Common Factors and Spatial Dependence: An Application to US House Prices

Cynthia Fan Yang

Florida State University

1 November 2017

Common Factors and Spatial Dependence: An Application to US House Prices*

Cynthia Fan Yang†

August 20, 2018

Abstract

This paper considers panel data models with cross-sectional dependence arising from both spatial autocorrelation and unobserved common factors. It derives conditions for model identification and proposes estimation methods that employ cross-sectional averages as factor proxies, including the 2SLS, Best 2SLS, and GMM estimations. The proposed estimators are robust to unknown heteroskedasticity and serial correlation in the disturbances, unrequired to estimate the number of unknown factors, and computationally tractable. The paper establishes the asymptotic distributions of these estimators and compares their consistency and efficiency properties. Extensive Monte Carlo experiments lend support to the theoretical findings and demonstrate the satisfactory finite sample performance of the proposed estimators. The empirical section of the paper finds strong evidence of spatial dependence of real house price changes across 377 Metropolitan Statistical Areas in the US from 1975Q1 to 2014Q4. The results also reveal that population and income growth have significantly positive direct and spillover effects on house price changes. These findings are robust to different specifications of the spatial weights matrix constructed based on distance, migration flows, and pairwise correlations.

Keywords: Cross-sectional dependence, Common factors, Spatial panel data models, Generalized method of moments, House prices

JEL Classifications: C13, C23, R21, R31

*I am grateful to M. Hashem Pesaran for his invaluable advice. I greatly appreciate helpful comments from Cheng Hsiao, Wenguang Sun, Hyungsik Roger Moon, Yu-Wei Hsieh, and participants at the 2017 China Meeting of the Econometric Society, the Singapore Economic Review Conference 2017, the Third Annual Conference of International Association for Applied Econometrics (IAAE), and the 2017 J-WEN Mentoring Event. I would like to thank Natalia Bailey for helpful email correspondence regarding her approach. This paper supersedes an earlier draft circulated under the title "Identification and Estimation of Spatial Autoregressive Models with Common Factors".

†Department of Economics, Florida State University, 113 Collegiate Loop, 281 Bellamy Building, Tallahassee, FL 32306, USA. Email: cynthia.yang@fsu.edu. This work was carried out during my doctoral study at the University of Southern California.

1 Introduction

The past decade has seen a growing attention to panel data models with cross-sectional dependence, which refers to the interaction between cross-section units such as households, firms, regions, and countries. Researchers have become increasingly aware that ignoring cross-sectional dependence in panel data analysis could lead to inconsistent estimates and misleading inferences. The interdependence among individual units is prevalent in all kinds of economic activities. It could arise from common factors that influence a large number of economic agents, such as technological change and oil price fluctuations. It could also originate from certain explicit correlation structures formed by spatial arrangements, production networks, and social interactions. Accordingly, two main modeling approaches have been proposed to characterize this phenomenon: the common factor models and the spatial econometric models. In the former, cross-sectional dependence is captured by a number of observable or latent factors (or common shocks); in the latter, it is represented by spatial weights matrices typically based on physical, economic, or social distance. Although describing the same phenomenon, these two strands of literature have been developing separately, with different sets of assumptions and emphases. Therefore, efforts are called for to investigate the connections and differences between these two modeling approaches.

This paper aims to bring together factor and spatial models for a unified characterization of cross-sectional dependence. The main contributions of the paper are twofold. First, it considers a joint modeling of the two sources of cross-sectional dependence in panel data models: common factors and spatial interactions. It establishes identification conditions and proposes estimation methods for the joint model. Second, the paper provides a detailed empirical application to house price changes in the US and finds strong evidence of spatial effects. The empirical findings are robust and could carry important policy and business implications.

Specifically, our model specifications allow the common effects to be unobservable and the spatial dependence to be an inherent property of the dependent variable. We begin by deriving the identification conditions for the joint model. In particular, a simple necessary condition is provided, which is both verifiable and of practical relevance, especially for large sparse networks. We then propose a number of estimators for the model and establish their asymptotic distributions. We are faced with two major challenges in devising an estimation strategy. One is related to the unobserved factors, and the other is associated with the endogenous spatial lags of the dependent variable. The estimators developed in this paper approximate the unobserved factors by cross-sectional averages of the dependent and independent variables, and then utilize instrumental variables and other moment conditions to resolve the endogeneity problem. These estimators do not require estimating the number of factors, which is well known to be a challenging task. Moreover, they are robust to both heteroskedasticity and serial correlations in the disturbances, and they are computationally attractive. We show that the proposed estimators, including the two-stage least squares (2SLS), Best 2SLS, and generalized method of moments (GMM) estimators, are consistent as long as the cross-section dimension (N) is large, irrespective of the size of the time series dimension (T). Furthermore, they are asymptotically normally distributed without nuisance parameters, provided that T is relatively smaller than N, as both N and T tend jointly towards infinity. The Monte Carlo simulation results support the identification conditions. A series of detailed experiments also demonstrate the satisfactory finite-sample properties of the proposed estimators.

The proposed estimation methods are applied in order to analyze changes in real house price the US across 377 Metropolitan Statistical Areas (MSAs) from 1975Q1 to 2014Q4. The study demonstrates the importance of the effective removal of common effects in evaluating the strength of spatial connections. It documents significant spatial dependence in house price changes. It also shows that population and income growth significantly increase house price growth through both direct effect and spillover effect. These findings are fairly robust to various specifications of the spatial weights, including weights based on distance, on migration flows, and on pairwise correlations of the de-factored observations.

Related Literature The theoretical analysis in this paper belongs to a recent and growing literature on panel data models with cross-sectional dependence (CSD). Chudik et al. (2011) introduce the notions of weak and strong CSD. Applying these concepts, a spatial model can be shown to be a form of weak CSD, whereas the standard factor model represents a form of strong CSD (Pesaran and Tosetti, 2011; Bailey, Holly, and Pesaran, 2016). Bailey, Kapetanios, and Pesaran (2016) propose measuring the degree of CSD by an exponent of dependence, which captures how fast the variance of the cross-sectional average declines with the cross-section dimension, N. Using this exponent of cross-sectional dependence, Pesaran (2015) further discusses testing for weak CSD in large panels.[1]

The characterization of CSD is divided into two areas of writing. On the one hand, there is a large body of literature on common factor models. Recent contributions on large panel data models with common factors include Pesaran (2006), Bai (2009), Bai and Li (2012), and Moon and Weidner (2015), just to name a few. Our study is particularly related to an influential paper by Pesaran (2006), who develops Common Correlated Effects (CCE) estimators for panel data models with multifactor error structure. The basic idea behind the CCE estimators is to filter the unobserved factors with cross-sectional averages. In follow-up studies, Kapetanios et al. (2011) show that the CCE estimators are still applicable if the unobserved factors follow unit root processes; Chudik and Pesaran (2015a) extend the estimation approach to models with lagged dependent variables and weakly exogenous regressors.

On the other hand, the present paper also draws from the spatial econometrics literature.[2] Two main classes of methods have been developed to estimate spatial models: the maximum likelihood (ML) techniques (Anselin, 1988; Lee, 2004; Yu et al., 2008; Lee and Yu, 2010a; Aquaro et al., 2015), and the instrumental variables (IV)/GMM approaches (Kelejian and Prucha, 1999, 2010; Lee, 2007; Lin and Lee, 2010; Lee and Yu, 2014). The estimation strategy in the current article is related to and builds on the GMM framework. Regarding the identification conditions of spatial models, a

[1] For overviews of the literature on panel data models with error cross-sectional dependence, see Sarafidis and Wansbeek (2012) and Chudik and Pesaran (2015b).

[2] Comprehensive reviews of spatial econometrics can be found in books including Anselin (1988) and Elhorst (2014). Also see the survey article by Lee and Yu (2010b) for the latest developments in spatial panel data models.

systematic discussion is provided in a recent study by Lee and Yu (2016) under the assumption that the sample size is finite. Aquaro et al. (2015) also conduct a detailed investigation of the identifiability of spatial models with heterogeneous coefficients. The present paper sheds new light on the identification of spatial models with factors, and it shows that the conditions in Lee and Yu (2016) cannot be applied when N tends to infinity.

The current paper is most closely related to a number of more recent studies that consider both common factors and spatial effects. Pesaran and Tosetti (2011) consider models where the idiosyncratic errors are spatially correlated and subject to common shocks. Bai and Li (2014) specify the spatial autocorrelation on the dependent variable while assuming the presence of unobserved common shocks. They advocate a pseudo-ML method that simultaneously estimates a large group of parameters, including the heterogeneous factor loadings and heterogeneous variances of the disturbances. A similar approach is considered by Bai and Li (2015) for dynamic models. Other studies within the ML framework include Shi and Lee (2017), and Lu (2017). However, besides computational complexities, the ML methods are not robust to serial correlation in the errors, and they require knowing or estimating the number of latent factors.[3] Instead of estimating the two effects jointly, Bailey, Holly, and Pesaran (2016) propose a two-stage approach that extracts the common factors in the first stage and then estimates the spatial connections in the second stage. Nonetheless, a formal distribution theory that takes into account the first-stage sampling errors is not yet available.

The empirical investigation in the present paper is concerned with the spatial dependence in house prices. The phenomenon that house price variations tend to exhibit spatial correlations has received increasing attention from economists over the past two decades, although little consensus has been reached regarding the spatial transmission mechanism. Possible explanations include migration, equity transfer, spatial arbitrage, and spatial patterns in the determinants of house prices (Meen, 1999). Researchers have obtained evidence on the spatial spillovers of house prices in the US at different levels of aggregation using various methods.[4] For example, Pollakowski and Ray (1997) examine nine US Census divisions as well as the New York metropolitan area using a vector autoregressive (VAR) model. Brady (2011) focuses on the diffusion of house prices across a panel of California counties by means of impulse response functions. Holly et al. (2010) analyze US house prices the State level using a spatial error model, where the importance of spatial effects is evaluated by fitting a spatial model to the residuals from a CCE estimation procedure. Brady (2014) also consider State level house prices but utilize spatial impulse response functions from a single-equation spatial autoregressive model. The current paper focuses on the extent to which house prices are interdependent among near 400 Metropolitan Statistical Areas (MSAs) in the US. Little research has investigated this issue at the MSA level. One exception is the study undertaken by Cohen et al. (2016), who incorporate geography into an autoregressive model via cross-lag effects

[3]Much is written on estimating the number of unobservable factors. See, for example, Bai and Ng (2002) (2007), Kapetanios (2010), and Stock and Watson (2011).

[4]International evidence on the spatial interconnections of house prices are provided by Luo et al. (2007) for Australia, Shi et al. (2009) for New Zealand, and Holly et al. (2011) for the UK, just to name a few.

and do not employ a spatial econometric approach.[5] Our empirical analysis is closely related to the inquiry by Bailey, Holly, and Pesaran (2016), who examine MSA level house price changes with a two-stage procedure. In comparison, besides using more recent data on updated MSA delineations, the present paper adopts a different estimation approach that jointly considers common factors and spatial dependence. It also explores the direct and indirect effects of possible determinant variables on house price growth. Another contribution of this paper involves the specification of spatial weights matrix based on migration flows.

Outline of the Paper The rest of the 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 investigates the identification conditions. Section 4 establishes the asymptotic distributions of the 2SLS, Best 2SLS, and GMM estimators. Section 5 reports the Monte Carlo experiments for the identification and estimation experiments. Section 6 presents an empirical application to US house prices, and finally, Section 7 concludes. The Appendix provides proofs of the main theorems and further details on data sources and variable transformations. The Online Supplement contains a list of lemmas used in the main proofs, and derivations of the identification conditions. The Supplement also gives additional results of Monte Carlo experiments and further empirical findings.

Notations

For an $N \times N$ real matrix $\mathbf{A} = (a_{ij})$, $||\mathbf{A}|| = \sqrt{tr(\mathbf{AA}')}$, $||\mathbf{A}||_\infty = \max\limits_{1\leq i\leq N} \sum_{j=1}^{N} |a_{ij,N}|$ 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 matrix $\mathbf{A}$, respectively. We say that the row (column) sums of a (sequence of) matrix $\mathbf{A}$ are uniformly bounded in absolute value, or $\mathbf{A}$ has bounded row (column) norm for short, if there exists a constant K, such that $||\mathbf{A}||_\infty < K < \infty$ $(||\mathbf{A}||_1 < K < \infty)$ for all N. $vec(\mathbf{A})$ is the column vector obtained by stacking the columns of $\mathbf{A}$. $Diag\,(\mathbf{A}) = Diag\,(a_{11}, a_{22}, \ldots, a_{NN})$ represents an $N \times N$ diagonal matrix formed with the diagonal entries of $\mathbf{A}$, whereas $diag(\mathbf{A}) = (a_{11}, a_{22}, \ldots, a_{NN})'$ denotes an $N \times 1$ vector. $\lambda_{\max}(\mathbf{A})$ and $\lambda_{\min}(\mathbf{A})$ are the largest and smallest eigenvalues of matrix $\mathbf{A}$, respectively. $tr(\mathbf{A})$ denotes the trace of matrix $\mathbf{A}$, and $\det(\mathbf{A})$ denotes the determinant of $\mathbf{A}$. $\odot$ stands for the Hadamard product, and $\otimes$ is the Kronecker product. $(N,T) \xrightarrow{j} \infty$ denotes joint convergence of N and T. Let $\{x_N\}_{N=1}^{\infty}$ be any real sequence and $\{y_N\}_{N=1}^{\infty}$ be a sequence of positive real numbers; we adopt the Landau's symbols and write $x_N = O\,(y_N)$ if there exists a positive finite constant K such that $|x_N| \leq K y_N$ for all N, and $x_N = o\,(y_N)$ if $x_N/y_N \to 0$ as $N \to \infty$. $O_p(.)$ and $o_p(.)$ are the equivalent stochastic orders in probability. $\lfloor x \rfloor$ denotes the integral part of a real number x. K is used generically for a finite positive constant.

[5] Cohen et al. (2016) also use a house price index different from ours. Specifically, the authors adopt the consolidated house price index by the Office of Federal Housing Enterprise Oversight (OFHEO) that covers 363 MSAs over the period of 1996-2013.

2 The Model and Assumptions

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

$$\begin{aligned} y_{it} &= \rho y_{it}^{*} + \boldsymbol{\beta}'\mathbf{x}_{it} + \boldsymbol{\gamma}_i'\mathbf{f}_t + e_{it}, \\ \mathbf{x}_{it} &= \mathbf{A}_i'\mathbf{f}_t + \mathbf{v}_{it}, \end{aligned} \tag{1}$$

for $i = 1, 2, \ldots, N$, and $t = 1, 2, \ldots, T$, where y_{it} is the dependent variable of unit i at time t, and $y_{it}^{*} = \sum_{j=1}^{N} w_{ij} y_{jt}$, which represents the endogenous interaction effects (or spatial lag effects) among the dependent variable. The matrix $\mathbf{W} = (w_{ij})_{N\times N}$ is a specified spatial weights matrix of known constants. It characterizes neighborhood relations, which are typically based on a geographical arrangement or on socio-economic connections of the cross-section units. The parameter ρ captures the strength of spatial dependence across observations on the dependent variable and is known as the spatial autoregressive coefficient. The $k \times 1$ vector $\mathbf{x}_{it} = (x_{it,1}, x_{it,2}, \ldots, x_{it,k})'$ contains individual-specific explanatory variables, and $\boldsymbol{\beta}$ is the corresponding vector of coefficients, where k is assumed to be a known fixed number. The variables e_{it} and $\mathbf{v}_{it} = (v_{it,1}, v_{it,2}, \ldots, v_{it,k})'$ are the idiosyncratic disturbances associated with y_{it} and $\mathbf{x}_{it}$ processes, respectively. The $m \times 1$ vector $\mathbf{f}_t = (f_{1t}, f_{2t}, \ldots, f_{mt})'$ represents unobserved common factors, where m is fixed but possibly unknown. The factor loadings $\boldsymbol{\gamma}_i$ and $\mathbf{A}_i$ capture heterogeneous impacts from the common effects on cross-section units.[6] Overall, the term ρy_{it}^{*} captures the spatial effect, while $\boldsymbol{\gamma}_i'\mathbf{f}_t$ captures the common factor effect. The latter is also referred to in the literature as an interactive effect, since it can be viewed as a generalization of the traditional additive fixed effect. The parameters of interest throughout this paper are $\boldsymbol{\delta} = \left(\rho, \boldsymbol{\beta}'\right)'$.

In model (1), the explanatory variables are specified so that they can be influenced by the same factors that affect the dependent variable. Such a specification is reasonable in practice and has been considered in studies including Pesaran (2006) and Bai and Li (2014). Also note that this model can be readily extended without additional complication to include observable factors such as intercepts, seasonal dummies, and deterministic trends;[7] here we focus on unobservable factors to facilitate exposition.

To cope with the unknown factors in model (1), we replace them with cross-sectional averages of the dependent and individual-specific independent variables, following the idea pioneered by Pesaran (2006). To see why this approximation works for the SAR model, we begin by rewriting model (1) as follows:

$$\begin{pmatrix} y_{it} - \rho \sum_{j=1}^{N} w_{ij} y_{jt} - \boldsymbol{\beta}'\mathbf{x}_{it} \\ \mathbf{x}_{it} \end{pmatrix} = \boldsymbol{\Phi}_i'\mathbf{f}_t + \mathbf{u}_{it}, \tag{2}$$

where $\boldsymbol{\Phi}_i = (\boldsymbol{\gamma}_i, \mathbf{A}_i)$, $\mathbf{u}_{it} = (e_{it}, \mathbf{v}_{it}')'$. Then, stacking (2) by individual unit for each time period,

[6] The heterogeneity in factor loadings may arise, for example, from differences in endowment, technical rigidities, or innate ability.

[7] See Remark 2 of Pesaran (2006).

the model can be expressed more compactly as

$$\boldsymbol{\Delta}\left(\rho,\boldsymbol{\beta}\right)\mathbf{z}_{.t}=\boldsymbol{\Phi}\mathbf{f}_t+\mathbf{u}_{.t},\quad \text{for } t=1,2,\ldots,T, \tag{3}$$

where $\mathbf{z}_{.t}=\left(\mathbf{z}'_{1t},\mathbf{z}'_{2t},\ldots,\mathbf{z}'_{Nt}\right)'$ is an $N\left(k+1\right)$-dimensional vector of observations, with $\mathbf{z}_{it}=\left(y_{it},\mathbf{x}'_{it}\right)'$, $\boldsymbol{\Phi}=\left(\boldsymbol{\Phi}_1,\boldsymbol{\Phi}_2,\ldots,\boldsymbol{\Phi}_N\right)'$, $\mathbf{u}_{.t}=\left(\mathbf{u}'_{1t},\mathbf{u}'_{2t},\ldots,\mathbf{u}'_{Nt}\right)'$, and $\boldsymbol{\Delta}=\boldsymbol{\Delta}\left(\rho,\boldsymbol{\beta}\right)$ is a square matrix, of which the $(i,j)^{th}$ subblock of size $(k+1)$, for $i,j=1,2,\ldots,N$, is given by

$$\boldsymbol{\Delta}_{ii}=\begin{pmatrix}1 & -\boldsymbol{\beta}'\\ 0 & \mathbf{I}_k\end{pmatrix},\ \text{if } i=j;\quad \text{and}\quad \boldsymbol{\Delta}_{ij}=\begin{pmatrix}-\rho w_{ij} & 0\\ 0 & 0\end{pmatrix},\ \text{if } i\neq j.$$

The way of stacking the equations in (2) follows that in Bai and Li (2014), who show that $\boldsymbol{\Delta}^{-1}=\boldsymbol{\Delta}^{-1}\left(\rho,\boldsymbol{\beta}\right)$ exists and its $(i,j)^{th}$ subblock is given by[8]

$$\boldsymbol{\Delta}_{ii}^{-1}=\begin{pmatrix}\check{s}_{ii} & \check{s}_{ii}\boldsymbol{\beta}'\\ \mathbf{0} & \mathbf{I}_k\end{pmatrix},\ \text{if } i=j;\quad \text{and}\quad \boldsymbol{\Delta}_{ij}^{-1}=\begin{pmatrix}\check{s}_{ij} & \check{s}_{ij}\boldsymbol{\beta}'\\ \mathbf{0} & \mathbf{0}\end{pmatrix},\ \text{if } i\neq j, \tag{4}$$

where $\check{s}_{ij}$ denotes the $(i,j)^{th}$ element of $\mathbf{S}^{-1}\left(\rho\right)$, and $\mathbf{S}\left(\rho\right)=\mathbf{I}_N-\rho\mathbf{W}$. The inverse of $\mathbf{S}\left(\rho\right)$ exists under certain regularity conditions, which will be discussed later. It then follows from (4) that (3) is equivalent to

$$\mathbf{z}_{.t}=\boldsymbol{\Delta}^{-1}\left(\boldsymbol{\Phi}\mathbf{f}_t+\mathbf{u}_{.t}\right)=\mathbf{C}'\mathbf{f}_t+\boldsymbol{\epsilon}_{.t}, \tag{5}$$

where $\mathbf{C}=\left(\boldsymbol{\Delta}^{-1}\boldsymbol{\Phi}\right)'$ and $\boldsymbol{\epsilon}_{.t}=\boldsymbol{\Delta}^{-1}\mathbf{u}_{.t}=\left(\boldsymbol{\epsilon}'_{1t},\boldsymbol{\epsilon}'_{2t},\ldots,\boldsymbol{\epsilon}'_{Nt}\right)'$ are the transformed new error terms.

Now letting $\boldsymbol{\Theta}_a=N^{-1}\boldsymbol{\tau}'_N\otimes\mathbf{I}_{k+1}$, where $\boldsymbol{\tau}_N$ is an $N\times 1$ vector of ones, it is easily verified that $\bar{\mathbf{z}}_{.t}=\boldsymbol{\Theta}_a\mathbf{z}_{.t}=\left(\bar{y}_{.t},\bar{\mathbf{x}}'_{.t}\right)'$, where $\bar{y}_{.t}=T^{-1}\sum_{i=1}^{N}y_{it}$ and $\bar{\mathbf{x}}_{.t}=T^{-1}\sum_{i=1}^{N}\mathbf{x}_{it}$. As shown,$\boldsymbol{\Theta}_a$ is a matrix that operates on any $N\left(k+1\right)$-dimensional vector that is stacked in the same order as $\mathbf{z}_{.t}$ and produces an $k\times 1$ vector of cross-sectional averages. Similarly, we have $\bar{\boldsymbol{\epsilon}}_{.t}=\boldsymbol{\Theta}_a\boldsymbol{\epsilon}_{.t}=T^{-1}\sum_{i=1}^{N}\boldsymbol{\epsilon}_{it}$. Premultiplying both sides of (5) with $\boldsymbol{\Theta}_a$ yields

$$\bar{\mathbf{z}}_{.t}=\bar{\mathbf{C}}'\mathbf{f}_t+\bar{\boldsymbol{\epsilon}}_{.t}, \tag{6}$$

where

$$\bar{\mathbf{C}}=\left(\boldsymbol{\Theta}_a\mathbf{C}'\right)'=N^{-1}\left[\sum_{i=1}^{N}\sum_{j=1}^{N}\check{s}_{ij}\left(\boldsymbol{\gamma}_j+\mathbf{A}_j\boldsymbol{\beta}\right),\ \sum_{j=1}^{N}\mathbf{A}_j\right], \tag{7}$$

Assuming that $\bar{\mathbf{C}}$ has full row rank, namely, $\text{Rank}\left(\bar{\mathbf{C}}\right)=m\leq k+1$, for all N including $N\to\infty$, we obtain

$$\mathbf{f}_t=\left(\bar{\mathbf{C}}\bar{\mathbf{C}}'\right)^{-1}\bar{\mathbf{C}}\left(\bar{\mathbf{z}}_{.t}-\bar{\boldsymbol{\epsilon}}_{.t}\right). \tag{8}$$

The task now is to show that $\bar{\boldsymbol{\epsilon}}_{.t}$ diminishes for sufficiently large N. We establish in Lemma A2 that $\bar{\boldsymbol{\epsilon}}_{.t}$ converges to zero in quadratic mean as $N\to\infty$, for any t. It follows from (8) that $\mathbf{f}_t$ can be

[8] See Lemma A.1 of Bai and Li (2014).

approximated by the cross-sectional averages $\bar{\mathbf{z}}_{.t}$ with an error of order $O_p(1/\sqrt{N})$. More formally, we have

$$\mathbf{f}_t \xrightarrow{p} \left(\mathbf{C}_0\mathbf{C}_0'\right)^{-1}\mathbf{C}_0\bar{\mathbf{z}}_{.t}, \quad \text{as } N \to \infty, \tag{9}$$

where

$$\mathbf{C}_0 = \lim_{N\to\infty} \bar{\mathbf{C}} = \left[E\left(\boldsymbol{\gamma}_i\right), E\left(\mathbf{A}_i\right)\right] \begin{pmatrix} \bar{\check{s}} & \mathbf{0} \\ \bar{\check{s}}\boldsymbol{\beta} & \mathbf{I}_k \end{pmatrix},$$

$$\bar{\check{s}} = N^{-1}\boldsymbol{\tau}_N'\mathbf{S}^{-1}\left(\rho\right)\boldsymbol{\tau}_N = N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\check{s}_{ij}.$$

It is clear from (9) that $\bar{\mathbf{z}}_{.t}$ serve fairly well as factor proxies as long as N is large.[9] Note that the use of equal weights in constructing the cross-sectional averages is nonessential to the asymptotic analysis, which can be readily carried through with other weighting schemes satisfying the granularity conditions.[10] Thus, the current paper will focus on simple cross-sectional averages for ease of exposition.

To facilitate formal analysis, it is convenient to define the infeasible de-factoring matrices (or residual maker) as follows:

$$\mathbf{M}_f = \mathbf{I}_T - \mathbf{F}\left(\mathbf{F}'\mathbf{F}\right)^{-}\mathbf{F}', \quad \mathbf{M}_f^b = \mathbf{M}_f \otimes \mathbf{I}_N, \tag{10}$$

where $\mathbf{F} = (\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_T)'$ is a $T \times m$ matrix of unobserved common factors, and $(\mathbf{F}'\mathbf{F})^{-}$ denotes the generalized inverse of $\mathbf{F}'\mathbf{F}$. The observable counterparts of (10) that utilize cross-sectional averages are given by

$$\bar{\mathbf{M}} = \mathbf{I}_T - \bar{\mathbf{Z}}\left(\bar{\mathbf{Z}}'\bar{\mathbf{Z}}\right)^{-}\bar{\mathbf{Z}}', \quad \mathbf{M}^b = \bar{\mathbf{M}} \otimes \mathbf{I}_N, \tag{11}$$

where $\bar{\mathbf{Z}} = (\bar{\mathbf{z}}_{.1}, \bar{\mathbf{z}}_{.2}, \ldots, \bar{\mathbf{z}}_{.T})'$. Note that $\mathbf{M}_f^b$ and $\mathbf{M}^b$ are de-factoring matrices of NT dimension that operate on the observations stacked as successive cross sections, namely, $\mathbf{Y} = (\mathbf{y}_{.1}', \mathbf{y}_{.2}', \ldots, \mathbf{y}_{.T}')'$ and $\mathbf{X} = (\mathbf{X}_{.1}', \mathbf{X}_{.2}', \ldots, \mathbf{X}_{.T}')'$, where $\mathbf{y}_{.t} = (y_{1t}, y_{2t}, \ldots, y_{NT})'$ and $\mathbf{X}_{.t} = (\mathbf{x}_{1t}, \mathbf{x}_{2t}, \ldots, \mathbf{x}_{Nt})'$, for $t = 1, 2, \ldots, T$. Throughout this paper, K is used generically to denote a finite positive constant.

In order to formally analyze model (1), we will make the following assumptions:

Assumption 1. *The unobserved common factors $\mathbf{f}_t$ are covariance stationary with absolutely summable autocovariances, and they 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*

(i) 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 $(\zeta_{it}, \boldsymbol{\varsigma}_{it}')' \sim IID\left(0_{k+1}, \mathbf{I}_{k+1}\right)$ with finite fourth-order moments. The errors e_{it} and $\mathbf{v}_{jt'}$ are distributed independently of

[9] In practice, it may also worth including $\bar{y}_t^*$ as factor proxies if $\bar{y}_t^*$ is not highly correlated with $\bar{y}_t$, where $\bar{y}_t^* = N^{-1}\sum_{i=1}^{N} y_t^*$.

[10] See Assumption 5 in Pesaran (2006).

each other, for all i, j, t, t'. In addition, $Var(e_{it}) = \sum_{l=0}^{\infty} a_{il}^2 = \sigma_i^2 < K$ and $Var(\mathbf{v}_{it}) = \sum_{l=0}^{\infty} \mathbf{\Xi}_{il}\mathbf{\Xi}_{il}' = \mathbf{\Sigma}_{v,i} < K$, where $\sigma_i^2 > 0$ and $\mathbf{\Sigma}_{v,i}$ is positive definite.

(ii) The error term e_{it} has absolutely summable cumulants up to the fourth order.

Assumption 3. *The factor loadings, $\boldsymbol{\gamma}_i$ and $\mathbf{A}_i$, are independently and identically distributed across i, and independent of e_{jt}, $\mathbf{v}_{jt}$, and $\mathbf{f}_t$, for all i, j, and t. Both $\boldsymbol{\gamma}_i$ and $\mathbf{A}_i$ have fixed means, which are given by $\boldsymbol{\gamma}$ and $\mathbf{A}$, respectively, and finite variances. In particular, for all i, $\boldsymbol{\gamma}_i = \boldsymbol{\gamma} + \boldsymbol{\eta}_i$, $\boldsymbol{\eta}_i \sim IID(\mathbf{0}, \mathbf{\Omega}_\eta)$, where $\mathbf{\Omega}_\eta$ is a symmetric non-negative definite matrix, $\|\boldsymbol{\gamma}\| < K$, $\|\mathbf{A}\| < K$, and $\|\mathbf{\Omega}_\eta\| < K$.*

Assumption 4. *The true parameter vector, $\boldsymbol{\delta}_0 = \left(\rho_0, \boldsymbol{\beta}_0'\right)'$, is in the interior of the parameter space, denoted by $\boldsymbol{\Delta}_{sp}$, which is a compact subset of the $(k+1)$-dimensional Euclidean space, $\mathbb{R}^{k+1}$.*

Assumption 5. *The matrix $\bar{\mathbf{C}}$, given by (7), has full row rank for all N, including $N \to \infty$.*

Assumption 6. *The $N \times N$ nonstochastic spatial weights matrix, $\mathbf{W} = (w_{ij})$, has bounded row and column sum norms, namely, $||\mathbf{W}||_\infty < K$ and $||\mathbf{W}||_1 < K$, respectively, and*

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

for all values of ρ. In addition, the diagonal entries of $\mathbf{W}$ are zero, that is, $w_{ii} = 0$, for all $i = 1, 2, \ldots, N$.

Assumption 7. *The $N \times q$ matrix of instrumental variables, $\mathbf{Q}_{.t}$, for $t = 1, 2, \ldots, T$, is composed of a subset of the columns of $\left(\mathbf{X}_{.t}, \mathbf{W}\mathbf{X}_{.t}, \mathbf{W}^2\mathbf{X}_{.t}, \ldots\right)$, and its column dimension q is fixed for all N and t. The matrix $\mathbf{Q} = (\mathbf{Q}_{.1}', \mathbf{Q}_{.2}', \ldots, \mathbf{Q}_{.T}')'$ represents the IV matrix of dimension $NT \times q$.*

Assumption 8. *(i) There exists N_0 and T_0, such that for all $N > N_0$ and $T > T_0$, the matrices $(NT)^{-1}\mathbf{Q}'\mathbf{M}^b\mathbf{Q}$ and $(NT)^{-1}\mathbf{Q}'\mathbf{M}_f^b\mathbf{Q}$ exist and are nonsingular.*

(ii) The matrix $p\lim_{N,T\to\infty}(NT)^{-1}\left(\mathbf{Q}'\mathbf{M}_f^b\mathbf{L}_0\right)$ is of full column rank, where $\mathbf{L}_0 = \left(\mathbf{G}_0^b\mathbf{X}\boldsymbol{\beta},\ \mathbf{X}\right)$, $\mathbf{G}_0^b = \mathbf{I}_T \otimes \mathbf{G}_0$, and $\mathbf{G}_0 = \mathbf{W}\mathbf{S}^{-1}(\rho_0)$.

(iii) $E|x_{it,p}|^{2+\delta} < K$, for some $\delta > 0$, and for all $i = 1, 2, \ldots, N$, $t = 1, 2, \ldots, T$, and $p = 1, 2, \ldots, k$.

Remark 1. An attractive feature of the model is that it allows for the presence of both heteroskedasticity and serial correlation in the disturbance processes, as stated in Assumption 2.[11] The asymptotic analysis in the current paper is conducted under this fairly general configuration, and the theoretical findings are corroborated by Monte Carlo evidence. Note that Assumption 2(ii) is only made for the limit theory of the GMM estimator. Under Assumption 2, we have $Var(\mathbf{u}_{.t}) = \mathbf{\Sigma}_u = Diag(\mathbf{\Sigma}_{u,1}, \mathbf{\Sigma}_{u,2}, \ldots, \mathbf{\Sigma}_{u,N})$ and $Var(\mathbf{u}_{it}) = \mathbf{\Sigma}_{u,i} = Diag\left(\sigma_i^2, \mathbf{\Sigma}_{v,i}\right)$, for $i = 1, 2, \ldots, N$; both $\mathbf{\Sigma}_u$ and $\mathbf{\Sigma}_{u,i}$ are block-diagonal matrices.

