A Spatial Panel Data Model with Time Varying Endogenous Weights Matrices and Common Factors

Wei Shi\textsuperscript{a,}\hspace{1em}*,\hspace{1em}Lung-fei Lee\textsuperscript{b}

\textsuperscript{a}Institute for Economic and Social Research, Jinan University, Guangzhou, Guangdong 510632, China.
\textsuperscript{b}Department of Economics, The Ohio State University, Columbus, Ohio 43210, United States.

Abstract

Many spatial panel data sets exhibit cross sectional and/or intertemporal dependence from spatial interactions or common factors. In an application of a spatial autoregressive model, a spatial weights matrix may be constructed from variables that may correlate with unobservables in the main equation and therefore is endogenous. Some common factors may be unobserved and correlate with included regressors in the equation. This paper presents a unified approach to model spatial panels with endogenous time varying spatial weights matrices and unobserved common factors. We show that the proposed QML estimator is consistent and asymptotically normal. As its limiting distribution may have a leading order bias, an analytical bias correction is proposed. Monte Carlo simulations demonstrate good finite sample properties of the estimators. This model is empirically applied to examine the effects of house price dynamics on reverse mortgage origination rates in the United States.

Keywords: Spatial panel data, endogenous spatial weighting matrix, multiplicative individual and time effects, QMLE, reverse mortgages

JEL classification: C13, C23, C51, G21

\hspace{1em}\hspace{1em}\hspace{1em}This version: 9/7/2016. We would like to thank Stephanie Moulton for sharing the HECM loan data, and participants at seminars in the Ohio State University, the 15th International Workshop on Spatial Econometrics and Statistics, the 2016 North American Summer Meeting of the Econometric Society, and the 2016 China Meeting of the Econometric Society for comments. The usual disclaimer applies. 
\hspace{1em}\hspace{1em}\hspace{1em}*Corresponding author.

Email addresses: shiweiemail@gmail.com (Wei Shi), lee.1777@osu.edu (Lung-fei Lee)
1. Introduction

The increasing availability of large panel data brings many opportunities that are not available with only cross sectional or time series data (see [Hsiao (2014)](#) for a good review). Individual heterogeneity can be better accounted for in a panel. In a two-way fixed effects model, observed regressors can correlate with an individual and time invariant component (the individual effect), and a common and time variant component (the time effect) of the error term. Recently the additive convention on individual and time effects is extended to allow for additional interactive effects, where unobserved time factors can have heterogeneous impacts on individuals (Pesaran (2006), Bai (2009), Ahn et al. (2013) and Moon and Weidner (2015)). In those models, the unobservables exhibit a factor structure and the covariates may correlate with these factors. The factors can capture omitted economy-wide shocks, etc, and the two-way fixed effects model is a special case of a general interactive fixed effects model.

Cross sectional dependence can be generated with unobserved time factors as they influence all cross sectional units. In addition, cross sectional dependence can also arise due to spatial effects. In terms of econometric modeling, spatial effects can be modeled with a spatially lagged dependent variable (Cliff and Ord (1973)). Spatial weights matrix provides a structure on spatial dependence and outcomes of individuals close by may be interdependent. Estimation methods of the spatial autoregressive (SAR) model have been well established for an exogenously given spatial weights matrix (see, inter alia, Kelejian and Prucha (2001), Kelejian and Prucha (2010), Lee (2004) and Yu et al. (2008)). In a SAR model, an explanatory variable may have both an direct effect on an individual’s outcome, as well as an indirect effect through spillovers on its neighbors. In addition, spatial correlation may be modeled with spatially dependent errors. For example, Pesaran and Tosetti (2011) consider cross sectionally dependent errors that have factor and spatial structures, which are interpreted, respectively, as strong and weak cross section dependence (also see Holly et al. (2010) and Holly et al. (2011) for empirical applications to house prices).

A spatial weights matrix can be based on geographic distances, or in many applications, economic theory may inform how individuals are connected with links constructed using economic variables (see the applications in Conley and Ligon (2002), Kelejian and Piras (2014)). However, when a spatial weights matrix is constructed from economic variables, the exogeneity assumption may not be satisfied if these included economic variables in the SAR equation correlate with disturbances of the equation. In a recent paper, Horrace et al. (2015) propose a network production model where a manager selects workers into the network and whose decision may depend on factors that correlate with the network unobserved characteristics in the production function. They show that the endogenous effect can be accounted for by using a Heckman type parametric or semi-parametric approach, or by including network fixed effects. Qu and Lee (2015) consider a cross section SAR model with an endogenously generated spatial weights matrix. Their approach caters to cases where individuals reside in a space, which is a common feature in many regional science applications. Using a control function approach, they provide consistent and asymptotically normal IV, QMLE and GMM estimators. Kelejian and Piras (2014) propose a 2SLS estimator for a panel data model with an endogenous spatial weights matrix and additive individual and time effects.

Panels with interactive effects also have wide empirical applications. For example, in regional policy evaluations, interactive effects can better control for time varying cross section heterogeneities (Gobillon
It has been used in the study of the effect of political and economic integration of Hong Kong with mainland China (Hsiao et al. (2012)), the effect of divorce law reforms on divorce rates (Kim and Oka (2014)), among others. As noted in Blundell et al. (2004), researchers need to weigh the benefits of using close neighbors as controls which are more likely to be otherwise the same as the treated, against the potential contaminating spatial spillover effects. In the regression panel with interactive effects, one might argue that spatial effects are absorbed into the factor loadings. However, this approach will likely be inadequate if spatial effects are time varying or endogenous. Furthermore, spatial correlation may be due to agents interacting with each other, but not just through spatial similarity in factor loading responses. In addition, the direct and indirect spillover effect of a policy may also be of interest.

This paper extends endogeneity of spatial weights in a cross section setting in Qu and Lee (2015) to a large $n, T$ panel data model with time varying spatial weights matrices and unobserved interactive individual and time effects. Bai and Li (2014) and Shi and Lee (2015) consider a panel SAR model with interactive effects and an exogenously given time invariant spatial weights matrix. In this paper, we have time varying endogenous spatial weights matrices constructed from variables that may correlate with the disturbances in the main equation. The interactive fixed effects are controlled for in equations and are treated as parameters. Even though spatial weights matrices are endogenous and time varying, interactive individual effects and time factors can be concentrated out. The profile objective function (concentrated likelihood function) includes a sum of certain eigenvalues of a random matrix. For a regression panel model, Moon and Weidner (2014, 2015) apply the perturbation method in Kato (1995) to derive its approximate gradient vector and Hessian matrix, so that asymptotic distributions of estimates of the regression coefficients can be derived. We show that the perturbation method can also be applied beyond regression panels to our model and that the QML estimator taking into account of endogenous spatial weights matrices is consistent and asymptotically normal.

We demonstrate the estimator’s good finite sample performance by a Monte Carlo study. We apply the model to analyze the effect of house price dynamics on reverse mortgage origination rates in the U.S. We find that in the reverse mortgage market, loan activity in one state influences activities in neighboring states, but that the spillover effect is smaller for states dominated by large lenders. This suggests that demand side factors such as common media exposure may be the driving force behind spatial spillovers in reverse mortgage activities.

This paper is organized as follows. We discuss our spatial panel model in the next section. We present the QML estimator in Section 3. Under some regularity assumptions, we investigate its asymptotic properties in Section 4 mainly its consistency and limiting distribution. As the limiting distribution may not be properly centered, an analytic bias correction is proposed. A Monte Carlo study is reported in Section 5. The empirical application on reverse mortgage origination rates is in Section 6. Section 7 concludes.

**Notation 1.** For a finite dimensional 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., $\|A\|_2 = \sqrt{\mu_1(A^TA)}$. In addition, $\|A\|_1 = \max_{1 \leq i \leq n} \sum_{j=1}^n |A_{ij}|$ is its maximum column sum norm, $\|A\|_\infty = \max_{1 \leq j \leq n} \sum_{i=1}^n |A_{ij}|$ is its maximum row sum norm, and $\|A\|_F = \sqrt{\text{tr}(AA^T)}$ is its Frobenius norm. Define the projection matrices $P_A = A(A^TA)^{-1}A'$ and $M_A = I - A(A^TA)^{-1}A'$. In general, lower case letters represent
vectors and upper case letters represent matrices.

2. The model

2.1. Model specification

This paper analyzes panel data where individuals are located in space through time. The data set contains \( n \) individuals indexed by \( i \), and each of them has observations for \( T \) time periods indexed by \( t \). Individuals are located on a possibly unevenly spaced lattice \( D \subset \mathbb{R}^{d_0} \) with \( d_0 \geq 1 \), which is time invariant and can base on geographic variables. The location \( l : \{1, \cdots, n\} \rightarrow D_n \subset D \) is a mapping of individual \( i \) to its location \( l(i) \in D \subset \mathbb{R}^{d_0} \). In addition, let \( D_T \subset \mathbb{Z} \) denote the set of time indexes. An individual \( i \) at time \( t \) has geographic location \( l(i) \) and time location \( t \). This is an adaptation of the topological structure of spatial processes in Jenish and Prucha (2009, 2012) and its extension by Qu and Lee (2015) to the panel data setting where the distance between \( it \) (individual \( i \) at time \( t \)) and \( js \) depends on both individual identities and time indexes. Define the metric \( \rho_{it, js} = \rho(i, js) = \max \{ \max_{1 \leq k \leq d_0} \{ |l(i)_k - l(j)_k| \}, |t - s| \} \) where \( l(i)_k \) is the \( k \)-th component of the \( d_0 \times 1 \) vector \( l(i) \). For asymptotic analysis, sample expansion is ensured by an unbounded expansion of the sample region (increasing domain asymptotics) along both the cross section and the time dimensions. Note that although individuals’ geographic locations are time invariant, the strength of a spatial link may depend also on economic variables, which can vary over time. Those will be made precise later.

Assumption 1. The lattice \( D \subset \mathbb{R}^{d_0} \), \( d_0 \geq 1 \), is infinitely countable. All elements in \( D \) are located at distances of at least \( \rho_0 > 0 \) from each other. Without loss of generality, we assume that \( \rho_0 = 1 \).

The equation of interest is

\[
y_{nt} = \lambda W_{nt}y_{nt} + X_{nty}\beta_y + \tilde{\Gamma}_{ny}f_{yt} + v_{nt},
\]

where \( y_{nt} \) is \( n \times 1 \); \( X_{nty} \) is \( n \times k_y \) matrix of \( k_y \) exogenous regressors. \( W_{nt} \) is a \( n \times n \) spatial weights matrix with elements denoted as \( w_{it, jt, nT} \). A spatial weights matrix measures degrees of connections between spatial units and is specified as nonnegative and with zero diagonals. Let \( y_{it} \) denote the \( i \)-th element of \( y_{nt} \). In Eq. 1, \( \lambda w_{it, jt, nT} \) measures the impact of \( y_{jt} \) on \( y_{it} \) and \( \lambda \) is the spatial interactions coefficient. Economic theories may suggest variables to be included and how the spatial weights matrix is constructed. For example, in the study of cross country growth spillovers (Conley and Ligon (2002)), a distance measure between two countries includes their geographic distance and transportation costs of goods and labor. In the analysis of cigarette demand (Kelejian and Piras (2014)), weights are based on the relative prices of cigarettes between two neighboring states. Essentially the spatial weights matrix provides information on how an individual is relatively affected by its neighbors. Note that individuals’ geographic locations in Assumption 1 may also limit degrees of spatial interactions among individuals (see Assumption 4). In the following, we drop the subscript \( nT \) in \( w_{it, jt, nT} \) and simplify it to \( w_{it, jt} \) when no confusion arises, while keeping in mind that elements like \( w_{it, jt} \) may depend on \( n \) and \( T \).

Remark 1. Eq. 1 can be interpreted as a best response function where individual \( i \)’s optimal action (e.g. contribution towards a public good) depends on its neighbors’ actions. The spatial interaction coefficient \( \lambda \) reflects the indirect effect of neighbors’ \( X_{nty} \), which can be seen from the reduced form of Eq. 1,
\[ y_{nt} = (I_n - \lambda W_{nt})^{-1} \left( X_{nt} \beta + \Gamma_{ny} \tilde{f}_{yt} + v_{nt} \right) = \sum_{s=0}^{\infty} \lambda^s W_{ts} \left( X_{nt} \beta + \Gamma_{ny} \tilde{f}_{yt} + v_{nt} \right), \]

assuming that \( I_n - \lambda W_{nt} \) is invertible and its inverse can be represented by a Neumann series. \( W_{ts} \) reflects the influence of \( s \)-th level neighbors. Also note that the model (1) and subsequent analysis can be easily extended to the case with multiple spatial weights matrices.

In an economic model, it is unlikely that included regressors capture all the information. When omitted variables exist and correlate with included regressors, a standard estimation method generally becomes invalid. In large panels with long time dimension (many time periods), to account for possible unobserved effects, we postulate that the error term in a SAR panel equation can be decomposed into a common factor component and an idiosyncratic component, where common factors can potentially correlate with included regressors. In Eq. (1), the dependent variable is affected by \( \tilde{R}_y \) unobserved factors, where the time factors, \( \tilde{f}_{yt} \), is \( R_y \times 1 \) and its loading matrix, \( \tilde{\Gamma}_{ny} \), is \( n \times R_y \) which allows time factors to have heterogeneous impacts on individuals. We treat factors and loadings as fixed effects to allow for hidden complex structures. \( v_{nt} = \left( v_{1t,nt} \cdots v_{nt,nt} \right)' \) is a \( n \times 1 \) vector of idiosyncratic shocks with zero mean and finite variance \( \sigma_v^2 \).

The factor structure is a flexible yet parsimonious way to model common shocks or a variance-covariance structure of a panel.

2.2. Sources of endogeneity

When the degree of spatial interaction is jointly determined with the outcome, \( W_{nt} \) becomes endogenous. Over time, the spatial weights matrix may also be time varying to reflect changing economic environments and network links. In this section, we illustrate the types of endogeneity covered in this paper.

