Spatial Dynamic Panel Data Models with Interactive Fixed Effects

Wei Shi
Department of Economics, The Ohio State University, Columbus, Ohio 43210.

Lung-Fei Lee
Department of Economics, The Ohio State University, Columbus, Ohio 43210.

Abstract

This paper studies the estimation of a dynamic spatial panel data model with interactive individual and time effects with large $n$ and $T$. The model has a rich spatial structure including contemporaneous spatial interaction and spatial heterogeneity. Dynamic features include individual time lag and spatial diffusion. The interactive effects capture heterogeneous impacts of time effects on cross sectional units. The interactive effects are treated as parameters, so as to allow correlations between the interactive effects and the regressors. We consider a quasi-maximum likelihood estimation and show estimator consistency and characterize its asymptotic distribution. The Monte Carlo experiment shows that the estimator performs well and the proposed bias correction is effective. We illustrate the empirical relevance of the model by applying it to examine the effects of house price dynamics on reverse mortgage origination rates in the US.

Keywords: Spatial panel, dynamics, multiplicative individual and time effects

JEL classification: C13, C23, C51

This version: 11/26/2015. We would like to thank the participants of the 2014 China Meeting of the Econometric Society at Xiamen University, the 2014 Shanghai Econometrics Workshop at Shanghai University of Finance and Economics, the 2015 MEA Annual Meeting, New York Camp Econometrics X, the 11th World Congress of the Econometric Society, and the Econometrics seminars at the Ohio State University for many valuable comments. We appreciate receiving valuable comments and suggestions from referees, an associate editor and a coeditor of this journal.

Email addresses: shi.271@osu.edu (Wei Shi), lee.1777@osu.edu (Lung-Fei Lee)
1. Introduction

Spatial interaction is present in many economic problems. When a state determines its tax rate, it takes into account not only its domestic constituent, but also what neighboring states might do (e.g., Han (2013)). Fluctuations in one industrial sector can spill to other sectors if the sectors are “close” in terms of using similar production technology, using similar inputs, etc (e.g., Conley and Dupor (2003)). Early contributions to spatial econometrics include Cliff and Ord (1973) and Anselin (1988). Kelejian and Prucha (2010) examine the GMM for estimation of spatial models. Lee (2004a) establishes asymptotic properties of the quasi-maximum likelihood (QML) estimator of a spatial autoregressive (SAR) model. A spatial panel can take into account dynamics and control for unobserved heterogeneity. Dynamic spatial panel data models with fixed individual and/or time effects where spatial effects appear as lags in time and in space have been studied in Yu et al. (2008) and Lee and Yu (2010b). Su and Yang (2014) examine QML estimation of dynamic panel data models with spatial effects in the errors. Lee and Yu (2013b) is a recent survey on spatial panel data models.

On the other hand, individual units can be differently affected by common factors. In the example of industrial sectors above, common factors like interest rate, demographic trend, etc., can simultaneously affect various industrial sectors, but magnitudes of the impact can be different across sectors. A few possibly unobserved common factors may drive much essential comovement between sectors although those sectors are far from each other according to economic distance measures. Essentially, a factor induces a time fixed effect which may affect individuals differently. Bai (2009) labels this an interactive effect.

In recent years, much progress has been made in the estimation and inference of panel data models with interactive effects. When interactive effects are viewed as fixed parameters, the model can be estimated by the nonlinear least squares (NLS) method involving principal components. Bai (2009) systematically studies asymptotic properties of the NLS estimator. The estimation is iterative where in each iteration, slope parameters are estimated given factors, and then factors are estimated by principal components given the estimated slope parameters. Moon and Weidner (2015) show that, under additional assumptions, the limiting distribution of the NLS estimator does not depend on the number of factors assumed in the estimation, as long as it does not fall below the true number of factors. They analyze asymptotic properties of the NLS estimator by perturbation method in Kato (1995), where the objective function is expanded involving its

\[ \text{In the literature, “common shocks”, “common factors” or “factors” refer to the time varying factors. “Factor loading” quantifies the magnitude of the effect of the time varying factors on an individual. There can be multiple factors. In this paper, they are treated as fixed parameters to be estimated. Because the main purpose of this paper is to consistently estimate slope coefficients, it is not necessary to separately identify time factors from their loadings when they are concentrated out in the estimation.} \]
approximated gradient vector and Hessian matrix. Assuming that factors and factor loadings are random and factors also enter regressors linearly, Pesaran (2006) proposes the common correlated effects (CCE) estimator. His idea is that variations due to factors can be captured by cross sectional averages of the dependent and explanatory variables. Ahn et al. (2013) propose a GMM estimator for fixed $T$ case. The interactive effects can be eliminated by a transformation involving the factors and then GMM can be applied.

Spatial interaction and common factors are two specifications explored in the literature where individuals’ activities and outcomes are not independently distributed. In this paper, an individual can be influenced by its neighbors’ actions or outcomes which is modeled by the SAR specification. Individuals are also exposed to unobserved common factors which are modeled as interactive effects following Bai (2009). By treating interactive effects as parameters to be estimated, this approach allows flexible correlation between the interactive effects and the regressors. As a data generating process involves spatial spillovers and interactive effects, both should be taken into account for estimators to be consistent.

Existing literature on factor models (Bai (2003), Forni et al. (2004), Stock and Watson (2002)) ignores spatial interactions. Literature on spatial interactions (Lee and Yu (2010a), Lee and Yu (2013b)) has not considered unobserved interactive effects. To the best of our knowledge, there are only a few papers that jointly model spatial correlation and interactive effects. Pesaran and Tosetti (2011) model different forms of error correlation, including spatial error correlation and common factor, and show that the CCE estimator continues to work well. Bai and Li (2014) consider a model with spatial correlation in the dependent variable and common factors.

This paper jointly models spatial interactions and interactive effects in a panel data with large $n$ and $T$. We consider spatial interaction in the dependent variable where the degree of spatial correlation is of interest, in which case the CCE estimator is not directly applicable. The spatial panel model under consideration is a general dynamic spatial panel data model where spatial effects can appear both in the form of lags and errors. In addition to contemporaneous spatial interaction, time lagged dependent variables, diffusion and spatially correlated and heterogeneous disturbances are included to allow a rich specification of the state dependence and guard against spurious spatial correlation. We do not impose specific structures on how interactive effects affect the regressors. The interactive effects are treated as nuisance parameters and are concentrated out in the estimation. Moon and Weidner (2015) show how to derive the approximated gradient vector and the Hessian matrix of the concentrated NLS objective function of a regression panel. In spatial panel setting, the log of the sum of squared residuals from the regression panel is a component of the likelihood function, and we adapt their approach to that component. We provide conditions for identification and show that the QML estimation method works well. The estimator is shown to be consistent and asymptotically normal.
Asymptotic biases of order $\frac{1}{\sqrt{nt}}$ exist due to incidental parameters, and a bias correction method is proposed.

The paper is organized as follows. Section 2 presents the model and discusses assumptions and identification. We prove consistency and derive the limiting distribution of the QML estimator in Section 3. We then illustrate the empirical relevance of the theory by demonstrating the estimator’s good finite sample performance and applying the model to analyze the effect of house price dynamics on reverse mortgage origination rates in the U.S. Section 6 concludes. Proofs of the main results are collected in the appendix. A supplementary file is available online which has detailed proofs on relevant useful lemmas.

In this paper, here are some essential notations. For a vector $\eta$, $||\eta||_1 = \sum_k |\eta_k|$ and $||\eta||_2 = \sqrt{\sum_k |\eta_k|^2}$. Let $\mu_i(M)$ denote the $i$-th largest eigenvalue of a symmetric matrix $M$ of dimension $n$ with eigenvalues listed in a decreasing order such that $\mu_n(M) \leq \mu_{n-1}(M) \leq \cdots \leq \mu_1(M)$. For a real matrix $A$, its spectral norm is $||A||_2$, i.e., $||A||_2 = \sqrt{\mu_1(A'A)}$. In addition, $||A||_1 = \max_{1 \leq j \leq n} \sum_{i=1}^m |A_{ij}|$ is its maximum column sum norm, $||A||_\infty = \max_{1 \leq i \leq m} \sum_{j=1}^n |A_{ij}|$ is its maximum row sum norm, and $||A||_F = \sqrt{\text{tr}(AA')}$ is its Frobenius norm. Denote the projection matrices $P_A = A(A'A)^{-1}A'$ and $M_A = I - P_A$. In cases where $A$ might not have full rank, we use $(A'A)^\dagger$ to denote the Moore-Penrose generalized inverse of $A'A$. For a real number $x$, $[x]$ is the smallest integer greater than or equal to $x$. “wpa 1” stands for “with probability approaching 1”.

2. The Spatial Dynamic Panel Data (SDPD) Model with Interactive Effects

2.1. The Model

There are $n$ individual units and $T$ time periods. The SDPD model has the following specification,

$$Y_{nt} = \lambda W_n Y_{nt} + \gamma Y_{n,t-1} + \rho W_n Y_{n,t-1} + X_{nt} \beta + \Gamma_n f_t + U_{nt}, \quad \text{and} \quad U_{nt} = \alpha \tilde{W}_n U_{nt} + \epsilon_{nt},$$

where $Y_{nt}$ is an $n$-dimensional column vector of observed dependent variables and $X_{nt}$ is an $n \times (K-2)$ matrix of exogenous regressors, so that the total number of variables in $Y_{n,t-1}$, $W_n Y_{n,t-1}$ and $X_{nt}$ is $K$. The model accommodates two types of cross sectional dependences, namely, local dependence and global (strong) dependence. Individual units are impacted by potentially time varying unknown common factors $f_t$, which captures global (strong) dependence. The effects of the factors can be heterogeneous on the cross section units, as described by the factor loading parameter matrix $\Gamma_n$. For example, in an earnings regression where $Y_{nt}$ is the wage rate, each row of $\Gamma_n$ may correspond to a vector of an individual’s skills and $f_t$ is the skill premium which may be time varying. The number of unobserved factors is assumed to be a fixed constant $r$ that is much smaller than $n$ and $T$.\footnote{In many empirical studies, the number of factors is much smaller than the dimension of the dataset. For example, in Stock and Watson (2002), 6 factors are used to model 215 macroeconomic time series.} The matrix of $n \times r$ factor loading $\Gamma_n$ and the $T \times r$ factors
\( F_T = (f_1, f_2, \cdots, f_T)' \) are not observed and are treated as parameters. The fixed effects approach is flexible and allows unknown correlation between the common factor components and the regressors. The \( n \times n \) spatial weights matrices \( W_n \) and \( \tilde{W}_n \) in Eqs. (1) are used to model spatial dependences. The term \( \lambda W_n Y_{nt} \) describes the contemporaneous spatial interactions. There are also dynamics in model (1). \( \gamma Y_{nt-1} \) captures the pure dynamic effect. \( \rho W_n Y_{nt-1} \) is a spatial time lag of interactions, which captures diffusion (Lee and Yu (2013b)).\(^3\) The idiosyncratic error \( \epsilon_{nt} \) with elements of \( \epsilon_{it} \) being i.i.d. \( (0, \sigma^2) \) also possesses a spatial structure \( \tilde{W}_n \), which may or may not be the same as \( W_n \).

The specification in (1) is general which encompasses many models of empirical interest.

- Additive fixed individual and time effects:

\[
Y_{nt} = \lambda W_n Y_{nt} + \gamma Y_{nt-1} + \rho W_n Y_{nt-1} + X_{nt} \beta + \tilde{\zeta}_n + \xi_{nt} + \epsilon_{nt},
\]

where \( \tilde{\zeta}_n = \left( \zeta_1, \zeta_2, \cdots, \zeta_n \right)' \) are individual effects, and \( \xi_{nt} \) are time effects with \( \xi_{nt} = (1, 1, \cdots, 1)' \).

Eq. (2) is a special case of Eq. (1) with \( \Gamma_n = \left( \begin{array}{cccc} \zeta_1 & \zeta_2 & \cdots & \zeta_n \\ 1 & 1 & \cdots & 1 \end{array} \right)' \) and \( F_T = \left( \begin{array}{cccc} \zeta_1 & \zeta_2 & \cdots & \zeta_T \end{array} \right)' \).

- Spatial panel data model with common shocks by Bai and Li (2014):

\[
Y_{nt} = \lambda W_n Y_{nt} + X_{nt} \beta + \tilde{\zeta}_{yn} + \Gamma_n f_t + \epsilon_{nt},
\]

\[
X_{nt,k} = \tilde{\zeta}_{nk,n} + \Gamma_{nk,n} f_t + \nu_{nt,k}, \quad k = 1, \cdots, K,
\]

where \( X_{nt,k} \) is the \( k \)-th column of \( X_{nt} \), \( \tilde{\zeta}_{yn} = \left( \zeta_{y1}, \zeta_{y2}, \cdots, \zeta_{yn} \right)' \) and \( \tilde{\zeta}_{nk,n} = \left( \zeta_{nk1}, \zeta_{nk2}, \cdots, \zeta_{nk,n} \right)' \) are fixed effects, \( f_t \) are \( r \times 1 \) common factors with loadings \( \Gamma_{yn} \) and \( \Gamma_{nk,n} \). While heteroscedasticity in \( \epsilon_{nt} \) cross \( i \) but invariant over \( t \) is allowed, their model is static and the common shocks are limited to impact \( X_{nt} \) linearly. Pesaran (2006) considers the case with \( \lambda = 0 \) and heterogeneous coefficients.

Define \( A_n = S_n^{-1}(\gamma I_n + \rho W_n) \), where \( S_n = I_n - \lambda W_n \). From Eq. (1), \( Y_{nt} = A_n Y_{nt-1} + S_n^{-1}(X_{nt} \beta + \Gamma_n f_t + U_{nt}) \).

Continuous substitution gives \( Y_{nt} = \sum_{h=0}^{\infty} A_n^h S_n^{-1} (X_{nt-h} \beta + \Gamma_n f_{t-h} + U_{nt-h}) \), assuming that the series converges. With \( \|A_n\|_2 < 1 \), \( \{A_n^h\}_{h=0}^{\infty} \) is absolutely summable and the initial condition \( Y_{n,0} \) does not affect the asymptotic analysis when \( T \to \infty \). Lee and Yu (2013b) discuss the parameter space of \( \gamma, \rho \) and \( \lambda \) and regularity conditions that guarantee \( \|A_n\|_2 < 1 \). Let \( \sigma_{ni} \) denote an eigenvalue of \( W_n \) and \( d_{ni} \) the corresponding eigenvalue of \( A_n \), we then have \( d_{ni} = \gamma^{h} \rho \sigma_{ni} \overline{1 - \lambda E_n} \). If the spatial weights matrix \( W_n \) is row normalized from a

\(^3\)In general, the spatial weights matrices for the contemporaneous spatial interactions and for the diffusion can be different. However it is straightforward to extend this paper’s analysis to such cases. QMLE is still consistent and asymptotically normal under assumptions that will be introduced in Section 2.3. Assuming identical spatial weights simplifies the notation.
symmetric matrix (as in Ord (1975)), all eigenvalues of $W_{n}$ are real. Furthermore, if $\text{tr}(W_{n}) = 0$, the condition $\frac{1}{\bar{\sigma}_{n,\min}} < \lambda < 1 = \sigma_{n,\max}$ implies that $I_{n} - \lambda W_{n}$ is invertible, where $\bar{\sigma}_{n,\min}$ and $\bar{\sigma}_{n,\max}$ are, respectively, the smallest and largest eigenvalues of $W_{n}$. Stationarity further requires that $|d_{ni}| < 1$ for all $i$. Lee and Yu (2013b) show that with a row normalized $W_{n}$, the parameter space for $\|A_{n}\|_{2} < 1$ can be characterized as a region enclosed by four linear hyperplanes

$$R_{s} = \{ (\lambda, \gamma, \rho) : \gamma + (\rho - \lambda) \bar{\sigma}_{n,\min} > -1, \gamma + \rho + \lambda < 1, \gamma + \rho - \lambda > -1, \gamma + (\lambda + \rho) \bar{\sigma}_{n,\min} < 1 \}.$$ 

Elements in $R_{s}$ already imply that $-1 < \gamma < 1$ and $\frac{1}{\bar{\sigma}_{n,\min}} < \lambda < 1$. Because $|\bar{\sigma}_{n,i}| < 1$, a sufficient condition for $\|A_{n}\|_{2} < 1$ is $|\lambda| + |\gamma| + |\rho| < 1$.

### 2.2. Estimation Method

The parameters for the model (1) are $\theta = (\delta', \lambda, \alpha)'$ with $\delta = (\gamma, \rho, \beta')'$, $\sigma^{2}$, $\Gamma_{n}$ and $F_{T}$. It is convenient to collect the predetermined variables and exogenous regressors by the $n \times K$ matrix $Z_{nt} = (Y_{n,t-1}, W_{n}Y_{n,t-1}, X_{nt})$, where $K = k + 2$, with $k$ being the number of exogenous regressors in $X_{nt}$. Denote $S_{n}(\lambda) = I_{n} - \lambda W_{n}$ and $R_{n}(\alpha) = I_{n} - \alpha \bar{W}_{n}$. The sample averaged quasi-log likelihood function is

$$Q_{nT}(\theta, \sigma^{2}, \Gamma_{n}, F_{T}) = -\frac{1}{2} \log 2\pi - \frac{1}{2} \log \sigma^{2} + \frac{1}{n} \log |S_{n}(\lambda)R_{n}(\alpha)|$$

$$- \frac{1}{2\sigma^{2}nT} \sum_{t=1}^{T} (S_{n}(\lambda)Y_{nt} - Z_{nt}\delta - \Gamma_{n,fi})^{\prime}R_{n}(\alpha)^{\prime}R_{n}(\alpha)(S_{n}(\lambda)Y_{nt} - Z_{nt}\delta - \Gamma_{n,fi}).$$ \quad (3)

Concentrating out $\sigma^{2}$ from the objective function (3) and dropping the overall constant term for simplicity,

$$Q_{nT}(\theta, \Gamma_{n}, F_{T}) = \frac{1}{n} \log |S_{n}(\lambda)R_{n}(\alpha)|$$

$$- \frac{1}{2} \log \left( \frac{1}{nT} \sum_{t=1}^{T} (S_{n}(\lambda)Y_{nt} - Z_{nt}\delta - \Gamma_{n,fi})^{\prime}R_{n}(\alpha)^{\prime}R_{n}(\alpha)(S_{n}(\lambda)Y_{nt} - Z_{nt}\delta - \Gamma_{n,fi}) \right)$$

is a concentrated sample averaged log likelihood function of $\theta, \Gamma_{n}$ and $F_{T}$. In view of the unobserved nature of the common factor component, we shall make minimal assumptions on their structures. The number of factors is set at $r$ in the estimation, which does not necessarily equal to the true number of factors $r_{0}$. Later sections will show that consistency requires that $r \geq r_{0}$ while the results on limiting distribution require $r = r_{0}$. Because for our estimation method, no restriction is imposed on $\Gamma_{n}$ and $R_{n}(\alpha)$ is assumed invertible.

---

4This paper does not require $W_{n}$ to be row normalized, see Assumption R2 which is a weaker condition. Although row normalized spatial weights matrix is convenient to work with, in some applications it is not appropriate. For example, in the analysis of social interactions where the effect of network structure (e.g. centrality) is of interest or there are individuals who influence others but are not influenced by others, row normalization might not be appropriate, see Liu and Lee (2010).
for $\alpha$ in its parameter space, optimizing with respect to $\Gamma_n \in \mathbb{R}^{n \times r}$ is equivalent to optimizing with respect to the transformed $\tilde{\Gamma}_n$ with $\tilde{\Gamma}_n = R_n(\alpha)\Gamma_n$. The objective function can be equivalently written as

$$Q_{nT}(\theta) = \max_{\tilde{\Gamma}_n \in \mathbb{R}^{n \times r}, F_T \in \mathbb{R}^{r \times T}} Q_{nT}(\theta, \tilde{\Gamma}_n, F_T) = \frac{1}{n} \log |S_n(\lambda)R_n(\alpha)| - \frac{1}{2} \log \left( \frac{1}{nT} \sum_{t=1}^{T} (R_n(\alpha) (S_n(\lambda)Y_{nt} - Z_m\delta) - \tilde{\Gamma}_n f_t)'(R_n(\alpha) (S_n(\lambda)Y_{nt} - Z_m\delta) - \tilde{\Gamma}_n f_t) \right).$$

(4)

As the sample expands, the number of parameters in the factors and their loadings also increases. Because the parameter of interest is $\theta$, we concentrate out factors and their loadings using the principal component theory: $\min_{F_T \in \mathbb{R}^{T \times r}, \tilde{\Gamma}_n \in \mathbb{R}^{n \times r}} \text{tr} \left( (H_{nT} - \tilde{\Gamma}_n f_T') (H_{nT} - \tilde{\Gamma}_n f_T')' \right) = \min_{F_T \in \mathbb{R}^{T \times r}} \text{tr} (H_{nT} M_{F_T} H_{nT}) = \sum_{i=r+1}^{n} \mu_i (H_{nT} H_{nT})$ for an $n \times T$ matrix $H_{nT}$. The concentrated log likelihood is

$$Q_{nT}(\theta) = \max_{F_T \in \mathbb{R}^{T \times r}} Q_{nT}(\theta, \tilde{\Gamma}_n, F_T) = \frac{1}{n} \log |S_n(\lambda)R_n(\alpha)| - \frac{1}{2} \log L_{nT}(\theta),$$

(5)

with $L_{nT}(\theta) = \frac{1}{nT} \sum_{i=r+1}^{n} \mu_i \left( R_n(\alpha) (S_n(\lambda) - \sum_{k=1}^{K} Z_k\delta_k) (S_n(\lambda) - \sum_{k=1}^{K} Z_k\delta_k)' R_n(\alpha)' \right)$. The QML estimator is $\hat{\theta}_{nT} = \arg \max_{\theta \in \Theta} Q_{nT}(\theta)$. The estimate for $\tilde{\Gamma}_n$ can be obtained as the eigenvectors associated with the first $r$ largest eigenvalues of $R_n(\alpha) (S_n(\lambda) - \sum_{k=1}^{K} Z_k\delta_k) (S_n(\lambda) - \sum_{k=1}^{K} Z_k\delta_k)' R_n(\alpha)'$. By switching $n$ and $T$, the estimate for $F_T$ can be similarly obtained. Note that the estimated $\tilde{\Gamma}_n$ and $F_T$ are not unique, as $\tilde{\Gamma}_n H H^{-1} F_T'$ is observationally equivalent to $\tilde{\Gamma}_n F_T'$ for any invertible $r \times r$ matrix $H$. However, the column spaces of $\tilde{\Gamma}_n$ and $F_T$ are invariant to $H$, hence the projectors $M_{\tilde{\Gamma}_n}$ and $M_{F_T}$ are uniquely determined.

2.3. Assumptions

The true values of $\theta$, $\tilde{\Gamma}_n$ and $F_T$ are denoted by $\theta_0$, $\tilde{\Gamma}_{n0}$ and $F_{T0}$. Note that the dimensions of $\tilde{\Gamma}_n$ and $F_T$ are $n \times r$ and $T \times r$, and may not equal to the dimensions of $\tilde{\Gamma}_{n0}$ and $F_{T0}$ which are $n \times r_0$ and $T \times r_0$ respectively. Denote $\varepsilon = \left( \varepsilon_{n1} \varepsilon_{n2} \ldots \varepsilon_{nT} \right)$, a $n \times T$ matrix.

Assumption E

1. The disturbances $\varepsilon_{it}$ are independently distributed across $i$ and over $t$ with $\mathbb{E} \varepsilon_{it} = 0$, $\mathbb{E} \varepsilon_{it}^2 = \sigma_0^2 > 0$ and has uniformly bounded moment $\mathbb{E} |\varepsilon_{it}|^{4+\eta}$ for some $\eta > 0$.

2. The disturbances in $\varepsilon$ are independently distributed from regressors $X_t$ and the factors $F_{0T}$ and $\Gamma_{0n}$.

The disturbances of the model have a spatial structure. From Eq. (1), $U = R_n(\alpha_0)^{-1} \varepsilon$. Its spatial heterogeneity is captured by $\tilde{W}_n$ and coefficient $\alpha$. In panels with factors, many estimation methods allow idiosyncratic errors to be cross sectionally correlated and heteroskedastic in an unknown form, up to a degree (Bai}
(2009), Pesaran (2006)). However, when a spatial autoregressive model is estimated by QML assuming homoskedastic error, the QMLE is generally inconsistent if errors are in fact heteroskedastic but ignored (Lin and Lee (2010)).\(^5\) In the current setting, consistency of the QMLE requires this stronger homoskedastic assumption in \(\varepsilon\). Latala (2005) show that, under Assumption E, \(\|\varepsilon\|_2 = O_P\left(\sqrt{\max(n, T)}\right)\).

To have more simplified notations, define \(n \times n\) matrices \(S_n = S_n(\lambda_0), G_n = W_n S_n^{-1}, R_n = R_n(\alpha_0), \tilde{G}_n = \tilde{W}_n R_n^{-1}\) and \(n \times T\) matrices \(Z_{K+1} = \left( G_n Z_{n1} \delta_0 \cdots G_n Z_{nT} \delta_0 \right) = \sum_{k=1}^{K} \delta_{0k} G_n Z_k, Y = \left( Y_{n1} \cdots Y_{nT} \right)\) and \(Y_{-1} = \left( Y_{n0} \cdots Y_{nT-1} \right)\). In the case of nonnormal disturbance, denote \(\mu^{(3)} = \mathbb{E} \varepsilon_{it}^3\) and \(\mu^{(4)} = \mathbb{E} \varepsilon_{it}^4\).

