25	0.18	3.45	3.70	4.10	4.50	5.20	5.90	29.10	44.70	67.40	93.80	100.00	100.00
500	-0.02	-0.01	0.00	0.00	0.00	0.00	0.61	0.48	0.36	0.25	0.11	0.08	4.35	4.30	4.40	4.40	5.25	5.60	89.10	98.25	100.00	100.00	100.00	100.00
1000	0.00	0.00	0.00	0.00	0.00	0.00	0.42	0.34	0.26	0.18	0.08	0.06	3.80	4.90	4.50	5.10	5.05	5.05	99.55	100.00	100.00	100.00	100.00	100.00
	2SLS estimator																							
20	-0.20	-0.10	-0.12	-0.09	-0.01	-0.02	3.56	2.81	2.21	1.55	0.97	0.88	6.80	7.60	9.40	9.20	21.70	30.05	10.40	14.20	19.60	32.90	82.90	93.40
30	-0.09	-0.02	-0.01	0.00	0.00	-0.01	2.81	2.21	1.66	1.17	0.66	0.57	6.45	6.45	7.50	7.55	14.10	21.95	12.30	18.40	27.00	45.80	95.85	99.00
50	0.00	0.03	0.00	0.00	0.01	0.00	2.03	1.61	1.21	0.85	0.42	0.33	5.60	5.95	5.50	5.30	9.60	13.40	16.70	25.00	39.85	68.60	99.80	99.95
100	0.00	0.01	0.01	0.01	0.00	0.00	1.41	1.11	0.83	0.58	0.26	0.20	4.60	4.70	4.85	5.45	6.40	7.40	28.25	43.35	67.35	93.10	100.00	100.00
500	-0.02	-0.01	0.00	0.00	0.00	0.00	0.63	0.49	0.37	0.26	0.11	0.08	5.05	4.85	4.90	4.90	5.95	6.00	87.35	97.90	100.00	100.00	100.00	100.00
1000	0.00	0.00	0.00	0.00	0.00	0.00	0.45	0.35	0.26	0.18	0.08	0.06	4.35	4.75	4.70	5.25	5.00	5.00	99.25	99.95	100.00	100.00	100.00	100.00
	B2SLS estimator																							
20	-0.28	-0.14	-0.15	-0.10	-0.02	-0.02	3.48	2.73	2.16	1.52	0.95	0.85	6.85	7.55	9.95	9.80	21.50	29.80	10.50	14.05	18.95	35.15	83.75	93.25
30	-0.13	-0.03	-0.01	0.00	0.00	-0.01	2.74	2.16	1.63	1.15	0.65	0.55	6.10	6.45	7.15	8.00	15.30	22.50	12.40	18.40	27.95	48.05	95.90	98.90
50	-0.01	0.03	0.00	0.00	0.01	0.00	1.99	1.57	1.19	0.83	0.41	0.33	5.30	6.10	5.25	5.45	9.25	12.60	18.10	25.95	41.30	70.95	99.80	99.95
100	-0.01	-0.01	0.01	0.01	0.00	0.00	1.38	1.08	0.80	0.56	0.26	0.19	4.75	5.05	4.45	5.20	5.60	7.25	29.70	45.05	69.20	94.15	100.00	100.00
500	-0.02	-0.01	0.00	0.00	0.00	0.00	0.62	0.48	0.36	0.25	0.11	0.08	4.95	4.55	4.85	4.60	5.70	5.50	88.65	98.05	100.00	100.00	100.00	100.00
1000	0.00	0.00	-0.01	0.00	0.00	0.00	0.44	0.34	0.25	0.18	0.08	0.06	4.50	4.65	5.05	5.60	5.25	5.20	99.45	99.95	100.00	100.00	100.00	100.00
	GMM estimator																							
20	-1.90	-1.74	-1.62	-1.50	-1.38	-1.37	3.50	2.95	2.44	1.99	1.59	1.54	12.65	15.05	20.20	30.20	71.75	87.45	7.20	8.70	9.90	11.50	35.60	51.05
30	-1.16	-1.02	-0.98	-0.93	-0.88	-0.89	2.60	2.10	1.72	1.36	1.04	1.00	9.15	9.85	14.25	20.85	60.90	79.95	8.20	10.80	14.90	25.40	75.80	91.70
50	-0.65	-0.59	-0.58	-0.54	-0.51	-0.52	1.89	1.53	1.21	0.90	0.63	0.59	7.80	9.25	11.05	13.35	41.70	64.95	15.20	21.50	32.00	57.10	99.20	99.90
100	-0.29	-0.28	-0.27	-0.26	-0.26	-0.26	1.25	0.98	0.76	0.56	0.34	0.31	6.35	6.15	6.65	8.70	23.80	40.60	32.20	46.45	69.45	94.90	100.00	100.00
500	-0.07	-0.06	-0.06	-0.06	-0.05	-0.05	0.54	0.41	0.32	0.22	0.11	0.08	5.90	5.00	5.70	6.70	8.70	11.80	95.15	99.60	100.00	100.00	100.00	100.00
1000	-0.03	-0.03	-0.03	-0.03	-0.02	-0.02	0.37	0.30	0.22	0.16	0.07	0.05	4.60	5.20	5.20	6.20	6.20	8.30	100.00	100.00	100.00	100.00	100.00	100.00
	MLE																							
20	0.33	0.21	0.14	0.14	0.19	0.18	3.06	2.31	1.69	1.15	0.54	0.40	15.05	11.65	9.55	8.55	9.45	11.45	27.15	29.15	34.55	54.85	99.55	100.00
30	0.30	0.23	0.18	0.16	0.15	0.14	2.32	1.79	1.36	0.92	0.43	0.31	11.80	10.20	8.65	7.85	8.85	9.70	30.65	34.95	46.80	71.20	100.00	100.00
50	0.35	0.23	0.14	0.11	0.11	0.10	1.79	1.39	1.02	0.70	0.32	0.24	12.45	10.00	8.30	7.30	7.85	9.30	41.45	49.05	64.00	88.70	100.00	100.00
100	0.29	0.17	0.11	0.09	0.07	0.07	1.26	0.95	0.70	0.49	0.22	0.17	11.55	9.25	7.00	6.95	7.00	8.65	59.75	71.85	89.05	99.25	100.00	100.00
500	0.22	0.11	0.05	0.04	0.03	0.04	0.59	0.43	0.31	0.22	0.10	0.08	13.40	9.30	7.20	7.70	6.50	8.70	99.00	100.00	100.00	100.00	100.00	100.00
1000	0.20	0.11	0.06	0.04	0.02	0.02	0.42	0.32	0.23	0.16	0.07	0.05	14.40	11.10	8.40	7.00	5.60	5.40	100.00	100.00	100.00	100.00	100.00	100.00

Notes: True parameter values are $\rho = 0.4$, $\beta_1 = 1$ and $\beta_2 = 2$, and the number of factors is $m = 2$. The spatial weights matrix is the 1-ahead-and-1-behind circular neighbors matrix. The naive estimator ignores the latent factors, and the infeasible estimator assumes the factors are known. The naive 2SLS, infeasible 2SLS, and 2SLS use IV matrix $(\mathbf{X}_{.t}, \mathbf{W}\mathbf{X}_{.t})$, and the B2SLS estimator is obtained with IV matrix $\hat{\mathbf{Q}}^* = \mathbf{M}^b\left[(\mathbf{I}_T \otimes \mathbf{G}(\hat{\rho}_{2sls}))\,\mathbf{X}\hat{\boldsymbol{\beta}}_{2sls}, \mathbf{X}\right]$. The efficient two-step GMM estimator uses $\mathbf{P}_1 = \mathbf{W}$ in the quadratic moments and $(\mathbf{X}_{.t}, \mathbf{W}\mathbf{X}_{.t})$ as IV in the linear moments. The MLE is attributed to Bai and Li (2014) and the number of factors is assumed to be known. The number of replication is 2000. The 95% confidence interval for size 5% is $[3.6\%, 6.4\%]$, and the power is computed under $H_1 : \rho = 0.38$.

Table 2: Small sample properties of estimators for slope parameter β_1 ($\rho = 0.4$, i.i.d. errors)

	Bias(×100)						RMSE(×100)						Size(×100)						Power(×100)					
N\T	20	30	50	100	500	1000	20	30	50	100	500	1000	20	30	50	100	500	1000	20	30	50	100	500	1000
	Naive 2SLS estimator (excluding factors)																							
20	10.36	10.57	10.66	10.79	10.86	10.89	14.15	13.74	13.46	13.26	13.04	13.03	46.30	56.75	68.50	79.95	92.35	94.95	65.70	76.60	86.65	92.50	97.60	98.75
30	10.52	10.79	10.80	10.94	11.08	11.10	13.28	13.12	12.77	12.66	12.55	12.54	57.00	67.70	78.20	87.85	95.35	96.95	78.25	86.45	92.70	96.90	99.30	99.35
50	10.45	10.55	10.72	10.94	11.12	11.15	12.53	12.21	12.04	12.01	12.00	12.02	68.90	78.90	87.85	94.45	98.60	98.85	88.60	94.30	97.70	99.50	99.95	99.85
100	10.67	10.77	10.87	11.09	11.23	11.27	12.06	11.90	11.73	11.77	11.73	11.76	82.90	90.75	96.00	99.00	99.90	99.90	97.15	99.15	99.75	99.95	100.00	100.00
500	10.78	10.91	10.99	11.19	11.32	11.34	11.62	11.53	11.40	11.44	11.44	11.45	97.70	99.05	99.95	100.00	100.00	100.00	99.90	100.00	100.00	100.00	100.00	100.00
1000	10.80	10.94	11.01	11.22	11.33	11.36	11.57	11.48	11.36	11.42	11.41	11.42	98.80	99.85	99.95	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	Infeasible 2SLS estimator (including factors)																							
20	-0.02	0.03	0.08	0.08	-0.02	-0.01	5.63	4.26	3.33	2.37	1.03	0.74	4.95	3.90	5.05	5.95	5.30	5.55	14.25	19.80	33.15	59.15	99.85	100.00
30	0.05	0.02	0.00	-0.01	-0.02	-0.03	4.50	3.57	2.73	1.88	0.84	0.58	4.45	4.40	5.00	4.95	4.90	4.70	19.20	28.15	46.25	75.85	100.00	100.00
50	-0.19	-0.17	-0.11	-0.09	-0.02	-0.02	3.45	2.66	2.01	1.40	0.63	0.45	4.20	3.90	4.30	4.15	4.60	4.75	25.25	39.85	64.60	92.40	100.00	100.00
100	-0.13	-0.04	-0.05	-0.05	-0.02	0.00	2.46	1.92	1.48	1.02	0.44	0.31	4.70	5.00	4.65	4.65	5.05	4.20	48.25	71.75	91.40	99.85	100.00	100.00
500	-0.06	-0.04	-0.02	-0.01	-0.01	0.00	1.07	0.84	0.66	0.47	0.20	0.14	3.90	3.85	4.80	5.35	4.40	4.55	99.70	100.00	100.00	100.00	100.00	100.00
1000	0.01	0.01	0.01	0.00	0.00	0.00	0.79	0.63	0.47	0.33	0.14	0.10	4.95	4.95	4.65	5.15	5.70	5.80	100.00	100.00	100.00	100.00	100.00	100.00
	2SLS estimator																							
20	0.08	0.08	0.04	0.05	-0.02	-0.01	6.04	4.63	3.63	2.71	1.67	1.49	6.80	6.25	8.35	10.20	22.05	30.85	17.00	23.10	34.75	60.20	96.35	98.55
30	0.06	0.04	0.04	0.03	0.00	-0.01	4.73	3.77	2.91	2.05	1.11	0.94	5.80	6.30	7.60	7.40	14.55	21.35	20.15	29.30	47.25	75.70	99.80	99.90
50	-0.18	-0.18	-0.09	-0.09	-0.02	-0.02	3.61	2.77	2.08	1.45	0.73	0.59	4.70	5.05	4.80	5.00	8.60	13.40	26.25	39.70	65.70	92.70	100.00	100.00
100	-0.13	-0.05	-0.06	-0.05	-0.02	0.00	2.54	1.96	1.51	1.03	0.47	0.35	5.45	5.10	5.00	5.10	6.35	7.15	47.00	70.50	90.65	99.85	100.00	100.00
500	-0.07	-0.05	-0.02	-0.01	-0.01	0.00	1.12	0.86	0.67	0.48	0.20	0.14	4.15	4.35	5.00	5.20	4.40	4.65	99.20	100.00	100.00	100.00	100.00	100.00
1000	0.02	0.01	0.01	0.00	0.00	0.00	0.82	0.65	0.48	0.33	0.14	0.10	5.60	5.40	5.35	5.20	6.00	5.85	100.00	100.00	100.00	100.00	100.00	100.00
	B2SLS estimator																							
20	0.09	0.09	0.05	0.05	-0.02	-0.01	6.04	4.63	3.63	2.71	1.67	1.49	6.65	6.25	8.30	10.15	22.15	31.05	17.10	23.25	35.20	60.30	96.35	98.50
30	0.07	0.04	0.04	0.03	0.00	-0.01	4.73	3.77	2.91	2.05	1.11	0.94	5.75	6.30	7.45	7.35	14.60	21.30	20.30	29.35	47.30	75.45	99.80	99.95
50	-0.18	-0.18	-0.09	-0.09	-0.02	-0.02	3.61	2.76	2.08	1.45	0.73	0.59	4.70	5.10	4.70	4.85	8.70	13.15	26.30	39.80	65.80	92.70	100.00	100.00
100	-0.13	-0.05	-0.06	-0.04	-0.02	0.00	2.54	1.96	1.51	1.03	0.47	0.34	5.35	5.10	4.75	5.10	6.55	7.10	47.00	70.85	90.75	99.85	100.00	100.00
500	-0.07	-0.05	-0.02	-0.01	-0.01	0.00	1.12	0.86	0.67	0.48	0.20	0.14	4.20	4.25	5.00	5.30	4.50	4.95	99.25	100.00	100.00	100.00	100.00	100.00
1000	0.02	0.01	0.01	0.00	0.00	0.00	0.82	0.64	0.48	0.33	0.14	0.10	5.45	5.30	5.25	5.25	6.05	5.75	100.00	100.00	100.00	100.00	100.00	100.00
	GMM estimator																							
20	0.14	0.20	0.18	0.20	0.13	0.14	6.03	4.64	3.64	2.72	1.68	1.50	6.20	5.90	7.85	10.05	22.20	32.55	16.45	23.05	35.55	62.00	96.70	98.70
30	0.15	0.16	0.17	0.16	0.13	0.12	4.75	3.79	2.92	2.05	1.12	0.95	5.15	6.50	6.85	7.10	14.65	21.95	19.45	29.40	48.30	77.35	99.85	99.95
50	-0.10	-0.08	0.00	0.00	0.08	0.08	3.62	2.77	2.07	1.45	0.74	0.59	4.05	4.45	4.45	4.80	8.45	13.90	25.70	40.00	66.55	93.35	100.00	100.00
100	-0.09	0.00	-0.01	0.01	0.03	0.05	2.54	1.96	1.50	1.03	0.47	0.35	4.35	4.65	4.65	4.55	6.35	7.20	45.25	70.35	90.65	99.85	100.00	100.00
500	-0.06	-0.04	-0.01	0.00	0.01	0.01	1.11	0.86	0.67	0.47	0.20	0.14	3.80	3.75	4.85	5.15	4.35	4.90	99.05	100.00	100.00	100.00	100.00	100.00
1000	0.02	0.02	0.02	0.01	0.00	0.00	0.81	0.64	0.48	0.33	0.14	0.10	4.55	4.75	5.20	5.10	5.90	6.20	100.00	100.00	100.00	100.00	100.00	100.00
	MLE																							
20	-0.06	-0.03	0.01	0.03	-0.06	-0.05	6.44	4.71	3.54	2.48	1.07	0.77	13.10	9.25	8.40	7.90	6.20	6.90	24.60	26.75	36.50	60.30	99.75	100.00
30	-0.01	-0.01	-0.06	-0.04	-0.04	-0.06	5.05	3.85	2.86	1.93	0.85	0.60	11.45	9.30	7.70	6.10	6.55	5.55	29.80	35.10	48.45	75.80	100.00	100.00
50	-0.20	-0.16	-0.13	-0.11	-0.05	-0.05	3.76	2.79	2.06	1.43	0.64	0.46	10.20	7.00	5.95	5.55	5.15	5.30	36.20	47.15	68.15	93.15	100.00	100.00
100	-0.14	-0.05	-0.07	-0.06	-0.04	-0.02	2.68	2.01	1.52	1.04	0.45	0.31	10.60	8.00	6.80	6.05	5.55	4.80	58.45	76.55	91.85	99.90	100.00	100.00
500	-0.02	-0.01	-0.01	-0.01	-0.02	-0.01	1.18	0.88	0.67	0.48	0.20	0.14	10.70	6.50	5.90	7.00	4.70	5.50	99.60	100.00	100.00	100.00	100.00	100.00
1000	0.04	0.00	0.02	-0.01	-0.02	-0.01	0.84	0.65	0.47	0.33	0.14	0.10	10.40	8.40	6.30	5.50	6.10	5.50	100.00	100.00	100.00	100.00	100.00	100.00

See notes to Table 1. The power is computed under $H_1 : \beta_1 = 0.95$.

Table 3: Small sample properties of estimators for spatial parameter ρ ($\rho = 0.4$, independent and heteroskedastic errors)

	Bias(×100)						RMSE(×100)						Size(×100)						Power(×100)					
N\T	20	30	50	100	500	1000	20	30	50	100	500	1000	20	30	50	100	500	1000	20	30	50	100	500	1000
	Infeasible 2SLS estimator (including factors)																							
20	-0.15	-0.10	-0.11	-0.06	0.00	0.00	3.13	2.45	1.85	1.26	0.56	0.40	5.10	4.80	5.65	4.80	4.70	5.30	9.70	12.60	17.65	33.15	94.70	99.90
30	-0.11	-0.06	-0.02	0.00	0.01	0.00	2.45	1.96	1.50	1.01	0.45	0.32	3.95	4.25	4.90	4.25	4.40	4.95	10.50	16.20	27.05	48.60	99.55	100.00
50	0.03	0.04	-0.01	0.00	0.01	0.00	1.88	1.50	1.14	0.79	0.35	0.24	4.45	4.55	4.65	4.60	5.35	4.75	17.25	26.10	39.70	69.55	100.00	100.00
100	0.01	0.01	0.01	0.01	0.00	0.00	1.32	1.05	0.80	0.57	0.25	0.18	3.45	3.80	3.95	4.25	5.20	5.35	29.65	45.20	67.80	94.00	100.00	100.00
500	-0.02	-0.01	0.00	0.00	0.00	0.00	0.61	0.47	0.36	0.25	0.11	0.08	4.40	4.20	4.50	4.65	5.70	5.70	89.35	98.40	100.00	100.00	100.00	100.00
1000	0.00	0.00	0.00	0.00	0.00	0.00	0.42	0.34	0.26	0.18	0.08	0.06	3.90	4.85	4.50	4.95	4.95	5.45	99.55	100.00	100.00	100.00	100.00	100.00
	2SLS estimator																							
20	-0.18	-0.10	-0.12	-0.09	-0.01	-0.01	3.57	2.81	2.21	1.55	0.97	0.88	7.00	7.30	9.35	9.45	21.15	30.45	10.80	14.05	19.95	32.90	83.05	93.00
30	-0.09	-0.02	-0.02	-0.01	0.00	-0.01	2.81	2.20	1.67	1.17	0.66	0.57	6.00	6.40	7.55	7.40	14.25	21.70	12.35	17.80	27.50	45.70	95.80	99.00
50	0.00	0.04	0.01	0.00	0.01	0.00	2.03	1.61	1.21	0.84	0.42	0.33	5.45	5.65	5.25	5.25	9.75	13.55	16.90	26.05	40.00	68.40	99.85	99.95
100	0.01	0.01	0.02	0.01	0.00	0.00	1.41	1.11	0.83	0.58	0.26	0.20	4.40	5.10	4.75	5.00	6.60	7.50	29.30	43.45	67.35	93.10	100.00	100.00
500	-0.02	-0.01	0.00	0.00	0.00	0.00	0.63	0.49	0.37	0.26	0.11	0.08	5.45	5.00	4.45	4.65	5.75	6.05	87.35	97.90	100.00	100.00	100.00	100.00
1000	0.00	0.00	0.00	0.00	0.00	0.00	0.45	0.35	0.26	0.18	0.08	0.06	4.50	4.45	4.60	5.05	5.00	5.20	99.45	99.95	100.00	100.00	100.00	100.00
	B2SLS estimator																							
20	-0.27	-0.14	-0.14	-0.09	-0.02	-0.02	3.49	2.73	2.15	1.52	0.94	0.85	6.85	7.50	9.00	9.10	20.75	29.50	10.60	14.40	19.50	34.70	83.80	93.30
30	-0.13	-0.04	-0.02	0.00	0.00	-0.01	2.74	2.15	1.63	1.15	0.65	0.55	6.10	6.10	7.05	7.75	14.75	22.25	12.55	17.70	28.10	47.90	95.95	99.05
50	0.00	0.04	0.01	0.00	0.01	0.00	1.98	1.57	1.19	0.82	0.41	0.33	5.10	5.80	5.30	5.30	9.75	13.05	17.80	26.50	41.70	70.35	99.85	99.95
100	0.00	0.00	0.01	0.01	0.00	0.00	1.38	1.08	0.80	0.57	0.26	0.19	4.75	5.50	4.50	5.15	5.90	7.45	30.65	45.20	68.95	94.30	100.00	100.00
500	-0.02	-0.01	0.00	0.00	0.00	0.00	0.62	0.48	0.36	0.25	0.11	0.08	4.65	4.60	4.70	4.35	5.90	5.60	88.55	98.30	100.00	100.00	100.00	100.00
1000	0.00	0.00	-0.01	0.00	0.00	0.00	0.44	0.34	0.25	0.18	0.08	0.06	4.35	4.40	5.05	5.60	5.35	5.05	99.55	99.95	100.00	100.00	100.00	100.00
	GMM estimator																							
20	-1.90	-1.74	-1.61	-1.50	-1.38	-1.37	3.50	2.94	2.43	1.99	1.59	1.55	12.25	15.00	19.70	30.85	72.90	87.45	7.20	8.50	9.30	11.55	34.80	49.40
30	-1.18	-1.03	-0.99	-0.94	-0.89	-0.89	2.59	2.10	1.74	1.37	1.04	1.00	9.10	10.35	14.75	21.40	61.65	80.35	8.25	10.60	14.65	24.60	74.55	91.60
50	-0.65	-0.59	-0.58	-0.54	-0.51	-0.52	1.88	1.52	1.21	0.90	0.62	0.59	8.60	9.00	10.90	13.20	41.45	65.20	15.25	20.75	31.70	57.10	99.30	99.90
100	-0.29	-0.28	-0.26	-0.25	-0.26	-0.26	1.25	0.98	0.75	0.56	0.34	0.31	6.45	6.15	6.30	8.70	23.30	40.70	32.55	46.20	69.05	94.55	100.00	100.00
500	-0.07	-0.06	-0.06	-0.06	-0.05	-0.05	0.53	0.41	0.32	0.22	0.11	0.08	4.95	4.55	5.25	6.00	8.10	12.25	95.65	99.65	100.00	100.00	100.00	100.00
1000	-0.03	-0.03	-0.03	-0.03	-0.02	-0.02	0.37	0.30	0.22	0.16	0.07	0.05	5.00	5.25	4.80	5.65	6.10	8.30	99.95	100.00	100.00	100.00	100.00	100.00
	MLE																							
20	0.31	0.20	0.14	0.15	0.19	0.18	2.97	2.23	1.62	1.11	0.52	0.39	15.45	12.30	10.30	8.25	9.75	11.95	28.25	30.05	36.80	57.50	99.60	100.00
30	0.29	0.23	0.18	0.15	0.15	0.14	2.24	1.72	1.31	0.89	0.42	0.30	12.05	9.95	9.30	7.65	8.75	10.15	32.00	36.30	49.15	74.65	100.00	100.00
50	0.32	0.21	0.14	0.11	0.11	0.11	1.72	1.34	0.98	0.67	0.31	0.23	12.35	9.90	8.40	7.30	8.05	10.05	43.60	51.70	66.55	90.55	100.00	100.00
100	0.26	0.16	0.11	0.09	0.07	0.07	1.21	0.91	0.68	0.47	0.21	0.16	11.75	9.35	7.25	6.85	7.25	9.60	61.80	73.95	91.15	99.70	100.00	100.00
500	0.20	0.10	0.05	0.04	0.04	0.04	0.56	0.41	0.30	0.21	0.10	0.07	13.00	9.20	7.40	7.90	7.00	9.00	99.50	100.00	100.00	100.00	100.00	100.00
1000	0.18	0.09	0.04	0.02	0.02	0.02	0.41	0.30	0.21	0.15	0.07	0.05	14.00	9.30	7.60	6.20	6.00	5.80	100.00	100.00	100.00	100.00	100.00	100.00

See notes to Table 1. The experimental design for this table only differs from that of Table 1 in the error generating process, which is now independent and heteroskedastic instead of i.i.d.