The fixed effects formulation allows correlation between \( W_{nt} \) and \( y_{nt} \) due to the common factors. For example, the spatial weights can be influenced by the same time factors \( \tilde{f}_{yt} \) that impact \( y_{nt} \), i.e., \( w_{it,jt} = h \left( \gamma_{w,ij} \tilde{f}_{yt} + \varepsilon_{it,jt} \right) \) with conformable factor loading \( \gamma_{w,ij} \) and even \( \varepsilon_{it,jt} \) being independent from \( v_{nt} \), where \( h(\cdot) \) is some transformation. This can be covered once time factors are treated as fixed time effects in estimation. More interestingly, the spatial weights matrix may be constructed by variables that may correlate with disturbances of the outcomes. Suppose there are \( p \) variables, \( z_{it1}, \cdots, z_{itp} \) used to construct \( W_{nt} \). The equations for \( z_{itl}, l = 1, \cdots, p \), are

\[ z_{itl} = x_{itl}' \beta_{zl} + \gamma_{zl} \tilde{f}_{zt} + \varepsilon_{itl}, \]

where \( x_{itl} \) are \( k_{zl} \times 1 \) regressors with corresponding coefficient vector \( \beta_{zl} \), and the unobservables have two components, \( f_{zt} \) consists of \( R_z \times 1 \) time factors with loading \( \gamma_{zl} \) and \( \varepsilon_{itl} \) is idiosyncratic error. Stacking Eq. (2) across \( i \) and then over \( l \) for time period \( t \), we have \( z_{nt} = \left( z_{1t1}, \cdots, z_{nt1}, z_{1t2}, \cdots, z_{nt2}, \cdots, z_{ntp} \right)' \) an \( np \times 1 \) vector.
which has the structure,

$$z_{nt} = X_{ntz} \beta_{z} + \tilde{\Gamma}_{nz} f_{nt} + \varepsilon_{nt}, \text{ where } X_{ntz} = \begin{pmatrix} X'_{1t} \varepsilon_{1t}, \ldots, X'_{nt} \varepsilon_{nt} \end{pmatrix}. \tag{3}$$

is $np \times k_{z}$ with $k_{z} = k_{z1} + \cdots + k_{zp}$; $\beta_{z}$ is a column vector of dimension $k_{z}$, $\beta_{z} = (\beta'_{z1}, \ldots, \beta'_{zp})'$; $\tilde{\Gamma}_{nz} = (\tilde{\gamma}_{1z1}, \ldots, \tilde{\gamma}_{ tz1}, \tilde{\gamma}_{1z2}, \ldots, \tilde{\gamma}_{ tz2}, \ldots, \tilde{\gamma}_{ tznp})'$ is $np \times R_{z}$; and $\varepsilon_{nt} = (\varepsilon_{1t1}, \ldots, \varepsilon_{nt1}, \varepsilon_{1t2}, \ldots, \varepsilon_{ntp})'$ is an $np \times 1$ vector. Note that $X_{nty}$ and $X_{ntz}$ may have some common regressors. Let $w_{it,j} = h_{it,j}(z_{it}, \rho_{it,j})$ for $i, j = 1, \ldots, n, i \neq j$, where $h(\cdot)$'s are uniformly bounded functions. $w_{it,j}$ exhibits limited spatial dependence via distance $\rho_{it,j}$ which will be made precise in Assumption 6.

The following conditional moment assumption specifies that the disturbances in $z_{nt}$ may correlate with $v_{it}$ of the $y_{nt}$ equation. Indirectly, as $W_{nt}$ is a function of $z_{nt}$, the spatial weights may be endogenous. For an individual $i$, collect the error term $\varepsilon_{itp}$ in its $z_{itp}$ equation and define the $p \times 1$ vector $\varepsilon_{it} = (\varepsilon_{it1}, \ldots, \varepsilon_{itp})'$.

**Assumption 2.** The error terms $(v_{it}, \varepsilon_{it}')'$ are independently and identically distributed across $i$ and over $t$, and have a joint distribution $(v_{it}, \varepsilon_{it}') \sim (0, \Sigma_{ve})$, where $\Sigma_{ve} = \begin{pmatrix} \sigma_{v}^{2} & \sigma_{ve} \varepsilon_{v} \\ \sigma_{ve} \varepsilon_{v} & \sigma_{e}^{2} \end{pmatrix}$ is positive definite, $\sigma_{v}^{2}$ is a scalar variance, the covariance $\sigma_{ve} = (\sigma_{vr}, \ldots, \sigma_{vz})'$ is a $p \times 1$ vector, and $\Sigma_{e}$ is a $p \times p$ matrix. Furthermore, sup$_{n,T}$ sup$_{i,t} E |v_{it}|^{4+\delta_{v}}$ and sup$_{n,T}$ sup$_{i,t} E \|\varepsilon_{it}\|^{4+\delta_{e}}$ exist for some $\delta_{v}, \delta_{e} > 0$. Denote $E(v_{it} | \varepsilon_{it}) = \varepsilon_{it}' \delta$ and define $\xi_{it} = v_{it} - \varepsilon_{it}' \delta$. Assuming that $E(\xi_{it}^{2}) = E(\varepsilon_{it}^{2})$ and $E(\xi_{it}^{4}) = E(\varepsilon_{it}^{4})$.

Assumption 2 is used in Qu and Lee (2015) to model the endogeneity of the spatial weights matrix. If $\sigma_{ve} \neq 0$, the spatial weights matrix correlates with disturbances in the outcome equation of the same period and estimation methods that rely on exogenous spatial weights matrices might no longer be valid. As $E(\xi_{it} | \varepsilon_{it}) = 0$ and var$(\xi_{it} | \varepsilon_{it}) = \sigma_{e}^{2}$, the disturbances in the model can therefore be rewritten in terms of $\tilde{\xi}_{it} = (\tilde{\xi}_{it} \varepsilon_{it}')'$. Note that the conditional moment assumptions do not require $\tilde{\xi}_{it}$ to be independent from $\varepsilon_{it}$ as they are sufficient for our analysis. Denote $E_{nt} = (\varepsilon_{nt1}, \ldots, \varepsilon_{nTp})$ and $\Xi_{nt} = (\tilde{\xi}_{nt1}, \ldots, \tilde{\xi}_{nTp})$ with $\Xi_{nt} = (\tilde{\xi}_{nt1}, \ldots, \tilde{\xi}_{ntp})'$, respectively $np \times T$ and $n \times T$ matrices of the disturbances.

**Claim 1.** Under Assumption 2, $\|\Xi_{nt}\|_{2} = O_{P}\left(\sqrt{\max(n,T)}\right)$ and $\|E_{nt}\|_{2} = O_{P}\left(\sqrt{\max(n,T)}\right)$. The claim is proved in Lemma S.2.1 of Moon and Weidner (2014) using the result of Latala (2005).

3. Estimation method

The parameters of interest of the model consisting of Eqs. (1) and (3) are $\theta = (\beta'_{z}, \beta'_{y}, \lambda, \sigma_{e}^{2}, \alpha', \delta')'$, where $\beta_{z}$ and $\beta_{y}$ are respectively the regression coefficients in Eqs. (3) and (1), $\lambda$ is the spatial interaction...
coefficient, and $\sigma^2_\xi$, $\alpha$ and $\delta$ are from the variance structure of the disturbances with $\alpha$ being a $J \times 1$ vector of distinct elements in $\Sigma_e$. The unobserved time factors are collected in matrices $\bar{F}_{Ty} = (\bar{f}_{y1}, \cdots, \bar{f}_{yT})'$ and $F_{Tz} = (f_{z1}, \cdots, f_{zT})'$, which have dimensions $T \times \bar{R}_y$ and $T \times R_z$ respectively. In view of the unobserved nature of the common factor components ($\bar{\Gamma}_{nc}, \bar{\Gamma}_{ny}, F_{Tz}, \bar{F}_{Ty}$), we shall make minimal assumptions on their structures and allow the $z_{nt}$ and $y_{nt}$ equations to be impacted by possibly different factors. As the common factors and loadings are allowed to correlate with the regressors, they are treated as parameters to be estimated. The sample average log quasi-likelihood function is

$$\frac{1}{nT} \log L_{ntT}(\theta, \bar{\Gamma}_{nc}, \bar{\Gamma}_{ny}, F_{Tz}, \bar{F}_{Ty})$$

$$= -\frac{p+1}{2} \log (2\pi) - \frac{1}{2} \log |\Sigma_e| - \frac{1}{2} \log \sigma^2_\xi + \frac{1}{nT} \sum_{t=1}^T \log |S_{nt}(\lambda)|$$

$$- \frac{1}{2nT} \sum_{t=1}^T (z_{nt} - X_{nt}\beta_c - \bar{\Gamma}_{nc} f_{zt})' (\sigma^{-1}_\xi \otimes I_n) (z_{nt} - X_{nt}\beta_c - \bar{\Gamma}_{nc} f_{zt})$$

$$- \frac{1}{2\sigma^2_\xi nT} \sum_{t=1}^T (S_{nt}(\lambda)y_{nt} - X_{nty}\beta_y - (\delta' \otimes I_n) (z_{nt} - X_{nt}\beta_c - \bar{\Gamma}_{ny} f_{yt}) - \bar{\Gamma}_{ny} f_{yt})'$$

$$\times (S_{nt}(\lambda)y_{nt} - X_{nty}\beta_y - (\delta' \otimes I_n) (z_{nt} - X_{nt}\beta_c - \bar{\Gamma}_{ny} f_{yt}) - \bar{\Gamma}_{ny} f_{yt})$$

where $S_{nt}(\lambda) = I_n - \lambda W_{nt}$. Let $\Gamma_{nc} = \left( \Sigma_{x}^{-\frac{1}{2}} \otimes I_n \right) \bar{\Gamma}_{nc}$. The common component in the $y_{nt}$ equation is $\sigma^{-1}_\xi \bar{\Gamma}_{ny} f_{yt} - \sigma^{-1}_\xi (\delta' \otimes I_n) \bar{\Gamma}_{nc} f_{zt} = \Gamma_{ny} f_{yt}$ with the factor loading $\Gamma_{ny}$ appropriately constructed from $\sigma^{-1}_\xi \bar{\Gamma}_{ny}$ and $\sigma^{-1}_\xi (\delta' \otimes I_n) \bar{\Gamma}_{nc}$, and $f_{zt}$ collects the factors in $\bar{f}_{zt}$ and $f_{zt}$. For example, if $f_{zt} = f_{zt}$, $\Gamma_{ny} = \sigma^{-1}_\xi \left[ \Gamma_{ny} - (\delta' \otimes I_n) \bar{\Gamma}_{nc} \right]$ and $f_{zt} = f_{yt}$. The dimension of $f_{zt}$ is $R_y \times 1$, and write $F_{Ty} = (f_{y1}, \cdots, f_{yT})'$. Then the above equation can be written as

$$\frac{1}{nT} \log L_{ntT}(\theta, \Gamma_{nc}, \Gamma_{ny}, F_{Tz}, \bar{F}_{Ty})$$

$$= -\frac{p+1}{2} \log (2\pi) - \frac{1}{2} \log |\Sigma_e| - \frac{1}{2} \log \sigma^2_\xi + \frac{1}{nT} \sum_{t=1}^T \log |S_{nt}(\lambda)|$$

$$- \frac{1}{2nT} \sum_{t=1}^T \left[ \left( \Sigma_{x}^{-\frac{1}{2}} \otimes I_n \right) (z_{nt} - X_{nt}\beta_c) - \Gamma_{nc} f_{zt} \right]' \left[ \left( \Sigma_{x}^{-\frac{1}{2}} \otimes I_n \right) (z_{nt} - X_{nt}\beta_c) - \Gamma_{nc} f_{zt} \right]$$

$$- \frac{1}{2nT} \sum_{t=1}^T \left[ \sigma^{-1}_\xi (S_{nt}(\lambda)y_{nt} - X_{nty}\beta_y - (\delta' \otimes I_n) (z_{nt} - X_{nt}\beta_c)) - \Gamma_{ny} f_{yt} \right]'$$

$$\times \left[ \sigma^{-1}_\xi (S_{nt}(\lambda)y_{nt} - X_{nty}\beta_y - (\delta' \otimes I_n) (z_{nt} - X_{nt}\beta_c)) - \Gamma_{ny} f_{yt} \right]$$

(4) In Eq. (5), $(\delta' \otimes I_n) (z_{nt} - X_{nt}\beta_c)$ controls for the correlation between $z_{nt}$ and the disturbances in the $y_{nt}$ equation. As sample expands in $n$ and $T$, the number of parameters in factors and their loadings also increases. Because the parameter of interest is $\theta$, we concentrate out factors and their loadings using the following lemma.

**Lemma 1.** Let $n$, $T$, $R$ be positive integers such that $R \leq n$ and $R \leq T$. Let $Z_{nT}$ be an $n \times T$ matrix, $\Gamma_n$ be an $n \times R$ matrix, and $F_T$ be an $T \times R$ matrix. Then $\min_{\Gamma_n \in \mathbb{R}^{n \times R}, \Gamma_n \in \mathbb{R}^{n \times R}} \text{tr} \left( (Z_{nT} - \Gamma_n F_T)' (Z_{nT} - \Gamma_n F_T) \right) = \min_{\Gamma_n \in \mathbb{R}^{n \times R}} \text{tr} \left( Z_{nT}' M_{TnT} Z_{nT} \right) = \sum_{t=R+1}^T \mu_t (Z_{nT} Z_{nT}' + Z_{nT}' Z_{nT}).$
This lemma relates the estimation of factor loadings as the problem of principle components in statistics. The proof can be found, for example, in Lemma A1 of Moon and Weidner (2014). The concentrated log likelihood is:

\[
Q_{nT}(\theta) = -\frac{p + 1}{2} \log (2\pi) - \frac{1}{2} \log |\Sigma_e| - \frac{1}{2} \log \sigma_v^2 + \frac{1}{nT} \sum_{t=1}^{T} \log |S_{nt}(\lambda)| - \frac{1}{2} L_{nT}^z(\theta) - \frac{1}{2} L_{nT}^y(\theta),
\]

where

\[
L_{nT}^z(\theta) = \frac{1}{nT} \sum_{r=R_t+1}^{nT} \mu_r \left( \sum_{t=1}^{T} d_{nt} d_{nt}' \right), \quad \text{with} \quad d_{nt} = \left( \Sigma_e^{-\frac{1}{2}} \otimes I_n \right) (z_{nt} - X_{ntz} \beta), \quad \text{and},
\]

\[
L_{nT}^y(\theta) = \frac{1}{nT} \sum_{r=R_t+1}^{nT} \mu_r \left( \sum_{t=1}^{T} g_{nt} g_{nt}' \right), \quad \text{with} \quad g_{nt} = \sigma_v^{-1} (S_{nt}(\lambda) y_{nt} - X_{nty} \beta_y - (\delta' \otimes I_n) (z_{nt} - X_{ntz} \beta)).
\]

The QML estimator is \( \hat{\theta} = \arg \max_{\theta \in \Theta} Q_{nT}(\theta) \). The estimate \( \hat{\Gamma}_{ny} (\hat{\Gamma}_{nz}) \) for \( \Gamma_{ny} (\Gamma_{nz}) \) can be obtained as the eigenvectors associated with the first \( R_y \) (\( R_z \)) largest eigenvalues of \( \sum_{t=1}^{T} g_{nt} g_{nt}' \) (\( \sum_{t=1}^{T} d_{nt} d_{nt}' \)). By switching \( n \) and \( T \), the estimate \( \hat{F}_{Ty} (\hat{F}_{Tz}) \) for \( F_{Ty} (F_{Tz}) \) can be similarly obtained. Note that \( \hat{\Gamma}_{ny} \) and \( \hat{\Gamma}_{nz} \) (\( \hat{F}_{Ty} \) and \( \hat{F}_{Tz} \)) are not unique as \( \hat{\Gamma}_{ny} H H^{-1} \hat{F}_{Ty} \) is observationally equivalent to \( \hat{\Gamma}_{ny} \hat{F}_{Ty} \) for an invertible \( R_y \times R_y \) matrix \( H \). However, the column spaces of \( \hat{\Gamma}_{ny} \) and \( \hat{\Gamma}_{nz} \) (\( \hat{F}_{Ty} \) and \( \hat{F}_{Tz} \)) are invariant to \( H \), hence the projectors \( M_{\hat{\Gamma}_{ny}}, M_{\hat{\Gamma}_{nz}}, \) and \( M_{\hat{F}_{Ty}}, M_{\hat{F}_{Tz}} \) are uniquely determined.

Our model also covers two simpler cases of interest.