[11] This model can be further extended to accommodate spatial correlations in the error processes.

Remark 2. The assumptions on the factors and factor loadings (Assumptions 1 and 3) follow the specifications in Pesaran (2006). The compactness of the parameter space in Assumption 4 is a condition to facilitate the theoretical analysis of the GMM estimation. This condition is usually assumed when the objective function for an estimator is highly nonlinear. The rank condition in Assumption 5 is imposed for analytical convenience and can be relaxed following similar arguments as in Pesaran (2006).[12]

Remark 3. Assumption 6 ensures that $\mathbf{S}(\rho)$ is nonsingular for all possible values of ρ, where $\mathbf{S}(\rho) = \mathbf{I}_N - \rho\mathbf{W}$. 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 6 also implies that $\mathbf{S}^{-1}(\rho)$ is uniformly bounded in row and column sums in absolute value for all values of ρ, since

$$||\mathbf{S}^{-1}||_1 = ||\mathbf{I}_N + \rho\mathbf{W} + \rho^2\mathbf{W}^2 + \ldots||_1 \leq 1 + |\rho|||\mathbf{W}||_1 + |\rho|^2||\mathbf{W}||_1^2 + \ldots = \frac{1}{1 - |\rho|||\mathbf{W}||_1} < K,$$

and similarly, it can be shown that $||\mathbf{S}^{-1}||_\infty < K$. The uniform boundedness assumption is standard in the spatial econometrics literature. It essentially imposes sparsity restrictions on $\mathbf{W}$ so that the degree of cross-sectional correlation is manageable. As we shall see, this assumption plays an important role in the asymptotic analysis. Also note that $\mathbf{W}$ need not to be row-standardized so that each row sums to unity, which is often performed in practice for ease of interpretation. If all the elements of $\mathbf{W}$ are non-negative, row-standardization implies that y_{it}^* is a weighted average of neighboring values. Lastly, the zero diagonal assumption for the $\mathbf{W}$ matrix is innocuous and only for notational convenience in discussing the GMM estimation. No unit has self-influence under this assumption, which is clearly satisfied if $\mathbf{W}$ represents geographical distance or social interactions.

Remark 4. The spatially lagged dependent variable, y_{it}^*, is in general correlated with the error term. The selection of the instrumental variables in Assumption 7 originates from Kelejian and Prucha (1998) for cross-sectional SAR models. This choice is motivated by the spatial power series expansion of the expectation of the spatial lag (see Kelejian and Prucha, 1998, p.104).

Remark 5. Assumptions 8(i) and 8(ii) are the standard rank conditions for the 2SLS and GMM estimators analyzed below to be well defined asymptotically. The existence of higher-than-second moments in Assumption 8(iii) is required for the GMM estimation to apply a central limit theorem (CLT) for the linear and quadratic form, which is an extension of Theorem 1 in Kelejian and Prucha (2001). For the 2SLS estimations, the existence of the second moments would be sufficient.

3 Identification

Before discussing how to estimate the joint model (1), it is important to make sure that the parameters are identified. Since we are only interested in estimating $\boldsymbol{\delta} = \left(\rho, \boldsymbol{\beta}'\right)'$, we will derive the

[12] Also see Kapetanios et al. (2011) and Chudik and Pesaran (2015a) for discussions about the Common Correlated Effects (CCE) estimators in the rank deficiency case.

identification conditions of $\boldsymbol{\delta}$ assuming the factors are known.[13] It should be noted that whether the factors are observable will not affect the identification conditions. If there are unobserved factors, replacing them with certain proxies will only affect the consistency and efficiency properties of an estimator. Furthermore, as has been seen from (9), the unknown factors can be well approximated by cross-sectional averages for all values of ρ and $\boldsymbol{\beta}$ under the given assumptions, with an approximation error of order $O_p(1/\sqrt{N})$. Hence, the following analysis on the identification problem is undertaken conditional on observable factors. We will begin by examining SAR models with factors but without exogenous explanatory variables, $\mathbf{x}_{it}$, and return to models with $\mathbf{x}_{it}$ afterwards.

Now let us consider the following model,

$$y_{it} = \rho y_{it}^* + \boldsymbol{\gamma}_i'\mathbf{f}_t + e_{it},\ i = 1, 2, \ldots, N;\ t = 1, 2, \ldots, T, \tag{12}$$

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 $0 < \sigma^2 < K$. Writing (12) in stacked form, we have

$$\mathbf{y}_{.t} = \rho\mathbf{y}_{.t}^* + \boldsymbol{\Gamma}\mathbf{f}_t + \mathbf{e}_{.t},\ t = 1, 2, \ldots, T,$$

where $\mathbf{y}_{.t}^* = \mathbf{W}\mathbf{y}_{.t} = (y_{1t}^*, y_{2t}^*, \ldots, y_{Nt}^*)'$, $\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \ldots, \boldsymbol{\gamma}_N)'$ is an $N \times m$ matrix of factor loadings, and $\mathbf{e}_{.t} = (e_{1t}, e_{2t}, \ldots, e_{Nt})'$. Define $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1', \boldsymbol{\gamma}_2', \ldots, \boldsymbol{\gamma}_N')'$, and let $\boldsymbol{\varphi}_0 = \left(\rho_0, \boldsymbol{\gamma}_0', \sigma_0^2\right)'$ denote the true value of $\boldsymbol{\varphi} = \left(\rho, \boldsymbol{\gamma}', \sigma^2\right)'$. We adopt the most general identification framework based on the likelihood function proposed by Rothenberg (1971). The (quasi) log-likelihood function of (12) is given by

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

and it follows that

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

where

$$\mathbf{H}_f\left(\rho_0, \boldsymbol{\gamma}_0'\right) = (NT)^{-1}E_0\sum_{t=1}^{T}\left(\boldsymbol{J}_{0,t}'\boldsymbol{J}_{0,t}\right), \quad \boldsymbol{J}_{0,t} = \begin{pmatrix} \mathbf{G}_0\boldsymbol{\Gamma}_0\mathbf{f}_t, & \mathbf{F}_t \end{pmatrix}, \tag{13}$$

$\mathbf{G}(\rho) = \mathbf{W}\mathbf{S}^{-1}(\rho)$, $\mathbf{G}_0 = \mathbf{G}(\rho_0) = \mathbf{W}\mathbf{S}_0^{-1}$, $\mathbf{F}_t = \mathbf{I}_N \otimes \mathbf{f}_t'$, and for the discussion of identification, we use E_0 to emphasize that the expectation is calculated using the true values of the parameters.

[13] The factor loadings are identified up to a rotation if factors are unobserved.

Letting $Q_{NT}(\boldsymbol{\psi}) = (NT)^{-1} E_0 [l(\boldsymbol{\varphi}_0) - l(\boldsymbol{\varphi})]$, 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

$$\begin{aligned} Q_{NT}(\boldsymbol{\psi}) = & -\frac{1}{2}[ln(1-\vartheta)+\vartheta] - \frac{1}{N} ln|\mathbf{I}_N - d\mathbf{G}_0| - \frac{1}{N}(1-\vartheta) d tr(\mathbf{G}_0) + \frac{1}{2}(1-\vartheta)d^2 \frac{tr(\mathbf{G}_0\mathbf{G}_0')}{N} \\ & + \frac{1}{2}\sigma_0^2(1-\vartheta)\left(d, \boldsymbol{\zeta}'\right)\mathbf{H}_f(\rho_0, \boldsymbol{\gamma}_0')\left(d, \boldsymbol{\zeta}'\right)'. \end{aligned} \quad (14)$$

Then, by a mean value expansion, and noting that $\partial Q_{NT}(\mathbf{0})/\partial\boldsymbol{\psi} = \mathbf{0}$, we have $Q_N(\boldsymbol{\psi}) = \frac{1}{2}\boldsymbol{\psi}'\boldsymbol{\Lambda}_{f,NT}\left(\bar{\boldsymbol{\psi}}\right)\boldsymbol{\psi}$, where $\boldsymbol{\Lambda}_{f,NT}(\boldsymbol{\psi}) = \partial^2 Q_{NT}(\boldsymbol{\psi})/\partial\boldsymbol{\psi}\boldsymbol{\psi}'$, a detailed expression of which is given by (S.15) in the Online Supplement. $\bar{\boldsymbol{\psi}} = \left(\bar{d}, \bar{\boldsymbol{\zeta}}', \bar{\vartheta}\right)' = \left[\bar{\rho} - \rho_0, \bar{\boldsymbol{\gamma}}' - \boldsymbol{\gamma}_0', (\bar{\sigma}^2 - \sigma_0^2)/\bar{\sigma}^2\right]'$, where $\bar{\rho}$, $\bar{\boldsymbol{\gamma}}$, and $\bar{\sigma}^2$ 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, the parameters $\boldsymbol{\psi}_0$ 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 (S.16) in the Online Supplement. This condition can be further simplified after some algebra.[14] We formally state the results in the following proposition.

Proposition 1. *Consider the model given by (12). For all N (including $N \to \infty$) and all T, the true parameter values ρ_0, $\boldsymbol{\gamma}_0$, and σ_0^2 are locally identified if and only if*

$$h_g \equiv \frac{tr\left(\mathbf{G}_0^2 + \mathbf{G}_0\mathbf{G}_0'\right)}{N} - \frac{2\left[tr\left(\mathbf{G}_0\right)\right]^2}{N^2} > 0, \quad (15)$$

and $T^{-1}E_0\left(\mathbf{f}_t\mathbf{f}_t'\right)$ is positive definite.

Notice that model (12) reduces to a pure SAR model if there are no common factors; the identification condition would become $h_g > 0$, for all N (including $N \to \infty$). This condition is in line with the findings in a recent study by Aquaro et al. (2015), who investigate 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 given by (15). To further our understanding of (15), we make the following four observations.

First, it is worth pointing out that a necessary condition for (15) is that there exists an $\varepsilon > 0$ such that

$$N^{-1} tr\left(\mathbf{G}_0\mathbf{G}_0'\right) > \varepsilon > 0, \text{ for all } N, \text{ including } N \to \infty. \quad (16)$$

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

$$\begin{aligned} \frac{tr\left(\mathbf{G}_0^2 + \mathbf{G}_0\mathbf{G}_0'\right)}{N} - \frac{2\left[tr\left(\mathbf{G}_0\right)\right]^2}{N^2} &= \left\{\frac{tr\left(\mathbf{G}_0\mathbf{G}_0'\right)}{N} - \frac{\left[tr\left(\mathbf{G}_0\right)\right]^2}{N^2}\right\} + \left\{\frac{tr\left(\mathbf{G}_0^2\right)}{N} - \frac{\left[tr\left(\mathbf{G}_0\right)\right]^2}{N^2}\right\} \\ &\leq 2\left\{\frac{tr\left(\mathbf{G}_0\mathbf{G}_0'\right)}{N} - \frac{\left[tr\left(\mathbf{G}_0\right)\right]^2}{N^2}\right\}. \end{aligned}$$

[14] See the theory section of Online Supplement for details.

Therefore, for (15) to hold it is necessary that

$$\frac{tr\left(\mathbf{G}_0\mathbf{G}_0'\right)}{N} > \frac{\left[tr\left(\mathbf{G_0}\right)\right]^2}{N^2}. \tag{17}$$

However, 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, (16) 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 $\mathbf{I}_N$, namely, $\mathbf{G}_0 \neq c\mathbf{I}_N$ for all $c \neq 0$.

Second, under Assumption 6, a necessary and sufficient condition for (16) is that there exists an $\varepsilon > 0$ such that

$$N^{-1}tr\left(\mathbf{W}'\mathbf{W}\right) > \varepsilon > 0, \text{ for all } N, \text{ including } N \to \infty. \tag{18}$$

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

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

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[15]

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

which establishes necessity, and

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

which establishes sufficiency. As a simple necessary condition for identification, (18) does not depend on any unknown parameters and can be easily employed to check identifiability in practice.

Third, (16) is both a necessary and a sufficient identification condition if $\rho_0 = 0$. This can be seen by replacing $\mathbf{G}_0$ with $\mathbf{W}$ in (15) and by using $tr(\mathbf{G}_0) = 0$.

Finally, it should be noted that the condition (18) requires $N^{-1}tr\left(\mathbf{W}'\mathbf{W}\right)$ to be strictly positive for $N \to \infty$. This is an important consideration because the distinction between strong and weak cross-sectional dependence relies on N approaching infinity (Chudik et al., 2011). Notice that model (12) can be seen as a special case of the spatial Durbin models if there are no common factors. Lee and Yu (2016) investigates the identification conditions of these models but restrict their attention to finite sample sizes. The authors conclude that the parameter ρ_0 is identifiable if $\mathbf{I}_N$, $\mathbf{W} + \mathbf{W}'$, and $\mathbf{W}'\mathbf{W}$ are linearly independent. However, it is possible that this condition is met whereas

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

(18) is violated as $N \to \infty$. In such a case, our findings suggest that ρ_0 cannot be identified. An example is provided in Section 5.1 to demonstrate the necessity of (18) for identification.

We now proceed to include exogenous regressors $\mathbf{x}_{it}$ in (12) and consider the following model,

$$y_{it} = \rho y_{it}^* + \boldsymbol{\beta}'\mathbf{x}_{it} + \boldsymbol{\gamma}_i'\mathbf{f}_t + e_{it}. \tag{19}$$

In contrast with model (1), here we assume that $\mathbf{x}_{it}$ are uncorrelated with $\mathbf{f}_t$ for all i and t, and $e_{it} \sim IIDN(0, \sigma^2)$. With a slight abuse of notation, we use the same letter $\boldsymbol{\varphi}$ to denote the parameters of this model, $\boldsymbol{\varphi} = \left(\rho, \boldsymbol{\beta}', \boldsymbol{\gamma}', \sigma^2\right)'$, and their true values are denoted by $\boldsymbol{\varphi}_0 = \left(\rho_0, \boldsymbol{\beta}_0', \boldsymbol{\gamma}_0', \sigma_0^2\right)'$. By similar reasoning, we proclaim the following identification proposition, the proof of which is provided in the Online Supplement.

Proposition 2. *Consider the model given by (19), 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 identified if $h_g > 0$, or/and if $\mathbf{H}\left(\rho_0, \boldsymbol{\beta}_0'\right)$ is positive definite, where h_g is given by (15),*

$$\mathbf{H}\left(\rho_0, \boldsymbol{\beta}_0'\right) = (NT)^{-1} E_0\left(\mathbf{L}_0'\mathbf{L}_0\right), \tag{20}$$

$$\mathbf{L}_0 = \left(\mathbf{G}_0^b\mathbf{X}\boldsymbol{\beta}_0,\ \mathbf{X}\right),\ \ \textit{and}\ \ \mathbf{G}_0^b = \mathbf{I}_T \otimes \mathbf{G}_0. \tag{21}$$

Provided that ρ_0 is identifiable, the parameter vector $\boldsymbol{\beta}_0$ is identified if $(NT)^{-1} E_0\left(\mathbf{X}'\mathbf{X}\right)$ is positive definite. The vector $\boldsymbol{\gamma}_0$ is identified if $T^{-1}E_0\left(\mathbf{f}_t\mathbf{f}_t'\right)$ is positive definite.

Remark 6. Note that if $\mathbf{H}\left(\rho_0, \boldsymbol{\beta}_0'\right)$ is positive definite, both ρ_0 and $\boldsymbol{\beta}_0$ are identified; if it is not, the identification of ρ_0 can be achieved by $h_g > 0$. Comparing with the identification conditions for the pure SAR model, including individual-specific exogenous variables $\mathbf{x}_{it}$ introduces an additional means to identify ρ_0; however, including common factors does not help. This is not surprising, because common factors do not contain information regarding cross-sectional variations.

Remark 7. If there were no common factors, model (1) would reduce to a SAR model with exogenous regressors. Proposition 2 provides the identification conditions of parameters ρ_0, $\boldsymbol{\beta}_0$ and σ_0^2. Note that these conditions are valid even if $N \to \infty$.

Finally, let us return to model (1). Writing it in stacked form for each time period, we obtain

$$\mathbf{y}_{.t} = \rho\mathbf{y}_{.t}^* + \mathbf{X}_{.t}\boldsymbol{\beta} + \boldsymbol{\Gamma}\mathbf{f}_t + \mathbf{e}_{.t}, \quad t = 1, 2, \ldots, T. \tag{22}$$

Supposing that we are only interested in identifying ρ_0 and $\boldsymbol{\beta}_0$, as is the case in the following analysis, we can remove the effects of $\mathbf{f}_t$ by premultiplying (22) with $\mathbf{M}_f$. The identification conditions can be established as a corollary to Proposition 2.

Corollary 1. *Consider the model given by (1). For all N (including $N \to \infty$) and all T, the true parameter value ρ_0 is locally identified if $h_g > 0$, or/and if $\mathring{\mathbf{H}}\left(\rho_0, \boldsymbol{\beta}_0'\right)$ is positive definite, where h_g*

is given by (15) and $\mathring{\mathbf{H}}\left(\rho_0, \boldsymbol{\beta}_0'\right)$ *is defined by*

$$\mathring{\mathbf{H}}(\rho_0, \boldsymbol{\beta}_0') = (NT)^{-1} E_0 \left(\mathbf{L}_0' \mathbf{M}_f^b \mathbf{L}_0\right). \tag{23}$$

Provided that ρ_0 *is identifiable, the parameter vector* $\boldsymbol{\beta}_0$ *is identified if* $(NT)^{-1} E_0 \left(\mathbf{X}' \mathbf{M}_f^b \mathbf{X}\right)$ *is positive definite, which is ensured if* $\mathring{\mathbf{H}}\left(\rho_0, \boldsymbol{\beta}_0'\right)$ *is positive definite.*

4 Estimation

Having established the identification conditions, we now turn to considering the estimation of model (1). We suggest three estimation methods, including the 2SLS, Best 2SLS, and GMM estimations. This section formally establishes the asymptotic distributions of these estimators.

4.1 2SLS Estimation

The first estimation method we propose is the 2SLS estimation using the instrumental variables, $\mathbf{Q}$, as specified in Assumption 7. As before, $\boldsymbol{\delta}_0 = (\rho_0, \boldsymbol{\beta}_0')'$ denotes the true parameter vector. The 2SLS estimator of $\boldsymbol{\delta}_0$, denoted by $\hat{\boldsymbol{\delta}}_{2sls}$, is defined as

$$\hat{\boldsymbol{\delta}}_{2sls} = \left(\mathbf{L}' \mathbf{P}_Q \mathbf{L}\right)^{-1} \mathbf{L}' \mathbf{P}_Q \mathbf{Y}, \tag{24}$$

where $\mathbf{P}_Q = \mathbf{M}^b \mathbf{Q} \left(\mathbf{Q}' \mathbf{M}^b \mathbf{Q}\right)^{-1} \mathbf{Q}' \mathbf{M}^b$, $\mathbf{L} = (\mathbf{Y}^*, \mathbf{X})$ and $\mathbf{Y}^* = (\mathbf{I}_T \otimes \mathbf{W}) \mathbf{Y}$. There are two ways to interpret (24). One way is to de-factor the data with cross-sectional averages, namely, $\mathring{\mathbf{Y}} = \mathbf{M}^b \mathbf{Y}$ and $\mathring{\mathbf{L}} = \mathbf{M}^b \mathbf{L}$, and then apply the standard 2SLS procedure to the de-factored observations $\mathring{\mathbf{Y}}$ and $\mathring{\mathbf{L}}$. Alternatively, the matrix $\mathbf{M}^b \mathbf{Q}$ can be directly considered as instruments.

We begin by showing that the 2SLS estimator, $\hat{\boldsymbol{\delta}}_{2sls}$, is consistent as $N \to \infty$, for T fixed or $T \to \infty$. To see this, note that

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

and then

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

Applying Lemma A6, we have

$$\frac{1}{NT}\mathbf{Q}'\mathbf{M}^b\mathbf{Q} = \frac{1}{NT}\mathbf{Q}'\mathbf{M}_f^b\mathbf{Q} + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right),$$
$$\frac{1}{NT}\mathbf{Q}'\mathbf{M}^b\mathbf{L} = \frac{1}{NT}\mathbf{Q}'\mathbf{M}_f^b\mathbf{L}_0 + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right),$$

where $\mathbf{L}_0$ is given by (21), and it follows that

$$\frac{1}{NT}\mathbf{L}'\mathbf{P}_Q\mathbf{L} = \frac{1}{NT}\mathbf{L}_0'\mathbf{P}_{Q,f}\mathbf{L}_0 + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right),$$

where $\mathbf{P}_{Q,f} = \mathbf{M}_f^b\mathbf{Q}\left(\mathbf{Q}'\mathbf{M}_f^b\mathbf{Q}\right)^{-1}\mathbf{Q}'\mathbf{M}_f^b$. Under Assumption 8, $\underset{N\to\infty}{p\lim}\,(NT)^{-1}\,\mathbf{L}_0'\mathbf{P}_{Q,f}\mathbf{L}_0$ exists and is nonsingular. Furthermore, we have shown in Lemma A6 that $\underset{N\to\infty}{p\lim}\,(NT)^{-1}\,\mathbf{Q}'\mathbf{M}_f^b\left[(\mathbf{I}_T\otimes\boldsymbol{\Gamma}_0)\mathbf{f}+\mathbf{e}\right] = 0$. As a result, $\hat{\boldsymbol{\delta}}_{2sls}$ is consistent for $\boldsymbol{\delta}_0$, as $N\to\infty$.

For the asymptotic distribution of $\hat{\boldsymbol{\delta}}_{2sls}$, we show in Appendix A that as $(N,T)\xrightarrow{j}\infty$ and $T/N\to 0$, the term $(NT)^{-1/2}\,\mathbf{Q}'\mathbf{M}^b\left[(\mathbf{I}_T\otimes\boldsymbol{\Gamma}_0)\,\mathbf{f}\right]$ converges in probability to zero, and $(NT)^{-1/2}\,\mathbf{Q}'\mathbf{M}^b\mathbf{e}$ tends toward a normal distribution. The relative rate of expansion of T and N is imposed to eliminate the nuisance parameters from the limiting distribution.

The following theorem summarizes the limiting distribution of the 2SLS estimator.

Theorem 1. *Consider the panel data model given by (1) and suppose that Assumptions 1, 2(i), and 3–8 hold. The 2SLS estimator, $\hat{\boldsymbol{\delta}}_{2sls}$, defined by (24), is consistent for $\boldsymbol{\delta}_0$, as $N\to\infty$, for T fixed or $T\to\infty$. Moreover, as $(N,T)\xrightarrow{j}\infty$ and $T/N\to 0$, we have*

$$\sqrt{NT}\left(\hat{\boldsymbol{\delta}}_{2sls}-\boldsymbol{\delta}_0\right)\xrightarrow{d} N\left(\mathbf{0},\boldsymbol{\Sigma}_{2sls}\right), \tag{25}$$

where

$$\boldsymbol{\Sigma}_{2sls} = \boldsymbol{\Psi}_{LPL}^{-1}\boldsymbol{\Omega}_{LPe}\boldsymbol{\Psi}_{LPL}^{-1}, \tag{26}$$

$$\boldsymbol{\Psi}_{LPL} = \underset{N,T\to\infty}{plim}\,(NT)^{-1}\,\mathbf{L}_0'\mathbf{P}_{Q,f}\mathbf{L}_0,\quad \boldsymbol{\Omega}_{LPe} = \boldsymbol{\Psi}_{QML}'\boldsymbol{\Psi}_{QMQ}^{-1}\boldsymbol{\Omega}_{QMe}\boldsymbol{\Psi}_{QMQ}^{-1}\boldsymbol{\Psi}_{QML}, \tag{27}$$

$$\boldsymbol{\Psi}_{QMQ} = \underset{N,T\to\infty}{plim}\,(NT)^{-1}\,\mathbf{Q}'\mathbf{M}_f^b\mathbf{Q},\qquad \boldsymbol{\Psi}_{QML} = \underset{N,T\to\infty}{plim}\,(NT)^{-1}\,\mathbf{Q}'\mathbf{M}_f^b\mathbf{L}_0, \tag{28}$$

$$\boldsymbol{\Omega}_{QMe} = \lim_{N\to\infty}\left(N^{-1}\sum_{i=1}^{N}\boldsymbol{\Omega}_{iQMe}\right),\qquad \boldsymbol{\Omega}_{iQMe} = \underset{T\to\infty}{p\lim}\,T^{-1}\mathbf{Q}_{i.}'\mathbf{M}_f\boldsymbol{\Omega}_{e,i}\mathbf{M}_f\mathbf{Q}_{i.}, \tag{29}$$

$\boldsymbol{\Omega}_{e,i} = E\left(\mathbf{e}_{i.}\mathbf{e}_{i.}'\right)$, *and* $\mathbf{Q}_{i.} = \left(\mathbf{Q}_{i1},\mathbf{Q}_{i2},\ldots,\mathbf{Q}_{iT}\right)'$.

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

$$\hat{\boldsymbol{\Sigma}}_{2sls} = \left(\frac{1}{NT}\mathbf{L}\mathbf{P}_Q\mathbf{L}\right)^{-1}\hat{\boldsymbol{\Omega}}_{LPe}\left(\frac{1}{NT}\mathbf{L}\mathbf{P}_Q\mathbf{L}\right)^{-1}, \tag{30}$$

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

$$\hat{\mathbf{\Omega}}_{LPe} = N^{-1}\sum_{i=1}^{N}\hat{\mathbf{\Omega}}_{iLPe}, \tag{31}$$
$$\hat{\mathbf{\Omega}}_{iLPe} = \hat{\mathbf{\Omega}}_{iLPe,0} + \sum_{h=1}^{M_l}\left(1-\frac{h}{M_l+1}\right)\left(\hat{\mathbf{\Omega}}_{iLPe,h} + \hat{\mathbf{\Omega}}'_{iLPe,h}\right),$$
$$\hat{\mathbf{\Omega}}_{iLPe,h} = T^{-1}\sum_{t=h+1}^{T}\hat{e}_{it}\hat{e}_{i,t-h}\hat{\boldsymbol{l}}_{it}\hat{\boldsymbol{l}}'_{i,t-h},$$

where M_l is the the window size (or bandwidth) of the Bartlett kernel, $\hat{\mathbf{e}} = \mathbf{M}^b\left(\mathbf{Y} - \mathbf{L}\hat{\boldsymbol{\delta}}_{2sls}\right) = (\hat{\mathbf{e}}'_{.1}, \hat{\mathbf{e}}'_{.2}, \ldots, \hat{\mathbf{e}}'_{.T})'$, $\hat{e}_{it}$ is the t^{th} element of $\hat{\mathbf{e}}_{.t}$, $\hat{\mathbf{L}} = \mathbf{P}_Q\mathbf{L} = \left(\hat{\mathbf{L}}'_{.1}, \hat{\mathbf{L}}'_{.2}, \ldots, \hat{\mathbf{L}}'_{.T}\right)'$, and $\hat{\boldsymbol{l}}'_{it}$ is the i^{th} row of $\hat{\mathbf{L}}_{.t}$.

Remark 8. Although our interest lies in the parameters $\boldsymbol{\delta}$, we can gain insight into the variability of the factor loadings after obtaining estimates of $\boldsymbol{\delta}$. This can be done by regressing $y_{it} - \boldsymbol{l}'_{it}\hat{\boldsymbol{\delta}}$ on $\bar{\mathbf{z}}_{.t}$ and an intercept for each cross-section unit i, where $\boldsymbol{l}_{it} = (y^*_{it}, \mathbf{x}'_{it})'$, and $\bar{\mathbf{z}}_{.t} = (\bar{y}_{.t}, \bar{\mathbf{x}}'_{.t})'$.

4.2 Best 2SLS Estimation

Having established the asymptotic distribution of the 2SLS estimator, the question then naturally arises whether optimal instrumental variables are available for model (1). An instrument is considered optimal or "best" if it produces an estimator that has the smallest asymptotic variance among all the IV estimators for the model. For cross-sectional spatial models, Lee (2003) suggests a best generalized spatial 2SLS estimator, and he shows that it is asymptotically optimal under a set of regularity conditions. In this section, we generalize this estimation procedure to spatial models with common factors. Specifically, let $\hat{\boldsymbol{\delta}} = \left(\hat{\rho}, \hat{\boldsymbol{\beta}}\right)$ denote some consistent initial estimate of $\boldsymbol{\delta}_0$, possibly obtained by the 2SLS estimation described in the previous section. We will investigate if the IV estimator, $\hat{\boldsymbol{\delta}}_{b2sls}$ can achieve the smallest asymptotic variance for model (1), where

$$\hat{\boldsymbol{\delta}}_{b2sls} = \left(\hat{\mathbf{Q}}^{*\prime}\mathbf{L}\right)^{-1}\hat{\mathbf{Q}}^{*\prime}\mathbf{Y}, \tag{32}$$

$$\hat{\mathbf{Q}}^* = \mathbf{M}^b\left[\left(\mathbf{I}_T \otimes \hat{\mathbf{G}}\right)\mathbf{X}\hat{\boldsymbol{\beta}},\ \mathbf{X}\right], \tag{33}$$

and $\hat{\mathbf{G}} = \mathbf{G}(\hat{\rho})$. We refer to $\hat{\boldsymbol{\delta}}_{b2sls}$ as the best 2SLS (B2SLS) estimator and $\hat{\mathbf{Q}}^*$ as the (feasible) best IV.

The intuition behind the formulation of $\hat{\mathbf{Q}}^*$ is to exploit the part of $\mathbf{Y}^*$ that is uncorrelated with the errors. To see this, suppose for simplicity that there are no common factors. The structural equation (22) implies the following reduced form equation: $\mathbf{y}_{.t} = \mathbf{S}_0^{-1}\mathbf{X}_{.t}\boldsymbol{\beta}_0 + \mathbf{S}_0^{-1}\mathbf{e}_{.t}$, which further leads to $\mathbf{y}^*_{.t} = \mathbf{G}_0\mathbf{X}_{.t}\boldsymbol{\beta}_0 + \mathbf{G}_0\mathbf{e}_{.t}$. It is readily seen that the term $\mathbf{G}_0\mathbf{X}_{.t}\boldsymbol{\beta}$ is correlated with $\mathbf{y}^*_{.t}$ but uncorrelated with $\mathbf{e}_{.t}$ given that $\mathbf{X}_{.t}$ is exogenous. Since $\mathbf{G}_0\mathbf{X}_{.t}\boldsymbol{\beta}_0$ depends on the unknown parameters ρ_0 and $\boldsymbol{\beta}_0$, a feasible IV for $\mathbf{y}^*_{.t}$ can be constructed as $\hat{\mathbf{G}}\mathbf{X}_{.t}\hat{\boldsymbol{\beta}}$. Accordingly, the B2SLS estimation can be implemented in two steps: first obtaining some preliminary consistent estimates of the parameters, and then conducting an IV estimation using the best IV based on the first-step estimates. A similar argument applies to model (1) with common factors. Equation (33) indicates that in constructing the best IV, we need to filter out the common effects from the observations using the de-factoring matrix $\mathbf{M}^b$.

The following theorem states the asymptotic properties of the B2SLS estimator and shows that it is the best IV estimator provided that the error terms are independently and identically distributed. The proof is given in Appendix A.

Theorem 2. *Consider the panel data model given by (1). Suppose that Assumptions 1, 2(i), and 3–8 hold and $\mathring{\mathbf{H}}(\rho_0, \boldsymbol{\beta}_0')$ is positive definite, where $\mathring{\mathbf{H}}(\rho_0, \boldsymbol{\beta}_0')$ is given by (23). Then, the best 2SLS (B2SLS) estimator, $\hat{\boldsymbol{\delta}}_{b2sls}$, defined by (32), is consistent for $\boldsymbol{\delta}_0$, as $N \to \infty$, for T fixed or $T \to \infty$; as $(N,T) \xrightarrow{j} \infty$ and $T/N \to 0$, it has the following distribution:*

$$\sqrt{NT}\left(\hat{\boldsymbol{\delta}}_{b2sls} - \boldsymbol{\delta}_0\right) \xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{\Sigma}_{b2sls}\right), \tag{34}$$

where

$$\boldsymbol{\Sigma}_{b2sls} = \boldsymbol{\Psi}_{LML}^{-1}\boldsymbol{\Omega}_{LMe}\boldsymbol{\Psi}_{LML}^{-1}, \tag{35}$$

$$\boldsymbol{\Psi}_{LML} = \underset{N,T\to\infty}{plim}\ (NT)^{-1}\,\mathbf{L}_0'\mathbf{M}_f^b\mathbf{L}_0,$$

$$\boldsymbol{\Omega}_{LMe} = \lim_{N\to\infty}\left(N^{-1}\sum_{i=1}^{N}\boldsymbol{\Omega}_{iLMe}\right), \quad \boldsymbol{\Omega}_{iLMe} = \underset{T\to\infty}{plim}\, T^{-1}\mathbf{L}_{0,i.}'\mathbf{M}_f\boldsymbol{\Omega}_{e,i}\mathbf{M}_f\mathbf{L}_{0,i.}, \tag{36}$$

$\mathbf{L}_0$ is given by (21), $\mathbf{L}_{0,i.} = (\boldsymbol{l}_{0,i1}, \boldsymbol{l}_{0,i2}, \ldots, \boldsymbol{l}_{0,iT})'$, and $\boldsymbol{l}_{0,it}'$ is the $[N(t-1)+i]^{th}$ row of $\mathbf{L}_0$. The B2SLS estimator is the best IV estimator if the disturbances $\{e_{it}\}$ are independently and identically distributed with mean zero and variance σ_e^2.