Table 4: Small sample properties of estimators for slope parameter β_1 ($\rho = 0.4$, independent and heteroskedastic errors)

	Bias(×100)						RMSE(×100)						Size(×100)						Power(×100)					
N\T	20	30	50	100	500	1000	20	30	50	100	500	1000	20	30	50	100	500	1000	20	30	50	100	500	1000
	Infeasible 2SLS estimator (including factors)																							
20	-0.02	0.03	0.07	0.07	-0.02	-0.01	5.61	4.26	3.34	2.36	1.03	0.74	4.75	3.80	5.45	6.15	5.00	5.45	14.05	19.50	33.80	58.65	99.75	100.00
30	0.01	-0.01	-0.02	-0.02	-0.02	-0.03	4.45	3.57	2.72	1.88	0.84	0.58	4.35	4.40	5.10	5.10	5.30	4.85	18.60	28.00	45.25	75.70	100.00	100.00
50	-0.19	-0.17	-0.11	-0.10	-0.02	-0.02	3.45	2.68	2.03	1.41	0.63	0.45	4.45	3.95	4.45	4.10	4.05	4.50	25.40	39.70	65.10	92.50	100.00	100.00
100	-0.13	-0.05	-0.05	-0.04	-0.02	0.00	2.45	1.93	1.48	1.02	0.44	0.31	4.60	5.10	4.85	4.65	5.15	4.10	47.85	71.55	91.25	99.80	100.00	100.00
500	-0.05	-0.03	-0.02	-0.01	-0.01	0.00	1.07	0.84	0.66	0.47	0.20	0.14	3.45	3.55	4.50	5.10	5.05	4.60	99.75	100.00	100.00	100.00	100.00	100.00
1000	0.01	0.01	0.02	0.00	0.00	0.00	0.79	0.63	0.47	0.33	0.14	0.10	5.20	5.15	4.60	4.40	5.50	5.05	100.00	100.00	100.00	100.00	100.00	100.00
	2SLS estimator																							
20	0.06	0.08	0.04	0.06	-0.01	0.00	6.03	4.64	3.65	2.72	1.68	1.50	6.90	6.15	8.40	9.95	21.55	31.00	16.85	23.50	34.75	60.80	96.20	98.45
30	0.03	0.02	0.03	0.02	0.00	-0.01	4.71	3.78	2.92	2.06	1.13	0.96	5.35	6.10	7.10	7.90	14.95	22.30	20.10	28.95	47.30	75.70	99.80	99.90
50	-0.19	-0.17	-0.09	-0.09	-0.01	-0.01	3.62	2.78	2.09	1.46	0.73	0.59	4.90	5.20	5.15	4.85	8.30	13.25	26.85	40.35	66.10	92.25	100.00	100.00
100	-0.13	-0.05	-0.05	-0.04	-0.02	0.00	2.53	1.97	1.52	1.04	0.47	0.35	5.45	5.35	5.40	5.10	6.25	7.45	46.80	70.15	90.20	99.80	100.00	100.00
500	-0.07	-0.05	-0.02	-0.01	-0.01	0.00	1.12	0.86	0.67	0.48	0.20	0.14	4.05	3.95	4.85	5.60	4.90	4.60	99.10	100.00	100.00	100.00	100.00	100.00
1000	0.02	0.01	0.01	0.00	0.00	0.00	0.82	0.65	0.48	0.33	0.14	0.10	5.50	5.60	4.95	4.95	5.55	5.20	100.00	100.00	100.00	100.00	100.00	100.00
	B2SLS estimator																							
20	0.08	0.09	0.05	0.06	-0.01	0.00	6.03	4.64	3.65	2.72	1.68	1.50	7.00	6.15	8.20	9.85	21.50	30.75	16.95	23.60	34.80	60.65	96.20	98.45
30	0.04	0.02	0.03	0.02	0.00	-0.01	4.70	3.77	2.92	2.06	1.13	0.95	5.30	6.05	7.00	7.95	15.10	22.35	20.10	29.00	47.60	75.70	99.75	99.90
50	-0.19	-0.18	-0.09	-0.09	-0.01	-0.01	3.62	2.77	2.09	1.46	0.73	0.59	4.90	5.20	5.00	4.85	8.35	13.25	26.80	40.15	65.90	92.40	100.00	100.00
100	-0.13	-0.05	-0.05	-0.04	-0.02	0.00	2.53	1.97	1.52	1.04	0.47	0.34	5.55	5.40	5.40	5.10	6.30	7.45	46.75	70.35	90.30	99.80	100.00	100.00
500	-0.07	-0.05	-0.02	-0.01	-0.01	0.00	1.12	0.86	0.67	0.48	0.20	0.14	4.25	3.95	4.85	5.45	5.00	4.95	99.15	100.00	100.00	100.00	100.00	100.00
1000	0.02	0.01	0.02	0.00	0.00	0.00	0.82	0.65	0.48	0.33	0.14	0.10	5.50	5.60	5.00	4.80	5.45	5.10	100.00	100.00	100.00	100.00	100.00	100.00
	GMM estimator																							
20	0.14	0.20	0.18	0.20	0.14	0.15	6.03	4.66	3.66	2.73	1.69	1.51	6.25	5.90	8.20	10.05	22.00	31.30	16.15	22.90	35.55	62.70	96.75	98.50
30	0.13	0.14	0.16	0.16	0.14	0.13	4.72	3.79	2.93	2.07	1.14	0.96	4.95	6.10	6.95	7.60	15.35	22.30	19.45	29.25	49.20	77.30	99.85	100.00
50	-0.10	-0.08	0.00	0.00	0.08	0.08	3.62	2.77	2.09	1.45	0.74	0.59	4.15	4.85	5.00	4.85	8.15	13.75	26.10	40.35	66.40	93.35	100.00	100.00
100	-0.09	0.00	0.00	0.01	0.03	0.05	2.53	1.96	1.51	1.03	0.47	0.35	4.80	4.85	5.30	4.95	6.50	7.55	45.45	70.10	90.35	99.80	100.00	100.00
500	-0.06	-0.04	-0.01	0.00	0.01	0.01	1.11	0.86	0.67	0.47	0.20	0.14	3.45	3.55	4.55	5.30	4.70	4.90	99.00	100.00	100.00	100.00	100.00	100.00
1000	0.02	0.02	0.02	0.01	0.00	0.00	0.82	0.64	0.48	0.33	0.14	0.10	4.65	5.00	5.05	4.90	5.50	5.45	100.00	100.00	100.00	100.00	100.00	100.00
	MLE																							
20	-0.06	-0.03	0.01	0.02	-0.07	-0.06	6.18	4.52	3.38	2.38	1.02	0.74	13.30	8.80	8.65	7.45	6.45	6.55	26.30	27.50	39.10	63.05	99.85	100.00
30	0.02	0.00	-0.06	-0.02	-0.04	-0.06	4.88	3.70	2.75	1.85	0.81	0.57	11.80	8.75	7.55	6.20	6.25	5.20	31.25	37.10	52.50	79.70	100.00	100.00
50	-0.17	-0.15	-0.12	-0.10	-0.05	-0.05	3.59	2.66	1.95	1.37	0.61	0.43	10.05	6.55	5.45	5.55	5.05	4.95	38.45	50.95	72.65	94.85	100.00	100.00
100	-0.14	-0.05	-0.07	-0.06	-0.04	-0.02	2.56	1.92	1.45	0.99	0.43	0.30	10.60	7.65	6.50	5.75	5.45	5.25	61.90	79.80	94.25	99.95	100.00	100.00
500	-0.03	-0.02	-0.02	-0.01	-0.02	-0.01	1.13	0.85	0.65	0.46	0.19	0.14	10.60	6.70	6.10	6.90	4.50	5.90	99.60	100.00	100.00	100.00	100.00	100.00
1000	0.03	0.01	0.02	-0.00	-0.02	-0.01	0.79	0.61	0.45	0.31	0.14	0.10	9.70	7.20	6.40	5.80	5.70	6.10	100.00	100.00	100.00	100.00	100.00	100.00

See notes to Table 3. The power is computed under $H_1 : \beta_1 = 0.95$.

Table 5: Small sample properties of estimators for spatial parameter ρ ($\rho = 0.4$, serially correlated and heteroskedastic errors), Bartlett window size $2\sqrt{T}$

	Bias(×100)						RMSE(×100)						Size(×100)						Power(×100)					
N\T	20	30	50	100	500	1000	20	30	50	100	500	1000	20	30	50	100	500	1000	20	30	50	100	500	1000
	Infeasible 2SLS estimator (including factors)																							
20	-0.12	-0.08	-0.11	-0.03	0.01	-0.01	3.66	2.93	2.26	1.59	0.73	0.52	6.35	6.10	6.35	6.05	5.70	6.70	10.35	12.15	15.50	25.90	80.30	97.00
30	-0.17	-0.10	-0.03	-0.01	0.01	0.00	2.94	2.40	1.85	1.30	0.58	0.41	5.30	5.80	5.85	5.95	4.55	5.70	10.55	15.40	20.90	35.15	92.40	99.70
50	0.04	0.04	-0.02	0.01	0.01	0.01	2.23	1.81	1.42	0.99	0.45	0.31	5.65	5.65	5.65	5.70	5.35	5.70	15.55	22.35	31.25	54.45	99.20	100.00
100	0.03	0.03	0.03	0.02	0.00	0.00	1.61	1.30	1.00	0.72	0.32	0.23	5.10	5.00	4.80	5.95	5.30	5.50	25.70	35.80	53.70	81.05	100.00	100.00
500	-0.04	-0.03	-0.02	-0.02	0.00	0.00	0.72	0.58	0.45	0.32	0.15	0.10	4.60	5.35	5.20	4.70	5.65	5.90	77.15	92.95	99.40	100.00	100.00	100.00
1000	-0.01	-0.01	-0.01	0.00	0.00	0.00	0.51	0.41	0.32	0.23	0.10	0.07	4.90	5.25	6.15	6.15	5.30	5.90	97.55	99.85	100.00	100.00	100.00	100.00
	2SLS estimator																							
20	-0.15	-0.07	-0.12	-0.07	-0.01	-0.02	4.04	3.23	2.55	1.82	1.08	0.94	7.30	7.85	9.05	7.95	14.30	21.70	11.40	13.65	16.60	25.85	70.65	85.80
30	-0.15	-0.08	-0.04	-0.01	0.00	-0.01	3.20	2.58	1.99	1.42	0.76	0.62	6.60	7.45	7.70	7.75	11.15	15.05	12.40	16.95	22.40	34.60	86.20	97.10
50	0.04	0.05	0.00	0.02	0.02	0.01	2.33	1.90	1.47	1.04	0.51	0.39	6.05	6.70	5.40	6.20	8.60	11.30	15.55	23.00	33.45	53.60	98.25	99.90
100	0.03	0.03	0.03	0.02	0.00	0.00	1.64	1.32	1.01	0.73	0.33	0.24	5.55	5.85	6.50	6.20	6.35	6.65	25.50	36.60	54.75	80.25	100.00	100.00
500	-0.04	-0.03	-0.02	-0.02	0.00	0.00	0.73	0.58	0.45	0.32	0.15	0.10	5.45	5.30	5.75	5.15	5.80	5.80	77.35	92.55	99.40	100.00	100.00	100.00
1000	-0.01	-0.01	-0.01	0.00	0.00	0.00	0.52	0.41	0.32	0.23	0.10	0.07	5.45	5.70	5.85	6.40	4.95	6.15	97.25	99.85	100.00	100.00	100.00	100.00
	B2SLS estimator																							
20	-0.26	-0.13	-0.17	-0.08	-0.01	-0.02	3.95	3.12	2.48	1.78	1.05	0.91	7.15	7.90	9.05	8.05	14.55	21.05	11.00	13.05	16.85	26.55	71.80	86.55
30	-0.20	-0.11	-0.05	-0.01	0.00	-0.01	3.10	2.52	1.94	1.39	0.75	0.61	6.95	7.40	7.55	8.05	11.05	15.40	12.40	17.10	23.00	35.85	87.40	97.15
50	0.03	0.04	0.00	0.01	0.01	0.00	2.27	1.84	1.44	1.01	0.49	0.38	5.95	6.75	6.30	5.95	8.20	10.55	16.90	23.10	34.85	55.50	98.50	99.95
100	0.02	0.02	0.03	0.02	0.00	0.00	1.61	1.30	0.99	0.71	0.32	0.23	6.35	5.90	5.45	5.60	5.95	6.50	26.80	37.00	56.30	82.90	100.00	100.00
500	-0.05	-0.03	-0.02	-0.02	0.00	0.00	0.71	0.57	0.45	0.31	0.14	0.10	5.50	6.05	6.25	5.40	6.00	6.15	78.60	93.15	99.70	100.00	100.00	100.00
1000	-0.01	-0.01	-0.01	0.00	0.00	0.00	0.51	0.40	0.31	0.23	0.10	0.07	5.60	5.15	5.95	6.30	5.05	5.55	97.70	99.95	100.00	100.00	100.00	100.00
	GMM estimator																							
20	-1.56	-1.45	-1.46	-1.42	-1.47	-1.55	3.70	3.09	2.58	2.10	1.73	1.73	9.65	9.85	14.25	19.65	60.45	83.05	7.10	7.80	8.45	11.20	23.50	32.25
30	-1.04	-0.94	-0.90	-0.89	-0.91	-0.93	2.89	2.38	1.94	1.50	1.10	1.06	8.25	9.55	11.80	15.05	45.70	68.70	8.95	11.65	14.05	19.80	57.65	78.00
50	-0.51	-0.48	-0.50	-0.48	-0.50	-0.52	2.08	1.71	1.37	1.01	0.66	0.61	6.75	7.40	8.20	9.75	28.30	48.40	13.20	19.15	27.40	44.45	95.60	99.60
100	-0.22	-0.22	-0.22	-0.23	-0.25	-0.26	1.42	1.15	0.89	0.66	0.38	0.33	5.85	6.25	6.45	7.70	16.95	29.25	24.70	36.70	56.75	82.05	100.00	100.00
500	-0.08	-0.07	-0.06	-0.06	-0.05	-0.05	0.62	0.49	0.39	0.27	0.13	0.10	5.35	4.75	5.55	5.25	7.60	8.75	87.60	97.85	99.90	100.00	100.00	100.00
1000	-0.03	-0.03	-0.03	-0.03	-0.02	-0.02	0.44	0.35	0.27	0.20	0.09	0.06	4.25	4.90	5.85	6.30	5.30	6.60	99.35	99.90	100.00	100.00	100.00	100.00
	MLE																							
20	0.48	0.26	0.15	0.17	0.20	0.18	3.34	2.60	1.93	1.34	0.65	0.48	24.45	20.35	18.00	16.80	18.55	20.60	37.10	36.80	40.35	58.40	98.35	100.00
30	0.38	0.22	0.19	0.15	0.15	0.15	2.63	2.08	1.59	1.09	0.51	0.37	21.20	19.20	17.50	16.15	16.30	18.10	39.90	42.15	50.30	71.25	99.80	100.00
50	0.45	0.25	0.14	0.13	0.11	0.11	2.02	1.57	1.19	0.84	0.38	0.28	22.10	18.05	15.50	15.70	15.35	17.10	50.95	55.55	65.05	86.25	100.00	100.00
100	0.39	0.22	0.14	0.10	0.07	0.07	1.46	1.13	0.85	0.59	0.27	0.20	22.55	18.80	15.80	15.35	14.25	16.65	66.60	74.95	87.60	98.50	100.00	100.00
500	0.27	0.12	0.05	0.03	0.03	0.04	0.69	0.50	0.36	0.26	0.12	0.09	23.20	17.00	14.60	13.10	14.90	17.50	99.40	99.80	100.00	100.00	100.00	100.00
1000	0.26	0.20	0.04	0.03	0.02	0.02	0.50	0.46	0.26	0.19	0.08	0.06	28.50	24.50	14.90	14.80	11.80	14.90	100.00	100.00	100.00	100.00	100.00	100.00

Notes: The true values are $\rho = 0.4$, $\beta_1 = 1$ and $\beta_2 = 2$, and the number of factors is $m = 2$. The maximum lag of the robust variance estimator is chosen to be $2\sqrt{T}$. Refer to notes to Table 1 for a description of all estimation methods. The number of replication is 2000. The 95% confidence interval for size 5% is $[3.6\%, 6.4\%]$, and the power is computed under $H_1 : \rho = 0.38$.

Table 6: Small sample properties of estimators for slope parameter β_1 ($\rho = 0.4$, serially correlated and heteroskedastic errors), Bartlett window size $2\sqrt{T}$

	Bias(×100)						RMSE(×100)						Size(×100)						Power(×100)					
N\T	20	30	50	100	500	1000	20	30	50	100	500	1000	20	30	50	100	500	1000	20	30	50	100	500	1000
	Infeasible 2SLS estimator (including factors)																							
20	0.17	0.12	0.18	0.11	0.00	-0.02	6.56	5.13	4.14	2.96	1.34	0.95	5.85	6.00	5.70	5.60	6.20	5.85	13.40	17.55	27.85	47.00	96.15	99.95
30	0.10	0.04	-0.01	-0.01	-0.02	-0.03	5.35	4.44	3.41	2.38	1.08	0.75	6.30	6.55	6.75	6.65	5.30	5.20	18.25	23.65	35.85	58.65	99.40	100.00
50	-0.23	-0.17	-0.09	-0.11	-0.02	-0.02	4.12	3.33	2.53	1.77	0.81	0.57	5.00	5.65	4.85	5.35	5.10	5.00	22.10	33.55	50.00	77.25	100.00	100.00
100	-0.17	-0.07	-0.04	-0.04	-0.02	0.00	2.95	2.37	1.86	1.30	0.56	0.40	5.70	5.95	6.35	5.50	5.20	4.75	38.80	57.50	78.20	97.40	100.00	100.00
500	-0.03	-0.02	-0.01	0.00	-0.01	0.00	1.26	1.04	0.82	0.59	0.26	0.18	4.70	4.75	5.60	5.45	5.15	5.50	97.10	99.65	100.00	100.00	100.00	100.00
1000	0.02	0.01	0.02	0.00	0.00	0.00	0.95	0.76	0.59	0.42	0.18	0.13	5.55	5.75	5.95	5.70	5.70	5.45	100.00	100.00	100.00	100.00	100.00	100.00
	2SLS estimator																							
20	0.21	0.18	0.15	0.10	0.02	0.00	6.84	5.43	4.37	3.22	1.88	1.61	7.85	7.80	8.65	9.85	16.35	23.45	16.55	21.45	29.85	47.90	91.95	96.75
30	0.13	0.07	0.05	0.04	0.00	-0.01	5.46	4.49	3.51	2.52	1.30	1.07	6.70	7.00	8.20	8.30	11.55	15.90	19.85	25.95	37.50	59.90	98.55	99.80
50	-0.20	-0.15	-0.06	-0.11	-0.01	-0.02	4.17	3.35	2.54	1.79	0.89	0.69	5.90	6.60	6.10	5.10	7.20	10.15	24.65	33.95	51.00	78.80	100.00	100.00
100	-0.14	-0.06	-0.05	-0.03	-0.02	0.00	2.93	2.38	1.86	1.30	0.58	0.43	5.65	6.85	6.75	5.85	5.85	6.00	40.15	57.15	78.75	97.45	100.00	100.00
500	-0.05	-0.04	-0.02	-0.01	-0.01	0.00	1.29	1.05	0.83	0.59	0.26	0.18	5.25	5.10	5.90	5.90	5.00	5.50	96.70	99.65	100.00	100.00	100.00	100.00
1000	0.02	0.00	0.02	0.00	0.00	0.00	0.95	0.76	0.60	0.42	0.19	0.13	6.30	6.40	6.40	5.85	5.85	5.35	100.00	100.00	100.00	100.00	100.00	100.00
	B2SLS estimator																							
20	0.23	0.20	0.16	0.11	0.02	0.00	6.84	5.43	4.36	3.21	1.88	1.61	8.05	7.65	8.65	9.90	16.55	23.30	16.45	21.35	30.05	48.05	91.75	96.75
30	0.14	0.08	0.06	0.04	0.00	-0.01	5.45	4.49	3.51	2.51	1.30	1.06	6.70	7.15	8.30	8.50	11.65	16.00	19.85	26.20	37.50	60.15	98.55	99.80
50	-0.20	-0.15	-0.06	-0.11	-0.01	-0.02	4.17	3.34	2.54	1.79	0.89	0.69	5.85	6.50	6.05	5.00	7.00	10.25	24.45	34.20	51.45	78.55	100.00	100.00
100	-0.14	-0.06	-0.05	-0.03	-0.02	0.00	2.93	2.37	1.86	1.30	0.58	0.43	5.75	6.90	6.65	5.85	6.00	6.05	40.00	57.25	78.90	97.50	100.00	100.00
500	-0.05	-0.04	-0.02	-0.01	-0.01	0.00	1.29	1.05	0.83	0.59	0.26	0.18	5.30	5.15	5.80	5.95	5.05	5.65	96.70	99.65	100.00	100.00	100.00	100.00
1000	0.02	0.00	0.02	0.00	0.00	0.00	0.95	0.76	0.60	0.42	0.19	0.13	6.35	6.35	6.40	5.95	5.75	5.50	100.00	100.00	100.00	100.00	100.00	100.00
	GMM estimator																							
20	0.24	0.29	0.26	0.25	0.18	0.16	6.82	5.45	4.41	3.23	1.89	1.61	7.20	7.45	8.95	9.95	17.10	23.15	15.45	21.45	30.35	49.10	92.85	97.55
30	0.21	0.19	0.18	0.17	0.13	0.13	5.48	4.52	3.54	2.53	1.31	1.07	6.30	7.65	8.45	8.95	12.00	16.90	19.00	26.15	38.35	62.15	98.65	99.85
50	-0.15	-0.08	0.01	-0.03	0.08	0.08	4.18	3.34	2.54	1.79	0.90	0.69	5.25	5.95	5.80	4.95	7.35	10.85	23.80	33.85	52.10	80.05	100.00	100.00
100	-0.11	-0.02	0.00	0.01	0.03	0.06	2.94	2.37	1.86	1.30	0.58	0.43	5.05	6.30	6.85	6.00	5.90	6.15	38.15	56.25	79.05	97.85	100.00	100.00
500	-0.05	-0.04	-0.01	0.00	0.00	0.01	1.29	1.04	0.82	0.59	0.26	0.18	4.80	4.45	5.50	5.95	4.90	5.60	96.40	99.65	100.00	100.00	100.00	100.00
1000	0.02	0.01	0.02	0.01	0.00	0.00	0.95	0.76	0.60	0.42	0.18	0.13	5.30	6.05	6.40	6.00	5.75	5.45	100.00	100.00	100.00	100.00	100.00	100.00
	MLE																							
20	0.09	0.02	0.09	0.06	-0.05	-0.07	6.94	5.36	4.18	2.95	1.30	0.93	22.00	16.90	17.55	16.15	14.35	14.75	34.65	36.45	45.05	62.15	98.90	99.95
30	0.10	-0.01	-0.05	-0.03	-0.05	-0.06	5.61	4.47	3.41	2.35	1.04	0.72	20.80	18.25	17.10	14.60	13.40	13.65	38.15	42.60	53.20	74.20	99.95	100.00
50	-0.16	-0.11	-0.11	-0.12	-0.05	-0.05	4.17	3.25	2.43	1.74	0.79	0.56	18.65	16.15	12.95	13.00	13.20	13.90	45.10	54.15	68.85	89.95	100.00	100.00
100	-0.10	-0.06	-0.06	-0.06	-0.03	-0.01	2.99	2.32	1.82	1.26	0.54	0.38	19.85	17.15	16.00	14.10	11.25	12.25	64.50	77.25	89.60	99.40	100.00	100.00
500	0.03	0.02	-0.01	-0.00	-0.02	-0.01	1.32	1.04	0.79	0.56	0.25	0.18	19.60	15.40	15.40	14.50	12.80	13.70	99.50	99.90	100.00	100.00	100.00	100.00
1000	0.07	0.01	0.04	0.01	-0.01	-0.01	0.96	0.73	0.57	0.40	0.18	0.13	20.50	17.00	16.60	13.50	14.30	14.10	100.00	100.00	100.00	100.00	100.00	100.00

See notes to Table 5. The power is computed under $H_1 : \beta_1 = 0.95$.