- **Seemingly unrelated regression (SUR) equations panel with factors**

By considering the equations of \( z_{nt} \) alone, one is considering the estimation of a \( p \) equation SUR model with unobserved common factors, which extends the single equation model (e.g., Pesaran (2006) and Bai (2009)) by allowing cross equation correlations. The likelihood component of \( z_{nt}, t = 1, \cdots, T \), is

\[
-\frac{p}{2} \log (2\pi) - \frac{1}{2} \log |\Sigma_e| - \frac{1}{2nT} \sum_{t=1}^{T} \left[ \left( \Sigma_e^{-\frac{1}{2}} \otimes I_n \right) (z_{nt} - X_{ntz} \beta) - \Gamma_{nz} f_{zt} \right]' \left[ \left( \Sigma_e^{-\frac{1}{2}} \otimes I_n \right) (z_{nt} - X_{ntz} \beta) - \Gamma_{nz} f_{zt} \right].
\]

- **Exogenous spatial weights matrices**

In the event that \( v_{it} \) is independent from \( e_{it}, \delta = 0 \) and the model becomes a spatial panel with unobserved common factors but weights matrices are exogenous. In this case, the outcome equation can be estimated separately from the \( z_{nt} \) equation. Imposing \( \delta = 0 \), the likelihood component of \( y_{nt}, t = 1, \cdots, T \), from (6) becomes

\[
-\frac{1}{2} \log (2\pi) - \frac{1}{2} \log \sigma_v^2 + \frac{1}{nT} \sum_{t=1}^{T} \log |S_{nt}(\lambda)|
\]

\[\text{In} \quad \text{Shi and Lee (2015), we also concentrate out the variance parameter. Because the correlations between equations are also of interest in this paper, the variance parameters are not concentrated out here. For example, when the disturbances are jointly normal and} \ p = 1, \ \text{the correlation coefficient} (\rho) \text{between} \ e_{it} \text{and} \ v_{it} \text{can be solved from} \ \delta = \rho \sigma_v \sigma_e^{-\frac{1}{2}} \text{and} \ \sigma_e^2 = (1 - \rho^2) \sigma_v^2.\]}
\[-\frac{1}{2nT}\sum_{t=1}^{T} \left[ \sigma_\xi^{-1}(S_n(\lambda)y_{nt} - X_{nt}\beta_\lambda) - \Gamma_{ny}f_{yt} \right]' \left[ \sigma_\xi^{-1}(S_n(\lambda)y_{nt} - X_{nt}\beta_\lambda) - \Gamma_{ny}f_{yt} \right], \]

with $\Gamma_{ny} = \sigma_\xi^{-1}\Gamma_{ny}$. If the spatial weights matrix is time invariant, the case is covered in Shi and Lee (2015). Here we allow for the more general case with time varying spatial weights matrices. Spatial panel models with time varying weights matrices have been considered in Lee and Yu (2012); however, those models have only additive individual and time effects but not interactive effects.

4. Asymptotic properties of the QML estimator

4.1. Assumptions

Assumption 3. (1) The parameter vector $\Theta$ is in a compact set $\Theta$ in the Euclidean space $\mathbb{R}^{k_\theta}$ with $k_\theta = k_z + k_y + 2 + J + p$, where $k_z$ is the dimension of $\beta_z$, $k_y$ is the dimension of $\beta_y$, $J$ is the dimension of $\alpha$ which contains distinct elements in $\Sigma_\varepsilon$, and $p$ is the dimension of $\delta$. The other two parameters are the spatial interaction coefficient $\lambda$ and the variance $\sigma_\xi^2$. The true parameter vector $\Theta_0$ is in the interior of $\Theta$.

(2) For any $i, j, n$ and $t$, the spatial weights $w_{it,ij} \geq 0, w_{it,ij} = 0$. Furthermore, $W_{nt}$ and $S_{nt}^{-1}$ are uniformly bounded in absolute value in both row and column sums, where $S_{nt} = I_n - \lambda_0 W_{nt}$. Furthermore, let $\sup_{nT} \|W_{nt}\| \leq c_w$ with probability 1 for some finite constant $c_w$ and $\sup_{\lambda \in A} |\lambda| c_w < 1$, where $A$ is the parameter space for $\lambda$.

(3) All elements in $X_{ntz}$ and $X_{nty}$ are deterministic and bounded in absolute value.

(4) There are respectively $R_{z0}$ and $R_{y0}$ factors in the $z_{nt}$ and $y_{nt}$ equations. The factors and their loadings are non-stochastic and uniformly bounded.

(5) $n$ is an increasing function of $T$ and $T$ goes to infinity.

Assumption 3(2) limits the spatial interaction to a manageable degree. The non-stochastic assumptions in (3) and (4) are used here to simplify the analysis in order to concentrate on the key issue of endogenous spatial dependence in the presence of common time factors, which generate another form of cross sectional dependence, but these assumptions can be relaxed by imposing moment conditions, see Moon and Weidner (2014). The asymptotics of this paper are derived with both $n$ and $T$ jointly increasing, as in (5). The following assumptions rule out collinearity among the regressors and are critical for parameter identification.

Notation 2. Define the following $n \times T$ dimensional matrices of regressors in the $z$ and $y$ equations. For $k = 1, \cdots, k_z$, $Z_{ntz,k} = (X_{n1z,k} \cdots X_{nTz,k})$ and $X_{ntz,k}$ is the $k$-th column of $X_{ntz}$. For $k = 1, \cdots, k_y$, $Z_{nty,k} = (X_{n1y,k} \cdots X_{nTy,k})$ and $X_{nty,k}$ is the $k$-th column of $X_{nty}$.

Assumption 4. There exists positive constants $c_z$, $c_y$, such that

\[
\min_{l_z \in B_{k_z}} \sum_{r=2k_z+1}^{n+k_z} \mu_r \left( \frac{1}{nT} \langle l_z \cdot Z_{ntz}^r \rangle \right) \geq c_z > 0, \\
\min_{l_y \in B_{k_y}} \sum_{r=2k_y+1}^{n+k_y} \mu_r \left( \frac{1}{nT} \langle l_y \cdot Z_{nty}^r \rangle \right) \geq c_y > 0,
\]

wpa 1 as $n,T \to \infty$, where $B_{k_z}$ ($B_{k_y}$) is a unit ball in the $k_z$ ($k_y + 1 + p$)-dimensional Euclidean space; $l_z$ ($l_y$) is a $k_z \times 1$ ($k_y + 1 + p \times 1$) vector with $\|l_z\|_2 = \|l_y\|_2 = 1$; furthermore, $l_z \cdot Z_{ntz}^r = \sum_{k=1}^{k_z} l_z(k)Z_{ntz,k}^r$ and $l_y \cdot Z_{nty}^r = \sum_{k=1}^{k_y} l_y(k)Z_{nty,k}^r$, where for $k = 1, \cdots, k_y$, $Z_{nty,k}^r = X_{nty,k}; Z_{ntz,k}^{k_z+1,T} = (G_{n1}(X_{n1z,\beta_0} + \tilde{\epsilon}_{n1} \delta_0 + \tilde{\Gamma}_{n0,\tilde{f}_{y10}}) \cdots G_{nT}(X_{nTy,\beta_0} + \tilde{\epsilon}_{nT} \delta_0 + \tilde{\Gamma}_{n0,\tilde{f}_{yT0}}))$ with $G_{nt} = W_{nt}S_{nt}^{-1}$.
and \( \mathbf{\tilde{e}}_{nt} = (\varepsilon_{nt1}, \cdots, \varepsilon_{ntp}) \) is \( n \times p \); and for \( k = 1, \cdots, p \), \( Z_{nt, k}^i = (\tilde{\varepsilon}_{n1,k}, \cdots, \tilde{\varepsilon}_{nt,k}) \) with \( \tilde{\varepsilon}_{nt,k} \) being the \( k \)-th column of \( \mathbf{\tilde{e}}_{nt} \).

Note that despite the simultaneity in the model, no exclusion restrictions are necessary, because of the nonlinear interactions of \( W_{it} \) and \( y_{it} \) and the nonlinearity of \( z_{it} \) in \( W_{it} \). Assumption 4 will not be satisfied if \( \text{rank} \left( Z_{nt}^\mathcal{I} \right) = 2R_y \), which can happen if \( \beta_0 = 0 \), \( \delta_0 = 0 \) and the spatial weights matrices are time invariant. This assumption will also be violated if \( G_{nt} \left( X_{nt} \beta_0 + \mathbf{\tilde{e}}_{nt} \delta_0 + \Gamma_{\eta_0} \phi_0 \right) \) is linearly dependent on \( X_{nt} \), and \( \mathbf{\tilde{e}}_{nt,k} \). We provide the following alternative conditions, which do not include the above term with \( G_{nt} \), but with additional condition on variance structures in terms of the reduced form disturbances. The later condition is motivated in [Lee (2004)], which is relevant for the identification of \( \lambda \).

**Assumption 5.** (1) There exists positive constants \( c_z, c_y \), such that

\[
\min_{l_z \in \mathbb{B}_{k_z}} \sum_{r = 2R_y + 1}^{n \times R_y} \mu_r \left( \frac{1}{nT} (l_z \cdot Z^z_n) (l_z \cdot Z^z_n) \right) \geq c_z > 0, \tag{5.1}
\]

\[
\min_{l_y \in \mathbb{B}_{k_y}} \sum_{r = 2R_y + 1}^{n \times R_y} \mu_r \left( \frac{1}{nT} (l_y \cdot Z^y_n) (l_y \cdot Z^y_n) \right) \geq c_y > 0, \tag{5.2}
\]

w.p.1 as \( n, T \to \infty \), where \( B_{k_z} \) is a unit ball in the \( k_z \)-dimensional Euclidean space; \( l_z \) is a \( k_z \times 1 \) (\( k_y \times 1 \)) nonzero vector with \( \|l_z\|_2 = \sqrt{T/l_z} = 1 \) (\( \|l_y\|_2 = 1 \)). Furthermore, \( l_z \cdot Z^z_n = \sum_{k=1}^{k_z} l_z \tilde{X}^z_{nt,k} \). In addition, \( l_y \cdot Z^y_n = \sum_{k=1}^{k_y} l_y \tilde{Z}^y_{k,nT} \). For \( k = 1, \cdots, k_y \), \( Z^y_{k,nT} = \tilde{X}^y_{nT,k} \); for \( k = 1, \cdots, p \), \( Z^z_{k+nT} = (\tilde{\varepsilon}_{n1,k}, \cdots, \tilde{\varepsilon}_{nt,k}) \) and \( \tilde{\varepsilon}_{nt,k} \) is the \( k \)-th column of \( \mathbf{\tilde{e}}_{nt} \), with \( \mathbf{\tilde{e}}_{nt} = (\varepsilon_{nt1}, \cdots, \varepsilon_{ntp}) \) an \( n \times p \) matrix.

(2) For any \( \lambda \in \Theta_\lambda \), \( \lambda \neq \lambda_0 \),

\[
\lim_{n,T \to \infty} \left( \frac{1}{nT} \sum_{t=1}^{T} \text{tr} \left( S_n(\lambda)' S_n(\lambda) S_n^{-1} S_n^{-1} \right) - \prod_{t=1}^{T} \left| \text{tr} \left( S_n(\lambda)' S_n(\lambda) S_n^{-1} S_n^{-1} \right) \right|^{\frac{1}{2}} \right) > 0. \tag{5.3}
\]

Assumption 5(2) ensures the identification of \( \lambda \). By the arithmetic and geometric means inequality,

\[
\frac{1}{nT} \sum_{t=1}^{T} \text{tr} \left( S_n(\lambda)' S_n(\lambda) S_n^{-1} S_n^{-1} \right) \geq \frac{1}{T} \sum_{t=1}^{T} \left| \text{tr} \left( S_n(\lambda)' S_n(\lambda) S_n^{-1} S_n^{-1} \right) \right|^{\frac{1}{2}} \geq \left( \prod_{t=1}^{T} \left| \text{tr} \left( S_n(\lambda)' S_n(\lambda) S_n^{-1} S_n^{-1} \right) \right|^{\frac{1}{2}} \right)^{\frac{1}{T}}.
\]

So this assumption essentially requires that the inequality holds strictly in the limit for \( \lambda \neq \lambda_0 \).

[Qu and Lee (2015)] provides a useful LLN for some key statistics for a cross sectional spatial model with an endogenous spatial weights matrix, which is derived using properties of spatial NED processes ([Jenish and Prucha (2012)]). That result will also be useful for our panel model by extending spatial location into the time-space location setting. Write \( z_{it} = (z_{i1}, \cdots, z_{ip}) \).

**Assumption 6.** (1) The spatial weights satisfy \( w_{it,jt} \geq 0 \), and \( w_{it,jt} = 0 \) if \( \rho_{it,j} > \rho_c \), i.e., there exists a threshold \( \rho_c > 1 \) such that the weight is zero if the geographic distance exceeds \( \rho_c \). For \( i \neq j \), \( w_{it,jt} = h_{ij}(z_{it}, z_{jt}) \) is the row normalized version that \( w_{it,jt} = h_{ij}(z_{it}, z_{jt})/\sum_{k=1}^{n} h_{kj}(z_{it}, z_{kt}) \), where \( h_{ij}(\cdot) \) are non-negative, uniformly bounded functions. (2) The function \( h_{ij}(\cdot) \) satisfies the Lipschitz condition, \( |h_{ij}(a_1, b_1) - h_{ij}(a_2, b_2)| \leq c_0 (|a_1 - a_2| + |b_1 - b_2|) \) for some finite constant \( c_0 \).

The underlying topological structure provides a bound on the degree of spatial dependence. Intuitively, this assumption requires that an individual is connected to other individuals at most \( \rho_c \) distance away and the functions that construct the spatial weights are uniformly bounded and smooth. Many functions satisfy
those properties. For the LLNs in Lemmas\(^2\) and \(^4\) below, the terms that involve the spatial weights need to satisfy the NED property, which can be derived from the NED property of \(z\) with smooth function \(h(\cdot)\)’s, as in Qu and Lee (2015) and Qu et al. (2015).

4.2. Consistency

To accommodate time varying spatial effects, time is considered as an additional dimension in an individual’s location vector, as Assumption \(1\) shows. Viewing individuals at different time points as different individuals, the panel data can be structured as a large collection of nodes with interactions via space-time locations. Define \(\tilde{G}_{nT} = \text{diag} (G_{n1}, \ldots, G_{nT})\) which is an \(nT \times nT\) diagonal block matrix with each block \(n \times n\) submatrix. Let \(\xi_{it}\) be a real valued function such that \(\xi_{it} = f_{it}(\xi_{it}, \theta_{it}, X_{it}, \Gamma_{nc}, \Gamma_{ny}, F_{Ty}, \theta_0)\), and denote \(\xi_{nT}^* = (\xi_{n1}^*, \ldots, \xi_{n1}^*, \xi_{n1}^*, \ldots, \xi_{nT}^*)'\) a \(nT \times 1\) vector. To prove the consistency of the QML estimator, the LLN and CLT in Qu and Lee (2015), Propositions 1 and 2, are reproduced here with modification for the panel data setting.