Note that under Assumption 6, $(\mathbf{I}_N - \rho_0\mathbf{W})^{-1}\mathbf{X}_{.t}\boldsymbol{\beta}_0 = \sum_{s=1}^{N}\rho_0^s\mathbf{W}^s\mathbf{X}_{.t}\boldsymbol{\beta}_0$. Hence, the term $\mathbf{G}_0\mathbf{X}_{.t}\boldsymbol{\beta}_0$ can be approximated by linear combinations of $\mathbf{X}_{.t}\boldsymbol{\beta}$, $\mathbf{W}\mathbf{X}_{.t}\boldsymbol{\beta}$, $\mathbf{W}^2\mathbf{X}_{.t}\boldsymbol{\beta}, \ldots$. Clearly, the higher the power of $\mathbf{W}$ included, the better the approximation. However, the efficiency gain by including more instruments may not be significant. In practice, the 2SLS estimator with instruments $\left(\mathbf{X}_{.t}, \mathbf{W}\mathbf{X}_{.t}, \mathbf{W}^2\mathbf{X}_{.t}\right)$ is often found to perform well enough. The finite sample properties of $\hat{\boldsymbol{\delta}}_{b2sls}$ will be compared with that of $\hat{\boldsymbol{\delta}}_{2sls}$ using Monte Carlo techniques in Section 5.

4.3 GMM Estimation

The third estimator we propose is the GMM estimator that utilizes quadratic moment conditions based on the properties of the idiosyncratic errors in addition to the 2SLS-type linear moments. The use of the quadratic moments for SAR models is proposed by Lee (2007) and later extended by Lin and Lee (2010) and Lee and Yu (2014). The advantages of adopting the quadratic moments lie both in improving efficiency and in making it possible to estimate the spatial autoregressive coefficient when there are no exogenous regressors. In this section, we show that this idea can be extended to spatial models with common factors.

Specifically, we consider the following sample moment conditions, which consist of r quadratic moments and q linear moments:[16]

$$\mathbf{g}_{NT}(\boldsymbol{\delta}) = \begin{pmatrix} \boldsymbol{\xi}'(\boldsymbol{\delta})\mathbf{M}^b\mathbf{P}_1^b\mathbf{M}^b\boldsymbol{\xi}(\boldsymbol{\delta}) \\ \vdots \\ \boldsymbol{\xi}'(\boldsymbol{\delta})\mathbf{M}^b\mathbf{P}_r^b\mathbf{M}^b\boldsymbol{\xi}(\boldsymbol{\delta}) \\ \mathbf{Q}'\mathbf{M}^b\boldsymbol{\xi}(\boldsymbol{\delta}) \end{pmatrix}, \tag{37}$$

where $\mathbf{M}^b$ is the de-factoring matrix defined by (11), $\boldsymbol{\xi}(\boldsymbol{\delta})$ is the vector of residuals given by

$$\boldsymbol{\xi}(\boldsymbol{\delta}) = [\mathbf{I}_T \otimes \mathbf{S}(\rho)]\,\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}, \tag{38}$$

and $\boldsymbol{\delta} = (\rho, \boldsymbol{\beta}')'$ represents the unknown parameters in the parameter space, $\boldsymbol{\Delta}_{sp}$. For $l = 1, 2, \ldots, r$, we define $\mathbf{P}_l^b = \mathbf{I}_T \otimes \mathbf{P}_l$, where $\mathbf{P}_l = (p_{l,ij})$ is an N-dimensional square matrix with zero diagonal, namely, $diag\,(\mathbf{P}_l) = (p_{l,11}, p_{l,22}, \ldots, p_{l,NN})' = \mathbf{0}$.

Intuitively, the idea behind the quadratic moments is to use some matrix $\mathbf{P}_l$ to eliminate the correlations among the elements of the idiosyncratic error $\mathbf{e}_{.t}$ in order to achieve zero expectations. To see this, consider the simpler scenario where there are no common factors: the l^{th} population quadratic moment at $\boldsymbol{\delta}_0$ will be reduced to

$$E\left(\mathbf{e}'\mathbf{P}_l^b\mathbf{e}\right) = \sum_{i=1}^{N}\sum_{j=1}^{N} p_{l,ji}E\left(\mathbf{e}_{i.}'\mathbf{e}_{j.}\right) = \sum_{i=1}^{N} p_{l,ii}E\left(\mathbf{e}_{i.}'\mathbf{e}_{i.}\right) = 0,$$

where $p_{l,ji}$ is the $(j,i)^{th}$ element of matrix $\mathbf{P}_l$, and the last equality follows from the assumption that $diag(\mathbf{P}_l) = \mathbf{0}$. It is worth noting that the moment conditions are built on the key assumption of the cross-sectional uncorrelatedness between e_{it} and $e_{jt'}$, $i \neq j$, for all t and t'. Also note that since we allow for unknown heteroskedasticity, we need all diagonal elements of $\mathbf{P}_l$ to be zero in order 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 GMM estimator, $\hat{\boldsymbol{\delta}}_{GMM}$, is then defined as

$$\hat{\boldsymbol{\delta}}_{GMM} = \underset{\boldsymbol{\delta}\in\boldsymbol{\Delta}_{sp}}{\operatorname{argmin}}\; \mathbf{g}_{NT}'(\boldsymbol{\delta})\,\mathbf{A}_{NT}^{w\prime}\mathbf{A}_{NT}^{w}\mathbf{g}_{NT}(\boldsymbol{\delta}), \tag{39}$$

where $\mathbf{g}_{NT}(\boldsymbol{\delta})$ is given by (37), and $\mathbf{A}_{NT}^{w}$ is a constant full row rank matrix of $k_a \times (r+q)$ dimension, where $k_a \geq k+1$, and $\mathbf{A}_{NT}^{w\prime}\mathbf{A}_{NT}^{w}$ is assumed to converge to a positive definite matrix $\mathbf{A}^{w\prime}\mathbf{A}$. The following additional assumption is needed for the asymptotic analysis of the GMM estimator.

[16] We use the aggregated moment conditions over time instead of a moment condition for each period separately, since the latter approach may induce the many-moment bias problem and is beyond the scope of the current paper. See Lee and Yu (2014) for a discussion of this issue for spatial models.

Assumption 9. *The matrices $\mathbf{P}_l$, for $l = 1, 2, \ldots, r$, used in the moment conditions given by (37), are nonstochastic and have bounded maximum row and column sum norms, namely, $||\mathbf{P}_l||_\infty < K$ and $||\mathbf{P}_l||_1 < K$.*

Theorem 3. *Consider the panel data model given by (1) and suppose that Assumptions 1–9 hold. The GMM estimator, $\hat{\boldsymbol{\delta}}_{GMM}$, defined by (39), is consistent for $\boldsymbol{\delta}_0$ as $N \to \infty$, for T fixed or $T \to \infty$. Furthermore, as $(N, T) \overset{j}{\to} \infty$ and $T/N \to 0$, we have*

$$\sqrt{NT}\left(\hat{\boldsymbol{\delta}}_{GMM} - \boldsymbol{\delta}_0\right) \overset{d}{\to} N\left(\mathbf{0}, \boldsymbol{\Sigma}_{GMM}\right), \tag{40}$$

where

$$\boldsymbol{\Sigma}_{GMM} = \left(\mathbf{D}'\mathbf{A}^{w\prime}\mathbf{A}^w\mathbf{D}\right)^{-1}\mathbf{D}'\mathbf{A}^{w\prime}\mathbf{A}^w\boldsymbol{\Sigma}_g\mathbf{A}^{w\prime}\mathbf{A}^w\mathbf{D}\left(\mathbf{D}'\mathbf{A}^{w\prime}\mathbf{A}^w\mathbf{D}\right)^{-1}, \tag{41}$$

$$\mathbf{D} = \left(\mathbf{D}_p', \boldsymbol{\Psi}_{QML}'\right)', \qquad \mathbf{D}_p = \begin{pmatrix} \mathbf{d}_p, & \mathbf{0}_{r\times k} \end{pmatrix}, \tag{42}$$

$$\mathbf{d}_p = \lim_{N\to\infty} N^{-1}\left(\sum_{i=1}^N \tilde{g}_{ii,1}^s\sigma_i^2,\ \sum_{i=1}^N \tilde{g}_{ii,2}^s\sigma_i^2,\ \ldots,\ \sum_{i=1}^N \tilde{g}_{ii,r}^s\sigma_i^2\right)', \tag{43}$$

$$\boldsymbol{\Sigma}_g = \begin{pmatrix} \boldsymbol{\Sigma}_p & \mathbf{0}_{r\times(k+1)} \\ \mathbf{0}_{(k+1)\times p} & \boldsymbol{\Omega}_{QMe} \end{pmatrix}, \tag{44}$$

$$\boldsymbol{\Sigma}_p = \lim_{N\to\infty} N^{-1}\begin{pmatrix} tr\left[\left(\mathbf{P}_1 \odot \mathbf{P}_1^s\right)\boldsymbol{\Sigma}_e\right] & \cdots & tr\left[\left(\mathbf{P}_1 \odot \mathbf{P}_r^s\right)\boldsymbol{\Sigma}_e\right] \\ \vdots & & \vdots \\ tr\left[\left(\mathbf{P}_r \odot \mathbf{P}_1^s\right)\boldsymbol{\Sigma}_e\right] & \cdots & tr\left[\left(\mathbf{P}_r \odot \mathbf{P}_r^s\right)\boldsymbol{\Sigma}_e\right] \end{pmatrix}, \tag{45}$$

where $\tilde{g}_{ii,l}^s$ ($l = 1, 2, \ldots, r$) is the i^{th} diagonal element of matrix $\tilde{\mathbf{G}}_l(\rho_0) = \mathbf{P}_l^s\mathbf{G}_0$, $\mathbf{P}_l^s = \mathbf{P}_l + \mathbf{P}_l'$, $\boldsymbol{\Sigma}_e = (\varsigma_{e,ij})$ is an $N \times N$ matrix of which the $(i,j)^{th}$ element is given by $\varsigma_{e,ij} = \lim_{T\to\infty} T^{-1} tr\left(\boldsymbol{\Omega}_{e,i}\boldsymbol{\Omega}_{e,j}\right)$, $\boldsymbol{\Psi}_{QML}$ and $\boldsymbol{\Omega}_{QMe}$ are given by (28) and (29), respectively, and $\odot$ denotes the Hadamard (or entrywise) product.

The (infeasible) efficient GMM estimator can be obtained using the optimal weighting matrix, $\boldsymbol{\Sigma}_g^{-1}$, in the usual fashion, namely,

$$\hat{\boldsymbol{\delta}}_{GMM}^* = \underset{\boldsymbol{\delta}\in\boldsymbol{\Delta}_{sp}}{\operatorname{argmin}}\ \mathbf{g}_{NT}'(\boldsymbol{\delta})\boldsymbol{\Sigma}_g^{-1}\mathbf{g}_{NT}(\boldsymbol{\delta}). \tag{46}$$

The asymptotic distribution of $\hat{\boldsymbol{\delta}}_{GMM}^*$ is formally stated in the next theorem.

Theorem 4. *Under the same assumptions as in Theorem 3, the efficient GMM estimator, $\hat{\boldsymbol{\delta}}_{GMM}^*$, defined by (46), has the following asymptotic distribution as $(N, T) \overset{j}{\to} \infty$ and $T/N \to 0$:*

$$\sqrt{NT}\left(\hat{\boldsymbol{\delta}}_{GMM}^* - \boldsymbol{\delta}_0\right) \overset{d}{\to} N\left(\mathbf{0}, \boldsymbol{\Sigma}_{GMM}^*\right), \tag{47}$$

where $\boldsymbol{\Sigma}_{GMM}^* = \left(\mathbf{D}'\boldsymbol{\Sigma}_g^{-1}\mathbf{D}\right)^{-1}$.

A consistent estimator of $\mathbf{\Sigma}_{GMM}$ can be obtained by replacing $\mathbf{D}$ and $\mathbf{\Sigma}_g$ in (41) with $\hat{\mathbf{D}}$ and $\hat{\mathbf{\Sigma}}_g$, respectively, where

$$\hat{\mathbf{D}} = \left(\hat{\mathbf{D}}_p', \hat{\mathbf{\Psi}}_{QML}'\right)', \quad \hat{\mathbf{D}}_p = \begin{pmatrix} \hat{\mathbf{d}}_p & \mathbf{0}_{r\times k} \end{pmatrix}, \quad \hat{\mathbf{\Psi}} = (NT)^{-1}\mathbf{Q}'\mathbf{M}^b\mathbf{L},$$

$$\hat{\mathbf{d}}_p = (NT)^{-1}\left(\sum_{i=1}^N \hat{\tilde{g}}_{ii,1}^s \hat{\mathbf{e}}_{i.}'\hat{\mathbf{e}}_{i.},\ \sum_{i=1}^N \hat{\tilde{g}}_{ii,2}^s \hat{\mathbf{e}}_{i.}'\hat{\mathbf{e}}_{i.},\ \ldots,\ \sum_{i=1}^N \hat{\tilde{g}}_{ii,r}^s \hat{\mathbf{e}}_{i.}'\hat{\mathbf{e}}_{i.}\right)',$$

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

$\hat{\mathbf{e}} = \mathbf{M}^b\left(\mathbf{Y} - \mathbf{L}\hat{\boldsymbol{\delta}}_{GMM}\right)$, $\hat{\tilde{g}}_{ii,l}^s$ is the i^{th} diagonal element of $\tilde{\mathbf{G}}_l(\hat{\rho})$,

$$\hat{s}_{e,ij} = T\hat{\gamma}_{e,i}(0)\hat{\gamma}_{e,j}(0) + 2\sum_{h=1}^{M_l}(T-h)\left(1 - \frac{h}{M_l+1}\right)\hat{\gamma}_{e,i}(h)\hat{\gamma}_{e,j}(h),$$

$\hat{\gamma}_{e,i}(h) = T^{-1}\sum_{t=h+1}^T \hat{e}_{it}\hat{e}_{i,t-h}$, and M_l is the maximum lag length (or window size). 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 $\hat{\mathbf{\Sigma}}_g^*$ would be computed using $\hat{\mathbf{e}}^* = \mathbf{M}^b\left(\mathbf{Y} - \mathbf{L}\hat{\boldsymbol{\delta}}_{GMM}^*\right)$ instead of $\hat{\mathbf{e}}$.

It is straightforward to see that the 2SLS estimator $\hat{\boldsymbol{\delta}}_{2sls}$ is asymptotically less efficient than $\hat{\boldsymbol{\delta}}_{GMM}$, since the latter makes use of quadratic moments in addition to the linear moments. 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

$$\mathbf{D}'\mathbf{\Sigma}_g^{-1}\mathbf{D} = \begin{pmatrix} \mathbf{d}_p'\mathbf{\Sigma}_p^{-1}\mathbf{d}_p & \mathbf{0}_{1\times k} \\ \mathbf{0}_{k\times 1} & \mathbf{0}_{k\times k} \end{pmatrix} + \frac{1}{NT}\left(\mathbf{Q}'\mathbf{M}_f^b\mathbf{L}_0\right)'\mathbf{\Omega}_{QMe}\left(\mathbf{Q}'\mathbf{M}_f^b\mathbf{L}_0\right). \tag{48}$$

It can be seen from (48) that, ideally, one should choose $\mathbf{P}_l$ $(l = 1, 2\ldots, r)$ to maximize $\mathbf{d}_p'\mathbf{\Sigma}_p^{-1}\mathbf{d}_p$. However, this term depends on the unknown variance structure of the disturbances. If the disturbances are independent and identically distributed (i.i.d.), it is known in the spatial literature that the best $\mathbf{P}_l$ within the class of matrices with zero diagonal is given by $\mathbf{P}^* = \mathbf{G}_0 - Diag(\mathbf{G}_0)$ (Lee, 2007; Lee and Yu, 2014). Using similar arguments, the results can be 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 can be obtained by minimizing the optimally weighted moments (37), where $\mathbf{P}$ is set to $\hat{\mathbf{P}}^* = \mathbf{G}(\hat{\rho}) - Diag\left(\mathbf{G}(\hat{\rho})\right)$, and $\mathbf{Q}$ is replaced by $\hat{\mathbf{Q}}^*$ given in (33). Nonetheless, in the presence of unknown heteroskedasticity and serial correlations, the BGMM estimator in general will not be the most efficient. This conclusion can be drawn by applying similar reasoning as in the proof of Theorem 2 for the B2SLS estimator. The present paper omits further discussions on the BGMM estimator in view of the strong assumption required for it to have optimal properties.

5 Monte Carlo Experiments

This section first provides Monte Carlo evidence in support of the identification conditions, then documents the finite sample properties of the proposed estimators under various specifications of the disturbance process and under different intensities of spatial dependence. It also compares the performance of the proposed estimators with that of alternative estimators.

5.1 Identification Experiments

We now construct an example to show that the condition given by (18), namely,

$$N^{-1}tr\left(\mathbf{W}'\mathbf{W}\right) > \varepsilon > 0, \text{ for all } N, \text{ including } N \to \infty, \tag{49}$$

is indeed necessary for identification. Consider the following data generating process (DGP),

$$y_{it} = \rho y_{it}^{*} + e_{it}, \tag{50}$$

for $i = 1, 2, \ldots, N$, and $t = 1, 2, \ldots, T$, where $y_{it}^{*} = \sum_{j=1}^{N} w_{ij} y_{jt}$ and $e_{it} \sim IIDN(0, \sigma^2)$. Suppose that $N_1 = \lfloor N^{\alpha} \rfloor$ rows of $\mathbf{W}$ are nonzero and that the other $N_2 = N - N_1$ rows are all zeros, in which $\lfloor N^{\alpha} \rfloor$ denotes the integer part of N^{α}, and α is a constant that does not depend on N and lies in the range$[0, 1]$. In other words, we allow the number of nonzero rows of $\mathbf{W}$ to rise more slowly than the sample size, N, and the rate at which it rises with N is measured by α.

Note that the identification condition, (18), fails to hold if $\alpha < 1$. To see this, there is no loss of generality in assuming that the first N_1 rows of $\mathbf{W}$ are nonzero, and it follows that

$$\frac{tr\left(\mathbf{W}'\mathbf{W}\right)}{N} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}^2}{N} = \frac{\sum_{i=1}^{N_1} K_i + \sum_{i=N_1+1}^{N} 0}{N} \leq K\frac{\lfloor N^{\alpha} \rfloor}{N} \leq KN^{\alpha-1},$$

where the second equality follows from $\sum_{j=1}^{N} w_{ij}^2 = K_i < \infty$, for all i. Hence, $N^{-1}tr(\mathbf{W}'\mathbf{W}) \to 0$, as $N \to \infty$, if $\alpha < 1$, and it approaches zero faster for smaller α.

In the Monte Carlo experiments, we consider the q-ahead-and-q-behind circular neighbors spatial weights, which are commonly employed in the literature. An m-ahead-and-m-behind matrix is motivated to capture spatial relations in which all units are located in a circle; the q units immediately ahead of and behind a particular unit are considered "neighbors" and assigned equal weights. For example, for the 2-ahead-and-2-behind spatial matrix, the i^{th} row of $\mathbf{W}$ has nonzero elements in the positions $i-2, i-1, i+1, i+2$, and each weigh $1/4$ due to row normalization. Without loss of generality, we adopt the 5-ahead-and-5-behind spatial weights in the first N_1 rows of $\mathbf{W}$, and we set the remaining rows to zeros.

It is worth noting that the identification condition for model (50) proposed by Lee and Yu (2016), which states that the matrices, $\mathbf{I}_N$, $\mathbf{W} + \mathbf{W}'$, and $\mathbf{W}'\mathbf{W}$ are linearly independent, is satisfied in this case. To see this, let c_1, c_2, and c_3 be constants such that

$$c_1 + 2c_2 w_{ii} + c_3 \sum_{k=1}^{N} w_{ki}^2 = 0, \quad \text{for all } i = 1, 2, \ldots, N, \tag{51}$$

$$c_2 \left(w_{ij} + w_{ji}\right) + c_3 \sum_{k=1}^{N} w_{ki} w_{kj} = 0, \quad \text{for all } i, j = 1, 2, \ldots, N, \text{ and } i \neq j. \tag{52}$$

Then $\mathbf{I}_N$, $\mathbf{W} + \mathbf{W}'$, and $\mathbf{W}'\mathbf{W}$ are linearly independent if and only if $c_1 = c_2 = c_3 = 0$. Suppose first that $c_3 = 0$. From (52) we must have $c_2 = 0$, since $w_{ij} + w_{ji} > 0$ exists for some i and j. Then, using (51), we obtain $c_1 = 0$. If, on the other hand, $c_3 \neq 0$, then it can be easily verified that there are no constants $\tilde{c}_1 > 0$ and $\tilde{c}_2 \neq 0$, such that

$$\begin{cases} \sum_{k=1}^{N} w_{ki}^2 = \tilde{c}_1, & \text{for all } i \\ \sum_{k=1}^{N} w_{ki} w_{kj} = \tilde{c}_2 (w_{ij} + w_{ji}), & \text{for all } i, j, \text{ and } i \neq j \end{cases}. \tag{53}$$

This establishes the linear independence of $\mathbf{I}_N$, $\mathbf{W} + \mathbf{W}'$, and $\mathbf{W}'\mathbf{W}$.

In sum, we have shown that the $\mathbf{W}$ matrix as described above meets the independence condition by Lee and Yu (2016), but it violates the necessary condition for identification given by (18) if $\alpha < 1$. Using this spatial weights matrix, we generate data following (50) for combinations of $N = 20$, 50, 100, 500, $1,000$, and $T = 1$, 20, 50, 100, under $\alpha = 1$, $1/2$, $1/3$, $1/4$, respectively. The true values of the parameters are set to $\rho = 0.2$ and $\sigma^2 = 1$. Each experiment is replicated $2,000$ times.

Model (50) can be estimated by the standard maximum likelihood approach for SAR models,[17] and Table 1 reports the bias, root mean squared error (RMSE), size, and power of the MLE under different values of α. We first observe that when $\alpha = 1$, the MLE performs properly with declining bias and RMSE as N and/or T increases, and with correct empirical size and good power. Nonetheless, as expected, the bias and RMSE are substantial when $\alpha < 1$, and they are especially severe if T is small. Even when both N and T are large, there is considerably greater variation in the estimates when $\alpha < 1$ as compared to $\alpha = 1$. For instance, for $N = 1,000$ and $T = 100$, the bias ($\times 100$) is -1.10 when $\alpha = 1/4$, whereas it is 0 when $\alpha = 1$; in addition, the RMSE ($\times 100$) is 11.19 when $\alpha = 1/4$, which by contrast is only 0.68 when $\alpha = 1$. It is also evident that the smaller the value of α, the greater the RMSE. Overall, these results corroborate our finding that $tr\left(\mathbf{W}'\mathbf{W}\right)/N > \varepsilon > 0$ for all N (including $N \to \infty$) is essential for the identification of the spatial autoregressive models.

[17] See, for example, Anselin (1988), Chapter 6.

5.2 Estimation Experiments

For the estimation experiments, we follow the Monte Carlo design of Pesaran (2006) and consider the following DGP:

$$\begin{aligned} y_{it} =& \rho y_{it}^* + \beta_1 x_{it1} + \beta_2 x_{it2} + \boldsymbol{\gamma}_{y,i}' f_t + e_{it}, \\ x_{itp} =& \boldsymbol{\gamma}_{x,ip}' f_t + \upsilon_{itp}, \quad p = 1, 2, \end{aligned} \tag{54}$$

for $i = 1, 2, \ldots, N$, and $t = 1, 2, \ldots, T$. The unobserved factors are generated by

$$\begin{aligned} f_{lt} =& \rho_{fl} f_{l,t-1} + \varsigma_{f_{lt}}, \;\; l = 1, 2, \ldots, m; \;\; t = -49, -48, \ldots, 0, 1, \ldots, T, \\ \varsigma_{f_{lt}} \sim& IIDN\left(0, 1 - \rho_{fl}^2\right), \quad \rho_{fl} = 0.5, \quad f_{l,-50} = 0, \end{aligned}$$

where 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

$$\left(\begin{array}{cc} \gamma_{x,i11} & \gamma_{x,i12} \\ \gamma_{x,i21} & \gamma_{x,i22} \end{array} \right) \sim IID \left(\begin{array}{cc} N(0.5, 0.5) & N(0, 0.5) \\ N(0, 0.5) & N(0.5, 0.5) \end{array} \right).$$

The idiosyncratic errors of the x_{itp} processes, $(\upsilon_{it1}, \upsilon_{it2})'$, are generated as

$$\begin{aligned} \upsilon_{it,p} =& \rho_{\upsilon_{ip}} \upsilon_{it-1,p} + \vartheta_{it,p}, \quad t = -49, -48, \ldots, 0, 1, \ldots, T, \\ \vartheta_{it,p} \sim& N\left(0, 1 - \rho_{\vartheta_{ip}}^2\right), \quad \upsilon_{ip,-50} = 0, \\ \rho_{\vartheta_{ip}} \sim& IIDU(0.05, 0.95), \quad p = 1, 2, \end{aligned}$$

where the first 50 observations are discarded.

We consider two different designs for the idiosyncratic errors of y_{it}:[18]

- The errors e_{it} are generated from $IIDN(0,1)$. The main goal of this baseline setup is to compare the efficiency properties of the competing estimators. In particular, it is of interest to examine if the B2SLS and GMM estimators are more efficient than the 2SLS estimator.

- The errors 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

[18] We have also examined the case where the errors are independent over time and heteroskedastic across individual units. The results are presented in the Online Supplement.

half:

$$
\begin{aligned}
e_{it} =& \rho_{ie} e_{i,t-1} + \sigma_i \left(1 - \rho_{ie}^2\right)^{1/2} \zeta_{it}, \quad i = 1, 2, \ldots, \lfloor N/2 \rfloor, & (55) \\
e_{it} =& \sigma_i \left(1 + \theta_{ie}^2\right)^{1/2} \left(\zeta_{it} + \theta_{ie} \zeta_{i,t-1}\right), \quad i = \lfloor N/2 \rfloor + 1, \lfloor N/2 \rfloor + 2, \ldots, N, & (56) \\
\zeta_{it} \sim& IIDN\left(0, 1\right), \quad \sigma_i^2 \sim IIDU\left(0.5, 1.5\right), \\
\rho_{ie} \sim& IIDU\left(0.05, 0.95\right), \quad e_{i,-50} = 0.
\end{aligned}
$$

The spatial weights matrix is specified as the q-ahead-and-q-behind circular neighbors weights matrix; without loss of generality, we set $q = 1$. In all experiments, the true number of factors is set to $m = 2$; the true values of slope coefficients are $\beta_1 = 1$ and $\beta_2 = 2$; the true value of the spatial autoregressive coefficient is $\rho = 0.4$.[19] The sample sizes are $N = 30$, 50, 100, 500, $1{,}000$; and $T = 20$, 30, 50, 100. Each experiment is replicated $2{,}000$ times.

The parameters of interest for model (54) are $(\rho, \beta_1, \beta_2)'$, which are estimated by the following methods:

- The naive 2SLS estimator, which ignores the latent factors and applies a standard 2SLS estimation procedure directly with instruments $\mathbf{Q}_{.t}^{(2)} = \left(\mathbf{X}_{.t}, \mathbf{W}\mathbf{X}_{.t}, \mathbf{W}^2\mathbf{X}_{.t}\right)$, for $t = 1, 2, \ldots, T$, where the superscript of $\mathbf{Q}_{.t}$ denotes that the highest power of $\mathbf{W}$ used in constructing the instruments.

- The infeasible 2SLS estimator, which assumes the factors are known and utilizes instruments $\mathbf{Q}_{.t}^{(2)}$, for $t = 1, 2, \ldots, T$.

- The 2SLS estimator given by (24) with instruments $\mathbf{Q}_{.t}^{(2)}$, for $t = 1, 2, \ldots, T$.

- The B2SLS estimator given by (32), which is implemented in two steps. In the first step, we compute a preliminary 2SLS estimate following (24) using instruments $\mathbf{Q}_{.t}^{(2)}$, for $t = 1, 2, \ldots, T$. In the second step, the B2SLS estimate is obtained by using the estimated best IV matrix $\hat{\mathbf{Q}}^*$ in (32), where

$$
\hat{\mathbf{Q}}^* = \mathbf{M}^b \left[\left(\mathbf{I}_T \otimes \hat{\mathbf{G}}\right) \mathbf{X}\hat{\boldsymbol{\beta}}_{2sls}, \ \mathbf{X}\right], \tag{57}
$$

  and $\hat{\mathbf{G}} = \mathbf{G}\left(\hat{\rho}_{2sls}\right)$.

- The efficient GMM estimator given by (46) that uses $\mathbf{P}_1 = \mathbf{W}$ and $\mathbf{P}_2 = \mathbf{W}^2 - Diag\left(\mathbf{W}^2\right)$ in the quadratic moments and $\mathbf{Q}_{.t}^{(2)}$ as IVs in the linear moments. It is obtained by a two-step procedure. In the first step, we take the identity matrix as the moments weighting matrix and compute a preliminary GMM estimate. In the second step, the estimated inverse of the covariance of moments is used as the weighting matrix, and the model is re-estimated using the same $\mathbf{P}_1$, $\mathbf{P}_2$, and IVs.

[19] We have also considered $\rho = 0.8$, which represents high intensity of spatial dependence, and the results are provided in the Online Supplement.

- The MLE developed by Bai and Li (2014). This procedure assumes that the disturbances of the model are independently distributed with heteroskedastic variances and explicitly estimates all of 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 individual factors.[20] We compute the MLE following the Expectation-Maximization (EM) algorithm suggested by Bai and Li (2014). The number of factors is assumed known in the experiments to reduce the computational burden.[21] The size and power properties of the MLE are not reported in their paper.

For the robust variance estimation of the above methods (except the MLE), the Bartlett window width is chosen to be $\left\lfloor 2\sqrt{T} \right\rfloor$.[22]

Tables 2a to 3b collect the results of the estimation experiments. Each table reports the estimates of bias, root mean squared error (RMSE), size, and power for the aforementioned estimators. Sub-table a reports the estimates of the spatial coefficient, ρ, and Sub-table b reports the estimates of the slope coefficient, β_1. 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 enormous 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, 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 it provides little or no improvement for the slope parameter $\boldsymbol{\beta}$. This implies that the IV matrix $\mathbf{Q}_{.t}^{(2)} = \left(\mathbf{X}_{.t}, \mathbf{W}\mathbf{X}_{.t}, \mathbf{W}^2\mathbf{X}_{.t}\right)$ used in computing the 2SLS estimates approximates the best IV quite well in our experimental designs. The GMM estimator for ρ outperforms the 2SLS and B2SLS estimator in reducing the RMSEs, and it even beats the infeasible 2SLS estimator for modest to large sample size ($N \geq 100$). Finally, the MLE developed 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 be weakened if the number of factors is estimated, especially when the estimated number of factors is smaller than the true value.

Turning to size and power properties, as anticipated by the theoretical findings, the 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.

[20] Bai and Li (2014) point out that one could switch the role of N and T if T is much smaller than N. We do not report results under this interchange, since it involves different stringent assumptions on the disturbances and does not improve the performance of MLE under our Monte Carlo designs.

[21] Bai and Li (2014) propose using an information criterion to estimate the number of factors in their Monte Carlo experiments.

[22] We have also considered $\left\lfloor T^{1/3} \right\rfloor$ as the window size. The results are close, but using $\left\lfloor 2\sqrt{T} \right\rfloor$ has slightly better size properties.

Table 1: Small sample properties of the maximum likelihood estimator of the spatial autoregressive coefficient, ρ, for the identification experiments under different values of α

	Bias($\times$100)				RMSE($\times$100)				Size($\times$100)				Power($\times$100)			
$N\backslash T$	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
1,000	-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
1,000	-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
1,000	-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
1,000	-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: The DGP is given by (50). The true value of ρ is 0.2, and ρ is estimated by the maximum likelihood method. 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 the rest $N_2 = N - N_1$ rows of $\mathbf{W}$ are all zeros. The number of replications is $2,000$. The 95% confidence interval for size 5% is $[3.6\%, 6.4\%]$. The power is calculated under the alternative $H_1 : \rho = 0.1$.