Table 7: Small sample properties of estimators for spatial parameter ρ ($\rho = 0.8$, independent and heteroskedastic errors)

	Bias(×100)						RMSE(×100)						Size(×100)						Power(×100)					
N\T	20	30	50	100	500	1000	20	30	50	100	500	1000	20	30	50	100	500	1000	20	30	50	100	500	1000
	Infeasible 2SLS estimator (including factors)																							
20	-0.08	-0.05	-0.06	-0.03	0.00	0.00	1.73	1.35	1.02	0.69	0.30	0.22	5.05	4.80	5.70	4.70	4.70	5.20	22.40	31.65	48.70	80.85	100.00	100.00
30	-0.06	-0.03	-0.01	0.00	0.01	0.00	1.35	1.07	0.82	0.55	0.24	0.17	3.95	4.40	4.80	4.20	4.30	4.95	29.55	44.20	67.70	93.60	100.00	100.00
50	0.02	0.02	0.00	0.00	0.00	0.00	1.03	0.82	0.62	0.43	0.19	0.13	4.65	4.65	4.65	4.60	5.25	4.60	46.10	67.55	87.75	99.20	100.00	100.00
100	0.01	0.01	0.01	0.00	0.00	0.00	0.72	0.57	0.44	0.31	0.14	0.10	3.50	3.65	3.90	4.35	5.25	5.35	74.20	92.00	99.60	100.00	100.00	100.00
500	-0.01	-0.01	0.00	0.00	0.00	0.00	0.33	0.26	0.20	0.14	0.06	0.04	4.35	4.35	4.50	4.60	5.70	5.70	100.00	100.00	100.00	100.00	100.00	100.00
1000	0.00	0.00	0.00	0.00	0.00	0.00	0.23	0.18	0.14	0.10	0.04	0.03	3.95	4.85	4.55	4.95	4.95	5.45	100.00	100.00	100.00	100.00	100.00	100.00
	2SLS estimator																							
20	-0.11	-0.06	-0.07	-0.05	-0.01	-0.01	2.17	1.70	1.33	0.93	0.58	0.53	7.10	7.40	9.35	9.45	21.20	30.25	19.35	26.80	40.85	66.05	98.70	99.35
30	-0.05	-0.01	-0.01	0.00	0.00	-0.01	1.64	1.28	0.96	0.68	0.38	0.33	6.15	6.25	7.55	7.55	14.30	21.70	28.45	40.40	59.65	87.25	99.95	99.95
50	0.00	0.02	0.00	0.00	0.01	0.00	1.15	0.91	0.68	0.47	0.24	0.19	5.50	5.65	5.20	5.20	9.75	13.60	42.45	62.00	83.70	98.95	100.00	100.00
100	0.01	0.00	0.01	0.01	0.00	0.00	0.78	0.62	0.46	0.32	0.15	0.11	4.65	5.05	4.55	5.05	6.55	7.45	71.40	89.40	99.05	100.00	100.00	100.00
500	-0.01	-0.01	0.00	0.00	0.00	0.00	0.35	0.27	0.20	0.14	0.06	0.04	5.35	4.95	4.40	4.65	5.75	6.05	99.90	100.00	100.00	100.00	100.00	100.00
1000	0.00	0.00	0.00	0.00	0.00	0.00	0.24	0.19	0.14	0.10	0.04	0.03	4.55	4.45	4.60	5.05	5.00	5.20	100.00	100.00	100.00	100.00	100.00	100.00
	B2SLS estimator																							
20	-0.21	-0.11	-0.09	-0.06	-0.01	-0.02	1.89	1.46	1.15	0.83	0.50	0.45	6.50	6.75	8.75	10.00	20.90	29.05	22.10	31.65	48.50	75.15	99.55	99.75
30	-0.10	-0.03	-0.01	0.00	0.00	-0.01	1.42	1.11	0.84	0.59	0.33	0.28	5.40	6.30	6.50	7.55	15.10	21.80	33.15	49.15	69.35	93.85	99.95	99.95
50	0.00	0.01	0.00	0.00	0.00	0.00	1.00	0.79	0.60	0.41	0.20	0.16	5.55	5.50	5.95	6.25	9.00	12.35	53.60	73.00	92.10	99.65	100.00	100.00
100	-0.01	-0.01	0.00	0.00	0.00	0.00	0.68	0.53	0.39	0.27	0.12	0.09	4.95	5.60	4.65	5.05	5.85	6.85	82.25	94.85	99.90	100.00	100.00	100.00
500	-0.01	0.00	0.00	0.00	0.00	0.00	0.30	0.23	0.18	0.12	0.05	0.04	5.10	5.10	5.15	4.60	5.90	5.75	100.00	100.00	100.00	100.00	100.00	100.00
1000	0.00	0.00	0.00	0.00	0.00	0.00	0.21	0.16	0.12	0.09	0.04	0.03	4.40	4.85	4.95	5.25	5.20	4.90	100.00	100.00	100.00	100.00	100.00	100.00
	GMM estimator																							
20	-1.12	-1.03	-0.96	-0.91	-0.83	-0.83	2.13	1.79	1.48	1.21	0.96	0.94	13.15	16.00	21.30	32.40	73.95	88.00	13.20	17.75	25.15	41.05	90.85	96.25
30	-0.67	-0.59	-0.57	-0.54	-0.52	-0.52	1.52	1.23	1.01	0.80	0.61	0.58	9.55	10.55	15.45	21.95	63.05	80.90	23.20	33.85	49.10	76.90	99.85	99.85
50	-0.36	-0.33	-0.32	-0.30	-0.29	-0.29	1.07	0.86	0.69	0.51	0.35	0.33	8.75	9.15	11.30	13.50	42.20	66.05	43.00	61.65	82.40	98.75	100.00	100.00
100	-0.16	-0.15	-0.15	-0.14	-0.14	-0.14	0.70	0.55	0.42	0.31	0.19	0.17	6.30	6.15	6.40	8.70	23.75	41.00	78.30	94.65	99.55	100.00	100.00	100.00
500	-0.04	-0.03	-0.03	-0.03	-0.03	-0.03	0.29	0.23	0.17	0.12	0.06	0.05	5.10	4.60	5.30	6.00	8.15	12.30	100.00	100.00	100.00	100.00	100.00	100.00
1000	-0.02	-0.02	-0.02	-0.02	-0.01	-0.01	0.20	0.16	0.12	0.09	0.04	0.03	5.20	5.25	4.75	5.65	6.10	8.35	100.00	100.00	100.00	100.00	100.00	100.00
	MLE																							
20	0.35	0.23	0.17	0.16	0.17	0.17	1.72	1.21	0.87	0.60	0.31	0.25	22.50	17.25	14.05	12.80	17.50	24.90	57.30	66.75	83.90	98.05	100.00	100.00
30	0.27	0.19	0.16	0.13	0.13	0.12	1.21	0.90	0.68	0.46	0.23	0.18	17.05	13.20	11.70	10.50	15.50	19.55	69.75	81.25	94.25	99.85	100.00	100.00
50	0.25	0.16	0.11	0.09	0.09	0.09	0.89	0.67	0.49	0.34	0.17	0.13	15.30	11.75	9.70	8.75	12.60	17.00	86.70	94.75	99.50	100.00	100.00	100.00
100	0.19	0.11	0.08	0.07	0.06	0.06	0.62	0.46	0.34	0.24	0.11	0.09	14.05	10.25	8.75	8.00	10.80	16.40	98.45	99.80	100.00	100.00	100.00	100.00
500	0.14	0.07	0.04	0.03	0.03	0.04	0.30	0.21	0.15	0.10	0.05	0.05	17.80	10.30	8.10	8.90	12.30	19.60	100.00	100.00	100.00	100.00	100.00	100.00
1000	0.12	0.06	0.03	0.01	0.01	0.01	0.23	0.16	0.11	0.07	0.03	0.02	21.60	12.10	8.40	7.40	7.60	8.10	100.00	100.00	100.00	100.00	100.00	100.00

Notes: The true values are $\rho = 0.8$, $\beta_1 = 1$ and $\beta_2 = 2$, and the number of factors is $m = 2$. The spatial weights matrix is the 1-ahead-and-1-behind circular neighbors matrix. Refer to notes to Table 1 for a description of all estimation methods. The number of replication is 2000. The 95% confidence interval for size 5% is $[3.6\%, 6.4\%]$, and the power is computed under $H_1 : \rho = 0.78$.

Table 8: Small sample properties of estimators for slope parameter β_1 ($\rho = 0.8$, independent and heteroskedastic errors)

	Bias(×100)						RMSE(×100)						Size(×100)						Power(×100)					
N\T	20	30	50	100	500	1000	20	30	50	100	500	1000	20	30	50	100	500	1000	20	30	50	100	500	1000
	Infeasible 2SLS estimator (including factors)																							
20	0.02	0.05	0.10	0.09	-0.02	-0.01	5.76	4.37	3.42	2.42	1.05	0.76	4.65	4.00	5.40	5.85	5.00	5.40	13.85	19.40	32.75	57.55	99.70	100.00
30	0.04	0.00	-0.01	-0.02	-0.02	-0.03	4.54	3.64	2.78	1.92	0.86	0.60	3.90	4.35	4.95	4.65	5.00	4.85	17.65	27.20	43.80	73.80	100.00	100.00
50	-0.20	-0.18	-0.11	-0.09	-0.02	-0.02	3.52	2.74	2.08	1.44	0.64	0.46	4.65	4.30	4.25	4.40	4.10	4.70	23.70	37.75	62.85	91.40	100.00	100.00
100	-0.14	-0.05	-0.05	-0.05	-0.02	0.00	2.50	1.98	1.52	1.05	0.46	0.32	4.60	5.30	5.05	4.20	5.15	4.30	46.35	69.35	90.10	99.75	100.00	100.00
500	-0.05	-0.03	-0.02	-0.01	-0.01	0.00	1.09	0.86	0.67	0.48	0.21	0.15	3.45	3.55	4.80	5.20	5.05	5.05	99.50	100.00	100.00	100.00	100.00	100.00
1000	0.01	0.01	0.02	0.00	0.00	0.00	0.81	0.64	0.49	0.34	0.15	0.10	5.05	5.20	4.90	4.70	5.60	4.95	100.00	100.00	100.00	100.00	100.00	100.00
	2SLS estimator																							
20	0.10	0.10	0.07	0.07	-0.01	0.00	6.17	4.72	3.71	2.77	1.70	1.52	6.80	5.85	8.40	9.70	21.55	31.30	16.80	22.45	34.05	58.65	95.85	98.40
30	0.05	0.02	0.03	0.03	0.00	0.00	4.79	3.85	2.98	2.10	1.15	0.97	5.55	6.35	7.15	7.50	15.00	21.60	19.60	28.50	45.55	74.00	99.60	99.90
50	-0.19	-0.18	-0.09	-0.09	-0.02	-0.01	3.69	2.84	2.14	1.49	0.75	0.60	4.55	5.15	4.95	5.05	9.00	13.50	24.85	38.45	63.35	91.45	100.00	100.00
100	-0.13	-0.06	-0.06	-0.04	-0.02	0.00	2.58	2.02	1.56	1.07	0.48	0.35	5.10	5.35	5.55	4.90	6.35	7.35	45.35	67.90	89.05	99.70	100.00	100.00
500	-0.06	-0.04	-0.02	-0.01	-0.01	0.00	1.15	0.88	0.68	0.49	0.21	0.15	4.15	3.75	5.05	5.40	4.90	5.30	98.55	99.95	100.00	100.00	100.00	100.00
1000	0.02	0.01	0.02	0.00	0.00	0.00	0.84	0.66	0.50	0.34	0.15	0.11	5.65	5.30	5.25	4.65	5.85	4.85	99.95	100.00	100.00	100.00	100.00	100.00
	B2SLS estimator																							
20	0.17	0.13	0.08	0.08	0.00	0.00	6.13	4.71	3.70	2.75	1.71	1.53	7.00	6.00	8.15	9.50	22.00	30.55	16.75	23.20	34.55	59.15	95.70	98.40
30	0.08	0.03	0.03	0.02	0.00	0.00	4.76	3.82	2.96	2.08	1.14	0.96	5.60	6.25	7.15	7.80	14.85	21.55	20.00	28.80	46.25	74.50	99.65	99.95
50	-0.19	-0.18	-0.09	-0.09	-0.01	-0.01	3.66	2.81	2.12	1.47	0.74	0.60	4.65	5.30	5.10	4.70	8.45	13.60	25.80	38.65	64.80	92.15	100.00	100.00
100	-0.12	-0.04	-0.05	-0.04	-0.02	0.00	2.57	2.00	1.55	1.06	0.48	0.35	5.35	4.95	5.30	5.20	6.25	7.45	45.90	68.90	89.75	99.70	100.00	100.00
500	-0.06	-0.04	-0.02	-0.01	-0.01	0.00	1.14	0.88	0.68	0.48	0.21	0.14	4.25	3.85	4.95	5.40	5.10	5.05	98.85	100.00	100.00	100.00	100.00	100.00
1000	0.02	0.01	0.02	0.00	0.00	0.00	0.83	0.66	0.49	0.34	0.15	0.10	5.40	5.50	5.10	4.70	5.75	5.45	99.95	100.00	100.00	100.00	100.00	100.00
	GMM estimator																							
20	0.53	0.56	0.53	0.53	0.44	0.45	6.15	4.76	3.75	2.81	1.75	1.58	6.45	5.65	8.35	10.70	24.00	34.30	17.40	24.55	38.05	64.65	97.40	98.85
30	0.37	0.36	0.38	0.36	0.33	0.33	4.79	3.86	3.00	2.12	1.19	1.02	4.85	6.15	7.35	7.95	16.25	24.75	19.60	29.85	50.60	79.35	99.85	99.95
50	0.04	0.05	0.13	0.12	0.20	0.20	3.68	2.83	2.13	1.48	0.77	0.63	4.45	4.70	4.70	4.95	9.75	15.05	26.20	40.40	67.35	93.95	100.00	100.00
100	-0.03	0.06	0.06	0.07	0.09	0.11	2.56	1.99	1.54	1.06	0.49	0.37	4.70	4.75	5.10	4.95	7.00	8.85	45.00	70.30	90.60	99.70	100.00	100.00
500	-0.05	-0.02	0.00	0.01	0.02	0.02	1.13	0.87	0.67	0.48	0.21	0.15	3.45	3.20	4.65	5.10	4.85	5.20	98.80	100.00	100.00	100.00	100.00	100.00
1000	0.03	0.02	0.03	0.01	0.01	0.01	0.83	0.65	0.49	0.34	0.15	0.10	4.55	4.80	5.30	5.10	5.35	5.70	99.95	100.00	100.00	100.00	100.00	100.00
	MLE																							
20	-0.22	-0.15	-0.09	-0.07	-0.16	-0.15	6.25	4.57	3.43	2.40	1.04	0.76	13.20	9.25	8.50	7.70	6.70	7.15	25.20	26.85	37.50	60.90	99.75	100.00
30	-0.12	-0.10	-0.14	-0.09	-0.11	-0.12	4.93	3.73	2.79	1.86	0.83	0.58	11.90	8.55	7.40	6.70	6.20	5.50	29.50	36.05	50.65	77.50	100.00	100.00
50	-0.30	-0.23	-0.18	-0.15	-0.10	-0.10	3.62	2.68	1.97	1.38	0.62	0.45	10.00	6.60	5.20	5.35	5.45	5.45	35.90	49.35	70.50	94.25	100.00	100.00
100	-0.23	-0.11	-0.12	-0.10	-0.07	-0.05	2.60	1.95	1.47	1.00	0.44	0.31	10.75	8.10	6.80	6.00	5.70	5.40	60.20	78.15	93.40	99.90	100.00	100.00
500	-0.10	-0.05	-0.03	-0.03	-0.04	-0.03	1.15	0.86	0.66	0.46	0.20	0.14	10.30	6.40	6.40	7.10	4.70	6.60	99.60	100.00	100.00	100.00	100.00	100.00
1000	-0.03	-0.02	0.01	-0.01	-0.02	-0.01	0.80	0.62	0.46	0.32	0.14	0.10	10.60	7.70	6.30	5.60	5.70	5.90	100.00	100.00	100.00	100.00	100.00	100.00

See notes to Table 7. The power is computed under $H_1 : \beta_1 = 0.95$.

Table 9: Identification experiments of SAR model under different α : small sample performance of MLE

	Bias($\times$100)				RMSE($\times$100)				Size($\times$100)				Power($\times$100)			
N\T	1	20	50	100	1	20	50	100	1	20	50	100	1	20	50	100
	$\alpha = 1$															
20	-19.63	-1.35	-0.61	-0.38	51.24	9.85	6.17	4.36	3.50	5.30	4.95	5.25	6.25	17.05	37.20	60.60
50	-9.51	-0.59	-0.34	-0.08	31.42	6.25	3.92	2.77	4.85	5.50	5.05	5.50	7.45	38.40	69.75	94.05
100	-4.87	-0.41	-0.14	0.01	21.27	4.41	2.78	1.95	5.40	5.30	5.10	5.15	10.10	59.50	93.40	99.90
500	-0.97	0.03	0.06	0.03	8.84	1.95	1.24	0.90	5.00	5.15	5.10	6.45	21.60	99.90	100.00	100.00
1000	-0.64	0.06	0.04	0.00	6.21	1.38	0.90	0.68	5.30	5.10	6.10	6.65	37.80	100.00	100.00	100.00
	$\alpha = 1/2$															
20	-31.73	-4.64	-2.28	-1.13	85.60	31.07	19.83	13.80	0.00	5.80	5.80	6.00	0.00	7.30	9.10	12.00
50	-30.17	-3.10	-1.27	-0.59	73.08	20.45	12.56	8.71	0.00	5.60	5.55	4.75	0.00	8.70	12.95	19.70
100	-26.41	-2.64	-1.14	-0.60	64.30	15.76	9.82	7.01	1.90	5.25	5.25	6.00	3.55	9.85	16.80	28.90
500	-17.32	-0.89	-0.23	-0.05	47.67	9.74	6.17	4.34	2.35	5.20	5.40	4.90	4.55	17.95	39.25	64.65
1000	-13.43	-0.84	-0.36	-0.20	39.93	8.39	5.19	3.56	5.05	6.00	6.15	5.30	7.00	23.30	48.10	78.70
	$\alpha = 1/3$															
20	-25.27	-4.63	-2.09	-1.01	91.58	46.45	31.01	21.58	0.00	3.40	6.15	6.20	0.00	6.65	7.15	8.20
50	-28.33	-4.65	-1.87	-0.54	87.65	37.21	23.65	16.54	0.00	6.00	6.00	5.20	0.00	6.60	8.45	10.85
100	-28.66	-5.13	-1.96	-1.10	82.56	30.50	19.34	13.30	0.00	4.70	5.55	5.20	0.00	6.20	8.55	11.05
500	-30.78	-2.71	-0.73	-0.19	72.72	19.92	12.46	8.74	0.00	5.35	5.25	4.55	0.00	9.20	13.90	23.10
1000	-28.90	-2.27	-0.89	-0.63	68.08	17.35	10.70	7.43	2.15	5.75	5.40	4.90	3.70	11.55	18.20	25.05
	$\alpha = 1/4$															
20	-25.27	-4.63	-2.09	-1.01	91.58	46.45	31.01	21.58	0.00	3.40	6.15	6.20	0.00	6.65	7.15	8.20
50	-22.18	-4.42	-1.38	-0.41	90.22	46.33	30.66	21.17	0.00	3.25	5.80	4.80	0.00	7.20	6.95	7.80
100	-27.96	-5.64	-2.03	-1.23	87.23	37.53	23.76	16.27	0.00	4.85	5.15	5.65	0.00	6.65	8.25	9.10
500	-30.66	-3.97	-1.26	-0.58	83.89	30.54	18.91	13.30	0.00	5.65	5.40	5.80	0.00	6.85	8.45	11.60
1000	-31.69	-3.58	-1.59	-1.10	80.58	26.16	16.11	11.19	0.00	6.20	5.65	5.15	0.00	7.55	10.10	13.85

Notes: True value of ρ is 0.2. The spatial weights matrix $\mathbf{W}$ is constructed such that the first $N_1 = \lfloor N^{\alpha} \rfloor$ rows contain the 5-ahead-and-5-behind spatial weights, where $\alpha \in [0, 1]$ and $\lfloor N^{\alpha} \rfloor$ is the integer part of N^{α}, and the rest $N_2 = N - N_1$ rows of $\mathbf{W}$ are all zeros. The number of replication is 2000. The 95% confidence interval for size 5% is $[3.6\%, 6.4\%]$. The power is calculated under the alternative $H_1 : \rho = 0.1$.

A Appendix: Notations and Lemmas

Some frequently used notations:

$\mathbf{u}_{it} = (e_{it}, \mathbf{v}_{it}')'$, $\boldsymbol{\epsilon}_{.t} = \boldsymbol{\Delta}^{-1}\mathbf{u}_{.t}$, $\bar{\boldsymbol{\epsilon}}_{.t} = \boldsymbol{\Theta}_a\boldsymbol{\epsilon}_{.t}$, $\boldsymbol{\Theta}_a = \frac{1}{N}\boldsymbol{\tau}_N' \otimes \mathbf{I}_{k+1}$,
$\mathbf{S}(\rho) = \mathbf{I}_N - \rho\mathbf{W}$, $\mathbf{G}(\rho) = \mathbf{W}\mathbf{S}^{-1}(\rho)$, $\mathbf{G}_0^b = \mathbf{I}_T \otimes \mathbf{G}_0$,
$\mathbf{X} = (\mathbf{X}_{.1}', \mathbf{X}_{.2}', ..., \mathbf{X}_{.T}')'$, $\mathbf{Y} = (\mathbf{Y}_{.1,}', \mathbf{Y}_{.2}', ..., \mathbf{Y}_{.T}')'$,
$\mathbf{L} = [(\mathbf{I}_T \otimes \mathbf{W})\mathbf{Y}, \mathbf{X}]$, $\mathbf{L}_0 = [(\mathbf{I}_T \otimes \mathbf{G}_0)\mathbf{X}\boldsymbol{\beta}_0, \mathbf{X}]$,
$\mathbf{M}_f^b = \mathbf{M}_f \otimes \mathbf{I}_N$ with $\mathbf{M}_f = \mathbf{I}_T - \mathbf{F}(\mathbf{F}'\mathbf{F})^{-}\mathbf{F}'$,
$\mathbf{M}^b = \bar{\mathbf{M}} \otimes \mathbf{I}_N$ with $\bar{\mathbf{M}} = \mathbf{I}_T - \bar{\mathbf{Z}}(\bar{\mathbf{Z}}'\bar{\mathbf{Z}})^{-}\bar{\mathbf{Z}}'$,
$\boldsymbol{\delta} = (\rho, \boldsymbol{\beta}')'$.

Lemma 1. *Under Assumptions 3 and 5, the matrix $\boldsymbol{\Delta}^{-1}$ has bounded row and column norms, where the (i,j)th subblock of $\boldsymbol{\Delta}^{-1}$ is given in (8).*

Proof. Consider the row norm,

$$\begin{aligned} ||\boldsymbol{\Delta}^{-1}||_\infty &= \max\left\{1, \max_{1\le i\le N}\{\sum_{j=1}^N |s_{ij}^{-1}| + \sum_{j=1}^N\sum_{p=1}^k |s_{ij}^{-1}\beta_p|\}\right\} \\ &\le \max\left\{1, \max_{1\le i\le N}\{\sum_{j=1}^N |s_{ij}^{-1}| + K\sum_{j=1}^N |s_{ij}^{-1}|\}|\right\} \\ &\le \max\left\{1, K\max_{1\le i\le N}\{\sum_{j=1}^N |s_{ij}^{-1}|\}\right\} = \max\{1, K||\mathbf{S}^{-1}||_\infty\}, \end{aligned}$$

which is bounded as $||\mathbf{S}^{-1}||_\infty$ is bounded.

Likewise, we can show that $||\boldsymbol{\Delta}^{-1}||_1 \le \max\left\{||\mathbf{S}^{-1}||_1, 1 + K||\mathbf{S}^{-1}||_1\right\} < K < \infty$.

□

Lemma 2. *Under Assumption 2, for all t,*

(a) $E(\bar{\boldsymbol{\epsilon}}_{.t}) = 0$, $Var(\bar{\boldsymbol{\epsilon}}_{.t}) = O(\frac{1}{N})$, and hence $\bar{\boldsymbol{\epsilon}}_{.t} \overset{q.m.}{\to} 0$ as $N \to \infty$,

(b) $E||\bar{\boldsymbol{\epsilon}}_{.t}||^2 = O(\frac{1}{N})$, $E||\bar{\boldsymbol{\epsilon}}_{.t}|| = O(\frac{1}{\sqrt{N}})$,

where $\bar{\boldsymbol{\epsilon}}_{.t} = \boldsymbol{\Theta}_a\boldsymbol{\epsilon}_{.t}$, $\boldsymbol{\Theta}_a = \frac{1}{N}\boldsymbol{\tau}_N' \otimes \mathbf{I}_{k+1}$ and $\boldsymbol{\epsilon}_{.t} = \boldsymbol{\Delta}^{-1}\mathbf{u}_{.t}$.