**Assumption R**

1. The parameter \(\theta_0\) is in the interior of \(\Theta\), where \(\Theta\) is a compact subset of \(\mathbb{R}^{K+1}\). We use \(\Theta_\chi\) to denote the parameter space for parameter \(\chi, \chi = \lambda, \alpha, \) etc.

2. The spatial weights matrices \(W_n\) and \(\tilde{W}_n\) are non-stochastic. \(W_n, S_n^{-1}, \tilde{W}_n\) and \(R_n^{-1}\) are uniformly bounded in absolute value in both row and column sums (UB). \(S_n(\lambda)\) and \(R_n(\alpha)\) are invertible for any \(\lambda \in \Theta_\lambda\) and \(\alpha \in \Theta_\alpha\). Furthermore, \(\liminf_{n,T \to \infty} \inf_{\lambda \in \Theta_\lambda} |S_n(\lambda)| > 0\) and \(\liminf_{n,T \to \infty} \inf_{\alpha \in \Theta_\alpha} |R_n(\alpha)| > 0\).

3. The elements of \(X_{it}\) have uniformly bounded 4-th moments.

4. The number of factors is constant \(r_0\), and elements of \(\Gamma_{n0}\) and \(F_{T0}\) have uniformly bounded 4-th moments.

5. \(\sum_{n=1}^{\infty} \text{abs} (A_n)\) is UB, where \(\text{abs} (A_n)_{ij} = |A_{n,ij}|\). In addition, there exists a constant \(b < 1\) and \(n_0\), such that \(\|A_n\|_2 \leq b\) for all \(n \geq n_0\).

6. \(n\) is a nondecreasing function of \(T\). As \(T\) goes to infinity, so does \(n\).\(^6\)

Assumption R1 is standard. The sum \(\sum_{\ell=0}^{\infty} (\lambda W_n)^\ell\) is convergent if \(\|\lambda W_n\| < 1\) for some norm \(\|\cdot\|\).\(^7\) In this situation, \(S_n(\lambda)\) is invertible and \(S_n(\lambda)^{-1} = \sum_{\ell=0}^{\infty} (\lambda W_n)^\ell\) is Neumann’s series. Therefore a sufficient condition for the invertibility of \(S_n(\lambda)\) is \(|\lambda| < \frac{1}{\|W_n\|}\). Similar properties hold for \(R_n(\alpha)\).

2.4. Identification

The following Assumptions ID1 and ID2 are used to show that \(\theta_0, \Gamma_{01}F_{0T}\) and the number of factors can be uniquely recovered from the distribution of data on \(Y\) and \(Z\). Their sample counterparts are the subsequent

\(^5\)In Bai and Li (2014), disturbances can be heteroskedastic along the cross section but invariant over time. They are treated as parameters. They show that the estimates of spatial correlation and slope coefficients are consistent as \(T \to \infty\). Consistent estimation methods remain to be seen if variances depend on individual explanatory variables across units and time.

\(^6\)Nondecreasing function could be a constant function.

\(^7\)This condition does not depend on the type of norm used, since all norms in \(\mathbb{R}^{n \times n}\) are equivalent and therefore convergence of a series does not depend on a specific type of norm.
Assumptions NC1 and NC2. Assumptions ID1 and NC1 require that $Z_1, \cdots, Z_{K+1}$ are linearly independent, while Assumptions ID2 and NC2 relax this, but impose more restrictions on the variance structure. The linear independence conditions can fail if $G_t Z_{nt} \delta_t$ is linearly dependent on $Z_{nt}$ for all $t = 1, \cdots, T$. This can happen in a pure SAR model with no regressors in which case $\delta_t = 0$.

**Assumption ID1**

1. Let $z = \left( \text{vec}(Z_1) \cdots \text{vec}(Z_{K+1}) \right)$, which is an $nT \times (K + 1)$ matrix of regressors. The $(K + 1) \times (K + 1)$ matrix $E \left( z' \left( M_{F_{G_t}} \otimes \left( R_n(\alpha)' M_{F_{\delta}} R_n(\alpha) \right) \right) z \right)$ is positive definite for any $\alpha \in \Theta_\alpha$ and $\hat{\Gamma}_n \in \mathbb{R}^{n \times r}$ with some $r \geq r_0$ where $r_0$ is the true number of factors, $M_{F_{G_t}} = I_T - F_{0T} (F_{0T}' F_{0T})^{-1} F_{0T}'$ and $M_{F_\delta} = I_n - \hat{\Gamma}_n \hat{\Gamma}_n'$.

2. For any $\alpha \neq \alpha_0$, $R_n(\alpha)' R_n(\alpha)$ is linearly independent of $R_n' R_n$.

Assumption ID1(2) implies that $\frac{1}{n} \text{tr} \left( R_n^{-1} R_n(\alpha)' R_n(\alpha) R_n^{-1} \right) - \left| R_n^{-1} R_n(\alpha)' R_n(\alpha) R_n^{-1} \right|^\frac{1}{2} > 0$, by the inequality of arithmetic and geometric means. A sufficient condition for ID1(2) is that $I_n, \tilde{W}_n + \tilde{W}'_n$ and $W' \tilde{W}_n$ are linearly independent.\(^8\) Assumption ID1 generalizes those in Lee and Yu (2013a) on the identification of spatial panel models with additive individual and time effects.

**Assumption ID2**

1. Let $z = \left( \text{vec}(Z_1) \cdots \text{vec}(Z_K) \right)$, which is $nT \times K$. The $K \times K$ matrix $E \left( z' \left( M_{F_{G_t}} \otimes \left( R_n(\alpha)' M_{F_{\delta}} R_n(\alpha) \right) \right) z \right)$ is positive definite for any $\alpha \in \Theta_\alpha$ and $\hat{\Gamma}_n \in \mathbb{R}^{n \times r}$ with some $r \geq r_0$ where $r_0$ is the true number of factors, $M_{F_{G_t}} = I_T - F_{0T} (F_{0T}' F_{0T})^{-1} F_{0T}'$ and $M_{F_\delta} = I_n - \hat{\Gamma}_n \hat{\Gamma}_n'$.

2. For any $\lambda \in \Theta_\lambda$ and $\alpha \in \Theta_\alpha$, if $\lambda \neq \lambda_0$ or $\alpha \neq \alpha_0$, then $S_n(\lambda)' R_n(\alpha)' R_n(\alpha) S_n(\lambda)$ is linearly independent of $S_n' R_n' R_n S_n$\(^9\).

Assumption ID2(2) is equivalent to for any $\lambda \in \Theta_\lambda$ and $\alpha \in \Theta_\alpha$, if $\lambda \neq \lambda_0$ or $\alpha \neq \alpha_0$,

$$\frac{1}{n} \text{tr} \left( R_n^{-1} S_n^{-1} S_n(\lambda)' R_n(\alpha)' R_n(\alpha) S_n(\lambda) S_n^{-1} R_n^{-1} \right) - \left| R_n^{-1} S_n^{-1} S_n(\lambda)' R_n(\alpha)' R_n(\alpha) S_n(\lambda) S_n^{-1} R_n^{-1} \right|^\frac{1}{2} > 0.$$  

Assumption ID requires that regressors are not linearly dependent. For parameter identification, we need the concentrated expected objective function to be uniquely maximized at the truth. We assume that the

\(^8\)Let $c_1$ and $c_2$ be two scalars. For $\alpha \neq \alpha_0$, $c_1 R_n(\alpha)' R_n(\alpha) + c_2 R_n' R_n = (c_1 + c_2) I_n - (c_1 \alpha + c_2 \alpha_0) (W_n + W'_n) + (c_1 \alpha^2 + c_2 \alpha_0^2) W_n W'_n$ is equal to $0$ if $c_1 = c_2 = 0$ because $I_n, W_n + W'_n$ and $W_n W'_n$ are assumed to be linearly independent. Therefore for any $\alpha \neq \alpha_0$, $R_n(\alpha)' R_n(\alpha)$ is linearly independent of $R_n' R_n$. Notice that $W_n$ can be symmetric.

\(^9\)A sufficient condition is that the following 9 matrices are linearly independent, $I_n, \ W_n + W'_n, \ W_n + W'_n W_n, \ W_n W'_n, \ W'_n W_n, \ W_n W_n W_n + W'_n W_n + W_n W_n + W'_n W_n$ and $W_n W'_n W_n$, for the case that $W_n \neq W'_n$. In the event that $W_n = W'_n$. Assumption ID2(2) can only give local identification for $\lambda_0$ and $\alpha_0$ in the sense that $(\lambda_0, \alpha_0)$ cannot be distinguished from $(\alpha_0, \lambda_0)$. The latter situation is similar to the identification issue of a pure spatial autoregressive with spatial error process $Y_n = \lambda W_n Y_n + U_n$ with $U_n = \alpha W_n U_n + \epsilon_n.$
number of factors used in the concentrated expected objective function is not smaller than the true number of factors. Given that the number of latent factors is small in many empirical applications, it is reasonable to assume that an upper bound of the factor number is known. The estimated $\Gamma_n$ and $F_T$ need not have full column rank and the true number of factors, $r_0$, can be recovered from the rank of $\Gamma_n F_T$, as the following proposition shows.

**Proposition 1.** Under Assumptions E, R and ID1 (or ID2), $\theta_0, \Gamma_0 F_T'$ and $r_0$ are identified.

The proof is in Appendix B. Assumptions NC below are sample counterparts of Assumptions ID. They are specifically needed for the consistency of the proposed estimator. They can be slightly weakened, but will then involve the unobserved factors, as in Assumption A of Bai (2009).

**Assumption NC1**

1. There exists a positive constant $b$, such that $\min_{\eta \in B_{K+1}} \sum_{i=2r+1}^n \mu_i \left( \frac{1}{n^2} R_n(\alpha)(\eta \cdot Z)(\eta \cdot Z)' R_n(\alpha)' \right) \geq b > 0$ wpa 1 as $n, T \to \infty$, where $B_{K+1}$ is the unit ball of the $(K+1)$-dimensional Euclidean space; $\eta$ is a $(K+1) \times 1$ nonzero vector with $||\eta||_2 = \sqrt{\eta' \eta} = 1$; $\eta \cdot Z \equiv \sum_{k=1}^{K+1} \eta_k Z_k$ is a convex linear combination of those $n \times T$ matrices $Z_k$‘s.

2. For any $\alpha \in \Theta_\alpha$, $\alpha \neq \alpha_0$, $\liminf_{n,T \to \infty} \left( \frac{1}{n} \text{tr} \left( R_n^{-1} R_n(\alpha)' R_n(\alpha) R_n^{-1} \right) - \left| R_n^{-1} R_n(\alpha)' R_n(\alpha) R_n^{-1} \right|^\frac{1}{2} \right) > 0.$

The assumption in NC1(1) requires no perfect collinearity between regressors and sufficient variations for each regressor. Notice that this excludes constant regressors, because they are constant along $i$ or $t$, and

$$\sum_{i=2r+1}^n \mu_i \left( \frac{1}{n^2} R_n(\alpha)(\eta \cdot Z)(\eta \cdot Z)' R_n(\alpha)' \right) = 0$$

for them. Such regressors include those that do not vary over time, e.g., gender, race, and those that are common across the individuals, e.g., common time effects.

It is desirable to understand more about the condition in NC1(1) in terms of regressors implied by the SAR model. Define $n \times (K+1)$ matrices $Z_{nt} = (Z_{nt,1}, \cdots, Z_{nt,K}, G_n Z_{nt} \delta_0) = (Z_{nt}, G_n Z_{nt} \delta_0)$ for $t = 1, \cdots, T$; and the overall $n \times (K+1)T$ matrix $Z_{nT} = [Z_{n1}, \cdots, Z_{nT}]$. We have

$$R_n(\alpha)(\eta \cdot Z) = \sum_{k=1}^{K+1} \eta_k R_n(\alpha) Z_k = R_n(\alpha) \left( \sum_{k=1}^{K} \eta_k \begin{bmatrix} Z_{n1,k} & Z_{n2,k} & \cdots & Z_{nT,k} \end{bmatrix} + \eta_{K+1} \begin{bmatrix} G_n Z_{n1} \delta_0 \ G_n Z_{n2} \delta_0 \ \cdots \ G_n Z_{nT} \delta_0 \end{bmatrix} \right)$$

$$= R_n(\alpha) \sum_{k=1}^{K} Z_{n1,k} \eta_k + (G_n Z_{n1} \delta_0) \eta_{K+1}, \cdots, \sum_{k=1}^{K} Z_{nT,k} \eta_k + (G_n Z_{nT} \delta_0) \eta_{K+1}$$

$$= R_n(\alpha) \begin{bmatrix} Z_{n1} \eta \ Z_{n2} \eta \ \cdots \ Z_{nT} \eta \end{bmatrix} = R_n(\alpha) Z_{nT} (I_T \otimes \eta),$$

where $\eta = (\eta_1, \cdots, \eta_{K+1})'$. So the NC1(1) condition concerns about the smallest $(n-2r)$ eigenvalues of the $n \times n$ matrix $\frac{1}{n^2} R_n(\alpha)(\eta \cdot Z)(\eta \cdot Z)' R_n(\alpha)' = \frac{1}{n^2} R_n(\alpha) Z_{nT} (I_T \otimes \eta)(I_T \otimes \eta)' Z_{nT}' R_n(\alpha)'$ for each $\eta \in \Theta_\alpha$. 


\( B_{K+1} \) and \( \alpha \in \Theta_\alpha \). Because these matrices are nonnegative definite, their eigenvalues are nonnegative but some can be zero. If there were an \( \alpha \in \Theta_\alpha \) and \( \eta \in B_{K+1} \) with the \( n - 2r \) smallest eigenvalues of \( \frac{1}{n}R_n(\alpha)(\eta \cdot Z)(\eta \cdot Z)'R_n(\alpha)' \) being all zero, then the NC1 assumption would not be satisfied. So we need some sufficient conditions to rule out such cases.

**Proposition 2.** If \( \sum_{t=2r+1+KT}^n \mu_t(\frac{1}{n} \mathcal{Z}_{nt} \mathcal{Z}'_{nt}) > 0 \), i.e., the sum of the smallest \( n - 2r - 1 - KT \) eigenvalues of \( \frac{1}{n} \mathcal{Z}_{nt} \mathcal{Z}'_{nt} \) is positive, and \( \mu_t(R_n(\alpha)'R_n(\alpha)) > 0 \) for all \( \alpha \in \Theta_\alpha \), with probability approaching 1 as \( n, T \to \infty \), then Assumption NC1(1) is satisfied.

The proof is in Appendix B. In order for Proposition 2 to hold, it is necessary that \( \mathcal{Z}_{nt} \) has rank at least as large as \( KT + 2r + 1 \), that in turn, requires \( (2r + 1 + KT) \leq \min\{n, (K + 1)T\} \) because \( \mathcal{Z}_{nt} \) is a \( n \times (K + 1)T \) matrix. As the problem under consideration has both \( n \) and \( T \) tend to infinity and \( r \) is finite, the latter requires \( \frac{n}{T} > K \) for large enough \( n \) and \( T \).

The above analysis can be generalized to the case where \( G_nZ_{nt} \delta_0 \) is linearly dependent on \( Z_{nt} \) for all \( t = 1, \cdots, T \), such that \( G_nZ_{nt} \delta_0 = Z_{nt}C \) for a constant vector \( C \). As pointed out preceding Assumption ID1, this can happen in a pure SAR model. In this case, let \( \eta^* = (\eta_1, \cdots, \eta_K)' \). Here,

\[
R_n(\alpha)\eta \cdot Z = R_n(\alpha)[Z_{n1}\eta^* + (G_nZ_{n1}\delta_0)\eta_{K+1} + \cdots, Z_{nT}\eta^* + (G_nZ_{nT}\delta_0)\eta_{K+1}] = R_n(\alpha)[Z_{n1}, \cdots, Z_{nT}](I_T \otimes \tilde{\eta}),
\]

where \( \tilde{\eta} = \eta^* + \eta_{K+1}C \). The previous result can now be applied to the \( n \times KT \) matrix \( \mathcal{Z}_{nt}^* = [Z_{n1}, \cdots, Z_{nT}] \). In such case, we have an alternative set of conditions that guarantee consistency, as follows.

**Assumption NC2**

1. Suppose that \( G_nZ_{nt} \delta_0 = Z_{nt}C \) for a constant vector \( C \) for all \( t = 1, \cdots, T \). There exists a positive constant \( b \), such that \( \min_{\eta \in B_\kappa, \alpha \in \Theta_\alpha} \sum_{t=2r+1}^n \mu_t(\frac{1}{n}R_n(\alpha)(\eta \cdot Z)(\eta \cdot Z)'R_n(\alpha)') \geq b > 0 \) wpa 1 as \( n, T \to \infty \), where \( B_\kappa \) is the unit ball of the \( K \)-dimensional Euclidean space; \( \eta \) is a \( K \times 1 \) nonzero vector with \( ||\eta||_2 = \sqrt{\eta'\eta} = 1; \eta \cdot Z = \sum_{k=1}^K \eta_kZ_k \).

2. For any \( \lambda \in \Theta_\lambda \) and \( \alpha \in \Theta_\alpha \), if \( \lambda \neq \lambda_0 \) or \( \alpha \neq \alpha_0 \),

\[
\lim_{n,T \to \infty} \inf \left( \frac{1}{n} \text{tr} \left( R_n^{-1}S_n^{-1}S_n(\lambda)'R_n(\alpha)'R_n(\alpha)S_n(\lambda)S_n^{-1}R_n^{-1} \right) - \left| R_n^{-1}S_n^{-1}S_n(\lambda)'R_n(\alpha)'R_n(\alpha)S_n(\lambda)S_n^{-1}R_n^{-1} \right|^2 \right) > 0.
\]

When \( G_nZ_{nt} \delta_0 \) is linearly dependent on \( Z_{nt} \) for all \( t \), NC1(1) will not be satisfied. The additional condition (2) of NC2 on the variance structure will make up for it, as will be clear in Proposition 3 below.
3. Asymptotic Theory

3.1. Consistency

Standard argument for consistency of an extremum estimator consists of showing that, for any $\tau > 0$, $\limsup_{n,T \to \infty} \left( \max_{\theta \in \Theta(\theta_0, \tau)} E Q_{nT}(\theta) - E Q_{nT}(\theta_0) \right) < 0$, where $\Theta(\theta_0, \tau)$ is the complement of an open neighborhood of $\theta_0$ in $\Theta$ with radius $\tau$ (i.e., identification uniqueness); and $Q_{nT}(\theta) - E Q_{nT}(\theta)$ converges to zero uniformly on its parameter space $\Theta$.

In many situations, the objective function with a finite number of parameters contains averages, and LLN follows with regularity assumptions (e.g. Amemiya (1985)). With an additional smoothness condition, uniform LLN would also follow (Andrews (1987)). In our model, the concentrated likelihood function (Eq. (5)) involves sum of certain eigenvalues of a random matrix and is not in the direct form of sample averages. Furthermore, the number of parameters increases to infinity as $n$ (and $T$) tends to infinity. It turns out that for consistency proof, it is relatively easier to work with the objective function without concentrating out $\Gamma_n$ and $F_T$. The idea is to respectively find a lower bound and an upper bound of the objective function, and then to show that the former is strictly greater than the latter for any $\theta$ that is outside the $\tau-$neighborhood of $\theta_0$ for any $\tau > 0$, as $n$ and $T$ increase. Since we are maximizing the objective function, the upper bound at $\hat{\theta}$ must be not smaller than the lower bound at $\theta_0$, which implies that the distance between $\hat{\theta}_{nT}$ and $\theta_0$ is collapsing to 0 as $n$ and $T$ increase. Lemma 1 of Wu (1981) demonstrates this idea formally. Using this method, Moon and Weidner (2015) show consistency of an NLS estimator for a regression panel model with interactive effects. We adapt these arguments to the spatial panel setting.

**Proposition 3.** Under Assumptions NC1 (or NC2), $E$ and $R$, and assuming that the number of factors is not underestimated, i.e., $r \geq r_0$, then $\|\hat{\theta}_{nT} - \theta_0\|_1 = o_P(1)$.

The proof of Proposition 3 is in Appendix B. For consistency, we do not need to know the exact number of factors. It is enough that the true number of factors is less than or equal to the number of factors assumed in estimation. Intuitively, if the number of factors used is not fewer than the true number of factors, variations due to factors can be accounted for. This feature has been observed in Moon and Weidner (2015) for the panel regression model. This property turns out to hold also for spatial panels. A step in the consistency proof is based on the following inequality (Eq. B.9),

$$Q_{nT}(\theta_0) = \frac{1}{n} \log |S_n| - \frac{1}{2} \log \left( \min_{\Gamma_n \in \mathbb{R}^{r \times r}, F_T \in \mathbb{R}^{T \times r}} \frac{1}{nT} \sum_{t=1}^{T} \left( \Gamma_{0t} f_{0t} + U_{nt} - \Gamma_{nt} f_t \right)' R_n R_n \left( \Gamma_{0t} f_{0t} + U_{nt} - \Gamma_{nt} f_t \right) \right)$$

\[\geq \frac{1}{n} \log |S_n| - \frac{1}{2} \log \left( \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \epsilon_{it}^2 \right),\]
which together with an upper bound of $Q_{nT}(\hat{\theta}_{nT})$ justifies the condition of Eq. (B.10) for consistency. This inequality trivially holds if the true number of factors is at most $r$, but might not hold if the number of factors in the estimation is smaller than the number of true factors.

### 3.2. Limiting Distribution

In this section, we derive the limiting distribution of the QML estimator $\hat{\theta}_{nT}$. Recall that $\tilde{\Gamma}_n = R_n \Gamma_n$. Because $F_T$ and $\tilde{\Gamma}_n$ are concentrated out in estimation, only the true factor and loading will be needed via the analysis of first and second order derivatives of the concentrated objective function. Therefore, as no confusion will arise, in subsequent sections $F_T$ and $\tilde{\Gamma}_n$ refer to the true factor and loading with the subscript ‘0’ omitted for simplicity. For the consistency of $\hat{\theta}_{nT}$, we do not need to make limiting assumptions on $\tilde{\Gamma}_n$ and $F_T$. However, for the limiting distribution of $\hat{\theta}_{nT}$, additional assumptions on limiting behaviors of $\tilde{\Gamma}_n$ and $F_T$ are needed.

**Assumption SF**

The number of factors, $r_0$, is constant and known. $\text{plim}_{n,T \to \infty} \frac{1}{n} \tilde{\Gamma}_n^\prime \tilde{\Gamma}_n = \tilde{\Gamma}$ and $\text{plim}_{n,T \to \infty} \frac{1}{T} F_T^\prime F_T = \tilde{F}$ exist and are positive definite.

The above assumption implies that for large enough $n$ and $T$, all the eigenvalues of $\tilde{F}$ and $\tilde{\Gamma}$ are bounded away from zero and are bounded from above. For consistency, it is not necessary that the true number of factors is known, as long as it is constant and not larger than the number of factors specified in estimation. But for deriving the limiting distribution, the number of factors needs to be exact in order for asymptotic analysis to be tractable.\(^{10}\)

Define $d^2_{\text{min}}(\tilde{\Gamma}_n, F_T) = \frac{1}{n^2} \mu_0(\tilde{\Gamma}_n F_T^\prime F_T \tilde{\Gamma}_n^\prime)$ and $d^2_{\text{max}}(\tilde{\Gamma}_n, F_T) = \frac{1}{n^2} \mu_1(\tilde{\Gamma}_n F_T^\prime F_T \tilde{\Gamma}_n^\prime)$. Notice that $\frac{1}{n^2} \tilde{\Gamma}_n F_T^\prime F_T \tilde{\Gamma}_n^\prime$ has at most $r_0$ positive eigenvalues. As a consequence of Assumption SF, $\text{plim}_{n,T \to \infty} d^2_{\text{min}}(\tilde{\Gamma}_n, F_T) > 0$ and $\text{plim}_{n,T \to \infty} d^2_{\text{max}}(\tilde{\Gamma}_n, F_T) < \infty$.\(^{11}\) The total variation in $Y$ is $\frac{1}{n^2} \text{tr}(YY^\prime)$, and its component $\frac{1}{n^2} \text{tr}(\tilde{\Gamma}_n F_T^\prime F_T \tilde{\Gamma}_n^\prime)$ is due to common factors. Assumption SF guarantees that each of the $r_0$ factors has a nontrivial contribution towards $\frac{1}{n^2} \text{tr}(\tilde{\Gamma}_n F_T^\prime F_T \tilde{\Gamma}_n^\prime)$. Similar assumption is in Bai (2003, 2009). Moon and Weidner (2015) labels this “strong factor assumption”.

In deriving the limiting distribution of $\hat{\theta}_{nT}$, we need to express $L_{nT}(\theta)$ around $\theta_0$, where $L_{nT}(\theta) = \frac{1}{n^2} \text{tr}(YY^\prime)$.

---

\(^{10}\) In Moon and Weidner (2015), they show that with additional assumptions on regressors and error distribution, the additional term does not change the limiting distribution of the estimator. However, those additional assumptions are rather strong. In spatial models, relevant assumptions remain to be seen.

\(^{11}\) This is so because the $n \times n$ matrix $\frac{1}{n^2} \Gamma_n F_T^\prime F_T \Gamma_n$ and the $r_0 \times r_0$ matrix $\frac{1}{n^2} \Gamma_n F_T^\prime F_T \Gamma_n$ have the same nonzero eigenvalues, counting multiplicity. For large $n$ and $T$, $d^2_{\text{max}}(\Gamma_n, F_T) = \mu_1(\frac{1}{n^2} \Gamma_n F_T^\prime F_T \Gamma_n) \leq \mu_1(\frac{1}{n^2} \Gamma_n^\prime \Gamma_n) \mu_1(\frac{1}{T} F_T^\prime F_T) < \infty$, and $d^2_{\text{min}}(\Gamma_n, F_T) = \mu_0(\frac{1}{n^2} \Gamma_n F_T^\prime F_T \Gamma_n) \geq \mu_0(\frac{1}{n^2} \Gamma_n^\prime \Gamma_n) \mu_0(\frac{1}{T} F_T^\prime F_T) > 0$. See Theorem 8.12 (2) in Zhang (2011), which shows that for Hermitian and positive semidefinite $n \times n$ matrices $A$ and $B$, $\mu_i(A) \mu_i(B) \leq \mu_i(AB) \leq \mu_i(A) \mu_i(B)$. 