Table 2a: Small sample properties of estimators for the spatial parameter ρ ($\rho = 0.4$, i.i.d. errors)

	Bias($\times$100)				RMSE($\times$100)				Size($\times$100)				Power($\times$100)			
$N\backslash T$	20	30	50	100	20	30	50	100	20	30	50	100	20	30	50	100
	Naive 2SLS estimator (excluding factors)															
30	16.06	16.21	16.34	16.41	16.58	16.61	16.65	16.66	99.40	99.65	99.95	99.95	99.65	99.85	99.95	100.00
50	16.04	16.23	16.34	16.38	16.44	16.52	16.55	16.55	99.90	100.00	100.00	100.00	100.00	100.00	100.00	100.00
100	16.01	16.24	16.32	16.38	16.33	16.46	16.47	16.48	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
500	15.95	16.14	16.27	16.33	16.21	16.30	16.37	16.39	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
1,000	15.95	16.14	16.27	16.34	16.20	16.30	16.37	16.40	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	Infeasible 2SLS estimator (including factors)															
30	-0.10	-0.04	0.00	0.01	2.40	1.91	1.47	0.99	5.10	4.95	5.15	4.45	13.50	18.15	29.10	52.10
50	0.04	0.05	0.00	0.00	1.85	1.48	1.11	0.77	5.75	5.40	5.25	5.15	20.45	29.70	43.00	72.25
100	0.01	0.01	0.01	0.01	1.30	1.03	0.78	0.55	5.25	4.35	4.35	4.40	34.75	48.50	71.15	95.20
500	-0.02	-0.01	0.00	0.00	0.59	0.46	0.35	0.25	5.60	4.85	4.65	4.50	92.60	98.75	100.00	100.00
1,000	0.00	0.00	-0.01	0.00	0.42	0.33	0.25	0.18	4.90	5.65	5.30	5.70	99.80	99.95	100.00	100.00
	2SLS estimator															
30	-0.08	-0.01	0.00	0.01	2.75	2.16	1.64	1.15	6.10	6.45	7.30	8.10	12.40	18.35	28.30	48.30
50	0.02	0.05	0.01	0.00	1.99	1.58	1.20	0.83	5.35	5.95	5.30	5.70	18.05	26.15	41.55	71.00
100	0.01	0.00	0.01	0.01	1.38	1.08	0.81	0.56	4.50	5.10	4.45	5.20	30.25	45.30	69.40	94.10
500	-0.02	-0.01	0.00	0.00	0.62	0.48	0.36	0.25	5.00	4.55	4.95	4.75	89.05	97.95	100.00	100.00
1,000	0.00	0.00	-0.01	0.00	0.44	0.34	0.26	0.18	4.70	4.90	5.20	5.40	99.45	99.95	100.00	100.00
	B2SLS estimator															
30	-0.12	-0.03	-0.01	0.00	2.74	2.16	1.63	1.15	6.00	6.40	7.10	8.00	12.25	18.40	27.90	48.05
50	0.00	0.03	0.00	0.00	1.99	1.58	1.19	0.83	5.30	6.10	5.25	5.45	18.10	25.75	41.30	70.90
100	-0.01	-0.01	0.01	0.01	1.38	1.08	0.80	0.56	4.75	5.05	4.45	5.20	29.75	45.10	69.20	94.15
500	-0.02	-0.01	0.00	0.00	0.62	0.48	0.36	0.25	4.95	4.55	4.85	4.60	88.65	98.05	100.00	100.00
1,000	0.00	0.00	-0.01	0.00	0.44	0.34	0.25	0.18	4.50	4.65	5.05	5.60	99.45	99.95	100.00	100.00
	GMM estimator															
30	-1.25	-1.11	-1.07	-1.02	2.60	2.11	1.76	1.41	10.30	11.45	16.00	24.20	8.75	10.85	14.10	23.15
50	-0.69	-0.64	-0.64	-0.60	1.86	1.52	1.22	0.94	8.50	9.80	12.15	16.00	15.80	21.95	32.10	54.25
100	-0.33	-0.32	-0.31	-0.29	1.24	0.98	0.75	0.57	6.90	6.85	7.05	9.95	33.60	47.30	69.85	94.85
500	-0.08	-0.07	-0.07	-0.06	0.52	0.41	0.31	0.22	6.00	5.20	6.00	6.65	96.15	99.80	100.00	100.00
1,000	-0.03	-0.03	-0.04	-0.03	0.36	0.29	0.22	0.15	5.25	5.60	5.65	6.25	100.00	100.00	100.00	100.00
	MLE															
30	0.30	0.23	0.18	0.16	2.32	1.79	1.36	0.92	11.80	10.20	8.65	7.85	30.65	34.95	46.80	71.20
50	0.35	0.23	0.14	0.11	1.79	1.39	1.02	0.70	12.45	10.00	8.30	7.30	41.45	49.05	64.00	88.70
100	0.29	0.17	0.11	0.09	1.26	0.95	0.70	0.49	11.55	9.25	7.00	6.95	59.75	71.85	89.05	99.25
500	0.22	0.11	0.05	0.04	0.59	0.43	0.31	0.22	13.40	9.30	7.20	7.70	99.00	100.00	100.00	100.00
1,000	0.20	0.11	0.06	0.04	0.42	0.32	0.23	0.16	14.40	11.10	8.40	7.00	100.00	100.00	100.00	100.00

Notes: The DGP is given by (54), where $e_{it} \sim IIDN(0,1)$. The true parameter values are $\rho = 0.4$, $\beta_1 = 1$ and $\beta_2 = 2$. The true number of factors is 2. The spatial weights matrix is the 1-ahead-and-1-behind circular neighbors matrix. The naive estimator ignores latent factors, and the infeasible estimator treats factors as known. The naive 2SLS, infeasible 2SLS, and 2SLS estimators are computed using instruments $\mathbf{Q}_{.t}^{(2)} = \left(\mathbf{X}_{.t}, \mathbf{W}\mathbf{X}_{.t}, \mathbf{W}^2\mathbf{X}_{.t}\right)$, for $t = 1, 2, \ldots, T$. The best 2SLS (B2SLS) estimator is computed using $\hat{\mathbf{Q}}^*$ given by (57). The efficient two-step GMM estimator utilizes $\mathbf{P}_1 = \mathbf{W}$ and $\mathbf{P}_2 = \mathbf{W}^2 - Diag\left(\mathbf{W}^2\right)$ in the quadratic moments and $\mathbf{Q}_{.t}^{(2)}$ in the linear moments. The MLE is computed by the Expectation-Maximization (EM) algorithm described in Bai and Li (2014), assuming the number of factors is known. The number of replications is $2,000$. The 95% confidence interval for size 5% is $[3.6\%, 6.4\%]$, and the power is computed under $H_1 : \rho = 0.38$.

Table 2b: Small sample properties of estimators for the slope parameter β_1 ($\beta_1 = 1$, i.i.d. errors)

	Bias($\times$100)				RMSE($\times$100)				Size($\times$100)				Power($\times$100)			
$N\backslash T$	20	30	50	100	20	30	50	100	20	30	50	100	20	30	50	100
	Naive 2SLS estimator (excluding factors)															
30	8.82	9.09	9.11	9.24	11.71	11.56	11.23	11.10	53.55	63.40	72.35	83.00	76.15	83.80	90.45	95.65
50	8.77	8.88	9.05	9.25	10.91	10.60	10.45	10.40	65.15	74.30	84.40	91.25	87.45	92.60	96.50	99.05
100	9.03	9.13	9.22	9.42	10.43	10.28	10.11	10.14	79.90	86.45	93.35	97.80	96.70	98.60	99.70	99.90
500	9.15	9.27	9.34	9.53	9.93	9.85	9.73	9.78	97.00	98.90	99.80	100.00	99.85	100.00	100.00	100.00
1,000	9.17	9.30	9.36	9.55	9.87	9.80	9.69	9.74	98.15	99.70	99.95	100.00	100.00	100.00	100.00	100.00
	Infeasible 2SLS estimator (including factors)															
30	0.05	0.02	-0.01	-0.01	4.50	3.57	2.73	1.88	5.50	5.20	5.40	5.10	21.80	30.45	47.85	75.95
50	-0.19	-0.17	-0.11	-0.09	3.45	2.66	2.01	1.40	5.45	4.40	4.65	4.50	28.55	42.75	66.25	92.75
100	-0.13	-0.04	-0.05	-0.05	2.46	1.92	1.48	1.01	5.65	5.65	5.35	5.05	52.45	73.75	91.90	99.85
500	-0.06	-0.04	-0.02	-0.01	1.07	0.84	0.66	0.47	4.85	4.35	5.20	5.80	99.80	100.00	100.00	100.00
1,000	0.01	0.01	0.01	0.00	0.79	0.62	0.47	0.33	6.10	6.00	5.30	5.30	100.00	100.00	100.00	100.00
	2SLS estimator															
30	0.06	0.04	0.03	0.03	4.73	3.77	2.91	2.05	5.75	6.30	7.35	7.35	20.00	29.20	47.00	75.60
50	-0.19	-0.18	-0.09	-0.09	3.61	2.76	2.08	1.45	4.70	5.10	4.75	4.85	26.05	39.60	65.80	92.60
100	-0.13	-0.05	-0.06	-0.05	2.54	1.96	1.51	1.03	5.40	5.05	4.75	5.05	46.85	70.70	90.80	99.85
500	-0.07	-0.05	-0.02	-0.01	1.12	0.86	0.67	0.48	4.15	4.30	5.00	5.30	99.25	100.00	100.00	100.00
1,000	0.02	0.01	0.01	0.00	0.82	0.64	0.48	0.33	5.45	5.20	5.25	5.25	100.00	100.00	100.00	100.00
	B2SLS estimator															
30	0.07	0.04	0.04	0.03	4.73	3.77	2.91	2.05	5.75	6.25	7.40	7.35	19.95	29.25	47.20	75.45
50	-0.19	-0.18	-0.09	-0.09	3.61	2.76	2.08	1.45	4.70	5.10	4.70	4.85	26.10	39.75	65.75	92.70
100	-0.13	-0.05	-0.06	-0.04	2.54	1.96	1.51	1.03	5.35	5.10	4.75	5.10	46.90	70.75	90.75	99.85
500	-0.07	-0.05	-0.02	-0.01	1.12	0.86	0.67	0.48	4.15	4.25	5.00	5.30	99.25	100.00	100.00	100.00
1,000	0.02	0.01	0.01	0.00	0.82	0.64	0.48	0.33	5.45	5.30	5.25	5.25	100.00	100.00	100.00	100.00
	GMM estimator															
30	0.16	0.17	0.17	0.17	4.79	3.80	2.92	2.06	5.70	6.95	7.30	7.55	21.35	30.70	49.25	77.55
50	-0.09	-0.08	0.01	0.01	3.64	2.77	2.07	1.45	4.75	4.90	4.80	4.85	27.90	41.60	67.25	93.55
100	-0.08	0.01	0.00	0.01	2.54	1.96	1.50	1.03	5.45	5.05	4.80	4.45	47.70	71.95	90.95	99.85
500	-0.06	-0.04	-0.01	0.00	1.11	0.86	0.67	0.47	4.10	4.30	5.00	5.30	99.20	100.00	100.00	100.00
1,000	0.02	0.02	0.02	0.01	0.81	0.64	0.48	0.33	5.30	5.40	5.50	5.40	100.00	100.00	100.00	100.00
	MLE															
30	-0.01	-0.01	-0.06	-0.04	5.05	3.85	2.86	1.93	11.45	9.30	7.70	6.10	29.80	35.10	48.45	75.80
50	-0.20	-0.16	-0.13	-0.11	3.76	2.79	2.06	1.43	10.20	7.00	5.95	5.55	36.20	47.15	68.15	93.15
100	-0.14	-0.05	-0.07	-0.06	2.68	2.01	1.52	1.04	10.60	8.00	6.80	6.05	58.45	76.55	91.85	99.90
500	-0.02	-0.01	-0.01	-0.01	1.18	0.88	0.67	0.48	10.70	6.50	5.90	7.00	99.60	100.00	100.00	100.00
1,000	0.04	0.00	0.02	-0.01	0.84	0.65	0.47	0.33	10.40	8.40	6.30	5.50	100.00	100.00	100.00	100.00

Notes: The DGP is given by (54), where $e_{it} \sim IIDN(0,1)$. The true parameter values are $\rho = 0.4$, $\beta_1 = 1$ and $\beta_2 = 2$. The true number of factors is 2. The power is computed under $H_1 : \beta_1 = 0.95$. See the notes to Table 2a for other details.

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

	Bias(×100)				RMSE(×100)				Size(×100)				Power(×100)			
$N \backslash T$	20	30	50	100	20	30	50	100	20	30	50	100	20	30	50	100
	Infeasible 2SLS estimator (including factors)															
30	-0.16	-0.08	-0.01	0.00	2.85	2.34	1.81	1.26	5.85	7.00	6.65	5.85	13.60	16.95	23.20	37.55
50	0.07	0.05	-0.01	0.01	2.18	1.77	1.39	0.97	6.75	6.00	5.70	5.90	19.30	25.05	34.70	56.65
100	0.05	0.03	0.03	0.02	1.58	1.28	0.98	0.70	6.70	5.90	5.45	5.85	30.50	40.05	57.35	83.00
500	-0.04	-0.03	-0.02	-0.02	0.70	0.56	0.44	0.31	6.15	5.90	5.45	5.05	81.30	94.70	99.65	100.00
1,000	-0.01	-0.01	-0.01	0.00	0.50	0.40	0.31	0.22	6.05	5.95	6.40	6.75	98.45	99.95	100.00	100.00
	2SLS estimator															
30	-0.14	-0.07	-0.02	0.00	3.10	2.51	1.95	1.39	6.55	7.05	7.35	7.90	12.65	17.80	23.30	36.15
50	0.07	0.06	0.01	0.02	2.27	1.85	1.44	1.01	6.00	6.95	6.20	6.00	17.00	23.60	35.05	55.90
100	0.04	0.03	0.03	0.02	1.62	1.30	0.99	0.71	6.10	5.90	5.55	5.95	26.65	37.30	56.65	83.00
500	-0.04	-0.03	-0.02	-0.02	0.71	0.57	0.44	0.31	5.45	5.70	5.90	5.40	78.90	93.10	99.65	100.00
1,000	-0.01	-0.01	-0.01	0.00	0.51	0.40	0.31	0.23	5.65	5.25	6.05	6.55	97.65	99.95	100.00	100.00
	B2SLS estimator															
30	-0.19	-0.10	-0.04	-0.01	3.10	2.52	1.94	1.39	6.85	7.30	7.45	8.05	12.30	17.10	22.90	35.85
50	0.03	0.04	0.00	0.01	2.27	1.84	1.44	1.01	5.90	6.70	6.20	5.90	16.90	23.10	34.95	55.50
100	0.02	0.02	0.03	0.02	1.61	1.30	0.99	0.71	6.35	5.90	5.45	5.65	26.85	37.00	56.35	82.85
500	-0.05	-0.03	-0.02	-0.02	0.71	0.57	0.45	0.31	5.50	6.05	6.25	5.40	78.60	93.15	99.70	100.00
1,000	-0.01	-0.01	-0.01	0.00	0.51	0.40	0.31	0.23	5.60	5.15	5.95	6.30	97.70	99.95	100.00	100.00
	GMM estimator															
30	-1.11	-1.01	-0.97	-0.97	2.85	2.38	1.96	1.54	9.70	11.00	13.45	17.40	9.20	11.95	14.55	19.10
50	-0.54	-0.52	-0.55	-0.55	2.04	1.69	1.37	1.03	8.00	8.30	8.90	11.00	14.25	20.25	27.25	42.50
100	-0.24	-0.25	-0.25	-0.26	1.41	1.14	0.88	0.66	6.85	6.70	6.95	8.35	27.40	38.30	57.70	82.70
500	-0.08	-0.08	-0.07	-0.07	0.61	0.49	0.38	0.27	5.85	5.15	5.95	5.60	89.30	98.25	100.00	100.00
1,000	-0.04	-0.03	-0.04	-0.03	0.43	0.34	0.27	0.19	4.90	5.05	5.80	6.25	99.50	99.95	100.00	100.00
	MLE															
30	0.38	0.22	0.19	0.15	2.63	2.08	1.59	1.09	21.20	19.20	17.50	16.15	39.90	42.15	50.30	71.25
50	0.45	0.25	0.14	0.13	2.02	1.57	1.19	0.84	22.10	18.05	15.50	15.70	50.95	55.55	65.05	86.25
100	0.39	0.22	0.14	0.10	1.46	1.13	0.85	0.59	22.55	18.80	15.80	15.35	66.60	74.95	87.60	98.50
500	0.27	0.12	0.05	0.03	0.69	0.50	0.36	0.26	23.20	17.00	14.60	13.10	99.40	99.80	100.00	100.00
1,000	0.26	0.20	0.04	0.03	0.50	0.46	0.26	0.19	28.50	24.50	14.90	14.80	100.00	100.00	100.00	100.00

Notes: The DGP is given by (54), where e_{it} are given by (55) and (56). The true parameter values are $\rho = 0.4$, $\beta_1 = 1$ and $\beta_2 = 2$. The true number of factors is 2. The power is computed under $H_1 : \rho = 0.38$. The maximum lag of the robust variance estimator is set to be $2\sqrt{T}$. See also the notes to Table 2a.

Table 3b: Small sample properties of estimators for the slope parameter β_1 ($\beta_1 = 1$, serially correlated and heteroskedastic errors)

	Bias(×100)				RMSE(×100)				Size(×100)				Power(×100)			
$N \backslash T$	20	30	50	100	20	30	50	100	20	30	50	100	20	30	50	100
	Infeasible 2SLS estimator (including factors)															
30	0.10	0.04	-0.02	-0.02	5.35	4.43	3.41	2.38	7.40	7.55	7.30	6.80	21.00	25.75	36.75	59.35
50	-0.23	-0.17	-0.09	-0.12	4.11	3.32	2.53	1.77	6.40	6.45	5.35	5.55	25.55	36.05	51.35	77.65
100	-0.18	-0.07	-0.04	-0.04	2.95	2.37	1.86	1.30	7.05	6.75	7.15	5.70	42.25	60.20	79.90	97.55
500	-0.03	-0.02	-0.01	0.00	1.26	1.04	0.82	0.59	6.00	5.55	6.15	5.90	97.70	99.80	100.00	100.00
1,000	0.02	0.01	0.02	0.00	0.94	0.76	0.60	0.42	7.10	6.70	6.55	6.05	100.00	100.00	100.00	100.00
	2SLS estimator															
30	0.12	0.07	0.05	0.04	5.45	4.49	3.51	2.52	6.65	7.10	8.25	8.45	19.70	26.15	37.15	60.20
50	-0.21	-0.16	-0.06	-0.11	4.17	3.34	2.54	1.79	5.75	6.45	6.05	4.95	24.30	34.00	51.40	78.50
100	-0.14	-0.06	-0.05	-0.04	2.93	2.37	1.86	1.30	5.70	6.85	6.70	5.85	39.85	57.05	78.80	97.50
500	-0.05	-0.04	-0.02	-0.01	1.29	1.05	0.83	0.59	5.30	5.15	5.80	5.95	96.65	99.70	100.00	100.00
1,000	0.02	0.00	0.02	0.00	0.95	0.76	0.60	0.42	6.25	6.35	6.55	5.95	100.00	100.00	100.00	100.00
	B2SLS estimator															
30	0.13	0.07	0.05	0.04	5.45	4.49	3.51	2.51	6.65	7.10	8.30	8.45	19.65	26.20	37.25	60.10
50	-0.20	-0.15	-0.06	-0.11	4.17	3.34	2.54	1.79	5.75	6.50	6.05	5.00	24.40	34.00	51.40	78.55
100	-0.14	-0.06	-0.05	-0.03	2.93	2.37	1.86	1.30	5.75	6.75	6.55	5.90	39.85	57.10	78.85	97.50
500	-0.05	-0.04	-0.02	-0.01	1.29	1.05	0.83	0.59	5.30	5.15	5.80	5.95	96.70	99.65	100.00	100.00
1,000	0.02	0.00	0.02	0.00	0.95	0.76	0.60	0.42	6.35	6.35	6.40	5.95	100.00	100.00	100.00	100.00
	GMM estimator															
30	0.21	0.19	0.17	0.18	5.55	4.55	3.55	2.55	7.80	8.25	8.95	8.75	22.00	28.40	39.35	63.40
50	-0.15	-0.08	0.01	-0.02	4.21	3.34	2.53	1.79	6.40	6.75	6.20	5.40	26.00	35.65	53.40	80.95
100	-0.11	-0.01	0.01	0.02	2.94	2.38	1.86	1.30	6.00	6.70	7.10	5.75	40.90	58.45	79.40	98.05
500	-0.05	-0.03	-0.01	0.00	1.29	1.05	0.82	0.59	5.40	4.95	5.75	5.80	96.75	99.70	100.00	100.00
1,000	0.02	0.01	0.03	0.01	0.95	0.76	0.60	0.42	6.00	6.60	6.85	6.10	100.00	100.00	100.00	100.00
	MLE															
20	0.09	0.02	0.09	0.06	6.94	5.36	4.18	2.95	22.00	16.90	17.55	16.15	34.65	36.45	45.05	62.15
30	0.10	-0.01	-0.05	-0.03	5.61	4.47	3.41	2.35	20.80	18.25	17.10	14.60	38.15	42.60	53.20	74.20
50	-0.16	-0.11	-0.11	-0.12	4.17	3.25	2.43	1.74	18.65	16.15	12.95	13.00	45.10	54.15	68.85	89.95
100	-0.10	-0.06	-0.06	-0.06	2.99	2.32	1.82	1.26	19.85	17.15	16.00	14.10	64.50	77.25	89.60	99.40
500	0.03	0.02	-0.01	-0.00	1.32	1.04	0.79	0.56	19.60	15.40	15.40	14.50	99.50	99.90	100.00	100.00
1,000	0.07	0.01	0.04	0.01	0.96	0.73	0.57	0.40	20.50	17.00	16.60	13.50	100.00	100.00	100.00	100.00

Notes: The DGP is given by (54), where e_{it} are given by (55) and (56). The true parameter values are $\rho = 0.4$, $\beta_1 = 1$ and $\beta_2 = 2$. The true number of factors is 2. The power is computed under $H_1 : \beta_1 = 0.95$. The maximum lag of the robust variance estimator is set to be $2\sqrt{T}$. See also the notes to Table 2a.

In cases where N is much smaller than T, the rejection frequencies under the null of the 2SLS and B2SLS estimators are slightly higher than 5%, and the GMM estimator is more oversized than the 2SLS estimators. It is also evident that the size distortion is more pronounced for the spatial parameter than for the slope coefficients. In view of these results, it is worthwhile to bear in mind that the 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 MLE based tests are significantly over-sized in this case. For the combinations of N and T considered, the empirical sizes of the MLE range from 13% to 29%.

In summary, the proposed estimators exhibit robust performance to unknown heteroskedasticity and serial correlation in the errors. Furthermore, the estimators are also robust to different intensity of spatial dependence, as supported by the additional simulation results in the Online Supplement.

6 An Empirical Application to US House Prices

In this section, we apply the proposed estimation methods to analyzing the spatial dependence of real house price changes in the US at the level of Metropolitan Statistical Areas. Since neighboring regions are often influenced by the same aggregate supply and demand shocks, it is the purpose of this exercise to properly assess the strength of the spatial interconnections while netting out the effects of common factors. As we will see, the degree of spatial dependence will be exaggerated if the unobserved common effects are not effectively removed. In addition, we are also interested in the effects of possible determinant variables on house price growth, including both direct and indirect (spillover) effects.[23]

A Metropolitan Statistical Area (MSA) is defined by the United States Office of Management and Budget (OMB) as a core area with a relatively high population density ($50,000$ people or more), including surrounding territory displaying a high level of economic and social integration with the core, as measured by commuting ties. We consider a total of 377 MSAs using the February 2013 delineations, excluding two MSAs in Alaska and two in Hawaii.[24] For the house price data, we use the Freddie Mac House Price Index (FMHPI) at the MSA level covering the period of 1975Q1–2014Q4. The FMHPI is constructed using a repeat-transactions methodology and published by Freddie Mac every quarter. The nominal house prices are deflated by the Consumer Price Index (CPI) for each MSA, and the following analysis is centered on the quarterly rate of changes in real house prices. For the explanatory variables, we are interested in examining the impact of population growth and real per capita income growth on house price growth. See Appendix B for a detailed description of the data sources and transformations.

[23] Cohen et al. (2016) and Bailey, Holly, and Pesaran (2016) focus on house price series itself and do not consider any explanatory variables.

[24] The Office of Management and Budget (OMB) periodically revises the MSA delineations to reflect the changes in population counts and commuting patterns. There are 381 MSAs in the US as of February 2013. The terms "area" and "MSA" are used interchangeably in the following discussions.

As a preliminary examination of the data, we conduct the cross-sectional dependence (CD) test developed by Pesaran (2015) on the rate of changes in deseasonalized real house prices. The deseasonalization is performed by regressing the nominal house price changes on seasonal dummies and an intercept for each MSA. The CD statistic turns out to be 1364.110 (with the estimated average of the pairwise correlation coefficient being 0.406), which substantially exceeds the critical value of 1.96 at the 5% level and strongly rejects the null hypothesis of weak cross-sectional dependence. Additionally, we compute the exponent of cross-sectional dependence proposed by Bailey, Kapetanios, and Pesaran (2016) and obtain a value of 1.000 (with a standard error 0.024). The value of the exponent, if it lies within the range $[3/4, 1]$, would suggest that the cross-sectional dependence is fairly strong; lying in $[1/2, 3/4)$, it would imply weak dependence of different degrees. Accordingly, the values of both the CD statistic and the estimated exponent clearly indicate the existence of strong cross-sectional dependence in real house price changes; hence, it is imperative to incorporate common factors into the standard spatial models, which capture only weak cross-sectional dependence.

6.1 The Model

Let y_{it} denote the rate of changes in real house prices for area i at time t, which is computed by $y_{it} = log\,(P_{it}/\text{CPI}_{it}) - log\,(P_{i,t-1}/\text{CPI}_{i,t-1})$, where P_{it} is the house price index and CPI_{it} is the Consumer Price Index. We consider the following model for house price changes written in stacked form:

$$\mathbf{y}_{.t} = \rho \mathbf{W} \mathbf{y}_{.t} + (\beta_1 + \theta_1 \mathbf{W})\,\%\Delta\text{Population}_{.t} + (\beta_2 + \theta_2 \mathbf{W})\,\%\Delta\text{Income}_{.t} + \boldsymbol{\Upsilon}\mathbf{d}_t + \boldsymbol{\Gamma}\mathbf{f}_t + \mathbf{e}_{.t}, \quad (58)$$

for $t = 1, 2, \ldots, T$, where $\mathbf{y}_{.t} = (y_{1t}, y_{2t}, \ldots, y_{NT})'$ is a vector of observations on house price growth rates for all MSAs at period t; $\mathbf{d}_t$ signifies an $m_d \times 1$ vector of observed factors that includes quarterly dummies and an intercept; $\mathbf{f}_t$ represents an $m_f \times 1$ vector of unobserved factors; $\boldsymbol{\Upsilon} = (\boldsymbol{\upsilon}_1, \boldsymbol{\upsilon}_2, \ldots, \boldsymbol{\upsilon}_N)'$ and $\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \ldots, \boldsymbol{\gamma}_N)'$ are corresponding individual-specific factor loading matrices; and $\mathbf{e}_{.t}$ is a vector of idiosyncratic error terms. It should be noted that this model accommodates individual fixed effects by including a constant term in $\mathbf{d}_t$ and letting its loadings be heterogeneous. $\%\Delta\text{Population}_{.t}$ represents an $N \times 1$ vector of percentage changes in population at time t, and $\%\Delta\text{Income}_{.t}$ denotes a vector of percentage changes in real per capita income. Both variables are calculated as first differences of natural logarithms. $\mathbf{W}$ is the spatial weights matrix. For generality, model (58) also allows for spatial lags of the explanatory variables, namely $\mathbf{W}\%\Delta\text{Population}_{.t}$ and $\mathbf{W}\%\Delta\text{Income}_{.t}$, which are often referred to as Durbin terms in the literature and capture the interaction effects of exogenous variables.

When it comes to the specification of $\mathbf{W}$, it is common practice to adopt distance- or contiguity-based weighting scheme in the studies of spatial dependence in housing markets. We will follow this tradition first and then explore other possibilities in subsequent analysis. In particular, we assume that contiguity relations are determined by radial distance, and we define "neighbors" of an MSA as those units located within a threshold distance d (miles). The weights of neighbors take a value of one, and the weights of non-neighbors take a value of zero. Then, $\mathbf{W}$ is row-standardized so that the weights across each row sums to unity. The spatial weights matrix constructed in this way is denoted by $\mathbf{W}_d$. Our analysis takes $d = 100$ miles as a point of departure and examines potential dependencies within commuting and transport distances around an MSA.

The parameters of interest are $\boldsymbol{\delta} = \left(\rho, \boldsymbol{\beta}', \boldsymbol{\theta}'\right)'$, where $\boldsymbol{\beta} = (\beta_1, \beta_2)'$ and $\boldsymbol{\theta} = (\theta_1, \theta_2)'$. In what follows, we will focus on the efficient GMM estimator of $\boldsymbol{\delta}$ defined by (46).[25] Specifically, the estimation is implemented by utilizing $\mathbf{P}_1 = \mathbf{W}$ and $\mathbf{P}_2 = \mathbf{W}^2 - Diag\left(\mathbf{W}^2\right)$ in the quadratic moments and $\mathbf{Q}_{.t}^{(2)} = \left(\mathbf{X}_{.t}, \mathbf{W}\mathbf{X}_{.t}, \mathbf{W}^2\mathbf{X}_{.t}\right)$ as instruments in the linear moments, where $\mathbf{X}_{.t} = (\Delta\text{Population}_{.t}, \%\Delta\text{Income}_{.t})$, for $t = 1, 2, ..., T$.

Table 4 summarizes the estimation results of model (58) based on $\mathbf{W} = \mathbf{W}_{100}$. Findings using other specifications of $\mathbf{W}$ will be discussed later. In column (1), the Durbin terms are excluded and the unobserved factors are proxied by cross-sectional averages of both dependent and individual-specific regressors across all MSAs.[26] The estimated spatial coefficient is positive and highly significant, with a value of 0.730 (with a standard error of 0.004). Higher population and income growth are found to increase house price growth, as anticipated. We then include the Durbin terms, and we add to the list of factor proxies the cross-sectional averages of $\mathbf{X}_{.t}^*$ across all MSAs, where $\mathbf{X}_{.t}^* = \mathbf{W}\mathbf{X}_{.t}$. As can be seen from columns (3) and (5), population growth displays a positive and significant spatial interaction effect, but real income growth does not. Overall, the estimates of ρ and $\boldsymbol{\beta}$ are very close across columns (1), (3), and (5). The CD statistics on the residuals of these specifications range from -5.11 to -4.93, which are substantially reduced from the previous test statistic, 1364.110, of the house price growth series itself. The exponents of cross-section dependence of the residuals, however, are about 0.73–0.74, which suggests that a moderate degree of cross-section dependence may be unaccounted for. Therefore, we will next consider local (regional) unobserved factors in addition to global (national) factors, and we will investigate if strong dependence can be more effectively eliminated.[27]

Suppose now that all MSAs are classified into R regions. The model can still be represented by (58), but the observations are now ordered by regions. In specific, the $N \times 1$ vector of house prices changes, $\mathbf{y}_{.t}$, can be written as $\mathbf{y}_{.t} = (\mathbf{y}'_{.1t}, \mathbf{y}'_{.2t}, \ldots, \mathbf{y}'_{.Rt})'$, where $\mathbf{y}_{.rt} = (y_{1rt}, y_{2rt}, \ldots, y_{N_r rt})'$ is an $N_r \times 1$ vector of observations for the r^{th} region, for $r = 1, 2, \ldots, R$. N_r is the number of MSAs in region r, and clearly we have $\sum_{r=1}^{R} N_r = N$. Observations on independent variables and spatial weights are also sorted accordingly. Note that the latent factors, $\mathbf{f}_t$, are now assumed to have a hierarchical structure, namely, $\mathbf{f}_t = \left(\mathbf{f}'_{g,t}, \mathbf{f}'_{l,t}\right)'$, where $\mathbf{f}_{g,t}$ denotes an $m_g \times 1$ vector of

[25] The 2SLS estimates are omitted to save space, since they are very close to the GMM estimates but have larger standard errors, as expected.

[26] In the empirical analysis, $\bar{y}_t^*$ is also included as factor proxies since it may potentially improve the small sample properties of the estimator, where $\bar{y}_t^* = N^{-1}\sum_{i=1}^{N} y_t^*$ and $y_{it}^* = \sum_{j=1}^{N} w_{ij} y_{jt}$. However, it turns out that $\bar{y}_t^*$ and $\bar{y}_t$ are highly correlated for most the $\mathbf{W}$ matrices we considered; therefore, whether $\bar{y}_t^*$ is included makes little difference to the results.

[27] Bailey, Holly, and Pesaran (2016) also consider regional effects, but the authors do not show the impact of eliminating regional factors to the estimated intensity of spatial dependence.