**Lemma 2.** Under Assumptions \(1\), \(3\)(2) and \(6\) suppose \(\sup_{n,T} \sup_{i,t} \mathbb{E} \|\xi_{it}^*\|_2^4 < \infty\), then

\[
\frac{1}{nT} \left( \xi_{nT}^* ' \tilde{G}_{nT} \xi_{nT}^* - \mathbb{E} \left( \xi_{nT}^* ' \tilde{G}_{nT} \xi_{nT}^* \right) \right) = O_p \left( \frac{1}{\sqrt{nT}} \right),
\]
and

\[
\frac{1}{nT} \left( \xi_{nT}^* ' \tilde{G}_{nT} \tilde{G}_{nT} \xi_{nT}^* - \mathbb{E} \left( \xi_{nT}^* ' \tilde{G}_{nT} \tilde{G}_{nT} \xi_{nT}^* \right) \right) = O_p \left( \frac{1}{\sqrt{nT}} \right).
\]

The above is analogue to a cross section data with a spatial weights matrix. Because the asymptotics are derived with both \(n\) and \(T\) increasing, the preceding result originated in Qu and Lee (2015) holds by considering an “extended” cross section data of \(nT\) individuals with the \(nT \times nT\) matrix \(\tilde{G}_{nT}\). It is straightforward to check that \(\tilde{G}_{nT}\) with Assumptions \(3\)(2) and \(6\) satisfy Qu and Lee (2015)’s Assumptions 3(1) and 4 for the LLN. The following results on matrix norms are useful.

**Claim 2.** Under Assumptions \(1\), \(2\), \(3\) and \(6\) \(\|\tilde{X}_{nT,k}\|_2 = O(\sqrt{nT})\) for \(k = 1, \ldots, k_z\), \(\|Z_{k,nT}\|_2 = O_p(\sqrt{nT})\) for \(k = 1, \ldots, k_z + 1\), and \(\|Z_{k+1+k,nT}\|_2 = O_p(\max(\sqrt{n}, \sqrt{T}))\) for \(k = 1, \ldots, p\), where \(\tilde{X}_{nT,k}\) and \(Z_{k,nT}\) are defined in Notation \(2\) and Assumption \(3\).

**Claim 3.** Under Assumptions \(2\) and \(3\) (1) and (2), \(\|U_{nT}\|_2 = O_p(\sqrt{nT} \left(n^{-\frac{1}{4}} + T^{-\frac{1}{4}}\right))\), where \(U_{nT} = (u_{n1}, \ldots, u_{nT})\) with \(u_{nt} = G_{nt} \xi_{nt}\) or \(u_{nt} = S_{nt}(\lambda) S_{nt}^{-1} \xi_{nt}\).

The following proposition establishes that the QML estimator is consistent under appropriate regularity conditions. Notice that the consistency only requires the number of factors specified in the model to be not fewer than the true number of factors. Lemma \(2\) is used to show that the sum of squared residuals in the likelihood is dominated by \(\|\theta - \theta_0\|_2^2\). Note that for this purpose the relevant terms do not involve projectors. When we later show that the variance of the estimator can be consistently estimated, Lemma \(2\) will be extended to allow for projectors, see Lemma \(4\). We can also establish a possible rate of convergence. The convergence rate can be improved to \(\frac{1}{\sqrt{nT}}\) if the number of factors is correctly specified, as shown in the next subsection.
Proposition 1. Under Assumptions 1, 3, 4 or 5 and assuming that the number of factors is not underestimated, i.e., $R_x \geq R_{x0}$ and $R_y \geq R_{y0}$, then $\|\hat{\theta} - \theta\|_2 = O_p(1)$. Furthermore, under Assumptions 1, 3, 4

$\begin{align*}
0 & \leq R_x \geq R_{x0} \text{ and } R_y \geq R_{y0}, \quad \|\hat{\beta}_x - \beta_{x0}\|_2 = O_p\left(\frac{c_{nT}^{-1}}{nT}\right), \\
\|\hat{\beta}_y - \beta_{y0}\|_2 = O_p\left(\frac{c_{nT}^{-1}}{nT}\right), \\
\|\hat{\lambda} - \lambda_0\| = O_p\left(\frac{c_{nT}^{-1}}{nT}\right)
\end{align*}$

The proof is in the appendix.

4.3. Asymptotic normality of the QML estimator

To derive the asymptotic distribution of the QML estimator, the mean value theorem or a Taylor expansion is often used to derive a proper linear-quadratic expansion of the log quasi-likelihood at their true parameter values. Such a technique is difficult for this model as the log likelihood involves the sum of certain eigenvalues of the sum of squared residuals. However, because the eigenvalues in the objective function are 0 if $\theta = \theta_0$, $E_{nT} = 0$ and $\Sigma_{nT} = 0$, and they will be close to 0 when small perturbations are applied, the perturbation theory of linear operator are useful (see Kato (1995)). The perturbations are small asymptotically under assumptions on error disturbances and the consistency of the QML estimator. Moon and Weidner (2015) use the perturbation method to obtain an approximate gradient vector and an approximate Hessian matrix for a regression panel with factors. Shi and Lee (2015) apply the method to a QML objective function of a spatial panel data model. Here we extend our analysis in Shi and Lee (2015) to Eq. (6) by taking into account the presence of endogenous and time varying spatial weights matrices. An additional feature is the presence of variance matrix $\Sigma$ of the log likelihood function. In order to utilize the perturbation theory, we adopt the following reparameterization.

Assumption 8. Let $\Sigma_{ef}^{-\frac{1}{2}}$ represent the principal square root of $\Sigma_{ef}^{-1}$ and it can be linearly parameterized, as $\Sigma_{ef}^{-\frac{1}{2}} = \sum_{j=1}^J \Omega_j \alpha_j$ where each $\Omega_j$ does not involve unknown parameters, $\Omega_1, \cdots, \Omega_J$ are linearly independent, and the $\alpha$’s are $J$ unique unknown elements of $\Sigma_{ef}^{-\frac{1}{2}}$.

Example 1. Supposing $p = 2$ and no further restrictions are imposed on $\Sigma_{ef}^{-\frac{1}{2}}$, then $\Sigma_{ef}^{-\frac{1}{2}}$ has three distinct elements. So one way to parameterize $\Sigma_{ef}^{-\frac{1}{2}}$ is $\Sigma_{ef}^{-\frac{1}{2}} = \sum_{j=1}^3 \Omega_j \alpha_j$ with $\Omega_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, $\Omega_2 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$ and $\Omega_3 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.

Because the common factors and loadings are concentrated out in estimation, in the following, $\bar{F}_{nz}$, $\bar{F}_{ny}$ and $\bar{F}_{Tz}$ refer to their true values in the data generating process, as they are parts of the DGP for asymptotic analysis. The asymptotic distribution of the QMLE is derived under the asymptotics that $\lim_{n,T \to \infty} \frac{n}{T} \to \kappa > 0$.

Assumption 8. The numbers of factors, $R_{x0}$ and $R_{y0}$ are constants and known. $\Gamma_{nz}$, $\Gamma_{ny}$ and $\Gamma_{Tz}$ as defined in Eqs. 4 and 5 are nonstochastic. $\lim_{n \to \infty} \frac{1}{n} \Gamma_{nz}^\top \Gamma_{ny} = \Gamma_y$, $\lim_{n \to \infty} \frac{1}{n} \Gamma_{ny}^\top \Gamma_{nz} = \Gamma_z$, $\lim_{T \to \infty} \frac{1}{T} F_{nz}^\top F_{Tz} = F_z$ and $\lim_{T \to \infty} \frac{1}{T} F_{ny}^\top F_{Tz} = F_y$ exist and are positive definite.

The above assumption which has also been used in Bai (2009) and Moon and Weidner (2015), states that each factor contributes a nontrivial share towards the common factor component. When normalized
by the sample size $nT$, this assumption ensures that the eigenvalues from the common factor component are bounded away from 0 while those from the idiosyncratic errors are close to 0, and thus they are well separated.

Firstly consider $L_{nT}^z(\theta)$ in Eq. (6) for the spatial weights formation. Let $D_{nT} = (d_{n1}, \ldots, d_{nT})$ with $d_{nt} = \left(\sum_{r=0}^{\frac{1}{n^2}} I_n\right)(z_{nt} - X_{nt}^*\beta_z)$. We have $D_{nT} = \Gamma_{nc}F_{Tz} + \sum_{k=1}^{\frac{1}{n^2}+1} \eta_{k}^z\Sigma_{k} + \sum_{k_1,k_2=1}^{\frac{1}{n^2}+1} \eta_{k_1}^z\eta_{k_2}^zV_{k_1,k_2}^z$ where

1. For $k = 1, \ldots, k_z$, $\eta_{k}^z = \beta\overset{\text{def}}{=} \beta_0 - \beta_z$, $V_{k}^z = \left(\sum_{r=0}^{\frac{1}{n^2}} I_n\right)X_{nTz,k}$.
2. For $k = 1, \ldots, J$, $\eta_{k}^{z+k} = \alpha_0 - \alpha_k$ and $V_{k}^{z+k} = - (\Omega_k \otimes I_n) \left(E_{nT} + \tilde{\Gamma}_{nz}F_{Tz}\right)$.
3. $\eta_{k}^{z+j+1} = \frac{\|\varepsilon_{nt}\|}{\sqrt{nT}}$ and $V_{k}^{z+j+1} = \left(\sum_{r=0}^{\frac{1}{n^2}} I_n\right)\left(\frac{E_{nT}}{\sqrt{nT}}\right)$.
4. For $k_1 = 1, \ldots, k_z$ and $k_2 = 1, \ldots, J, V_{k_1,k_2}^{z} = - (\Omega_{k_2} \otimes I_n) \tilde{X}_{nTz,k_1}$ and $V_{kk'}^{z} = 0$ otherwise.

The component under analysis here is a SUR panel with $p$ equations and correlations between the equations. Items (2) and (4) come from the variance. It extends a single regression panel with factors in Moon and Weidner (2015). Then, we have

$$L_{nT}^z(\theta) = \frac{1}{nT} \sum_{r=0}^{nT} \mu_r(D_{nT}D_{nT}^*) = \frac{1}{nT} \sum_{r=0}^{nT} \mu_r\left(T^{(0)} + \sum_{k_1=1}^{\frac{1}{n^2}+1} \eta_{k_1}^z T_{k_1}^{(1)} + \sum_{k_1,k_2=1}^{\frac{1}{n^2}+1} \eta_{k_1}^z \eta_{k_2}^z T_{k_1,k_2}^{(2)} + \sum_{k_1,k_2,k_3=1}^{\frac{1}{n^2}+1} \eta_{k_1}^z \eta_{k_2}^z \eta_{k_3}^z T_{k_1,k_2,k_3}^{(3)} + \sum_{k_1,k_2,k_3,k_4=1}^{\frac{1}{n^2}+1} \eta_{k_1}^z \eta_{k_2}^z \eta_{k_3}^z \eta_{k_4}^z T_{k_1,k_2,k_3,k_4}^{(4)}\right),$$

where $T^{(0)} = \Gamma_{nc}F_{Tz}^*F_{Tz} \Gamma_{nc}^*$, $T_{k_1}^{(1)} = V_{k_1}^z V_{k_1}^z$, $T_{k_1}^{(2)} = V_{k_1}^z V_{k_1}^z + V_{k_1}^z V_{k_1}^z$, $T_{k_1,k_2}^{(3)} = V_{k_1}^z V_{k_2}^z + V_{k_2}^z V_{k_1}^z$, and $T_{k_1,k_2,k_3}^{(4)} = V_{k_1}^z V_{k_2}^z V_{k_3}^z$. Under Assumption 8, $\frac{1}{nT} \sum_{r=0}^{nT} \mu_r(T^{(0)}) = 0$, because $\frac{1}{nT} T^{(0)}$ has exactly $R_0$ positive eigenvalues and the remaining ones are zeros. For small $\eta_{k}^z$, $L_{nT}^z(\theta)$ can be approximated by a quadratic polynomial of $\eta_{k}^z$. Using Lemma D.1 and proper but simplified expressions of Lemma D.2 in [Shi and Lee (2015)], we have

$$L_{nT}^z(\theta) = L_{nT}^z(\theta_0) + 2 \sum_{k=1}^{\frac{1}{n^2}+1} \eta_{k}^z \frac{1}{nT} \text{tr}\left(M_{nc} \left(\sum_{r=0}^{\frac{1}{n^2}} I_n\right) E_{nT} M_{Fz} V_{k}^z\right)$$

$$- 2 \sum_{k=1}^{\frac{1}{n^2}+1} \eta_{k}^z \frac{1}{nT} \text{tr}\left(M_{nc} V_{k}^z M_{Fz} E_{nT} \left(\sum_{r=0}^{\frac{1}{n^2}} I_n\right) E_{nT} M_{Fz} V_{k}^z\right)$$

$$- 2 \sum_{k=1}^{\frac{1}{n^2}+1} \eta_{k}^z \frac{1}{nT} \text{tr}\left(M_{nc} \left(\sum_{r=0}^{\frac{1}{n^2}} I_n\right) E_{nT} M_{Fz} V_{k}^z P_{nc,Fz} E_{nT} \left(\sum_{r=0}^{\frac{1}{n^2}} I_n\right) P_{nc,Fz} V_{k}^z\right)$$

$$- 2 \sum_{k=1}^{\frac{1}{n^2}+1} \eta_{k}^z \frac{1}{nT} \text{tr}\left(M_{nc} \left(\sum_{r=0}^{\frac{1}{n^2}} I_n\right) E_{nT} M_{Fz} E_{nT} \left(\sum_{r=0}^{\frac{1}{n^2}} I_n\right) P_{nc,Fz} V_{k}^z\right)$$

$$+ \sum_{k_1,k_2=1}^{\frac{1}{n^2}+1} \eta_{k_1}^z \eta_{k_2}^z \frac{1}{nT} \text{tr}\left(M_{nc} V_{k_1}^z M_{Fz} V_{k_2}^z\right) + L_{nT}^{\text{rem}}(\theta),$$

where $M_{nc}$ and $M_{Fz}$ are the projection matrices, $P_{nc,Fz} = \Gamma_{nc} \left(\Gamma_{nc}^* \Gamma_{nc}\right)^{-1} \left(F_{Tz}^* F_{Tz}\right)^{-1} F_{Tz}$, and $L_{nT}^{\text{rem}}(\theta)$ captures remainder terms that can be shown to be $O_P\left(\|\theta - \theta_0\|^2_2\right) + O_P\left(\|\theta - \theta_0\|^2_2 n^{-\frac{1}{2}}\right) + O_P\left(\|\theta - \theta_0\|^2_2 n^{-\frac{1}{2}}\right) + O_P\left(\|\theta - \theta_0\|^2_2 n^{-\frac{1}{2}}\right)$. 