---

12
Theorems 1 and 3 characterize the asymptotic distribution, asymptotic bias and asymptotic variance of \( \hat{\theta}_{nT} \). The perturbation theory of linear operators is used. The technical details of perturbation are in Appendix C and the supplementary file.

We now provide the limiting distribution of \( \hat{\theta}_{nT} \). The detailed proofs are in Appendix D. Let \( \mathcal{G}_{nT} \) denote the sigma algebra generated by \( X_{n1}, \cdots, X_{nT}, \hat{\Gamma}_n \) and \( F_T \). Define the \( n \times T \) matrices \( \tilde{Z}_k = E(Z_k|\mathcal{G}_{nT}) \), \( Z_k \equiv M_{Tn}R_n\bar{Z}_kM_{F_T} + M_{Tn}R_n(Z_k - \bar{Z}_k), k = 1, \cdots, K + 1 \) and the \( n \times T \) matrix of lagged disturbances \( \bar{\varepsilon}_i = \left( \varepsilon_{n1-h} \cdots \varepsilon_{nT-h} \right), h \geq 1 \), where we drop subscripts \( n \) and \( T \) for those matrices for simplicity. Using the reduced form of the dynamic equation in (1), we have

\[
\begin{align*}
Z_1 - Z_1 &= \sum_{h=1}^{\infty} A_n^{h-1} S_n^{-1} R_n^{-1} \bar{\varepsilon}_h, \\
Z_2 - Z_2 &= W_n \sum_{h=1}^{\infty} A_n^{h-1} S_n^{-1} R_n^{-1} \bar{\varepsilon}_h, \\
Z_k - Z_k &= 0 \quad \text{for } k = 3, \cdots, K, \\
Z_{K+1} - Z_{K+1} &= (\gamma_0 G_n + \rho_0 G_n W_n) \sum_{h=1}^{\infty} A_n^{h-1} S_n^{-1} R_n^{-1} \bar{\varepsilon}_h.
\end{align*}
\]

(6)

Theorems 1 and 3 characterize the asymptotic distribution, asymptotic bias and asymptotic variance of \( \hat{\theta}_{nT} \).

**Theorem 1.** Assume that \( \frac{n}{T} \rightarrow \kappa > 0 \) and Assumptions NC1 (or NC2), E, R and SF hold, then

\[
\sqrt{nT} \left( \hat{\theta}_{nT} - \theta_0 \right) - \left( \sigma_0^2 D_{nT} \right)^{-1} \varphi_{nT} = \left( \sigma_0^2 D_{nT} \right)^{-1} \frac{1}{\sqrt{nT}} v_{nT} + o_p(1),
\]

(7)

where \( D_{nT} \) is defined in Eq. (9) and is assumed to be positive definite, \( \varphi_{nT} = \left( \varphi_{nT,\gamma}, \varphi_{nT,\rho}, \varphi_{nT,\lambda}, \varphi_{nT,\alpha} \right)' \) with \( \varphi_{nT,\gamma} = -\frac{\alpha_0^2}{\sqrt{nT}} \sum_{h=1}^{T-1} \text{tr} \left( J_0 P_{F_T} J'_h \right) \text{tr} \left( A_n^{h-1} S_n^{-1} \right), \varphi_{nT,\rho} = -\frac{\alpha_0^2}{\sqrt{nT}} \sum_{h=1}^{T-1} \text{tr} \left( J_0 P_{F_T} J'_h \right) \text{tr} \left( W_n A_n^{h-1} S_n^{-1} \right), \varphi_{nT,\lambda} = -\frac{\alpha_0^2}{\sqrt{nT}} \sum_{h=1}^{T-1} \text{tr} \left( J_0 P_{F_T} J'_h \right) \text{tr} \left( (\gamma_0 G_n + \rho_0 G_n W_n) A_n^{h-1} S_n^{-1} \right), \varphi_{nT,\alpha} = \frac{\alpha_0^2}{\sqrt{nT}} \left( \frac{n}{T} \text{tr} \left( \tilde{G}_n \right) - \text{tr} \left( P_{Tn} \tilde{G}_n \right) \right), J_h = \left( 0_{T \times (T-h)}, I_{T \times T}, 0_{T \times h} \right)', I_{T \times T} \text{ is the } T \times T \text{ identity matrix, and}

\[
v_{nT} = \left( \text{tr} \left( \bar{Z}_1 \varepsilon' \right), \cdots, \text{tr} \left( \bar{Z}_K \varepsilon' \right) \right),
\]

\[
\text{tr} \left( \bar{Z}_{K+1} \varepsilon' \right) + \frac{1}{\sqrt{nT}} \text{tr} \left( R_n G_n R_n^{-1} \varepsilon \varepsilon' \right) - \frac{1}{n} \text{tr} \left( G_n \right) - \frac{1}{\sqrt{nT}} \text{tr} \left( \tilde{G}_n \varepsilon \varepsilon' \right) - \frac{1}{n} \text{tr} \left( \tilde{G}_n \right).
\]

To derive the joint distribution of \( v_{nT} \), Cramér-Wold device can be used. Let \( c \in \mathbb{R}^{K+2} \),

\[
c' v_{nT} = \text{tr} \left( \sum_{k=1}^{K+1} c_k \bar{Z}_k \varepsilon' \right) + c_{K+1} \left( \text{tr} \left( R_n G_n R_n^{-1} \varepsilon \varepsilon' \right) - \frac{1}{n} \text{tr} \left( G_n \right) \text{tr} \left( \varepsilon \varepsilon' \right) \right) + c_{K+2} \left( \text{tr} \left( \tilde{G}_n \varepsilon \varepsilon' \right) - \frac{1}{n} \text{tr} \left( \tilde{G}_n \right) \text{tr} \left( \varepsilon \varepsilon' \right) \right)
\]

\[
= \text{vec} \left( \sum_{k=1}^{K+1} c_k \bar{Z}_k \right)' \text{vec} \left( \varepsilon \right) + \text{vec} \left( \varepsilon' \right) c_{K+1} \left( I_T \otimes \left( R_n G_n R_n^{-1} - \frac{1}{n} \text{tr} \left( G_n \right) \right) \right) \text{vec} \left( \varepsilon \right)
\]

\[
= b_{nT}^c \text{vec} \left( \varepsilon \right) + \omega_{nT}^c \text{vec} \left( \varepsilon \right) + \text{vec} \left( \varepsilon' \right) A_{nT}^c \text{vec} \left( \varepsilon \right),
\]

where \( b_{nT}^c = \sum_{k=1}^{K+1} c_k \text{vec} \left( M_{Tn} R_n \bar{Z}_k M_{F_T} \right), \omega_{nT}^c = \text{vec} \left( \sum_{h=1}^{\infty} P_{nh} \bar{\varepsilon}_h \right) \) with \( P_{nh} = B_n^c A_n^{h-1} S_n^{-1} R_n^{-1} \), and

\[
B_n^c = M_{Tn} R_n \left( c_1 I_T + c_2 W_n + c_{K+1} (\gamma_0 G_n + \rho_0 G_n W_n) \right), \quad A_{nT}^c = \frac{1}{2} \left( A_{nT}^c + A_{nT}^c' \right) \text{ is an } nT \times nT \text{ symmetric matrix with}
\]

\[
A_{nT}^c = c_{K+1} H_{K+1} + c_{K+2} H_{K+2}, H_{K+1} = I_T \otimes \left( R_n G_n R_n^{-1} - \frac{1}{n} \text{tr} \left( G_n \right) \right), \quad \text{and } H_{K+2} = I_T \otimes \left( \tilde{G}_n - \frac{1}{n} \text{tr} \left( \tilde{G}_n \right) \right).
\]

13
Under Assumptions R and SF, elements of $b_{nT}^c$ have uniformly bounded 4-th moments; $A_{nT}^c, B_{nT}^c, \sum_{h=1}^n \abs{A_{0h}^c}$, $S_{nT}^{-1}$ and $R_{nT}^{-1}$ are UB. Together with Assumption E, the CLT of the martingale difference array for linear-quadratic form (Kelejian and Prucha (2001) and Yu et al. (2008) Lemma 13) are applicable.

**Theorem 2.** Under Assumptions E, R and SF; $\Ec'v_{nT} = \sigma_0^2 \text{tr}(A_{nT}^c) = 0$,

\[
\text{var}(c'v_{nT}) = 4 \sigma_0^4 \text{tr}(\Sigma_0^{-1}) + 2 \sigma_0^4 \text{tr}(A_{nT}^c) + \left( \mu(4) - 3 \sigma_0^4 \right) \sum_{i=1}^{nT} \abs{[A_{nT,i}]^2} \]

\[
= nT \sigma_0^4 c' (D_{nT} + \Sigma_{nT} + op(1)) c, \tag{8}
\]

where the $(K+2) \times (K+2)$ matrix $D_{nT}$ is

\[
D_{nT} = \frac{1}{nT} \sum_{i=1}^{nT} (\Xi_i' \Xi_k + \Xi_i \Xi_k' + \Xi_i \Xi_k'),
\]

where $\Xi_i = (\Xi_1 \cdots \Xi_{K+1})$, with $\Xi_k = \text{vec} \left( M_{k} R_{k} A_{k} M_{k} \right)$, $k = 1, \cdots, K+1$;

\[
\psi_{K+1,K+1} = \frac{1}{nT} \text{tr} \left( R_{k} A_{k} R_{k} A_{k}^{-1} \right), \quad \psi_{K+1,K+2} = \frac{1}{nT} \text{tr} \left( R_{k} A_{k} R_{k}^{-1} \right), \quad \psi_{K+2,K+2} = \frac{1}{nT} \text{tr} \left( R_{k} A_{k} R_{k}^{-1} \right), \quad \text{and furthermore}
\]

\[
\Sigma_{nT} = \frac{\mu(3)}{\sigma_0^4} \left( \begin{array}{ccc}
0_{K \times K} & \Sigma_{1,K+1}^{T+A} & \Sigma_{1,K+2}^{T+A} \\
\Sigma_{1,K+1}^{T+A} & \Sigma_{K+1,K+1}^{T+A} & \Sigma_{K+1,K+2}^{T+A} \\
\Sigma_{1,K+2}^{T+A} & \Sigma_{K+1,K+2}^{T+A} & \Sigma_{K+2,K+2}^{T+A}
\end{array} \right) + \frac{\mu(4) - 3 \sigma_0^4}{\sigma_0^4} \left( \begin{array}{ccc}
0_{K \times K} & 0 & 0 \\
0 & 0_{T+B} & 0_{T+B} \\
0 & 0_{T+B} & 0_{T+B}
\end{array} \right),
\]

where $\Sigma_{i,k}^{T+A} = \frac{1}{nT} \sum_{i=1}^{nT} \text{vec} \left( M_{k} R_{k} Z_{k} A_{k} M_{k} \right)$, $i = 1, \cdots, K+1$ and $k = 1, \cdots, K + 2$; $\Sigma_{i,k}^{K+1,K+1} = \frac{1}{nT} \sum_{i=1}^{nT} (H_{k,i})^2$, $\Sigma_{i,k}^{K+1,K+2} = \frac{1}{nT} \sum_{i=1}^{nT} (H_{k+1,i})^2$, $\Sigma_{i,k}^{K+2,K+2} = \frac{1}{nT} \sum_{i=1}^{nT} (H_{k+2,i})^2$; and $\Sigma_{T+B} = \frac{1}{nT} \sum_{i=1}^{nT} H_{K+1,i} H_{K+2,i}$.

**Theorem 3.** Assume that $\frac{T}{n} \rightarrow \kappa^2 > 0$; $D = \text{plim}_{n,T \rightarrow \infty} D_{nT}$ is positive definite; $\Sigma = \text{plim}_{n,T \rightarrow \infty} \Sigma_{nT}$; $\varphi = \text{plim}_{n,T \rightarrow \infty} \varphi_{nT}$; and suppose that Assumptions NC1 (or NC2), E, R and SF hold, then $\sqrt{nT} (\hat{Q}_{nT} - Q_0) - (\sigma_0^2 D)^{-1} \varphi \overset{d}{\rightarrow} N(0, D^{-1}(D + \Sigma) D^{-1})$.

Theorem 3 shows that the limiting distribution of $\hat{Q}_{nT}$ may not center at $Q_0$, with an asymptotic bias term $(\sigma_0^2 D)^{-1} \varphi$. For a regression panel with factors, Moon and Weidner (2014) show that leading order biases are due to the correlation between the predetermined regressors with the disturbances, and heteroskedastic disturbances. In our model, the biases arise from the predetermined regressors and the interaction between
the spatial effects and the factor loadings. As the time factors and loadings contain infinite number of parameters when sample sizes $n$ and $T$ go to infinity, the asymptotic bias must be due to the incidental parameter problem. In the next subsection, a bias corrected estimator will be proposed. The variance matrix in the limiting distribution has a sandwich form to accommodate possible non-normal disturbances. When disturbances in the model are normally distributed, $\Sigma = 0$ and the limiting variance will be $D^{-1}$ in a single matrix form.

### 3.3. Bias Correction

This section proposes a method to correct the bias $\phi$ in the limiting distribution. The bias depends on $\theta$, $P_{\Gamma_n}$, $P_{\Gamma}$, $\sigma_0^2$ and $D^{-1}$. The parameter $\theta$ can be estimated by $\hat{\theta}_{nT}$. The projectors $P_{\Gamma_n}$ and $P_{\Gamma}$ can be estimated as follows. From Eq. (5), let $\hat{B}_{\Gamma_n}$ denote the $n \times r_0$ matrix of the eigenvectors associated with the largest $r_0$ eigenvalues of $\frac{1}{nT}R_n(\hat{\alpha}) \left(S_n(\hat{\lambda}) Y - \sum_{k=1}^{K} Z_k \hat{\delta}_k\right)' \left(S_n(\hat{\lambda}) Y - \sum_{k=1}^{K} Z_k \hat{\delta}_k\right)' R_n(\hat{\alpha})'$, where the subscripts $nT$ of parameter estimates are dropped for simplicity, then $\hat{P}_{\Gamma_n} = \hat{B}_{\Gamma_n} \hat{B}_{\Gamma_n}'$, $\tilde{M}_{\Gamma_n} = I_n - \hat{P}_{\Gamma_n}$. Interchanging the role of $n$ and $T$, the factors $P_{\Gamma}$ and $M_{\Gamma}$ can be similarly estimated. Let $\hat{B}_{F_n}$ denote the $T \times r_0$ matrix of the eigenvectors associated with the largest $r_0$ eigenvalues of $\frac{1}{nT} \left(S_n(\hat{\lambda}) Y - \sum_{k=1}^{K} Z_k \hat{\delta}_k\right)' \left(S_n(\hat{\lambda}) Y - \sum_{k=1}^{K} Z_k \hat{\delta}_k\right)' R_n(\hat{\alpha})' R_n(\hat{\alpha}) \left(S_n(\hat{\lambda}) Y - \sum_{k=1}^{K} Z_k \hat{\delta}_k\right)'$, and $\hat{P}_{F_n} = \hat{B}_{F_n} \hat{B}_{F_n}'$, $\tilde{M}_{F_n} = I_T - \hat{P}_{F_n}$. Lemma 11 in Appendix C shows that under the assumptions of Theorem 3, $\left\| \tilde{M}_{\Gamma_n} - M_{\Gamma_n} \right\|_2 = \left\| \hat{P}_{\Gamma_n} - P_{\Gamma_n} \right\|_2 = O_P(\frac{1}{\sqrt{n}})$ and $\left\| \tilde{M}_{F_n} - M_{F_n} \right\|_2 = \left\| \hat{P}_{F_n} - P_{F_n} \right\|_2 = O_P(\frac{1}{\sqrt{T}})$.

The variance $\sigma_0^2$ can be estimated by

$$\hat{\sigma}^2 = L_{nT}(\hat{\theta}_{nT}) = \frac{1}{nT} \sum_{i=r_0+1}^{n} \mu_i \left( R_n(\hat{\alpha}) \left(S_n(\hat{\lambda}) Y - \sum_{k=1}^{K} Z_k \hat{\delta}_k\right) \left(S_n(\hat{\lambda}) Y - \sum_{k=1}^{K} Z_k \hat{\delta}_k\right)' \right)' R_n(\hat{\alpha})'. \tag{11}$$

Finally, $\hat{D}_{nT}$ is an estimate of $D_{nT}$ (see Eq. (9)) with estimated $\hat{\theta}_{nT}$, $P_{\Gamma_n}$, $P_{\Gamma}$, and $\hat{\sigma}^2$, and hence $\hat{D}_{nT}$ also estimates $D$. The bias $\phi$ can then be estimated by plugging in these estimated elements. The following theorem shows that the bias corrected estimator $\hat{\phi}_{nT}$ is asymptotically normal and centered at $\theta_0$.

**Theorem 4.** Assume that $\frac{T}{n} \to \kappa^2 > 0$; $D = \text{plim}_{n,T \to \infty} D_{nT}$ is positive definite; $\Sigma = \text{lim}_{n,T \to \infty} \Sigma_{nT}$; and suppose that Assumptions NC1 (or NC2), E, R and SF hold, the bias corrected estimator is $\hat{\phi}_{nT} = \hat{\phi}_{nT} - \left(\hat{\sigma}_{nT}^2 \hat{D}_{nT}\right)^{-1} \frac{1}{\sqrt{nT}} \hat{\phi}_n$. Then $\sqrt{nT} \left(\hat{\phi}_{nT} - \theta_0\right) \overset{d}{\to} N\left(0, D^{-1} (D + \Sigma) D^{-1}\right)$.

Section 4 investigates the finite sample performance of the QML estimators in a Monte Carlo study.

### 3.4. The number of factors

As long as the number of factors specified is no fewer than the true number of factors, the QML estimator is consistent. However, the limiting distribution is under the premise that the number of factors is correctly specified. Although Moon and Weidner (2015) show that, under certain conditions, the limiting distribution
of the NLS estimator for a regression panel is invariant to the inclusion of redundant factors, its finite sample performance may suffer, as Lu and Su (2015) emphasize. If factors are interpreted as omitted variables, their detection is a first step in trying to measure them. In this section, we demonstrate how the factor number can be consistently determined. Denote \( \hat{\theta}_{nT} \) the QML estimator with \( r \geq r_0 \), which is a preliminary consistent estimator using a large number of factors. The residuals \( \hat{D}_{nT} = R_n(\hat{\theta}) \left( S_n \left( \hat{\lambda} \right) Y - \sum_{k=1}^{K} Z_k \hat{\delta}_k \right) \) have approximate factor structures as in Bai and Ng (2002). To see this, using the notation of Eq. (C.1), we have \( \hat{D}_{nT} = \hat{\Gamma}_n F_n^T + \varepsilon + \hat{E}_{nT}(\hat{\theta}) \), with \( \hat{E}_{nT}(\hat{\theta}) = \sum_{k=1}^{K+2} \hat{\eta}_k V_k + \sum_{k_1,k_2=1}^{K+2} \hat{\eta}_{k_1,k_2} V_{k_1} V_{k_2} \). Several criterion on factor number selection have been proposed in the literature, including Bai and Ng (2002)’s PC and IC criterion, Onatski (2010)’s Edge Distribution estimator, and Ahn and Horenstein (2013)’s eigenvalue ratio tests. We specifically show how Ahn and Horenstein (2013)’s eigenvalue ratio criterion is used here. Denote \( \beta_{nT,0} = V(0)/\log(\min(n, T)) \), \( \hat{\mu}_{nT,k} = \mu_k \left( \frac{n}{T} \hat{D}_{nT} \hat{D}_{nT}^T \right) \) for \( k \geq 1 \), \( V(k) = \sum_{j=k+1}^{n} \hat{\mu}_{nT,j} \) for \( k \geq -1 \). Define the “eigenvalue ratio” statistic, \( \text{ER}(k) = \frac{\hat{\beta}_{nT,k}}{\beta_{nT,k+1}} \), and the “growth ratio” statistic, \( \text{GR}(k) = \log \left( \frac{V(k+1)}{V(k)} \right) / \log \left( \frac{V(k)}{V(k+1)} \right) \). The number of factors can be selected according to \( \hat{k}_{\text{ER}} = \max_{0 \leq k \leq k_{\text{max}}} \text{ER}(k) \) or \( \hat{k}_{\text{GR}} = \max_{0 \leq k \leq k_{\text{max}}} \text{GR}(k) \) where \( k_{\text{max}} \) is a pre-specified constant.

**Theorem 5.** Assuming that Assumptions E, R and SF hold, \( \lim_{n,T \to \infty} \frac{\hat{\theta}_{nT} - \theta_0}{\sqrt{n}} \rightarrow \kappa > 0 \), the preliminary estimator \( \left\| \hat{\theta}_{nT} - \theta_0 \right\|_2 = o_p \left( n^{-\frac{1}{2}} \right) \), \( r_0 \geq 0 \), then \( \lim_{n,T \to \infty} \text{Pr}(\hat{k}_{\text{ER}} = r_0) = \lim_{n,T \to \infty} \text{Pr}(\hat{k}_{\text{GR}} = r_0) = 1 \).

Therefore the number of factors can be determined consistently. The proof, which is in Appendix D, checks that the relevant assumptions of Ahn and Horenstein (2013) are satisfied and their result then applies here. Note that the case with no factors \( (r_0 = 0) \) is covered.

4. Monte Carlo simulations

4.1. Design

We study the finite sample performance of the QML estimator and the accuracy of the factor number selection in samples of different sizes, different degrees of spatial interaction, and different ratios between the variances of the idiosyncratic error and the factors. We also illustrate finite sample biases due to misspecification when the DGP has either spatial or interactive effects but they are ignored.

The DGP is \( Y_{nt} = \lambda_0 W_n Y_{nt} + \gamma_0 Y_{nt-1} + \rho_0 W_n Y_{nt-1} + X_{nt} \beta_0 + \Gamma_0 n + U_{nt} + \varepsilon_{nt} \), with \( U_{nt} = \alpha_0 W_n U_{nt} + \varepsilon_{nt} \). The dependent variable is affected by 2 unobserved factors \( F_{0T} = \left( f_{01} \quad f_{02} \quad \cdots \quad f_{0T} \right)' \) which is \( T \times 2 \) and their \( n \times 2 \) loadings matrix \( \Gamma_{0n} = \left( \gamma_{01} \quad \gamma_{02} \quad \cdots \quad \gamma_{0n} \right)' \). \( X_{nt} = \left( X_{nt,1} \quad X_{nt,2} \right)' \) is a \( n \times 2 \) matrix of two regressors, which are generated according to \( X_{nt,1i} = 0.25 ( \gamma_{0i} f_{0i} + (\gamma_{0i} f_{0i})^2 + \ell' \eta_{it,1} + \ell' f_{0i} + \eta_{it,1} ) \) and \( X_{nt,2i} = \eta_{it,2} \), where \( \ell = \left( 1 \quad 1 \right)' \). Elements of \( \gamma_{0i}, f_{0i}, \eta_{it,1} \) and \( \eta_{it,2} \) are generated from independent standard normal variables. We see that \( X_{nt,1i} \) is correlated with the common component \( \gamma_{0i} f_{0i} \), its square, and the
factors and loadings separately. $X_{nt,2i}$ is not affected by the factors. The spatial weights matrix $W_n$ is generated from a rook matrix. Individual units are arranged row by row on an $\sqrt{n} \times \sqrt{n}$ chessboard where neighbors are defined as those who share a common border. Units in the interior of the chessboard have 4 neighbors, and units on the border and corner have respectively 3 and 2 neighbors. This design of the spatial weights matrix is motivated by the observation that regions in most observed regional structures have similar connectivity as units in the rook matrix. Define $n \times n$ matrix $\tilde{M}_n$, such that $\tilde{M}_{n,ij} = 1$ if and only if individuals $i$ and $j$ are neighbors on the chessboard, and $\tilde{M}_{n,ij} = 0$ otherwise. Then the spatial weights matrix $W_n$ is defined as a row normalized $\tilde{M}_n$.

The idiosyncratic errors $\varepsilon_{it}$’s are generated from independent normal $(0, \vartheta)$ distributions. Elements of the common factor components and the idiosyncratic errors have the same variances when $\vartheta = 1$, and the latter has more variation when $\vartheta > 1$. For each Monte Carlo experiment, $X_{nt}, \Gamma_{0n}$ and $F_{0T}$ are generated according to the above specification for $T + 1000$ time periods and the last $T$ periods are taken as our sample.

4.2. Monte Carlo results

1000 Monte Carlo replications are carried out for each design. The numerical maximization routine starts at multiple values, because the objective function is not concave and multiple local maxima might exist. The baseline specification is $\theta^a_0 = (0.3, 0.3, 0.3, 0.3, 1, 1)$ for $\theta = (\lambda, \gamma, \rho, \alpha, \beta_1, \beta_2)$. We then consider the cases with negative spatial interaction ($\theta^b_0 = (-0.3, 0.3, 0.3, 0.3, 1, 1)$), no spatial correlation in disturbances ($\theta^c_0 = (0.3, 0.3, 0.3, 0, 1, 1)$), and no contemporaneous spatial interaction ($\theta^d_0 = (0, 0.3, 0.3, 0.3, 1, 1)$).