Table 4: Efficient GMM estimation results of model (58)

%ΔHouse price	(1)	(2)	(3)	(4)	(5)	(6)
ρ [$\mathbf{W}\times$ %ΔHouse price]	0.730	0.643	0.732	0.648	0.731	0.648
	(0.004)	(0.005)	(0.004)	(0.005)	(0.004)	(0.005)
β_1 [%ΔPopulation]	0.380	0.366	0.383	0.432	0.369	0.417
	(0.035)	(0.040)	(0.037)	(0.048)	(0.036)	(0.048)
β_2 [%ΔIncome per capita]	0.099	0.093	0.106	0.096	0.111	0.094
	(0.007)	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)
θ_1 [$\mathbf{W}\times$%ΔPopulation]			0.078	0.063	0.063	0.069
			(0.031)	(0.036)	(0.031)	(0.037)
θ_2 [$\mathbf{W}\times$ %ΔIncome per capita]					-0.006	0.019
					(0.010)	(0.012)
Regional unobserved factors	No	Yes	No	Yes	No	Yes
National unobserved factors	Yes	Yes	Yes	Yes	Yes	Yes
MSA FE and seasonal dummies	Yes	Yes	Yes	Yes	Yes	Yes
Residuals						
CD test statistic	-4.946	-6.532	-4.927	-6.385	-5.111	-6.365
Exponent of cross-section	0.734	0.674	0.743	0.690	0.734	0.652
dependence	(0.031)	(0.019)	(0.030)	(0.019)	(0.027)	(0.019)
$\bar{R}^2$	0.808	0.837	0.813	0.844	0.817	0.847
Observations			$N = 377$, $T = 159$			

Notes: Dependent variable is the rate of changes in real house prices, which is computed by first difference of log of real house prices. The explanatory variables are population growth rate and real per capita income growth rate, as well as possibly their spatial lags. MSAs are classified into eight Bureau of Economic Analysis (BEA) Regions. All estimations consider national unobserved factors and include MSA fixed effects (FE) and quarterly dummies. To save space, factor estimates are not reported. The spatial weights matrix is $\mathbf{W} = \mathbf{W}_{100}$. The efficient GMM estimates are obtained by (46), using $\mathbf{P}_1 = \mathbf{W}$ and $\mathbf{P}_2 = \mathbf{W}^2 - Diag\left(\mathbf{W}^2\right)$ in the quadratic moments and $\mathbf{Q}_{.t}^{(2)} = (\mathbf{X}_{.t}, \mathbf{W}\mathbf{X}_{.t}, \mathbf{W}^2\mathbf{X}_{.t})$ as IVs in the linear moments. Standard errors are in parentheses. The standard errors for the slope estimates are heteroskedasticity and autocorrelation consistent with the maximum lag length set to $2\lfloor T^{1/2} \rfloor$.

global factors, $\mathbf{f}_{l,t}$ denotes an $m_l \times 1$ vector of local factors, and $m_g + m_l = m_f$. The associated factor loadings are partitioned as $\mathbf{\Gamma} = (\mathbf{\Gamma}_g, \mathbf{\Gamma}_l)$, where $\mathbf{\Gamma}_g = \left(\boldsymbol{\gamma}_{g,1}, \boldsymbol{\gamma}_{g,2}, \ldots, \boldsymbol{\gamma}_{g,N}\right)'$ is an $N \times m_g$ matrix of loadings for national factors, and $\mathbf{\Gamma}_l = \left(\mathbf{\Gamma}'_{l,1}, \mathbf{\Gamma}'_{l,2}, \ldots, \mathbf{\Gamma}'_{l,R}\right)'$ is an $N \times m_l$ matrix, with $\mathbf{\Gamma}_{l,r} = \left(\boldsymbol{\gamma}_{l,1r}, \boldsymbol{\gamma}_{l,2r}, \ldots, \boldsymbol{\gamma}_{l,N_r r}\right)'$ being the $N_r \times m_l$ factor loadings for the r^{th} region, $r = 1, 2, \ldots, R$. The proposed GMM estimation procedure can easily accommodate regional unobservable factors by replacing them with cross-sectional averages of observations on both dependent and individual-specific independent variables for that region.

Table 4 columns (2), (4), and (6) report the estimation results when both regional and national latent factors are taken into account. We group all 377 MSAs into $R = 8$ Regions based on the geographical classification by the Bureau of Economic Analysis (BEA).[28] Compared with the

[28] The eight BEA Regions are New England, Mideast, Great Lakes, Plains, Southeast, Southwest, Rocky Mountain, and Far West Regions. See the BEA web page, https://www.bea.gov/regional/docs/regions.cfm, for details.

earlier outcomes that did not assume regional factors, both the estimated spatial coefficients and the exponents of cross-section dependence of the residuals decline, suggesting that regional common shocks contribute to the strong cross-sectional dependence in house price changes in the US and that the strength of spatial connections will be overestimated if strong dependence is not effectively eliminated. In addition, after the inclusion of regional factors, the spatial interaction effect of population growth is no longer significant at the 5% level; the spatial interaction effect of income growth remains insignificant. Moreover, the values of $\bar{R}^2$ indicate that the model's goodness of fit improves if regional effects are considered, where $\bar{R}^2$ is computed by $\bar{R}^2 = 1 - \hat{\sigma}^2_{res}/\hat{\sigma}^2_{tot}$, with

$$\hat{\sigma}^2_{tot} = [N(T-1)]^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - \bar{y}_{i.}),$$

$$\hat{\sigma}^2_{res} = [N(T - k_{cs} - k_d) - k_z]^{-1} \sum_{i=1}^{N} \left(\mathbf{y}_{i.} - \mathbf{Z}_{i.}\hat{\boldsymbol{\delta}}\right)' \bar{\mathbf{M}} \left(\mathbf{y}_{i.} - \mathbf{Z}_{i.}\hat{\boldsymbol{\delta}}\right),$$

$\bar{y}_{i.} = T^{-1}\sum_{t=1}^{T} y_{it}$, $\mathbf{y}_{i.} = (y_{i1}, y_{i2}, \ldots, y_{iT})'$, $\mathbf{Z}_{i.} = (\mathbf{y}^*_{i.}, \mathbf{X}_{i.}, \mathbf{X}^*_{i.})$ is a $T \times k_z$ matrix of regressors, k_d is the number of observed factors, and $\bar{\mathbf{M}}$ represents the de-factoring matrix of $T \times k_{cs}$ dimension.[29] According to the above comparisons, we conclude that column (2) provides the best estimation results among all the specifications in Table 4, which points to a significant neighborhood effect in house price changes, with an estimated spatial coefficient of 0.643 (0.005).[30]

Care must be taken when interpreting the estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ in model (58), as they do not directly signify the marginal effects of the independent variables on house price variations. An important feature of SAR models is that a change in an explanatory variable of a unit will affect not only the dependent variable of that unit itself but also the dependent variables of other cross-section units. The former is known as the direct effect, the latter as indirect effect, or spillover effect. Also notice that both effects in general vary across cross-section units. Therefore, to find out the marginal effects of population and income growth on house price changes, we calculate the summary measures of direct and indirect effects proposed by LeSage and Pace (2009). The average direct effect of the k^{th} explanatory variable ($k = 1, 2$) is given by the average of the diagonal elements of $\boldsymbol{\Pi}_k$, where

$$\boldsymbol{\Pi}_k = (\mathbf{I}_N - \rho\mathbf{W})^{-1} (\beta_k \mathbf{I}_N + \theta_k \mathbf{W}), \tag{59}$$

and the average indirect effect is represented by the average row sum of the non-diagonal elements of $\boldsymbol{\Pi}_k$. It can be seen from (59) that imposing $\boldsymbol{\theta} = \mathbf{0}$ implies that the ratio of direct to indirect effects is the same for every explanatory variable, which may be too restrictive; hence, model (58) takes into account the Durbin terms. To test if the direct and spillover effects are significant, we compute the standard errors by simulation, due to the complex formula for the effects in terms of

[29] In specific, $k_d = 4$ because the observed factors consist of quarterly dummies and an intercept. The values of k_z and k_{cs} vary with detailed model specifications, that is, whether Durbin terms are included and if regional factors are considered. This measure of model fit in the presence of unobserved factors is in accordance with the suggestion by Holly et al. (2010, p.164).

[30] Standard error in parentheses.

Table 5: Average direct and indirect effects of population and income growth on house price changes

	Direct	Indirect	Total
Considering both national and regional factors			
%ΔPopulation	0.431	0.571	1.002
	(0.047)	(0.063)	(0.110)
%ΔIncome per capita	0.110	0.146	0.256
	(0.009)	(0.012)	(0.020)
Considering national factors only			
%ΔPopulation	0.518	1.153	1.672
	(0.046)	(0.112)	(0.149)
%ΔIncome per capita	0.135	0.249	0.384
	(0.009)	(0.017)	(0.026)

Notes: The effects of explanatory variables, taking both national and regional factors into account, are computed based on the estimates in column (2) of Table 4. When regional factors are neglected, the effects are computed using the estimates in column (3) of the same table. Bootstrapped standard errors based on $1,000$ iterations are in parentheses. See also the notes to Table 4.

the parameters.[31]

Table 5 shows the estimated average direct and spillover effects of population and income growth on house price growth based on the estimates in Table 4. The average total effect is the sum of average direct and indirect effects. When both national and regional unobserved factors are considered, the specification in column (2) of Table 4 outperforms its counterpart. When only national factors are taken into account, the preferred specification is given by column (3). We compute the effects of the explanatory variables based on these estimates, respectively. It is not surprising to see from Table 5 that the estimated indirect effects are much higher if regional factors are neglected, as there is a relatively stronger degree of cross-sectional dependence in house prices left uncontrolled for. Both population growth and per capita income growth are found to exert both positive and significant direct and indirect impact on house price changes. Specifically, using the estimates produced assuming a hierarchical factor structure, on average a 1% increase in population growth in an MSA is predicted to lead to a 0.43% increase in house price growth in the MSA itself, and a 0.57% increase in house price growth in its neighboring MSAs, while holding other covariates fixed. In comparison, a 1% increase in income growth has much smaller direct and indirect effects on house price growth, which are estimated to be 0.11% and 0.15%, respectively. The spillover effect of population growth to neighboring MSAs appears to be slightly higher than its direct effect, while both effects of income growth are of similar magnitude.

[31] See LeSage and Pace (2009) and Section 2.7 of (Elhorst, 2014) for detailed discussions on the computation.

6.2 Different Spatial Weights Matrix Specifications

We now turn to inspecting the robustness of our findings to various specifications of the spatial weights matrix. Three types of weights are considered, which are constructed based on distance, migration flows, and pairwise correlations, respectively. In all of the following analysis, we will control for unobserved factors at both national and regional levels, as the earlier discussion reveals the importance of both effects on the cross-sectional dependence in house price changes.

We start with comparing the estimation results of model (58) using different radial distance matrices, $\mathbf{W}_d$. In specific, we consider three threshold values, $d = 75$, 100, and 125 miles. The estimation results are presented in columns (1) to (3) of Table 6, respectively. Overall, the estimates are found to be very stable as the cutoff distance varies. The estimated strength of spatial dependence rises slightly from 0.573 (0.005) to 0.693 (0.005) as the neighborhood boundary expands from 75 to 125 miles. This change is reasonable because more units are considered as neighbors of an MSA and their influences are taken into account. The average number of neighbors per MSA is 3.31 when $\mathbf{W} = \mathbf{W}_{75}$, as compared to 8.65 when $\mathbf{W} = \mathbf{W}_{125}$. In addition, the estimated coefficients of population and income growth remain in relatively narrow ranges as d changes. Both variables are highly significant and of reasonable magnitude.[32]

Since many economic and demographic factors apart from geographical proximity may contribute to the cross-sectional dependence in house prices across MSAs, it is interesting to consider spatial weights based on other measures of closeness. In particular, the MSA-to-MSA migration flows are important indicators of the strength of interconnections. We construct a migration weights matrix, denoted by $\mathbf{W}_m$, of which the $(i, j)^{th}$ element represents the share of movers from area j to area i of the total number of movers to area i. We do not consider non-movers or migration flows from/to non-MSAs. Notice that $\mathbf{W}_m$ is an asymmetric matrix in which the immigration flow to each MSA is normalized to unity. The data on inter-MSA migration flows were introduced as part of the American Community Survey (ACS) dataset since 2009; therefore, $\mathbf{W}_m$ is constructed using the migration data from the 2010–2014 ACS 5-year estimates. After dropping the estimates with high margin of errors, each MSA ends up having an average of 4.46 "neighbors."[33] Since the most dominant migration ties are likely to be stable over a long period of time, the time invariability of $\mathbf{W}_m$ does not give cause for concern.

The estimation results of model (58) using $\mathbf{W}_m$ are reported in column (4) of Table 6. Not surprisingly, we find strong evidence of spatial dependence based on migration relations. The estimated spatial parameter is significantly positive, slightly higher than the estimates using distance-based weights. The estimated coefficients of population and income growth are very close to the previous results, and both are significantly different from zero. Both residual diagnostics and the value of $\bar{R}^2$ indicate that the model is a good fit. The similarities between the results using distance and migration weights are quite striking, given that around 65% of the migration flows occur between

[32] The spatially lagged population and income growth turn out to be insignificant in three cases and hence are excluded from the regressions.

[33] Details on the construction of $\mathbf{W}_m$ and its characterization are given in Appendix B and the Online Supplement.

Table 6: Efficient GMM estimation results of model (58) using different spatial weights matrices

	Spatial weights matrix				
	Distance			Migration	Pairwise correlations
%ΔHouse price	$\mathbf{W}_{75}$ (1)	$\mathbf{W}_{100}$ (2)	$\mathbf{W}_{125}$ (3)	$\mathbf{W}_{m}$ (4)	$\hat{\mathbf{W}}^{+}$, $\hat{\mathbf{W}}^{-}$ (5)
ρ [$\mathbf{W}\times$ %ΔHouse price]	0.573	0.643	0.693	0.772	
	(0.005)	(0.005)	(0.005)	(0.005)	
ρ^{+} [$\hat{\mathbf{W}}^{+}\times$ %ΔHouse price]					0.715
					(0.005)
ρ^{-} [$\hat{\mathbf{W}}^{-}\times$ %ΔHouse price]					-0.308
					(0.005)
β_1 [%ΔPopulation]	0.432	0.366	0.294	0.230	0.147
	(0.052)	(0.040)	(0.036)	(0.031)	(0.023)
β_2 [%ΔIncome per capita]	0.099	0.093	0.089	0.075	0.049
	(0.008)	(0.007)	(0.007)	(0.007)	(0.005)
Natl. & Rgnl. unobserved factors	Yes	Yes	Yes	Yes	Yes
MSA FE and seasonal dummies	Yes	Yes	Yes	Yes	Yes
Residuals					
CD test statistic	-6.678	-6.532	-7.127	-3.114	-6.846
Exponent of cross-section	0.668	0.674	0.624	0.728	0.631
dependence	(0.023)	(0.019)	(0.017)	(0.021)	(0.014)
Avg. no. neighbors	3.31	5.73	8.65	4.46	11.01 [$\hat{\mathbf{W}}^{+}$], 8.02 [$\hat{\mathbf{W}}^{-}$]
$\bar{R}^2$	0.833	0.837	0.833	0.840	0.908
Observations			$N = 377$, $T = 159$		

Notes: All estimations consider both national and regional (Natl. & Rgnl.) unobserved factors and also include MSA fixed effects (FE) and quarterly dummies. To save space, factor estimates are not reported. $\mathbf{W}_d$ denotes radial distance weights matrix with threshold distance d, where $d = 75$, 100, and 125 miles. $\mathbf{W}_m$ denotes weights matrix constructed from MSA-to-MSA migration flows. $\hat{\mathbf{W}}^{+}$ and $\hat{\mathbf{W}}^{-}$ denote weights matrices constructed from significant positive and negative pairwise correlations of de-factored house price changes, respectively. See also the notes to Table 4.

MSAs located 100 miles apart.[34]

The third type of spatial weights matrix we consider is created by a data-driven approach that detects significant bilateral relations using house price series itself. Essentially, this approach equates significant pairwise correlations with significant connections. Bailey, Holly, and Pesaran (2016) suggest filtering out strong cross-section dependence from a series first, then applying a regularization or thresholding method to create sparse weights matrices. We follow this idea and construct weights matrices based on significantly positive and negative pair-wise correlations of de-factored house price changes, which are denoted by $\hat{\mathbf{W}}^{+}$ and $\hat{\mathbf{W}}^{-}$, respectively. Specifically, the de-factoring process is conducted by regressing the house price growth rate for each MSA on an intercept, quarterly dummies, and cross-sectional averages of the dependent and explanatory

[34] See Figure S.1 in the Online Supplement for the distribution of distance between the area of origin and the area of destination.

variables at both national and regional levels. Then, significant connections are identified by applying the multiple testing procedure developed by Bailey et al. (2014) to the sample correlation matrix of the first-step residuals at the 5% significance level. If the corresponding correlation coefficient is positively significant, the element of $\hat{\mathbf{W}}^{+}$ is set to one, otherwise to zero. $\hat{\mathbf{W}}^{-}$ is created similarly but based on significantly negative correlations. $\hat{\mathbf{W}}^{+}$ and $\hat{\mathbf{W}}^{-}$ are then row-standardized so that each row sums to one.[35]

With the correlation-based weights matrices, we are able to distinguish between the intensity of positive and negative spatial connections. Let us now consider the following model,

$$\mathbf{y}_{.t} = \rho^{+}\hat{\mathbf{W}}^{+}\mathbf{y}_{.t} + \rho^{-}\hat{\mathbf{W}}^{-}\mathbf{y}_{.t} + \beta_1\%\Delta\text{Population}_{.t} + \beta_2\%\Delta\text{Income}_{.t} + \boldsymbol{\Upsilon}\mathbf{d}_t + \boldsymbol{\Gamma}\mathbf{f}_t + \mathbf{e}_{.t}, \qquad (60)$$

where, as before, $\mathbf{d}_t$ includes an intercept and quarterly dummies, and $\mathbf{f}_t$ contains national and regional unobserved factors.[36]

Table 6, column (5) presents the estimation results of model (60). The estimated ρ^{+} and ρ^{-} have the correct sign, and both are highly significant. The magnitude of the positive spatial effect is notably greater than that of the negative effect, with a value of $\hat{\rho}^{+}$ amounting to 0.715 (0.005) and a value of $\hat{\rho}^{-}$ being -0.308 (0.005). The coefficients of population and income growth are again found to be positive and significant, with slightly smaller magnitude than those obtained using distance and migration weights matrices. The CD statistic is low, and the cross-section exponent is close to the borderline case of 0.5, suggesting that only weak dependence is left in the residuals. The model fits the data very well, as implied from the value of $\bar{R}^2$.

7 Concluding Remarks

This paper considers panel data models in the presence of two sources of cross-sectional dependence: endogenous spatial interactions and common effects. It derives identification conditions and proposes a number of estimators for the joint model. The estimation approach replaces the unobserved common factors with cross-sectional averages and utilizes instrumental variables and quadratic moment conditions in order to cope with the endogenous spatial effects. The proposed estimators are shown to be consistent as long as N is large, irrespective of the size of T. The asymptotic distributions of these estimators are free of nuisance parameters, provided that T is of a smaller order of magnitude than N, as $(N, T) \to \infty$ jointly. Compared with the maximum likelihood approach, the number of latent factors need not be estimated, and more general forms of serial correlation in the disturbances are permitted. A wide range of Monte Carlo exercises lend further support to the theoretical results regarding identification and estimation.

A detailed empirical application to real house price changes reveals that significant spatial dependence exists across MSAs in the US, and it demonstrates the importance of adequately

[35]See Appendix B and the Online Supplement for a more detailed characterization and comparison of different spatial weights matrices.

[36]We have also considered the Durbin terms, but they are found to be insignificant.

removing common effects when analyzing the strength of spatial interconnections. The study also identifies significant effects of population and income growth on house price growth. Besides geographical proximity, we also consider spatial weights based on migration flows and on pairwise correlations of de-factored house price changes. The main findings remain valid under the different measures of connections. These empirical results highlight the need to consider the spatial spillover effects in housing markets when making policy and business decisions.

An important next step for future research is to incorporate rich spatio-temporal dynamics into the model specifications. Such extensions provide a full characterization of how an economic phenomenon transmits across space and over time, and they enable us to distinguish between short-term and long-term spillover effects. Another possible extension of the model is to include slope heterogeneity, which is especially relevant for studies covering different countries, regions, and industries. The present paper is also related to the recent study by Pesaran and Yang (2016), who consider networks with dominant units and common factors. The identification and estimation of these models, in which the spatial weights matrix may have unbounded column sums, are of practical importance and worth further investigation.

References

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*, Volume 4. Springer Science & Business Media.

Aquaro, M., N. Bailey, and M. H. Pesaran (2015). Quasi maximum likelihood estimation of spatial models with heterogeneous coefficients. USC-INET Research Paper, No. 15-17.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica 77*, 1229–1279.

Bai, J. and K. Li (2012). Statistical analysis of factor models of high dimension. *The Annals of Statistics 40*, 436–465.

Bai, J. and K. Li (2014). Spatial panel data models with common shocks. MPRA Paper 52786.

Bai, J. and K. Li (2015). Dynamic spatial panel data models with common shocks. Manuscript.

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica 70*, 191–221.

Bai, J. and S. Ng (2007). Determining the number of primitive shocks in factor models. *Journal of Business & Economic Statistics 25*, 52–60.

Bailey, N., S. Holly, and M. H. Pesaran (2016). A two-stage approach to spatio-temporal analysis with strong and weak cross-sectional dependence. *Journal of Applied Econometrics 31*(1), 249–280.

Bailey, N., G. Kapetanios, and M. H. Pesaran (2016). Exponent of cross-sectional dependence: Estimation and inference. *Journal of Applied Econometrics 31*, 929–960.

Bailey, N., M. H. Pesaran, and L. Smith (2014). A multiple testing approach to the regularisation of sample correlation matrices. CESifo working paper 4834.

Brady, R. R. (2011). Measuring the diffusion of housing prices across space and over time. *Journal of Applied Econometrics 26*(2), 213–231.

Brady, R. R. (2014). The spatial diffusion of regional housing prices across US states. *Regional Science and Urban Economics 46*, 150–166.

Chudik, A. and M. H. Pesaran (2015a). Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. *Journal of Econometrics 188*(2), 393–420.

Chudik, A. and M. H. Pesaran (2015b). Large panel data models with cross-sectional dependence: A survey. In B. H. Baltagi (Ed.), *The Oxford handbook of panel data*, Chapter 1. Oxford University Press.

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

Cohen, J. P., Y. M. Ioannides, and W. W. Thanapisitikul (2016). Spatial effects and house price dynamics in the USA. *Journal of Housing Economics 31*, 1–13.

Elhorst, J. P. (2014). *Spatial econometrics: From cross-sectional data to spatial panels*. Springer.

Holly, S., M. H. Pesaran, and T. Yamagata (2010). A spatio-temporal model of house prices in the USA. *Journal of Econometrics 158*(1), 160–173.

Holly, S., M. H. Pesaran, and T. Yamagata (2011). The spatial and temporal diffusion of house prices in the UK. *Journal of Urban Economics 69*(1), 2–23.

Kapetanios, G. (2010). A testing procedure for determining the number of factors in approximate factor models with large datasets. *Journal of Business & Economic Statistics 28*, 397–409.

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

Kelejian, H. H. and I. R. Prucha (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *The Journal of Real Estate Finance and Economics 17*(1), 99–121.

Kelejian, H. H. and I. R. Prucha (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International economic review 40*(2), 509–533.

Kelejian, H. H. and I. R. Prucha (2001). On the asymptotic distribution of the Moran I test statistic with applications. *Journal of Econometrics 104*(2), 219–257.

Kelejian, H. H. and I. R. Prucha (2010). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics 157*(1), 53–67.

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

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

Lee, L.-f. and J. Yu (2010a). Estimation of spatial autoregressive panel data models with fixed effects. *Journal of Econometrics 154*(2), 165–185.

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

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

Lee, L.-f. and J. Yu (2016). Identification of spatial durbin panel models. *Journal of Applied Econometrics 31*(1), 133–162.

LeSage, J. P. and R. K. Pace (2009). *Introduction to spatial econometrics*. CRC Press, Taylor & Francis Group, Boca Raton.

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

Lu, L. (2017). Simultaneous spatial panel data models with common shocks. Federal Reserve Bank of Boston Working Paper RPA 17-03.

Luo, Z. Q., C. Liu, and D. Picken (2007). Housing price diffusion pattern of Australia's state capital cities. *International Journal of Strategic Property Management 11*(4), 227–242.

Meen, G. (1999). Regional house prices and the ripple effect: A new interpretation. *Housing studies 14*(6), 733–753.

Moon, H. R. and M. Weidner (2015). Dynamic linear panel regression models with interactive fixed effects. *Econometric Theory*, 1–38.

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

Pesaran, M. H. (2015). Testing weak cross-sectional dependence in large panels. *Econometric Reviews 34*(6-10), 1089–1117.

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

Pesaran, M. H. and C. F. Yang (2016). Econometric analysis of production networks with dominant units. USC-INET Research Paper No. 16-25.

Pollakowski, H. O. and T. S. Ray (1997). Housing price diffusion patterns at different aggregation levels: An examination of housing market efficiency. *Journal of Housing Research*, 107–124.

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

Sarafidis, V. and T. Wansbeek (2012). Cross-sectional dependence in panel data analysis. *Econometric Reviews 31*(5), 483–531.

Shi, S., M. Young, and B. Hargreaves (2009). The ripple effect of local house price movements in New Zealand. *Journal of Property Research 26*(1), 1–24.

Shi, W. and L.-f. Lee (2017). Spatial dynamic panel data models with interactive fixed effects. *Journal of Econometrics 197*(2), 323–347.

Stock, J. H. and M. W. Watson (2011). Dynamic factor models. *Oxford handbook of economic forecasting 1*, 35–59.

Yu, J., R. De Jong, and L.-f. Lee (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.

A Appendix: Proofs of Main Theorems

The proofs are based on the lemmas in the Online Supplement.

Proof of Theorem 1

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 to the proof is to establish the distribution of $(NT)^{-1/2}\,\mathbf{Q}'\mathbf{M}^b\left[(\mathbf{I}_T\otimes\boldsymbol{\Gamma})\mathbf{f}+\mathbf{e}\right]$. Applying Lemma A6, we only need to derive the distribution of $(NT)^{-1/2}\sum_{i=1}^N \mathbf{X}_{i.}'\bar{\mathbf{M}}(\mathbf{F}\boldsymbol{\gamma}_i+\mathbf{e}_{i.})$, and then the distribution of $(NT)^{-1/2}\sum_{i=1}^N\sum_{l=1}^N w_{il}^s\mathbf{X}_{i.}'\bar{\mathbf{M}}(\mathbf{F}\boldsymbol{\gamma}_i+\mathbf{e}_{i.})$, for $s=1,2,\ldots$, will readily follow.

Let us first consider $(NT)^{-1/2}\sum_{i=1}^N \mathbf{X}_{i.}'\bar{\mathbf{M}}\mathbf{F}\boldsymbol{\gamma}_i$. Under Assumption 3, $\boldsymbol{\gamma}_i=\boldsymbol{\gamma}+\boldsymbol{\eta}_i$, and note that $N^{-1}\sum_{i=1}^N \mathbf{X}_{i.}'\bar{\mathbf{M}}\mathbf{F}\boldsymbol{\gamma}=\bar{\mathbf{X}}_{i.}'\bar{\mathbf{M}}\mathbf{F}\boldsymbol{\gamma}=\mathbf{0}$, we have

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

It is shown in Lemma A5 of the Online Supplement that[37]

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

Substituting this result into (A.1) yields

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

Using (S.6) and (S.7) of Lemma A5 in the Online Supplement, and noting that the norms of $\bar{\mathbf{C}}'(\bar{\mathbf{C}}\bar{\mathbf{C}}')^{-1}$ are bounded, we get

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

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 by Lemma A4, we obtain

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

Now that it is readily seen that $\bar{\mathbf{C}}'\left(\bar{\mathbf{C}}\bar{\mathbf{C}}'\right)^{-1}-\bar{\mathbf{C}}_{-i}'\left(\bar{\mathbf{C}}_{-i}\bar{\mathbf{C}}_{-i}'\right)^{-1}=O_p\left(N^{-1}\right)$, 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

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

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

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

We next turn to analyzing the distribution of $(NT)^{-1/2}\sum_{i=1}^{N}\mathbf{X}_{i.}'\bar{\mathbf{M}}\mathbf{e}_{i.}$. Let $\boldsymbol{\Pi}=\mathbf{F}\bar{\mathbf{C}}$. It can be shown that

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{X}_{i.}'\bar{\mathbf{M}}\mathbf{e}_{i.}=\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{X}_{i.}'\mathbf{M}_f\mathbf{e}_{i.}+\frac{1}{N}\sum_{i=1}^{N}\frac{\mathbf{X}_{i.}'\boldsymbol{\Pi}}{T}\left(\frac{\boldsymbol{\Pi}'\boldsymbol{\Pi}}{T}\right)^{-1}\left(\frac{\sqrt{N}\bar{\boldsymbol{\epsilon}}'\mathbf{e}_{i.}}{\sqrt{T}}\right)+O_p\left(\sqrt{\frac{T}{N}}\right). \tag{A.2}$$

[37]See (S.5) of Lemma A5.

The first term on the right-hand side of (A.2) follows a distribution

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

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

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

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

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

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

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

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

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

and it follows that

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

As a result, as $(N,T)\xrightarrow{j}\infty$ and $T/N\to 0$, we have $\sqrt{NT}\left(\hat{\boldsymbol{\delta}}_{2sls}-\boldsymbol{\delta}_0\right)\xrightarrow{d}N\left(\mathbf{0},\boldsymbol{\Sigma}_{2sls}\right)$, where $\boldsymbol{\Sigma}_{2sls}$ is given by (26).

Proof of Theorem 2

Note that

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

To establish the asymptotic distribution, it suffices to show that

$$\underset{N,T\to\infty}{p\lim}\frac{1}{NT}\hat{\mathbf{Q}}^{*\prime}\mathbf{L}=\underset{N,T\to\infty}{p\lim}\frac{1}{NT}\mathbf{L}_0'\mathbf{M}_f^b\mathbf{L}_0, \tag{A.3}$$

and

$$\frac{1}{\sqrt{NT}}\hat{\mathbf{Q}}^{*\prime}\left[(\mathbf{I}_T\otimes\boldsymbol{\Gamma}_0)\mathbf{f}+\mathbf{e}\right]\xrightarrow{d}N\left(\mathbf{0},\boldsymbol{\Omega}_{LMe}\right). \tag{A.4}$$

Substituting
$$\mathbf{Y}=\left(\mathbf{I}_T\otimes\mathbf{S}_0^{-1}\right)\left[\mathbf{X}\boldsymbol{\beta}_0+\left(\mathbf{I}_T\otimes\boldsymbol{\Gamma}_0\right)\mathbf{f}+\mathbf{e}\right]$$
into the definition of $\mathbf{L}$ yields
$$\mathbf{L}=\mathbf{L}_0+\left[\left(\mathbf{I}_T\otimes\boldsymbol{\Gamma}_0\right)\mathbf{f}+\mathbf{e},\mathbf{0}\right],$$
and it follows that
$$\frac{1}{NT}\hat{\mathbf{Q}}^{*\prime}\mathbf{L}=\frac{1}{NT}\left[\left(\mathbf{I}_T\otimes\mathbf{G}(\hat{\rho})\right)\mathbf{X}\hat{\boldsymbol{\beta}},\mathbf{X}\right]^{\prime}\mathbf{M}^b\mathbf{L}_0+\frac{1}{NT}\left[\left(\mathbf{I}_T\otimes\mathbf{G}(\hat{\rho})\right)\mathbf{X}\hat{\boldsymbol{\beta}},\mathbf{X}\right]^{\prime}\mathbf{M}^b\left[(\mathbf{I}_T\otimes\boldsymbol{\Gamma}_0)\mathbf{f}+\mathbf{e},\mathbf{0}\right].$$
Using the first-order Taylor expansion of $\mathbf{G}(\hat{\rho})$, we have
$$\mathbf{W}(\mathbf{I}_N-\hat{\rho}\mathbf{W})^{-1}=\mathbf{G}_0+\mathbf{W}(\mathbf{I}_N-\hat{\rho}\mathbf{W})^{-1}\mathbf{G}_0(\hat{\rho}-\rho_0).$$
Applying Lemma A6, and using $\hat{\rho}=\rho+o_p(1)$ and $\hat{\boldsymbol{\beta}}=\boldsymbol{\beta}_0+o_p(1)$, we obtain
$$\frac{1}{NT}\left[\left(\mathbf{I}_T\otimes\mathbf{G}(\hat{\rho})\right)\mathbf{X}\hat{\boldsymbol{\beta}},\mathbf{X}\right]^{\prime}\mathbf{M}^b\mathbf{L}_0=\frac{1}{NT}\mathbf{L}_0^{\prime}\mathbf{M}_f^b\mathbf{L}_0+o_p(1),$$
$$\frac{1}{NT}\left[\left(\mathbf{I}_T\otimes\mathbf{G}(\hat{\rho})\right)\mathbf{X}\hat{\boldsymbol{\beta}},\mathbf{X}\right]^{\prime}\mathbf{M}^b\left[(\mathbf{I}_T\otimes\boldsymbol{\Gamma}_0)\mathbf{f}+\mathbf{e},\mathbf{0}\right]=o_p(1).$$
Thus, the result in (A.3) is proved. The claim in (A.4) can be established using an argument similar to the one in the proof of Proposition 1.

Now in order to examine if $\hat{\mathbf{Q}}^*$ is the best IV, we need to compare the asymptotic variances $\boldsymbol{\Sigma}_{b2sls}$ with $\boldsymbol{\Sigma}_{2sls}$. Notice that
$$\mathbf{L}_0^{\prime}\mathbf{P}_{Q,f}\mathbf{L}_0=\mathbf{L}_0^{\prime}\mathbf{M}_f^b\mathbf{Q}(\mathbf{Q}^{\prime}\mathbf{M}_f^b\mathbf{Q})^{-1}\mathbf{Q}^{\prime}\mathbf{M}_f^b\mathbf{L}_0\leq\mathbf{L}_0^{\prime}\mathbf{M}_f^b\mathbf{L}_0,$$
and hence $\boldsymbol{\Psi}_{LPL}\leq\boldsymbol{\Psi}_{LML}$. If the disturbances $\{e_{it}\}$ are independently and identically distributed with mean zero and variance σ_e^2, then $\boldsymbol{\Sigma}_{b2sls}=\sigma_e^2\boldsymbol{\Psi}_{LML}^{-1}\leq\sigma_e^2\boldsymbol{\Psi}_{LPL}^{-1}=\boldsymbol{\Sigma}_{2sls}$. However, in general we cannot conclude that $\hat{\mathbf{Q}}^*$ is optimal, because $\boldsymbol{\Omega}_{e,i.}$ is unknown and $\boldsymbol{\Omega}_{LPe}$ could be greater than $\boldsymbol{\Omega}_{LMe}$.