\( O_p \left( n^{-\frac{3}{2}} \right) \). Note that the perturbations that are purely due to the idiosyncratic errors (\( \eta_{k_j+1}^i \bar{V}_{k_j+1} \)) are collected in \( L_{nT}^z (\theta_0) \), as

\[
L_{nT}^z (\theta_0) = \frac{1}{nT} \sum_{r=R_{0T}+1}^n \mu_r \left( \Gamma_{nyF_{rT}G_{ny}} + \left( \frac{1}{2} \sigma_0^{-1} \otimes I_n \right) E_{nT} F_{rT} \right) + \left( \frac{1}{2} \sigma_0^{-1} \otimes I_n \right) E_{nT} F_{rT} \left( \frac{1}{2} \sigma_0^{-1} \otimes I_n \right).
\]

Next consider \( g_{nt} = \sigma_{\xi}^{-1} (S_{nt} (\lambda) y_{nt} - X_{nt} \beta (\delta' \otimes I_n) (z_{nt} - X_{nt} \beta_z) \right) \) in Eq. (6) for the outcome equation. By reparameterization, let \( \alpha_\xi = \sigma_{\xi}^{-1} \). Define the following terms:

1. For \( k = 1, \ldots, k, \eta_{x_{k_1+1}}^y = \beta_{0,k} - \beta_{z,k} \) and \( V_k^y = -\sigma_{\xi}^{-1} (\delta_k' \otimes I_n) \tilde{X}_{nT,z,k} \).
2. For \( k = 1, \ldots, k, \eta_{x_{k_1+1}}^y = \beta_{0,k} - \beta_{y,k} \) and \( V_{k_1+k}^y = \sigma_{\xi}^{-1} \tilde{X}_{nT,y,k} \).
3. \( \eta_{x_{k_1+k_1+1}}^y = \lambda_0 - \lambda, V_{k_1+k_1+1}^y = \sigma_{\xi}^{-1} \tilde{X}_{nT,y,k_1} \) with the \( t \)-th column of \( \tilde{X}_{nT,y,k_1} \) given by \( X_{nT,y,k_1} = G_{nt} X_{nT,y,k_1} + \tilde{G}_{nt} \tilde{X}_{nT,y,k_1} \).
4. \( \eta_{x_{k_1+k_1+2}}^y = \alpha_{\xi} - \alpha_\xi \) and \( V_{k_1+k_1+2}^y = -\pi_{nt} \sigma_{\xi}^{-1} \Gamma_{ny} F_{nT} y_{k_1+k_1+2} \\)
5. For \( k = 1, \ldots, p, \eta_{x_{k_1+k_1+2+k}}^y = \delta_{0,k} - \delta_k, \) define \( e_k \) as a \( p \times 1 \) vector whose \( k \)-th element is 1 and all others are 0 and \( V_{k_1+k_1+2+k}^y = \sigma_{\xi}^{-1} (e_k' \otimes I_n) \left( \sum_{k=1}^{k_1+k_1+2+k} \tilde{X}_{nT,z,k} (\beta_{0,k} - \beta_{z,k}) + \tilde{G}_{nt} \tilde{X}_{nT,y,k} \right) \).

Item (3) comes from the spatial interaction and (5) comes from the control function. Similarly, the perturbation theory gives

\[
L_{nT}^y (\theta) = L_{nT}^y (\theta_0) + \sum_{k=1}^{k_1+k_1+2+p} \eta_{x_{k_1+k_1+2+p}}^y \ \frac{2}{nT \sigma_0^2} \text{tr} (M_{nt} \Gamma_{yx} M_{Fy} V_k^{y'}) - \sum_{k=1}^{k_1+k_1+2+p} \eta_{x_{k_1+k_1+2+p}}^y \ \frac{2}{nT \sigma_0^2} \text{tr} (M_{nt} \Gamma_{yx} M_{Fy} V_k^{y'}) + \frac{1}{nT} \text{tr} (M_{nt} \Gamma_{yx} M_{Fy} V_k^{y'}) + L_{nT}^{\text{rem,}y} (\theta),
\]

where

\[
L_{nT}^y (\theta_0) = \frac{1}{nT} \sum_{r=R_{0T}+1}^n \mu_r \left( \Gamma_{nyF_{rT}G_{ny}} + \sigma_{\xi}^{-1} \Gamma_{nyF_{rT}G_{ny}} + \sigma_{\xi}^{-1} \Gamma_{nyF_{rT}G_{ny}} + \sigma_{\xi}^{-1} \Gamma_{nyF_{rT}G_{ny}} \right),
\]

and the remainder term satisfies \( L_{nT}^{\text{rem,}y} (\theta) = O_P \left( \frac{|\theta - \theta_0|}{n^2} \sum_{x_{k_1+k_1+2}}^2 \right) + O_P \left( \frac{|\theta - \theta_0|^2}{n^{1/2}} \right) \). In Eq. (9), the term \( (\lambda_0 - \lambda) (\alpha_{\xi} - \alpha_\xi) \) is the interaction between the spatially generated regressor \( \tilde{X}_{nT,y,k} \) and the error term, while the interaction between the exogenous regressors and the error term can be absorbed into \( O_P \left( \frac{|\theta - \theta_0|^3}{n^{1/2}} \right) \) in the remainder.

The total number of parameters in \( \theta \) is \( k = k_x + k_y + 2 + J + p \). Substituting Eqs. (8) and (9) into the QML objective function (6), and expanding \( \log \left| \Sigma_x \right| \), \( \log \sigma_x^{-1} \), and \( \text{log} |S_{nt} (\lambda)| \) around their respective true
values, we have

\begin{equation}
Q_{nT}(\theta) = -\frac{D+1}{2} \log 2\pi - \frac{1}{2} \left( \log \sigma_{\xi_0}^2 - 2\sigma_{\xi_0} (\alpha_\xi - \alpha_{\xi_0}) + \sigma_{\xi_0}^2 (\alpha_\xi - \alpha_{\xi_0})^2 \right) \\
- \frac{1}{2} \left( \log |\Sigma_{\varepsilon_0}| - 2 \sum_{j=1}^J \text{tr} \left( \Sigma_{\varepsilon_0}^{-1} \Omega_j \right) (\alpha_j - \alpha_0) + \sum_{j,k=1}^J \text{tr} \left( \Sigma_{\varepsilon_0}^{-1} \Omega_j \Sigma_{\varepsilon_0}^{-1} \Omega_k \right) (\alpha_j - \alpha_0)(\alpha_k - \alpha_0) \right) \\
+ \frac{1}{nT} \sum_{t=1}^T \left( \log |S_{nt}| - \text{tr}(G_{nt}) (\lambda - \lambda_0) - \frac{1}{2} \text{tr}(G_{nt}^2) (\lambda - \lambda_0)^2 \right) \\
- \frac{1}{2} L_{nT}^Y(\theta_0) - \frac{1}{2} L_{nT}^Y(\theta_0) + \frac{1}{\sqrt{nT}} (\theta - \theta_0)' C_{nT}^{(1)} - \frac{1}{2} (\theta - \theta_0)' C_{nT} (\theta - \theta_0) + Q_{nT}^{\text{rem}}(\theta) \notag
\end{equation}

where \( C_{nT}^{(1)} \) and \( \tilde{C}_{nT}^{(1)} \) are \( k \times k \) matrices, and the remainder term term satisfies

\begin{equation}
Q_{nT}^{\text{rem}}(\theta) = O_P \left( \left\| \theta - \theta_0 \right\|_2^2 \right) + O_P \left( \left\| \theta - \theta_0 \right\|_2^2 n^{-\frac{3}{2}} \right) + O_P \left( \left\| \theta - \theta_0 \right\|_2 n^{-\frac{3}{2}} \right) + O_P \left( n^{-\frac{3}{2}} \right).
\end{equation}

The detailed expressions can be found in the appendix.

Define \( \gamma_{nT} = \frac{1}{\sqrt{nT}} \tilde{C}_{nT}^{-1} \tilde{C}_{nT}^{(1)} \), \( \tilde{e}_{nT} \) is defined in Proposition 2 below is the leading order bias term. The limiting distribution of the random vector \( \tilde{C}_{nT}^{(1)} \) can be derived using the Cramér-Wold device. Let \( v \) be an arbitrary conformable vector, which gives a linear combination \( R_{nT}^V = v' \tilde{C}_{nT}^{(1)} \) of columns of \( C_{nT}^{(1)} \). \( R_{nT}^V \) can be expressed in the following general form,

\begin{equation}
R_{nT}^V - \varphi_{nT}^V = \frac{1}{\sqrt{nT}} \left( a_{nT}' \tilde{e}_{nT} + \tilde{e}_{nT}' A_{nT} \tilde{e}_{nT} + b_{nT}' \tilde{e}_{nT} + \tilde{e}_{nT}' B_{nT} \tilde{e}_{nT} + \tilde{e}_{nT}' D_{nT} \tilde{e}_{nT} - \text{tr} \left( A_{nT} \right) - \text{tr} \left( B_{nT} \right) \right) + o_P(1),
\end{equation}

where \( \varphi_{nT}^V = v' \varphi_{nT} \) with \( \varphi_{nT} \) defined in Proposition 2 below is the leading order bias term. The \( b_{nT}^v \) and \( B_{nT}^v \) are in general functions of \( \tilde{e}_{nT} \), while \( a_{nT}^v, A_{nT}^v \) and \( D_{nT}^v \) are constant vectors or matrices. The limiting distribution of \( R_{nT}^V - \varphi_{nT}^V \) has been derived in [Qu et al. (2015)] using martingale differences CLT.

**Lemma 3.** Let \( \sigma_{nT}^2 \) denote the variance of \( R_{nT}^V - \varphi_{nT}^V \). Assuming that \( \sigma_{nT}^2 \) is bounded away from 0 and Assumptions 3 and 6 hold, then \( \frac{R_{nT}^V - \varphi_{nT}^V}{\sigma_{nT}} \overset{d}{\rightarrow} N(0, 1). \)

To show the probability limit of its variance, appropriate LLNs are needed. In [Qu et al. (2015)], in order to concentrate out additive individual and time effects, it gives rise to statistics of variables transformed by the projectors \( I_n - \frac{1}{n} \ell_n \ell_n' \) and \( I_T - \frac{1}{T} \ell_T \ell_T' \) where \( \ell_n \) and \( \ell_T \) are vectors of 1’s. Because individual and/or time observations of spatial units are averaged over space and/or time, the LLN in preceding Lemma 2 is not applicable, and they establish additional LLN in their Proposition 1. In this paper with interactive effects instead of additive effects, the projectors of loadings and factors are more general. However, \( M_{\Gamma_{ny}} = I_n - \frac{1}{n} \Gamma_{ny} \left( \frac{1}{n} \Gamma_{ny} \Gamma_{ny} \right)^{-1} \Gamma_{ny} = I_n - \frac{1}{n} \sum_{r=1}^{R_0} \zeta_{ny,r} \zeta_{ny,r}' \), and \( M_{F_T} = I_T - \frac{1}{T} F_T \left( \frac{1}{T} F_T F_T \right)^{-1} F_T = I_T - \frac{1}{T} \sum_{r=1}^{R_0} \zeta_{Ty,r} \zeta_{Ty,r}' \), where \( \zeta_{ny,r} \) and \( \zeta_{Ty,r} \) are respectively the \( r \)-th columns of \( \Gamma_{ny} \left( \frac{1}{n} \Gamma_{ny} \Gamma_{ny} \right)^{-1} \) and \( F_T \left( \frac{1}{T} F_T F_T \right)^{-1} \). With Assumptions 8 and 3, elements of \( \zeta_{ny,r} \) and \( \zeta_{Ty,r} \) are uniformly bounded and it can be checked that the LLNs in [Qu et al. (2015)] can be extended to statistics with these general projectors.
Lemma 4. Under Assumptions \[\text{[2][3][4]}\] and \[\text{[3][4][5]}\] define \(\varsigma_{nT}^\ast\) as in Lemma \[\text{[3]}\] then

\[
\frac{1}{nT} \left( \hat{\varsigma}_{nT}^\ast \gamma_{nT}^\ast (M_{F_T} \otimes M_{\Gamma_n}) \varsigma_{nT}^\ast - \mathbb{E} \left( \hat{\varsigma}_{nT}^\ast \gamma_{nT}^\ast (M_{F_T} \otimes M_{\Gamma_n}) \varsigma_{nT}^\ast \right) \right) = o_p(1), \quad \text{and}
\]

\[
\frac{1}{nT} \left( \hat{\varsigma}_{nT}^\ast \gamma_{nT}^\ast (M_{F_T} \otimes M_{\Gamma_n}) \tilde{\varsigma}_{nT}^\ast - \mathbb{E} \left( \hat{\varsigma}_{nT}^\ast \gamma_{nT}^\ast (M_{F_T} \otimes M_{\Gamma_n}) \tilde{\varsigma}_{nT}^\ast \right) \right) = o_p(1).
\]

Lemma 3 implies that \(\|\varsigma_{nT}^{(1)}\|_2 = O_P(1)\). Completing the squares on the right hand side of Eq. (10),

\[
Q_{nT}(\theta) = Q_{nT}(\theta_0) - \frac{1}{2} (\theta - \theta_0 - \gamma)' \hat{C}_{nT} (\theta - \theta_0 - \gamma) + \frac{1}{2} \gamma_{nT}' \hat{C}_{nT} \gamma_{nT} + Q_{nT}^{\text{rem}}(\theta).
\]

(11)

At \(\theta = \theta_0 + \gamma_{nT}\), \(Q_{nT}(\theta_0 + \gamma_{nT}) = Q_{nT}(\theta_0) + \frac{1}{2} \gamma_{nT}' \hat{C}_{nT} \gamma_{nT} + Q_{nT}^{\text{rem}}(\theta_0 + \gamma_{nT})\). Now let \(\hat{\theta}_{nT}\) be the QML estimator. From Eq. (11),

\[
Q_{nT}(\hat{\theta}_{nT}) = Q_{nT}(\theta_0) - \frac{1}{2} (\hat{\theta}_{nT} - \theta_0 - \gamma_{nT})' \hat{C}_{nT} (\hat{\theta}_{nT} - \theta_0 - \gamma_{nT}) + \frac{1}{2} \gamma_{nT}' \hat{C}_{nT} \gamma_{nT} + Q_{nT}^{\text{rem}}(\hat{\theta}_{nT}).
\]

Because \(\hat{\theta}_{nT}\) maximizes the objective function, \(Q_{nT}(\hat{\theta}_{nT}) \geq Q_{nT}(\theta_0 + \gamma_{nT})\), which implies that

\[
(\hat{\theta}_{nT} - \theta_0 - \gamma_{nT})' \hat{C}_{nT} (\hat{\theta}_{nT} - \theta_0 - \gamma_{nT}) \leq 2 \left( Q_{nT}^{\text{rem}}(\hat{\theta}_{nT}) - Q_{nT}^{\text{rem}}(\theta_0 + \gamma_{nT}) \right).
\]

(12)

Assuming that \(\hat{C}_{nT}\) is positive definite, it follows that \((\hat{\theta}_{nT} - \theta_0 - \gamma_{nT})' \hat{C}_{nT} (\hat{\theta}_{nT} - \theta_0 - \gamma_{nT}) \geq b \| \hat{\theta}_{nT} - \theta_0 - \gamma_{nT} \|_2^2\) for some \(b > 0\), and \(\| \gamma_{nT} \|_2 = O_P \left( \frac{1}{\sqrt{nT}} \right) \) from \(\| \hat{C}_{nT}^{(1)} \|_2 = O_P(1) \). Using the similar argument in \[\text{[2][3]}\] and \[\text{[4][5]}\], it follows that \(b \| \hat{\theta}_{nT} - \theta_0 - \gamma_{nT} \|_2 + O_P \left( \| \hat{\theta}_{nT} - \theta_0 - \gamma_{nT} \| n^{-\frac{1}{2}} \right) + O_P \left( n^{-\frac{1}{2}} \right) \leq 0\), which implies that \(\sqrt{nT} \| \hat{\theta}_{nT} - \theta_0 - \gamma_{nT} \|_2 = O_P \left( \frac{1}{n^{1/2}} \right)\) as \(n\) and \(T\) are proportional in the limit. Therefore,

\[
\sqrt{nT} (\hat{\theta}_{nT} - \theta_0) = \sqrt{nT} \gamma_{nT} + o_P(1) = \tilde{c}_{nT}^{-1} \varsigma_{nT}^{(1)} + o_P(1).
\]