We use combinations of $n = 25, 49, 81$ and $T = 25, 49, 81$. The Monte Carlo designs cover panels with small and moderate $n$ and $T$. Table 1 reports the Monte Carlo results for the QML estimator. The magnitude of biases generally decreases as $n$ and $T$ increase. The coverage probability (CP) is calculated using the asymptotic variance covariance matrix and the nominal coverage probability is set at 95%. The estimates for $\alpha$ have noticeable biases in finite sample, and as a result, their CPs are well below 95%. The CPs for other parameter estimates are also generally below 95% and therefore hypothesis tests will have over-rejection. The Monte Carlo results of the bias corrected estimator are reported in Table 2. The biases have been reduced significantly, especially for $\alpha$. The CPs also improve which indicate a more reliable statistical inference based on the bias corrected estimator. Due to limited space, Monte Carlo results for $\theta^c_0$ and $\theta^d_0$ are reported in the supplementary file.

Moon and Weidner (2015) show that in regression panel, the limiting distribution of the QML estimator does not change when the number of factors is overspecified. Such a result is not available for the spatial panel considered in this paper, but consistency is still possible as argued. In Table 3, we report the performance of the bias corrected estimators when the number of factors is over-specified by 1, i.e. $r = 3$ while the true $r_0 = 2$. Although estimates are still consistent, $\alpha$ has noticeable bias in small sample that is not
removed by the bias correction procedure. Tables S.5 and S.6 in the supplementary file report additional Monte Carlo results with more redundant factors. As the number of redundant factors increases, the CP deteriorates. Note that the biases and CP improve in large samples (e.g., \( n = 81 \) and \( T = 81 \)), and this is consistent with the results of Moon and Weidner (2015). Therefore for valid inference in small sample, it is important that a correct number of factors is chosen. The estimators are less sensitive to redundant factors in large samples.

We check the accuracy of factor selections given by the eigenvalue ratio (ER) and growth ratio (GR) criteria. Figure 1 shows the number of incorrect selection in 1000 simulations. The accuracy is almost 100% when the variances of the idiosyncratic error and the factors are the same \((\vartheta = 1)\), even in small sample \((n = 25, T = 25)\). We then make the idiosyncratic error to have up to 9 times more variation than the factors so factors are relatively weaker, and find that the selection errors increase as a result. However, the selection accuracy quickly improves as sample size increases, and close to 100% accuracy is achieved in the sample with \( n = 81 \) and \( T = 81 \). The ER and GR criteria have similar overall performances, and GR criterion performs slightly better when the factors are weak (high \( \vartheta \)).

For misspecification issues, Tables S.3 and S.4 in the supplementary file show that estimators might not be consistent as the biases are substantial if factors or spatial effects are ignored in the estimation but in fact they exist. Such biases are rather severe for the estimates \( \hat{\rho} \) and \( \hat{\alpha} \).

5. Empirical application: spatial spillovers in mortgage originations

Our empirical application is motivated by Haurin et al. (2014), which analyzes the effect of house prices on state-level origination rates of the Home Equity Conversion Mortgage (HECM), but does not consider spatial spillovers. HECM is the predominant type of reverse mortgages which enable senior homeowners to withdraw their home equity without home sale or monthly payments. HECMs are insured and regulated by the federal government, although the private market originates the loans. The insurance is provided by the Federal Housing Administration through the mutual mortgage insurance fund, which guarantees that the borrower can have access to the loan fund in the future even when the lender is no longer in business, and the lender can be fully repaid when the loan terminates even if the house value is less than the loan balance. The borrower pays mortgage insurance premium both at loan closing and monthly over the lifetime of the loan. Haurin et al. (2014) find that states with past volatile house prices and current house price levels above long term norms have higher origination rates. This is consistent with the hypothesis that households use HECMs to insure against house price declines and therefore the mortgage insurance should take into account this behavioral response to house price dynamics, as the insurance fund will face higher claim risk when the
Table 1: Performance of the QML Estimator

<table>
<thead>
<tr>
<th>$\theta_0$</th>
<th>$n$</th>
<th>$T$</th>
<th>$\theta$</th>
<th>$\lambda$</th>
<th>$\hat{\gamma}$</th>
<th>$\hat{\rho}$</th>
<th>$\hat{\alpha}$</th>
<th>$\hat{\beta}_1$</th>
<th>$\hat{\beta}_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_0^a$</td>
<td>25</td>
<td>25</td>
<td>Bias</td>
<td>-0.00023</td>
<td>-0.00135</td>
<td>0.00007</td>
<td>0.01910</td>
<td>0.00173</td>
<td>-0.00137</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.921</td>
<td>0.920</td>
<td>0.909</td>
<td>0.869</td>
<td>0.929</td>
<td>0.919</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>49</td>
<td>Bias</td>
<td>0.00026</td>
<td>-0.00076</td>
<td>-0.00049</td>
<td>0.02024</td>
<td>-0.00024</td>
<td>-0.00038</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.921</td>
<td>0.927</td>
<td>0.936</td>
<td>0.859</td>
<td>0.918</td>
<td>0.919</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>81</td>
<td>Bias</td>
<td>0.00074</td>
<td>-0.00047</td>
<td>-0.00039</td>
<td>0.02142</td>
<td>0.00057</td>
<td>-0.00093</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.933</td>
<td>0.931</td>
<td>0.924</td>
<td>0.830</td>
<td>0.932</td>
<td>0.933</td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>25</td>
<td>Bias</td>
<td>-0.00078</td>
<td>-0.00126</td>
<td>0.00103</td>
<td>0.00940</td>
<td>-0.00020</td>
<td>-0.00080</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.934</td>
<td>0.918</td>
<td>0.936</td>
<td>0.907</td>
<td>0.920</td>
<td>0.940</td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>49</td>
<td>Bias</td>
<td>0.00015</td>
<td>-0.00022</td>
<td>-0.00017</td>
<td>0.01268</td>
<td>-0.00054</td>
<td>0.00017</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.946</td>
<td>0.938</td>
<td>0.942</td>
<td>0.890</td>
<td>0.943</td>
<td>0.938</td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>81</td>
<td>Bias</td>
<td>-0.00037</td>
<td>-0.00000</td>
<td>0.00001</td>
<td>0.01080</td>
<td>0.00070</td>
<td>0.00087</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.944</td>
<td>0.934</td>
<td>0.944</td>
<td>0.901</td>
<td>0.939</td>
<td>0.937</td>
</tr>
<tr>
<td></td>
<td>81</td>
<td>25</td>
<td>Bias</td>
<td>-0.00031</td>
<td>0.00014</td>
<td>-0.00038</td>
<td>0.00526</td>
<td>0.00043</td>
<td>-0.00072</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.932</td>
<td>0.937</td>
<td>0.933</td>
<td>0.914</td>
<td>0.919</td>
<td>0.944</td>
</tr>
<tr>
<td></td>
<td>81</td>
<td>49</td>
<td>Bias</td>
<td>0.00013</td>
<td>0.00019</td>
<td>-0.00074</td>
<td>0.00488</td>
<td>0.00017</td>
<td>0.00058</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.944</td>
<td>0.935</td>
<td>0.955</td>
<td>0.937</td>
<td>0.929</td>
<td>0.927</td>
</tr>
<tr>
<td></td>
<td>81</td>
<td>81</td>
<td>Bias</td>
<td>-0.00037</td>
<td>-0.00017</td>
<td>0.00028</td>
<td>0.00692</td>
<td>0.00025</td>
<td>-0.00066</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.926</td>
<td>0.945</td>
<td>0.924</td>
<td>0.905</td>
<td>0.941</td>
<td>0.957</td>
</tr>
<tr>
<td>$\theta_0^b$</td>
<td>25</td>
<td>25</td>
<td>Bias</td>
<td>0.00238</td>
<td>-0.00173</td>
<td>-0.00118</td>
<td>0.02102</td>
<td>0.00194</td>
<td>-0.00087</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.915</td>
<td>0.920</td>
<td>0.923</td>
<td>0.857</td>
<td>0.929</td>
<td>0.921</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>49</td>
<td>Bias</td>
<td>0.00159</td>
<td>-0.00139</td>
<td>-0.00189</td>
<td>0.02106</td>
<td>-0.00003</td>
<td>-0.00001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.928</td>
<td>0.922</td>
<td>0.919</td>
<td>0.861</td>
<td>0.911</td>
<td>0.917</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>81</td>
<td>Bias</td>
<td>0.00253</td>
<td>-0.00038</td>
<td>-0.00042</td>
<td>0.02106</td>
<td>0.00101</td>
<td>-0.00032</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.921</td>
<td>0.922</td>
<td>0.917</td>
<td>0.837</td>
<td>0.933</td>
<td>0.926</td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>25</td>
<td>Bias</td>
<td>0.00108</td>
<td>-0.00141</td>
<td>-0.00039</td>
<td>0.00967</td>
<td>-0.00025</td>
<td>-0.00055</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.917</td>
<td>0.920</td>
<td>0.943</td>
<td>0.912</td>
<td>0.924</td>
<td>0.937</td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>49</td>
<td>Bias</td>
<td>0.00133</td>
<td>-0.00030</td>
<td>-0.00057</td>
<td>0.01255</td>
<td>-0.00037</td>
<td>0.00050</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.944</td>
<td>0.931</td>
<td>0.929</td>
<td>0.896</td>
<td>0.950</td>
<td>0.934</td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>81</td>
<td>Bias</td>
<td>0.00036</td>
<td>-0.00020</td>
<td>-0.00037</td>
<td>0.01064</td>
<td>0.00066</td>
<td>0.00100</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.944</td>
<td>0.936</td>
<td>0.951</td>
<td>0.902</td>
<td>0.945</td>
<td>0.946</td>
</tr>
<tr>
<td></td>
<td>81</td>
<td>25</td>
<td>Bias</td>
<td>0.00020</td>
<td>-0.00024</td>
<td>-0.00109</td>
<td>0.00587</td>
<td>0.00036</td>
<td>-0.00069</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.949</td>
<td>0.928</td>
<td>0.938</td>
<td>0.910</td>
<td>0.928</td>
<td>0.943</td>
</tr>
<tr>
<td></td>
<td>81</td>
<td>49</td>
<td>Bias</td>
<td>0.00081</td>
<td>0.00002</td>
<td>-0.00057</td>
<td>0.00482</td>
<td>0.00017</td>
<td>0.00078</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.934</td>
<td>0.945</td>
<td>0.952</td>
<td>0.920</td>
<td>0.931</td>
<td>0.929</td>
</tr>
<tr>
<td></td>
<td>81</td>
<td>81</td>
<td>Bias</td>
<td>0.00011</td>
<td>-0.00024</td>
<td>0.00003</td>
<td>0.00680</td>
<td>0.00019</td>
<td>-0.00000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.921</td>
<td>0.944</td>
<td>0.929</td>
<td>0.910</td>
<td>0.936</td>
<td>0.954</td>
</tr>
</tbody>
</table>

$\theta_0^a = (0.3, 0.3, 0.3, 0.3, 1, 1)$, $\theta_0^b = (-0.3, 0.3, 0.3, 0.3, 1, 1)$, and $\theta = (\lambda, \gamma, \rho, \alpha, \beta_1, \beta_2)$. $\vartheta = 1$. 

19
<table>
<thead>
<tr>
<th>$\theta_0$</th>
<th>$n$</th>
<th>$T$</th>
<th>$\theta$</th>
<th>$\hat{\lambda}$</th>
<th>$\hat{\gamma}$</th>
<th>$\hat{\rho}$</th>
<th>$\hat{\alpha}$</th>
<th>$\hat{\beta}_1$</th>
<th>$\hat{\beta}_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_0^a$</td>
<td>25</td>
<td>25</td>
<td>Bias</td>
<td>-0.00037</td>
<td>-0.00124</td>
<td>0.00010</td>
<td>0.00040</td>
<td>0.00177</td>
<td>-0.00136</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.921</td>
<td>0.928</td>
<td>0.910</td>
<td>0.905</td>
<td>0.931</td>
<td>0.919</td>
</tr>
<tr>
<td>25</td>
<td>49</td>
<td>Bias</td>
<td>0.00019</td>
<td>-0.00075</td>
<td>-0.00045</td>
<td>0.00159</td>
<td>-0.00023</td>
<td>-0.00039</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.920</td>
<td>0.929</td>
<td>0.936</td>
<td>0.919</td>
<td>0.919</td>
<td>0.920</td>
</tr>
<tr>
<td>25</td>
<td>81</td>
<td>Bias</td>
<td>0.00067</td>
<td>-0.00046</td>
<td>-0.00034</td>
<td>0.00199</td>
<td>0.00057</td>
<td>-0.00094</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.929</td>
<td>0.934</td>
<td>0.923</td>
<td>0.934</td>
<td>0.933</td>
<td>0.933</td>
</tr>
<tr>
<td>49</td>
<td>25</td>
<td>Bias</td>
<td>-0.00082</td>
<td>-0.00114</td>
<td>0.00097</td>
<td>-0.00026</td>
<td>-0.00016</td>
<td>-0.00078</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.928</td>
<td>0.931</td>
<td>0.936</td>
<td>0.925</td>
<td>0.922</td>
<td>0.941</td>
</tr>
<tr>
<td>49</td>
<td>49</td>
<td>Bias</td>
<td>0.00015</td>
<td>-0.00023</td>
<td>-0.00017</td>
<td>0.00258</td>
<td>-0.00053</td>
<td>0.00018</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.942</td>
<td>0.949</td>
<td>0.944</td>
<td>0.936</td>
<td>0.945</td>
<td>0.938</td>
</tr>
<tr>
<td>49</td>
<td>81</td>
<td>Bias</td>
<td>-0.00038</td>
<td>-0.00001</td>
<td>0.00003</td>
<td>0.00043</td>
<td>0.00070</td>
<td>0.00087</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.944</td>
<td>0.937</td>
<td>0.946</td>
<td>0.950</td>
<td>0.938</td>
<td>0.937</td>
</tr>
<tr>
<td>81</td>
<td>25</td>
<td>Bias</td>
<td>-0.00029</td>
<td>0.00015</td>
<td>-0.00042</td>
<td>-0.00077</td>
<td>0.00048</td>
<td>-0.00070</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.934</td>
<td>0.944</td>
<td>0.937</td>
<td>0.921</td>
<td>0.921</td>
<td>0.944</td>
</tr>
<tr>
<td>81</td>
<td>49</td>
<td>Bias</td>
<td>0.00015</td>
<td>0.00016</td>
<td>-0.00073</td>
<td>-0.00141</td>
<td>0.0001</td>
<td>0.00058</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.944</td>
<td>0.944</td>
<td>0.955</td>
<td>0.937</td>
<td>0.929</td>
<td>0.927</td>
</tr>
<tr>
<td>81</td>
<td>81</td>
<td>Bias</td>
<td>-0.00038</td>
<td>-0.00015</td>
<td>0.00027</td>
<td>0.00065</td>
<td>0.00025</td>
<td>-0.00006</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.926</td>
<td>0.947</td>
<td>0.925</td>
<td>0.938</td>
<td>0.942</td>
<td>0.957</td>
</tr>
<tr>
<td>$\theta_0^b$</td>
<td>25</td>
<td>25</td>
<td>Bias</td>
<td>0.00134</td>
<td>-0.00152</td>
<td>-0.00089</td>
<td>0.00276</td>
<td>0.00181</td>
<td>-0.00113</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.917</td>
<td>0.925</td>
<td>0.921</td>
<td>0.905</td>
<td>0.928</td>
<td>0.920</td>
</tr>
<tr>
<td>25</td>
<td>49</td>
<td>Bias</td>
<td>0.00074</td>
<td>-0.00130</td>
<td>-0.00168</td>
<td>0.00294</td>
<td>-0.00018</td>
<td>-0.00024</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.934</td>
<td>0.930</td>
<td>0.920</td>
<td>0.928</td>
<td>0.913</td>
<td>0.919</td>
</tr>
<tr>
<td>25</td>
<td>81</td>
<td>Bias</td>
<td>0.00150</td>
<td>-0.00026</td>
<td>-0.00013</td>
<td>0.00234</td>
<td>0.00083</td>
<td>-0.00060</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.928</td>
<td>0.929</td>
<td>0.915</td>
<td>0.935</td>
<td>0.932</td>
<td>0.929</td>
</tr>
<tr>
<td>49</td>
<td>25</td>
<td>Bias</td>
<td>0.00064</td>
<td>-0.00125</td>
<td>-0.00026</td>
<td>0.00028</td>
<td>-0.00029</td>
<td>-0.00065</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.919</td>
<td>0.930</td>
<td>0.942</td>
<td>0.925</td>
<td>0.926</td>
<td>0.937</td>
</tr>
<tr>
<td>49</td>
<td>49</td>
<td>Bias</td>
<td>0.00083</td>
<td>-0.00026</td>
<td>-0.00044</td>
<td>0.00285</td>
<td>-0.00045</td>
<td>0.00037</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.945</td>
<td>0.934</td>
<td>0.930</td>
<td>0.935</td>
<td>0.951</td>
<td>0.933</td>
</tr>
<tr>
<td>49</td>
<td>81</td>
<td>Bias</td>
<td>-0.00011</td>
<td>-0.00015</td>
<td>-0.00024</td>
<td>0.00065</td>
<td>0.00058</td>
<td>0.00087</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.944</td>
<td>0.941</td>
<td>0.952</td>
<td>0.950</td>
<td>0.945</td>
<td>0.944</td>
</tr>
<tr>
<td>81</td>
<td>25</td>
<td>Bias</td>
<td>-0.00001</td>
<td>-0.00020</td>
<td>-0.00105</td>
<td>0.00003</td>
<td>0.00038</td>
<td>-0.00072</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.949</td>
<td>0.933</td>
<td>0.938</td>
<td>0.920</td>
<td>0.930</td>
<td>0.944</td>
</tr>
<tr>
<td>81</td>
<td>49</td>
<td>Bias</td>
<td>0.00057</td>
<td>0.00002</td>
<td>-0.00051</td>
<td>-0.00124</td>
<td>0.00014</td>
<td>0.00072</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.938</td>
<td>0.950</td>
<td>0.954</td>
<td>0.945</td>
<td>0.930</td>
<td>0.930</td>
</tr>
<tr>
<td>81</td>
<td>81</td>
<td>Bias</td>
<td>-0.00015</td>
<td>-0.00020</td>
<td>0.00011</td>
<td>0.00075</td>
<td>0.00014</td>
<td>-0.00007</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.921</td>
<td>0.944</td>
<td>0.929</td>
<td>0.938</td>
<td>0.937</td>
<td>0.953</td>
</tr>
</tbody>
</table>

$\theta_0^a = (0.3, 0.3, 0.3, 0.3, 1, 1)$, $\theta_0^b = (-0.3, 0.3, 0.3, 0.3, 1, 1)$, and $\theta = (\lambda, \gamma, \rho, \alpha, \beta_1, \beta_2)$. $\varrho = 1$. The bias-corrected estimator is from Theorem 4.
### Table 3: Performance of the Bias Corrected Estimator When the Number of Factors is Overspecified by 1

<table>
<thead>
<tr>
<th>$\theta_0^a$</th>
<th>$n$</th>
<th>$T$</th>
<th>$\hat{\lambda}$</th>
<th>$\hat{\gamma}$</th>
<th>$\hat{\rho}$</th>
<th>$\hat{\alpha}$</th>
<th>$\hat{\beta}_1$</th>
<th>$\hat{\beta}_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>25 25 Bias</td>
<td>0.00001</td>
<td>-0.00178</td>
<td>0.00040</td>
<td>0.00327</td>
<td>0.00221</td>
<td>-0.00208</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CP</td>
<td>0.875</td>
<td>0.882</td>
<td>0.875</td>
<td>0.829</td>
<td>0.894</td>
<td>0.874</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 49 Bias</td>
<td>0.00054</td>
<td>-0.00088</td>
<td>-0.0068</td>
<td>0.00597</td>
<td>-0.0027</td>
<td>0.00013</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CP</td>
<td>0.896</td>
<td>0.890</td>
<td>0.903</td>
<td>0.872</td>
<td>0.879</td>
<td>0.891</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 81 Bias</td>
<td>0.00062</td>
<td>-0.00051</td>
<td>-0.003</td>
<td>0.00692</td>
<td>0.00061</td>
<td>-0.00073</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CP</td>
<td>0.904</td>
<td>0.914</td>
<td>0.907</td>
<td>0.863</td>
<td>0.912</td>
<td>0.907</td>
<td></td>
<td></td>
</tr>
<tr>
<td>49 25 Bias</td>
<td>-0.00025</td>
<td>-0.00125</td>
<td>0.00043</td>
<td>-0.00111</td>
<td>-0.00030</td>
<td>-0.00115</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CP</td>
<td>0.891</td>
<td>0.905</td>
<td>0.892</td>
<td>0.880</td>
<td>0.900</td>
<td>0.902</td>
<td></td>
<td></td>
</tr>
<tr>
<td>49 49 Bias</td>
<td>0.00015</td>
<td>-0.00038</td>
<td>-0.00063</td>
<td>0.00425</td>
<td>-0.00089</td>
<td>0.00033</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CP</td>
<td>0.930</td>
<td>0.917</td>
<td>0.911</td>
<td>0.902</td>
<td>0.929</td>
<td>0.916</td>
<td></td>
<td></td>
</tr>
<tr>
<td>49 81 Bias</td>
<td>-0.00026</td>
<td>-0.00001</td>
<td>-0.00099</td>
<td>0.00165</td>
<td>0.00073</td>
<td>0.00077</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CP</td>
<td>0.928</td>
<td>0.915</td>
<td>0.935</td>
<td>0.919</td>
<td>0.924</td>
<td>0.933</td>
<td></td>
<td></td>
</tr>
<tr>
<td>81 25 Bias</td>
<td>-0.00025</td>
<td>0.00000</td>
<td>-0.00043</td>
<td>-0.00082</td>
<td>0.00030</td>
<td>-0.00096</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CP</td>
<td>0.909</td>
<td>0.916</td>
<td>0.903</td>
<td>0.885</td>
<td>0.902</td>
<td>0.928</td>
<td></td>
<td></td>
</tr>
<tr>
<td>81 49 Bias</td>
<td>0.00020</td>
<td>0.00018</td>
<td>-0.00081</td>
<td>-0.00115</td>
<td>0.00022</td>
<td>0.00056</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CP</td>
<td>0.936</td>
<td>0.929</td>
<td>0.941</td>
<td>0.935</td>
<td>0.919</td>
<td>0.918</td>
<td></td>
<td></td>
</tr>
<tr>
<td>81 81 Bias</td>
<td>-0.00036</td>
<td>-0.00019</td>
<td>0.00028</td>
<td>0.00081</td>
<td>0.00025</td>
<td>-0.00018</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CP</td>
<td>0.919</td>
<td>0.937</td>
<td>0.908</td>
<td>0.914</td>
<td>0.929</td>
<td>0.947</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\theta_0^b$</td>
<td>25 25 Bias</td>
<td>0.00322</td>
<td>-0.00240</td>
<td>-0.00164</td>
<td>0.00509</td>
<td>0.00270</td>
<td>-0.00157</td>
<td></td>
</tr>
<tr>
<td>CP</td>
<td>0.869</td>
<td>0.898</td>
<td>0.882</td>
<td>0.829</td>
<td>0.882</td>
<td>0.889</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 49 Bias</td>
<td>0.00167</td>
<td>-0.00156</td>
<td>-0.00218</td>
<td>0.00738</td>
<td>-0.00004</td>
<td>0.00037</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CP</td>
<td>0.901</td>
<td>0.896</td>
<td>0.898</td>
<td>0.879</td>
<td>0.880</td>
<td>0.892</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 81 Bias</td>
<td>0.00212</td>
<td>-0.00041</td>
<td>-0.00031</td>
<td>0.00643</td>
<td>0.00093</td>
<td>-0.00028</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CP</td>
<td>0.905</td>
<td>0.916</td>
<td>0.897</td>
<td>0.855</td>
<td>0.913</td>
<td>0.903</td>
<td></td>
<td></td>
</tr>
<tr>
<td>49 25 Bias</td>
<td>0.00138</td>
<td>-0.00143</td>
<td>-0.00064</td>
<td>-0.00027</td>
<td>-0.00026</td>
<td>-0.00097</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CP</td>
<td>0.902</td>
<td>0.904</td>
<td>0.913</td>
<td>0.889</td>
<td>0.898</td>
<td>0.911</td>
<td></td>
<td></td>
</tr>
<tr>
<td>49 49 Bias</td>
<td>0.00091</td>
<td>-0.00042</td>
<td>-0.00042</td>
<td>0.00458</td>
<td>-0.00079</td>
<td>0.00054</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CP</td>
<td>0.934</td>
<td>0.922</td>
<td>0.924</td>
<td>0.909</td>
<td>0.932</td>
<td>0.918</td>
<td></td>
<td></td>
</tr>
<tr>
<td>49 81 Bias</td>
<td>0.00011</td>
<td>-0.00019</td>
<td>-0.00041</td>
<td>0.00190</td>
<td>0.00064</td>
<td>0.00080</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CP</td>
<td>0.937</td>
<td>0.926</td>
<td>0.934</td>
<td>0.918</td>
<td>0.925</td>
<td>0.936</td>
<td></td>
<td></td>
</tr>
<tr>
<td>81 25 Bias</td>
<td>0.00005</td>
<td>-0.00035</td>
<td>-0.00112</td>
<td>0.00004</td>
<td>0.00017</td>
<td>-0.00101</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CP</td>
<td>0.917</td>
<td>0.909</td>
<td>0.913</td>
<td>0.889</td>
<td>0.903</td>
<td>0.924</td>
<td></td>
<td></td>
</tr>
<tr>
<td>81 49 Bias</td>
<td>0.00051</td>
<td>0.00005</td>
<td>-0.00058</td>
<td>-0.00080</td>
<td>0.00017</td>
<td>0.00081</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CP</td>
<td>0.922</td>
<td>0.927</td>
<td>0.934</td>
<td>0.935</td>
<td>0.919</td>
<td>0.918</td>
<td></td>
<td></td>
</tr>
<tr>
<td>81 81 Bias</td>
<td>-0.00014</td>
<td>-0.00023</td>
<td>0.00010</td>
<td>0.00093</td>
<td>0.00015</td>
<td>-0.00019</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CP</td>
<td>0.919</td>
<td>0.933</td>
<td>0.923</td>
<td>0.917</td>
<td>0.931</td>
<td>0.951</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$\theta_0^a = (0.3, 0.3, 0.3, 0.3, 1, 1)$, $\theta_0^b = (-0.3, 0.3, 0.3, 0.3, 1, 1)$, and $\theta = (\lambda, \gamma, \rho, \alpha, \beta_1, \beta_2)$. $\vartheta = 1$. The DGP is the same as described in the text. The true number of factors is 2 and the estimation assumes 3 factors. $\hat{\theta}$ is the bias corrected QML estimator assuming 3 factors.
Figure 1: Frequencies of Incorrect Estimation

True parameter values: $\theta_0^d = (0.3, 0.3, 0.3, 0.3, 1, 1)$, where $\theta = (\lambda, \gamma, \rho, \alpha, \beta_1, \beta_2)$. $\vartheta = 1$. True number of factors: 2. Initial estimates assume 10 factors in both equations.
insured HECMs concentrate disproportionately in areas that more likely see house price declines.