Proof of Theorem 3

Consistency

Under the identification conditions for this model, it suffices to show that $(NT)^{-1}\mathbf{A}_{NT}^w\mathbf{g}_{NT}(\boldsymbol{\delta})$ converges to its mean uniformly in $\boldsymbol{\delta}\in\boldsymbol{\Delta}_{sp}$ and the limit equals zero at $\boldsymbol{\delta}_0$. Notice that
$$\boldsymbol{\xi}(\boldsymbol{\delta})\quad=\quad\left[\mathbf{I}_T\otimes\mathbf{S}(\rho)\right]\left(\mathbf{I}_T\otimes\mathbf{S}_0^{-1}\right)\left[\mathbf{X}\boldsymbol{\beta}_0+(\mathbf{I}_T\otimes\boldsymbol{\Gamma}_0)\mathbf{f}+\mathbf{e}\right]-\mathbf{X}\boldsymbol{\beta}.$$
Since $\mathbf{S}(\rho)\mathbf{S}_0^{-1}=\left[\mathbf{S}_0+(\rho_0-\rho)\mathbf{W}\right]\mathbf{S}_0^{-1}=\mathbf{I}_N+(\rho_0-\rho)\mathbf{G}_0$, where $\mathbf{G}_0=\mathbf{W}\mathbf{S}_0^{-1}$, we then obtain
$$\begin{aligned}\boldsymbol{\xi}(\boldsymbol{\delta})&=\left[\mathbf{I}_{NT}+(\rho_0-\rho)(\mathbf{I}_T\otimes\mathbf{G}_0)\right]\left[\mathbf{X}\boldsymbol{\beta}_0+(\mathbf{I}_T\otimes\boldsymbol{\Gamma}_0)\mathbf{f}\right]-\mathbf{X}\boldsymbol{\beta}+\mathbf{I}_T\otimes\left[\mathbf{S}(\rho)\mathbf{S}_0^{-1}\right]\mathbf{e}\\&=\mathbf{X}(\boldsymbol{\beta}_0-\boldsymbol{\beta})+(\rho_0-\rho)\left(\mathbf{I}_T\otimes\mathbf{G}_0\right)\left[\mathbf{X}\boldsymbol{\beta}_0+(\mathbf{I}_T\otimes\boldsymbol{\Gamma}_0)\mathbf{f}\right]+(\mathbf{I}_T\otimes\boldsymbol{\Gamma}_0)\mathbf{f}+\mathbf{I}_T\otimes\left[\mathbf{S}(\rho)\mathbf{S}_0^{-1}\right]\mathbf{e}\\&=\mathbf{d}(\boldsymbol{\delta})+\mathbf{r}_{\xi}(\boldsymbol{\delta}).\end{aligned}\tag{A.5}$$

where

$$\mathbf{d}(\boldsymbol{\delta}) = \mathbf{X}(\boldsymbol{\beta}_0 - \boldsymbol{\beta}) + \mathbf{J}(\rho_0 - \rho) \tag{A.6}$$

$$\mathbf{J} = (\mathbf{I}_T \otimes \mathbf{G}_0)[\mathbf{X}\boldsymbol{\beta}_0 + (\mathbf{I}_T \otimes \boldsymbol{\Gamma}_0)\mathbf{f}], \tag{A.7}$$

$$\mathbf{r}_\xi(\boldsymbol{\delta}) = (\mathbf{I}_T \otimes \boldsymbol{\Gamma}_0)\mathbf{f} + \mathbf{I}_T \otimes \left[\mathbf{S}(\rho)\mathbf{S}_0^{-1}\right]\mathbf{e}. \tag{A.8}$$

Let $\mathbf{A}_{NT}^w = \left(\mathbf{a}^{(1)}, \mathbf{a}^{(2)}, \ldots, \mathbf{a}^{(r)}, \mathbf{A}^{(Q)}\right)$, where $\mathbf{a}^{(l)}$, for $l = 1, 2, \ldots, r$, are $k_a \times 1$ vectors, and $\mathbf{A}^{(Q)}$ is a $k_a \times q$ matrix. By definition, we have

$$\frac{1}{NT}\mathbf{A}_{NT}^w \mathbf{g}_{NT}(\boldsymbol{\delta}) = \frac{1}{NT}\sum_{l=1}^{r}\mathbf{a}^{(l)}\left[\boldsymbol{\xi}'(\boldsymbol{\delta})\mathbf{M}^b\mathbf{P}_l^b\mathbf{M}^b\boldsymbol{\xi}(\boldsymbol{\delta})\right] + \frac{1}{NT}\mathbf{A}^{(Q)}\mathbf{Q}'\mathbf{M}^b\boldsymbol{\xi}(\boldsymbol{\delta}). \tag{A.9}$$

Expanding the first term of (A.9) produces

$$\frac{1}{NT}\sum_{l=1}^{r}\mathbf{a}^{(l)}\left[\boldsymbol{\xi}'(\boldsymbol{\delta})\mathbf{M}^b\mathbf{P}_l^b\mathbf{M}^b\boldsymbol{\xi}(\boldsymbol{\delta})\right] = \boldsymbol{\varpi}_1 + 2\boldsymbol{\varpi}_2 + \boldsymbol{\varpi}_3,$$

where $\boldsymbol{\varpi}_1 = \frac{1}{NT}\sum_{l=1}^{r}\mathbf{a}^{(l)}\left[\mathbf{d}'(\boldsymbol{\delta})\mathbf{M}^b\mathbf{P}_l^b\mathbf{M}^b\mathbf{d}(\boldsymbol{\delta})\right]$, $\boldsymbol{\varpi}_2 = \frac{1}{NT}\sum_{l=1}^{r}\mathbf{a}^{(l)}\left[\mathbf{d}'(\boldsymbol{\delta})\mathbf{M}^b\mathbf{P}_l^b\mathbf{M}^b\mathbf{r}_\xi(\boldsymbol{\delta})\right]$, and $\boldsymbol{\varpi}_3 = \frac{1}{NT}\sum_{l=1}^{r}\mathbf{a}^{(l)}\left[\mathbf{r}_\xi'(\boldsymbol{\delta})\mathbf{M}^b\mathbf{P}_l^b\mathbf{M}^b\mathbf{r}_\xi(\boldsymbol{\delta})\right]$. Note that $\mathbf{S}(\rho)\mathbf{S}_0^{-1}$ has bounded row and column norms, and so do the products of $\mathbf{S}(\rho)\mathbf{S}_0^{-1}$, $\mathbf{P}_l$, and $\left[\mathbf{S}(\rho)\mathbf{S}_0^{-1}\right]'$. Also notice that $\mathbf{M}^b\mathbf{P}_l^b\mathbf{M}^b = \bar{\mathbf{M}} \otimes \mathbf{P}_l$. Applying Lemma A6 and A7, we obtain that $\boldsymbol{\varpi}_1$, $\boldsymbol{\varpi}_2$ and $\boldsymbol{\varpi}_3$ converge uniformly to their means, respectively. In addition, the second term in (A.9) converges uniformly to zero. Hence, we establish the uniform convergence of $(NT)^{-1}\mathbf{A}_{NT}^w\mathbf{g}_{NT}(\boldsymbol{\delta})$. Furthermore, its limit equals zero at the true value $\boldsymbol{\delta}_0$. This 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 $\frac{\partial \mathbf{g}_{NT}'(\hat{\boldsymbol{\delta}})}{\partial}\mathbf{A}_{NT}^{w\prime}\mathbf{A}_{NT}^w\mathbf{g}_{NT}\left(\hat{\boldsymbol{\delta}}\right) = 0$ around the true value, $\boldsymbol{\delta}_0$, we obtain

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

where $\ddot{\boldsymbol{\delta}}$ is a point between $\hat{\boldsymbol{\delta}}$ and $\boldsymbol{\delta}_0$. For any $\boldsymbol{\delta}$ in the parameter space $\boldsymbol{\Delta}_{sp}$, we have $\partial\boldsymbol{\xi}(\boldsymbol{\delta})/\partial\boldsymbol{\delta}' = -[(\mathbf{I}_T \otimes \mathbf{W})\mathbf{Y}, \mathbf{X}]$, and it follows that

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

where $\mathbf{P}_l^{sb} = \mathbf{I}_T \otimes \mathbf{P}_l^s$ and $\mathbf{P}_l^s = \mathbf{P}_l + \mathbf{P}_l'$, for $l = 1, 2, \ldots, r$. Since

$$\begin{aligned}\frac{1}{NT}\boldsymbol{\xi}'(\boldsymbol{\delta})\mathbf{M}^b\mathbf{P}_l^{sb}\mathbf{M}^b(\mathbf{I}_T \otimes \mathbf{W})\mathbf{Y} =& \frac{1}{NT}\boldsymbol{\xi}'(\boldsymbol{\delta})\left(\bar{\mathbf{M}} \otimes \mathbf{P}_l^s\mathbf{G}_0\right)\mathbf{X}\boldsymbol{\beta}_0 \\ &+ \frac{1}{NT}\boldsymbol{\xi}'(\boldsymbol{\delta})\left[\left(\bar{\mathbf{M}} \otimes \mathbf{P}_l^s\mathbf{G}_0\right)\mathbf{e} + \left(\bar{\mathbf{M}} \otimes \mathbf{P}_l^s\mathbf{G}_0\boldsymbol{\Gamma}_0\right)\mathbf{f}\right],\end{aligned} \tag{A.10}$$

by Lemma A6 and A7, at true value $\boldsymbol{\delta}_0$ the above equation (A.10) can be rewritten as

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

where $\tilde{g}_{ii,l}^{s}$ is the i^{th} diagonal element of matrix $\tilde{\mathbf{G}}_l(\rho_0) = \mathbf{P}_l^s \mathbf{G}_0$, and $(NT)^{-1}\, \mathbf{e}' \mathbf{M}^b \mathbf{P}_l^s \mathbf{M}^b \mathbf{X} = o_p(1)$. In addition, we have

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

It then follows that $(NT)^{-1}\, \partial \mathbf{g}_{NT}'\left(\boldsymbol{\delta}\right)/\partial \boldsymbol{\delta} = -\mathbf{D} + o_p(1)$, where $\mathbf{D}$ is given by (42).

Finally, applying the Central Limit Theorem given by Lemma A9 for the linear and quadratic forms establishes

$$\begin{aligned} \frac{1}{\sqrt{NT}}\mathbf{A}_{NT}^{w}\mathbf{g}_{NT}\left(\boldsymbol{\delta}_0\right) =& \frac{1}{\sqrt{NT}}\left[\mathbf{r}_\xi'\left(\boldsymbol{\delta}_0\right)\left(\sum_{l=1}^{r}\mathbf{a}^{(l)}\mathbf{M}^b\mathbf{P}_l^b\mathbf{M}^b\right)\mathbf{r}_\xi\left(\boldsymbol{\delta}_0\right) + \mathbf{A}^{(Q)}\mathbf{Q}'\mathbf{M}^b\mathbf{r}_\xi\left(\boldsymbol{\delta}_0\right)\right] \\ &\xrightarrow{d} N\left(\mathbf{0}, \mathbf{A}^{w\prime}\boldsymbol{\Sigma}_g\mathbf{A}^{w}\right), \end{aligned} \tag{A.11}$$

where $\boldsymbol{\Sigma}_g$ is given by (44), and this completes the proof.

B Data Appendix

The house price indices for Metropolitan Statistical Areas (MSAs) at monthly frequency are obtained from the website of Freddie Mac: http://www.freddiemac.com/finance/fmhpi/archive.html. The quarterly values are computed by taking the three-month arithmetic averages.

The annual Consumer Price Index (CPI) series for all urban areas is sourced from website of the Bureau of Labor Statistics: http://data.bls.gov/pdq/querytool.jsp?survey=cu. The CPI for each MSA is constructed from the corresponding state CPI, and the missing observations for a few area-year combinations are replaced by the US averages.

The data on annual income per capita and population at the MSA level are obtained from the website of the Bureau of Economic Analysis (BEA): http://bea.gov/regional/downloadzip.cfm.

The quarterly values of CPI, income, and population are computed from annual series, following the interpolation method given in Appendix B.3 of the Global Vector AutoRegressive (GVAR) Toolbox User Guide, which is available at the GVAR modeling website:
https://sites.google.com/site/gvarmodelling/gvar-toolbox/download.

The geodesic distance between MSAs is calculated by the Haversine formula, using the Latitude-Longitude of zip codes of the corresponding MSAs. The data on MSA-to-MSA migration flows are sourced from the 2010-2014 American Community Survey (ACS) 5-year estimates by the United States Census Bureau. The flow estimates with coefficients of variation higher than 20% are dropped from the sample. Table B.1 reports the summary measures of different spatial weights matrices used in the analysis. Further details about these weights matrices are provided in the Online Supplement.

Table B.1: Summary of the spatial weights matrices

	$\mathbf{W}_{75}$	$\mathbf{W}_{100}$	$\mathbf{W}_{125}$	$\mathbf{W}_{m}$	$\hat{\mathbf{W}}^{+}$	$\hat{\mathbf{W}}^{-}$
Links						
Mean	3.31	5.73	8.65	4.46	10.41	7.53
Max	12	20	27	59	35	35
Total	1,246	2,162	3,260	1,681	3,926	2,838
Network density	0.88%	1.53%	2.30%	1.19%	2.77%	2.00%
Isolated MSAs	39	15	10	53	3	45
Dimensions			377×377			

Notes: $\mathbf{W}_d$ denotes radial distance weights matrix with threshold distance d (miles). $\mathbf{W}_m$ represents weights matrix based on MSA-to-MSA migration flows. $\hat{\mathbf{W}}^{+}$ ($\hat{\mathbf{W}}^{-}$) is constructed from significantly positive (negative) pairwise correlations of de-factored house price changes. The total number of links equals the number of nonzero elements in the weights matrix. Network density is computed by dividing the sum of existing links by the number of all possible links.

Online Supplement to "Common Factors and Spatial Dependence: An Application to US House Prices"

Cynthia Fan Yang†

August 20, 2018

This Online Supplement is organized into three sections. Section S1 provides supplementary lemmas for the main proofs and derivations of the identification conditions. Section S2 reports additional results of Monte Carlo experiments. Section S3 presents more empirical findings and further description of the spatial weights matrices.

S1 Theory Supplement

S1.1 Supplementary Lemmas

The following lemmas summarize some useful results under Assumptions 1–7 in the main paper.

Lemma A1. *Under Assumptions 4 and 6, the matrix $\mathbf{\Delta}^{-1}$ has bounded row and column norms, where the $(i,j)^{th}$ subblock of $\mathbf{\Delta}^{-1}$, for $i,j=1,2,\ldots,N$, is given by (4).*

Proof. Consider first the row norm. By definition, we have

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

which is bounded as $||\mathbf{S}^{-1}||_\infty < K < \infty$. Likewise, we can show that the column norm of $\mathbf{\Delta}^{-1}$ is bounded, since

$$||\mathbf{\Delta}^{-1}||_1 \le \max\left\{||\mathbf{S}^{-1}||_1, 1+K||\mathbf{S}^{-1}||_1\right\} < K.$$

□

Lemma A2. *Under Assumptions 2, 4 and 6, for all t,*

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

(b) $E||\bar{\boldsymbol{\epsilon}}_{.t}||^2=O\left(N^{-1}\right)$, $E||\bar{\boldsymbol{\epsilon}}_{.t}||=O\left(N^{-1/2}\right)$,

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

†Department of Economics, Florida State University, 113 Collegiate Loop, 281 Bellamy Building, Tallahassee, FL 32306, USA. Email: cynthia.yang@fsu.edu. This work was carried out during my doctoral study at the University of Southern California.

Proof. This lemma is a direct counterpart of Lemma 1 of Pesaran (2006). Although the error terms are defined differently, 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,

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

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

$$\boldsymbol{\theta}_a \boldsymbol{\Delta}^{-1} \boldsymbol{\Sigma}_u \boldsymbol{\Delta}^{-1\prime} \boldsymbol{\theta}'_a \leq \left(\boldsymbol{\theta}_a \boldsymbol{\theta}'_a\right) \lambda_{\max}\left(\boldsymbol{\Delta}^{-1} \boldsymbol{\Sigma}_u \boldsymbol{\Delta}^{-1\prime}\right) = N^{-1} \lambda_{\max}\left(\boldsymbol{\Delta}^{-1} \boldsymbol{\Sigma}_u \boldsymbol{\Delta}^{-1\prime}\right).$$

Since $\boldsymbol{\Delta}^{-1}$ has bounded row and column norms by Lemma A1, and so does $\boldsymbol{\Sigma}_u$ under Assumption 2, it follows that the product, $\boldsymbol{\Delta}^{-1} \boldsymbol{\Sigma}_u \boldsymbol{\Delta}^{-1\prime}$, has bounded row and column norms, and consequently $\lambda_{\max}\left(\boldsymbol{\Delta}^{-1} \boldsymbol{\Sigma}_u \boldsymbol{\Delta}^{-1\prime}\right)$ is bounded, which proves that $Var(\bar{\boldsymbol{\epsilon}}_{.t})$ is of order $O\left(N^{-1}\right)$. The last statement is readily established by the definition of convergence in quadratic mean.

(b) Note that

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

and then, $E||\bar{\boldsymbol{\epsilon}}_{.t}|| \leq \left(E||\bar{\boldsymbol{\epsilon}}_{.t}||^2\right)^{1/2} = O\left(N^{-1/2}\right)$. ☐

Lemma A3. *Under Assumptions 1, 2, 3, 4 and 6, for all i,*

(a) $\frac{\bar{\boldsymbol{\epsilon}}'\bar{\boldsymbol{\epsilon}}}{T} = O_p\left(\frac{1}{N}\right)$, *(b)* $\frac{\mathbf{F}'\bar{\boldsymbol{\epsilon}}}{T} = O_p\left(\frac{1}{\sqrt{NT}}\right)$, *(c)* $\frac{\mathbf{V}'_{i.}\mathbf{F}}{T} = O_p\left(\frac{1}{\sqrt{T}}\right)$,

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

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

where $\bar{\boldsymbol{\epsilon}} = (\bar{\boldsymbol{\epsilon}}_{.1}, \bar{\boldsymbol{\epsilon}}_{.2}, \ldots, \bar{\boldsymbol{\epsilon}}_{.T})'$ *is of dimension* $T \times (k+1)$*, with* $\bar{\boldsymbol{\epsilon}}_{.t} = \boldsymbol{\Theta}_a \boldsymbol{\epsilon}_{.t}$, $\mathbf{F} = (\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_T)'$, $\mathbf{V}_{i.} = (\mathbf{v}_{i1}, \mathbf{v}_{i2}, \ldots, \mathbf{v}_{iT})'$, $\mathbf{e}_{i.} = (e_{i1}, e_{i2}, \ldots, e_{iT})'$*, and* $\mathbf{X}_{i.} = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{iT})'$.

Proof. Having established Lemma A2, results (a), (b), and (c) can be proved following similar arguments as those for (A.10)-(A.12) in Lemma 2 of Pesaran (2006), so here we give only the proofs of (d) and (e).

(d) Notice that $T^{-1}\mathbf{e}'_{i.}\bar{\boldsymbol{\epsilon}}$ is a $(k+1)$-dimensional row vector. Let $T^{-1}\mathbf{e}'_{i.}\bar{\boldsymbol{\epsilon}} = (\tilde{e}_1, \tilde{e}_2, \ldots, \tilde{e}_{k+1})$. We will consider separately its first entry and the rest, due to the composition of $\bar{\boldsymbol{\epsilon}}$.

Expanding $\tilde{e}_1$ by definition, we have

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

where $\check{s}_{ij}$ is the $(i,j)^{th}$ element of the matrix $\mathbf{S}^{-1}(\rho) = (\mathbf{I}_N - \rho\mathbf{W})^{-1}$, and $\check{s}_{.h} = \sum_{q=1}^{N} \check{s}_{qh} = O(1)$. It follows that

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

and

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

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 are given by

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

$$
E\left(e_{it}e_{is}\mathbf{v}_{ht}'\boldsymbol{\beta}\mathbf{v}_{ls}'\boldsymbol{\beta}\right) = \begin{cases} E\left(e_{it}e_{is}\right)\boldsymbol{\beta}'E\left(\mathbf{v}_{lt}\mathbf{v}_{ls}'\right)\boldsymbol{\beta}, & \text{if } h=l \\ 0, & \text{otherwise} \end{cases}.
$$

Furthermore, since e_{it} and $\mathbf{v}_{js}$ have finite fourth-order moments and their autocovariances are absolutely summable, we thus have

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

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

We now turn to the rest of the elements of $T^{-1}\mathbf{e}_{i.}'\bar{\boldsymbol{\epsilon}}$. Note that $\tilde{e}_r = (NT)^{-1}\sum_{t=1}^{T}\sum_{q=1}^{N}e_{it}v_{qt,r}$, for $r=2,3,\ldots,k+1$. Clearly $E\left(\tilde{e}_r\right)=0$, and

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

Therefore, $\tilde{e}_r = O_p\left(1/\sqrt{NT}\right)$, for $r=2,3,\ldots,k+1$. Together with the results for $\tilde{e}_1$, we conclude that $T^{-1}\mathbf{e}_{i.}'\bar{\boldsymbol{\epsilon}} = O_p\left(1/N\right)+O_p\left(1/\sqrt{NT}\right)$. The second result in (d) can be proved in a similar manner.

(e) Note that $T^{-1}\mathbf{X}_{i.}'\bar{\boldsymbol{\epsilon}} = \mathbf{A}_i'\left(T^{-1}\mathbf{F}'\bar{\boldsymbol{\epsilon}}\right)+T^{-1}\mathbf{V}_{i.}'\bar{\boldsymbol{\epsilon}}$. The claim readily follows from the results (b), (d), and the assumption that $||\mathbf{A}_i||<K$. □

Lemma A4. *Let* $\mathbf{\Pi} = \mathbf{F}\bar{\mathbf{C}}$. *Under Assumptions 1, 2, 3, 4 and 6,*

(a) $\frac{\mathbf{\Pi}'\mathbf{\Pi}}{T} = O_p(1)$, *(b)* $\frac{\mathbf{\Pi}'\bar{\boldsymbol{\epsilon}}}{T} = O_p\left(\frac{1}{\sqrt{NT}}\right)$, *(c)* $\frac{\bar{\mathbf{Z}}'\bar{\mathbf{Z}}}{T} = O_p(1)$,

(d) $\frac{\bar{\mathbf{Z}}'\mathbf{F}}{T} = O_p(1)$, *(e)* $\frac{\bar{\mathbf{Z}}'\mathbf{V}_{i.}}{T} = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{T}}\right)$, *(f)* $\frac{\bar{\mathbf{Z}}'\mathbf{X}_{i.}}{T} = O_p(1)$,

(g) $\frac{\mathbf{\Pi}'\mathbf{X}_{i.}}{T} = O_p(1)$, *(h)* $\frac{\bar{\mathbf{Z}}'\mathbf{e}_{i.}}{T} = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{T}}\right)$, *(i)* $\frac{\bar{\mathbf{Z}}'\bar{\boldsymbol{\epsilon}}}{T} = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right)$.

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) $\frac{\mathbf{\Pi}'\bar{\boldsymbol{\epsilon}}}{T} = \bar{\mathbf{C}}'\frac{\mathbf{F}'\bar{\boldsymbol{\epsilon}}}{T} = O_p\left(\frac{1}{\sqrt{NT}}\right)$, as the elements of $\bar{\mathbf{C}}$ are bounded and $\frac{\mathbf{F}'\bar{\boldsymbol{\epsilon}}}{T} = O_p\left(\frac{1}{\sqrt{NT}}\right)$ by Lemma A3.

(c) Since $\bar{\mathbf{Z}} = \mathbf{\Pi} + \bar{\boldsymbol{\epsilon}}$, we have $\frac{\bar{\mathbf{Z}}'\bar{\mathbf{Z}}}{T} = \frac{\mathbf{\Pi}'\mathbf{\Pi}}{T} + \frac{\bar{\boldsymbol{\epsilon}}'\bar{\boldsymbol{\epsilon}}}{T} + 2\frac{\mathbf{\Pi}'\bar{\boldsymbol{\epsilon}}}{T} + \frac{\bar{\boldsymbol{\epsilon}}'\mathbf{\Pi}}{T} = O_p(1) + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right) = O_p(1)$.

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

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

(f) Recall that $\mathbf{X}_{i.} = \mathbf{F}\mathbf{A}_i + \mathbf{V}_{i.}$, and then it follows that $\frac{\bar{\mathbf{Z}}'\mathbf{X}_{i.}}{T} = \frac{\bar{\mathbf{Z}}'\mathbf{F}}{T}\mathbf{A}_i + \frac{\bar{\mathbf{Z}}'\mathbf{V}_{i.}}{T} = O_p(1) + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{T}}\right) = O_p(1)$.

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

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

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

Lemma A5. *Under Assumptions 1-6, for any* i *and* j,

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

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

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

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

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

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

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

Proof. (a) From (8), we have

$$\mathbf{F} = \left(\bar{\mathbf{C}}\bar{\mathbf{C}}'\right)^{-1}\bar{\mathbf{C}}\left(\bar{\mathbf{Z}} - \bar{\boldsymbol{\epsilon}}\right). \tag{S.1}$$

Since $\bar{\mathbf{M}}\bar{\mathbf{Z}} = \mathbf{0}$, it follows that

$$\mathbf{X}_{i.}'\bar{\mathbf{M}}\mathbf{F} = -\left(\mathbf{X}_{i.}'\bar{\mathbf{M}}\bar{\boldsymbol{\epsilon}}\right)\bar{\mathbf{C}}'\left(\bar{\mathbf{C}}\bar{\mathbf{C}}'\right)^{-1}, \tag{S.2}$$

As $\mathbf{X}_{i.} = \mathbf{F}\mathbf{A}_i + \mathbf{V}_{i.}$, we have

$$\mathbf{X}_{i.}'\bar{\mathbf{M}}\bar{\boldsymbol{\epsilon}} = \mathbf{A}_i'\mathbf{F}'\bar{\mathbf{M}}\bar{\boldsymbol{\epsilon}} + \mathbf{V}_{i.}'\bar{\mathbf{M}}\bar{\boldsymbol{\epsilon}}. \tag{S.3}$$

Using (S.1) again gives

$$\bar{\boldsymbol{\epsilon}}'\bar{\mathbf{M}}\mathbf{F} = -\left(\bar{\boldsymbol{\epsilon}}'\bar{\mathbf{M}}\bar{\boldsymbol{\epsilon}}\right)\bar{\mathbf{C}}'\left(\bar{\mathbf{C}}\bar{\mathbf{C}}'\right)^{-1}. \tag{S.4}$$

Then substituting (S.3) and (S.4) into (S.2) yields

$$\mathbf{X}_{i.}'\bar{\mathbf{M}}\mathbf{F} = -\mathbf{A}_i'\left(\bar{\mathbf{C}}\bar{\mathbf{C}}'\right)^{-1}\bar{\mathbf{C}}\bar{\boldsymbol{\epsilon}}'\bar{\mathbf{M}}\bar{\boldsymbol{\epsilon}}\bar{\mathbf{C}}'\left(\bar{\mathbf{C}}\bar{\mathbf{C}}'\right)^{-1} - \mathbf{V}_{i.}'\bar{\mathbf{M}}\bar{\boldsymbol{\epsilon}}\bar{\mathbf{C}}'\left(\bar{\mathbf{C}}\bar{\mathbf{C}}'\right)^{-1}. \tag{S.5}$$

Since $\mathbf{A}_i'$ and $\left(\bar{\mathbf{C}}\bar{\mathbf{C}}'\right)^{-1}\bar{\mathbf{C}}$ have bounded norms, now we only need to establish the probability orders of $\left\|T^{-1}\bar{\boldsymbol{\epsilon}}'\bar{\mathbf{M}}\bar{\boldsymbol{\epsilon}}\right\|$ and $\left\|T^{-1}\mathbf{V}_{i.}'\bar{\mathbf{M}}\bar{\boldsymbol{\epsilon}}\right\|$, and then the assertion in (a) will follow. Expanding $\bar{\mathbf{M}}$ by definition and applying Lemma A3(a), Lemma A4(c) and (i) leads to

$$\frac{\bar{\boldsymbol{\epsilon}}'\bar{\mathbf{M}}\bar{\boldsymbol{\epsilon}}}{T} = \frac{\bar{\boldsymbol{\epsilon}}'\bar{\boldsymbol{\epsilon}}}{T} - \left(\frac{\bar{\boldsymbol{\epsilon}}'\bar{\mathbf{Z}}}{T}\right)\left(\frac{\bar{\mathbf{Z}}'\bar{\mathbf{Z}}}{T}\right)^{-1}\left(\frac{\bar{\mathbf{Z}}'\bar{\boldsymbol{\epsilon}}}{T}\right) = O_p\left(\frac{1}{N}\right). \tag{S.6}$$

Similarly, by Lemma A3(d), Lemma A4(c), (e) and (i), we have

$$\frac{\mathbf{V}_{i.}'\bar{\mathbf{M}}\bar{\boldsymbol{\epsilon}}}{T} = \frac{\mathbf{V}_{i.}'\bar{\boldsymbol{\epsilon}}}{T} - \left(\frac{\mathbf{V}_{i.}'\bar{\mathbf{Z}}}{T}\right)\left(\frac{\bar{\mathbf{Z}}'\bar{\mathbf{Z}}}{T}\right)^{-1}\left(\frac{\bar{\mathbf{Z}}'\bar{\boldsymbol{\epsilon}}}{T}\right) = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right). \tag{S.7}$$

Hence the result in (a) follows.

The rest of Lemma A5 can be proved by applying Lemma A4 and using similar reasoning as that for Lemma 3 in Kapetanios et al. (2011). To save space, we only give the proof of (b) to illustrate the main idea.

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

$$\left\|\frac{\mathbf{X}_{i.}'\bar{\mathbf{M}}\mathbf{X}_{j.}}{T} - \frac{\mathbf{X}_{i.}'\mathbf{M}_\Pi\mathbf{X}_{j.}}{T}\right\| = \left\|\frac{\mathbf{X}_{i.}'\bar{\mathbf{Z}}\left(\bar{\mathbf{Z}}'\bar{\mathbf{Z}}\right)^{-1}\bar{\mathbf{Z}}'\mathbf{X}_{j.}}{T} - \frac{\mathbf{X}_{i.}'\boldsymbol{\Pi}\left(\boldsymbol{\Pi}'\boldsymbol{\Pi}\right)^{-1}\boldsymbol{\Pi}'\mathbf{X}_{j.}}{T}\right\| \leq d_1 + d_2 + d_3,$$

where

$$d_1 \equiv \left\|T^{-1}\left(\mathbf{X}_{i.}'\bar{\mathbf{Z}} - \mathbf{X}_i'\boldsymbol{\Pi}\right)\left(\bar{\mathbf{Z}}'\bar{\mathbf{Z}}\right)^{-1}\bar{\mathbf{Z}}'\mathbf{X}_{j.}\right\| \leq \left\|\frac{\mathbf{X}_{i.}'\bar{\boldsymbol{\epsilon}}}{T}\right\|\left\|\left(\frac{\bar{\mathbf{Z}}'\bar{\mathbf{Z}}}{T}\right)^{-1}\frac{\bar{\mathbf{Z}}'\mathbf{X}_{j.}}{T}\right\| = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right),$$

$$\begin{aligned}
d_2 &\equiv \left\|T^{-1}\mathbf{X}_{i.}'\boldsymbol{\Pi}\left[\left(\bar{\mathbf{Z}}'\bar{\mathbf{Z}}\right)^{-1} - \left(\boldsymbol{\Pi}'\boldsymbol{\Pi}\right)^{-1}\right]\bar{\mathbf{Z}}'\mathbf{X}_{j.}\right\| \\
&\leq \left\|-\frac{\bar{\boldsymbol{\epsilon}}'\bar{\boldsymbol{\epsilon}}}{T} - \frac{\boldsymbol{\Pi}'\bar{\boldsymbol{\epsilon}}}{T} - \frac{\bar{\boldsymbol{\epsilon}}'\boldsymbol{\Pi}}{T}\right\|\left\|\frac{\mathbf{X}_{i.}'\boldsymbol{\Pi}}{T}\left(\frac{\bar{\mathbf{Z}}'\bar{\mathbf{Z}}}{T}\right)^{-1}\right\|\left\|\left(\frac{\boldsymbol{\Pi}'\boldsymbol{\Pi}}{T}\right)^{-1}\frac{\bar{\mathbf{Z}}'\mathbf{X}_{j.}}{T}\right\| \\
&= O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right),
\end{aligned}$$

$$d_3 \equiv \left\|T^{-1}\mathbf{X}_{i.}'\boldsymbol{\Pi}\left(\boldsymbol{\Pi}'\boldsymbol{\Pi}\right)^{-1}\left(\bar{\mathbf{Z}}'\mathbf{X}_{j.} - \boldsymbol{\Pi}'\mathbf{X}_{j.}\right)\right\| \leq \left\|\frac{\mathbf{X}_{i.}'\boldsymbol{\Pi}}{T}\left(\frac{\boldsymbol{\Pi}'\boldsymbol{\Pi}}{T}\right)^{-1}\right\|\left\|\frac{\bar{\boldsymbol{\epsilon}}'\mathbf{X}_{j.}}{T}\right\| = O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right).$$

Under the full rank condition given in Assumption 5, $\mathbf{M}_{\Pi} = \mathbf{M}_f$, and hence we have

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

$\square$

Lemma A6. *Under Assumptions 1-7,*

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

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

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

where $\mathbf{B} = (b_{ij})$ *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

$$\mathbf{Q}_c = \left[\left(\mathbf{W}^r \mathbf{x}_{.1,p}\right)', \left(\mathbf{W}^r \mathbf{x}_{.2,p}\right)', \ldots, \left(\mathbf{W}^r \mathbf{x}_{.T,p}\right)'\right]',$$

where $r = 0, 1, 2, \ldots$, $p = 1, 2, \ldots, k$, $\mathbf{x}_{.t,p} = (x_{1t,p}, x_{2t,p}, ., x_{Nt,p})'$, and $\mathbf{W}^0 \equiv \mathbf{I}_N$, we have

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

Evidently, the claim in (a) readily follows Lemma A5(a), and the assumptions that $\boldsymbol{\gamma}_i$ is bounded and $\mathbf{W}$ has bounded row and column norms.