Recall that the parameter vector is \(\theta = (\beta', \delta')'\). In the Proposition 2 below, denote accordingly,

\[
\varphi_{nT} = \begin{pmatrix}
0 \\
0 \\
- \frac{1}{\sqrt{nT}} \text{tr} \left[ (P_{F_T} \otimes I_n) \hat{G}_{nT} + (I_T \otimes P_{I_{\Gamma_n}}) \hat{G}_{nT} \right] \\
R_{\alpha_0} \sqrt{\frac{T}{n}} \text{tr} \left[ \Sigma_{\alpha_0}^\frac{1}{2} \Omega_1 \right] + \sqrt{\frac{T}{n}} \text{tr} \left[ \Sigma_{\alpha_0}^\frac{1}{2} \Omega_1 \otimes I_n \right] + \sqrt{\frac{T}{n}} \text{tr} \left[ \Sigma_{\alpha_0}^\frac{1}{2} \Omega_1 \otimes I_n \right] \\
R_{\alpha_0} \sqrt{\frac{T}{n}} \text{tr} \left[ \Sigma_{\alpha_0}^\frac{1}{2} \Omega_1 \right] + \sqrt{\frac{T}{n}} \text{tr} \left[ \Sigma_{\alpha_0}^\frac{1}{2} \Omega_1 \otimes I_n \right] + \sqrt{\frac{T}{n}} \text{tr} \left[ \Sigma_{\alpha_0}^\frac{1}{2} \Omega_1 \otimes I_n \right] \\
\vdots \\
0
\end{pmatrix}
\]

Proposition 2. Assuming that \(\lim_{n,T \to \infty} \frac{n}{T} \to \kappa > 0\), \(\Sigma = \lim_{n,T \to \infty} \Sigma_{nT}\) is positive definite, \(\Pi = \lim_{n,T \to \infty} \Pi_{nT}\), which are given in Eq. (A.30) of the appendix, and Assumptions \[\text{[3][4][5]}\] or \[\text{[3][4][6]}\] hold, then

\[
\sqrt{nT} \left( \hat{\theta}_{nT} - \theta_0 - \frac{1}{\sqrt{nT}} \tilde{c}_{nT}^{-1} \varphi_{nT} \right) \frac{d}{d} \sim N(0, \Sigma^{-1} (\Sigma + \Pi) \Sigma^{-1}).
\]
If the disturbance terms are normally distributed, $\Pi = 0$ and $\sqrt{nT} \left( \theta_nT - \theta_0 - \frac{1}{\sqrt{nT}} \tilde{C}_{nT}^{-1} \phi_{nT} \right) \overset{d}{\rightarrow} N(0, \Sigma^{-1})$.

It can be shown that the variance of $\tilde{C}_{nT}^{(1)} - \phi_{nT}$ is $\Sigma_nT + \Pi_nT$. Comparing the terms of $\tilde{C}_{nT}$ with $\Sigma_nT$ reveals that $p \lim_{nT \to \infty} (\tilde{C}_{nT} - \Sigma_nT) = 0$, which gives the limiting variance matrix $\Sigma^{-1} (\Sigma + \Pi) \Sigma^{-1}$.

Proposition 2 shows that the limiting distributions of the QML estimates do not center at their true values but with a bias term of the order $O_p \left( \frac{1}{\sqrt{nT}} \right)$. The bias term is due to incidental parameters, namely factor loadings and time factors, as the estimates for them have slower convergence rates. As in Shi and Lee (2015), the bias can be corrected by substituting in the QML estimates $\hat{\theta}_{nT}$ and the estimated projectors of factors and factor loadings in $\phi_{nT}$ and $\Sigma_nT$. The bias corrected estimator is $\hat{\theta}^c_{nT} = \hat{\theta}_{nT} - \frac{1}{\sqrt{nT}} \tilde{C}_{nT}^{-1} \phi_{nT}$.

**Corollary 1.** Assuming that $\lim_{nT \to \infty} \frac{nT}{T} \to \kappa > 0$, $\Sigma = \lim_{nT \to \infty} \Sigma_nT$ is positive definite, $\Pi = \lim_{nT \to \infty} \Pi_nT$, and Assumptions 2, 3, 4, or 6, 8 hold, then $\sqrt{nT} \left( \hat{\theta}^c_{nT} - \theta_0 \right) \overset{d}{\rightarrow} N(0, \Sigma^{-1} (\Sigma + \Pi) \Sigma^{-1})$. If the disturbance terms are normally distributed, $\sqrt{nT} \left( \hat{\theta}^c_{nT} - \theta_0 \right) \overset{d}{\rightarrow} N(0, \Sigma^{-1})$.

### 4.4. The number of factors

As long as the number of factors specified is not 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 estimator for a regression panel is invariant to the inclusion of redundant factors, its finite sample performance may suffer, as Lu and Su (2013) 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.

Several criteria on factor number selection have been proposed in the literature, including Bai and Ng (2002)’s PC and IC criteria, Onatski (2010)’s Edge Distribution estimator, and Ahn and Horenstein (2013)’s eigenvalue ratio and growth ratio criteria. We specifically show how Ahn and Horenstein (2013)’s eigenvalue ratio and growth ratio criteria can be extended here. Denote $\hat{\theta}_{nT}$ the QML estimator with $R_\gamma \geq R_0$, which is a preliminary consistent estimator using a larger number of factors than the true one. For the $z$ equation, let $\tilde{D}_{nT} = (d_{n1}, \ldots, d_{nT})$ with $\tilde{D}_{nT} = \left( \tilde{\Sigma}^{-\frac{1}{2}} \otimes I_n \right) \left( z_{nT} - X_{nT} \hat{\beta}_T \right)$. Using the notation of Eq. (7), we have $\tilde{D}_{nT} = \Gamma_n F^T_{\gammaT} + \left( \Sigma_{\gamma \vee 0} \otimes I_n \right) E_{nT} + \tilde{E}_{nT}(\hat{\theta}_{nT})$, with $\tilde{E}_{nT}(\hat{\theta}_{nT}) = \sum_{k=1}^{k+y} \hat{\eta}^y_{k} \tilde{V}^y_{z} + \sum_{k_1, k_2 = 1}^{k+y} \hat{\eta}^y_{k_1} \tilde{V}^y_{z} k_1 k_2$. Denote $\tilde{\mu}_{nT,k} = \mu_k \left( \frac{1}{nT} \tilde{D}_{nT} \tilde{D}_{nT}^T \right)$ for $k \geq 1$, $V^z(k) = \sum_{j=k+1}^{\infty} \mu_j$, and for $k \geq 0$, and $\tilde{\mu}^z_{nT,0} = V^z(0)/\log(\min(n, T))$. Define the “eigenvalue ratio” statistic, $ER^z(k) = \frac{\tilde{\mu}^z_{nT,k}}{\mu_{nT,k+1}}$, and the “growth ratio” statistic, $GR^z(k) = \log \left( \frac{V^z(k+1)}{V^z(k)} \right) / \log \left( \frac{V^z(k+1)}{V^z(k)} \right)$. The number of factors in the $z$ equation can be selected according to $\hat{\kappa}_{ER}^z(k) = \max_{0 \leq k \leq k_{\text{max}}} ER^z(k)$ or $\hat{\kappa}_{GR}^z(k) = \max_{0 \leq k \leq k_{\text{max}}} GR^z(k)$.

For the $y$ equation, let $\tilde{G}_{nT} = (\tilde{g}_{n1}, \ldots, \tilde{g}_{nT})$ with $\tilde{g}_{nT} = \tilde{\Sigma}^{-\frac{1}{2}} \left( S_{nT} \tilde{\lambda} - X_{nT} \hat{\beta}_T \right) \left( z_{nT} - X_{nT} \hat{\beta}_T \right)$. From Eq. (8), $\tilde{G}_{nT} = \Gamma_n F^T_{\gammaT} + \sigma_{\gamma \vee 1} \tilde{Z}_{nT} + \tilde{F}_{nT}(\hat{\theta}_{nT})$, with $\tilde{F}_{nT}(\hat{\theta}_{nT}) = \sum_{k=1}^{k+y} \hat{\eta}^y_{k} \tilde{V}^y_{z} + \sum_{k_1, k_2 = 1}^{k+y} \hat{\eta}^y_{k_1} \tilde{V}^y_{z} k_1 k_2$. $\hat{\kappa}_{GR}^y(k)$ and $\hat{\kappa}_{ER}^y(k)$ are defined similarly via $\tilde{\mu}^y_{nT,k}$, $\tilde{\mu}^y_{nT,k}$.

**Proposition 3.** Assuming that $\lim_{nT \to \infty} \frac{nT}{T} \to \kappa > 0$, Assumptions 2, 3, and 5 hold, the preliminary estimator $\| \hat{\theta}_{nT} - \theta_0 \|_2 = o_p(n^{-\frac{1}{2}})$, $R_0 \geq 0$, $R_0 \geq 0$, then $\lim_{nT \to \infty} Pr(\hat{\kappa}_{GR}^y = R_0) = \lim_{nT \to \infty} Pr(\hat{\kappa}_{ER}^y = R_0) = \lim_{nT \to \infty} Pr(\hat{\kappa}_{GR}^z = R_0) = \lim_{nT \to \infty} Pr(\hat{\kappa}_{ER}^z = R_0) = 1$. 

16
Therefore the number of factors can be determined consistently. The proof, which is in the appendix, 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 ($R_{\alpha0} = 0$ or $R_{\gamma0} = 0$) is covered.

5. Monte Carlo simulations

5.1. 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 endogeneity, and different ratios between the variances of the idiosyncratic error and the factors. We also check the finite sample performance of the QML estimator in a data specification that tailors to the empirical application in Section 4.

In the main data specification, the outcome equation is $y_{it} = \lambda \sum_{j=1}^{n} w_{it,j} y_{jt} + x_{it} \beta_{x} + \gamma_{it} f_{it} + v_{it}$. There are two unobserved factors in the $y_{it}$ equation. The factors ($f_{it}$) and factor loadings ($\gamma_{it}$) are generated from independent uniform random variables on $[-2, 2]$. The observed regressor ($x_{it}$) is a scalar and correlates with factors through $x_{it} = \frac{1}{3} (\gamma_{it} f_{it} + \gamma_{it} f_{jt} + f_{it} \ell_{2}) + \eta_{it}$, where $\eta_{it}$ is a uniform random variable on $[-2, 2]$ and $\ell_{2}$ is a $2 \times 1$ vector of 1’s. The spatial weights are correlated with $v_{it}$ according to the following process.

1. Individuals indexed by 1 to $n$ are located successively on a chessboard of dimension $\sqrt{n} \times \sqrt{n}$. The neighborhood structure follows the rook pattern, i.e. two individuals are neighbors if they share a common border. Therefore individuals in the interior of the chessboard have 4 neighbors, and those on the border and the corner have 3 and 2 neighbors, respectively. Let $w_{ij} = 1$ if $i$ and $j$ are neighbors and $w_{ij} = 0$ otherwise, and denote the $n \times n$ matrix $W = \begin{bmatrix} w_{ij} \end{bmatrix}$. Note that in this experiment, the geographic locations do not change over time.

2. Let $z_{it} = x_{it} \beta_{c} + \gamma_{ij} f_{ij} + \epsilon_{it}$. There is one unobserved factor in the $z_{it}$ equation. The scalar $\gamma_{ij}$ is generated from uniform random variable on $[-2, 2]$, and $f_{ij}$ is the first element of $f_{ij}$. Let $w_{ij} = w_{ij} \times \min \left( \frac{1}{|z_{it} - z_{jt}|} ; 2 \right)$. The spatial dependence is stronger for neighbors with similar $z$’s. In line with the condition that the functions $h_{ij}(\cdot)$ are uniformly bounded, $w_{ij}^{c}$ is capped at 2 such that individuals whose $z$’s are very similar ($|z_{it} - z_{jt}| < 0.5$) have the same degree of spatial effect. The spatial weights in the outcome equation are row-normalized, $w_{it,j} = \frac{w_{it,j}^{c}}{\sum_{j=1}^{n} w_{it,j}^{c}}$. Note that the spatial weights matrix complies with Assumption [6] for all $i \neq j$.

3. The idiosyncratic errors $v_{it}$ and $\epsilon_{it}$ are generated from i.i.d. bivariate normal random variables with mean 0 and variance $\frac{1}{\vartheta} \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. The inverse of $\vartheta$ is a measure of the signal to noise ratio. Each common factor and the idiosyncratic error have the same variance when $\vartheta = 1$, and the latter is more variable when $\vartheta > 1$.

Our empirical application considers multiple spatial weights matrices. To provide evidence that our estimator can perform well, in an alternative data specification, the outcome equation is replaced by $y_{it} = \lambda_{1} \sum_{j=1}^{n} w_{1,ij} y_{jt} + \lambda_{2} \sum_{j=1}^{n} w_{it,j} y_{jt} + x_{it} \beta_{x} + \gamma_{it} f_{it} + v_{it}$, where $w_{it,j}$ is the same as in the main data specification representing time varying spatial relations, and $W_{1} = \begin{bmatrix} w_{1,ij} \end{bmatrix}$ is row-normalized $W$ representing geographic contiguity. Other parts of the data specification are the same.
5.2. Monte Carlo results

1000 Monte Carlo replications are carried out. The numerical maximization routine starts at multiple values, because the objective function is not concave and multiple local maxima may exist. We vary the degree of spatial interaction ($\lambda = 0.2$ or $\lambda = 0.6$) and endogeneity ($\rho = 0.2$ or $\rho = 0.6$). $\alpha_\xi$ denotes $\sigma_\xi^{-1}$ and $\alpha$ denotes $\Sigma^{-1}$. 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 $\Sigma$ in Eq. (A.30), and the nominal coverage probability is set at 95%. The estimates for the variance parameters ($\alpha_\xi$ and $\alpha$) have noticeable biases in finite sample, and as a result, their CPs are well below 95%. The CPs for other parameter estimates are generally slightly below 95% and are better. Therefore overall, hypothesis tests might be over-rejected. The Monte Carlo results of the bias corrected estimator are reported in Table 2. The biases for the variance parameters have been significantly reduced. The CPs are more accurate so a more reliable statistical inference can be based on the bias corrected estimator.

Table 3 shows that estimates suffer from substantial bias if the spatial weights matrix is treated erroneously as exogenous in the estimation, where only the outcome equation is estimated. The biases still remain in large samples. The estimate for the spatial effect is biased upward, and the bias is larger in samples with a higher degree of endogeneity.

Proposition 1 shows that the QML estimates are consistent even when the number of factors is over-specified. In Table 4, we report the performance of the bias corrected estimators when the number of factors in the outcome equation is over-specified by 1, i.e. $R_y = 3$ while the true $R_{y0} = 2$. Although estimates are still consistent, but for these finite samples, the variance parameter ($\alpha_\xi$) has noticeable bias that is not removed by the bias correction procedure. Tables 2 and 10 in the appendix report additional Monte Carlo results. As the number of redundant factors increases, the CP deteriorates. Therefore for valid inference, it seems important that a correct number of factors is chosen.