Proof. This lemma is a direct counterpart of Lemma 1 in Pesaran (2006). Although the way we define the stacked error terms is different, we will demonstrate that the same properties can be established.

(a) $E(\bar{\boldsymbol{\epsilon}}_{.t}) = 0$ immediately follows $E(\mathbf{u}_{.t}) = 0$. As for the variance,

$$Var(\bar{\boldsymbol{\epsilon}}_{.t}) = \boldsymbol{\Theta}_a\boldsymbol{\Delta}^{-1}E(\mathbf{u}_{.t}\mathbf{u}_{.t}')\boldsymbol{\Delta}^{-1'}\boldsymbol{\Theta}_a' = \boldsymbol{\Theta}_a\boldsymbol{\Delta}^{-1}\boldsymbol{\Sigma}_u\boldsymbol{\Delta}^{-1'}\boldsymbol{\Theta}_a'. \tag{A.1}$$

For any row vector of $\boldsymbol{\Theta}_a$, denoted as $\boldsymbol{\theta}_a$, we have

$$\boldsymbol{\theta}_a\boldsymbol{\Delta}^{-1}\boldsymbol{\Sigma}_u\boldsymbol{\Delta}^{-1'}\boldsymbol{\theta}_a' \le (\boldsymbol{\theta}_a\boldsymbol{\theta}_a')\lambda_{\max}(\boldsymbol{\Delta}^{-1}\boldsymbol{\Sigma}_u\boldsymbol{\Delta}^{-1'}) = \frac{1}{N}\lambda_{\max}(\boldsymbol{\Delta}^{-1}\boldsymbol{\Sigma}_u\boldsymbol{\Delta}^{-1'}). \tag{A.2}$$

Since $\boldsymbol{\Delta}^{-1}$ has bounded row and column norms by Lemma 1, and so does $\boldsymbol{\Sigma}_u$ under Assumption 2, it follows that the product $\boldsymbol{\Delta}^{-1}\boldsymbol{\Sigma}_u\boldsymbol{\Delta}^{-1'}$ has bounded row and column norms, and consequently $\lambda_{\max}(\boldsymbol{\Delta}^{-1}\boldsymbol{\Sigma}_u\boldsymbol{\Delta}^{-1'})$ is bounded, which proves the order of the $Var(\bar{\boldsymbol{\epsilon}}_{.t})$ is $O(\frac{1}{N})$. The last statement is readily established by the definition of convergence in quadratic mean.

(b) Note that

$$E||\bar{\boldsymbol{\epsilon}}_{.t}||^2 = Etr\left[(\boldsymbol{\Theta}_a\boldsymbol{\Delta}^{-1}\boldsymbol{u}_{.t}\boldsymbol{u}'_{.t}\boldsymbol{\Delta}^{-1'}\boldsymbol{\Theta}'_a)\right] = tr\left[\boldsymbol{\Theta}_a\boldsymbol{\Delta}^{-1}\boldsymbol{\Sigma}_u\boldsymbol{\Delta}^{-1'}\boldsymbol{\Theta}'_a\right] = O(\frac{1}{N}), \quad \text{(A.3)}$$

and then, $E||\bar{\boldsymbol{\epsilon}}_{.t}|| \leq [E||\bar{\boldsymbol{\epsilon}}_{.t}||^2]^{1/2} = O(\frac{1}{\sqrt{N}})$. □

Lemma 3. *Under Assumptions 1, 2 and 3, for all i,*

(a) $\frac{\bar{\boldsymbol{\epsilon}}'\bar{\boldsymbol{\epsilon}}}{T} = O_p(\frac{1}{N})$,

(b) $\frac{\mathbf{F}'\bar{\boldsymbol{\epsilon}}}{T} = O_p(\frac{1}{\sqrt{NT}})$,

(c) $\frac{\mathbf{V}'_{i.}\mathbf{F}}{T} = O_p(\frac{1}{\sqrt{T}})$,

(d) $\frac{\mathbf{e}'_{i.}\bar{\boldsymbol{\epsilon}}}{T} = O_P(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})$, $\frac{\mathbf{V}'_{i.}\bar{\boldsymbol{\epsilon}}}{T} = O_P(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})$,

(e) $\frac{\mathbf{X}'_{i.}\bar{\boldsymbol{\epsilon}}}{T} = O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})$,

where $\bar{\boldsymbol{\epsilon}} = (\bar{\boldsymbol{\epsilon}}_{.1}, ..., \bar{\boldsymbol{\epsilon}}_{.T})'$ *with* $\bar{\boldsymbol{\epsilon}}_{.t} = \boldsymbol{\Theta}_a\boldsymbol{\epsilon}_{.t}$, $\mathbf{F} = (\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_T)'$, $\mathbf{V}_{i.} = (\mathbf{v}_{i1}, ..., \mathbf{v}_{iT})'$ *and* $\mathbf{e}_{i.} = (e_{i1}, ..., e_{iT})'$.

Proof. Given Lemma 2, (a)(b)(c) can be proved following similar arguments as those for (A.10)-(A.12) in Lemma 2 in Pesaran (2006).

(d) By definition, $\frac{1}{T}\mathbf{e}'_{i.}\bar{\boldsymbol{\epsilon}} = \frac{1}{T}\mathbf{e}'_{i.}\mathbf{U}\boldsymbol{\Delta}^{-1'}\boldsymbol{\Theta}'_a$, where $\mathbf{U} = (\mathbf{u}_{.1}, \mathbf{u}_{.2}, ..., \mathbf{u}_{.T})'$ is a $T \times N(k+1)$ matrix. Notice that $\frac{1}{T}\mathbf{e}'_{i.}\bar{\boldsymbol{\epsilon}}$ is a $(k+1)$-dimensional row vector. For ease of exposition, we denote $\frac{1}{T}\mathbf{e}'_{i.}\bar{\boldsymbol{\epsilon}} = (\tilde{\mathbf{e}}_1, \tilde{\mathbf{e}}_2, ..., \tilde{\mathbf{e}}_{k+1})$, and let s^{-1}_{ij} be the (i,j)th element of the matrix $\mathbf{S}^{-1}$.

We begin by considering the first column of $\frac{1}{T}\mathbf{e}'_{i.}\bar{\boldsymbol{\epsilon}}$. By expansion,

$$\begin{aligned}\tilde{\mathbf{e}}_1 &= \frac{1}{NT}\sum_{t=1}^{T}\sum_{h=1}^{N}\sum_{q=1}^{N} e_{it}(e_{ht}s^{-1}_{qh} + \mathbf{v}'_{ht}s^{-1}_{qh}\boldsymbol{\beta}) \\ &= \frac{1}{NT}\sum_{t=1}^{T}\sum_{h=1}^{N} s^{-1}_{.h}(e_{it}e_{ht} + e_{it}\mathbf{v}'_{ht}\boldsymbol{\beta}),\end{aligned}$$

where $s^{-1}_{.h} = \sum_{q=1}^{N} s^{-1}_{qh} = O(1)$. It follows that

$$\begin{aligned}E(\tilde{\mathbf{e}}_1) &= \frac{1}{NT}\sum_{t=1}^{T}\sum_{h=1}^{N} s^{-1}_{.h}\left[E(e_{it}e_{ht}) + E(e_{it}\mathbf{v}_{ht})'\boldsymbol{\beta}\right] \\ &= \frac{1}{NT}\sum_{t=1}^{T} s^{-1}_{.i}E(e^2_{it}) = \frac{1}{NT}\sum_{t=1}^{T} s^{-1}_{.i}\sigma^2_i \\ &\leq \frac{K}{NT}\sum_{t=1}^{T} s^{-1}_{.i} = O(\frac{1}{N}),\end{aligned}$$

and

$$
\begin{aligned}
Var(\tilde{\mathbf{e}}_1) &= \frac{1}{N^2T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{h=1}^{N}\sum_{l=1}^{N} s_{.h}^{-1}s_{.l}^{-1}E[(e_{it}e_{ht}+e_{it}\mathbf{v}_{ht}'\boldsymbol{\beta})(e_{is}e_{ls}+e_{is}\mathbf{v}_{ls}'\boldsymbol{\beta})] \\
&= \frac{1}{N^2T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{h=1}^{N}\sum_{l=1}^{N} s_{.h}^{-1}s_{.l}^{-1}[E(e_{it}e_{ht}e_{is}e_{ls})+E(e_{it}\mathbf{v}_{ht}'\boldsymbol{\beta}e_{is}\mathbf{v}_{ls}'\boldsymbol{\beta}) \\
&\quad +E(e_{it}e_{ht}e_{is}\mathbf{v}_{ls}'\boldsymbol{\beta})+E(e_{it}\mathbf{v}_{ht}'\boldsymbol{\beta}e_{is}e_{ls})],
\end{aligned}
$$

where the last two terms are zeros due to independence between e_{it} and $\mathbf{v}_{js}$ for all (i, j, t, s), and the first two terms equal

$$
E(e_{it}e_{ht}e_{is}e_{ls}) = \begin{cases} E(e_{it}^2e_{is}^2) & if\ h=l=i \\ E(e_{it}e_{is})E(e_{lt}e_{ls}) & if\ h=l\neq i \\ 0 & otherwise \end{cases}, \tag{A.4}
$$

$$
E(e_{it}\mathbf{v}_{ht}'\boldsymbol{\beta}e_{is}\mathbf{v}_{ls}'\boldsymbol{\beta}) = \begin{cases} E(e_{it}e_{is})\boldsymbol{\beta}'E[(\mathbf{v}_{lt}\mathbf{v}_{ls}')]\boldsymbol{\beta} & if\ h=l \\ 0 & otherwise \end{cases}. \tag{A.5}
$$

Furthermore, since e_{it} and $\mathbf{v}_{js}$ have finite fourth moment and their autocovariance functions decay exponentially, we thus have

$$
\begin{aligned}
Var(\tilde{\mathbf{e}}_1) &= \frac{1}{N^2T^2}\sum_{t=1}^{T}\sum_{s=1}^{T} s_{.i}^{-1}s_{.i}^{-1}E(e_{it}^2e_{is}^2) + \frac{1}{N^2T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{h=1,h\neq l}^{N} s_{.h}^{-1}s_{.h}^{-1}E(e_{it}e_{is})E(e_{lt}e_{ls}) \\
&\quad + \frac{1}{N^2T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{h=1}^{N} s_{.h}^{-1}s_{.h}^{-1}E(e_{it}e_{is})\boldsymbol{\beta}'E[(\mathbf{v}_{lt}\mathbf{v}_{ls}')]\boldsymbol{\beta} \\
&= O(\frac{1}{N^2}) + O(\frac{1}{NT}),
\end{aligned}
$$

which implies that $\tilde{\mathbf{e}}_1 = O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})$.

We now turn to rest of the columns of $\frac{1}{T}\mathbf{e}_{i.}'\bar{\boldsymbol{\epsilon}}$. Pick any $\tilde{\mathbf{e}}_r$ $(r = 2, ..., k+1)$, we have

$$
\tilde{\mathbf{e}}_r = \frac{1}{NT}\sum_{t=1}^{T}\sum_{q=1}^{N} e_{it}\mathbf{v}_{qt}. \tag{A.6}
$$

Clearly $E(\tilde{\mathbf{e}}_r) = 0$, and

$$
Var(\tilde{\mathbf{e}}_r) = \frac{1}{N^2T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{q=1}^{N}\sum_{h=1}^{N} E(e_{it}e_{is})E(\mathbf{v}_{qt}\mathbf{v}_{hs}') = O(\frac{1}{NT}).
$$

Therefore, $\tilde{\mathbf{e}}_r = O_p(\frac{1}{\sqrt{NT}})$ and we establish that $\frac{\mathbf{e}_{i.}'\bar{\boldsymbol{\epsilon}}}{T} = O_P(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})$.

The second result in (d) can be proved in a similar manner.

(e) Note that $\frac{\mathbf{X}_{i.}'\bar{\boldsymbol{\epsilon}}}{T} = \mathbf{A}_i'(\frac{\mathbf{F}'\bar{\boldsymbol{\epsilon}}}{T}) + \frac{\mathbf{V}_{i.}'\bar{\boldsymbol{\epsilon}}}{T}$. The claim readily follows (b), (d) and the assumption that $||\mathbf{A}_i|| < K$. □

Lemma 4. *Let $\mathbf{\Pi} = \mathbf{F}\bar{\mathbf{C}}$, and then $\bar{\mathbf{Z}} = \mathbf{\Pi} + \bar{\boldsymbol{\epsilon}}$. Under Assumptions 1, 2, 3 and 4,*

(a) $\frac{\mathbf{\Pi}'\mathbf{\Pi}}{T} = O_p(1)$,

(b) $\frac{\mathbf{\Pi}'\bar{\boldsymbol{\epsilon}}}{T} = O_p(\frac{1}{\sqrt{NT}})$,

(c) $\frac{\bar{\mathbf{Z}}'\bar{\mathbf{Z}}}{T} = O_p(1)$,

(d) $\frac{\bar{\mathbf{Z}}'\mathbf{F}}{T} = O_p(1)$,

(e) $\frac{\bar{\mathbf{Z}}'\mathbf{V}_{i.}}{T} = O_P(\frac{1}{N}) + O_p(\frac{1}{\sqrt{T}})$,

(f) $\frac{\bar{\mathbf{Z}}'\mathbf{X}_{i.}}{T} = O_p(1)$,

(g) $\frac{\mathbf{\Pi}'\mathbf{X}_{i.}}{T} = O_p(1)$,

(h) $\frac{\bar{\mathbf{Z}}'\mathbf{e}_{i.}}{T} = O_P(\frac{1}{N}) + O_p(\frac{1}{\sqrt{T}})$,

(i) $\frac{\bar{\mathbf{Z}}'\bar{\boldsymbol{\epsilon}}}{T} = O_P(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})$.

Proof. (a) $\frac{\mathbf{\Pi}'\mathbf{\Pi}}{T} = \bar{\mathbf{C}}'\frac{\mathbf{F}'\mathbf{F}}{T}\bar{\mathbf{C}} = O_p(1)$, since the elements of $\bar{\mathbf{C}}$ are bounded and $\frac{\mathbf{F}'\mathbf{F}}{T} = O_p(1)$.

(b) $\frac{\mathbf{\Pi}'\bar{\boldsymbol{\epsilon}}}{T} = \bar{\mathbf{C}}'\frac{\mathbf{F}'\bar{\boldsymbol{\epsilon}}}{T} = O_p(\frac{1}{\sqrt{NT}})$, as the elements of $\bar{\mathbf{C}}$ are bounded and $\frac{\mathbf{F}'\bar{\boldsymbol{\epsilon}}}{T} = O_p(\frac{1}{\sqrt{NT}})$ by Lemma 3.

(c) $\frac{\bar{\mathbf{Z}}'\bar{\mathbf{Z}}}{T} = \frac{\mathbf{\Pi}'\mathbf{\Pi}}{T} + \frac{\bar{\boldsymbol{\epsilon}}'\bar{\boldsymbol{\epsilon}}}{T} + \frac{\mathbf{\Pi}'\bar{\boldsymbol{\epsilon}}}{T} + \frac{\bar{\boldsymbol{\epsilon}}'\mathbf{\Pi}}{T} = O_p(1) + O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{\sqrt{NT}}) = O_p(1)$.

(d) $\frac{\bar{\mathbf{Z}}'\mathbf{F}}{T} = \bar{\mathbf{C}}'\frac{\mathbf{F}'\mathbf{F}}{T} + \frac{\bar{\boldsymbol{\epsilon}}'\mathbf{F}}{T} = O_p(1) + O_p(\frac{1}{\sqrt{NT}}) = O_p(1)$.

(e) $\frac{\bar{\mathbf{Z}}'\mathbf{V}_{i.}}{T} = \bar{\mathbf{C}}'\frac{\mathbf{F}'\mathbf{V}_{i.}}{T} + \frac{\bar{\boldsymbol{\epsilon}}'\mathbf{V}_{i.}}{T} = O_p(\frac{1}{\sqrt{T}}) + \Big[O_P(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})\Big] = O_P(\frac{1}{N}) + O_p(\frac{1}{\sqrt{T}})$.

(f) Recall that $\mathbf{X}_{i.} = \mathbf{F}\mathbf{A}_i + \mathbf{V}_{i.}$. Then, $\frac{\bar{\mathbf{Z}}'\mathbf{X}_{i.}}{T} = \frac{\bar{\mathbf{Z}}'\mathbf{F}}{T}\mathbf{A}_i + \frac{\bar{\mathbf{Z}}'\mathbf{V}_{i.}}{T} = O_p(1) + O_P(\frac{1}{N}) + O_p(\frac{1}{\sqrt{T}}) = O_p(1)$.

(g) $\frac{\mathbf{\Pi}'\mathbf{X}_{i.}}{T} = \bar{\mathbf{C}}'\frac{\mathbf{F}'\mathbf{F}}{T}\mathbf{A}_i + \bar{\mathbf{C}}'\frac{\mathbf{F}'\mathbf{V}_{i.}}{T} = O_p(1) + O_p(\frac{1}{\sqrt{T}}) = O_p(1)$.

(h) $\frac{\bar{\mathbf{Z}}'\mathbf{e}_{i.}}{T} = \bar{\mathbf{C}}'\frac{\mathbf{F}'\mathbf{e}_{i.}}{T} + \bar{\mathbf{C}}'\frac{\bar{\boldsymbol{\epsilon}}'\mathbf{e}_{i.}}{T} = O_p(\frac{1}{\sqrt{T}}) + \Big[O_P(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})\Big] = O_P(\frac{1}{N}) + O_p(\frac{1}{\sqrt{T}})$.

(i) $\frac{\bar{\mathbf{Z}}'\bar{\boldsymbol{\epsilon}}}{T} = \frac{\mathbf{\Pi}'\bar{\boldsymbol{\epsilon}}}{T} + \frac{\bar{\boldsymbol{\epsilon}}'\bar{\boldsymbol{\epsilon}}}{T} = O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{N})$. □

Lemma 5. *Under Assumptions 1, 2, 3 and 4, for any i and j,*

(a) $\frac{\mathbf{X}_{i.}'\bar{\mathbf{M}}\mathbf{F}}{T} = O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})$,

(b) $\frac{\mathbf{X}_{i.}'\bar{\mathbf{M}}\mathbf{X}_{j.}}{T} = \frac{\mathbf{X}_{i.}'\mathbf{M}_f\mathbf{X}_{j.}}{T} + O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})$,

(c) $\frac{\mathbf{X}_{i.}'\bar{\mathbf{M}}\mathbf{e}_{j.}}{T} = \frac{\mathbf{X}_{i.}'\mathbf{M}_f\mathbf{e}_{j.}}{T} + O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})$,

(d) $\frac{\mathbf{e}_{i.}'\bar{\mathbf{M}}\mathbf{e}_{j.}}{T} = \frac{\mathbf{e}_{i.}'\mathbf{M}_f\mathbf{e}_{j.}}{T} + O_p(\frac{1}{N\sqrt{T}}) + O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{N^2})$,

(e) $\frac{\mathbf{e}_{i.}'\bar{\mathbf{M}}\mathbf{F}}{T} = O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})$,

(f) $\frac{\mathbf{F}'\bar{\mathbf{M}}\mathbf{F}}{T} = O_p(\frac{1}{N})$,

(g) $\frac{\mathbf{X}_{i.}'\bar{\mathbf{M}}\bar{\boldsymbol{\epsilon}}}{T} = \frac{\mathbf{X}_{i.}'\mathbf{M}_f\bar{\boldsymbol{\epsilon}}}{T} + O_p(\frac{1}{N})$.

Proof. (a) This result follows the same logic as that in Pesaran (2006) and Pesaran and Tosetti (2011). Here we only emphasize a few important relations and orders of probability for our setting, and for more details we refer readers to Lemma 4 in Kapetanios et al. (2011). Notice that

$$\mathbf{X}_{i.}'\bar{\mathbf{M}}\mathbf{F} = -(\mathbf{X}_{i.}'\bar{\mathbf{M}}\bar{\boldsymbol{\epsilon}})\bar{\mathbf{C}}'(\bar{\mathbf{C}}\bar{\mathbf{C}}')^{-1}, \tag{A.7}$$

$$\mathbf{X}_{i.}'\bar{\mathbf{M}}\bar{\boldsymbol{\epsilon}} = \mathbf{A}_i'\mathbf{F}'\bar{\mathbf{M}}\bar{\boldsymbol{\epsilon}} + \mathbf{V}_{i.}'\bar{\mathbf{M}}\bar{\boldsymbol{\epsilon}}, \tag{A.8}$$

$$\bar{\boldsymbol{\epsilon}}'\bar{\mathbf{M}}\mathbf{F} = -(\bar{\boldsymbol{\epsilon}}'\bar{\mathbf{M}}\bar{\boldsymbol{\epsilon}})\bar{\mathbf{C}}'(\bar{\mathbf{C}}\bar{\mathbf{C}}')^{-1}. \tag{A.9}$$

By substitution,

$$\mathbf{X}_{i.}'\bar{\mathbf{M}}\mathbf{F} = -\mathbf{A}_i'(\bar{\mathbf{C}}\bar{\mathbf{C}}')^{-1}\bar{\mathbf{C}}\bar{\boldsymbol{\epsilon}}'\bar{\mathbf{M}}\bar{\boldsymbol{\epsilon}}\bar{\mathbf{C}}'(\bar{\mathbf{C}}\bar{\mathbf{C}}')^{-1} - \mathbf{V}_{i.}'\bar{\mathbf{M}}\bar{\boldsymbol{\epsilon}}\bar{\mathbf{C}}'(\bar{\mathbf{C}}\bar{\mathbf{C}}')^{-1}. \tag{A.10}$$

Because

$$\begin{aligned}\frac{\bar{\boldsymbol{\epsilon}}'\bar{\mathbf{M}}\bar{\boldsymbol{\epsilon}}}{T} &= \frac{\bar{\boldsymbol{\epsilon}}'\bar{\boldsymbol{\epsilon}}}{T} - \left(\frac{\bar{\boldsymbol{\epsilon}}'\bar{\mathbf{Z}}}{T}\right)\left(\frac{\bar{\mathbf{Z}}'\bar{\mathbf{Z}}}{T}\right)^{-1}\left(\frac{\bar{\mathbf{Z}}'\bar{\boldsymbol{\epsilon}}}{T}\right) \\ &= O_p(\frac{1}{N}) + \left[O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})\right] O_p(1) \left[O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})\right] \\ &= O_p(\frac{1}{N}),\end{aligned} \tag{A.11}$$

$$\begin{aligned}\frac{\mathbf{V}_{i.}'\bar{\mathbf{M}}\bar{\boldsymbol{\epsilon}}}{T} &= \frac{\mathbf{V}_{i.}'\bar{\boldsymbol{\epsilon}}}{T} - \left(\frac{\mathbf{V}_{i.}'\bar{\mathbf{Z}}}{T}\right)\left(\frac{\bar{\mathbf{Z}}'\bar{\mathbf{Z}}}{T}\right)^{-1}\left(\frac{\bar{\mathbf{Z}}'\bar{\boldsymbol{\epsilon}}}{T}\right) \\ &= \left[O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})\right] + \left[O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{T}})\right] O_p(1) \left[O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})\right] \\ &= O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}}),\end{aligned} \tag{A.12}$$

and the norms of $\mathbf{A}_i'$ and $(\bar{\mathbf{C}}\bar{\mathbf{C}}')^{-1}\bar{\mathbf{C}}$ are bounded, the probability order in (a) follows.

The rest of the lemma can be proved based on Lemma 4 and using similar reasoning as that for Lemma 3 and 4 in Kapetanios et al. (2011). To save space, we only give the proof of (b) as an illustration.