(b) Taking the p^{th} column from $\mathbf{Q}$, $p = 1, 2, \ldots, k$, as in the proof of (a) we can show that

$$\begin{aligned}
\frac{1}{NT}\mathbf{Q}_c'\left(\bar{\mathbf{M}} \otimes \mathbf{B}\right)\mathbf{e} =& \frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N} b_{ij} w_{il}^r \mathbf{x}_{l.,p}'\bar{\mathbf{M}}\mathbf{e}_{j.} \\
=& \frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N} b_{ij} w_{il}^r \mathbf{x}_{l.,p}'\mathbf{M}_f\mathbf{e}_{j.} + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right) \\
=& \frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N} b_{ij} w_{il}^r \mathbf{v}_{l.,p}'\mathbf{e}_{j.} + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right), \qquad \text{(S.8)}
\end{aligned}$$

where the second equality follows by Lemma A5(c) and the assumption that $\mathbf{B}$ and $\mathbf{W}$ have bounded row and column norms.

Consider the first term in (S.8). Its mean is zero and its variance is given by

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

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

$$T^{-1}\mathbf{v}'_{m.,p}\boldsymbol{\Omega}_{e,j}\mathbf{v}_{n.,p} \le T^{-1}\lambda_{\max}(\boldsymbol{\Omega}_{e,j})\mathbf{v}'_{m.,p}\mathbf{v}_{n.,p} \le KT^{-1}\mathbf{v}'_{m.,p}\mathbf{v}_{n.,p} = O(1).$$

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

$$Var\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N} b_{ij}w_{il}^{r}\mathbf{v}'_{l.,p}\mathbf{e}_{j.}\right) = O\left(\frac{1}{NT}\right),$$

and consequently the order of the first term in (S.8) is $O_p\left(1/\sqrt{NT}\right)$, which completes the proof.

(c) Let $\mathbf{C} = \mathbf{B}'\mathbf{W}^r = (c_{ij})$. For any column of $\mathbf{Q}$, we have

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

Again, the last line follows by Lemma A5 and $\sum_{j=1}^{N} c_{ji} = O(1)$.

□

Lemma A7. *Under Assumptions 1-6, for any $N\times N$ nonstochastic matrix $\mathbf{B} = (b_{ij})$ with bounded row and column norms,*

(a) $\frac{1}{NT}\mathbf{e}'\left(\bar{\mathbf{M}}\otimes\mathbf{B}\right)\mathbf{e} - \frac{1}{N}\sum_{i=1}^{N} b_{ii}\sigma_i^2 = o_p(1)$,

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

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

Proof. (a) Applying Lemma A5, we have

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

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

$$E\left[\frac{1}{NT}\mathbf{e}'\left(\mathbf{I}_T\otimes\mathbf{B}\right)\mathbf{e}\right] = \frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}b_{ji}E\left(\mathbf{e}'_{i.}\mathbf{e}_{j.}\right) = \frac{1}{N}\sum_{i=1}^{N}b_{ii}\sigma_i^2 = O(1), \tag{S.9}$$

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

Next, consider the second moment

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

where

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

It follows that

$$\begin{aligned}Var\left[\frac{1}{NT}\mathbf{e}'\left(\mathbf{I}_T\otimes\mathbf{B}\right)\mathbf{e}\right]=&\frac{1}{N^2T^2}\Bigg\{\sum_{i=1}^{N}b_{ii}^2E\left[\left(\mathbf{e}_{i.}'\mathbf{e}_{i.}\right)^2\right]+\sum_{i=1}^{N}\sum_{l=1,l\neq i}^{N}b_{ii}b_{ll}E\left(\mathbf{e}_{i.}'\mathbf{e}_{i.}\right)E\left(\mathbf{e}_{l.}'\mathbf{e}_{l.}\right)\\&+\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\left(b_{ji}^2+b_{ji}b_{ij}\right)E\left[\left(\mathbf{e}_{i.}'\mathbf{e}_{j.}\right)^2\right]\Bigg\}-\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{l=1}^{N}b_{ii}b_{ll}E\left(\mathbf{e}_{i.}'\mathbf{e}_{i.}\right)E\left(\mathbf{e}_{l.}'\mathbf{e}_{l.}\right)\\=&\frac{1}{N^2T^2}\sum_{i=1}^{N}b_{ii}^2\left\{E\left[\left(\mathbf{e}_{i.}'\mathbf{e}_{i.}\right)^2\right]-\left[E\left(\mathbf{e}_{i.}'\mathbf{e}_{i.}\right)\right]^2-2tr\left(\mathbf{\Omega}_{e,i}\mathbf{\Omega}_{e,i}\right)\right\}\\&+\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}b_{ji}\left(b_{ji}+b_{ij}\right)tr\left(\mathbf{\Omega}_{e,i}\mathbf{\Omega}_{e,j}\right).\end{aligned}\tag{S.10}$$

It is readily seen that if $b_{ii}=0$, for $i=1,2,\ldots,N$, then

$$Var\left[\frac{1}{NT}\mathbf{e}'\left(\mathbf{M}_f\otimes\mathbf{B}\right)\mathbf{e}\right]=\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}b_{ji}\left(b_{ji}+b_{ij}\right)tr\left(\mathbf{\Omega}_{e,i}\mathbf{\Omega}_{e,j}\right)=O\left(\frac{1}{NT}\right),\tag{S.11}$$

where the second equality follows from the assumption that $\mathbf{B}$ and $\mathbf{\Omega}_{e,i}$, for all i, are uniformly bounded in row and column sums. In general, when $diag(\mathbf{B})\neq\mathbf{0}$, the first term in (S.10) does not equal zero but is of order $O(N^{-1}T^{-1})$ since

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

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

$$Var\left[\frac{1}{NT}\mathbf{e}'\left(\mathbf{I}_T\otimes\mathbf{B}\right)\mathbf{e}\right]=O\left(\frac{1}{NT}\right),$$

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

(b) Let $\mathbf{C}=\mathbf{\Gamma}'\mathbf{B}$, then $(NT)^{-1}\,\mathbf{f}'\left(\bar{\mathbf{M}}\otimes\mathbf{\Gamma}'\mathbf{B}\right)\mathbf{e}=(NT)^{-1}\sum_{i=1}^{N}\sum_{j=1}^{m}\mathbf{e}_{i.}'\bar{\mathbf{M}}\mathbf{F}c_{ji}$. Its probability order is immediately established by applying Lemma A5 and noting that all elements c_{ij} are uniformly bounded.

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

Lemma A8. *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) $E\left[\mathbf{e}'\left(\mathbf{I}_T\otimes\mathbf{B}\right)\mathbf{e}\right]=0$,

(b) $E\left\{\left[\mathbf{e}'\left(\mathbf{I}_T\otimes\mathbf{B}\right)\mathbf{e}\right]^2\right\}=Ttr\left[\left(\mathbf{B}\odot\mathbf{B}^s\right)\boldsymbol{\Sigma}_{eT}\right]=T\sum_{i=1}^{N}\sum_{j=1}^{N}b_{ji}(b_{ij}+b_{ji})\varsigma_{eT,ij}$,

(c) $E\left[\mathbf{e}'\left(\mathbf{I}_T\otimes\mathbf{B}\right)\mathbf{e}\mathbf{e}'\left(\mathbf{I}_T\otimes\mathbf{D}\right)\mathbf{e}\right]=Ttr\left[\left(\mathbf{B}\odot\mathbf{D}^s\right)\boldsymbol{\Sigma}_{eT}\right]=T\sum_{i=1}^{N}\sum_{j=1}^{N}b_{ji}(d_{ij}+d_{ji})\varsigma_{eT,ij}$,

where $\mathbf{B}^s=\mathbf{B}+\mathbf{B}'$, $\mathbf{D}^s$ is defined similarly, and $\boldsymbol{\Sigma}_{eT}=(\varsigma_{eT,ij})$ is an $N\times N$ matrix of which the $(i,j)^{th}$ element is given by $\varsigma_{eT,ij}=T^{-1}tr\left(\mathbf{\Omega}_{e,i}\mathbf{\Omega}_{e,j}\right)$.

Proof. Results (a) and (b) follow from (S.9) and (S.11) in the proof of Lemma A7(a). The result in (c) can be verified similarly. □

Lemma A9. *Consider the following linear-quadratic form: $h=\mathbf{e}'\left(\mathbf{I}_T\otimes\mathbf{B}\right)\mathbf{e}+\mathbf{c}'\mathbf{e}$, where $\mathbf{e}$ is an $NT\times 1$ vector of disturbances following the data generating process 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*

$$\sigma_h^2=\sum_{i=1}^{N}\sum_{j=1}^{N}b_{ji}\left(b_{ij}+b_{ji}\right)tr\left(\mathbf{\Omega}_{e,i}\mathbf{\Omega}_{e,j}\right)+\sum_{i=1}^{N}\mathbf{c}_{i.}'\mathbf{\Omega}_{e,i}\mathbf{c}_{i.}.$$

If $(NT)^{-1}\sigma_h^2$ is bounded away from zero, we have $h/\sigma_h\overset{d}{\to}N(0,1)$ 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, $i=1,2,\ldots,N$, forms a martingale difference array with respect to the $\sigma-$field generated by $\{\mathbf{e}_{1.},\mathbf{e}_{2.},\ldots,\mathbf{e}_{i-1.}\}$, since

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

To apply a martingale difference central limit theorem (CLT), we only need to show that the following two sufficient conditions hold (see, for example, Kelejian and Prucha, 2001, Theorem A.1): (i) $\frac{1}{\sigma_h^{2+\delta}}\sum_{i=1}^{N}E|h_i|^{2+\delta}\to 0$, for some $\delta>0$, and (ii) $\frac{1}{\sigma_h^2}\sum_{i=1}^{N}E\left(h_i^2|1,2,\ldots,i-1\right)\overset{p}{\to}1$.

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

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

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

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

By the C_r inequality, $E\left(|\mathbf{e}_{i.}'\mathbf{e}_{j.}|^q\right) \leq T^{q-1}\sum_{t=1}^T E|e_{it}|^q E|e_{jt}|^q = O\left(T^{q-1}\right)$, 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, $E\left(|\mathbf{c}_{i.}'\mathbf{e}_{i.}|^q\right) \leq T^{q-1}\sum_{t=1}^T |c_{it}|^q E|e_{it}|^q = O\left(T^{q-1}\right)$. As a result, $\sum_{i=1}^N E|h_i|^{2+\delta} = O\left(NT^{1+\delta}\right)$, 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),

$$\sum_{i=1}^N E\left(h_i^2|1,2,\ldots,i-1\right) - \sigma_h^2 = r_1 + 2r_2,$$

where

$$\begin{aligned}
r_1 &= \sum_{i=1}^N\sum_{j=1}^{i-1} b_{ji}\left(b_{ij}+b_{ji}\right)\left[E\left(\mathbf{e}_i'\mathbf{e}_{j.}\mathbf{e}_j'\mathbf{e}_{i.}|1,2,\ldots,i-1\right) - E\left(\mathbf{e}_i'\mathbf{e}_{j.}\mathbf{e}_j'\mathbf{e}_{i.}\right)\right] \\
&= \sum_{i=1}^N\sum_{j=1}^{i-1} b_{ji}\left(b_{ij}+b_{ji}\right)\sum_{t=1}^T\sum_{s=1}^T E(e_{it}e_{is})\left[e_{jt}e_{js} - E(e_{jt}e_{js})\right], \\
r_2 &= \sum_{i=1}^N E\left[\left(\sum_{j=1}^N b_{ji}\mathbf{e}_i'\mathbf{e}_{j.}\right)\left(\mathbf{c}_{i.}'\mathbf{e}_{i.}\right)\middle| 1,2,\ldots,i-1\right] = \sum_{i=1}^N\sum_{j=1}^{i-1}\sum_{t=1}^T\sum_{s=1}^T b_{ji}c_{is}E\left(e_{it}e_{is}\right)e_{jt}.
\end{aligned}$$

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 boundedness of the row and column norms of $\mathbf{B}$, we have $Var(r_1) = O(NT)$. Next,

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

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

$$\frac{1}{\sigma_h^2}\sum_{i=1}^{N} E\left(h_i^2|1,2,\ldots,i-1\right) - 1 = \frac{\frac{1}{NT}\left[\sum_{i=1}^{N} E\left(h_i^2|1,\ldots,i-1\right) - \sigma_h^2\right]}{\frac{1}{NT}\sigma_h^2} = O\left(\frac{1}{\sqrt{NT}}\right),$$

which proves (ii). □

Lemma A10. *Let* $\mathbf{A} = (a_{ij})$ *be an* $N \times N$ *matrix. Then,*

$$\frac{tr\left(\mathbf{A}^2 + \mathbf{A}\mathbf{A}'\right)}{N} - \frac{2\left[tr\left(\mathbf{A}\right)\right]^2}{N^2} \geq 0, \tag{S.12}$$

for all N*, including* $N \to \infty$.

Proof. It is clear from the definition of trace that

$$\begin{aligned}\frac{tr\left(\mathbf{A}^2 + \mathbf{A}\mathbf{A}'\right)}{N} - \frac{2\left[tr\left(\mathbf{A}\right)\right]^2}{N^2} &= \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(a_{ij}^2 + a_{ij}a_{ji} - \frac{2}{N}a_{ii}a_{jj}\right) \\ &= 2\left[\frac{1}{N}\sum_i a_{ii}^2 - \left(\frac{1}{N}\sum_i a_{ii}\right)\left(\frac{1}{N}\sum_j a_{jj}\right)\right] + \frac{1}{N}\sum_i\sum_{j\neq i}\left(a_{ij}^2 + a_{ij}a_{ji}\right).\end{aligned}$$

By applying the Cauchy-Schwarz inequality,

$$\frac{1}{N}\sum_i a_{ii}^2 \geq \left(\frac{1}{N}\sum_i a_{ii}\right)\left(\frac{1}{N}\sum_j a_{jj}\right), \tag{S.13}$$

and noting that

$$\begin{aligned}\frac{1}{N}\sum_i\sum_{j\neq i}\left(a_{ij}^2 + a_{ij}a_{ji}\right) &= \frac{1}{N}\left(\sum_i\sum_{j>i}a_{ij}^2 + \sum_i\sum_{j<i}a_{ij}^2 + 2\sum_i\sum_{j>i}a_{ij}a_{ji}\right) \\ &= \frac{1}{N}\left(\sum_i\sum_{j>i}a_{ij}^2 + \sum_i\sum_{j>i}a_{ji}^2 + 2\sum_i\sum_{j>i}a_{ij}a_{ji}\right) \\ &= \frac{1}{N}\sum_i\sum_{j>i}(a_{ij} + a_{ji})^2 \geq 0, \qquad (S.14)\end{aligned}$$

the result given by (S.12) follows immediately. The equality in (S.12) is reached if and only if both equalities in (S.13) and (S.14) hold true. In particular, when N is finite, (S.13) becomes an equality if and only if $a_{11} = a_{22} = \ldots = a_{NN}$, and (S.14) becomes an equality if and only if $a_{ij} = -a_{ji}$, for $i \neq j$.

$\square$

S1.2 Derivations of Identification Conditions

Model (12) in the main paper: Without exogenous variables $\mathbf{x}_{it}$

Consider $Q_{NT}(\boldsymbol{\psi})$ as defined in (14) of the main text. The first derivatives are

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

where $\mathbf{g}_{0,i}'$ is the i^{th} row of $\mathbf{G}_0$, and $\mathbf{H}_f\left(\rho_0, \boldsymbol{\gamma}_0'\right)$ is given by

$$\mathbf{H}_f\left(\rho_0, \boldsymbol{\gamma}_0'\right) = (NT)^{-1} E_0 \sum_{t=1}^{T} \left(\boldsymbol{J}_{0,t}'\boldsymbol{J}_{0,t}\right), \quad \boldsymbol{J}_{0,t} = \begin{pmatrix} \mathbf{G}_0\boldsymbol{\Gamma}_0\mathbf{f}_t, & \mathbf{F}_t \end{pmatrix}.$$

The second derivatives are given by

$$\begin{aligned}
\boldsymbol{\Lambda}_{f,NT}(\boldsymbol{\psi}) =& \frac{\partial^2 Q_{NT}(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}\boldsymbol{\psi}'} = \begin{pmatrix} \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial^2 d} & \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial d\partial\boldsymbol{\zeta}'} & \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial d\partial\vartheta} \\ . & \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial\boldsymbol{\zeta}\partial\boldsymbol{\zeta}'} & \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial\boldsymbol{\zeta}\partial\vartheta} \\ . & . & \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial^2\vartheta} \end{pmatrix} \\
=& \begin{pmatrix} \boldsymbol{\Lambda}_{f,11} & \boldsymbol{\Lambda}_{f,12} & \boldsymbol{\Lambda}_{f,13} \\ . & \boldsymbol{\Lambda}_{f,22} & \boldsymbol{\Lambda}_{f,23} \\ . & . & \boldsymbol{\Lambda}_{f,33} \end{pmatrix}, \qquad \text{(S.15)}
\end{aligned}$$

where

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

$$\boldsymbol{\Lambda}_{f,22} = \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial\boldsymbol{\zeta}\partial\boldsymbol{\zeta}'} = \left\{\frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial\boldsymbol{\zeta}_i\partial\boldsymbol{\zeta}_j'}\right\} = \left\{\frac{\sigma_0^2(1-\vartheta)}{NT}E_0\sum_{t=1}^T \mathbf{f}_t\mathbf{f}_t', \text{ if } i=j; \text{ and } \mathbf{0}, \text{ if } i\neq j\right\},$$

$$\boldsymbol{\Lambda}_{f,23} = \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial\boldsymbol{\zeta}\partial\vartheta} = \left\{\frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial\boldsymbol{\zeta}_i\partial\vartheta}\right\} = \left\{-\frac{\sigma_0^2}{NT}E_0\sum_{t=1}^T \left(d\mathbf{f}_t\mathbf{g}_{0,i}'\boldsymbol{\Gamma}_0\mathbf{f}_t + \mathbf{f}_t\mathbf{f}_t'\boldsymbol{\zeta}_i\right)\right\},$$

$$\boldsymbol{\Lambda}_{f,33} = \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial^2\vartheta} = \frac{1}{2(1-\vartheta)^2}.$$

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

$$\boldsymbol{\Lambda}_{f,NT}(\mathbf{0}) = \begin{pmatrix} \boldsymbol{\Lambda}_{f,11}(\mathbf{0}) & \frac{\sigma_0^2}{NT}E_0\sum_{t=1}^T\mathbf{g}_{0,1}'\boldsymbol{\Gamma}_0\mathbf{f}_t\mathbf{f}_t' & \frac{\sigma_0^2}{NT}E_0\sum_{t=1}^T\mathbf{g}_{0,2}'\boldsymbol{\Gamma}_0\mathbf{f}_t\mathbf{f}_t' & \cdots & \frac{\sigma_0^2}{NT}E_0\sum_{t=1}^T\mathbf{g}_{0,N}'\boldsymbol{\Gamma}_0\mathbf{f}_t\mathbf{f}_t' & \frac{tr(\mathbf{G}_0)}{N} \\ . & \frac{\sigma_0^2}{NT}E_0\sum_{t=1}^T\mathbf{f}_t\mathbf{f}_t' & \mathbf{0}_{m\times m} & \cdots & \mathbf{0}_{m\times m} & \mathbf{0}_{m\times 1} \\ . & \mathbf{0}_{m\times m} & \frac{\sigma_0^2}{NT}E_0\sum_{t=1}^T\mathbf{f}_t\mathbf{f}_t' & \mathbf{0}_{m\times m} & \mathbf{0}_{m\times m} & \mathbf{0}_{m\times 1} \\ \vdots & \vdots & \mathbf{0}_{m\times m} & \ddots & \vdots & \vdots \\ . & \mathbf{0}_{m\times m} & \mathbf{0}_{m\times m} & \cdots & \frac{\sigma_0^2}{NT}E_0\sum_{t=1}^T\mathbf{f}_t\mathbf{f}_t' & \mathbf{0}_{m\times 1} \\ . & 0 & 0 & \cdots & 0 & \frac{1}{2} \end{pmatrix}, \tag{S.16}$$

where

$$\boldsymbol{\Lambda}_{f,11}(\mathbf{0}) = \frac{tr\left(\mathbf{G}_0^2+\mathbf{G}_0\mathbf{G}_0'\right)}{N} + \frac{\sigma_0^2}{NT}E_0\sum_{t=1}^T\boldsymbol{R}_{0,t}'\boldsymbol{R}_{0,t},$$

and $\boldsymbol{R}_{0,t} = \mathbf{G}_0\boldsymbol{\Gamma}_0\mathbf{f}_t$. The determinant of $\boldsymbol{\Lambda}_{f,11}(\mathbf{0})$ can be computed as follows:

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

where $\bar{\boldsymbol{g}}_0 = \begin{bmatrix} N^{-1}tr(\mathbf{G}_0), & \mathbf{0}_{1\times Nm} \end{bmatrix}'$ and $h_g = N^{-1}tr\left(\mathbf{G}_0^2+\mathbf{G}_0\mathbf{G}_0'\right) - 2N^{-2}\left[tr\left(\mathbf{G}_0\right)\right]^2$. Also note that $\mathbf{F}_t'\mathbf{F}_t = \mathbf{I}_N\otimes\mathbf{f}_t\mathbf{f}_t'$. Therefore, $\det\left[\boldsymbol{\Lambda}_{f,NT}(\mathbf{0})\right] > 0$ if and only if $h_g > 0$ and $T^{-1}E_0\left(\mathbf{f}_t\mathbf{f}_t'\right)$ is positive definite. This establishes the identification conditions in Proposition 1 of the main paper.

Model (19) in the main paper: With exogenous variables $\mathbf{x}_{it}$

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

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

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

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

where $\mathbf{G}_0^b = \mathbf{I}_T \otimes \mathbf{G}_0 = \mathbf{I}_T \otimes \left(\mathbf{W}\mathbf{S}_0^{-1}\right)$, $\mathbf{H}_f(\rho_0, \boldsymbol{\gamma}_0')$ is given by (13), and $\mathbf{H}(\rho_0, \boldsymbol{\beta}_0')$ is given by (20). Hence, we obtain

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

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

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

The first derivatives are given by

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

The second derivatives are given by

$$\boldsymbol{\Lambda}_{NT}(\boldsymbol{\psi}) = \frac{\partial^2 Q_{NT}(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}\boldsymbol{\psi}'} = \begin{pmatrix} \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial^2 d} & \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial d\partial\boldsymbol{\zeta}'} & \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial d\partial\boldsymbol{\chi}'} & \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial d\partial\vartheta} \\ . & \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial\boldsymbol{\zeta}\partial\boldsymbol{\zeta}'} & \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial\boldsymbol{\zeta}\partial\boldsymbol{\chi}'} & \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial\boldsymbol{\zeta}\partial\vartheta} \\ . & . & \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial\boldsymbol{\chi}\partial\boldsymbol{\chi}'} & \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial\boldsymbol{\chi}\partial\vartheta} \\ . & . & . & \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial^2\vartheta} \end{pmatrix}$$

$$= \begin{pmatrix} \boldsymbol{\Lambda}_{11} & \boldsymbol{\Lambda}_{12} & \boldsymbol{\Lambda}_{13} & \boldsymbol{\Lambda}_{14} \\ . & \boldsymbol{\Lambda}_{22} & \boldsymbol{\Lambda}_{23} & \boldsymbol{\Lambda}_{24} \\ . & . & \boldsymbol{\Lambda}_{33} & \boldsymbol{\Lambda}_{34} \\ . & . & . & \boldsymbol{\Lambda}_{44} \end{pmatrix}, \tag{S.17}$$

where

$$\boldsymbol{\Lambda}_{11} = \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial^2 d} = \frac{1}{N} tr\left[(\mathbf{I}_N - d\mathbf{G}_0)^{-1}\mathbf{G}_0(\mathbf{I}_N - d\mathbf{G}_0)^{-1}\mathbf{G}_0\right] + (1-\vartheta)\frac{tr(\mathbf{G}_0\mathbf{G}_0')}{N}$$
$$+ \frac{\sigma_0^2(1-\vartheta)}{NT} E_0\left[(\mathbf{G}_0^b\mathbf{X}\boldsymbol{\beta}_0)'\mathbf{G}_0^b\mathbf{X}\boldsymbol{\beta}_0 + (\mathbf{G}_0\boldsymbol{\Gamma}_0\mathbf{f}_t)'\mathbf{G}_0\boldsymbol{\Gamma}_0\mathbf{f}_t\right],$$

$$\boldsymbol{\Lambda}_{12} = \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial d\partial\boldsymbol{\zeta}'} = \frac{\sigma_0^2(1-\vartheta)}{NT} E_0\left[(\mathbf{G}_0^b\mathbf{X}\boldsymbol{\beta}_0)'\mathbf{X}\right],$$

$$\boldsymbol{\Lambda}_{13} = \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial d\partial\boldsymbol{\chi}'} = \left\{\frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial d\partial\boldsymbol{\chi}_i'}\right\} = \left\{\frac{\sigma_0^2(1-\vartheta)}{NT} E_0 \sum_{t=1}^{T} \mathbf{g}_{0,i}'\boldsymbol{\Gamma}_0\mathbf{f}_t\mathbf{f}_t'\right\}$$

$$\boldsymbol{\Lambda}_{14} = \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial d\partial\vartheta} = \frac{tr(\mathbf{G}_0)}{N} - d\frac{tr(\mathbf{G}_0\mathbf{G}_0')}{N}$$
$$- \frac{\sigma_0^2}{NT} E_0\left[d(\mathbf{G}_0^b\mathbf{X}\boldsymbol{\beta}_0)'\mathbf{G}_0^b\mathbf{X}\boldsymbol{\beta}_0 + (\mathbf{G}_0^b\mathbf{X}\boldsymbol{\beta}_0)'\mathbf{X}\boldsymbol{\zeta} + d(\mathbf{G}_0\boldsymbol{\Gamma}_0\mathbf{f}_t)'\mathbf{G}_0\boldsymbol{\Gamma}_0\mathbf{f}_t + (\mathbf{G}_0\boldsymbol{\Gamma}_0\mathbf{f}_t)'\mathbf{F}_t\boldsymbol{\chi}\right],$$

$$\boldsymbol{\Lambda}_{22} = \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial\boldsymbol{\zeta}\partial\boldsymbol{\zeta}'} = \frac{\sigma_0^2(1-\vartheta)}{NT} E_0\left(\mathbf{X}'\mathbf{X}\right),$$

$$\boldsymbol{\Lambda}_{23} = \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial\boldsymbol{\zeta}\partial\boldsymbol{\chi}'} = \mathbf{0},$$

$$\boldsymbol{\Lambda}_{24} = \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial\boldsymbol{\zeta}\partial\vartheta} = -\frac{\sigma_0^2}{NT} E_0\left(d\mathbf{X}'\mathbf{G}_0^b\mathbf{X}\boldsymbol{\beta}_0 + \mathbf{X}'\mathbf{X}\boldsymbol{\zeta}\right),$$

$$\boldsymbol{\Lambda}_{33} = \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial\boldsymbol{\chi}\partial\boldsymbol{\chi}'} = \left\{\frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial\boldsymbol{\chi}_i\partial\boldsymbol{\chi}_j'}\right\} = \left\{\frac{\sigma_0^2(1-\vartheta)}{NT} E_0 \sum_{t=1}^{T} \mathbf{f}_t\mathbf{f}_t', \text{ if } i=j; \text{ and } \mathbf{0}, \text{ if } i \neq j\right\},$$

$$\boldsymbol{\Lambda}_{34} = \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial\boldsymbol{\chi}\partial\vartheta} = \left\{\frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial\boldsymbol{\chi}_i\partial\vartheta}\right\} = \left\{-\frac{\sigma_0^2}{NT} E_0 \sum_{t=1}^{T}\left(d\mathbf{f}_t\mathbf{g}_{0,i}'\boldsymbol{\Gamma}_0\mathbf{f}_t + \mathbf{f}_t\mathbf{f}_t'\boldsymbol{\chi}_i\right)\right\},$$

$$\boldsymbol{\Lambda}_{44} = \frac{\partial Q_{NT}(\boldsymbol{\psi})}{\partial^2\vartheta} = \frac{1}{2(1-\vartheta)^2}.$$

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

$$\boldsymbol{\Lambda}_{NT}(\mathbf{0})=\begin{pmatrix}\frac{tr(\mathbf{G}_0^2+\mathbf{G}_0\mathbf{G}_0')}{N} & \mathbf{0}_{1\times k} & \mathbf{0}_{1\times Nm} & \frac{tr(\mathbf{G}_0)}{N}\\ \mathbf{0}_{k\times 1} & \mathbf{0}_{k\times k} & \mathbf{0}_{k\times Nm} & \mathbf{0}_{k\times 1}\\ \mathbf{0}_{Nm\times 1} & \mathbf{0}_{Nm\times k} & \mathbf{0}_{Nm\times Nm} & \mathbf{0}_{Nm\times 1}\\ \frac{tr(\mathbf{G}_0)}{N} & \mathbf{0}_{1\times k} & \mathbf{0}_{1\times Nm} & \frac{1}{2}\end{pmatrix}$$
$$+\sigma_0^2\begin{pmatrix}\mathbf{H}(\rho_0,\boldsymbol{\beta}_0') & \mathbf{0}_{(k+1+Nm)\times 1}\\ \mathbf{0}_{1\times(k+1+Nm)} & 0\end{pmatrix}+\sigma_0^2\begin{pmatrix}h_{f,11} & \mathbf{0}_{1\times k} & \mathbf{h}_{f,21}' & 0\\ \mathbf{0}_{k\times 1} & \mathbf{0}_{k\times k} & \mathbf{0}_{k\times Nm} & \mathbf{0}_{k\times 1}\\ \mathbf{h}_{f,21} & \mathbf{0}_{Nm\times k} & \mathbf{H}_{f,22} & \mathbf{0}_{Nm\times 1}\\ 0 & \mathbf{0}_{1\times k} & \mathbf{0}_{1\times Nm} & 0\end{pmatrix}, \tag{S.18}$$

where $\mathbf{H}_f\left(\rho_0,\boldsymbol{\gamma}_0'\right)$ is partitioned as

$$\mathbf{H}_f\left(\rho_0,\boldsymbol{\gamma}_0'\right)=\begin{pmatrix}h_{f,11} & \left(\mathbf{h}_{f,21}'\right)_{1\times Nm}\\ \left(\mathbf{h}_{f,21}\right)_{Nm\times 1} & \left(\mathbf{H}_{f,22}\right)_{Nm\times Nm}\end{pmatrix}.$$

Notice that all three terms on the right-hand side of (S.18) are positive semidefinite, which can be seen by applying Lemma A10 and by noting that both $\mathbf{H}(\rho_0,\boldsymbol{\beta}_0')$ and $\mathbf{H}_f(\rho_0,\boldsymbol{\gamma}_0')$ are positive semidefinite. Recall that the true parameter vector $\boldsymbol{\psi}_0$ is locally identified if and only if $\lambda_{\min}\left[\boldsymbol{\Lambda}_{NT}(\mathbf{0})\right]>0$. Hence, if $\mathbf{H}(\rho_0,\boldsymbol{\beta}_0')$ is positive definite, then both ρ_0 and $\boldsymbol{\beta}_0$ are identified. Given that ρ_0 is identifiable, σ_0 can be identified through the first term in (S.18). On the other hand, if the first term is positive definite, which is equivalent to $h_g>0$, then both ρ_0 and σ_0 are identified; and if in addition $(NT)^{-1}E_0(\mathbf{X}'\mathbf{X})$ is positive definite, the parameter vector $\boldsymbol{\beta}_0$ is identified. In both cases, $\boldsymbol{\gamma}_0$ is identified if $T^{-1}E_0\left(\mathbf{f}_t\mathbf{f}_t'\right)$ is positive definite. These findings are summarized in Proposition 2 in the main paper.