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

The empirical application in the next section uses two spatial weights matrices and the estimation results there reveal different degrees of spatial interactions. Table 5 based on the empirical feature of the empirical application confirms that our bias corrected estimator still performs well in the alternative specification with multiple spatial weights matrices.
Table 1: Performance of the QML Estimator $\hat{\theta}_{QT}$

<table>
<thead>
<tr>
<th>$n$</th>
<th>$T$</th>
<th>$\lambda_0$</th>
<th>$\rho_0$</th>
<th>Bias</th>
<th>$\hat{\beta}_z$</th>
<th>$\hat{\beta}_y$</th>
<th>$\hat{\lambda}$</th>
<th>$\hat{\alpha}_z$</th>
<th>$\hat{\alpha}$</th>
<th>$\hat{\delta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>25</td>
<td>25</td>
<td>0.2</td>
<td>0.2</td>
<td>0.000252</td>
<td>-0.00113</td>
<td>-0.00211</td>
<td>0.08269</td>
<td>0.03813</td>
<td>0.00099</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.94</td>
<td>0.898</td>
<td>0.903</td>
<td>0.159</td>
<td>0.674</td>
<td>0.928</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>25</td>
<td>0.2</td>
<td>0.2</td>
<td>0.00013</td>
<td>0.00142</td>
<td>-0.00143</td>
<td>0.05931</td>
<td>0.02786</td>
<td>0.00086</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.942</td>
<td>0.926</td>
<td>0.918</td>
<td>0.119</td>
<td>0.675</td>
<td>0.955</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>49</td>
<td>0.2</td>
<td>0.2</td>
<td>-0.00115</td>
<td>0.00001</td>
<td>-0.00190</td>
<td>0.05873</td>
<td>0.02836</td>
<td>0.00128</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.938</td>
<td>0.938</td>
<td>0.923</td>
<td>0.140</td>
<td>0.673</td>
<td>0.934</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>49</td>
<td>0.2</td>
<td>0.2</td>
<td>0.00028</td>
<td>0.00146</td>
<td>-0.00180</td>
<td>0.03846</td>
<td>0.01868</td>
<td>0.00029</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.945</td>
<td>0.946</td>
<td>0.916</td>
<td>0.187</td>
<td>0.696</td>
<td>0.942</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2)</td>
<td>25</td>
<td>25</td>
<td>0.2</td>
<td>0.6</td>
<td>0.00108</td>
<td>-0.00334</td>
<td>0.00113</td>
<td>0.10114</td>
<td>0.03806</td>
<td>0.00003</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.941</td>
<td>0.913</td>
<td>0.906</td>
<td>0.150</td>
<td>0.687</td>
<td>0.912</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>25</td>
<td>0.2</td>
<td>0.6</td>
<td>0.00020</td>
<td>0.00063</td>
<td>-0.00104</td>
<td>0.07260</td>
<td>0.02746</td>
<td>0.00051</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.936</td>
<td>0.931</td>
<td>0.928</td>
<td>0.114</td>
<td>0.693</td>
<td>0.927</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>49</td>
<td>0.2</td>
<td>0.6</td>
<td>-0.00119</td>
<td>-0.00074</td>
<td>0.00076</td>
<td>0.07245</td>
<td>0.02773</td>
<td>0.00063</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.935</td>
<td>0.940</td>
<td>0.922</td>
<td>0.122</td>
<td>0.681</td>
<td>0.921</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>49</td>
<td>0.2</td>
<td>0.6</td>
<td>0.00046</td>
<td>0.00128</td>
<td>-0.00143</td>
<td>0.04699</td>
<td>0.01856</td>
<td>0.00038</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.938</td>
<td>0.943</td>
<td>0.929</td>
<td>0.187</td>
<td>0.706</td>
<td>0.924</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3)</td>
<td>25</td>
<td>25</td>
<td>0.6</td>
<td>0.2</td>
<td>0.000252</td>
<td>-0.00003</td>
<td>-0.00284</td>
<td>0.08201</td>
<td>0.03813</td>
<td>0.00137</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.941</td>
<td>0.901</td>
<td>0.907</td>
<td>0.183</td>
<td>0.674</td>
<td>0.926</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>25</td>
<td>0.6</td>
<td>0.2</td>
<td>0.00013</td>
<td>0.00176</td>
<td>-0.00129</td>
<td>0.05904</td>
<td>0.02786</td>
<td>0.00092</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.942</td>
<td>0.927</td>
<td>0.930</td>
<td>0.141</td>
<td>0.675</td>
<td>0.956</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>49</td>
<td>0.6</td>
<td>0.2</td>
<td>-0.00115</td>
<td>0.00082</td>
<td>-0.00232</td>
<td>0.05822</td>
<td>0.02836</td>
<td>0.00154</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.938</td>
<td>0.938</td>
<td>0.917</td>
<td>0.156</td>
<td>0.673</td>
<td>0.933</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>49</td>
<td>0.6</td>
<td>0.2</td>
<td>0.00028</td>
<td>0.00176</td>
<td>-0.00144</td>
<td>0.03821</td>
<td>0.01868</td>
<td>0.00034</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.945</td>
<td>0.948</td>
<td>0.926</td>
<td>0.198</td>
<td>0.696</td>
<td>0.940</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(4)</td>
<td>25</td>
<td>25</td>
<td>0.6</td>
<td>0.6</td>
<td>0.00108</td>
<td>-0.00236</td>
<td>-0.00139</td>
<td>0.10062</td>
<td>0.03806</td>
<td>0.00072</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.941</td>
<td>0.910</td>
<td>0.912</td>
<td>0.171</td>
<td>0.687</td>
<td>0.916</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>25</td>
<td>0.6</td>
<td>0.6</td>
<td>0.00020</td>
<td>0.00082</td>
<td>-0.00083</td>
<td>0.07239</td>
<td>0.02746</td>
<td>0.00060</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.936</td>
<td>0.931</td>
<td>0.930</td>
<td>0.126</td>
<td>0.693</td>
<td>0.933</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>49</td>
<td>0.6</td>
<td>0.6</td>
<td>-0.00119</td>
<td>-0.00010</td>
<td>-0.00063</td>
<td>0.07216</td>
<td>0.02773</td>
<td>0.00107</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.935</td>
<td>0.937</td>
<td>0.929</td>
<td>0.144</td>
<td>0.681</td>
<td>0.919</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>49</td>
<td>0.6</td>
<td>0.6</td>
<td>0.00046</td>
<td>0.00150</td>
<td>-0.00112</td>
<td>0.04676</td>
<td>0.01856</td>
<td>0.00050</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.938</td>
<td>0.941</td>
<td>0.928</td>
<td>0.205</td>
<td>0.706</td>
<td>0.920</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

True parameter values: $\beta_0 = 1$, $\beta_0 = 1$, $\sigma_\epsilon = \sqrt{\frac{2}{3}}$, $\sigma_\epsilon = \sqrt{\frac{2}{3}}$ (hence $\alpha_0 = \left(\frac{1}{4}\right)^{-1}$). In the low endogeneity case, $\rho_0 = 0.2$, $\alpha_{\epsilon_0} = 0.88$ and $\delta_0 = 0.2$. In the high endogeneity case, $\rho_0 = 0.6$, $\alpha_{\epsilon_0} = 1.08$ and $\delta_0 = 0.6$. $\hat{\theta}_{QT}$ is the QML estimator. The coverage probabilities (CP) are calculated using the theoretical standard deviations obtained from the diagonal elements of $\frac{1}{n^2\Sigma^{-1}}$. 

19
Table 2: Performance of the Bias-Corrected QML Estimator $\hat{\theta}_{\delta T}$

<table>
<thead>
<tr>
<th>n</th>
<th>T</th>
<th>$\lambda_0$</th>
<th>$\rho_0$</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\lambda$</th>
<th>$\hat{\alpha}$</th>
<th>$\hat{\delta}$</th>
<th>Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>25</td>
<td>25</td>
<td>0.2</td>
<td>0.2</td>
<td>0.00252</td>
<td>-0.00145</td>
<td>-0.00096</td>
<td>0.00543</td>
<td>0.00197</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.952</td>
<td>0.934</td>
<td>0.934</td>
<td>0.928</td>
<td>0.945</td>
</tr>
<tr>
<td>(2)</td>
<td>49</td>
<td>25</td>
<td>0.2</td>
<td>0.2</td>
<td>0.00013</td>
<td>0.00117</td>
<td>-0.00026</td>
<td>0.00241</td>
<td>0.00086</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.945</td>
<td>0.948</td>
<td>0.932</td>
<td>0.936</td>
<td>0.943</td>
</tr>
<tr>
<td>(3)</td>
<td>25</td>
<td>49</td>
<td>0.2</td>
<td>0.2</td>
<td>-0.00115</td>
<td>-0.00034</td>
<td>-0.00059</td>
<td>0.00188</td>
<td>0.00135</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.943</td>
<td>0.956</td>
<td>0.939</td>
<td>0.915</td>
<td>0.940</td>
</tr>
<tr>
<td>(4)</td>
<td>49</td>
<td>25</td>
<td>0.2</td>
<td>0.2</td>
<td>0.00028</td>
<td>0.00122</td>
<td>-0.00076</td>
<td>0.00088</td>
<td>0.00063</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.948</td>
<td>0.954</td>
<td>0.936</td>
<td>0.933</td>
<td>0.945</td>
</tr>
</tbody>
</table>

True parameter values: $\beta_2 = 1$, $\beta_0 = 1$, $\sigma_0 = \sqrt{\frac{2}{n}}$, $\sigma_0 = \sqrt{\frac{2}{n}}$ (hence $\alpha_0 = \left(\frac{1}{2}\right)^{-\frac{1}{2}}$). In the low endogeneity case, $\rho_3 = 0.2$, $\alpha_3 = 0.88$ and $\delta_3 = 0.2$. In the high endogeneity case, $\rho_3 = 0.6$, $\alpha_3 = 1.08$ and $\delta_3 = 0.6$. $\hat{\theta}_{\delta T}$ is the bias-corrected QML estimator. The coverage probabilities (CP) are calculated using the theoretical standard deviations obtained from the diagonal elements of $\frac{1}{nT} \sum_{nT}^{-1}$. 

20
Table 3: Estimation under Misspecification: Assuming Exogenous Spatial Weights Matrix

<table>
<thead>
<tr>
<th>$n$</th>
<th>$T$</th>
<th>$\lambda_0$</th>
<th>$\rho_0$</th>
<th>$\hat{\beta}_z$</th>
<th>$\lambda$</th>
<th>$\rho_0$</th>
<th>$\hat{\beta}_y$</th>
<th>$\lambda$</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>25</td>
<td>0.2</td>
<td>0.2</td>
<td>-0.00739</td>
<td>0.02007</td>
<td>0.6</td>
<td>-0.03103</td>
<td>0.08171</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>E-SD 0.04823</td>
<td>0.03597</td>
<td>E-SD</td>
<td>0.04883</td>
<td>0.03516</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RMSE 0.04877</td>
<td>0.04117</td>
<td>RMSE</td>
<td>0.05783</td>
<td>0.08895</td>
</tr>
<tr>
<td>49</td>
<td>25</td>
<td>0.2</td>
<td>0.2</td>
<td>-0.00263</td>
<td>0.01775</td>
<td>0.6</td>
<td>-0.01165</td>
<td>0.06913</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>E-SD 0.03099</td>
<td>0.02466</td>
<td>E-SD</td>
<td>0.03095</td>
<td>0.02386</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RMSE 0.03109</td>
<td>0.03038</td>
<td>RMSE</td>
<td>0.03513</td>
<td>0.07313</td>
</tr>
<tr>
<td>25</td>
<td>49</td>
<td>0.2</td>
<td>0.2</td>
<td>-0.00552</td>
<td>0.01789</td>
<td>0.6</td>
<td>-0.02391</td>
<td>0.07239</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>E-SD 0.03121</td>
<td>0.02381</td>
<td>E-SD</td>
<td>0.03150</td>
<td>0.02282</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RMSE 0.03168</td>
<td>0.02977</td>
<td>RMSE</td>
<td>0.03953</td>
<td>0.07590</td>
</tr>
<tr>
<td>49</td>
<td>49</td>
<td>0.2</td>
<td>0.2</td>
<td>-0.00347</td>
<td>0.01879</td>
<td>0.6</td>
<td>-0.02013</td>
<td>0.07490</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>E-SD 0.02143</td>
<td>0.01679</td>
<td>E-SD</td>
<td>0.02157</td>
<td>0.01620</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RMSE 0.02170</td>
<td>0.02519</td>
<td>RMSE</td>
<td>0.02949</td>
<td>0.07664</td>
</tr>
<tr>
<td>25</td>
<td>25</td>
<td>0.6</td>
<td>0.2</td>
<td>-0.00572</td>
<td>0.01010</td>
<td>0.6</td>
<td>-0.01931</td>
<td>0.03910</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>E-SD 0.03159</td>
<td>0.01654</td>
<td>E-SD</td>
<td>0.03136</td>
<td>0.01548</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RMSE 0.03177</td>
<td>0.01937</td>
<td>RMSE</td>
<td>0.03682</td>
<td>0.04205</td>
</tr>
<tr>
<td>49</td>
<td>25</td>
<td>0.6</td>
<td>0.2</td>
<td>-0.00572</td>
<td>0.00895</td>
<td>0.6</td>
<td>-0.02432</td>
<td>0.03767</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>E-SD 0.03173</td>
<td>0.01579</td>
<td>E-SD</td>
<td>0.03195</td>
<td>0.01456</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RMSE 0.03222</td>
<td>0.01814</td>
<td>RMSE</td>
<td>0.04014</td>
<td>0.04038</td>
</tr>
<tr>
<td>49</td>
<td>49</td>
<td>0.6</td>
<td>0.2</td>
<td>-0.00456</td>
<td>0.01102</td>
<td>0.6</td>
<td>-0.02298</td>
<td>0.04313</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>E-SD 0.02174</td>
<td>0.01164</td>
<td>E-SD</td>
<td>0.02179</td>
<td>0.01075</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>RMSE 0.02220</td>
<td>0.01603</td>
<td>RMSE</td>
<td>0.03166</td>
<td>0.04445</td>
</tr>
</tbody>
</table>