Observing that the origination rates exhibit spatial clustering, it is of interest to quantify the spatial spillover effect. If spatial effects are present, the HECM activity in a state can be affected by developments in the neighboring states. Our data covers 51 states and 52 quarters from 2001 to 2013. Let $y_{it}$ denote the HECM origination rate, defined as the number of newly originated HECM loans in state $i$ at quarter $t$ as a percentage of the senior population (age 65 plus) in state $i$ from the 2010 census. The $n \times n$ spatial weights matrices is $W_n$ and $W_{n,ij} = 1$ if states $i$ and $j$ share the same border and $W_{1,ij} = 0$ otherwise. House price dynamic variables are constructed using the Federal Housing Finance Agency’s quarterly all-transactions house price indexes (HPI) deflated by the CPI, and include deviations from the previous 9 year averages ($hpi_{dev}$), standard deviations of house price changes in the previous 9 years ($hpi_v$) and the interaction between the two. Figures 2 and 3 show the averages of these variables by U.S. regions in our sample period. It is likely that the origination rates are affected by some macroeconomic factors which are captured by
Table 4: Estimation of State-Level Origination Rates

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contemporaneous Spatial Effect $\lambda$</td>
<td>-0.05527*** 0.01426</td>
</tr>
<tr>
<td>Own Time Lag $\gamma$</td>
<td>0.68981*** 0.01263</td>
</tr>
<tr>
<td>Spatial Diffusion $\rho$</td>
<td>0.05405*** 0.00998</td>
</tr>
<tr>
<td>Spatial Effect in Disturbances $\alpha$</td>
<td>0.12180*** 0.00808</td>
</tr>
<tr>
<td>House Price Deviation $\beta_1$</td>
<td>-0.00025 0.00055</td>
</tr>
<tr>
<td>House Price Volatility $\beta_2$</td>
<td>0.01346*** 0.00153</td>
</tr>
<tr>
<td>Deviation $\times$ Volatility $\beta_3$</td>
<td>0.03203*** 0.00847</td>
</tr>
</tbody>
</table>

The sample size is $n = 51$ and $T = 52$. Bias corrected estimates are reported. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. $\hat{\sigma} = 0.0063$.

The results reveal interesting spatial patterns. Higher HECM activity in a state negatively influences neighboring states ($\lambda$). This is consistent with the hypothesis that lenders shift resources towards states with higher activity, resulting in lower activity in the neighboring states. On the other hand, spatial diffusion ($\rho$) and spatial error ($\alpha$) have positive effects, reflecting spatially correlated demand and supply effects. The own time lag ($\gamma$) has positive effect, capturing serially correlated effects. In addition, the results show that states with high past house price volatilities and current house prices above long term averages have higher origination rates, consistent with the findings of Haurin et al. (2014).

6. Conclusion

Dynamic spatial panels with interactive effects are of practical interest. Outcomes of spatial units are correlated due to spatial interactions and common factors. The model under consideration has a rich spatial structure, which includes contemporaneous spatial interaction, spatial diffusion and spatial disturbances. Unobserved interactive effects of individual loadings and time factors account for additional cross sectional
dependence and may correlate with observed regressors. This paper shows that the QML method provides consistent and asymptotically normal estimators. There are asymptotic biases arising from the predeter-
mined regressors and the interaction between the spatial effects and the factor loadings, but bias correction is possible. The Monte Carlo study shows that the proposed bias correction is effective in reducing bias. Consistency of estimators may still hold when the number of factors is overspecified. There are various criteria which can determine the number of factors in a sample. An application of the model to reverse mortgage originations reveals interesting spatial patterns.

7. References


_ , Young H. Lee, and Peter Schmidt, “Panel Data Models with Multiple Time-Varying Individual Ef-


Appendix A. Some Matrix Algebra and Useful Lemmas

This section lists some results in matrix algebra used frequently throughout the paper. All matrices are real. Lemma 1 on the uniform boundedness of some matrices is in Lee (2004b).

Lemma 1. Supposing that $$\|W_n\|$$ and $$\|S_n^{-1}\|$$ are uniformly bounded for some matrix norm $$\|\cdot\|$$, where $$S_n = I_n - \lambda_0 W_n$$. Then $$\|S_n(\lambda)^{-1}\|$$ is uniformly bounded for $$|\lambda - \lambda_0| < \frac{1}{2}$$, where $$S_n(\lambda) = I_n - \lambda W_n$$ and $$c = \max \left( \limsup_n \|W_n\|, \limsup_n \|S_n^{-1}\| \right)$$.

The following results on matrix norms are standard, see Bernstein (2009). Let $$\mu_i(M)$$ denote the $$i$$-th largest eigenvalue of a symmetric matrix $$M$$, $$\mu_n(M) \leq \mu_{n-1}(M) \leq \cdots \leq \mu_1(M)$$.

Lemma 2. A is a $$n \times m$$ matrix, B is a $$m \times p$$ matrix, and C and D are $$n \times n$$ matrices. Then

1. $$\|A\|_2 \leq \|A\|_F \leq \|A\|_2 \text{rank}(A)^{\frac{1}{2}}$$; and $$\|A\|_2^2 \leq \|A\|_1 \|A\|_\infty$$;
2. $$\|AB\|_F \leq \|A\|_F \|B\|_2 \leq \|A\|_F \|B\|_F$$; and $$\|\text{tr}(AB)\| \leq \|A\|_F \|B\|_F$$, if $$n = p$$;
3. $$\|C\| \leq \text{rank}(C)\|C\|$$, where $$\|\cdot\|$$ is an induced matrix norm;
4. $$\mu_i(C)\mu_1(D) \geq \mu_i(CD) \geq \mu_i(C)\mu_i(D)$$, where $$C, D$$ are $$n \times n$$ symmetric positive semidefinite matrices.

We have following results on matrix norms for the $$n \times T$$ dimensional explanatory variables, of which proofs are in the supplementary file.

Lemma 3. Under Assumptions E and R, $$\|Z_k\|_2 = O_P\left(\sqrt{nT}\right)$$, $$\|Z_k\|_F = O_P(\sqrt{nT})$$, and $$\text{tr}\left(\left(A_n Z_k\right)' \varepsilon\right) = O_P(\sqrt{nT})$$ for $$k = 1, \cdots, K + 1$$, where $$A_n$$ is an $$n \times n$$ nonstochastic UB matrix.

Lemma 4. Define $$\tilde{Z}_k = \mathbb{E}(Z_k | \mathcal{G}_{nT})$$ where the conditioning set $$\mathcal{G}_{nT}$$ is the sigma algebra generated by $$X_{n1}, \cdots, X_{nT}$$ and $$\tilde{\Gamma}_n$$ and $$F_T$$. Under Assumptions E, R and SF, we have

1. $$\left\|P_{\tilde{\Gamma}_n F_T}\right\|_2 = O_P\left(\frac{1}{\sqrt{nT}}\right)$$, where $$P_{\tilde{\Gamma}_n F_T} = \tilde{\Gamma}_n (\tilde{\Gamma}_n'\tilde{\Gamma}_n)^{-1} (\tilde{F}_T'\tilde{F}_T)^{-1} \tilde{F}_T'$$; and $$\left\|P_{\tilde{\Gamma}_n \varepsilon P_{F_T}}\right\|_2 = O_P(1)$$;
2. For $$k = 1, \cdots, K + 1$$, $$\|Z_k - \tilde{Z}_k\|_2 = O_P\left( \max\left(\sqrt{n}, \sqrt{T}\right) \right)$$; and $$\|Z_k'\tilde{\Gamma}_n\|_2 = O_P\left( \max\left( n, T \right) \right)$$.

Lemma 5. Under Assumptions E, R and SF, $$\frac{1}{\sqrt{nT}} \left[ \text{tr}\left( Q_n (Z_k - \tilde{Z}_k) P_{F_T} \varepsilon' \right) - \mathbb{E}\left( \text{tr}\left( Q_n (Z_k - \tilde{Z}_k) P_{F_T} \varepsilon' \right) | \mathcal{G}_{nT} \right) \right] = O_P\left( \frac{1}{\sqrt{T}} \right)$$ for all $$k = 1, \cdots, K + 1$$, where $$Q_n$$ is any $$n \times n$$ UB matrix.

Lemma 6. Under Assumptions E, R and SF, we have

1. $$\frac{1}{\sqrt{nT}} \left( \text{tr}\left( P_{F_T} \varepsilon' \varepsilon' \right) - \sigma_0^2 T \sigma_0^2 \right) = O_P\left( \frac{1}{\sqrt{n}} \right)$$;
2. $$\frac{1}{\sqrt{nT}} \left( \text{tr}\left( P_{F_T} R_n G_n R_n^{-1} \varepsilon' \varepsilon' \right) - \sigma_0^2 T \text{tr}\left( P_{F_T} R_n G_n R_n^{-1} \right) \right) = O_P \left( \frac{1}{\sqrt{n}} \right)$$.
\begin{align*}
\frac{1}{\sqrt{nT}} \left( \text{tr} \left( R_n G_n R_n^{-1} P_{nT} \varepsilon \varepsilon' \right) - \sigma_n^2 T \text{tr} \left( P_{nT} R_n G_n R_n^{-1} \right) \right) &= O_P \left( \frac{1}{\sqrt{n}} \right) \quad \text{and} \\
\frac{1}{\sqrt{nT}} \left( \text{tr} \left( P_{nT} R_n G_n R_n^{-1} P_{nT} \varepsilon \varepsilon' \right) - \sigma_n^2 T \text{tr} \left( P_{nT} R_n G_n R_n^{-1} \right) \right) &= O_P \left( \frac{1}{\sqrt{n}} \right) .
\end{align*}

(3) \begin{align*}
\frac{1}{\sqrt{nT}} \left( \text{tr} \left( P_{nT} \tilde{G}_n \varepsilon \varepsilon' \right) - \sigma_n^2 T \text{tr} \left( P_{nT} \tilde{G}_n \right) \right) &= O_P \left( \frac{1}{\sqrt{n}} \right) ; \\
\frac{1}{\sqrt{nT}} \left( \text{tr} \left( \tilde{G}_n P_{nT} \varepsilon \varepsilon' \right) - \sigma_n^2 T \text{tr} \left( P_{nT} \tilde{G}_n \right) \right) &= O_P \left( \frac{1}{\sqrt{n}} \right) ; \\
\text{and} \quad \frac{1}{\sqrt{nT}} \left( \text{tr} \left( P_{nT} \tilde{G}_n \varepsilon \varepsilon' \right) - \sigma_n^2 T \text{tr} \left( P_{nT} \tilde{G}_n \right) \right) &= O_P \left( \frac{1}{\sqrt{n}} \right) .
\end{align*}

Appendix B. Identification and Consistency: Proof of Propositions 1-3
Proof of Proposition 1.

Dropping constant terms, the expected objective function is equivalent to
\begin{align*}
\mathcal{Q}_{nT} (\theta, \tilde{\Gamma}_n, F_T) &= \frac{1}{n} \log |S_n(\lambda) R_n(\alpha)| \\
- \frac{1}{2} \log \left( \mathbb{E} \frac{1}{n T} \text{tr} \left[ \left( R_n(\alpha) \left( S_n(\lambda) Y - \sum_{k=1}^{K} Z_k \delta_k \right) - \tilde{\Gamma}_n F_T' \right) \left( R_n(\alpha) \left( S_n(\lambda) Y - \sum_{k=1}^{K} Z_k \delta_k \right) - \tilde{\Gamma}_n F_T' \right) \right] \right),
\end{align*}
with \( \tilde{\Gamma}_n = R_n(\alpha) \Gamma_n, \tilde{\Gamma}_n \in \mathbb{R}^{n \times r} \), because no restriction is imposed on \( \Gamma_n \) and \( R_n(\alpha) \) is invertible for \( \alpha \in \Theta_\alpha \).

Denote \( \tilde{\Gamma}_0 = R_0 \Gamma_0 \). Thus, \( \mathcal{Q}_{nT} (\theta, \tilde{\Gamma}_0, F_{0T}) = \frac{1}{n} \log |S_n R_n| - \frac{1}{2} \log \sigma_0^2 \). Substituting in the DGP of \( Y \),
\begin{align*}
R_n(\alpha) \left( S_n(\lambda) Y - \sum_{k=1}^{K} Z_k \delta_k \right) - \tilde{\Gamma}_n F_T' = R_n(\alpha) \left( S_n(\lambda) S_n^{-1} \Gamma_0 F_{0T} + S_n(\lambda) S_n^{-1} U + \sum_{k=1}^{K+1} Z_k (\theta_{0k} - \theta_k) \right) - \tilde{\Gamma}_n F_T',
\end{align*}
where for simplicity, \( \theta_k = \delta_k \), for \( k = 1, \cdots, K \), \( \theta_{K+1} = \lambda \), and \( Z_{K+1} = G_n \sum_{k=1}^{K} Z_k \theta_{0k} \). Under Assumption E, as the idiosyncratic disturbances are independent from the interactive effects and regressors,
\begin{align*}
\mathcal{Q}_{nT} (\theta, \tilde{\Gamma}_n, F_T) \\
\leq \frac{1}{n} \log |S_n(\lambda) R_n(\alpha)| - \frac{1}{2} \log \left( \frac{\sigma_n^2}{n} \text{tr} \left( R_n(\alpha) S_n(\lambda) S_n^{-1} R_n^{-1} R_n^{-1} \right) \right) + \\
\mathbb{E} \frac{1}{n T} \text{tr} \left[ M_{\Gamma_n} R_n(\alpha) \left( S_n(\lambda) S_n^{-1} \Gamma_0 F_{0T} + \sum_{k=1}^{K+1} Z_k (\theta_{0k} - \theta_k) \right) \left( S_n(\lambda) S_n^{-1} \Gamma_0 F_{0T} + \sum_{k=1}^{K+1} Z_k (\theta_{0k} - \theta_k) \right)' R_n(\alpha)' M_{\Gamma_n} \right] \\
\leq \frac{1}{n} \log |S_n(\lambda) R_n(\alpha)| - \frac{1}{2} \log \left( \frac{\sigma_n^2}{n} \text{tr} \left( R_n(\alpha) S_n(\lambda) S_n^{-1} R_n^{-1} R_n^{-1} \right) \right) + \\
\mathbb{E} \frac{1}{n T} \text{tr} \left[ M_{\Gamma_n} R_n(\alpha) \left( \sum_{k=1}^{K+1} Z_k (\theta_{0k} - \theta_k) \right) M_{\Gamma_n} \right] \\
\leq \frac{1}{n} \log |S_n(\lambda) R_n(\alpha)| - \frac{1}{2} \log \left( \frac{\sigma_n^2}{n} \text{tr} \left( R_n(\alpha) S_n(\lambda) S_n^{-1} R_n^{-1} R_n^{-1} \right) \right) + \\
\mathbb{E} \sum_{k_1, k_2=1}^{K+1} \text{vec} \left( Z_{k_1} \right)' \left( M_{\Gamma_n} \otimes R_n(\alpha)' M_{\Gamma_n} R_n(\alpha) \right) \text{vec} \left( Z_{k_2} \right) \left( \theta_{0k_1} - \theta_{k_1} \right) \left( \theta_{0k_2} - \theta_{k_2} \right).
\end{align*}

- Case 1: Assumption ID1 holds.
If $\delta \neq \delta_0$ or $\lambda \neq \lambda_0$, because $E[z' (M_F n \otimes R_n(\alpha)' M_{F_n} R_n(\alpha)) z]$ with $z = \left( \text{vec} \,(Z_1) \right.$ $\cdots$ $\text{vec} \,(Z_{K+1}) \big)$ is positive definite,

$$\mathcal{D}_{nT}(\theta, \Gamma_n, F_T) < \frac{1}{n} \log |S_n(\lambda)R_n(\alpha)| - \frac{1}{2} \log \left( \frac{\sigma^2_0}{n} \text{tr} (R_n(\alpha) S_n(\lambda) S_n^{-1} R_n^{-1} R_n^{-1} S_n^{-1} S_n(\lambda)' R_n(\alpha)') \right)$$

$$\leq \frac{1}{n} \log |S_n(\lambda)R_n(\alpha)| - \frac{1}{2} \log \left( \sigma^2_0 |R_n(\alpha) S_n(\lambda) S_n^{-1} R_n^{-1} R_n^{-1} S_n^{-1} S_n(\lambda)' R_n(\alpha)'| \frac{1}{2} \right)$$

$$= - \frac{1}{2} \log \sigma^2_0 + \frac{1}{n} \log |S_n R_n| = \mathcal{D}_{nT}(\theta_0, \hat{\Gamma}_0, F_{0T}).$$

Therefore $\delta_0$ and $\lambda_0$ are identified. At $\delta = \delta_0$ and $\lambda = \lambda_0$,

$$\mathcal{D}_{nT}(\theta, \Gamma_n, F_T) \leq \frac{1}{n} \log |S_n| + \frac{1}{n} \log |R_n(\alpha)| - \frac{1}{2} \log \left( \sigma^2_0 |R_n(\alpha) R_n^{-1} R_n^{-1} | \frac{1}{2} \right) = \mathcal{D}_{nT}(\theta_0, \hat{\Gamma}_0, F_{0T}).$$

If $\alpha \neq \alpha_0$, this inequality will be strict by Assumption ID1(2). Therefore $\alpha_0$ is identified.

- Case 2: Assumption ID2 holds.

Because $z' (M_F n \otimes R_n(\alpha)' M_{F_n} R_n(\alpha)) z$ is positive semi-definite,

$$\mathcal{D}_{nT}(\theta, \Gamma_n, F_T) \leq \frac{1}{n} \log |S_n(\lambda)R_n(\alpha)| - \frac{1}{2} \log \left( \sigma^2_0 |R_n(\alpha) S_n(\lambda) S_n^{-1} R_n^{-1} R_n^{-1} S_n^{-1} S_n(\lambda)' R_n(\alpha)'| \frac{1}{2} \right)$$

$$= - \frac{1}{2} \log \sigma^2_0 + \frac{1}{n} \log |S_n R_n| = \mathcal{D}_{nT}(\theta_0, \hat{\Gamma}_0, F_{0T}).$$

$\lambda$ and $\alpha$ are identified, because this inequality holds strictly for $\lambda \neq \lambda_0$ or $\alpha \neq \alpha_0$ under Assumption ID2(2).

With $\lambda = \lambda_0$, $\delta$ is then identified, because this inequality is also strict for $\delta \neq \delta_0$, due to $E \left( z' (M_F n \otimes R_n(\alpha)' M_{F_n} R_n(\alpha)) z \right)$ with $z = \left( \text{vec} \,(Z_1) \right.$ $\cdots$ $\text{vec} \,(Z_{K+1}) \big)$ being positive definite under Assumption ID2(1).

Therefore $\delta_0$, $\lambda_0$ and $\alpha_0$ are identified under either Assumption ID1 or ID2. With $\delta = \delta_0$, $\lambda = \lambda_0$, and $\alpha = \alpha_0$, $\mathcal{D}_{nT}(\theta_0, \Gamma_n, F_T) = \frac{1}{n} \log |S_n R_n| - \frac{1}{2} \log \left( \sigma^2_0 + \frac{1}{n^2} \text{tr} \left[ \left( \hat{F}_{0T}' \hat{F}'_T - F_T' \hat{F}_T - \hat{F}_T' F_T \right) \right] \right)$, which is strictly less than $\mathcal{D}_{nT}(\theta_0, \hat{\Gamma}_0, F_{0T})$ unless $F_T' = \hat{F}_T$. As a result, the number of factors $r_0$ is identified from the rank of $\hat{F}_{0T}' = R_0 \Gamma_0 F_{0T}'$.

Proof of Proposition 2.

Instead of considering the positive eigenvalues of $\frac{1}{n^2} R_n(\alpha)(\eta \cdot Z)(\eta \cdot Z)' R_n(\alpha)'$, one may consider the relevant eigenvalues of $\frac{1}{n^2} (\eta \cdot Z)' R_n(\alpha)' R_n(\alpha)(\eta \cdot Z)$. Because these two matrices have the same nonzero eigenvalues, counting multiplicity.\(^{12}\) There are $T$ eigenvalues of $\frac{1}{n^2} (\eta \cdot Z)' R_n(\alpha)' R_n(\alpha)(\eta \cdot Z) = \frac{1}{n^2} (I_T \otimes \eta)' \mathcal{Z}_{nT}(\alpha)' R_n(\alpha) \cdot \mathcal{Z}_{nT}(I_T \otimes \eta)$. Because $\eta \in B_{K+1}$, $(I_T \otimes \eta)' (I_T \otimes \eta) = ||\eta||^2 I_T = I_T$. By the Poincaré

\(^{12}\)See Theorem 2.8 in Zhang (2011).
interlacing theorem, \( \mu_{1+KT}(\frac{1}{nT}Z'_{nT}R_n(\alpha)'R_n(\alpha)Z_{nT}) \leq \mu_i(\frac{1}{nT}(I_T \otimes \eta')Z'_{nT}R_n(\alpha)'R_n(\alpha)Z_{nT}(I_T \otimes \eta)) \), for \( i = 1, \cdots, T \). Because \( \frac{1}{nT}Z'_{nT}R_n(\alpha)'R_n(\alpha)Z_{nT} \) is of dimension \((T + KT) \times (T + KT)\), its \( T \) smallest eigenvalues (including multiplicity) are less than or equal to the \( T \) corresponding eigenvalues of \( \frac{1}{nT}(I_T \otimes \eta')Z'_{nT}R_n(\alpha)'R_n(\alpha)Z_{nT}(I_T \otimes \eta) \) uniformly in \( \eta \in B_{K+1} \). A way to see this is the following:

1. For each \( \eta \), let \( n_\eta \) be the number of positive eigenvalues of \( \frac{1}{nT}R_n(\alpha)(\eta \cdot Z)(\eta \cdot Z)'R_n(\alpha)' \). Suppose \( n_\eta \geq 2r + 1 \), then \( \sum_{i=2r+1}^{n_\eta} \mu_i(\frac{1}{nT}R_n(\alpha)(\eta \cdot Z)(\eta \cdot Z)'R_n(\alpha)') \leq \sum_{i=2r+1}^{n_\eta} \mu_i(\frac{1}{nT}R_n(\alpha)(\eta \cdot Z)(\eta \cdot Z)'R_n(\alpha)') \).

2. There are \( n_\eta \) positive eigenvalues of \( \frac{1}{nT}(\eta \cdot Z)'R_n(\alpha)'R_n(\alpha)(\eta \cdot Z) \) and \( \mu_i(\frac{1}{nT}(\eta \cdot Z)'R_n(\alpha)'R_n(\alpha)(\eta \cdot Z)) \) for \( i = 1, \cdots, n_\eta \). The remaining \( T - n_\eta \) eigenvalues of \( \frac{1}{nT}(\eta \cdot Z)'R_n(\alpha)'R_n(\alpha)(\eta \cdot Z) \) are zero. Note that it is necessary that \( n_\eta \leq \min\{n, T\} \).