S2 Monte Carlo Supplement

This section provides additional simulation results of the estimation experiments. The Data Generating Process (DGP) follows the same design as in the main paper, namely,

$$\begin{aligned} y_{it}=&\rho y_{it}^*+\beta_1x_{it1}+\beta_2x_{it2}+\boldsymbol{\gamma}_{y,i}'f_t+e_{it}, \\ x_{itp}=&\boldsymbol{\gamma}_{x,ip}'f_t+\upsilon_{itp},\quad p=1,2,\end{aligned} \tag{S.1}$$

for $i=1,2,\ldots,N$, and $t=1,2,\ldots,T$. The unobserved factors are generated by

$$\begin{aligned} f_{lt}=&\rho_{fl}f_{l,t-1}+\varsigma_{f_{lt}},\ l=1,2,\ldots,m;\ t=-49,-48,\ldots,0,1,\ldots,T,\\ \varsigma_{f_{lt}}\sim& IIDN\left(0,1-\rho_{fl}^2\right),\quad \rho_{fl}=0.5,\quad f_{l,-50}=0,\end{aligned}$$

where the first 50 observations are dropped. The factor loadings are given by $\gamma_{y,i1}\sim IIDN\left(1,0.2\right)$, $\gamma_{y,i2}\sim IIDN\left(1,0.2\right)$, and

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

The idiosyncratic errors of the x_{itp} processes, $(\upsilon_{it1}, \upsilon_{it2})'$, are generated as

$$\begin{aligned}
\upsilon_{it,p} =& \rho_{\upsilon_{ip}} \upsilon_{it-1,p} + \vartheta_{it,p}, \quad t = -49, -48, \ldots, 0, 1, \ldots, T, \\
\vartheta_{it,p} \sim& N\left(0, 1 - \rho^2_{\vartheta_{ip}}\right), \quad \upsilon_{ip,-50} = 0, \\
\rho_{\vartheta_{ip}} \sim& IIDU\,(0.05, 0.95)\,, \qquad p = 1, 2,
\end{aligned}$$

For the idiosyncratic errors of the y_{it} process, in addition to the two generating processes we have considered in the main paper, we now assume that the errors e_{it} are independent over time and heteroskedastic across cross-section units. Specifically,

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

The spatial weights matrix is specified as the 1-ahead-and-1-behind circular neighbors weights matrix. The true number of factors is $m = 2$; the true values of slope coefficients are $\beta_1 = 1$ and $\beta_2 = 2$; for the spatial autoregressive coefficient, we consider $\rho = 0.4$ and 0.8, which represent low and high intensity of spatial dependence, respectively. The sample sizes are $N = 30$, 50, 100, 500, 1,000; and $T = 20$, 30, 50, 100. The number of replications is 2,000.

As in the main paper, we compare the performance of a number of estimators: the infeasible 2SLS estimator, 2SLS estimator, B2SLS estimator, GMM estimator and MLE. Tables S.1a and S.1b show the results under a low intensity of spatial dependence; Tables S.2a and S.2b present the results under a high intensity of spatial dependence. The results for β_2 are omitted since they are similar to those for β_1. Each table reports the estimates of bias, root mean squared error (RMSE), size, and power for the estimators. These results demonstrate that the proposed estimators have robust performance in the presence of heteroskedastic errors when N is relatively large compared to T. Moreover, the finite sample properties remain satisfactory even when there exists a high level of spatial dependence ($\rho = 0.8$).

Table S.1a: Small sample properties of estimators for the spatial parameter ρ ($\rho = 0.4$, independent and heteroskedastic errors)

	Bias(×100)				RMSE(×100)				Size(×100)				Power(×100)			
$N \backslash T$	20	30	50	100	20	30	50	100	20	30	50	100	20	30	50	100
	Infeasible 2SLS estimator (including factors)															
30	-0.10	-0.04	-0.01	0.00	2.40	1.91	1.48	0.99	5.55	5.15	5.65	4.50	13.50	18.25	29.20	51.75
50	0.05	0.06	0.00	0.00	1.85	1.48	1.11	0.77	5.45	5.60	5.40	5.10	20.60	30.10	43.60	72.90
100	0.02	0.01	0.01	0.01	1.30	1.03	0.78	0.55	4.80	3.95	4.40	4.10	34.90	49.10	70.95	95.50
500	-0.02	-0.01	0.00	0.00	0.59	0.46	0.35	0.25	5.15	4.95	4.30	4.70	92.55	98.85	100.00	100.00
1,000	0.00	0.00	-0.01	0.00	0.42	0.33	0.25	0.18	4.70	5.55	5.35	5.75	99.80	100.00	100.00	100.00
	2SLS estimator															
30	-0.09	-0.01	-0.01	0.01	2.74	2.15	1.64	1.15	5.90	6.05	7.10	7.90	12.70	18.25	28.35	48.40
50	0.03	0.06	0.02	0.01	1.99	1.57	1.19	0.82	5.15	5.90	5.45	5.05	18.05	27.00	42.35	71.00
100	0.01	0.01	0.02	0.01	1.38	1.08	0.81	0.57	4.85	5.55	4.55	5.30	30.45	45.25	69.20	94.30
500	-0.02	-0.01	0.00	0.00	0.62	0.48	0.36	0.25	4.70	4.50	4.95	4.40	88.80	98.15	100.00	100.00
1,000	0.00	0.00	-0.01	0.00	0.44	0.34	0.25	0.18	4.20	4.30	5.40	5.45	99.50	99.95	100.00	100.00
	B2SLS estimator															
30	-0.13	-0.03	-0.02	0.00	2.74	2.15	1.63	1.15	6.05	6.05	7.05	7.75	12.45	17.60	28.15	47.95
50	0.01	0.04	0.01	0.00	1.98	1.57	1.19	0.82	5.10	5.80	5.35	5.35	17.65	26.45	41.65	70.40
100	0.00	0.00	0.01	0.01	1.38	1.08	0.80	0.57	4.65	5.50	4.50	5.15	30.60	45.30	68.95	94.30
500	-0.02	-0.01	0.00	0.00	0.62	0.48	0.36	0.25	4.65	4.60	4.70	4.35	88.55	98.30	100.00	100.00
1,000	0.00	0.00	-0.01	0.00	0.44	0.34	0.25	0.18	4.35	4.40	5.05	5.60	99.55	99.95	100.00	100.00
	GMM estimator															
30	-1.27	-1.12	-1.08	-1.03	2.59	2.11	1.77	1.42	10.15	11.30	16.75	24.75	9.20	10.85	14.45	22.75
50	-0.69	-0.63	-0.64	-0.60	1.86	1.51	1.22	0.93	8.90	9.75	12.15	15.35	15.55	20.75	31.35	55.30
100	-0.32	-0.32	-0.30	-0.29	1.24	0.97	0.75	0.57	6.80	6.15	7.10	10.05	33.50	46.45	69.65	94.70
500	-0.08	-0.07	-0.07	-0.06	0.52	0.41	0.31	0.22	5.50	4.75	5.60	6.20	96.45	99.80	100.00	100.00
1,000	-0.03	-0.03	-0.04	-0.03	0.36	0.29	0.22	0.15	5.80	5.25	5.75	5.85	99.95	100.00	100.00	100.00
	MLE															
30	0.29	0.23	0.18	0.15	2.24	1.72	1.31	0.89	12.05	9.95	9.30	7.65	32.00	36.30	49.15	74.65
50	0.32	0.21	0.14	0.11	1.72	1.34	0.98	0.67	12.35	9.90	8.40	7.30	43.60	51.70	66.55	90.55
100	0.26	0.16	0.11	0.09	1.21	0.91	0.68	0.47	11.75	9.35	7.25	6.85	61.80	73.95	91.15	99.70
500	0.20	0.10	0.05	0.04	0.56	0.41	0.30	0.21	13.00	9.20	7.40	7.90	99.50	100.00	100.00	100.00
1,000	0.18	0.09	0.04	0.02	0.41	0.30	0.21	0.15	14.00	9.30	7.60	6.20	100.00	100.00	100.00	100.00

Notes: The DGP is given by (S.1), where e_{it} are generated by (S.2). The true parameter values are $\rho = 0.4$, $\beta_1 = 1$ and $\beta_2 = 2$. The true number of factors is 2. The naive estimator ignores latent factors, and the infeasible estimator treats factors as known. The naive 2SLS, infeasible 2SLS, and 2SLS estimators are computed using instruments $\mathbf{Q}_{.t}^{(2)} = \left(\mathbf{X}_{.t}, \mathbf{W}\mathbf{X}_{.t}, \mathbf{W}^2\mathbf{X}_{.t}\right)$, for $t = 1, 2, \ldots, T$. The best 2SLS (B2SLS) estimator is computed using $\hat{\mathbf{Q}}^*$ given by (57) in the main text. The efficient two-step GMM estimator utilizes $\mathbf{P}_1 = \mathbf{W}$ and $\mathbf{P}_2 = \mathbf{W}^2 - Diag\left(\mathbf{W}^2\right)$ in the quadratic moments and $\mathbf{Q}_{.t}^{(2)}$ in the linear moments. The MLE is computed by the Expectation-Maximization (EM) algorithm described in Bai and Li (2014), assuming the number of factors is known. The number of replications is $2,000$. The 95% confidence interval for size 5% is $[3.6\%, 6.4\%]$, and the power is computed under $H_1 : \rho = 0.38$.

Table S.1b: Small sample properties of estimators for the slope parameter β_1 ($\beta_1 = 1$, independent and heteroskedastic errors)

	Bias(×100)				RMSE(×100)				Size(×100)				Power(×100)			
$N \backslash T$	20	30	50	100	20	30	50	100	20	30	50	100	20	30	50	100
	Infeasible 2SLS estimator (including factors)															
30	0.01	-0.01	-0.02	-0.03	4.44	3.56	2.72	1.88	5.25	5.05	5.45	5.05	21.20	30.10	46.95	75.90
50	-0.20	-0.18	-0.12	-0.10	3.45	2.68	2.03	1.40	5.60	4.70	4.85	4.25	29.70	42.55	66.60	92.85
100	-0.14	-0.05	-0.05	-0.04	2.45	1.93	1.48	1.02	6.05	6.00	5.45	4.75	52.35	73.80	91.80	99.85
500	-0.05	-0.03	-0.02	-0.01	1.07	0.84	0.66	0.47	4.50	4.10	5.15	5.60	99.90	100.00	100.00	100.00
1,000	0.01	0.01	0.02	0.00	0.79	0.63	0.47	0.33	6.15	5.90	4.95	4.70	100.00	100.00	100.00	100.00
	2SLS estimator															
30	0.03	0.02	0.03	0.02	4.70	3.77	2.92	2.06	5.30	6.00	7.00	7.90	20.05	28.95	47.50	75.50
50	-0.19	-0.18	-0.09	-0.09	3.62	2.77	2.09	1.46	4.85	5.25	4.90	4.90	26.75	40.15	65.80	92.55
100	-0.13	-0.05	-0.05	-0.04	2.53	1.97	1.52	1.04	5.55	5.30	5.30	5.15	46.80	70.30	90.30	99.80
500	-0.07	-0.05	-0.02	-0.01	1.12	0.86	0.67	0.48	4.25	3.95	4.80	5.55	99.15	100.00	100.00	100.00
1,000	0.02	0.01	0.02	0.00	0.82	0.65	0.48	0.33	5.50	5.60	5.00	4.80	100.00	100.00	100.00	100.00
	B2SLS estimator															
30	0.04	0.02	0.03	0.02	4.70	3.77	2.92	2.06	5.30	5.95	7.00	7.95	20.05	28.90	47.45	75.65
50	-0.19	-0.18	-0.09	-0.09	3.62	2.77	2.09	1.46	4.85	5.20	4.95	4.85	26.80	40.15	65.80	92.40
100	-0.13	-0.05	-0.05	-0.04	2.53	1.97	1.52	1.04	5.55	5.35	5.40	5.05	46.75	70.25	90.30	99.80
500	-0.07	-0.05	-0.02	-0.01	1.12	0.86	0.67	0.48	4.20	3.95	4.85	5.45	99.15	100.00	100.00	100.00
1,000	0.02	0.01	0.02	0.00	0.82	0.65	0.48	0.33	5.50	5.60	5.00	4.80	100.00	100.00	100.00	100.00
	GMM estimator															
30	0.14	0.15	0.16	0.16	4.76	3.81	2.93	2.07	5.65	6.70	7.30	8.05	21.05	30.05	49.25	77.70
50	-0.09	-0.07	0.01	0.01	3.64	2.78	2.09	1.45	5.00	5.30	5.25	5.10	28.45	41.85	68.10	93.75
100	-0.08	0.00	0.01	0.02	2.53	1.96	1.51	1.03	5.80	5.10	5.30	5.10	47.65	71.30	90.95	99.80
500	-0.06	-0.04	-0.01	0.00	1.11	0.86	0.67	0.47	3.95	3.70	5.00	5.45	99.05	100.00	100.00	100.00
1,000	0.02	0.02	0.02	0.01	0.81	0.64	0.48	0.33	5.30	5.60	5.25	5.10	100.00	100.00	100.00	100.00
	MLE															
30	0.02	0.00	-0.06	-0.02	4.88	3.70	2.75	1.85	11.80	8.75	7.55	6.20	31.25	37.10	52.50	79.70
50	-0.17	-0.15	-0.12	-0.10	3.59	2.66	1.95	1.37	10.05	6.55	5.45	5.55	38.45	50.95	72.65	94.85
100	-0.14	-0.05	-0.07	-0.06	2.56	1.92	1.45	0.99	10.60	7.65	6.50	5.75	61.90	79.80	94.25	99.95
500	-0.03	-0.02	-0.02	-0.01	1.13	0.85	0.65	0.46	10.60	6.70	6.10	6.90	99.60	100.00	100.00	100.00
1,000	0.03	0.01	0.02	-0.00	0.79	0.61	0.45	0.31	9.70	7.20	6.40	5.80	100.00	100.00	100.00	100.00

Notes: The DGP is given by (S.1), where e_{it} are generated by (S.2). The true parameter values are $\rho = 0.4$, $\beta_1 = 1$ and $\beta_2 = 2$. The true number of factors is 2. The power is computed under $H_1 : \beta_1 = 0.95$. See also the notes to Table S.1a.

Table S.2a: Small sample properties of estimators for the spatial parameter ρ ($\rho = 0.8$, independent and heteroskedastic errors)

	Bias($\times$100)				RMSE($\times$100)				Size($\times$100)				Power($\times$100)			
$N \backslash T$	20	30	50	100	20	30	50	100	20	30	50	100	20	30	50	100
	Infeasible 2SLS estimator (including factors)															
30	-0.04	-0.01	0.00	0.01	1.21	0.95	0.74	0.49	5.25	5.15	5.50	4.85	39.00	55.10	77.70	97.25
50	0.04	0.04	0.01	0.00	0.93	0.75	0.56	0.38	5.45	5.65	5.25	5.45	60.80	78.35	93.70	99.90
100	0.01	0.01	0.01	0.00	0.65	0.51	0.39	0.27	5.35	4.45	4.40	4.40	85.85	97.15	99.90	100.00
500	-0.01	-0.01	0.00	0.00	0.29	0.23	0.18	0.12	5.20	5.30	4.40	4.30	100.00	100.00	100.00	100.00
1,000	0.00	0.00	0.00	0.00	0.21	0.17	0.13	0.09	5.00	5.45	5.45	6.15	100.00	100.00	100.00	100.00
	2SLS estimator															
30	-0.04	0.00	0.01	0.01	1.45	1.14	0.87	0.61	6.20	6.45	6.90	8.10	32.90	47.85	68.20	92.00
50	0.03	0.04	0.01	0.01	1.03	0.81	0.62	0.42	5.15	5.85	5.50	6.20	51.90	72.05	89.70	99.65
100	0.01	0.00	0.01	0.00	0.71	0.55	0.41	0.29	5.00	4.95	4.40	5.40	80.50	94.70	99.90	100.00
500	-0.01	-0.01	0.00	0.00	0.31	0.24	0.18	0.13	4.70	4.30	4.90	4.40	99.90	100.00	100.00	100.00
1,000	0.00	0.00	0.00	0.00	0.22	0.17	0.13	0.09	4.70	4.45	5.30	6.05	100.00	100.00	100.00	100.00
	B2SLS estimator															
30	-0.10	-0.03	-0.01	0.00	1.42	1.11	0.84	0.59	5.40	6.30	6.35	7.60	33.20	49.15	69.55	93.85
50	0.01	0.02	0.00	0.00	1.00	0.79	0.60	0.41	5.45	5.50	5.95	6.20	53.80	73.00	92.10	99.65
100	-0.01	-0.01	0.00	0.00	0.68	0.53	0.39	0.27	4.90	5.60	4.65	5.05	82.35	94.90	99.90	100.00
500	-0.01	0.00	0.00	0.00	0.30	0.23	0.18	0.12	5.10	5.10	5.15	4.60	100.00	100.00	100.00	100.00
1,000	0.00	0.00	0.00	0.00	0.21	0.16	0.12	0.09	4.45	4.85	4.95	5.25	100.00	100.00	100.00	100.00
	GMM estimator															
30	-0.77	-0.68	-0.65	-0.62	1.45	1.18	1.00	0.82	11.95	13.20	19.65	30.05	24.15	35.05	53.45	80.50
50	-0.39	-0.37	-0.37	-0.36	0.98	0.81	0.66	0.52	9.15	10.20	13.30	18.95	49.50	68.90	88.00	99.25
100	-0.19	-0.19	-0.18	-0.17	0.64	0.51	0.39	0.30	7.65	7.05	7.80	11.55	84.95	97.20	99.80	100.00
500	-0.04	-0.04	-0.04	-0.04	0.26	0.21	0.16	0.11	5.25	5.35	5.95	6.20	100.00	100.00	100.00	100.00
1,000	-0.02	-0.02	-0.02	-0.02	0.18	0.15	0.11	0.08	5.75	5.45	6.25	6.10	100.00	100.00	100.00	100.00
	MLE															
30	0.27	0.19	0.16	0.13	1.21	0.90	0.68	0.46	17.05	13.20	11.70	10.50	69.75	81.25	94.25	99.85
50	0.25	0.16	0.11	0.09	0.89	0.67	0.49	0.34	15.30	11.75	9.70	8.75	86.70	94.75	99.50	100.00
100	0.19	0.11	0.08	0.07	0.62	0.46	0.34	0.24	14.05	10.25	8.75	8.00	98.45	99.80	100.00	100.00
500	0.14	0.07	0.04	0.03	0.30	0.21	0.15	0.10	17.80	10.30	8.10	8.90	100.00	100.00	100.00	100.00
1,000	0.12	0.06	0.03	0.01	0.23	0.16	0.11	0.07	21.60	12.10	8.40	7.40	100.00	100.00	100.00	100.00

Notes: The DGP is given by (S.1), where e_{it} are given by (S.2). The true parameter values are $\rho = 0.8$, $\beta_1 = 1$ and $\beta_2 = 2$. The true number of factors is 2. The power is computed under $H_1 : \rho = 0.78$. See also the notes to Table S.1a.

Table S.2b: Small sample properties of estimators for the slope parameter β_1 ($\beta_1 = 1$, $\rho = 0.8$, independent and heteroskedastic errors)

	Bias(×100)				RMSE(×100)				Size(×100)				Power(×100)			
$N \backslash T$	20	30	50	100	20	30	50	100	20	30	50	100	20	30	50	100
	Infeasible 2SLS estimator (including factors)															
30	0.02	-0.01	-0.03	-0.03	4.51	3.61	2.76	1.92	5.20	5.00	5.10	5.30	20.50	29.45	45.50	74.35
50	-0.22	-0.20	-0.12	-0.10	3.50	2.72	2.06	1.43	5.65	4.85	4.95	4.55	28.00	40.90	64.60	91.50
100	-0.15	-0.05	-0.05	-0.04	2.49	1.96	1.51	1.04	5.70	6.00	5.55	4.75	50.90	72.15	90.95	99.80
500	-0.05	-0.03	-0.02	-0.01	1.09	0.85	0.67	0.48	4.55	4.25	5.20	5.65	99.75	100.00	100.00	100.00
1,000	0.01	0.01	0.02	0.00	0.80	0.64	0.48	0.33	5.90	5.45	5.00	5.00	100.00	100.00	100.00	100.00
	2SLS estimator															
30	0.04	0.01	0.02	0.01	4.77	3.82	2.96	2.08	5.50	6.10	7.20	7.75	19.45	28.40	46.10	74.50
50	-0.22	-0.20	-0.10	-0.09	3.67	2.82	2.12	1.48	4.65	5.40	5.05	4.75	25.40	38.45	64.25	91.70
100	-0.14	-0.05	-0.06	-0.04	2.58	2.01	1.55	1.06	5.35	5.00	5.50	4.85	45.50	68.45	89.45	99.75
500	-0.06	-0.04	-0.02	-0.01	1.14	0.88	0.68	0.48	4.30	3.95	4.90	5.40	98.90	100.00	100.00	100.00
1,000	0.02	0.01	0.02	0.00	0.83	0.66	0.49	0.34	5.45	5.55	5.15	4.85	99.95	100.00	100.00	100.00
	B2SLS estimator															
30	0.08	0.03	0.03	0.02	4.76	3.82	2.96	2.08	5.50	6.20	7.10	7.80	19.90	28.75	46.20	74.50
50	-0.19	-0.18	-0.09	-0.09	3.66	2.81	2.12	1.47	4.65	5.25	4.95	4.70	25.75	38.50	64.80	92.15
100	-0.13	-0.04	-0.05	-0.04	2.57	2.00	1.55	1.06	5.30	4.95	5.30	5.20	45.85	68.90	89.75	99.70
500	-0.06	-0.05	-0.02	-0.01	1.14	0.88	0.68	0.48	4.25	3.85	4.95	5.40	98.85	100.00	100.00	100.00
1,000	0.02	0.01	0.02	0.00	0.83	0.66	0.49	0.34	5.40	5.50	5.10	4.70	99.95	100.00	100.00	100.00
	GMM estimator															
30	0.42	0.41	0.40	0.40	4.82	3.86	2.99	2.12	5.70	6.90	7.35	8.35	22.40	31.95	51.30	80.25
50	0.06	0.07	0.16	0.15	3.69	2.81	2.11	1.48	5.10	5.20	5.35	5.10	28.90	42.55	69.25	94.20
100	-0.01	0.08	0.08	0.09	2.56	1.99	1.53	1.05	5.50	5.45	5.35	5.15	47.55	71.75	91.25	99.80
500	-0.04	-0.02	0.01	0.02	1.13	0.87	0.67	0.48	4.05	3.50	4.90	5.45	98.90	100.00	100.00	100.00
1,000	0.03	0.03	0.03	0.02	0.82	0.65	0.49	0.34	5.15	5.60	5.40	5.15	100.00	100.00	100.00	100.00
	MLE															
30	-0.12	-0.10	-0.14	-0.09	4.93	3.73	2.79	1.86	11.90	8.55	7.40	6.70	29.50	36.05	50.65	77.50
50	-0.30	-0.23	-0.18	-0.15	3.62	2.68	1.97	1.38	10.00	6.60	5.20	5.35	35.90	49.35	70.50	94.25
100	-0.23	-0.11	-0.12	-0.10	2.60	1.95	1.47	1.00	10.75	8.10	6.80	6.00	60.20	78.15	93.40	99.90
500	-0.10	-0.05	-0.03	-0.03	1.15	0.86	0.66	0.46	10.30	6.40	6.40	7.10	99.60	100.00	100.00	100.00
1,000	-0.03	-0.02	0.01	-0.01	0.80	0.62	0.46	0.32	10.60	7.70	6.30	5.60	100.00	100.00	100.00	100.00

Notes: The DGP is given by (S.1), where e_{it} are given by (S.2). The true parameter values are $\rho = 0.8$, $\beta_1 = 1$ and $\beta_2 = 2$. The true number of factors is 2. The power is computed under $H_1 : \beta = 0.95$. See also the notes to Table S.1a.

S3 Empirical Application Supplement

S3.1 Comparison of Alternative De-Factoring and Estimation Methods

The estimators developed in the main text take advantage of cross-sectional averages to approach the unknown factors. It is of interest to compare the results with those obtained by principal components analysis. Bailey, Holly, and Pesaran (2016) lay out a two-stage estimation procedure that consists of extracting common factors in the first stage and then applying a spatial model to the first-stage residuals. The de-factoring step can be performed using either cross-sectional averages or principal components (PC), and the estimation step can be implemented by the MLE or GMM/2SLS routines described in the spatial literature.[S1] From this perspective, the GMM estimator developed in this paper can also be achieved by a two-step estimation, but its inference under a two-step procedure would be invalid without some adjustments. No asymptotic theories are available yet for the two-step estimators involving PC and MLE, and it is unclear how the first-stage estimation error affects the second-stage results. Hence, in what follows, we will provide some useful accounts of the relative performance of these different methods.

As before, we consider unknown factors at both national and regional levels, and all MSAs are classified into eight BEA Regions. Both dependent and independent variables are purged of common effects, using the same procedure under consideration. If cross-sectional averages are adopted as factor surrogates, they include national and regional averages of both dependent and individual-specific regressors. If principal components analysis is performed, the strongest PC is extracted from the full sample, then from the residuals the strongest PC is extracted for each of the eight Regions separately.

Table S3.1 summarizes the estimation results of model (58) in the main text, using different combinations of de-factoring and estimation schemes. The results suggest that both de-factoring approaches have similar effectiveness in filtering out the strong dependence. All procedures yield very close estimates of the spatial autoregressive coefficient as well as the coefficients for population and income growth. All estimates are significant and of reasonable magnitude. The MLE has a smaller standard error, since it neglects the sampling variation in the first-stage, and it also disregards serial correlation and heteroskedasticity in the disturbances.

S3.2 Further Characterization of the Spatial Weights Matrices

Figure S.1 shows the histogram of distance between area of origin and area of destination based on the migration weights matrix, $\mathbf{W}_m$. It is readily seen that significant migration flows occur between places farther than 100 miles apart, and hence the migration-based weights captures a very different connections among MSAs from distance-based measures.

Figure S.2 presents the intensity plots for different spatial weights matrices that are considered

[S1] Bailey, Holly, and Pesaran (2016) analyze the residuals obtained by de-factoring with cross-sectional averages, using a quasi-MLE that allows for heterogeneous spatial coefficients; they do not consider other estimation procedures. By contrast, here the MLE refers to the standard estimation method for spatial models with homogeneous coefficients. (See, for example, Anselin, 1988, Chapter 6.)

Table S3.1: Comparison of estimation results of model (58) by alternative de-factoring and estimation methods

De-factoring method	CS		PC	
	GMM (1)	MLE (2)	GMM (3)	MLE (4)
ρ [$\mathbf{W} \times \%\Delta$House price]	0.643	0.612	0.606	0.568
	(0.005)	(0.003)	(0.006)	(0.004)
β_1 [$\%\Delta$Population]	0.366	0.261	0.304	0.324
	(0.040)	(0.009)	(0.025)	(0.011)
β_2 [$\%\Delta$Income per capita]	0.093	0.082	0.075	0.077
	(0.007)	(0.004)	(0.007)	(0.004)
Natl. & Rgnl. unobserved factors	Yes	Yes	Yes	Yes
MSA FE and seasonal dummies	Yes	Yes	Yes	Yes
Residuals				
CD test statistic	-6.532	-6.291	6.229	8.680
Exponent of cross-section	0.674	0.694	0.739	0.738
dependence	(0.019)	(0.019)	(0.019)	(0.019)
$\bar{R}^2$	0.837	0.835	0.815	0.812
Observations	$N = 377,\ T = 159$			

Notes: All estimations consider both national and regional (Natl. & Rgnl.) unobserved factors, and also include MSA fixed effects (FE) and quarterly dummies. To save space, factor estimates are not reported. CS stands for the de-factoring procedure using cross-sectional averages at both national and regional levels. PC refers to the de-factoring procedure that extracts one principal component from the full sample and one from each Region. The spatial weights matrix is $\mathbf{W} = \mathbf{W}_{100}$. Standard errors are in parentheses.

in the paper. This figure complements Table B.1 in Appendix B and provides a direct visualization of these weights matrices. In each plot, all MSAs are sorted first by Region and then by State.[S2] As expected, all weights matrices are sparse. Compared with distance-based matrices, the migration- and correlation-based matrices contain more non-zero elements farther away from the diagonal.

[S2]The identities of the MSAs corresponding to the plots are available upon request.

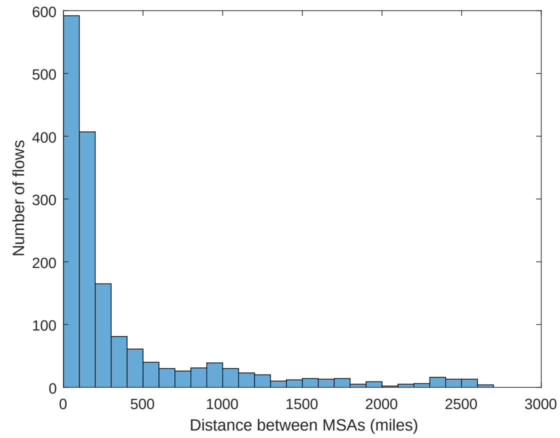
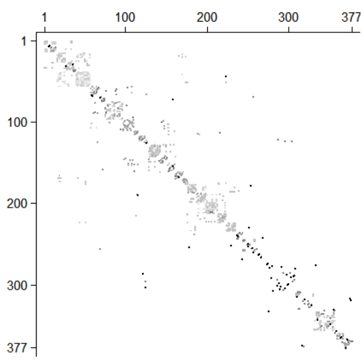
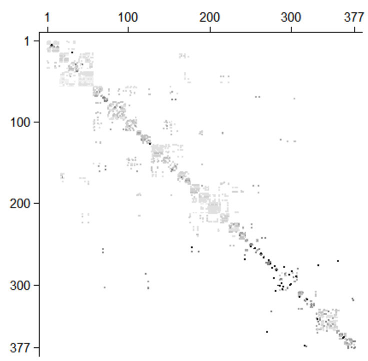


Figure S.1: Histogram of inter-MSA migration distance, based on the migration flow matrix $\mathbf{W}_m$

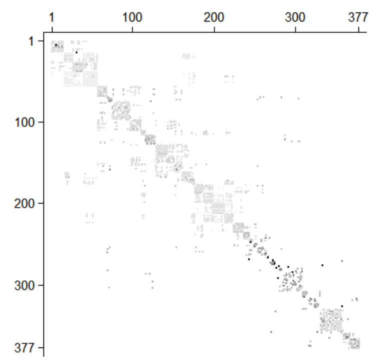
Notes: This figure shows the distribution of distance between two MSAs that have a migration flow between them, as indicated by a nonzero entry in the migration weights matrix $\mathbf{W}_m$. The inflows and outflows between two MSAs are considered as two flows.



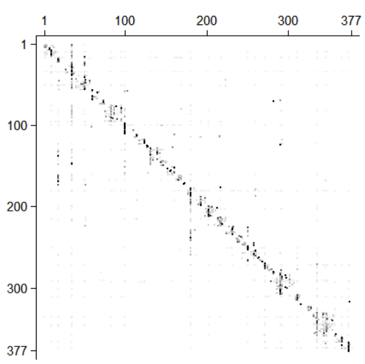
(a) $\mathbf{W}_{75}$



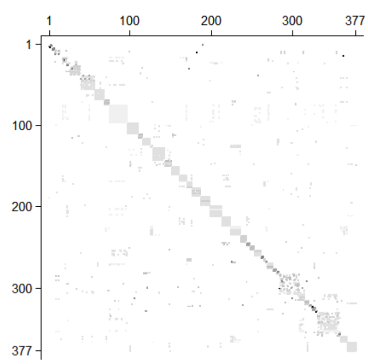
(b) $\mathbf{W}_{100}$



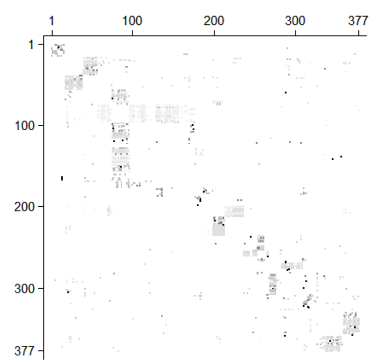
(c) $\mathbf{W}_{125}$



(d) $\mathbf{W}_m$



(e) $\hat{\mathbf{W}}^{+}$



(f) $\hat{\mathbf{W}}^{-}$

Figure S.2: Intensity plots of the spatial weights matrices

Notes: The 377 MSAs are sorted first by Region and then by State. The Regions and States are ordered from the East Coast to the West Coast. Zero elements of the weights matrix are plotted in white. Higher values are represented by darker colors. $\mathbf{W}_d$ denotes radial distance weights matrix with threshold distance d (miles). $\mathbf{W}_m$ represents weights matrix based on MSA-to-MSA migration flows. $\hat{\mathbf{W}}^{+}$ ($\hat{\mathbf{W}}^{-}$) is constructed from significantly positive (negative) pairwise correlations of de-factored house price changes.

References

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*, Volume 4. Springer Science & Business Media.

Bai, J. and K. Li (2014). Spatial panel data models with common shocks. MPRA Paper 52786.

Bailey, N., S. Holly, and M. H. Pesaran (2016). A two-stage approach to spatio-temporal analysis with strong and weak cross-sectional dependence. *Journal of Applied Econometrics 31* (1), 249–280.

Kapetanios, G. (2010). A testing procedure for determining the number of factors in approximate factor models with large datasets. *Journal of Business & Economic Statistics 28*, 397–409.

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