(b) Let $\mathbf{\Pi} = \mathbf{F}\bar{\mathbf{C}}$ and $\mathbf{M}_\Pi = \mathbf{\Pi}(\mathbf{\Pi}'\mathbf{\Pi})^{-1}\mathbf{\Pi}'$. Then we have $\bar{\mathbf{Z}} = \mathbf{\Pi} + \bar{\boldsymbol{\epsilon}}$, and

$$\begin{aligned} & \left\| \frac{\mathbf{X}_{i.}'\bar{\mathbf{M}}\mathbf{X}_{j.}}{T} - \frac{\mathbf{X}_{i.}'\mathbf{M}_\Pi\mathbf{X}_{j.}}{T} \right\| \\ = & \left\| \frac{\mathbf{X}_{i.}'\bar{\mathbf{Z}}(\bar{\mathbf{Z}}'\bar{\mathbf{Z}})^{-1}\bar{\mathbf{Z}}'\mathbf{X}_{j.}}{T} - \frac{\mathbf{X}_{i.}'\mathbf{\Pi}(\mathbf{\Pi}'\mathbf{\Pi})^{-1}\mathbf{\Pi}'\mathbf{X}_{j.}}{T} \right\| \\ \leq & \left\| \tfrac{1}{T}(\mathbf{X}_{i.}'\bar{\mathbf{Z}} - \mathbf{X}_i'\mathbf{\Pi})(\bar{\mathbf{Z}}'\bar{\mathbf{Z}})^{-1}\bar{\mathbf{Z}}'\mathbf{X}_{j.} \right\| + \left\| \tfrac{1}{T}\mathbf{X}_{i.}'\mathbf{\Pi}\left((\bar{\mathbf{Z}}'\bar{\mathbf{Z}})^{-1} - (\mathbf{\Pi}'\mathbf{\Pi})^{-1}\right)\bar{\mathbf{Z}}'\mathbf{X}_{j.} \right\| \\ & + \left\| \tfrac{1}{T}\mathbf{X}_{i.}'\mathbf{\Pi}(\mathbf{\Pi}'\mathbf{\Pi})^{-1}(\bar{\mathbf{Z}}'\mathbf{X}_{j.} - \mathbf{\Pi}'\mathbf{X}_{j.}) \right\|. \end{aligned} \tag{A.13}$$

We examine the three terms above in turn:

$$\begin{aligned} & \left\| \tfrac{1}{T}(\mathbf{X}_{i.}'\bar{\mathbf{Z}} - \mathbf{X}_i'\mathbf{\Pi})(\bar{\mathbf{Z}}'\bar{\mathbf{Z}})^{-1}\bar{\mathbf{Z}}'\mathbf{X}_{j.} \right\| \leq \left\| \frac{\mathbf{X}_{i.}'\bar{\boldsymbol{\epsilon}}}{T} \right\| \left\| \left(\frac{\bar{\mathbf{Z}}'\bar{\mathbf{Z}}}{T}\right)^{-1} \frac{\bar{\mathbf{Z}}'\mathbf{X}_{j.}}{T} \right\| \\ = & \left[O_p(\tfrac{1}{N}) + O_p(\tfrac{1}{\sqrt{NT}})\right] O_p(1)O_p(1) = O_p(\tfrac{1}{N}) + O_p(\tfrac{1}{\sqrt{NT}}), \end{aligned} \tag{A.14}$$

$$
\begin{aligned}
&\left\|\tfrac{1}{T}\mathbf{X}_{i.}'\boldsymbol{\Pi}\left((\bar{\mathbf{Z}}'\bar{\mathbf{Z}})^{-1}-(\boldsymbol{\Pi}'\boldsymbol{\Pi})^{-1}\right)\bar{\mathbf{Z}}'\mathbf{X}_{j.}\right\| \\
\leq\ & \left\|-\tfrac{\bar{\boldsymbol{\epsilon}}'\bar{\boldsymbol{\epsilon}}}{T}-\tfrac{\boldsymbol{\Pi}'\bar{\boldsymbol{\epsilon}}}{T}-\tfrac{\bar{\boldsymbol{\epsilon}}'\boldsymbol{\Pi}}{T}\right\|\left\|\tfrac{\mathbf{X}_{i.}'\boldsymbol{\Pi}}{T}\left(\tfrac{\bar{\mathbf{Z}}'\bar{\mathbf{Z}}}{T}\right)^{-1}\right\|\left\|\left(\tfrac{\boldsymbol{\Pi}'\boldsymbol{\Pi}}{T}\right)^{-1}\tfrac{\bar{\mathbf{Z}}'\mathbf{X}_{j.}}{T}\right\| \\
=\ & \left[O_p(\tfrac{1}{N})+O_p(\tfrac{1}{\sqrt{NT}})+O_p(\tfrac{1}{\sqrt{NT}})\right][O_p(1)O_p(1)]\,[O_p(1)O_p(1)] \\
=\ & O_p(\tfrac{1}{N})+O_p(\tfrac{1}{\sqrt{NT}}),
\end{aligned}
\tag{A.15}
$$

$$
\begin{aligned}
&\left\|\tfrac{1}{T}\mathbf{X}_{i.}'\boldsymbol{\Pi}(\boldsymbol{\Pi}'\boldsymbol{\Pi})^{-1}(\bar{\mathbf{Z}}'\mathbf{X}_{j.}-\boldsymbol{\Pi}'\mathbf{X}_{j.})\right\| \\
\leq\ & \left\|\tfrac{\mathbf{X}_{i.}'\boldsymbol{\Pi}}{T}\left(\tfrac{\boldsymbol{\Pi}'\boldsymbol{\Pi}}{T}\right)^{-1}\right\|\left\|\tfrac{\bar{\boldsymbol{\epsilon}}'\mathbf{X}_{j.}}{T}\right\| = [O_p(1)O_p(1)]\left[O_p(\tfrac{1}{N})+O_p(\tfrac{1}{\sqrt{NT}})\right] \\
=\ & O_p(\tfrac{1}{N})+O_p(\tfrac{1}{\sqrt{NT}}).
\end{aligned}
\tag{A.16}
$$

Under the full rank condition (Assumption 4), $\mathbf{M}_{\Pi}=\mathbf{M}_f$, and hence $\left\|\frac{\mathbf{X}_{i.}'\bar{\mathbf{M}}\mathbf{X}_{j.}}{T}-\frac{\mathbf{X}_{i.}'\mathbf{M}_f\mathbf{X}_{j.}}{T}\right\|=O_p(\frac{1}{N})+O_p(\frac{1}{\sqrt{NT}})$.

□

Lemma 6. *Under Assumptions 1-6,*

(a) $\frac{1}{NT}\mathbf{Q}'\mathbf{M}^b(\mathbf{I}_T\otimes\boldsymbol{\Gamma})\mathbf{f}=O_p(\frac{1}{N})+O_p(\frac{1}{\sqrt{NT}})$,

(b) $\frac{1}{NT}\mathbf{Q}'\mathbf{M}^b(\mathbf{I}_T\otimes\mathbf{B})\mathbf{e}=O_p(\frac{1}{N})+O_p(\frac{1}{\sqrt{NT}})$,

(c) $\frac{1}{NT}\mathbf{Q}'\mathbf{M}^b(\mathbf{I}_T\otimes\mathbf{B})\mathbf{X}=\frac{1}{NT}\mathbf{Q}'\mathbf{M}_f^b(\mathbf{I}_T\otimes\mathbf{B})\mathbf{X}+O_p(\frac{1}{N})+O_p(\frac{1}{\sqrt{NT}})$,

where $\mathbf{B}$ *is any* $N\times N$ *nonstochastic matrix with bounded row and column norms.*

Proof. (a) Taking a column from $\mathbf{Q}$ and expressing it generically as $\mathbf{Q}_c=[(\mathbf{W}^s\mathbf{X}_{.1,p})',(\mathbf{W}^s\mathbf{X}_{.2,p})',...,(\mathbf{W}^s\mathbf{X}_{.T,p})']'$, where $p=1,...,k,\ s=0,1,2,...,$ and $\mathbf{W}^0\equiv\mathbf{I}_N$, we have

$$
\begin{aligned}
\frac{1}{NT}\mathbf{Q}_c'\mathbf{M}^b(\mathbf{I}_T\otimes\boldsymbol{\Gamma})\mathbf{f} &= \frac{1}{NT}\mathbf{Q}_c'vec(\boldsymbol{\Gamma}\mathbf{F}'\bar{\mathbf{M}}') \\
&= \frac{1}{NT}tr\left\{\mathbf{W}^s[\mathbf{X}_{1.,p},\mathbf{X}_{2.,p},...,\mathbf{X}_{N.,p}]'\bar{\mathbf{M}}\mathbf{F}\boldsymbol{\Gamma}'\right\} \\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{l=1}^{N}w_{il}^s\mathbf{X}_{l.,p}'\bar{\mathbf{M}}\mathbf{F}\boldsymbol{\gamma}_i.
\end{aligned}
$$

Evidently, the claim in (a) readily follows Lemma 5(a), and the assumptions that $\boldsymbol{\gamma}_i$ is bounded and $\mathbf{W}$ has bounded row and column norms.

(b) Let $\mathbf{e}_{TN}=(\mathbf{e}_{1.},\mathbf{e}_{2.},...,\mathbf{e}_{N.})$ be the $T\times N$ matrix of errors, and b_{ij} denote the (i,j)th element of $\mathbf{B}$. Taking a column from $\mathbf{Q}$, as in the proof of (a) we can show that

$$\begin{aligned}\frac{1}{NT}\mathbf{Q}_c'(\bar{\mathbf{M}}\otimes\mathbf{B})\mathbf{e} &= \frac{1}{NT}\sum_{i=1}^N\sum_{j=1}^N\sum_{l=1}^N b_{ij}w_{il}^s\mathbf{X}_{l.,p}'\bar{\mathbf{M}}\mathbf{e}_{j.} \\ &= \frac{1}{NT}\sum_{i=1}^N\sum_{j=1}^N\sum_{l=1}^N b_{ij}w_{il}^s\mathbf{X}_{l.,p}'\mathbf{M}_f\mathbf{e}_{j.}+O_p(\frac{1}{N})+O_p(\frac{1}{\sqrt{NT}}) \\ &= \frac{1}{NT}\sum_{i=1}^N\sum_{j=1}^N\sum_{l=1}^N b_{ij}w_{il}^s\mathbf{V}_{l.,p}'\mathbf{e}_{j.}+O_p(\frac{1}{N})+O_p(\frac{1}{\sqrt{NT}}), \qquad (A.17)\end{aligned}$$

where the second equality follows by Lemma 5(c) and the assumption that $\mathbf{B}$ and $\mathbf{W}$ have bounded row and column norms.

Consider the first term in (A.17). Its mean is zero and its variance is given by

$$\begin{aligned}&\tfrac{1}{N^2T^2}E\left\{\textstyle\sum_{i=1}^N\sum_{l=1}^N\sum_{j=1}^N b_{ij}b_{lj}\mathbf{W}_i^{s\prime}\left[\mathbf{V}_{1.,p},...,\mathbf{V}_{N.,p}\right]'\mathbf{e}_{j.}\mathbf{e}_{j.}'\left[\mathbf{V}_{1.,p},...,\mathbf{V}_{N.,p}\right]\mathbf{W}_l^s\right\} \\ =\ &\tfrac{1}{N^2T^2}\textstyle\sum_{i=1}^N\sum_{l=1}^N\sum_{j=1}^N b_{ij}b_{lj}\sum_{m=1}^N\sum_{n=1}^N w_{im}w_{ln}E(\mathbf{V}_{m.,p}'\mathbf{\Omega}_{e,j}\mathbf{V}_{n.,p}),\end{aligned} \qquad (A.18)$$

where $\mathbf{\Omega}_{e,j}$ is the variance-covariance matrix of $\mathbf{e}_{j.}$. Because e_{jt} is stationary with absolutely summable autocovariances, $\mathbf{\Omega}_{e,j}$ has bounded row and column norms. It follows that

$$\mathbf{V}_{m.,p}'\mathbf{\Omega}_{e,j}\mathbf{V}_{n.,p}\leq\lambda_1(\mathbf{\Omega}_{e,j})(\mathbf{V}_{m.}'\mathbf{V}_{n.})\leq K(\mathbf{V}_{m.}'\mathbf{V}_{n.})=O(T). \qquad (A.19)$$

Also, $\sum_{l=1}^N\sum_{j=1}^N b_{ij}b_{lj}=O(1)$, since $\mathbf{BB}'$ has bounded row and column norms. Hence,

$$Var(\frac{1}{NT}\sum_{i=1}^N\sum_{j=1}^N\sum_{l=1}^N b_{ij}w_{il}^s\mathbf{V}_{l.,p}'\mathbf{e}_{j.})=O(\frac{1}{NT}), \qquad (A.20)$$

and consequently the order of the first term in (A.17) is $O_p(\frac{1}{\sqrt{NT}})$, which completes the proof.

(c) Let $\mathbf{C}=\mathbf{B}'\mathbf{W}^s$ and c_{ij} be its (i,j)th entry. For any column of $\mathbf{Q}$,

$$\begin{aligned}\tfrac{1}{NT}\mathbf{Q}_c'\mathbf{M}^b(\mathbf{I}_T\otimes\mathbf{B})\mathbf{X} &= \frac{1}{NT}tr\left\{\left[\mathbf{X}_{1.,p},...,\mathbf{X}_{N.,p}\right]'\bar{\mathbf{M}}\left[\mathbf{X}_{1.,p},...,\mathbf{X}_{N.,p}\right]\mathbf{B}'\mathbf{W}^s\right\} \\ &= \tfrac{1}{NT}\textstyle\sum_{i=1}^N\sum_{j=1}^N\mathbf{X}_i'\bar{\mathbf{M}}\mathbf{X}_jc_{ji} \\ &= \tfrac{1}{NT}\textstyle\sum_{i=1}^N\sum_{j=1}^N\mathbf{X}_i'\mathbf{M}_f\mathbf{X}_fc_{ji}\quad+O_p(\frac{1}{N})+O_p(\frac{1}{\sqrt{NT}}).\end{aligned}$$

Again, the last line follows by Lemma 5 and $\sum_{j=1}^N c_{ji}=O(1)$.

□

Lemma 7. *Under Assumptions 1-5, for any $N\times N$ nonstochastic matrix $\mathbf{B}$ with bounded row and column norms,*

(a) $\frac{1}{NT}\mathbf{e}'(\bar{\mathbf{M}}\otimes\mathbf{B})\mathbf{e}-\frac{1}{N}\sum_{i=1}^N b_{ii}\sigma_i^2=o_p(1)$. *In particular,* $\frac{1}{NT}\mathbf{e}'(\bar{\mathbf{M}}\otimes\mathbf{B})\mathbf{e}=o_p(1)$ *if* $diag(\mathbf{B})=\mathbf{0}$.

(b) $\frac{1}{NT}\mathbf{f}'(\bar{\mathbf{M}}\otimes\mathbf{\Gamma}'\mathbf{B})\mathbf{e}=O_p(\frac{1}{N})+O_p(\frac{1}{\sqrt{NT}})$,

(c) $\frac{1}{NT}\mathbf{f}'(\bar{\mathbf{M}}\otimes\mathbf{\Gamma}'\mathbf{B}\mathbf{\Gamma})\mathbf{f}=O_p(\frac{1}{N})$.

Proof. (a) Applying Lemma 5, we have

$$\begin{aligned}\frac{1}{NT}\mathbf{e}'(\bar{\mathbf{M}}\otimes\mathbf{B})\mathbf{e} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbf{e}_{i.}^{'}\bar{\mathbf{M}}\mathbf{e}_{j.}b_{ji} \\ &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbf{e}_{i.}^{'}\mathbf{M}_f\mathbf{e}_{j.}b_{ji}+\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}b_{ji}T\left[O_p(\frac{1}{N\sqrt{T}})+O_p(\frac{1}{\sqrt{NT}})+O_p(\frac{1}{N^2})\right] \\ &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbf{e}_{i.}^{'}\mathbf{e}_{j.}b_{ji}+O_p(\frac{1}{T})+O_p(\frac{1}{N\sqrt{T}})+O_p(\frac{1}{\sqrt{NT}})+O_p(\frac{1}{N^2}) \\ &= \frac{1}{NT}\mathbf{e}'(\mathbf{I}_T\otimes\mathbf{B})\mathbf{e}+O_p(\frac{1}{T})+O_p(\frac{1}{N\sqrt{T}})+O_p(\frac{1}{N^2}).\end{aligned}$$

Clearly, it suffices to show that $\frac{1}{NT}\mathbf{e}'(\mathbf{I}_T\otimes\mathbf{B})\mathbf{e}$ converges to its mean uniformly. First,

$$\begin{aligned}E\left(\frac{1}{NT}\mathbf{e}'(\mathbf{I}_T\otimes\mathbf{B})\mathbf{e}\right) &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}b_{ji}E(\mathbf{e}_{i.}^{'}\mathbf{e}_{j.}) \\ &= \frac{1}{N}\sum_{i=1}^{N}b_{ii}\sigma_i^2=O(1), \qquad \text{(A.21)}\end{aligned}$$

since e_{it} is independent from $e_{jt'}$ for any $i\neq j$, and obviously the mean is zero if $b_{ii}=0$ for all i.

Next, consider the second moment

$$E\left[\left(\frac{1}{NT}\mathbf{e}'(\mathbf{I}_T\otimes\mathbf{B})\mathbf{e}\right)^2\right]=\frac{1}{N^2T^2}\left\{\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{q=1}^{N}b_{ji}b_{ql}E\left[(\mathbf{e}_{i.}^{'}\mathbf{e}_{j.})(\mathbf{e}_{l.}^{'}\mathbf{e}_{q.})\right]\right\}, \qquad \text{(A.22)}$$

where

$$E\left[(\mathbf{e}_{i.}^{'}\mathbf{e}_{j.})(\mathbf{e}_{l.}^{'}\mathbf{e}_{q.})\right]=\begin{cases}E[(\mathbf{e}_{i.}^{'}\mathbf{e}_{i.})^2]=\sum_{t=1}^{T}\sum_{s=1}^{T}E(e_{it}^2e_{is}^2) & if\ i=j=l=q\\ E[(\mathbf{e}_{i.}^{'}\mathbf{e}_{j.})^2]=\sum_{t=1}^{T}\sum_{s=1}^{T}E(e_{it}e_{is})E(e_{jt}e_{js})=tr\left[\mathbf{\Omega}_{e,i}\mathbf{\Omega}_{e,j}\right] & if\ i=l\neq j=q\\ E[(\mathbf{e}_{i.}^{'}\mathbf{e}_{j.})^2] & if\ i=q\neq j=l\\ E(\mathbf{e}_{i.}^{'}\mathbf{e}_{i.})E(\mathbf{e}_{l.}^{'}\mathbf{e}_{l.})=T^2\sigma_i^2\sigma_l^2 & if\ i=j\neq l=q\\ 0 & otherwise\end{cases}. \qquad \text{(A.23)}$$

It follows that

$$
\begin{aligned}
Var\left[\frac{1}{NT}\mathbf{e}'(\mathbf{I}_T\otimes\mathbf{B})\mathbf{e}\right] &= \frac{1}{N^2T^2}\Bigg\{\sum_{i=1}^{N} b_{ii}^2 E[(\mathbf{e}_{i.}^{'}\mathbf{e}_{i.})^2] \\
&\quad +\sum_{i=1}^{N}\sum_{l=1,l\neq i}^{N} b_{ii}b_{ll}E(\mathbf{e}_{i.}^{'}\mathbf{e}_{i.})E(\mathbf{e}_{l.}^{'}\mathbf{e}_{l.}) \\
&\quad +\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}(b_{ji}^2+b_{ji}b_{ij})E[(\mathbf{e}_{i.}^{'}\mathbf{e}_{j.})^2]\Bigg\} \\
&\quad -\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{l=1}^{N} b_{ii}b_{ll}E(\mathbf{e}_{i.}^{'}\mathbf{e}_{i.})E(\mathbf{e}_{l.}^{'}\mathbf{e}_{l.}) \\
&= \frac{1}{N^2T^2}\sum_{i=1}^{N} b_{ii}^2\left\{E\left[(\mathbf{e}_{i.}^{'}\mathbf{e}_{i.})^2\right]-\left[E(\mathbf{e}_{i.}^{'}\mathbf{e}_{i.})\right]^2-2tr\left[\mathbf{\Omega}_{e,i}\mathbf{\Omega}_{e,i}\right]\right\} \\
&\quad +\frac{1}{N^2T^2}\left\{\sum_{i=1}^{N}\sum_{j=1}^{N} b_{ji}(b_{ji}+b_{ij})tr\left[\mathbf{\Omega}_{e,i}\mathbf{\Omega}_{e,j}\right]\right\}. \qquad \text{(A.24)}
\end{aligned}
$$

It is readily seen that

$$
Var\left[\frac{1}{NT}\mathbf{e}'(\mathbf{M}_f\otimes\mathbf{B})\mathbf{e}\right] = \frac{1}{N^2T^2}\left\{\sum_{i=1}^{N}\sum_{j=1}^{N} b_{ji}(b_{ji}+b_{ij})tr\left[\mathbf{\Omega}_{e,i}\mathbf{\Omega}_{e,j}\right]\right\} = O(\frac{1}{NT}) \qquad \text{(A.25)}
$$

if $b_{ii}=0$ for all i, where the second equality follows from the assumption that $\mathbf{B}$ and $\mathbf{\Omega}_{e,i}$ $(\forall i)$ are uniformly bounded in row and column sums. In general, when $diag(\mathbf{B})\neq\mathbf{0}$, the first term in (A.24) does not equal zero but is of order $O(\frac{1}{NT})$ because

$$
\begin{aligned}
E\left[(\mathbf{e}_{i.}^{'}\mathbf{e}_{i.})^2\right]-\left[E(\mathbf{e}_{i.}^{'}\mathbf{e}_{i.})\right]^2-2tr\left[\mathbf{\Omega}_{e,i}\mathbf{\Omega}_{e,i}\right] &= \sum_{t_1=1}^{T}\sum_{t_2=1}^{T} cum(e_{t_1},e_{t_1},e_{t_2},e_{t_2}) \\
&= T\sum_{t=1}^{T} cum(e_0,e_0,e_t,e_t) \\
&\leq T\sum_{t_1,t_2,t_3=1}^{T}|cum(e_0,e_{t_1},e_{t_2},e_{t_3})| = O(T),
\end{aligned}
$$

where $cum(.)$ denotes the cumulant, the first equality follows the definition of the fourth cumulant, the second equality follows by the stationarity of e_{it}, and the final result follows by Assumption 2 that the fourth-order cumulant of e_{it} is absolutely summable. We thus establish that

$$
Var\left[\left(\frac{1}{NT}\mathbf{e}'(\mathbf{I}_T\otimes\mathbf{B})\mathbf{e}\right)^2\right] = O(\frac{1}{NT}), \qquad \text{(A.26)}
$$

and by the Chebyshev's inequality $\frac{1}{NT}\mathbf{e}'(\mathbf{I}_T \otimes \mathbf{B})\mathbf{e}$ converges to its zero uniformly at the rate of $O_p(\frac{1}{\sqrt{NT}})$, and this finishes the proof.

(b) Let $\mathbf{C} = \mathbf{\Gamma}'\mathbf{B}$, then $\frac{1}{NT}\mathbf{f}'(\bar{\mathbf{M}} \otimes \mathbf{\Gamma}'\mathbf{B})\mathbf{e} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{m}\mathbf{e}_{i.}^{'}\overline{\mathbf{M}}\mathbf{F}c_{ji}$. Its probability order is immediately established by applying Lemma 5 and noting that all elements c_{ij}'s are uniformly bounded.

(c) The proof is similar to that of (b). □

Lemma 8. *Under Assumption 2, for any two $N \times N$ nonstochastic matrices $\mathbf{B}$ and $\mathbf{D}$ with bounded row and column norms and satisfying diag$(\mathbf{B}) = $ diag$(\mathbf{D}) = \mathbf{0}$,*

(a) $E\left[\mathbf{e}'(\mathbf{I}_T \otimes \mathbf{B})\mathbf{e}\right] = 0$,

(b) $E\left[(\mathbf{e}'(\mathbf{I}_T \otimes \mathbf{B})\mathbf{e})^2\right] = tr\left[(\mathbf{B} \odot \mathbf{B}^s)\,\mathbf{\Sigma}_{ee}\right] = \sum_{i=1}^{N}\sum_{j=1}^{N} b_{ji}(b_{ij}+b_{ji})tr[\mathbf{\Omega}_{e,i}\mathbf{\Omega}_{e,j}]$,

(c) $E\left[\mathbf{e}'(\mathbf{I}_T \otimes \mathbf{B})\mathbf{e}\mathbf{e}'(\mathbf{I}_T \otimes \mathbf{D})\mathbf{e}\right] = tr\left[(\mathbf{B} \odot \mathbf{D}^s)\,\mathbf{\Sigma}_{ee}\right] = \sum_{i=1}^{N}\sum_{j=1}^{N} b_{ji}(d_{ij}+d_{ji})tr[\mathbf{\Omega}_{e,i}\mathbf{\Omega}_{e,j}]$,

where $\mathbf{B}^s = \mathbf{B} + \mathbf{B}^{'}$, $\mathbf{D}^s$ is defined likewise, and $\mathbf{\Sigma}_{ee}$ is an $N \times N$ matrix of which the (i,j)th element is $tr[\mathbf{\Omega}_{e,i}\mathbf{\Omega}_{e,j}]$.