3. Because \( \mu_{1+KT}(\frac{1}{nT}Z'_{nT}R_n(\alpha)'R_n(\alpha)Z_{nT}) \leq \mu_i(\frac{1}{nT}(I_T \otimes \eta')Z'_{nT}R_n(\alpha)'R_n(\alpha)Z_{nT}(I_T \otimes \eta)) = \mu_i(\frac{1}{nT}(\eta \cdot Z)'R_n(\alpha)'R_n(\alpha)(\eta \cdot Z)) \) for \( i = 2r + 1, \cdots, n_\eta \), it follows that

\[
\sum_{i=2r+1}^{T+KT} \mu_i(\frac{1}{nT}Z'_{nT}R_n(\alpha)'R_n(\alpha)Z_{nT}) \leq \sum_{i=2r+1}^{n_\eta} \mu_i(\frac{1}{nT}(\eta \cdot Z)'R_n(\alpha)'R_n(\alpha)(\eta \cdot Z)) = \sum_{i=2r+1}^{n_\eta} \mu_i(\frac{1}{nT}(\eta \cdot Z)'R_n(\alpha)'R_n(\alpha)(\eta \cdot Z)).
\]

Because \( \frac{1}{nT}Z'_{nT}R_n(\alpha)'R_n(\alpha)Z_{nT} \) and \( \frac{1}{nT}Z'_{nT}Z_{nT}R_n(\alpha)'R_n(\alpha) \) have the same positive eigenvalues, counting multiplicity,

\[
\sum_{i=2r+1}^{T+KT} \mu_i(\frac{1}{nT}Z'_{nT}R_n(\alpha)'R_n(\alpha)Z_{nT}) = \sum_{i=2r+1}^{n_\eta} \mu_i(\frac{1}{nT}Z'_{nT}Z_{nT}R_n(\alpha)'R_n(\alpha)) \geq \sum_{i=2r+1}^{n_\eta} \mu_i(\frac{1}{nT}Z'_{nT}Z_{nT}R_n(\alpha)'R_n(\alpha)) = \sum_{i=2r+1}^{n_\eta} \mu_i(\frac{1}{nT}(\eta \cdot Z)'R_n(\alpha)'R_n(\alpha)(\eta \cdot Z)),
\]

where the last inequality is due to Lemma 2(4). Therefore, a set of sufficient conditions for Assumption NC1(1) is \( \sum_{i=2r+1}^{n_\eta} \mu_i(\frac{1}{nT}Z'_{nT}Z_{nT}R_n(\alpha)'R_n(\alpha)) > 0 \) and \( \mu_n(\eta \cdot Z)'R_n(\alpha)'R_n(\alpha)) > 0 \) wpa 1 as \( n, T \to \infty \).

**Proof of Proposition 3.**

Maximizing Eq. (4) with respect to \( \Gamma_n \) and \( F_T \) is equivalent to

\[
\min_{\Gamma_n \in \mathbb{R}^{n \times n}, F_T \in \mathbb{R}^{T \times n}} \sum_{i=1}^{T} \left( S_n(\lambda)S_n^{-1}(Z_n \delta_0 + U_n) - Z_n \delta + S_n(\lambda)S_n^{-1}\Gamma_0 f_0 - \Gamma_n f_1 \right)'R_n(\alpha)'R_n(\alpha) \\
\times \left( S_n(\lambda)S_n^{-1}(Z_n \delta_0 + U_n) - Z_n \delta + S_n(\lambda)S_n^{-1}\Gamma_0 f_0 - \Gamma_n f_1 \right)
\]

\(^{13}\text{See Zhang (2011), Theorem 8.10 and p.271}\)
\[
\begin{align*}
\geq & \min_{\tilde{\Gamma}_n \in \mathbb{R}^{n \times \tau}} \frac{1}{nT} \sum_{t=1}^{T} \left[ R_n(\alpha) \left( S_n(\lambda)S_n^{-1}(Z_{mt}\delta_0 + U_{mt}) - Z_{mt}\delta \right) - \Gamma_n f_t^i \right]^\top \left[ R_n(\alpha) \left( S_n(\lambda)S_n^{-1}(Z_{mt}\delta_0 + U_{mt}) - Z_{mt}\delta \right) - \Gamma_n f_t^i \right] \\
= & \min_{\tilde{\Gamma}_n \in \mathbb{R}^{n \times \tau}} \frac{1}{nT} \sum_{t=1}^{T} \left( S_n(\lambda)S_n^{-1}Z_{mt}\delta_0 - Z_{mt}\delta \right)^\top R_n(\alpha)^\top M_{\Gamma_n} R_n(\alpha) \left( S_n(\lambda)S_n^{-1}Z_{mt}\delta_0 - Z_{mt}\delta \right) \\
\geq & \min_{\tilde{\Gamma}_n \in \mathbb{R}^{n \times \tau}} \text{tr} \left( \frac{1}{nT} \sum_{t=1}^{T} R_n(\alpha) \left( S_n(\lambda)S_n^{-1}Z_{mt}\delta_0 - Z_{mt}\delta \right)^\top R_n(\alpha)^\top M_{\Gamma_n} R_n(\alpha) \left( S_n(\lambda)S_n^{-1}Z_{mt}\delta_0 - Z_{mt}\delta \right) - \frac{2}{nT} \sum_{t=1}^{T} R_n(\alpha) \left( S_n(\lambda)S_n^{-1}Z_{mt}\delta_0 - Z_{mt}\delta \right)^\top R_n(\alpha)^\top R_n(\alpha) \left( S_n(\lambda)S_n^{-1}Z_{mt}\delta_0 - Z_{mt}\delta \right) \right) \\
& - \max_{\tilde{\Gamma}_n \in \mathbb{R}^{n \times \tau}} \text{tr} \left( \frac{1}{nT} \sum_{t=1}^{T} R_n(\alpha) \left( S_n(\lambda)S_n^{-1}Z_{mt}\delta_0 - Z_{mt}\delta \right)^\top R_n(\alpha)^\top R_n(\alpha) \left( S_n(\lambda)S_n^{-1}Z_{mt}\delta_0 - Z_{mt}\delta \right) \right).
\end{align*}
\]

The first inequality above follows because the value of the minimization problem can be no less than the case where we were also able to optimally choose \( S_n(\lambda)S_n^{-1}\Gamma_0 \) and \( F_0T \). Eq. (B.1) is obtained by concentrating out the factor \( f_t \). Because there is no restriction on \( \tilde{\Gamma}_n \in \mathbb{R}^{n \times \tau} \), optimization with respect to \( R_n(\alpha) \tilde{\Gamma}_n \) is equivalent to optimization with respect to \( \tilde{\Gamma}_n \) in Eq. (B.1). Now we examine the terms in Eq. (B.2) one by one. Denote \( \eta = \left( \delta_0^\prime - \delta', \lambda_0 - \lambda \right)^\prime \). Then

\[
\begin{align*}
\min_{\tilde{\Gamma}_n \in \mathbb{R}^{n \times \tau}} & \text{tr} \left( \frac{1}{nT} \sum_{t=1}^{T} R_n(\alpha) \left( S_n(\lambda)S_n^{-1}Z_{mt}\delta_0 - Z_{mt}\delta \right) \left( S_n(\lambda)S_n^{-1}Z_{mt}\delta_0 - Z_{mt}\delta \right)^\top R_n(\alpha)^\top M_{\Gamma_n} \right) \\
= & \sum_{i=2T+1}^{n} \mu_i \left( \frac{1}{nT} R_n(\alpha) \left( \eta \cdot Z \right) (\eta \cdot Z)^\top R_n(\alpha)^\top \right) \geq b \| \eta \|_{1}^2,
\end{align*}
\]

for some constants \( b > 0 \) wpa 1 as \( n, T \to \infty \), by Assumption NC1. When \( G_n Z_{mt}\delta_0 = Z_{mt}C \) for a constant vector \( C \), then Assumption NC2 can be used in (B.3). In this case, \( S_n(\lambda)S_n^{-1}Z_{mt}\delta_0 - Z_{mt}\delta = Z_{mt}(\delta_0 - \delta) + Z_{mt}(\lambda_0 - \lambda) = Z_{mt}\eta^* \), where \( \eta^* = \delta_0 - \delta + C(\lambda_0 - \lambda) \). Denote \( \eta^* \cdot Z^* = \sum_{k=1}^{K} \eta^*_k Z_k \), by Assumption NC2,

\[
\begin{align*}
\min_{\tilde{\Gamma}_n \in \mathbb{R}^{n \times \tau}} & \text{tr} \left( \frac{1}{nT} \sum_{t=1}^{T} R_n(\alpha) \left( S_n(\lambda)S_n^{-1}Z_{mt}\delta_0 - Z_{mt}\delta \right) \left( S_n(\lambda)S_n^{-1}Z_{mt}\delta_0 - Z_{mt}\delta \right)^\top R_n(\alpha)^\top M_{\Gamma_n} \right) \\
= & \sum_{i=2T+1}^{n} \mu_i \left( \frac{1}{nT} R_n(\alpha) \left( \eta^* \cdot Z^* \right) (\eta^* \cdot Z^*)^\top R_n(\alpha)^\top \right) \geq b \| \eta^* \|_{1}^2.
\end{align*}
\]

for some constant \( b \geq 0 \) wpa 1 as \( n, T \to \infty \). Consider the next term,

\[
\begin{align*}
\frac{2}{nT} \sum_{t=1}^{T} \left( S_n(\lambda)S_n^{-1}Z_{mt}\delta_0 - Z_{mt}\delta \right)^\top R_n(\alpha)^\top R_n(\alpha) \left( S_n(\lambda)S_n^{-1}U_{mt} \right) \\
= \frac{2}{nT} \sum_{k=1}^{K+1} \eta_k \text{tr} \left( \left( R_n^{-1} S_n^{-1} \right)^\top R_n(\alpha)^\top R_n(\alpha) \left( Z_k \right) \varepsilon \right) = O_P \left( \frac{\| \eta \|_{1}}{\sqrt{nT}} \right),
\end{align*}
\]

(B.4)
by Lemma 3. For the next term, by using a law of large numbers for quadratic form (Lemma 9 in Yu et al. (2008)),

\[ \frac{1}{nT} \sum_{t=1}^{T} \left( S_n(\lambda)S_n^{-1}U_{nt} \right)' R_n(\alpha)' R_n(\alpha) (S_n(\lambda)S_n^{-1}U_{nt}) \]

\[ = \frac{\sigma_0^2}{n} \text{tr} \left( R_n^{-1} S_n^{-1} S_n(\lambda) R_n(\alpha)' R_n(\alpha) S_n(\lambda) S_n^{-1} R_n^{-1} \right) + O_P \left( \frac{1}{\sqrt{nT}} \right). \] (B.5)

From Lemma 2 (3), for an \( n \times n \) matrix \( A \), \( |\text{tr}(A)| \leq \text{rank}(A) \|A\| \), where \( \|\cdot\| \) is an induced matrix norm,

\[ \frac{2}{nT} \max_{\Gamma_n \in \mathbb{R}^{n \times 2r}} \text{tr} \left( \sum_{i=1}^{T} R_n(\alpha) \left( S_n(\lambda)S_n^{-1}U_{nt} \right) (S_n(\lambda)S_n^{-1}Z_{nt} - Z_{nt} \delta)' R_n(\alpha)' P_{T_n} \right) \]

\[ \leq 4r \frac{\sqrt{T}}{nT} \sum_{k=1}^{K+1} \|\eta_k\| \left\| R_n(\alpha) S_n(\lambda) S_n^{-1} R_n^{-1} \epsilon Z_k R_n(\alpha)' \right\|_2 = O_P \left( \frac{\|\eta\|_1}{\sqrt{\min(n,T)}} \right). \] (B.6)

The last equality is due to the sub-multiplicative property of \( \|\cdot\|_2 \) matrix norm, and that for a matrix \( A \),

\[ \|A\|_2^2 \leq \|A\|_1 \|A\|_\infty ; \] furthermore, Assumptions E and R guarantee that \( \|\epsilon\|_2 = O_P(\sqrt{\max(n,T)}) \) and \( \|Z_k\|_2 = O_P(\sqrt{nT}) \). Similarly,

\[ \frac{1}{nT} \max_{\Gamma_n \in \mathbb{R}^{n \times 2r}} \text{tr} \left( \sum_{i=1}^{T} R_n(\alpha) \left( S_n(\lambda)S_n^{-1}U_{nt} \right) (S_n(\lambda)S_n^{-1}U_{nt})' R_n(\alpha)' P_{T_n} \right) \]

\[ = O_P \left( \frac{1}{\min(n,T)} \right). \] (B.7)

Under Assumption NC1, by substituting Eqs. (B.3)-(B.7) into Eq. (B.2),

\[ Q_{nT}(\theta) \leq \frac{1}{n} \log |S_n(\lambda)R_n(\alpha)| \]

\[ - \frac{1}{2} \log \left( b\|\eta\|_1^2 + \frac{\sigma_0^2}{n} \text{tr} \left( R_n^{-1} S_n^{-1} S_n(\lambda)' R_n(\alpha)' R_n(\alpha) S_n(\lambda) S_n^{-1} R_n^{-1} \right) + O_P \left( \frac{1}{\sqrt{\min(n,T)}} \right) \right) \]

\[ = \tilde{Q}_{nT}(\theta). \] (B.8)

At \( \theta = \theta_0 \),

\[ Q_{nT}(\theta_0) = \frac{1}{n} \log |S_nR_n| - \frac{1}{2} \log \left( \min_{\Gamma_n \in \mathbb{R}^{n \times r}, F_\gamma \in \mathbb{R}^{T \times r}} \frac{1}{nT} \sum_{i=1}^{T} \left( \Gamma_n f_{0i} + U_{nt} - \Gamma_n f_i \right)' R_n' R_n \left( \Gamma_n f_{0i} + U_{nt} - \Gamma_n f_i \right) \right) \]

\[ \geq \frac{1}{n} \log |S_nR_n| - \frac{1}{2} \log \left( \frac{1}{nT} \sum_{i=1}^{T} \epsilon_{nt}' \epsilon_{nt} \right) \]

\[ = \frac{1}{n} \log |S_nR_n| - \frac{1}{2} \log \left( \frac{\sigma_0^2 + O_P \left( \frac{1}{\sqrt{nT}} \right) }{n} \right) \equiv Q(\theta_0). \] (B.9)

Lemma 1 in Wu (1981) provides a criterion for consistency of \( \hat{\theta}_{nT} = \arg \max_{\theta \in \Theta} Q_{nT}(\theta) \). To show that

\[ \|\hat{\theta}_{nT} - \theta_0\|_1 \overset{p}{\to} 0, \] it is sufficient to show that for all \( \tau > 0 \),

\[ \liminf_{n,T \to \infty} P \left( \inf_{\theta \in \Theta} \|\theta - \theta_0\|_1 \geq \tau (Q_{nT}(\theta_0) - Q_{nT}(\theta)) > 0 \right) = \]
1. From Eqs. (B.8) and (B.9),

\[
Q_{nT}(\theta_0) - Q_{nT}(\theta) \geq Q_{nT}(\theta_0) - \tilde{Q}_{nT}(\theta) = \frac{1}{2} \log \left( b ||\eta||_1^2 + \frac{\sigma^2}{n} \text{tr} \left( R_n^{-1} S_n^{-1} (\alpha) R_n (\alpha) S_n (\lambda) S_n^{-1} R_n^{-1} \right) \right) \\
- \frac{1}{2} \log \left( \sigma^2_n R_n^{-1} S_n^{-1} (\alpha) R_n (\alpha) S_n (\lambda) S_n^{-1} R_n^{-1} \right) + O_P \left( \frac{1}{\sqrt{\min(n,T)}} \right) \\
\geq \frac{1}{2} \log \left( b ||\eta||_1^2 + \frac{\sigma^2}{n} \text{tr} \left( R_n^{-1} S_n^{-1} (\alpha) R_n (\alpha) S_n (\lambda) S_n^{-1} R_n^{-1} \right) \right) \\
- \frac{1}{2} \log \left( \sigma^2_n \text{tr} \left( R_n^{-1} S_n^{-1} (\alpha) R_n (\alpha) S_n (\lambda) S_n^{-1} R_n^{-1} \right) \right) + o_P(1),
\]  

(B.10)

because \( \frac{1}{n} \text{tr} \left( R_n^{-1} S_n^{-1} (\alpha) R_n (\alpha) S_n (\lambda) S_n^{-1} R_n^{-1} \right) \geq \left| R_n^{-1} S_n^{-1} (\alpha) R_n (\alpha) S_n (\lambda) S_n^{-1} R_n^{-1} \right|^{-\frac{1}{2}} \) by the inequality of arithmetic and geometric means. Under Assumption NC1, when \( ||\delta - \delta_0||_1 \geq \tau \) or \( |\lambda - \lambda_0| \geq \tau \), \( b > 0 \) and \( \liminf_{n,T \to \infty} P (\inf_{\theta \in \Theta} ||\theta - \theta_0|| \geq \tau (Q_{nT}(\theta_0) - Q_{nT}(\theta)) > 0) = 1 \) holds. In addition, (B.10) is strict when \( \alpha \neq \alpha_0 \) even if \( \lambda = \lambda_0 \), as guaranteed by Assumption NC1(2). In either case, consistency follows from Lemma 1 of Wu (1981).

When Assumption NC2 holds instead of NC1, the inequality in (B.10) is strict when \( \alpha \neq \alpha_0 \) or \( \lambda \neq \lambda_0 \) from Assumption NC2(2), and even with \( \lambda = \lambda_0 \), \( b ||\eta||_1^2 > 0 \) when \( ||\delta - \delta_0||_1 \geq \tau \). Therefore consistency follows from \( \liminf_{n,T \to \infty} P (\inf_{\theta \in \Theta} ||\theta - \theta_0|| \geq \tau (Q_{nT}(\theta_0) - Q_{nT}(\theta)) > 0) = 1 \).

\[ \square \]

Appendix C. Perturbation Theory and Series Expansion of the Concentrated Log Likelihood Function

Kato (1995) has a systematic presentation of perturbation theory. Moon and Weidner (2015) applies the perturbation theory to a regression panel with common factors. Here we show specifically how to expand \( L_nT(\theta_0) \) itself and \( L_nT(\theta) \) around \( \theta_0 \). Firstly,

\[
R_n(\alpha) \left( S_n(\lambda) Y - \sum_{k=1}^K Z_k \delta_k \right) \\
= \left( R_n + (\alpha_0 - \alpha) W_n \right) \left( \Gamma_n F_T + R_n^{-1} \varepsilon + \sum_{k=1}^K Z_k (\delta_k - \delta_k) + (Z_{K+1} + G_n \Gamma_n F_T + G_n R_n^{-1} \varepsilon) (\lambda_0 - \lambda) \right) \\
= \tilde{\Gamma}_n F_T + \sum_{k=0}^{K+2} \xi_k V_k + \sum_{k=1, k_0=0}^{K+2} \xi_{k_0} \xi_{k_2} V_{k_0 k_2},
\]

(C.1)

where \( \tilde{\Gamma}_n = R_n \Gamma_n \), \( \xi_0 = \frac{||\varepsilon||_2}{\sqrt{nT}} \), \( \xi_k = \delta_k - \delta_k \) for \( k = 1, \cdots, K \), \( \xi_{K+1} = \lambda_0 - \lambda \), \( \xi_{K+2} = \alpha_0 - \alpha \), \( V_0 = \frac{\sqrt{nT}}{||\varepsilon||_2} \), \( V_k = R_n Z_k \) for \( k = 1, \cdots, K \), \( V_{K+1} = R_n (Z_{K+1} + G_n \Gamma_n F_T) \), and \( V_{K+2} = W_n R_n^{-1} \tilde{\Gamma}_n F_T \); and also, \( V_{k_2} = R_n G_n R_n^{-1} \frac{\sqrt{nT}}{||\varepsilon||_2} \) for \( k_1 = 0, k_2 = K + 1; V_{k_1 k_2} = \tilde{\Gamma}_n \frac{\sqrt{nT}}{||\varepsilon||_2} \) for \( k_1 = 0, k_2 = K + 2; V_{k_1 k_2} = \tilde{W}_n Z_{k_2} \) for \( k_1 = 0, k_2 = K + 3 \).
addition, our objective function (Eq. (5)) has the Jacobian term \( \log \xi \) on the remainder can be found in the supplementary file. Therefore, if there is no perturbation, i.e., \( n \) is small such that the condition on \( \xi \) in Lemma 7 holds. All the \( V_k \) and \( V_{k_2} \) above are \( n \times T \) matrices. In Moon and Weidner (2015), for the NLS estimator of a factor panel regression, only the similar term \( \Gamma_n F_T^\prime \) appears. For our model, \( \xi_{K+1}, V_{K+1}, \xi_{K+2}, V_{K+2} \) and \( V_{k_2} \) are extra terms due to the contemporaneous spatial lag and spatial disturbances. In addition, our objective function (Eq. (5)) has the Jacobian term \( \log |S_n(\lambda)R_n(\alpha)| \).

For our spatial model, the part \( L_{nT}(\theta) \) of the objective function without the Jacobian term is

\[
L_{nT}(\xi) = \sum_{i=r_0+1}^{n} \mu_i \left( \frac{1}{nT} T^{(0)} + \frac{1}{nT} \sum_{k_1=0}^{K+2} \xi_{k_1} T_{k_1}^{(1)} + \frac{1}{nT} \sum_{k_1=0}^{K+2} \sum_{k_2=0}^{K+2} \xi_{k_1} \xi_{k_2} T_{k_1k_2}^{(2)} + \frac{1}{nT} \sum_{k_1=0}^{K+2} \sum_{k_2=0}^{K+2} \sum_{k_3=0}^{K+2} \xi_{k_1} \xi_{k_2} \xi_{k_3} T_{k_1k_2k_3}^{(3)} + \frac{1}{nT} \sum_{k_1=0}^{K+2} \sum_{k_2=0}^{K+2} \sum_{k_3=0}^{K+2} \sum_{k_4=0}^{K+2} \xi_{k_1} \xi_{k_2} \xi_{k_3} \xi_{k_4} T_{k_1k_2k_3k_4}^{(4)} \right),
\]

where \( T^{(0)} = \tilde{\Gamma}_n F_T^\prime F_T \tilde{\Gamma}_n, T_{k_1}^{(1)} = V_{k_1} F_T \tilde{\Gamma}_n + \Gamma_n F_T^\prime V_{k_1}, T_{k_1k_2}^{(2)} = V_{k_1k_2} F_T \tilde{\Gamma}_n + \tilde{\Gamma}_n F_T^\prime V_{k_1k_2} + V_{k_1k_2}, T_{k_1k_2k_3}^{(3)} = V_{k_1k_2k_3} V_{k_1k_2k_3}, T_{k_1k_2k_3k_4}^{(4)} = V_{k_1k_2k_3k_4}, \) with \( k_j = 0, \cdots, K + 2 \) and \( j = 1, 2, 3, 4 \). In our model, \( T^{(0)} \) is the unperturbed operator and it has exactly \( n - r_0 \) zero eigenvalues and the rest \( r_0 \) eigenvalues are strictly positive. Therefore, if there is no perturbation, i.e., \( \xi_k = 0 \) for \( k = 0, \cdots, K + 2 \), \( L_{nT}(\xi) = 0 = 0 \). The objective is to expand \( L_{nT}(\xi) \) around \( L_{nT}(\xi) = 0 \) in terms of \( \xi \), \( T^{(1)} \), \( T^{(2)} \), \( T^{(3)} \) and \( T^{(4)} \) using the formula in Kato (1995). Define

\[
b_{nT} = \max \left( \sum_{k_1=0}^{K+2} \sum_{k_2=0}^{K+2} \sum_{k_3=0}^{K+2} \sum_{k_4=0}^{K+2} |\xi_{k_1}| |\xi_{k_2}| |\xi_{k_3}| |\xi_{k_4}| \left( \frac{1}{nT} \left\| T_{k_1k_2k_3k_4}^{(4)} \right\| \right)^{\frac{1}{2}} \left( \frac{1}{d_{\max}^2(\tilde{\Gamma}_n, F_T)} \right)^{\frac{1}{2}} \right) \]

Because \( \left\| T_{k_1}^{(1)} \right\|_2 = O_P(nT), \left\| T_{k_1k_2}^{(2)} \right\|_2 = O_P(nT), \left\| T_{k_1k_2k_3}^{(3)} \right\|_2 = O_P(nT), \left\| T_{k_1k_2k_3k_4}^{(4)} \right\|_2 = O_P(nT), \) and \( d_{\max}^2(\tilde{\Gamma}_n, F_T) \) is converging to some positive constant, \( b_{nT} = O_P(\| \xi \|_1) \). The following lemma provides an expansion of \( L_{nT}(\xi) \) as a power series in \( \xi \). The expansion is valid if \( \xi \) is small such that the condition on \( b_{nT} \) in Lemma 7 below holds. We list the expansion below for its use, but the detailed derivation of this lemma and bound on the remainder can be found in the supplementary file.

**Lemma 7.** Under Assumption SF and assume that \( d_{\max}^2(\tilde{\Gamma}_n, F_T) b_{nT} > 0 \), then \( L_{nT}(\xi) \) has a convergent
series expansion \( \tilde{L}_{nT}(\xi) = \frac{1}{nT} \sum_{g=1}^{\infty} \sum_{k_1=0}^{K+2} \sum_{k_2=0}^{K+2} \cdots \sum_{k_g=0}^{K+2} \xi_{k_1} \xi_{k_2} \cdots \xi_{k_g} \text{tr} \left( \tilde{T}^{(g)}_{k_1 k_2 \cdots k_g} \right) \), where

\[
\tilde{T}^{(g)}_{k_1 k_2 \cdots k_g} = - \sum_{p=\left[\frac{g}{2}\right]}^{g} (-1)^p \sum_{v_1+\cdots+v_p=g} S^{(m_1)} \cdots S^{(m_p)} T^{(v_1)}_{k_1} \cdots T^{(v_p)}_{k_g}
\]

with \( S^{(0)} = -M_{\Gamma_n}, \quad S^{(m)} = S^{(m)}_0 \) for \( m \geq 1 \) where \( S_0 = \tilde{\Gamma}_n (\tilde{\Gamma}_n)^{-1} (F_T)^{-1} (\tilde{\Gamma}_n)^{-1} \tilde{\Gamma}_n \). For \( g \geq 5, \)

\[
\frac{1}{nT} \sum_{k_1=0}^{K+2} \cdots \sum_{k_g=0}^{K+2} \xi_{k_1} \cdots \xi_{k_g} \text{tr} \left( \tilde{T}^{(g)}_{k_1 k_2 \cdots k_g} \right) \leq \frac{16 \rho d_{\max}^2 (\tilde{\Gamma}_n, F_T) d_{\min}^2 (\tilde{\Gamma}_n, F_T)}{(16 \rho d_{\max}^2 (\tilde{\Gamma}_n, F_T) - d_{\min}^2 (\tilde{\Gamma}_n, F_T)) \left( 1 - \frac{16 \rho d_{\max}^2 (\tilde{\Gamma}_n, F_T) b_{nT}}{d_{\min}^2 (\tilde{\Gamma}_n, F_T)} \right)^G} = O_P(\|\xi\|^{G+1}).
\]

The condition on \( b_{nT} \) can be satisfied when \( n, T \to \infty \), because \( b_{nT} = O_P(\|\xi\|_1), \|\xi_0\| = O_P(\frac{1}{\sqrt{\min(n, T)}}) \), and \( \|\tilde{\theta} - \theta_0\| = O_P(1) \) by Proposition 3. At \( \theta_0 \), \( L_{nT}(\theta_0) \) has the perturbation \( \xi_0 = \frac{\|\varepsilon\|}{\sqrt{nT}} \) while other \( \xi_k \), \( k = 1, \cdots, K+2 \) are zero. \( L_{nT}(\theta_0) \) has an expansion in terms of \( \xi_0 \) as,

\[
L_{nT}(\theta_0) = \frac{1}{nT} \sum_{i=0}^{n} \mu_i \left( (\Gamma_n F_T + \varepsilon) (\Gamma_n F_T' + \varepsilon)' \right) = \frac{1}{nT} \text{tr} \left( M_{F_T} \varepsilon M_{F_T} \varepsilon' \right) - \frac{2}{nT} \text{tr} \left( M_{F_T} \varepsilon M_{F_T} \varepsilon' F_{T_n} F_T \varepsilon' \right) + J_{nT} + O_P \left( \left( \frac{\|\varepsilon\|_2}{\sqrt{nT}} \right)^5 \right) \quad (C.4)
\]

\[
= \sigma_0^2 + O_P(\min(n, T)^{-1}), \quad (C.5)
\]

where in Eq. (C.4), the series expansion is truncated at order 4 with

\[
J_{nT} = - \frac{1}{nT} \text{tr} \left( S^{(0)} T^{(2)}_{00} S^{(1)} T^{(2)}_{00} S^{(0)} \right) + \frac{1}{nT} \text{tr} \left( S^{(0)} T^{(2)}_{00} S^{(0)} T^{(1)} T^{(1)} s^{(0)} \right) + \frac{1}{nT} \text{tr} \left( S^{(0)} T^{(2)}_{00} S^{(1)} T^{(1)} S^{(1)} T^{(1)} s^{(0)} \right) + \frac{1}{nT} \text{tr} \left( S^{(0)} T^{(2)}_{00} S^{(1)} T^{(1)} S^{(1)} T^{(1)} s^{(0)} \right) + \frac{1}{nT} \text{tr} \left( S^{(0)} T^{(2)}_{00} S^{(1)} T^{(1)} S^{(1)} T^{(1)} s^{(0)} \right)
\]

\[
= - \frac{1}{nT} \text{tr} \left( S^{(0)} T^{(2)}_{00} S^{(1)} T^{(1)} s^{(0)} \right) - \frac{1}{nT} \text{tr} \left( S^{(0)} T^{(2)}_{00} S^{(1)} T^{(1)} s^{(0)} \right) - \frac{1}{nT} \text{tr} \left( S^{(0)} T^{(2)}_{00} S^{(1)} T^{(1)} s^{(0)} \right)
\]

\[\text{Notice that } v_1 + \cdots + v_p = g. \text{ When } v = 1, T^{(v)}_{kk'} \text{ has a single } k \text{ subscript; when } v = 2, T^{(v)}_{kk'} \text{ has two subscripts, } k, k'. \text{ For example, when } g = 4, \text{ and in a particular summand, } v_1 = 1, v_2 = 2, v_3 = 1, \text{ a typical term in the summand is } S^{(m_1)} T^{(1)}_{00} S^{(m_2)} T^{(2)}_{00} S^{(m_3)} T^{(1)}_{00}. \]
\[-\frac{1}{nT} \text{tr} \left( S^{(1)} T_0^{(1)} S^{(0)} T_0^{(1)} S^{(1)} T_0^{(1)} S^{(1)} T_0^{(1)} S^{(1)} T_0^{(1)} S^{(1)} T_0^{(0)} S^{(0)} \right) - \frac{1}{nT} \text{tr} \left( S^{(0)} T_0^{(1)} S^{(1)} T_0^{(1)} S^{(1)} T_0^{(1)} S^{(1)} T_0^{(0)} S^{(0)} \right). \tag{C.6}\]

It follows that \( \frac{1}{nT} \text{tr}(\theta_0) = \frac{1}{\sigma_0^2} + O_P(\min(n, T)^{-1}). \)

For the derivation of the asymptotic distribution of the QMLE, it is sufficient to consider the series expansion of \( L_{nT}(\theta) \) truncated at order \( G = 4 \). Define the following \((K + 2) \times 1\) vectors,

\[
C^{(1)} = \left( \frac{1}{\sqrt{nT}} \text{tr} \left( M_{\Gamma_n} R_n Z_1 M_{F_n} \varepsilon' \right), \cdots, \frac{1}{\sqrt{nT}} \text{tr} \left( M_{\Gamma_n} R_n Z_K M_{F_n} \varepsilon' \right), \frac{1}{\sqrt{nT}} \text{tr} \left( M_{\Gamma_n} R_n Z_1 M_{F_n} \varepsilon' \right), 0 \right)^T, \tag{C.7}
\]

\[
C^{(2)} = \left( \frac{1}{\sqrt{nT}} \text{tr} \left( M_{\Gamma_n} M_{F_n} \varepsilon' P_{\Gamma_n} Z_1 M_{F_n} \varepsilon' \right), \cdots, \frac{1}{\sqrt{nT}} \text{tr} \left( M_{\Gamma_n} M_{F_n} \varepsilon' Z_n M_{F_n} \varepsilon' \right), \frac{1}{\sqrt{nT}} \text{tr} \left( M_{\Gamma_n} M_{F_n} \varepsilon' P_{\Gamma_n} Z_1 M_{F_n} \varepsilon' \right), 0 \right)^T, \tag{C.8}
\]

\[
C^{(3)} = \left( 0, \cdots, 0, \frac{1}{\sqrt{nT}} \text{tr} \left( M_{\Gamma_n} M_{F_n} \varepsilon' P_{\Gamma_n} Z_1 M_{F_n} \varepsilon' \right), \frac{1}{\sqrt{nT}} \text{tr} \left( M_{\Gamma_n} \tilde{G}_n M_{\Gamma_n} \varepsilon' M_{F_n} \varepsilon' \right) \right)^T, \tag{C.9}
\]

where \( P_{\Gamma_n} = \tilde{\Gamma}_n (\tilde{\Gamma}_n' \tilde{\Gamma}_n)^{-1} (F_T' F_T)^{-1} F_T' \). Define the following \((K + 2) \times (K + 2)\) matrices,

\[
C_1 = \begin{pmatrix}
\frac{1}{nT} \text{tr} \left( M_{\Gamma_n} R_n Z_1 M_{F_n} Z_1' R_n' \right) & \cdots & \frac{1}{nT} \text{tr} \left( M_{\Gamma_n} R_n Z_1 M_{F_n} Z_K' R_n' \right) & \frac{1}{nT} \text{tr} \left( M_{\Gamma_n} R_n Z_1 M_{F_n} Z_1' R_n' \right)

\vdots & \ddots & \vdots & \vdots

\frac{1}{nT} \text{tr} \left( M_{\Gamma_n} R_n Z_1 M_{F_n} Z_1' R_n' \right) & \cdots & \frac{1}{nT} \text{tr} \left( M_{\Gamma_n} R_n Z_1 M_{F_n} Z_K' R_n' \right) & \frac{1}{nT} \text{tr} \left( M_{\Gamma_n} R_n Z_1 M_{F_n} Z_1' R_n' \right)

0 & \cdots & 0 & 0
\end{pmatrix},
\]

\[
C_2 = \begin{pmatrix}
0 & \cdots & 0 & 0

\vdots & \ddots & \vdots & \vdots

0 & \cdots & 0 & 0
\end{pmatrix}, \tag{C.10}
\]

and

\[
C = C_1 + C_2.
\]

**Lemma 8.** Assume that \( \frac{\eta}{\gamma} \rightarrow \kappa^2 > 0 \), Assumptions NC1 (or NC2), E, R and SF hold, then with \( \theta = (\delta', \lambda, \alpha)' \) in a small neighborhood of \( \theta_0 \),

\[
L_{nT}(\theta) = L_{nT}(\theta_0) - \frac{2}{\sqrt{nT}} (\theta - \theta_0)' \left( C^{(1)} + C^{(2)} + C^{(3)} \right) + (\theta - \theta_0)' C (\theta - \theta_0) + L_{\text{rem}}(\theta),
\]

38
where \( C^{(1)}, C^{(2)}, C^{(3)} \) and \( C \) are defined in Eqs. (C.7)-(C.10), and the remainder term is \( L_{nT}^{\text{rem}}(\theta) = O_P\left(\|\theta - \theta_0\|_1^3\right) + O_P\left(\|\theta - \theta_0\|_1^2 (nT)^{-\frac{1}{2}}\right) + O_P\left(\|\theta - \theta_0\|_1 (nT)^{-\frac{3}{2}}\right) + O_P\left((nT)^{-\frac{3}{2}}\right)\).

Lemma 8 is an application of Lemma 7 by rearrangement.

**Lemma 9.** Under Assumptions E, R and SF, and assuming that \( \frac{2}{n} \to \kappa > 0 \), we have \( C = O_P(1) \); also for \( k = 1, \ldots, K, K + 1 \),

\[
C_k^{(1)} = \frac{1}{\sqrt{nT}} \text{tr}(\bar{Z}_k \varepsilon') - \frac{1}{\sqrt{nT}} \text{tr}(\hat{M}_n R_n (Z_k - \hat{Z}_k) P_n \varepsilon') = O_P(1), C_k^{(2)} = o_P(1),
\]

and

\[
C^{(3)} = \left(0, \cdots, 0, \frac{1}{\sqrt{nT}} \text{tr}(M_{n} R_n G_n \varepsilon' \varepsilon') + O_P(1) \right) = \left(0, \cdots, 0, \sqrt{T} \text{tr}(G_n) \sigma_0^2 + O_P(1) \right).
\]

**Lemma 10.** Assume that \( \frac{2}{n} \to \kappa^2 > 0 \) and Assumptions E, R and SF hold, then \( \tilde{D}_{nT} - D_{nT} = o_P(1) \), where \( D_{nT} \) and \( \tilde{D}_{nT} \) are respectively defined in Eq. (9) and Eq. (D.9).

**Lemma 11.** Under the assumptions of Theorem 3, \( \|M_{n} - M_n\|_2 = \|P_{n} - P_n\|_2 = O_P\left(\frac{1}{\sqrt{nT}}\right) \) and \( \|M_{F} - M_{F_n}\|_2 = \|P_{F} - P_{F_n}\|_2 = O_P\left(\frac{1}{\sqrt{T}}\right) \), where \( M_{n} \) and \( M_{F_n} \) are defined in Section 3.3.

**Appendix D. Asymptotic Distributions: Proof of Theorems 1-5**

**Proof of Theorem 1.**

Lemma 8 shows that, for \( \theta \) close to \( \theta_0 \), \( L_{nT}(\theta) = L_{nT}(\theta_0) - \frac{2}{\sqrt{nT}} (\theta - \theta_0)' (C^{(1)} + C^{(2)} + C^{(3)}) + (\theta - \theta_0)' C(\theta - \theta_0) + L_{nT}^{\text{rem}}(\theta) \), where \( L_{nT}^{\text{rem}}(\theta) = O_P\left(\|\theta - \theta_0\|_1^3\right) + O_P\left(\|\theta - \theta_0\|_1^2 (nT)^{-\frac{1}{2}}\right) + O_P\left(\|\theta - \theta_0\|_1 (nT)^{-\frac{3}{2}}\right) + O_P\left((nT)^{-\frac{3}{2}}\right) \). In the following, we provide concise expressions for those terms in the expansion of \( Q_{nT}(\tilde{\theta}_{nT}) \) around \( \theta_0 \), where \( \tilde{\theta}_{nT} \) satisfies \( \|\tilde{\theta}_{nT} - \theta_0\|_1 = o_P(1) \).

Before proceeding further, let's examine the \((K + 2) \times 1\) vectors \( C^{(1)}, C^{(2)}, C^{(3)} \) and the \((K + 2) \times (K + 2)\) matrix \( C \) more closely. Notice that the \( K + 1 \) and \( K + 2 \) entries of \( C^{(3)} \) have \( C_{K+1}^{(3)} = O_P\left(\sqrt{nT}\right) \) and \( C_{K+2}^{(3)} = O_P\left(\sqrt{nT}\right) \), which are of a higher stochastic order than \( C^{(1)} \) and \( C^{(2)} \). However, as Theorem 1 shows, the higher order parts of \( C^{(3)} \) will be canceled with terms from the log Jacobian determinant in the log likelihood function.

Recall that \( \bar{Z}_k = M_{n} R_n \bar{Z}_k M_{F} + M_{n} R_n (Z_k - \hat{Z}_k) \) for \( k = 1, \cdots, K + 1 \). Notice that \( \bar{Z}_{k,i} \) is independent from \( \varepsilon_{li} \). The concentrated likelihood function is

\[
Q_{nT}(\tilde{\theta}_{nT}) = \frac{1}{n} \log \left| S_n(\tilde{\lambda}_{nT}) R_n(\tilde{\alpha}_{nT}) \right| - \frac{1}{2} \log L_{nT}(\tilde{\theta}_{nT})
\]
\[
\begin{align*}
&= \frac{1}{n} \log \left| S_n(\tilde{\lambda}_{nT}) R_n(\tilde{\alpha}_{nT}) \right| \\
&\quad - \frac{1}{2} \log \left( L_{nT}(\theta_0) - \frac{2}{\sqrt{nT}} (\tilde{\theta}_{nT} - \theta_0)' \left( C^{(1)} + C^{(2)} + C^{(3)} \right) + (\tilde{\theta}_{nT} - \theta_0)' C (\tilde{\theta}_{nT} - \theta_0) + L_{nT}^{\text{rem}}(\tilde{\theta}_{nT}) \right) \\
&= Q_{nT}(\theta_0) - \frac{1}{n} \text{tr}(G_n) (\tilde{\lambda}_{nT} - \lambda_0) - \frac{1}{2n} \text{tr}(G_n^2) (\tilde{\lambda}_{nT} - \lambda_0)^2 + O(\|\tilde{\lambda}_{nT} - \lambda_0\|^3) \\
&\quad - \frac{1}{n} \text{tr}(\tilde{G}_n) (\tilde{\alpha}_{nT} - \alpha_0) - \frac{1}{2n} \text{tr}(\tilde{G}_n^2) (\tilde{\alpha}_{nT} - \alpha_0)^2 + O(\|\tilde{\alpha}_{nT} - \alpha_0\|^3) \\
&\quad - \frac{1}{2} \log \left( 1 - \frac{2}{\sqrt{nT}} (\tilde{\theta}_{nT} - \theta_0)' \left( \frac{C^{(1)} + C^{(2)} + C^{(3)}}{L_{nT}(\theta_0)} \right) + (\tilde{\theta}_{nT} - \theta_0)' \frac{C}{L_{nT}(\theta_0)} (\tilde{\theta}_{nT} - \theta_0) + \frac{L_{nT}^{\text{rem}}(\tilde{\theta}_{nT})}{L_{nT}(\theta_0)} \right),
\end{align*}
\]

(D.1)

where the Taylor expansions of $\log |S_n(\tilde{\lambda}_{nT})|$ and $\log |R_n(\tilde{\alpha}_{nT})|$ are used. From Lemma 9 and by using log$(1 + x) = x - \frac{x^2}{2} + O(x^3)$ with

\[
x = - \frac{2}{nT L_{nT}(\theta_0)} \left[ \text{tr} \left( M_{\Gamma_n}^T R_n G_n R_n^{-1} M_{\Gamma_n} \epsilon \epsilon' \right) (\tilde{\lambda}_{nT} - \lambda_0) + \text{tr} \left( M_{\Gamma_n}^T \tilde{G}_n M_{\Gamma_n} \epsilon \epsilon' \right) (\tilde{\alpha}_{nT} - \alpha_0) \right] \\
+ O_P \left( \|\tilde{\theta}_{nT} - \theta_0\|_1 (nT)^{-\frac{1}{2}} \right) + O_P \left( \|\tilde{\theta}_{nT} - \theta_0\|_1^2 (nT)^{-\frac{1}{2}} \right) + O_P \left( \|\tilde{\theta}_{nT} - \theta_0\|_1^3 (nT)^{-\frac{1}{2}} \right),
\]

(D.2)

where $D^{(1)} = \left( 0, \ldots, 0, \sqrt{\frac{T}{n}} \text{tr}(G_n), \sqrt{\frac{T}{n}} \text{tr}(\tilde{G}_n) \right)'$ is a $(K + 2) \times 1$ vector and the $(K + 2) \times (K + 2)$ matrix $D_{nT}$ is the one in Eq. (9). The remainder term is

\[
Q_{nT}^{\text{rem}}(\tilde{\theta}_{nT}) = O_P \left( (nT)^{-\frac{3}{2}} \left( 1 + nT \|\tilde{\theta}_{nT} - \theta_0\|_1^2 + 2\sqrt{nT} \|\tilde{\theta}_{nT} - \theta_0\|_1 \right) \right) + O_P \left( \|\tilde{\theta}_{nT} - \theta_0\|_1^3 \right) \\
= o_P \left( (nT)^{-1} \left( 1 + \sqrt{nT} \|\tilde{\theta}_{nT} - \theta_0\|_1 \right)^2 \right).
\]

(D.3)
In Eq. (D.3), $O_P \left( \left\| \tilde{\theta}_{nT} - \theta_0 \right\|_1^2 \right) = o_P \left( \left\| \tilde{\theta}_{nT} - \theta_0 \right\|_1^2 \right)$, because $\left\| \tilde{\theta}_{nT} - \theta_0 \right\|_1 = o_P(1)$.

Lemma 9 shows that $\frac{C^{(1)} + C^{(2)} + C^{(3)}}{L_{nT}(\theta_0)} - D^{(1)} = O_P(1)$. Define $\gamma = \frac{1}{\sqrt{nT}} D_{nT}^{-1} \left( \frac{C^{(1)} + C^{(2)} + C^{(3)}}{L_{nT}(\theta_0)} - D^{(1)} \right)$ which is $O_P \left( \frac{1}{\sqrt{nT}} \right)$. The rest of the proof is similar to Corollary 4.3 of Moon and Weidner (2015). Completing the squares in Eq. (D.2), $Q_{nT}(\tilde{\theta}_{nT}) = Q_{nT}(\theta_0) - \frac{1}{2} \left( \tilde{\theta}_{nT} - \theta_0 - \gamma \right)' D_{nT} \left( \tilde{\theta}_{nT} - \theta_0 - \gamma \right) + \frac{1}{2} \gamma' D_{nT} \gamma + Q_{nT}^{\text{rem}}(\tilde{\theta}_{nT})$.

Consider the following two cases, $\tilde{\theta}_{nT} = \hat{\theta} = \arg\max Q_{nT}(\theta)$ and $\tilde{\theta}_{nT} = \theta_0 + \gamma$. Notice that both $\hat{\theta}$ and $\theta_0 + \gamma$ satisfy the condition that $\left\| \tilde{\theta}_{nT} - \theta_0 \right\|_1 = o_P(1)$. As $Q_{nT}(\hat{\theta}_{nT}) = Q_{nT}(\theta_0) - \frac{1}{2} \left( \hat{\theta}_{nT} - \theta_0 - \gamma \right)' D_{nT} \left( \hat{\theta}_{nT} - \theta_0 - \gamma \right) + \frac{1}{2} \gamma' D_{nT} \gamma + Q_{nT}^{\text{rem}}(\hat{\theta}_{nT})$, $Q_{nT}(\theta_0 + \gamma) = Q_{nT}(\theta_0) + \frac{1}{2} \gamma' D_{nT} \gamma + Q_{nT}^{\text{rem}}(\theta_0 + \gamma)$, and $Q_{nT}(\hat{\theta}_{nT}) \geq Q_{nT}(\theta_0 + \gamma)$, it follows that $\left( \hat{\theta}_{nT} - \theta_0 - \gamma \right)' D_{nT} \left( \hat{\theta}_{nT} - \theta_0 - \gamma \right) \leq 2Q_{nT}^{\text{rem}}(\hat{\theta}_{nT}) - 2Q_{nT}^{\text{rem}}(\theta_0 + \gamma)$. Using Eq. (D.4), $Q_{nT}^{\text{rem}}(\hat{\theta}_{nT}) - Q_{nT}^{\text{rem}}(\theta_0 + \gamma) \leq o_P \left( (nT)^{-1} \left( (1 + \sqrt{nT} \left\| \tilde{\theta}_{nT} - \theta_0 \right\|_1 + (1 + \sqrt{nT} \left\| \gamma \right\|_1) \right)^2 \right)$.

Because $D_{nT}$ is assumed to be positive definite,

$$\sqrt{nT} \left\| \tilde{\theta}_{nT} - \theta_0 - \gamma \right\|_1 \leq o_P \left( 1 + \sqrt{nT} \left\| \tilde{\theta}_{nT} - \theta_0 \right\|_1 \right) + o_P \left( 1 + \sqrt{nT} \left\| \gamma \right\|_1 \right)$$

$$= o_P \left( 1 + \sqrt{nT} \left\| \tilde{\theta}_{nT} - \theta_0 \right\|_1 \right) + o_P \left( 1 + \sqrt{nT} \left\| \gamma \right\|_1 \right).$$

Because $\gamma = O_P(\frac{1}{\sqrt{nT}}), o_P(1 + \sqrt{nT} \left\| \gamma \right\|_1) = o_P(1)$, therefore $\sqrt{nT} \left\| \tilde{\theta}_{nT} - \theta_0 - \gamma \right\|_1 = o_P(1)$, and $\sqrt{nT} \left( \tilde{\theta}_{nT} - \theta_0 \right) - \sqrt{nT} \gamma = o_P(1)$, which implies $\sqrt{nT} \left( \tilde{\theta}_{nT} - \theta_0 \right) = D_{nT}^{-1} \left( \frac{C^{(1)} + C^{(2)} + C^{(3)}}{L_{nT}(\theta_0)} - D^{(1)} \right) + o_P(1)$. Because $L_{nT}(\theta_0) = \sigma_0^2 + O_P(\frac{1}{n})$,

$$\sqrt{nT} \left( \tilde{\theta}_{nT} - \theta_0 \right) = D_{nT}^{-1} \left( \frac{C^{(1)} + C^{(2)} + C^{(3)}}{L_{nT}(\theta_0)} - D^{(1)} \right) + o_P(1)$$

$$= (L_{nT}(\theta_0)D_{nT})^{-1} \left( C^{(1)} + C^{(2)} + C^{(3)} - D^{(1)}L_{nT}(\theta_0) \right) + o_P(1)$$

$$= (L_{nT}(\theta_0)D_{nT})^{-1} \left( C^{(1)} + C^{(2)} + C^{(3)} - D^{(1)}\sigma_0^2 \right) + (L_{nT}(\theta_0)D_{nT})^{-1} D^{(1)} \left( \sigma_0^2 - L_{nT}(\theta_0) \right) + o_P(1)$$

$$= \left( \sigma_0^2 D_{nT} \right)^{-1} \left( C^{(1)} + C^{(2)} + C^{(3)} - D^{(1)} \sigma_0^2 \right)$$

$$+ \left( \sigma_0^2 D_{nT} \right)^{-1} D^{(1)} \left( \sigma_0^2 - \frac{1}{nT} \text{tr} (\varepsilon \varepsilon') + \frac{1}{nT} \text{tr} (P_{nT} \varepsilon \varepsilon') + \frac{1}{nT} \text{tr} (\varepsilon P_{nT} \varepsilon') \right) + o_P(1)$$

$$= \left( \sigma_0^2 D_{nT} \right)^{-1} \left( \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} Z_{i,t} \varepsilon_{it} \frac{1}{\sqrt{nT}} \text{tr} (M_{nT} R_n (Z_1 - \tilde{Z}_1) P_{F_{nT}} \varepsilon') \right)$$

$$+ \left( \sigma_0^2 D_{nT} \right)^{-1} \left( \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} Z_{2,i} \varepsilon_{it} \frac{1}{\sqrt{nT}} \text{tr} (M_{nT} R_n (Z_2 - \tilde{Z}_2) P_{F_{nT}} \varepsilon') \right)$$

$$+ \left( \sigma_0^2 D_{nT} \right)^{-1} \left( \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} Z_{K+1,i} \varepsilon_{it} \frac{1}{\sqrt{nT}} \text{tr} (M_{nT} R_n (Z_{K+1} - \tilde{Z}_{K+1}) P_{F_{nT}} \varepsilon') \right) \right)$$

$$= \left( \sigma_0^2 D_{nT} \right)^{-1} \left( \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} Z_{K+1,i} \varepsilon_{it} \frac{1}{\sqrt{nT}} \text{tr} (M_{nT} R_n (Z_{K+1} - \tilde{Z}_{K+1}) P_{F_{nT}} \varepsilon') \right) \right)$$