Proof. The derivation of (a) and (b) are part of the proof in Lemma 7(a). The result in (c) can be verified similarly. □

Lemma 9. *Consider the linear-quadratic form: $h = \mathbf{e}^{'}(\mathbf{I}_T \otimes \mathbf{B})\mathbf{e} + \mathbf{c}^{'}\mathbf{e}$, where $\mathbf{e}$ is an $NT \times 1$ vector of disturbances specified in Assumption 2, $\mathbf{B}$ is an $N \times N$ nonstochastic matrix with bounded row and column norms and satisfies diag$(\mathbf{B}) = \mathbf{0}$, and $\mathbf{c}$ is an $NT \times 1$ nonstochastic vector such that $sup_{N,T}(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}|c_{it}|^{2+\delta} < \infty$ for some $\delta > 0$. Then the variance of h is given by $\sigma_h^2 = \sum_{i=1}^{N}\sum_{j=1}^{N} b_{ji}(b_{ij}+b_{ji})tr[\mathbf{\Omega}_{e,i}\mathbf{\Omega}_{e,j}] + \sum_{i=1}^{N}\mathbf{c}_{i.}^{'}\mathbf{\Omega}_{e,i}\mathbf{c}_{i.}$. If $\frac{1}{NT}\sigma_h^2$ is bounded away from zero, we have $\frac{h}{\sigma_h} \xrightarrow{d} N(0,1)$ as $N \to \infty$ and $T/N \to 0$.*

Proof. Let $h_i = \sum_{j=1}^{N} b_{ji}\mathbf{e}_{i.}^{'}\mathbf{e}_{j.} + \mathbf{c}_{i.}^{'}\mathbf{e}_{i.}$, and then $h = \sum_{i=1}^{N} h_i$. Note that h_i, $i = 1,2,...N$, forms a martingale difference array with respect to the $\sigma-$field generated by $\{\mathbf{e}_{1.}, \mathbf{e}_{2.}, ..., \mathbf{e}_{i-1.}\}$, because

$$\begin{aligned} E(h_i|1,2,...,i-1) &= \sum_{j=1}^{N} b_{ji}E(\mathbf{e}_{i.}^{'}\mathbf{e}_{j.}|1,...,i-1) + E(\mathbf{c}_{i.}^{'}\mathbf{e}_{i.}|1,...,i-1) \\ &= \sum_{j=1}^{i-1} b_{ji}E(\mathbf{e}_{i.}^{'})\mathbf{e}_{j} + \sum_{i+1}^{N} b_{ji}E(\mathbf{e}_{i.}^{'}\mathbf{e}_{j.}) + 0 \\ &= 0. \end{aligned}$$

To apply a martingale difference CLT, we only need to show that the following two sufficient conditions hold (see **?**, **?** and **?**): (i) $\frac{1}{\sigma_h^{2+\delta}}\sum_{i=1}^{N} E|h_i|^{2+\delta} \to 0$ for some $\delta > 0$, and (ii) $\frac{1}{\sigma_h^2}\sum_{i=1}^{N} E(h_i^2|1,...,i-1) \xrightarrow{p} 1$.

For (i), let $q = 2+\delta$ and $\frac{1}{p}+\frac{1}{q}=1$, we have

$$\begin{aligned}|h_i|^q &\le |\sum_{j=1}^{N} b_{ji}\mathbf{e}_{i.}^{'}\mathbf{e}_{j.}|^q + |\mathbf{c}_{i.}^{'}\mathbf{e}_{i.}|^q \\ &\le \left(\sum_{j=1}^{N}|b_{ji}|\right)^{\frac{q}{p}}\left(\sum_{j=1}^{N}|b_{ji}||\mathbf{e}_{i.}^{'}\mathbf{e}_{j.}|^q\right) + |\mathbf{c}_{i.}^{'}\mathbf{e}_{i.}|^q,\end{aligned}$$

where the second equality follows by the Holder's inequality, and then

$$\sum_{i=1}^{N} E|h_i|^q \le \sum_{i=1}^{N}\left(\sum_{j=1}^{N}|b_{ji}|\right)^{\frac{q}{p}}\left[\sum_{j=1}^{N}|b_{ji}|E\left(|\mathbf{e}_{i.}^{'}\mathbf{e}_{j.}|^q\right)\right] + \sum_{i=1}^{N} E\left(|\mathbf{c}_{i.}^{'}\mathbf{e}_{i.}|^q\right). \tag{A.27}$$

By the C_r inequality, $E\left(|\mathbf{e}_{i.}^{'}\mathbf{e}_{j.}|^q\right) \le T^{q-1}\sum_{t=1}^{T} E|e_{it}|^q E|e_{jt}|^q = O(T^{q-1})$, where the order follows by the assumption that the third cumulant of e_{it} is absolutely summable and the third central moment of a random variable is the same as the third cumulant. Similarly, $E\left(|\mathbf{c}_{i.}^{'}\mathbf{e}_{i.}|^q\right) \le T^{q-1}\sum_{t=1}^{T}|c_{it}|^q E|e_{it}|^q = O(T^{q-1})$. As a result, $\sum_{i=1}^{N} E|h_i|^{2+\delta} = O(NT^{1+\delta})$, and the assertion in (i) follows as $\sigma_h^{2+\delta} = O\left(N^{1+\frac{\delta}{2}}T^{1+\frac{\delta}{2}}\right)$.

For (ii),

$$\sum_{i=1}^{N} E(h_i^2|1,...,i-1) - \sigma_h^2 = r_1 + 2r_2, \tag{A.28}$$

where

$$\begin{aligned}r_1 &= \sum_{i=1}^{N}\sum_{j=1}^{i-1} b_{ji}(b_{ij}+b_{ji})\left\{E\left(\mathbf{e}_{i}^{'}\mathbf{e}_{j.}\mathbf{e}_{j}^{'}\mathbf{e}_{i.}|1,...,i-1\right) - E\left(\mathbf{e}_{i}^{'}\mathbf{e}_{j.}\mathbf{e}_{j}^{'}\mathbf{e}_{i.}\right)\right\} \\ &= \sum_{i=1}^{N}\sum_{j=1}^{i-1} b_{ji}(b_{ij}+b_{ji})\left\{\sum_{t=1}^{T}\sum_{s=1}^{T} E(e_{it}e_{is})\left[e_{jt}e_{js} - E(e_{jt}e_{js})\right]\right\},\end{aligned}$$

$$\begin{aligned}r_2 &= \sum_{i=1}^{N} E\left\{\left(\sum_{j=1}^{N} b_{ji}\mathbf{e}_{i}^{'}\mathbf{e}_{j.}\right)\left(\mathbf{c}_{i.}^{'}\mathbf{e}_{i.}\right)1,...,i-1\right\} \\ &= \sum_{i=1}^{N}\sum_{j=1}^{i-1}\sum_{t=1}^{T}\sum_{s=1}^{T} b_{ji}c_{is}E(e_{it}e_{is})e_{jt}.\end{aligned}$$

Clearly $E(r_1) = E(r_2) = 0$, and

$$\begin{aligned} Var(r_1) &= \sum_{i=1}^{N}\sum_{j=1}^{i-1}\sum_{l=1}^{N} b_{ji}(b_{ij}+b_{ji})b_{li}(b_{lj}+b_{jl}) \\ &\times \sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{t'=1}^{T}\sum_{s'=1}^{T} E(e_{it}e_{is})E(e_{lt'}e_{ls'})\left[E(e_{jt}e_{js}e_{jt'}e_{js'}) - E(e_{jt}e_{js})E(e_{jt'}e_{js'})\right] \\ &= \sum_{i=1}^{N}\sum_{j=1}^{i-1}\sum_{l=1}^{N} b_{ji}(b_{ij}+b_{ji})b_{li}(b_{lj}+b_{jl}) \\ &\times \sum_{t,s,t',s'=1}^{T} E(e_{it}e_{is})E(e_{lt'}e_{ls'})\left[cum(e_{jt},e_{js},e_{jt'},e_{js'}) + E(e_{jt}e_{jt'})E(e_{js}e_{js'}) + E(e_{jt}e_{js'})E(e_{js}e_{jt'})\right]. \end{aligned}$$

Under Assumption 2, e_{it} is stationary with absolutely summable autocovariance and fourth-order cumulant, and also in light of the uniform boundedness of the row and column norms of $\mathbf{B}$, we have $Var(r_1) = O(NT)$. Next,

$$Var(r_2) = \sum_{i=1}^{N}\sum_{j=1}^{i-1} b_{ji}^2 \sum_{t,s,t',s'=1}^{T} c_{is}c_{is'}E(e_{it}e_{is})E(e_{it'}e_{is'})E(e_{jt}e_{jt'}) = O(NT), \tag{A.29}$$

where we have used the uniform boundedness of c_{is} and the absolute summability of autocovariance of e_{it}. Accordingly,

$$\frac{1}{\sigma_h^2}\sum_{i=1}^{N} E(h_i^2|1,...,i-1) - 1 = \frac{\frac{1}{NT}\left\{\sum_{i=1}^{N} E(h_i^2|1,...,i-1) - \sigma_h^2\right\}}{\frac{1}{NT}\sigma_h^2} = O(\frac{1}{\sqrt{NT}}), \tag{A.30}$$

which proves (ii). □

B Appendix: Proofs of Main Theorems

Proof of Theorem 3

For ease of notation, in this proof we omit the subscript "0" and use $\boldsymbol{\gamma}_i$, $\boldsymbol{\Gamma}$, etc. to denote the true parameters. The key is to derive the distribution of $\frac{1}{\sqrt{NT}}\mathbf{Q}'\mathbf{M}^b\left[(\mathbf{I}_T \otimes \boldsymbol{\Gamma})\mathbf{f} + \mathbf{e}\right]$. According to Lemma 6, we only need to find out the distribution of $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{X}_{i.}'\bar{\mathbf{M}}(\mathbf{F}\boldsymbol{\gamma}_i + \mathbf{e}_{i.})$, and then the distribution of $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{l=1}^{N} w_{il}^s\mathbf{X}_{i.}'\bar{\mathbf{M}}(\mathbf{F}\boldsymbol{\gamma}_i + \mathbf{e}_{i.})$, $s = 1, 2, ...,$ readily follows.

We first consider $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{X}_{i.}'\bar{\mathbf{M}}\mathbf{F}\boldsymbol{\gamma}_i$. Under Assumption 3, $\boldsymbol{\gamma}_i = \boldsymbol{\gamma} + \boldsymbol{\eta}_i$, and note that $\frac{1}{N}\sum_{i=1}^{N}\mathbf{X}_{i.}'\bar{\mathbf{M}}\mathbf{F}\boldsymbol{\gamma} = \bar{\mathbf{X}}_{i.}'\bar{\mathbf{M}}\mathbf{F}\boldsymbol{\gamma} = \mathbf{0}$, we have

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{X}_{i.}'\bar{\mathbf{M}}\mathbf{F}\boldsymbol{\gamma}_i = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{X}_{i.}'\bar{\mathbf{M}}\mathbf{F}\boldsymbol{\eta}_i, \tag{B.1}$$

Substituting (A.10) into this expression yields

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{X}_{i.}^{'}\bar{\mathbf{M}}\mathbf{F}\boldsymbol{\gamma}_i \quad = \quad \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left\{-\mathbf{A}_i^{'}(\bar{\mathbf{C}}\bar{\mathbf{C}}^{'})^{-1}\bar{\mathbf{C}}\bar{\boldsymbol{\epsilon}}^{'}\bar{\mathbf{M}}\bar{\boldsymbol{\epsilon}}\bar{\mathbf{C}}^{'}(\bar{\mathbf{C}}\bar{\mathbf{C}}^{'})^{-1} - \mathbf{V}_{i.}^{'}\bar{\mathbf{M}}\bar{\boldsymbol{\epsilon}}\bar{\mathbf{C}}^{'}(\bar{\mathbf{C}}\bar{\mathbf{C}}^{'})^{-1}\right\}\boldsymbol{\eta}_i.$$

Since $\frac{\bar{\boldsymbol{\epsilon}}^{'}\bar{\mathbf{M}}\bar{\boldsymbol{\epsilon}}}{T} = O_p(\frac{1}{N})$, $\frac{\mathbf{V}_{i.}^{'}\bar{\mathbf{M}}\bar{\boldsymbol{\epsilon}}}{T} = O_p(\frac{1}{N}) + O_p(\frac{1}{\sqrt{NT}})$ (see (A.11) and (A.12)), and $\bar{\mathbf{C}}^{'}(\bar{\mathbf{C}}\bar{\mathbf{C}}^{'})^{-1}$ and $\boldsymbol{\eta}_i$ are uniformly bounded, it follows that

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{X}_{i.}^{'}\bar{\mathbf{M}}\mathbf{F}\boldsymbol{\eta}_i = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\left\{-\mathbf{V}_{i.}^{'}\bar{\mathbf{M}}\bar{\boldsymbol{\epsilon}}\bar{\mathbf{C}}^{'}(\bar{\mathbf{C}}\bar{\mathbf{C}}^{'})^{-1}\right\}\boldsymbol{\eta}_i + O_p(\sqrt{\frac{T}{N}}). \tag{B.2}$$

Further using $\frac{\mathbf{V}_{i.}^{'}\bar{\mathbf{M}}\bar{\boldsymbol{\epsilon}}}{T} = \frac{\mathbf{V}_{i.}^{'}\bar{\boldsymbol{\epsilon}}}{T} - \left(\frac{\mathbf{V}_{i.}^{'}\bar{\mathbf{Z}}}{T}\right)\left(\frac{\bar{\mathbf{Z}}^{'}\bar{\mathbf{Z}}}{T}\right)^{-1}\left(\frac{\bar{\mathbf{Z}}^{'}\bar{\boldsymbol{\epsilon}}}{T}\right)$ and noticing that its probability order is dominated by the first term on the right hand side, we obtain

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{X}_{i.}^{'}\bar{\mathbf{M}}\mathbf{F}\boldsymbol{\eta}_i = -\frac{1}{N}\sum_{i=1}^{N}\frac{\sqrt{N}\mathbf{V}_{i.}^{'}\bar{\boldsymbol{\epsilon}}}{\sqrt{T}}\bar{\mathbf{C}}^{'}(\bar{\mathbf{C}}\bar{\mathbf{C}}^{'})^{-1}\boldsymbol{\eta}_i + O_p(\sqrt{\frac{T}{N}}). \tag{B.3}$$

Now that it is easy to see $\bar{\mathbf{C}}^{'}(\bar{\mathbf{C}}\bar{\mathbf{C}}^{'})^{-1} - \bar{\mathbf{C}}_{-i}^{'}(\bar{\mathbf{C}}_{-i}\bar{\mathbf{C}}_{-i}^{'})^{-1} = O_p(\frac{1}{N})$, where $\bar{\mathbf{C}}_{-i}$ is constructed in a similar way as $\bar{\mathbf{C}}$ but excluding $\boldsymbol{\Phi}_i$, and by a weak law of large numbers for martingale difference triangular array we can establish that

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\sqrt{N}\mathbf{V}_{i.}^{'}\bar{\boldsymbol{\epsilon}}}{\sqrt{T}}\bar{\mathbf{C}}_{-i}^{'}(\bar{\mathbf{C}}_{-i}\bar{\mathbf{C}}_{-i}^{'})^{-1}\boldsymbol{\eta}_i \xrightarrow{p} \mathbf{0} \text{ as } N \to \infty \text{ and } T/N \to 0, \tag{B.4}$$

because $\boldsymbol{\eta}_i$ are i.i.d. with zero mean and are independent of all the stochastic quantities in the model, and $E\|\left(\sqrt{N}\mathbf{V}_{i.}^{'}\bar{\boldsymbol{\epsilon}}/\sqrt{T}\right)\bar{\mathbf{C}}_{-i}^{'}(\bar{\mathbf{C}}_{-i}\bar{\mathbf{C}}_{-i}^{'})^{-1}\boldsymbol{\eta}_i\|^2 < \infty$. Hence, it follows that

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{X}_{i.}^{'}\bar{\mathbf{M}}\mathbf{F}\boldsymbol{\gamma}_i \xrightarrow{p} \mathbf{0} \text{ as } N \to \infty \text{ and } T/N \to 0. \tag{B.5}$$

Next, we turn to analyze the distribution of $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{X}_{i.}^{'}\bar{\mathbf{M}}\mathbf{e}_{i.}$. From the proof of Lemma 5(c), we see that

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{X}_{i.}^{'}\bar{\mathbf{M}}\mathbf{e}_{i.} = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{X}_{i.}^{'}\mathbf{M}_f\mathbf{e}_{i.} + \frac{1}{N}\sum_{i=1}^{N}\frac{\mathbf{X}_{i.}^{'}\boldsymbol{\Pi}}{T}\left(\frac{\boldsymbol{\Pi}^{'}\boldsymbol{\Pi}}{T}\right)^{-1}\left(\frac{\sqrt{N}\bar{\boldsymbol{\epsilon}}^{'}\mathbf{e}_{i.}}{\sqrt{T}}\right) + O_p(\sqrt{\frac{T}{N}}). \tag{B.6}$$

The first term on the right-hand side of (B.6) follows a distribution

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\mathbf{X}_{i.}^{'}\mathbf{M}_f\mathbf{e}_{i.}}{\sqrt{T}} \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Omega}_{XMe}), \tag{B.7}$$

where $\mathbf{\Omega}_{XMe} = \lim_{N\to\infty}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{S}_{iXMe}\right)$,$\mathbf{S}_{iXMe} = \underset{T\to\infty}{plim}\left[(1/T)\mathbf{X}_{i.}^{'}\mathbf{M}_f E(\mathbf{e}_{i.}\mathbf{e}_{i.}^{'})\mathbf{M}_f\mathbf{X}_{i.}\right]$, because

$$\frac{\mathbf{X}_{i.}^{'}\mathbf{M}_f\mathbf{e}_{i.}}{\sqrt{T}} = \frac{\mathbf{V}_{i.}^{'}\mathbf{e}_{i.}}{\sqrt{T}} - \frac{1}{\sqrt{T}}\left(\frac{\mathbf{V}_{i.}^{'}\mathbf{F}}{\sqrt{T}}\right)\left(\frac{\mathbf{F}^{'}\mathbf{F}}{T}\right)^{-1}\left(\frac{\mathbf{F}^{'}\mathbf{e}_{i.}}{\sqrt{T}}\right) = \frac{\mathbf{V}_{i.}^{'}\mathbf{e}_{i.}}{\sqrt{T}} + O_p(\frac{1}{\sqrt{T}}), \tag{B.8}$$

and $\frac{\mathbf{V}_{i.}^{'}\mathbf{e}_{i.}}{\sqrt{T}} = O_p(1)$ under Assumption 2. For the second term on the right-hand side of (B.6), we have

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\mathbf{X}_{i.}^{\prime}\mathbf{\Pi}}{T}\left(\frac{\mathbf{\Pi}^{\prime}\mathbf{\Pi}}{T}\right)^{-1}\left(\frac{\sqrt{N}\bar{\boldsymbol{\epsilon}}^{'}\mathbf{e}_{i.}}{\sqrt{T}}\right) = \frac{1}{N}\sum_{i=1}^{N}\frac{\mathbf{X}_{i.}^{\prime}\mathbf{\Pi}}{T}\left(\frac{\mathbf{\Pi}^{\prime}\mathbf{\Pi}}{T}\right)^{-1}\left(\frac{\sqrt{N}\bar{\boldsymbol{\epsilon}}_{-i}^{'}\mathbf{e}_{i.}}{\sqrt{T}}\right) + O_p(\sqrt{\frac{T}{N}}), \tag{B.9}$$

where we used that $\frac{\bar{\boldsymbol{\epsilon}}^{'}\mathbf{e}_{i.}}{T} - \frac{\bar{\boldsymbol{\epsilon}}_{-i}^{'}\mathbf{e}_{i.}}{T} = O_p(\frac{1}{N})$. Applying a weak law of large numbers for a martingale difference triangular array with finite second moment, we establish

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\mathbf{X}_{i.}^{\prime}\mathbf{\Pi}}{T}\left(\frac{\mathbf{\Pi}^{\prime}\mathbf{\Pi}}{T}\right)^{-1}\left(\frac{\sqrt{N}\bar{\boldsymbol{\epsilon}}_{-i}^{'}\mathbf{e}_{i.}}{\sqrt{T}}\right) \xrightarrow{p} \mathbf{0} \text{ as } N\to\infty. \tag{B.10}$$

Thus, as $(N,T)\xrightarrow{j}\infty$ and $T/N\to 0$,

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\mathbf{X}_{i.}^{\prime}\mathbf{\Pi}}{T}\left(\frac{\mathbf{\Pi}^{\prime}\mathbf{\Pi}}{T}\right)^{-1}\left(\frac{\sqrt{N}\bar{\boldsymbol{\epsilon}}^{'}\mathbf{e}_{i.}}{\sqrt{T}}\right) \xrightarrow{p} \mathbf{0}, \tag{B.11}$$

and it follows that

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{X}_{i.}^{'}\bar{\mathbf{M}}\mathbf{e}_{i.} \xrightarrow{d} N(\mathbf{0},\mathbf{\Omega}_{XMe}). \tag{B.12}$$

As a result, as $(N,T)\xrightarrow{j}\infty$ and $T/N\to 0$, we have

$$\sqrt{NT}(\hat{\boldsymbol{\delta}}_{2sls} - \boldsymbol{\delta}_0) \xrightarrow{d} N(\mathbf{0},\mathbf{\Sigma}_{2sls}), \tag{B.13}$$

where $\mathbf{\Sigma}_{2sls}$ is given by (40).

Proof of Theorem 4

As

$$\sqrt{NT}(\hat{\boldsymbol{\delta}}_{b2sls} - \boldsymbol{\delta}_0) = \left[\frac{1}{NT}\hat{\mathbf{Q}}^{*'}\mathbf{L}\right]^{-1}\frac{1}{\sqrt{NT}}\hat{\mathbf{Q}}^{*'}\left[(\mathbf{I}_T\otimes\mathbf{\Gamma}_0)\mathbf{f} + \mathbf{e}\right], \tag{B.14}$$

to establish the asymptotic distribution, it suffices to show that

$$\underset{N\to\infty}{plim}\frac{1}{NT}\hat{\mathbf{Q}}^{*'}\mathbf{L} = \underset{N\to\infty}{plim}\frac{1}{NT}\mathbf{L}_0^{\prime}\mathbf{M}_f^b\mathbf{L}_0, \tag{B.15}$$

and

$$\frac{1}{\sqrt{NT}}\hat{\mathbf{Q}}^{*\prime}\left[(\mathbf{I}_T\otimes\mathbf{\Gamma}_0)\mathbf{f}+\mathbf{e}\right]\xrightarrow{d}N(\mathbf{0},\mathbf{\Omega}_{LMe}). \tag{B.16}$$

Substituting

$$\mathbf{Y}=[\mathbf{I}_T\otimes\mathbf{S}_0^{-1}][\mathbf{X}\boldsymbol{\beta}_0+(\mathbf{I}_T\otimes\mathbf{\Gamma}_0)\mathbf{f}+\mathbf{e}] \tag{B.17}$$

into the definition of $\mathbf{L}$ yields

$$\mathbf{L}=\mathbf{L}_0+[(\mathbf{I}_T\otimes\mathbf{\Gamma}_0)\mathbf{f}+\mathbf{e},\mathbf{0}], \tag{B.18}$$

and it follows that

$$\frac{1}{NT}\hat{\mathbf{Q}}^{*\prime}\mathbf{L}=\frac{1}{NT}\left[(\mathbf{I}_T\otimes\mathbf{G}(\hat{\rho}))\,\mathbf{X}\hat{\boldsymbol{\beta}},\mathbf{X}\right]^{\prime}\mathbf{M}^b\mathbf{L}_0+\frac{1}{NT}\left[(\mathbf{I}_T\otimes\mathbf{G}(\hat{\rho}))\,\mathbf{X}\hat{\boldsymbol{\beta}},\mathbf{X}\right]^{\prime}\mathbf{M}^b\left[(\mathbf{I}_T\otimes\mathbf{\Gamma}_0)\mathbf{f}+\mathbf{e},\mathbf{0}\right]. \tag{B.19}$$

Using the first-order Taylor expansion of $\mathbf{G}(\hat{\rho})$,

$$\mathbf{W}(\mathbf{I}_N-\hat{\rho}\mathbf{W})^{-1}=\mathbf{G}_0+\mathbf{W}(\mathbf{I}_N-\hat{\rho}\mathbf{W})^{-1}\mathbf{G}_0(\hat{\rho}-\rho_0), \tag{B.20}$$

also $\hat{\rho}=\rho+o_p(1)$, $\hat{\boldsymbol{\beta}}=\boldsymbol{\beta}_0+o_p(1)$, and applying Lemma 6, we obtain

$$\frac{1}{NT}\left[(\mathbf{I}_T\otimes\mathbf{G}(\hat{\rho}))\,\mathbf{X}\hat{\boldsymbol{\beta}},\mathbf{X}\right]^{\prime}\mathbf{M}^b\mathbf{L}_0=\frac{1}{NT}\mathbf{L}_0^{\prime}\mathbf{M}_f^b\mathbf{L}_0+o_p(1), \tag{B.21}$$

$$\frac{1}{NT}\left[(\mathbf{I}_T\otimes\mathbf{G}(\hat{\rho}))\,\mathbf{X}\hat{\boldsymbol{\beta}},\mathbf{X}\right]^{\prime}\mathbf{M}^b\left[(\mathbf{I}_T\otimes\mathbf{\Gamma}_0)\mathbf{f}+\mathbf{e},\mathbf{0}\right]=o_p(1). \tag{B.22}$$

Thus, the claim in (B.15) is established.