41
\[
+ \left( \sigma_0^2 D_{\eta T} \right)^{-1} \begin{pmatrix}
0 \\ \\ 0 \\
\frac{1}{\sqrt{nT}} \text{tr} \left( M_{\Gamma_n} R_n G_n R_n^{-1} M_{\Gamma_n} \varepsilon \varepsilon' \right) - \frac{1}{\sqrt{nT}} \text{tr} \left( M_{\Gamma_n} R_n G_n R_n^{-1} M_{\Gamma_n} \varepsilon P_F \varepsilon' \right) - \sqrt{\frac{T}{nT}} \text{tr} \left( G_n \right) \sigma_0^2 \\
\frac{1}{\sqrt{nT}} \text{tr} \left( M_{\tilde{\Gamma}_n} G_n M_{\tilde{\Gamma}_n} \varepsilon \varepsilon' \right) - \frac{1}{\sqrt{nT}} \text{tr} \left( M_{\tilde{\Gamma}_n} \tilde{G}_n M_{\tilde{\Gamma}_n} \varepsilon P_F \varepsilon' \right) - \sqrt{\frac{T}{nT}} \text{tr} \left( \tilde{G}_n \right) \sigma_0^2 \\
0 \\
\end{pmatrix}
\]

where the results of Lemma 9 are used. Eq. (D.5) can be further simplified. Notice that

\[
\frac{1}{n} \text{tr} \left( G_n \right) \text{tr} \left( \varepsilon P_F \varepsilon' \right) - \text{tr} \left( M_{\Gamma_n} R_n G_n R_n^{-1} M_{\Gamma_n} \varepsilon P_F \varepsilon' \right) = \text{vec} \left( \varepsilon \right)' \mathcal{M} \text{vec} \left( \varepsilon \right),
\]

where \( \mathcal{M} = P_F \otimes \frac{1}{2} \left( \frac{2}{n} \text{tr} \left( G_n \right) I_n - M_{\Gamma_n} R_n G_n R_n^{-1} M_{\Gamma_n} - M_{\Gamma_n} R_n^{-1} G_n R_n M_{\Gamma_n} \right) \). \( \mathcal{M} \) is symmetric, \( \text{tr} \left( \mathcal{M} \right) = O(1) \), and \( \text{tr} \left( \mathcal{M}^2 \right) = O(n) \). We have \( \mathbb{E} \left( \text{vec} \left( \varepsilon \right)' \mathcal{M} \text{vec} \left( \varepsilon \right) \right)^2 = \left( \mu^4 - 3 \sigma_0^4 \right) \sum_{i=1}^{nT} \left[ \mathcal{M}^2 \right]_{ii} + 2 \sigma_0^4 \text{tr} \left( \mathcal{M}^2 \right) + \left( \sigma_0^2 \text{tr} \left( \mathcal{M} \right) \right)^2 = O(n) \). Therefore \( \frac{1}{\sqrt{nT}} \left[ \frac{1}{n} \text{tr} \left( G_n \right) \text{tr} \left( \varepsilon P_F \varepsilon' \right) - \text{tr} \left( M_{\Gamma_n} R_n G_n R_n^{-1} M_{\Gamma_n} \varepsilon P_F \varepsilon' \right) \right] = O_p \left( \frac{1}{\sqrt{nT}} \right) \). Similarly, we have \( \frac{1}{\sqrt{nT}} \left[ \frac{1}{n} \text{tr} \left( \tilde{G}_n \right) \text{tr} \left( \varepsilon P_F \varepsilon' \right) - \text{tr} \left( M_{\tilde{\Gamma}_n} \tilde{G}_n M_{\tilde{\Gamma}_n} \varepsilon P_F \varepsilon' \right) \right] = O_p \left( \frac{1}{\sqrt{nT}} \right) \). Putting the above two terms into the remainder, Eq. (D.5) becomes,

\[
\sqrt{nT} \left( \hat{\theta}_n - \theta_0 - \left( \sigma_0^2 D_{\eta T} \right)^{-1} \frac{1}{\sqrt{nT}} \Delta_{\eta T} \right) = \left( \sigma_0^2 D_{\eta T} \right)^{-1} \left( \frac{1}{\sqrt{nT}} \sum_{i=1}^{nT} \sum_{t=1}^{K} Z_{1, it} \varepsilon_{it} \right) + \left( \sigma_0^2 D_{\eta T} \right)^{-1} \left( \frac{1}{\sqrt{nT}} \text{tr} \left( R_n G_n R_n^{-1} \varepsilon \varepsilon' \right) - \frac{1}{\sqrt{nT}} \text{tr} \left( \varepsilon \varepsilon' \right) \frac{1}{n} \text{tr} \left( G_n \right) \right) + o_p(1),
\]

(D.6)
where

\[
\Delta_{nT} = \frac{1}{\sqrt{nT}} \begin{pmatrix}
-\text{tr} \left( M_{\bar{r}} R_{n} (Z_1 - \bar{Z}_1) P_{\bar{r}} \epsilon' \right) \\
-\text{tr} \left( M_{\bar{r}} R_{n} (Z_2 - \bar{Z}_2) P_{\bar{r}} \epsilon' \right) \\
0 \\
\vdots \\
0 \\
-\text{tr} \left( M_{\bar{r}} R_{n} (Z_{K+1} - \bar{Z}_{K+1}) P_{\bar{r}} \epsilon' \right) \\
0
\end{pmatrix}
\]

\[
+ \frac{1}{\sqrt{nT}} \begin{pmatrix}
0 \\
0 \\
\vdots \\
0 \\
-\text{tr} \left( P_{\bar{r}} R_n G_n R_n^{-1} \epsilon \epsilon' \right) - \text{tr} \left( R_n G_n R_n^{-1} P_{\bar{r}} \epsilon \epsilon' \right) + \text{tr} \left( P_{\bar{r}} R_n G_n R_n^{-1} P_{\bar{r}} \epsilon \epsilon' \right) + \frac{1}{n} \text{tr} \left( (P_{\bar{r}}) \epsilon \epsilon' \right) \\
-\text{tr} \left( P_{\bar{r}} \tilde{G}_n \epsilon \epsilon' \right) - \text{tr} \left( \tilde{G}_n P_{\bar{r}} \epsilon \epsilon' \right) + \text{tr} \left( P_{\bar{r}} \tilde{G}_n P_{\bar{r}} \epsilon \epsilon' \right) + \frac{1}{n} \text{tr} \left( \tilde{G}_n \epsilon \epsilon' \right)
\end{pmatrix}
\]

(D.7)

From (D.6) it is clear that the limiting distribution is not centered but with the bias \( \left( \sigma^2 D_{nT} \right)^{-1} \frac{1}{nT} \Delta nT \).

The bias arises from the predetermined regressors (D.7) and the interactions between spatial effects and the factor loadings (D.8). The bias terms would disappear if all regressors were strictly exogenous and no spatial effects were present.

Applying Lemmas 5 and 6, \( \Delta_{nT} = \varphi_{nT} + O_P \left( \frac{1}{\sqrt{n}} \right) \), where

\[
\varphi_{nT} = \begin{pmatrix}
-\frac{\sigma^3}{\sqrt{nT}} \sum_{h=1}^{T-1} \text{tr} \left( J_0 P_{\bar{r}} J_h' \right) \text{tr} \left( A_h^{-1} S_n^{-1} \right) \\
-\frac{\sigma^3}{\sqrt{nT}} \sum_{h=1}^{T-1} \text{tr} \left( J_0 P_{\bar{r}} J_h' \right) \text{tr} \left( W_n A_n^{-1} S_n^{-1} \right) \\
0 \\
\vdots \\
0 \\
-\frac{\sigma^3}{\sqrt{nT}} \sum_{h=1}^{T-1} \text{tr} \left( J_0 P_{\bar{r}} J_h' \right) \text{tr} \left( (\gamma G_n + \rho G_n W_n) A_n^{-1} S_n^{-1} \right) + \sqrt{\frac{T}{n}} \sigma^2 \left( \frac{\text{tr} \left( G_n \right) - \text{tr} \left( P_{\bar{r}} R_n G_n R_n^{-1} \right) }{n} \right) \\
\sqrt{\frac{T}{n}} \sigma^2 \left( \frac{\text{tr} \left( \tilde{G}_n \right) - \text{tr} \left( P_{\bar{r}} \tilde{G}_n \right) }{n} \right)
\end{pmatrix}
\]

(D.8)
\[ J_h = \left(0_{T \times (T-h)}, I_T, 0_{T \times h}\right)' \] for \( h = 0, \ldots, T-1 \), and \( I_T \times T \) is \( T \times T \) identity matrix. It follows that

\[
\sqrt{nT} (\hat{\theta}_{nT} - \theta_0) - (\sigma_0^2 D_{nT})^{-1} \phi_{nT} = (\sigma_0^2 D_{nT})^{-1} \begin{pmatrix} \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} Z_{1,i} \epsilon_{it} \\ \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} Z_{2,i} \epsilon_{it} \\ \vdots \\ \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} Z_{K,i} \epsilon_{it} \\ 0^{T-1} \end{pmatrix} + (\sigma_0^2 D_{nT})^{-1} \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \frac{1}{\sqrt{nT}} \text{tr} (R_n G_n R_n^{-1} \epsilon \epsilon') - \frac{1}{\sqrt{nT}} \text{tr} (\epsilon \epsilon') \frac{1}{\sqrt{m}} \text{tr} (G_n) \\ \frac{1}{\sqrt{nT}} \text{tr} (\tilde{G}_n \epsilon \epsilon') - \frac{1}{\sqrt{nT}} \text{tr} (\epsilon \epsilon') \frac{1}{\sqrt{m}} \text{tr} (\tilde{G}_n) \end{pmatrix} + o_p(1).
\]

**Proof of Theorem 2.**

The object of interest is \( c' v_{nT} = b_{nT}' \text{vec}(\epsilon) + \omega_{nT}' \text{vec}(\epsilon) + \text{vec}(\epsilon)' A_{nT} \text{vec}(\epsilon) \), where \( \omega_{nT} = \text{vec} \left( \sum_{h=1}^{n} P_n \tilde{\epsilon}_h \right) \) and \( A_{nT} \) is a nonstochastic symmetric matrix. From Yu et al. (2008), p.128, \( \mathbb{E} c' v_{nT} = \sigma_0^2 \text{tr} (A_{nT}) = 0 \),

\[
\text{var} (c' v_{nT}) = T \sigma_0^4 \mathbb{E} \text{tr} \left( \sum_{h=1}^{n} P_n \epsilon_{n'h} \epsilon_{n'h} \right) + \sigma_0^2 \mathbb{E} b_{nT}' \omega_{nT} + 2 \sigma_0^4 \text{tr} (A_{nT}^{-2}) + 2 \mu^3 \mathbb{E} \sum_{i=1}^{nT} [b_{nT}']_i [A_{nT}^{-1}]_{ii} + \left( \mu^4 - 3 \sigma_0^4 \right) \sum_{i=1}^{nT} [A_{nT}]_{ii}^2.
\]

Explicitly, \( \frac{1}{nT} \text{var} (c' v_{nT}) = \sigma_0^4 c' (\tilde{D}_{nT} + \Sigma_{nT} + o(1)) c \), where \( \tilde{D}_{nT} = \lim_{n,T \to \infty} \tilde{D}_{nT} \), and

\[
\tilde{D}_{nT} = \frac{1}{\sigma_0^2} \begin{pmatrix} \frac{1}{nT} \text{tr} (M_{T_n} R_n \tilde{Z}_1 M_{F'} \tilde{Z}_1' R_n') \cdots \frac{1}{nT} \text{tr} (M_{T_n} R_n \tilde{Z}_1 M_{F'} \tilde{Z}_K' R_n') \frac{1}{nT} \text{tr} (M_{T_n} R_n \tilde{Z}_1 M_{F'} \tilde{Z}_{K+1}' R_n') \\ \vdots \\ \frac{1}{nT} \text{tr} (M_{T_n} R_n \tilde{Z}_K M_{F'} \tilde{Z}_1' R_n') \frac{1}{nT} \text{tr} (M_{T_n} R_n \tilde{Z}_K M_{F'} \tilde{Z}_K' R_n') \frac{1}{nT} \text{tr} (M_{T_n} R_n \tilde{Z}_K M_{F'} \tilde{Z}_{K+1}' R_n') \\ \frac{1}{nT} \text{tr} (M_{T_n} R_n \tilde{Z}_{K+1} M_{F'} \tilde{Z}_1' R_n') \cdots \frac{1}{nT} \text{tr} (M_{T_n} R_n \tilde{Z}_{K+1} M_{F'} \tilde{Z}_K' R_n') \frac{1}{nT} \text{tr} (M_{T_n} R_n \tilde{Z}_{K+1} M_{F'} \tilde{Z}_{K+1}' R_n') \\ 0 \cdots 0 \frac{1}{nT} \text{tr} (M_{T_n} R_n \tilde{Z}_1 M_{F'} \tilde{Z}_1' R_n') \frac{1}{nT} \text{tr} (M_{T_n} R_n \tilde{Z}_1 M_{F'} \tilde{Z}_K' R_n') \\ 0 \cdots 0 \frac{1}{nT} \text{tr} (M_{T_n} R_n \tilde{Z}_K M_{F'} \tilde{Z}_1' R_n') \frac{1}{nT} \text{tr} (M_{T_n} R_n \tilde{Z}_K M_{F'} \tilde{Z}_K' R_n') \\ 0 \cdots 0 \frac{1}{nT} \text{tr} (M_{T_n} R_n \tilde{Z}_{K+1} M_{F'} \tilde{Z}_1' R_n') \frac{1}{nT} \text{tr} (M_{T_n} R_n \tilde{Z}_{K+1} M_{F'} \tilde{Z}_K' R_n') \\ 0 \cdots 0 \frac{1}{nT} \text{tr} (M_{T_n} R_n \tilde{Z}_{K+1} M_{F'} \tilde{Z}_{K+1}' R_n') \frac{1}{nT} \text{tr} (M_{T_n} R_n \tilde{Z}_{K+1} M_{F'} \tilde{Z}_{K+1}' R_n') \end{pmatrix}
\]

\[
+ \begin{pmatrix} \phi_{1,1} & \phi_{1,2} & 0 & \cdots & & \phi_{1,K+1} & 0 \\ \phi_{1,2} & \phi_{2,2} & 0 & \cdots & & \phi_{2,K+1} & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \psi_{K+1,K+1} & \psi_{K+1,K+2} \\ 0 & 0 & 0 & \cdots & \psi_{K+2,K+1} & \psi_{K+2,K+2} \end{pmatrix},
\]

and \( \phi_{1,1} = \frac{1}{nT} \text{tr} (R_n P_n R_n') \), \( \phi_{1,2} = \frac{1}{nT} \text{tr} (R_n P_n W_n' R_n') \), \( \phi_{2,2} = \frac{1}{nT} \text{tr} (R_n W_n P_n W_n' R_n') \), \( \phi_{1,K+1} = \frac{1}{nT} \text{tr} (R_n P_n (\gamma_0 G_n^* + \rho_0 W_n' G_n') R_n') \), \( \phi_{2,K+1} = \frac{1}{nT} \text{tr} (R_n W_n P_n (\gamma_0 G_n^* + \rho_0 W_n' G_n') R_n') \), \( \phi_{K+1,K+1} = \frac{1}{nT} \text{tr} (R_n (\gamma_0 G_n + \rho_0 G_n W_n) P_n (\gamma_0 G_n + \rho_0 W_n' G_n') R_n') \),
where \( \mathbf{P}_n = \sum_{k=0}^{n-1} A_{0n}^{-1} S_n^{-1} R_n^{-1} S_n^{-1} A_{0n}^{-1} \); and \( \psi_{K+1,K+1} = \frac{1}{n} \text{tr} \left( R_n G_n R_n^{-1} G_n R_n' \right) + \frac{1}{n} \text{tr} \left( G_n^2 \right) - 2 \left( \frac{1}{n} \text{tr} \left( G_n \right) \right)^2 \). 

\( \psi_{K+1,K+2} = \frac{1}{n} \text{tr} \left( G_n \tilde{G}_n \right) + \frac{2}{n} \text{tr} \left( G_n \text{tr}(\tilde{G}_n) \right) \) and \( \psi_{K+2,K+2} = \frac{1}{n} \text{tr} \left( \tilde{G}_n G_n' \right) + \frac{1}{n} \text{tr} \left( \tilde{G}_n^2 \right) - 2 \left( \frac{1}{n} \text{tr} \left( \tilde{G}_n \right) \right)^2 \).

\( \Sigma_{nT} \) is defined in Eq. (10).

Lemma 10 shows that \( D_{nT} = \tilde{D}_{nT} + o_P(1) \), therefore \( \frac{1}{nT} \text{var} \left( c' \nu_{nT} \right) = \sigma_0^2 c' \left( D_{nT} + \Sigma_{nT} + o_P(1) \right) c. \)

**Proof of Theorem 3.**

The theorem follows from Theorem 2 by applying the CLT of the martingale difference array.

**Proof of Theorem 4.**

We have the following useful results.

1. \( \sigma_0^2 \) can be estimated by \( L_{nT}(\hat{\theta}_{nT}) \). This is because from Lemmas 8, and \( \| \hat{\theta}_{nT} - \theta_0 \|_1 = O_P \left( \frac{1}{\sqrt{nT}} \right) \).

2. \( L_{nT}(\hat{\theta}_{nT}) = L_{nT}(\theta_0) + O_P \left( \frac{1}{\sqrt{nT}} \right) \). From Eq. (C.5), \( L_{nT}(\theta_0) = \sigma_0^2 + O_P \left( \frac{1}{\sqrt{nT}} \right) \).

3. By the mean value theorem, \( \hat{G}_n - \tilde{G}_n = \tilde{W}_n R_n (\alpha - \hat{\alpha}) \tilde{W}_n R_n (\alpha - \hat{\alpha})^{-1} \), for some \( \hat{\alpha} \) in between of \( \alpha_0 \) and \( \hat{\alpha} \). Because \( \| \tilde{W}_n \|_2 \) and \( \| R_{n-1} \|_2 \) are uniformly bounded, \( \| R_{n-1}(\alpha) \|_2 \) is uniformly bounded in a neighborhood of \( \alpha_0 \) by Lemma 1. \( \| \hat{G}_n - \tilde{G}_n \|_2 = O_P \left( \frac{1}{\sqrt{nT}} \right) \). Similarly, \( \| \hat{G}_n - G_n \|_2 = O_P \left( \frac{1}{\sqrt{nT}} \right) \) and \( \| \hat{R}_{n-1} - R_{n-1} \|_2 = O_P \left( \frac{1}{\sqrt{nT}} \right) \).

4. \( \| \hat{Z}_{K+1} - Z_{K+1} \|_2 = O_P(1) \), because \( \hat{Z}_{K+1} - Z_{K+1} = \sum_{k=1}^{K} \hat{\delta}_k \hat{G}_n Z_k - \sum_{k=1}^{K} \delta_k G_n Z_k = \sum_{k=1}^{K} \hat{\delta}_k (\hat{G}_n - G_n) Z_k + \sum_{k=1}^{K} (\hat{\delta}_k - \delta_k) G_n Z_k \), \( \| \hat{\delta} - \delta \|_1 = O_P \left( \frac{1}{\sqrt{nT}} \right) \), and from (3), \( \| \hat{G}_n - G_n \|_2 = O_P \left( \frac{1}{\sqrt{nT}} \right) \).

To prove Theorem 4, it is sufficient to show that \( \hat{\phi}_{nT} - \varphi \xrightarrow{p} 0 \), which immediately follows from (1) to (4) above. For example,

\[
\left| \text{tr} \left( P_{T_n} \hat{G}_n \right) - \text{tr} \left( P_{T_n} \tilde{G}_n \right) \right| \leq \left| \text{tr} \left( \left( P_{T_n} - P_{T_n} \right) \hat{G}_n \right) \right| + \left| \text{tr} \left( P_{T_n} \left( \hat{G}_n - \tilde{G}_n \right) \right) \right| \\
\leq 2r_0 \left\| P_{T_n} - P_{T_n} \right\|_2 \left\| \hat{G}_n \right\|_2 + r_0 \left\| P_{T_n} \right\|_2 \left\| \hat{G}_n - \tilde{G}_n \right\|_2 = O_P \left( \frac{1}{\sqrt{nT}} \right),
\]

where \( \left\| P_{T_n} - P_{T_n} \right\|_2 = O_P \left( \frac{1}{\sqrt{nT}} \right) \) by Lemma 11 and \( \left\| \hat{G}_n - G_n \right\|_2 = O_P \left( \frac{1}{\sqrt{nT}} \right) \) from (3).

Furthermore, \( \tilde{D}_{nT} = D_{nT} + o_P(1) \), because, for example, \( \frac{1}{nT} \left| \text{tr} \left( M_{T_n} R_{Z_k Z_k} M_{T_n} \hat{Z}_{K+1} \hat{R}_n' \right) - \text{tr} \left( M_{T_n} R_{Z_k Z_k} M_{T_n} \hat{Z}_{K+1} \hat{R}_n' \right) \right| = o_P(1) \) for \( k = 1, \ldots, K \) by (1) to (4) above. \( \frac{1}{nT} \left| \text{tr} \left( \hat{G}_n \right) - \left( G_n \right) \right| = O_P \left( \frac{1}{\sqrt{nT}} \right), \frac{1}{nT} \left| \text{tr} \left( \tilde{G}_n \right) - \left( G_n \right) \right| = O_P \left( \frac{1}{\sqrt{nT}} \right) \)

and \( \left| \text{tr} \left( \hat{R}_{n-1} P_{T_n} \tilde{R}_n \hat{G}_n \right) - \text{tr} \left( R_{n-1} P_{T_n} R_{n} G_n \right) \right| = O_P \left( \frac{1}{\sqrt{nT}} \right) \). Finally, \( \tilde{D}_{nT} = D_{nT} + o_P(1) \).

**Proof of Theorem 5.**

We show that Corollary 1 of Ahn and Horenstein (2013) holds by checking that their Assumption A-D are satisfied. Assuming that \( n \) and \( T \) are proportional, and the preliminary estimator satisfies \( \| \hat{\theta} - \theta_0 \|_2 = o_p \left( n^{-1/2} \right) \), we have,
Horenstein (2013)’s result applies here. which is satisfied by (3), and their Assumption A is satisfied by our Assumption SF. Therefore, Ahn and Horenstein (2013)’s result applies here.

\[ (1) \| \epsilon + \bar{E}_{nT}(\tilde{\theta}) \|_2 = O_p(\sqrt{n}), \text{because} \| \epsilon \|_2 = O_p(\sqrt{n}) \text{by Assumption E and} \| \bar{E}_{nT}(\tilde{\theta}) \|_2 \leq \sum_{k=1}^{K+2} \| \tilde{\eta}_k \|_2 \| V_k \|_2 + \sum_{k_1,k_2=1}^{K+2} \| \tilde{\eta}_{k_1} \|_2 \| \tilde{\eta}_{k_2} \|_2 \| V_{k_1,k_2} \|_2 = O_p(\sqrt{n}). \]

(2) \[ \mu_n \left( \frac{1}{n} (\epsilon + \bar{E}_{nT}(\tilde{\theta}))(\epsilon + \bar{E}_{nT}(\tilde{\theta}))' \right) \geq c + o_P(1) \text{for some positive constant } c. \text{This is because} \]

\[ \mu_n \left( \frac{1}{n} (\epsilon + \bar{E}_{nT}(\tilde{\theta}))(\epsilon + \bar{E}_{nT}(\tilde{\theta}))' \right) \geq \mu_n \left( \frac{1}{n} \epsilon \epsilon' \right) + \mu_n \left( \frac{1}{n} \bar{E}_{nT}(\tilde{\theta}) \bar{E}_{nT}(\tilde{\theta})' \right) + \mu_n \left( \frac{1}{n} \epsilon \bar{E}_{nT}(\tilde{\theta})' + \frac{1}{n} \bar{E}_{nT}(\tilde{\theta}) \epsilon' \right) \]

\[ \geq \mu_n \left( \frac{1}{n} \epsilon \epsilon' \right) - \frac{1}{n} \| \bar{E}_{nT}(\tilde{\theta}) \bar{E}_{nT}(\tilde{\theta})' \|_2 - \frac{1}{n} \| \epsilon \bar{E}_{nT}(\tilde{\theta})' + \bar{E}_{nT}(\tilde{\theta}) \epsilon' \|_2 \]

\[ = \mu_n \left( \frac{1}{n} \epsilon \epsilon' \right) + o_P(1) \geq c + o_P(1), \]

where the last inequality is from Lemma A.1 of Ahn and Horenstein (2013) (due to Bai and Yin (1993)).

(3) For any matrix \( A_{T \times q} = (a_1, \ldots, a_q) \) such that \( A'A = TI_q \),

\[ \frac{1}{n^3} \left| \text{tr} \left( A'(\epsilon + \bar{E}_{nT}(\tilde{\theta})) A \right) \right| \leq \frac{r_0}{n^3} \| AA' \|_2 \| F_T \bar{\Gamma}_n' \|_2 \| \epsilon + \bar{E}_{nT}(\tilde{\theta}) \|_2 = O_P \left( n^{-\frac{3}{2}} \right). \]

Similarly, we have

\[ \frac{1}{n^3} \left| \text{tr} \left( A'(\epsilon + \bar{E}_{nT}(\tilde{\theta}))' \bar{\Gamma}_n (\bar{\Gamma}_n' \bar{\Gamma}_n)^{-1} \bar{\Gamma}_n' (\epsilon + \bar{E}_{nT}(\tilde{\theta})) A \right) \right| \]

\[ \leq \frac{r_0}{n} \| AA' \|_2 \| \epsilon + \bar{E}_{nT}(\tilde{\theta}) \|_2^2 \| \bar{\Gamma}_n (\bar{\Gamma}_n' \bar{\Gamma}_n)^{-1} \bar{\Gamma}_n' \|_2 = O_P(n^{-1}). \]

In Ahn and Horenstein (2013), Assumptions C and D can be replaced by the conditions in their Eqs. (2) and (3), which are satisfied by (1) and (2) above. Their Assumption B is used to prove their Lemma A.10, which is satisfied by (3), and their Assumption A is satisfied by our Assumption SF. Therefore, Ahn and Horenstein (2013)’s result applies here. \( \square \)