The proof of(B.16) is similar to that of (3).

Now to examine if $\hat{\mathbf{Q}}^*$ is the best IV, we need to compare the asymptotic variances $\mathbf{\Sigma}_{b2sls}$ with $\mathbf{\Sigma}_{2sls}$. Notice that

$$\mathbf{L}_0^{\prime}\mathbf{P}_{Q,f}\mathbf{L}_0=\mathbf{L}_0^{\prime}\mathbf{M}_f^b\mathbf{Q}(\mathbf{Q}^{\prime}\mathbf{M}_f^b\mathbf{Q})^{-1}\mathbf{Q}^{\prime}\mathbf{M}_f^b\mathbf{L}_0\leq\mathbf{L}_0^{\prime}\mathbf{M}_f^b\mathbf{L}_0, \tag{B.23}$$

hence $\mathbf{\Psi}_{LPL}\leq\mathbf{\Psi}_{LML}$. If the disturbances $\{e_{it}\}$ are independently and identically distributed with mean zero and variance σ_e^2, then $\mathbf{\Sigma}_{b2sls}=\sigma_e^2\mathbf{\Psi}_{LML}^{-1}\leq\sigma_e^2\mathbf{\Psi}_{LPL}^{-1}=\mathbf{\Sigma}_{2sls}$. In general, however, we cannot conclude that $\hat{\mathbf{Q}}^*$ is optimal because $\mathbf{\Omega}_{LPe}$ may be greater than $\mathbf{\Omega}_{LMe}$ as they depend on the unknown $\mathbf{\Omega}_{e,i\cdot}$.

Proof of Theorem 5

Consistency

Under the identification conditions for this model, it suffices to show that $\frac{1}{NT}\mathbf{A}_{NT}^w\mathbf{g}_{NT}(\boldsymbol{\delta})$ converges to its mean uniformly in $\boldsymbol{\delta}\in\Delta_{sp}$ and the limit equals zero at $\boldsymbol{\delta}_0$. Notice that

$$\boldsymbol{\xi}(\boldsymbol{\delta})\;=\;[\mathbf{I}_T\otimes\mathbf{S}(\rho)][\mathbf{I}_T\otimes\mathbf{S}_0^{-1}][\mathbf{X}\boldsymbol{\beta}_0+(\mathbf{I}_T\otimes\mathbf{\Gamma}_0)\mathbf{f}+\mathbf{e}]-\mathbf{X}\boldsymbol{\beta}.$$

Since $\mathbf{S}(\rho)\mathbf{S}_0^{-1} = [\mathbf{S}_0 + (\rho_0 - \rho)\mathbf{W}]\mathbf{S}_0^{-1} = \mathbf{I}_N + (\rho_0 - \rho)\mathbf{G}_0$, where $\mathbf{G}_0 = \mathbf{W}\mathbf{S}_0^{-1}$, we then obtain

$$\begin{aligned}\boldsymbol{\xi}(\boldsymbol{\delta}) &= [\mathbf{I}_{NT} + (\rho_0 - \rho)(\mathbf{I}_T \otimes \mathbf{G}_0)][\mathbf{X}\boldsymbol{\beta}_0 + (\mathbf{I}_T \otimes \boldsymbol{\Gamma}_0)\mathbf{f}]-\mathbf{X}\boldsymbol{\beta} + \mathbf{I}_T \otimes [\mathbf{S}(\rho)\mathbf{S}_0^{-1}]\mathbf{e} \\ &= \mathbf{X}(\boldsymbol{\beta}_0-\boldsymbol{\beta}) + (\rho_0 - \rho)[\mathbf{I}_T \otimes \mathbf{G}_0][\mathbf{X}\boldsymbol{\beta}_0 + (\mathbf{I}_T \otimes \boldsymbol{\Gamma}_0)\mathbf{f}] + (\mathbf{I}_T \otimes \boldsymbol{\Gamma}_0)\mathbf{f} + \mathbf{I}_T \otimes [\mathbf{S}(\rho)\mathbf{S}_0^{-1}]\mathbf{e} \\ &= \mathbf{d}(\boldsymbol{\delta}) + \mathbf{r}_\xi(\boldsymbol{\delta}). \qquad \text{(B.24)}\end{aligned}$$

where $\mathbf{d}(\boldsymbol{\delta}) = \mathbf{X}(\boldsymbol{\beta}_0-\boldsymbol{\beta}) + \mathbf{J}(\rho_0 - \rho)$ with

$$\mathbf{J} = [\mathbf{I}_T \otimes \mathbf{G}_0][\mathbf{X}\boldsymbol{\beta}_0 + (\mathbf{I}_T \otimes \boldsymbol{\Gamma}_0)\mathbf{f}], \qquad \text{(B.25)}$$

and $\mathbf{r}_\xi(\boldsymbol{\delta}) = (\mathbf{I}_T \otimes \boldsymbol{\Gamma}_0)\mathbf{f} + \mathbf{I}_T \otimes [\mathbf{S}(\rho)\mathbf{S}_0^{-1}]\mathbf{e}$.

By definition,

$$\frac{1}{NT}\mathbf{A}_{NT}^w \mathbf{g}_{NT}(\boldsymbol{\delta}) = \frac{1}{NT}\boldsymbol{\xi}'(\boldsymbol{\delta})\left(\sum_{l=1}^r \mathbf{a}^{(l)}\mathbf{M}^b\mathbf{P}_l^b\mathbf{M}^b\right)\boldsymbol{\xi}(\boldsymbol{\delta}) + \frac{1}{NT}\mathbf{A}^{(Q)}\mathbf{Q}'\mathbf{M}^b\boldsymbol{\xi}(\boldsymbol{\delta}). \qquad \text{(B.26)}$$

Expanding the first term of (B.26), we get

$$\frac{1}{NT}\boldsymbol{\xi}'(\boldsymbol{\delta})\left(\sum_{l=1}^r \mathbf{a}^{(l)}\mathbf{M}^b\mathbf{P}_l^b\mathbf{M}^b\right)\boldsymbol{\xi}(\boldsymbol{\delta}) = \mathbf{A}_1 + 2\mathbf{A}_2 + \mathbf{A}_3, \qquad \text{(B.27)}$$

where

$$\mathbf{A}_1 = \frac{1}{NT}\mathbf{d}'(\boldsymbol{\delta})\left(\sum_{l=1}^r \mathbf{a}^{(l)}\mathbf{M}^b\mathbf{P}_l^b\mathbf{M}^b\right)\mathbf{d}(\boldsymbol{\delta}), \qquad \text{(B.28)}$$

$$\mathbf{A}_2 = \frac{1}{NT}\mathbf{d}'(\boldsymbol{\delta})\left(\sum_{l=1}^r \mathbf{a}^{(l)}\mathbf{M}^b\mathbf{P}_l^b\mathbf{M}^b\right)\mathbf{r}_\xi(\boldsymbol{\delta}), \qquad \text{(B.29)}$$

$$\mathbf{A}_3 = \frac{1}{NT}\mathbf{r}_\xi(\boldsymbol{\delta})'\left(\sum_{l=1}^r \mathbf{a}^{(l)}\mathbf{M}^b\mathbf{P}_l^b\mathbf{M}^b\right)\mathbf{r}_\xi(\boldsymbol{\delta}).$$

Because $\mathbf{S}(\rho)\mathbf{S}_0^{-1}$ has bounded row and column norms, $\mathbf{M}^b\mathbf{P}_l^b\mathbf{M}^b = \bar{\mathbf{M}} \otimes \mathbf{P}_l$, and the products of $\mathbf{S}(\rho)\mathbf{S}_0^{-1}$, $\mathbf{P}_l$, and $[\mathbf{S}(\rho)\mathbf{S}_0^{-1}]'$ also have bounded row and column norms, applying Lemma 6 and 7 we obtain that $\mathbf{A}_1$, $\mathbf{A}_2$ and $\mathbf{A}_3$ converge uniformly to their means respectively. The second term in (B.26) converges uniformly to zero. Hence, we have established the uniform convergence of $\frac{1}{NT}\mathbf{A}_{NT}^w\mathbf{g}_{NT}(\boldsymbol{\delta})$. Its limit equals zero at true value $\boldsymbol{\delta}_0$, which can be verified by noticing that $\boldsymbol{\xi}(\boldsymbol{\delta}_0) = (\mathbf{I}_T \otimes \boldsymbol{\Gamma}_0)\mathbf{f} + \mathbf{e}$ and $E\left[\mathbf{e}'\mathbf{M}_f^b\mathbf{P}_l^b\mathbf{M}_f^b\mathbf{e}\right] = tr\left\{E(\mathbf{M}_f \otimes \mathbf{P}_l)E(\mathbf{e}\mathbf{e}')\right\} = 0$.

Asymptotic distribution

We omit the subscript and let $\hat{\boldsymbol{\delta}}$ denote the GMM estimator in this proof. By a mean value expansion of $\frac{\partial \mathbf{g}'_{NT}(\hat{\boldsymbol{\delta}})}{\partial} \mathbf{A}^{w\prime}_{NT} \mathbf{A}^{w}_{NT} \mathbf{g}_{NT}(\hat{\boldsymbol{\delta}}) = 0$ around $\boldsymbol{\delta}_0$,

$$\sqrt{NT}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0) = -\left[\frac{1}{NT} \frac{\partial \mathbf{g}'_{NT}(\hat{\boldsymbol{\delta}})}{\partial \boldsymbol{\delta}} \mathbf{A}^{w\prime}_{NT} \mathbf{A}^{w}_{NT} \frac{1}{NT} \frac{\partial \mathbf{g}_{NT}(\ddot{\boldsymbol{\delta}})}{\partial \boldsymbol{\delta}'} \right]^{-1} \frac{1}{NT} \frac{\partial \mathbf{g}'_{NT}(\hat{\boldsymbol{\delta}})}{\partial \boldsymbol{\delta}} \mathbf{A}^{w\prime}_{NT} \frac{1}{\sqrt{NT}} \mathbf{A}^{w}_{NT} \mathbf{g}_{NT}(\boldsymbol{\delta}_0), \tag{B.30}$$

where $\ddot{\boldsymbol{\delta}}$ lies between $\hat{\boldsymbol{\delta}}$ and $\boldsymbol{\delta}_0$. For any $\boldsymbol{\delta}$ in the parameter space, we have

$$\frac{\partial \boldsymbol{\xi}(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}'} = -\left[(\mathbf{I}_T \otimes \mathbf{W})\mathbf{Y}, \mathbf{X} \right], \tag{B.31}$$

therefore

$$\frac{\partial \mathbf{g}(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}'} = -\left[\mathbf{M}^b \mathbf{P}^{sb}_1 \mathbf{M}^b \boldsymbol{\xi}(\boldsymbol{\delta}), ..., \mathbf{M}^b \mathbf{P}^{sb}_r \mathbf{M}^b \boldsymbol{\xi}(\boldsymbol{\delta}), \mathbf{M}^b \mathbf{Q} \right]' \left[(\mathbf{I}_T \otimes \mathbf{W})\mathbf{Y}, \mathbf{X} \right], \tag{B.32}$$

where $\mathbf{P}^{sb}_l = \mathbf{I}_T \otimes \mathbf{P}^s_l$ and $\mathbf{P}^s_l = \mathbf{P}_l + \mathbf{P}'_l$, for $l = 1, ..., r$. Since

$$\begin{aligned} \frac{1}{NT} \boldsymbol{\xi}'(\boldsymbol{\delta}) \mathbf{M}^b \mathbf{P}^{sb}_l \mathbf{M}^b (\mathbf{I}_T \otimes \mathbf{W}) \mathbf{Y} &= \frac{1}{NT} \boldsymbol{\xi}'(\boldsymbol{\delta}) [\bar{\mathbf{M}} \otimes \mathbf{P}^s_l \mathbf{G}_0] \mathbf{X} \boldsymbol{\beta}_0 \\ &+ \frac{1}{NT} \boldsymbol{\xi}'(\boldsymbol{\delta}) \left\{ [\bar{\mathbf{M}} \otimes \mathbf{P}^s_l \mathbf{G}_0] \mathbf{e} + [\bar{\mathbf{M}} \otimes \mathbf{P}^s_l \mathbf{G}_0 \boldsymbol{\Gamma}_0] \mathbf{f} \right\}, \end{aligned} \tag{B.33}$$

by Lemma 6 and 7, at true value $\boldsymbol{\delta}_0$ the above equation (B.33) can be rewritten as

$$\frac{1}{NT} \left[(\mathbf{I}_T \otimes \boldsymbol{\Gamma}_0) \mathbf{f} + \mathbf{e} \right]' \mathbf{M}^b \mathbf{P}^{sb}_l \mathbf{M}^b (\mathbf{I}_T \otimes \mathbf{W}) \mathbf{Y} = \frac{1}{N} \sum_{i=1}^{N} \tilde{g}^s_{ii,l} \sigma^2_i + o_p(1), \tag{B.34}$$

where $\tilde{g}^s_{ii,l}$ is the $(i,i)th$ element of matrix $\tilde{\boldsymbol{G}}_l(\rho_0) = \mathbf{P}^s_l \mathbf{G}_0$, and

$$\frac{1}{NT} \mathbf{e}' \mathbf{M}^b \mathbf{P}^s_l \mathbf{M}^b \mathbf{X} = o_p(1). \tag{B.35}$$

In addition, we have

$$\frac{1}{NT} \mathbf{Q}' \mathbf{M}^b (\mathbf{I}_T \otimes \mathbf{W}) \mathbf{Y} = \frac{1}{NT} \mathbf{Q}' [\mathbf{M}_f \otimes \mathbf{G}_0] \mathbf{X} \boldsymbol{\beta}_0 + o_p(1). \tag{B.36}$$

Consequently,

$$\frac{1}{NT} \frac{\partial \mathbf{g}'_{NT}(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}} = -\mathbf{D} + o_p(1), \tag{B.37}$$

where $\mathbf{D}$ is given by (60).

Finally, we apply a triangular array CLT for linear and quadratic forms (Lemma 9) to establish

$$\frac{1}{\sqrt{NT}}\mathbf{A}_{NT}^{w}\mathbf{g}_{NT}(\boldsymbol{\delta}_0) = \frac{1}{\sqrt{NT}}\left\{\mathbf{r}_{\xi}^{'}(\boldsymbol{\delta}_0)(\sum_{l=1}^{r}\mathbf{a}^{(l)}\mathbf{M}^{b}\mathbf{P}_{l}^{b}\mathbf{M}^{b})\mathbf{r}_{\xi}(\boldsymbol{\delta}_0) + \mathbf{A}^{(Q)}\mathbf{Q}'\mathbf{M}^{b}\mathbf{r}_{\xi}(\boldsymbol{\delta}_0)\right\}$$
$$\xrightarrow{d} N(\mathbf{0}, \underset{N,T\to\infty}{plim}\mathbf{A}_{NT}^{w'}\boldsymbol{\Sigma}_{g}\mathbf{A}_{NT}^{w}), \quad \text{(B.38)}$$

where $\boldsymbol{\Sigma}_g$ is the variance matrix of $\frac{1}{\sqrt{NT}}\mathbf{g}_{NT}(\boldsymbol{\delta}_0)$, which is given by (61).

C Appendix: Derivation of Identification Conditions

Model (19): without exogenous variables $\mathbf{x}_{it}$

Given $Q_{NT}(\boldsymbol{\psi})$ defined in (23), the first derivatives are

$$\frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial d} = \frac{1}{N}tr\left[(\mathbf{I}_N - d\mathbf{G}_0)^{-1}\mathbf{G}_0\right] - (1-\vartheta)\frac{tr(\mathbf{G}_0)}{N} + (1-\vartheta)d\frac{tr(\mathbf{G}_0\mathbf{G}_0^{'})}{N} + \frac{\sigma_0^2(1-\vartheta)}{NT}E_0\sum_{t=1}^{T}\left[d(\mathbf{G}_0\boldsymbol{\Gamma}_0\mathbf{f}_t)^{'}\mathbf{G}_0\boldsymbol{\Gamma}_0\mathbf{f}_t + (\mathbf{G}_0\boldsymbol{\Gamma}_0\mathbf{f}_t)^{'}\mathbf{F}_t\boldsymbol{\zeta}\right],$$

$$\frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial \boldsymbol{\zeta}_i} = \frac{\sigma_0^2(1-\vartheta)}{NT}E_0\sum_{t=1}^{T}\left[d\mathbf{f}_t\mathbf{g}_{0,i}^{'}\boldsymbol{\Gamma}_0\mathbf{f}_t + \mathbf{f}_t\mathbf{f}_t^{'}\boldsymbol{\zeta}_i\right], \text{ for } i = 1, 2, ..., N,$$

$$\frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial \vartheta} = \frac{1}{2}\frac{\vartheta}{1-\vartheta} - \frac{1}{2} + \frac{1}{N}dtr(\mathbf{G}_0) - \frac{1}{2}d^2\frac{tr(\mathbf{G}_0\mathbf{G}_0^{'})}{N} - \frac{1}{2}\sigma_0^2\left(d, \boldsymbol{\zeta}^{'}\right)\mathbf{H}_f(\rho_0, \boldsymbol{\gamma}_0^{'})\left(d, \boldsymbol{\zeta}^{'}\right)^{'},$$

where $\mathbf{g}_{0,i}^{'}$ is the i-th row of $\mathbf{G}_0$, and the second derivatives are given by

$$\boldsymbol{\Lambda}_{f,NT}(\boldsymbol{\psi}) \equiv \frac{\partial^2 Q_{NT}(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}\boldsymbol{\psi}^{'}} = \begin{bmatrix} \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial^2 d} & \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial d\partial\boldsymbol{\zeta}^{'}} & \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial d\partial\vartheta} \\ . & \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial\boldsymbol{\zeta}\partial\boldsymbol{\zeta}^{'}} & \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial\boldsymbol{\zeta}\partial\vartheta} \\ . & . & \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial^2\vartheta} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Lambda}_{f,11} & \boldsymbol{\Lambda}_{f,12} & \boldsymbol{\Lambda}_{f,13} \\ . & \boldsymbol{\Lambda}_{f,22} & \boldsymbol{\Lambda}_{f,23} \\ . & . & \boldsymbol{\Lambda}_{f,33} \end{bmatrix} \quad \text{(C.1)}$$

where

$$
\begin{aligned}
\boldsymbol{\Lambda}_{f,11} &= \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial^2 d} = \frac{1}{N} tr\left\{(\mathbf{I}_N - d\mathbf{G}_0)^{-1}\mathbf{G}_0(\mathbf{I}_N - d\mathbf{G}_0)^{-1}\mathbf{G}_0\right\} + (1-\vartheta)\frac{tr(\mathbf{G}_0\mathbf{G}_0^{'})}{N} + \frac{\sigma_0^2(1-\vartheta)}{NT}E_0\sum_{t=1}^{T}\left[(\mathbf{G}_0\boldsymbol{\Gamma}_0\mathbf{f}_t)^{'}\mathbf{G}_0\boldsymbol{\Gamma}_0\mathbf{f}_t\right], \\
\boldsymbol{\Lambda}_{f,12} &= \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial d \partial \boldsymbol{\zeta}^{'}} = \left\{\frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial d \partial \boldsymbol{\zeta}_i^{'}}\right\} = \left\{\frac{\sigma_0^2(1-\vartheta)}{NT}E_0\sum_{t=1}^{T}\mathbf{g}_{0,i}^{'}\boldsymbol{\Gamma}_0\mathbf{f}_t\mathbf{f}_t^{'}\right\}, \\
\boldsymbol{\Lambda}_{f,13} &= \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial d \partial \vartheta} = \frac{tr(\mathbf{G}_0)}{N} - d\frac{tr(\mathbf{G}_0\mathbf{G}_0^{'})}{N} - \frac{\sigma_0^2}{NT}E_0\sum_{t=1}^{T}\left[d(\mathbf{G}_0\boldsymbol{\Gamma}_0\mathbf{f}_t)^{'}\mathbf{G}_0\boldsymbol{\Gamma}_0\mathbf{f}_t + (\mathbf{G}_0\boldsymbol{\Gamma}_0\mathbf{f}_t)^{'}\mathbf{F}_t\boldsymbol{\zeta}\right], \\
\boldsymbol{\Lambda}_{f,22} &= \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial \boldsymbol{\zeta}\partial \boldsymbol{\zeta}^{'}} = \left\{\frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial \boldsymbol{\zeta}_i \partial \boldsymbol{\zeta}_j^{'}}\right\} = \left\{\frac{\sigma_0^2(1-\vartheta)}{NT}E_0\sum_{t=1}^{T}\mathbf{f}_t\mathbf{f}_t^{'},\ \text{if } i=j;\ \text{and } \mathbf{0},\ \text{if } i\neq j\right\}, \\
\boldsymbol{\Lambda}_{f,23} &= \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial \boldsymbol{\zeta}\partial \vartheta} = \left\{\frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial \boldsymbol{\zeta}_i \partial \vartheta}\right\} = \left\{-\frac{\sigma_0^2}{NT}E_0\sum_{t=1}^{T}\left[d\mathbf{f}_t\mathbf{g}_{0,i}^{'}\boldsymbol{\Gamma}_0\mathbf{f}_t + \mathbf{f}_t\mathbf{f}_t^{'}\boldsymbol{\zeta}_i\right]\right\}, \\
\boldsymbol{\Lambda}_{f,33} &= \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial^2 \vartheta} = \frac{1}{2(1-\vartheta)^2}.
\end{aligned}
$$

At $\boldsymbol{\psi} = \mathbf{0}$, we have

$$
\boldsymbol{\Lambda}_{f,NT}(\mathbf{0}) = \begin{bmatrix}
\frac{tr(\mathbf{G}_0^2+\mathbf{G}_0\mathbf{G}_0^{'})}{N} + \frac{\sigma_0^2}{NT}E_0\sum_{t=1}^{T}\left[(\mathbf{G}_0\boldsymbol{\Gamma}_0\mathbf{f}_t)^{'}\mathbf{G}_0\boldsymbol{\Gamma}_0\mathbf{f}_t\right] & \frac{\sigma_0^2}{NT}E_0\sum_{t=1}^{T}\mathbf{g}_{0,1}^{'}\boldsymbol{\Gamma}_0\mathbf{f}_t\mathbf{f}_t^{'} & \frac{\sigma_0^2}{NT}E_0\sum_{t=1}^{T}\mathbf{g}_{0,2}^{'}\boldsymbol{\Gamma}_0\mathbf{f}_t\mathbf{f}_t^{'} & \cdots & \frac{\sigma_0^2}{NT}E_0\sum_{t=1}^{T}\mathbf{g}_{0,N}^{'}\boldsymbol{\Gamma}_0\mathbf{f}_t\mathbf{f}_t^{'} & \frac{tr(\mathbf{G}_0)}{N} \\
. & \frac{\sigma_0^2}{NT}E_0\sum_{t=1}^{T}\mathbf{f}_t\mathbf{f}_t^{'} & \mathbf{0}_{m\times m} & \cdots & \mathbf{0}_{m\times m} & 0 \\
. & \mathbf{0}_{m\times m} & \frac{\sigma_0^2}{NT}E_0\sum_{t=1}^{T}\mathbf{f}_t\mathbf{f}_t^{'} & \mathbf{0}_{m\times m} & \mathbf{0}_{m\times m} & 0 \\
\vdots & \vdots & \mathbf{0}_{m\times m} & \ddots & \vdots & \vdots \\
. & \mathbf{0}_{m\times m} & \mathbf{0}_{m\times m} & \cdots & \frac{\sigma_0^2}{NT}E_0\sum_{t=1}^{T}\mathbf{f}_t\mathbf{f}_t^{'} & 0 \\
. & 0 & 0 & \cdots & 0 & \frac{1}{2}
\end{bmatrix}
$$

$$
= \begin{bmatrix}
\frac{tr(\mathbf{G}_0^2+\mathbf{G}_0\mathbf{G}_0^{'})}{N} & \mathbf{0}_{1\times Nm} & \frac{tr(\mathbf{G}_0)}{N} \\
\mathbf{0}_{Nm\times 1} & \mathbf{0}_{Nm\times Nm} & \mathbf{0}_{Nm\times 1} \\
\frac{tr(\mathbf{G}_0)}{N} & \mathbf{0}_{1\times Nm} & \frac{1}{2}
\end{bmatrix} + \sigma_0^2 \begin{bmatrix}
\mathbf{H}_f(\rho_0, \boldsymbol{\gamma}_0^{'}) & \mathbf{0}_{(Nm+1)\times 1} \\
\mathbf{0}_{1\times(Nm+1)} & 0
\end{bmatrix}. \tag{C.2}
$$

Note that $\lambda_{\min}\left[\boldsymbol{\Lambda}_{f,NT}(\mathbf{0})\right] > 0$ is equivalent to $\det\left[\boldsymbol{\Lambda}_{NT}(\mathbf{0})\right] > 0$, which can be computed as follows (assuming $\frac{1}{NT}\sum_{t=1}^{T}\left[\mathbf{F}_t^{'}\mathbf{F}_t\right]$ to be positive definite):

$$
\begin{aligned}
\det\left[\boldsymbol{\Lambda}_{NT}(\mathbf{0})\right] &= \frac{1}{2}\det\left\{\begin{bmatrix}
\frac{tr(\mathbf{G}_0^2+\mathbf{G}_0\mathbf{G}_0^{'})}{N} + \frac{\sigma_0^2}{NT}E_0\sum_{t=1}^{T}\left[(\mathbf{G}_0\boldsymbol{\Gamma}_0\mathbf{f}_t)^{'}\mathbf{G}_0\boldsymbol{\Gamma}_0\mathbf{f}_t\right] & \frac{\sigma_0^2}{NT}E_0\sum_{t=1}^{T}\left[(\mathbf{G}_0\boldsymbol{\Gamma}_0\mathbf{f}_t)^{'}\mathbf{F}_t\right] \\
\frac{\sigma_0^2}{NT}E_0\sum_{t=1}^{T}E_0\left[\mathbf{F}_t^{'}\mathbf{G}_0\boldsymbol{\Gamma}_0\mathbf{f}_t\right] & \frac{\sigma_0^2}{NT}E_0\sum_{t=1}^{T}\left[\mathbf{F}_t^{'}\mathbf{F}_t\right]
\end{bmatrix} - 2\begin{bmatrix}\frac{tr(\mathbf{G}_0)}{N} \\ \mathbf{0}_{Nm\times 1}\end{bmatrix}\begin{bmatrix}\frac{tr(\mathbf{G}_0)}{N} & \mathbf{0}_{1\times Nm}\end{bmatrix}\right\} \\
&= \frac{1}{2}\det\left\{\begin{bmatrix}
\frac{tr(\mathbf{G}_0^2+\mathbf{G}_0\mathbf{G}_0^{'})}{N} - \frac{2[tr(\mathbf{G}_0)]^2}{N^2} + \frac{\sigma_0^2}{NT}E_0\sum_{t=1}^{T}\left[(\mathbf{G}_0\boldsymbol{\Gamma}_0\mathbf{f}_t)^{'}\mathbf{G}_0\boldsymbol{\Gamma}_0\mathbf{f}_t\right] & \frac{\sigma_0^2}{NT}E_0\sum_{t=1}^{T}\left[(\mathbf{G}_0\boldsymbol{\Gamma}_0\mathbf{f}_t)^{'}\mathbf{F}_t\right] \\
\frac{\sigma_0^2}{NT}E_0\sum_{t=1}^{T}\left[\mathbf{F}_t^{'}\mathbf{G}_0\boldsymbol{\Gamma}_0\mathbf{f}_t\right] & \frac{\sigma_0^2}{NT}E_0\sum_{t=1}^{T}\left[\mathbf{F}_t^{'}\mathbf{F}_t\right]
\end{bmatrix}\right\} \\
&= \frac{1}{2}\det\left\{\frac{\sigma_0^2}{NT}E_0\sum_{t=1}^{T}\left[\mathbf{F}_t^{'}\mathbf{F}_t\right]\right\}\left\{\frac{tr(\mathbf{G}_0^2+\mathbf{G}_0\mathbf{G}_0^{'})}{N} - \frac{2\left[tr(\mathbf{G}_0)\right]^2}{N^2}\right\},
\end{aligned}
$$

where in the last line we have used

$$E_0\sum_{t=1}^{T}\left[(\mathbf{G}_0\mathbf{\Gamma}_0\mathbf{f}_t)'\mathbf{G}_0\mathbf{\Gamma}_0\mathbf{f}_t\right]-E_0\sum_{t=1}^{T}\left[(\mathbf{G}_0\mathbf{\Gamma}_0\mathbf{f}_t)'\mathbf{F}_t\right]\left(E_0\sum_{t=1}^{T}\left[\mathbf{F}_t'\mathbf{F}_t\right]\right)^{-1}E_0\sum_{t=1}^{T}\left[\mathbf{F}_t'\mathbf{G}_0\mathbf{\Gamma}_0\mathbf{f}_t\right]=0. \quad \text{(C.3)}$$

Therefore, $\det\left[\mathbf{\Lambda}_{f,NT}(\mathbf{0})\right]>0$ if and only if $\frac{tr(\mathbf{G}_0^2+\mathbf{G}_0\mathbf{G}_0')}{N}-\frac{2[tr(\mathbf{G}_0)]^2}{N^2}>0$ and $\frac{1}{T}E_0\left(\mathbf{f}_t\mathbf{f}_t'\right)$ is positive definite.

Model (32): with exogenous variables $\mathbf{x}_{it}$

Suppose that the disturbances $e_{it}\sim IIDN(0,\sigma^2)$, the (quasi) log-likelihood function is given by

$$l(\boldsymbol{\varphi})=-\frac{NT}{2}ln(2\pi)-\frac{NT}{2}ln\sigma^2+Tln|\mathbf{S}(\rho)|-\frac{1}{2\sigma^2}\sum_{t=1}^{T}\left(\mathbf{S}(\rho)\mathbf{y}_{.t}-\mathbf{X}_{.t}\boldsymbol{\beta}-\mathbf{\Gamma}\mathbf{f}_t\right)'\left(\mathbf{S}(\rho)\mathbf{y}_{.t}-\mathbf{X}_{.t}\boldsymbol{\beta}-\mathbf{\Gamma}\mathbf{f}_t\right), \quad \text{(C.4)}$$

where $\boldsymbol{\varphi}=(\rho,\boldsymbol{\beta}',\boldsymbol{\gamma}',\sigma^2)'$. Under the assumption that $\mathbf{x}_{it}$ and $\mathbf{f}_t$ are uncorrelated, it follows that

$$\begin{aligned}\frac{1}{NT}E_0l(\boldsymbol{\varphi}) &= -\frac{1}{2}ln(2\pi)-\frac{1}{2}ln\sigma^2+\frac{1}{N}ln|\mathbf{S}(\rho)|-\frac{1}{2\sigma^2}\Bigg\{\left(\rho-\rho_0,(\boldsymbol{\beta}-\boldsymbol{\beta}_0)'\right)\mathbf{H}(\rho_0,\boldsymbol{\beta}_0')\left(\rho-\rho_0,(\boldsymbol{\beta}-\boldsymbol{\beta}_0)'\right)' \\ &\quad +\left(\rho-\rho_0,(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)'\right)\mathbf{H}_f(\rho_0,\boldsymbol{\gamma}_0')\left(\rho-\rho_0,(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)'\right)'+\frac{\sigma_0^2}{N}tr\left[\mathbf{S}_0^{-1}\mathbf{S}(\rho)\mathbf{S}'(\rho)\mathbf{S}_0^{-1'}\right]\Bigg\}, \\ \frac{1}{NT}E_0l(\boldsymbol{\varphi}_0) &= -\frac{1}{2}\left[ln(2\pi)+1\right]-\frac{1}{2}ln\sigma_0^2+\frac{1}{N}ln|\mathbf{S}_0|,\end{aligned}$$

where

$$\begin{aligned}\mathbf{H}(\rho_0,\boldsymbol{\beta}_0') &= \frac{1}{NT}E_0\left\{\begin{bmatrix}(\mathbf{G}_0^b\mathbf{X}\boldsymbol{\beta}_0)' \\ \mathbf{X}'\end{bmatrix}\begin{bmatrix}\mathbf{G}_0^b\mathbf{X}\boldsymbol{\beta}_0, & \mathbf{X}\end{bmatrix}\right\} \\ &= E_0\begin{bmatrix}\frac{1}{NT}\sum_{t=1}^{T}(\mathbf{G}_0\mathbf{X}_{.t}\boldsymbol{\beta}_0)'(\mathbf{G}_0\mathbf{X}_{.t}\boldsymbol{\beta}_0) & \frac{1}{NT}\sum_{t=1}^{T}(\mathbf{G}_0\mathbf{X}_{.t}\boldsymbol{\beta}_0)'\mathbf{X}_{.t} \\ \frac{1}{NT}\sum_{t=1}^{T}\mathbf{X}_{.t}'(\mathbf{G}_0\mathbf{X}_{.t}\boldsymbol{\beta}_0) & \frac{1}{NT}\sum_{t=1}^{T}\mathbf{X}_{.t}'\mathbf{X}_{.t}\end{bmatrix}, \quad \text{(C.5)}\end{aligned}$$

$\mathbf{G}_0^b=\mathbf{I}_T\otimes\mathbf{G}_0=\mathbf{I}_T\otimes\left(\mathbf{W}\mathbf{S}_0^{-1}\right)$, and $\mathbf{H}_f(\rho_0,\boldsymbol{\gamma}_0')$ is defined in (22). Hence, we obtain

$$\begin{aligned}\frac{E_0l(\boldsymbol{\varphi}_0)-E_0l(\boldsymbol{\varphi})}{NT} &= -\frac{1}{2}\left[ln\left(\frac{\sigma_0^2}{\sigma^2}\right)+\left(1-\frac{\sigma_0^2}{\sigma^2}\right)\right]-\frac{1}{N}\left\{ln|\mathbf{I}_N-(\rho-\rho_0)\mathbf{G}_0|+\frac{\sigma_0^2}{\sigma^2}(\rho-\rho_0)tr(\mathbf{G}_0)\right\} \\ &\quad +\frac{1}{2}\frac{\sigma_0^2}{\sigma^2}(\rho-\rho_0)^2\frac{tr(\mathbf{G}_0\mathbf{G}_0')}{N}+\frac{1}{2\sigma^2}\left(\rho-\rho_0,(\boldsymbol{\beta}-\boldsymbol{\beta}_0)'\right)\mathbf{H}(\rho_0,\boldsymbol{\beta}_0')\left(\rho-\rho_0,(\boldsymbol{\beta}-\boldsymbol{\beta}_0)'\right)' \\ &\quad +\frac{1}{2\sigma^2}\left(\rho-\rho_0,(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)'\right)\mathbf{H}_f(\rho_0,\boldsymbol{\gamma}_0')\left(\rho-\rho_0,(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)'\right)'.\end{aligned}$$

Denoting $Q_{NT}(\boldsymbol{\psi}) = (NT)^{-1}E_0[l(\boldsymbol{\varphi}_0) - l(\boldsymbol{\varphi})]$, where $\boldsymbol{\psi} = (d, \boldsymbol{\zeta}^{'}, \boldsymbol{\chi}^{'}, \vartheta)^{'}$ with $d = \rho - \rho_0$, $\boldsymbol{\zeta} = \boldsymbol{\beta} - \boldsymbol{\beta}_0$, $\boldsymbol{\chi} = \boldsymbol{\gamma} - \boldsymbol{\gamma}_0$, and $\vartheta = (\sigma^2 - \sigma_0^2)/\sigma^2 < 1$, we get

$$\begin{aligned} Q_{NT}(\boldsymbol{\psi}) &= -\frac{1}{2}\left[ln(1-\vartheta)+\vartheta\right] - \frac{1}{N}ln|\mathbf{I}_N - d\mathbf{G}_0| - \frac{1}{N}(1-\vartheta)dtr(\mathbf{G}_0) + \frac{1}{2}(1-\vartheta)d^2\frac{tr(\mathbf{G}_0\mathbf{G}_0^{'})}{N} \\ &\quad + \frac{1}{2}\sigma_0^2(1-\vartheta)\left(d, \boldsymbol{\zeta}^{'}\right)\mathbf{H}(\rho_0, \boldsymbol{\beta}_0^{'})\left(d, \boldsymbol{\zeta}^{'}\right)^{'} + \frac{1}{2}\sigma_0^2(1-\vartheta)\left(d, \boldsymbol{\chi}^{'}\right)\mathbf{H}_f(\rho_0, \boldsymbol{\gamma}_0^{'})\left(d, \boldsymbol{\chi}^{'}\right)^{'}. \end{aligned}$$

The first derivatives are given by

$$\begin{aligned} \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial d} &= \frac{1}{N}tr\left[(\mathbf{I}_N - d\mathbf{G}_0)^{-1}\mathbf{G}_0\right] - (1-\vartheta)\frac{tr(\mathbf{G}_0)}{N} + (1-\vartheta)d\frac{tr(\mathbf{G}_0\mathbf{G}_0^{'})}{N} \\ &\quad + \frac{\sigma_0^2(1-\vartheta)}{NT}E_0\sum_{t=1}^{T}\left[d(\mathbf{G}_0^b\mathbf{X}\boldsymbol{\beta}_0)^{'}\mathbf{G}_0^b\mathbf{X}\boldsymbol{\beta}_0 + (\mathbf{G}_0^b\mathbf{X}\boldsymbol{\beta}_0)^{'}\mathbf{X}\boldsymbol{\zeta} + d(\mathbf{G}_0\boldsymbol{\Gamma}_0\mathbf{f}_t)^{'}\mathbf{G}_0\boldsymbol{\Gamma}_0\mathbf{f}_t + (\mathbf{G}_0\boldsymbol{\Gamma}_0\mathbf{f}_t)^{'}\mathbf{F}_t\boldsymbol{\chi}\right], \\ \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial \boldsymbol{\zeta}} &= \frac{\sigma_0^2(1-\vartheta)}{NT}E_0\left[d\mathbf{X}^{'}\mathbf{G}_0^b\mathbf{X}\boldsymbol{\beta}_0 + \mathbf{X}^{'}\mathbf{X}\boldsymbol{\zeta}\right], \\ \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial \boldsymbol{\chi}_i} &= \frac{\sigma_0^2(1-\vartheta)}{NT}E_0\sum_{t=1}^{T}\left[d\mathbf{f}_t\mathbf{g}_{0,i}^{'}\boldsymbol{\Gamma}_0\mathbf{f}_t + \mathbf{f}_t\mathbf{f}_t^{'}\boldsymbol{\chi}_i\right], \text{ for } i = 1, 2, ..., N, \\ \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial \vartheta} &= \frac{1}{2}\frac{\vartheta}{1-\vartheta} - \frac{1}{2} + \frac{1}{N}dtr(\mathbf{G}_0) - \frac{1}{2}d^2\frac{tr(\mathbf{G}_0\mathbf{G}_0^{'})}{N} - -\frac{1}{2}\sigma_0^2\left(d, \boldsymbol{\zeta}^{'}\right)\mathbf{H}(\rho_0, \boldsymbol{\beta}_0^{'})\left(d, \boldsymbol{\zeta}^{'}\right)^{'} - \frac{1}{2}\sigma_0^2\left(d, \boldsymbol{\chi}^{'}\right)\mathbf{H}_f(\rho_0, \boldsymbol{\gamma}_0^{'})\left(d, \boldsymbol{\chi}^{'}\right)^{'}. \end{aligned}$$

The second derivatives are given by

$$\boldsymbol{\Lambda}_{NT}(\boldsymbol{\psi}) \equiv \frac{\partial^2 Q_{NT}(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}\boldsymbol{\psi}^{'}} = \begin{bmatrix} \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial^2 d} & \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial d\partial\boldsymbol{\zeta}^{'}} & \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial d\partial\boldsymbol{\chi}^{'}} & \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial d\partial\vartheta} \\ . & \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial\boldsymbol{\zeta}\partial\boldsymbol{\zeta}^{'}} & \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial\boldsymbol{\zeta}\partial\boldsymbol{\chi}^{'}} & \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial\boldsymbol{\zeta}\partial\vartheta} \\ . & . & \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial\boldsymbol{\chi}\partial\boldsymbol{\chi}^{'}} & \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial\boldsymbol{\chi}\partial\vartheta} \\ . & . & . & \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial^2\vartheta} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Lambda}_{11} & \boldsymbol{\Lambda}_{12} & \boldsymbol{\Lambda}_{13} & \boldsymbol{\Lambda}_{14} \\ . & \boldsymbol{\Lambda}_{22} & \boldsymbol{\Lambda}_{23} & \boldsymbol{\Lambda}_{24} \\ . & . & \boldsymbol{\Lambda}_{33} & \boldsymbol{\Lambda}_{34} \\ . & . & . & \boldsymbol{\Lambda}_{44} \end{bmatrix}, \quad \text{(C.6)}$$

where

$$
\begin{aligned}
\boldsymbol{\Lambda}_{11} &= \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial^2 d} = \frac{1}{N} tr\left\{(\mathbf{I}_N - d\mathbf{G}_0)^{-1}\mathbf{G}_0 (\mathbf{I}_N - d\mathbf{G}_0)^{-1}\mathbf{G}_0\right\} + (1-\vartheta)\frac{tr(\mathbf{G}_0\mathbf{G}_0^{'})}{N} \\
&\quad + \frac{\sigma_0^2(1-\vartheta)}{NT} E_0\left[(\mathbf{G}_0^b\mathbf{X}\boldsymbol{\beta}_0)^{'}\mathbf{G}_0^b\mathbf{X}\boldsymbol{\beta}_0 + (\mathbf{G}_0\boldsymbol{\Gamma}_0\mathbf{f}_t)^{'}\mathbf{G}_0\boldsymbol{\Gamma}_0\mathbf{f}_t\right], \\
\boldsymbol{\Lambda}_{12} &= \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial d\partial \boldsymbol{\zeta}^{'}} = \frac{\sigma_0^2(1-\vartheta)}{NT} E_0\left[(\mathbf{G}_0^b\mathbf{X}\boldsymbol{\beta}_0)^{'}\mathbf{X}\right], \\
\boldsymbol{\Lambda}_{13} &= \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial d\partial \boldsymbol{\chi}^{'}} = \left\{\frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial d\partial \boldsymbol{\chi}_i^{'}}\right\} = \left\{\frac{\sigma_0^2(1-\vartheta)}{NT} E_0 \sum_{t=1}^{T} \mathbf{g}_{0,i}^{'}\boldsymbol{\Gamma}_0\mathbf{f}_t\mathbf{f}_t^{'}\right\} \\
\boldsymbol{\Lambda}_{14} &= \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial d\partial \vartheta} = \frac{tr(\mathbf{G}_0)}{N} - d\frac{tr(\mathbf{G}_0\mathbf{G}_0^{'})}{N} - \frac{\sigma_0^2}{NT} E_0\left[d(\mathbf{G}_0^b\mathbf{X}\boldsymbol{\beta}_0)^{'}\mathbf{G}_0^b\mathbf{X}\boldsymbol{\beta}_0 + (\mathbf{G}_0^b\mathbf{X}\boldsymbol{\beta}_0)^{'}\mathbf{X}\boldsymbol{\zeta} + d(\mathbf{G}_0\boldsymbol{\Gamma}_0\mathbf{f}_t)^{'}\mathbf{G}_0\boldsymbol{\Gamma}_0\mathbf{f}_t + (\mathbf{G}_0\boldsymbol{\Gamma}_0\mathbf{f}_t)^{'}\mathbf{F}_t\boldsymbol{\chi}\right], \\
\boldsymbol{\Lambda}_{22} &= \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial \boldsymbol{\zeta}\partial \boldsymbol{\zeta}^{'}} = \frac{\sigma_0^2(1-\vartheta)}{NT} E_0\left[\mathbf{X}^{'}\mathbf{X}\right], \\
\boldsymbol{\Lambda}_{23} &= \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial \boldsymbol{\zeta}\partial \boldsymbol{\chi}^{'}} = \mathbf{0}, \\
\boldsymbol{\Lambda}_{24} &= \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial \boldsymbol{\zeta}\partial \vartheta} = -\frac{\sigma_0^2}{NT} E_0\left[d\mathbf{X}^{'}\mathbf{G}_0^b\mathbf{X}\boldsymbol{\beta}_0 + \mathbf{X}^{'}\mathbf{X}\boldsymbol{\zeta}\right], \\
\boldsymbol{\Lambda}_{33} &= \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial \boldsymbol{\chi}\partial \boldsymbol{\chi}^{'}} = \left\{\frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial \boldsymbol{\chi}_i\partial \boldsymbol{\chi}_j^{'}}\right\} = \left\{\frac{\sigma_0^2(1-\vartheta)}{NT} E_0 \sum_{t=1}^{T} \mathbf{f}_t\mathbf{f}_t^{'}, \text{ if } i=j; \text{ and } \mathbf{0}, \text{ if } i \neq j\right\}, \\
\boldsymbol{\Lambda}_{34} &= \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial \boldsymbol{\chi}\partial \vartheta} = \left\{\frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial \boldsymbol{\chi}_i\partial \vartheta}\right\} = \left\{-\frac{\sigma_0^2}{NT} E_0 \sum_{t=1}^{T}\left[d\mathbf{f}_t\mathbf{g}_{0,i}^{'}\boldsymbol{\Gamma}_0\mathbf{f}_t + \mathbf{f}_t\mathbf{f}_t^{'}\boldsymbol{\chi}_i\right]\right\}, \\
\boldsymbol{\Lambda}_{44} &= \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial^2 \vartheta} = \frac{1}{2(1-\vartheta)^2}.
\end{aligned}
$$

At $\boldsymbol{\psi} = \mathbf{0}$, we have

$$
\begin{aligned}
\boldsymbol{\Lambda}_{NT}(\mathbf{0}) &= \begin{bmatrix} \frac{tr(\mathbf{G}_0^2 + \mathbf{G}_0\mathbf{G}_0^{'})}{N} & \mathbf{0}_{1\times k} & \mathbf{0}_{1\times Nm} & \frac{tr(\mathbf{G}_0)}{N} \\ \mathbf{0}_{k\times 1} & \mathbf{0}_{k\times k} & \mathbf{0}_{k\times Nm} & \mathbf{0}_{k\times 1} \\ \mathbf{0}_{Nm\times 1} & \mathbf{0}_{Nm\times k} & \mathbf{0}_{Nm\times Nm} & \mathbf{0}_{Nm\times 1} \\ \frac{tr(\mathbf{G}_0)}{N} & \mathbf{0}_{1\times k} & \mathbf{0}_{1\times Nm} & \frac{1}{2} \end{bmatrix}, \\
&\quad + \sigma_0^2 \begin{bmatrix} \mathbf{H}(\rho_0, \boldsymbol{\beta}_0^{'}) & \mathbf{0}_{(k+1+Nm)\times 1} \\ \mathbf{0}_{1\times(k+1+Nm)} & 0 \end{bmatrix} + \sigma_0^2 \begin{bmatrix} \mathbf{H}_{f,11} & \mathbf{0}_{1\times k} & \mathbf{H}_{f,12} & 0 \\ \mathbf{0}_{k\times 1} & \mathbf{0}_{k\times k} & \mathbf{0}_{k\times Nm} & \mathbf{0}_{k\times 1} \\ \mathbf{H}_{f,21} & \mathbf{0}_{Nm\times k} & \mathbf{H}_{f,22} & \mathbf{0}_{Nm\times 1} \\ 0 & \mathbf{0}_{1\times k} & \mathbf{0}_{1\times Nm} & 0 \end{bmatrix}, \quad \text{(C.7)}
\end{aligned}
$$

where for brevity $\mathbf{H}_f(\rho_0, \boldsymbol{\gamma}_0^{'})$ is partitioned as

$$
\mathbf{H}_f(\rho_0, \boldsymbol{\gamma}_0^{'}) = \begin{bmatrix} (\mathbf{H}_{f,11})_{1\times 1} & (\mathbf{H}_{f,12})_{1\times Nm} \\ (\mathbf{H}_{f,21})_{Nm\times 1} & (\mathbf{H}_{f,22})_{Nm\times Nm} \end{bmatrix}. \quad \text{(C.8)}
$$

$\boldsymbol{\psi}_0$ are locally identified if and only if $\lambda_{\min}\left[\boldsymbol{\Lambda}_{NT}(\mathbf{0})\right] > 0$. Notice that all three terms in (C.7) are positive semidefinite. It follows that if $\mathbf{H}(\rho_0, \boldsymbol{\beta}_0^{'})$ is positive definite, then ρ_0 and $\boldsymbol{\beta}_0$ are identified. Given ρ_0, σ_0 can be identified through the first term in (C.7). On the other hand, if the first term is positive definite, which is the same identification condition (25) for the pure SAR model without exogenous regressors, then ρ_0 and σ_0 are identified, and if in addition $\frac{1}{NT}E_0(\mathbf{X}'\mathbf{X})$ is positive definite, $\boldsymbol{\beta}_0$ are identified. In both cases, $\boldsymbol{\gamma}_0$ is identified if $\frac{1}{T}E_0\left(\mathbf{f}_t\mathbf{f}_t'\right)$ is positive definite.