The estimation mistakenly assumes the spatial weights matrix to be exogenous, hence only the outcome equation is estimated. Interactive effects are still included. True parameter values: $\beta_0 = 1$, $\beta_y = 1$, $\sigma_0 = \sqrt{\frac{2}{3}}$, $\sigma_0 = \sqrt{\frac{2}{3}}$ (hence $\alpha_0 = \left(\frac{2}{3}\right)^{-\frac{1}{2}}$). In the low endogeneity case, $\rho_0 = 0.2$, $\alpha_{\delta_0} = 0.88$ and $\delta_0 = 0.2$. In the high endogeneity case, $\rho_0 = 0.6$, $\alpha_{\delta_0} = 1.08$ and $\delta_0 = 0.6$. E-SD is the empirical standard deviation of the estimates.
Table 4: Performance of the Bias-Corrected QML Estimator $\hat{\theta}_{qT}$ when $R_y = R_{z0} + 1$.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$T$</th>
<th>$\lambda_0$</th>
<th>$p_0$</th>
<th>$\hat{\beta}_\xi$</th>
<th>$\hat{\beta}_\eta$</th>
<th>$\hat{\lambda}$</th>
<th>$\hat{\alpha}_\zeta$</th>
<th>$\hat{\alpha}$</th>
<th>$\hat{\delta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>25</td>
<td>0.2</td>
<td>0.2</td>
<td>0.00252</td>
<td>-0.00201</td>
<td>-0.00158</td>
<td>0.04184</td>
<td>0.00197</td>
<td>0.00055</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CP</td>
<td></td>
<td>0.952</td>
<td>0.905</td>
<td>0.897</td>
<td>0.633</td>
<td>0.945</td>
<td>0.922</td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>0.2</td>
<td>0.2</td>
<td>0.00013</td>
<td>0.00093</td>
<td>-0.00090</td>
<td>0.02825</td>
<td>0.00086</td>
<td>0.00139</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CP</td>
<td></td>
<td>0.945</td>
<td>0.921</td>
<td>0.912</td>
<td>0.678</td>
<td>0.943</td>
<td>0.940</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.2</td>
<td>0.2</td>
<td>-0.00115</td>
<td>-0.00044</td>
<td>-0.00088</td>
<td>0.02740</td>
<td>0.00135</td>
<td>0.00158</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CP</td>
<td></td>
<td>0.943</td>
<td>0.938</td>
<td>0.911</td>
<td>0.677</td>
<td>0.940</td>
<td>0.938</td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>0.2</td>
<td>0.2</td>
<td>0.00028</td>
<td>0.00138</td>
<td>-0.00082</td>
<td>0.01865</td>
<td>0.00063</td>
<td>0.00054</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CP</td>
<td></td>
<td>0.948</td>
<td>0.941</td>
<td>0.934</td>
<td>0.699</td>
<td>0.945</td>
<td>0.942</td>
</tr>
<tr>
<td>(2)</td>
<td>25</td>
<td>0.2</td>
<td>0.6</td>
<td>0.00018</td>
<td>-0.00316</td>
<td>-0.00067</td>
<td>0.05076</td>
<td>0.00189</td>
<td>-0.00002</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CP</td>
<td></td>
<td>0.948</td>
<td>0.919</td>
<td>0.907</td>
<td>0.642</td>
<td>0.950</td>
<td>0.914</td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>0.2</td>
<td>0.6</td>
<td>0.00020</td>
<td>0.00079</td>
<td>-0.00063</td>
<td>0.03452</td>
<td>0.00047</td>
<td>0.00048</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CP</td>
<td></td>
<td>0.948</td>
<td>0.932</td>
<td>0.929</td>
<td>0.668</td>
<td>0.949</td>
<td>0.921</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.2</td>
<td>0.6</td>
<td>-0.00119</td>
<td>-0.00033</td>
<td>-0.00015</td>
<td>0.03395</td>
<td>0.00073</td>
<td>0.00091</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CP</td>
<td></td>
<td>0.947</td>
<td>0.937</td>
<td>0.930</td>
<td>0.672</td>
<td>0.938</td>
<td>0.926</td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>0.2</td>
<td>0.6</td>
<td>0.00047</td>
<td>0.00138</td>
<td>-0.00074</td>
<td>0.02277</td>
<td>0.00051</td>
<td>0.00059</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CP</td>
<td></td>
<td>0.942</td>
<td>0.940</td>
<td>0.929</td>
<td>0.708</td>
<td>0.950</td>
<td>0.929</td>
</tr>
<tr>
<td>(3)</td>
<td>25</td>
<td>0.6</td>
<td>0.2</td>
<td>0.00252</td>
<td>-0.00130</td>
<td>-0.00188</td>
<td>0.04149</td>
<td>0.00197</td>
<td>0.00147</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CP</td>
<td></td>
<td>0.952</td>
<td>0.907</td>
<td>0.903</td>
<td>0.648</td>
<td>0.945</td>
<td>0.923</td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>0.6</td>
<td>0.2</td>
<td>0.00013</td>
<td>0.00127</td>
<td>-0.00112</td>
<td>0.02801</td>
<td>0.00086</td>
<td>0.00148</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CP</td>
<td></td>
<td>0.945</td>
<td>0.920</td>
<td>0.930</td>
<td>0.689</td>
<td>0.943</td>
<td>0.941</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.6</td>
<td>0.2</td>
<td>-0.00115</td>
<td>-0.00007</td>
<td>-0.00103</td>
<td>0.02721</td>
<td>0.00135</td>
<td>0.00222</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CP</td>
<td></td>
<td>0.943</td>
<td>0.940</td>
<td>0.911</td>
<td>0.698</td>
<td>0.940</td>
<td>0.934</td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>0.6</td>
<td>0.2</td>
<td>0.00028</td>
<td>0.00167</td>
<td>-0.00092</td>
<td>0.01849</td>
<td>0.00063</td>
<td>0.00060</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CP</td>
<td></td>
<td>0.948</td>
<td>0.940</td>
<td>0.931</td>
<td>0.730</td>
<td>0.945</td>
<td>0.940</td>
</tr>
<tr>
<td>(4)</td>
<td>25</td>
<td>0.6</td>
<td>0.6</td>
<td>0.00108</td>
<td>-0.00257</td>
<td>-0.00126</td>
<td>0.05043</td>
<td>0.00189</td>
<td>0.00123</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CP</td>
<td></td>
<td>0.947</td>
<td>0.919</td>
<td>0.911</td>
<td>0.649</td>
<td>0.950</td>
<td>0.917</td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>0.6</td>
<td>0.6</td>
<td>0.00020</td>
<td>0.00102</td>
<td>-0.00075</td>
<td>0.03433</td>
<td>0.00047</td>
<td>0.00062</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CP</td>
<td></td>
<td>0.948</td>
<td>0.934</td>
<td>0.935</td>
<td>0.682</td>
<td>0.949</td>
<td>0.923</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.6</td>
<td>0.6</td>
<td>-0.00119</td>
<td>-0.00003</td>
<td>-0.00049</td>
<td>0.03383</td>
<td>0.00073</td>
<td>0.00182</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CP</td>
<td></td>
<td>0.947</td>
<td>0.940</td>
<td>0.926</td>
<td>0.701</td>
<td>0.938</td>
<td>0.925</td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>0.6</td>
<td>0.6</td>
<td>0.00047</td>
<td>0.00160</td>
<td>-0.00077</td>
<td>0.02259</td>
<td>0.00051</td>
<td>0.00070</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CP</td>
<td></td>
<td>0.942</td>
<td>0.940</td>
<td>0.935</td>
<td>0.722</td>
<td>0.950</td>
<td>0.927</td>
</tr>
</tbody>
</table>

The true number of factor is 1 in the $z$ equation and 2 in the $y$ equation. The estimation, the number of factors is set at 1 for the $z$ equation and 3 for the $y$ equation. True parameter values: $\beta_0 = 1, \beta_\xi = 1, \sigma_0 = \sqrt{2}, \sigma_\xi = \sqrt{2}$ (hence $\alpha_0 = (\frac{1}{4})^{-\frac{1}{2}}$). In the low endogeneity case, $\rho_0 = 0.2, \alpha_\zeta = 0.88$ and $\delta_\zeta = 0.2$. In the high endogeneity case, $\rho_0 = 0.6, \alpha_\zeta = 1.08$ and $\delta_\zeta = 0.6$. $\hat{\theta}_{qT}$ is the bias-corrected QML estimator. The coverage probabilities (CP) are calculated using the theoretical standard deviations obtained from the diagonal elements of $\Sigma_{nT}^{-\frac{1}{2}}$. 

22
Figure 1: Frequencies of Incorrect Estimation

True parameter values: \( \beta_0 = 1, \beta_{00} = 1, \sigma_{00} = 1, \sigma_{000} = 1 \) (hence \( \alpha_0 = 1 \)), \( \rho_0 = 0.2, \alpha_{00} = 0.88 \) and \( \delta_0 = 0.2 \). True number of factors: 1 in the \( z \) equation and 2 in the \( y \) equation. Initial estimates assume 10 factors in both equations.
The outcome equation has two spatial weights matrices, one time invariant ($\lambda_1$) and the other is time variant ($\lambda_2$). True parameter values: $\beta_0 = 1$, $\beta_1 = 1$, $\sigma_0 = \sqrt{\frac{2}{3}}$, $\sigma_0 = \sqrt{\frac{2}{3}}$ (hence $\alpha_0 = (\frac{2}{3})^{-\frac{1}{2}}$). In the low endogeneity case, $\rho_0 = 0.2$, $\alpha_{0}^{\omega} = 0.88$ and $\delta_0 = 0$. In the high endogeneity case, $\rho_0 = 0.6$, $\alpha_{0}^{\omega} = 1.08$ and $\delta_0 = 0.6$. $\hat{\theta}_{nT}^{\omega}$ is the bias-corrected QML estimator. The coverage probabilities (CP) are calculated using the theoretical standard deviations obtained from the diagonal elements of $\frac{1}{mT} \Sigma_{nT}^{-1}$.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$T$</th>
<th>$\lambda_{10}$</th>
<th>$\lambda_{20}$</th>
<th>$\rho_0$</th>
<th>$\hat{\beta}_1$</th>
<th>$\hat{\beta}_2$</th>
<th>$\hat{\lambda}_1$</th>
<th>$\hat{\lambda}_2$</th>
<th>$\hat{\alpha}_{\omega}$</th>
<th>$\hat{\alpha}$</th>
<th>$\hat{\delta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>25</td>
<td>25</td>
<td>-0.2</td>
<td>0.2</td>
<td>Bias</td>
<td>-0.00076</td>
<td>-0.00004</td>
<td>0.00359</td>
<td>-0.00211</td>
<td>0.00688</td>
<td>0.00216</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.964</td>
<td>0.957</td>
<td>0.956</td>
<td>0.954</td>
<td>0.922</td>
<td>0.943</td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>25</td>
<td>-0.2</td>
<td>0.2</td>
<td>Bias</td>
<td>0.00026</td>
<td>0.00022</td>
<td>0.00115</td>
<td>-0.00090</td>
<td>0.00313</td>
<td>0.00098</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.952</td>
<td>0.932</td>
<td>0.938</td>
<td>0.944</td>
<td>0.919</td>
<td>0.941</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>49</td>
<td>-0.2</td>
<td>0.2</td>
<td>Bias</td>
<td>0.00044</td>
<td>-0.00089</td>
<td>0.00175</td>
<td>-0.00065</td>
<td>0.00234</td>
<td>0.00125</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.957</td>
<td>0.955</td>
<td>0.946</td>
<td>0.946</td>
<td>0.927</td>
<td>0.939</td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>49</td>
<td>-0.2</td>
<td>0.2</td>
<td>Bias</td>
<td>0.00077</td>
<td>-0.00012</td>
<td>0.00023</td>
<td>-0.00052</td>
<td>0.00132</td>
<td>0.00063</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.950</td>
<td>0.943</td>
<td>0.939</td>
<td>0.950</td>
<td>0.936</td>
<td>0.948</td>
</tr>
<tr>
<td>(2)</td>
<td>25</td>
<td>25</td>
<td>-0.2</td>
<td>0.6</td>
<td>Bias</td>
<td>-0.00083</td>
<td>-0.00029</td>
<td>0.00170</td>
<td>-0.00178</td>
<td>0.00783</td>
<td>0.00195</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.963</td>
<td>0.949</td>
<td>0.954</td>
<td>0.950</td>
<td>0.930</td>
<td>0.943</td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>25</td>
<td>-0.2</td>
<td>0.6</td>
<td>Bias</td>
<td>-0.00008</td>
<td>-0.00025</td>
<td>0.00083</td>
<td>-0.00085</td>
<td>0.00383</td>
<td>0.00049</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.952</td>
<td>0.936</td>
<td>0.950</td>
<td>0.945</td>
<td>0.925</td>
<td>0.947</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>49</td>
<td>-0.2</td>
<td>0.6</td>
<td>Bias</td>
<td>-0.00003</td>
<td>-0.00134</td>
<td>0.00198</td>
<td>-0.00044</td>
<td>0.00300</td>
<td>0.00093</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.955</td>
<td>0.954</td>
<td>0.934</td>
<td>0.941</td>
<td>0.935</td>
<td>0.940</td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>49</td>
<td>-0.2</td>
<td>0.6</td>
<td>Bias</td>
<td>0.00048</td>
<td>-0.00014</td>
<td>-0.00070</td>
<td>0.00027</td>
<td>0.00173</td>
<td>0.00034</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.948</td>
<td>0.951</td>
<td>0.947</td>
<td>0.940</td>
<td>0.942</td>
<td>0.947</td>
</tr>
<tr>
<td>(3)</td>
<td>25</td>
<td>25</td>
<td>-0.3</td>
<td>0.2</td>
<td>Bias</td>
<td>-0.00076</td>
<td>-0.00007</td>
<td>0.00333</td>
<td>-0.00193</td>
<td>0.00689</td>
<td>0.00216</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.964</td>
<td>0.956</td>
<td>0.957</td>
<td>0.956</td>
<td>0.920</td>
<td>0.943</td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>25</td>
<td>-0.3</td>
<td>0.2</td>
<td>Bias</td>
<td>0.00026</td>
<td>0.00020</td>
<td>0.00104</td>
<td>-0.00077</td>
<td>0.00314</td>
<td>0.00098</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.952</td>
<td>0.934</td>
<td>0.940</td>
<td>0.946</td>
<td>0.920</td>
<td>0.941</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>49</td>
<td>-0.3</td>
<td>0.2</td>
<td>Bias</td>
<td>0.00044</td>
<td>-0.00091</td>
<td>0.00155</td>
<td>-0.00052</td>
<td>0.00235</td>
<td>0.00125</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.957</td>
<td>0.953</td>
<td>0.944</td>
<td>0.946</td>
<td>0.926</td>
<td>0.939</td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>49</td>
<td>-0.3</td>
<td>0.2</td>
<td>Bias</td>
<td>0.00077</td>
<td>-0.00013</td>
<td>0.00016</td>
<td>-0.00049</td>
<td>0.00133</td>
<td>0.00063</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.950</td>
<td>0.942</td>
<td>0.939</td>
<td>0.949</td>
<td>0.937</td>
<td>0.948</td>
</tr>
<tr>
<td>(4)</td>
<td>25</td>
<td>25</td>
<td>-0.3</td>
<td>0.6</td>
<td>Bias</td>
<td>-0.00083</td>
<td>-0.00030</td>
<td>0.00158</td>
<td>-0.00175</td>
<td>0.00784</td>
<td>0.00195</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.963</td>
<td>0.948</td>
<td>0.954</td>
<td>0.952</td>
<td>0.930</td>
<td>0.943</td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>25</td>
<td>-0.3</td>
<td>0.6</td>
<td>Bias</td>
<td>-0.00008</td>
<td>-0.00026</td>
<td>0.00076</td>
<td>-0.00077</td>
<td>0.00383</td>
<td>0.00049</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.952</td>
<td>0.936</td>
<td>0.950</td>
<td>0.940</td>
<td>0.925</td>
<td>0.947</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>49</td>
<td>-0.3</td>
<td>0.6</td>
<td>Bias</td>
<td>-0.00003</td>
<td>-0.00135</td>
<td>0.00183</td>
<td>-0.00036</td>
<td>0.00300</td>
<td>0.00093</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.955</td>
<td>0.955</td>
<td>0.933</td>
<td>0.940</td>
<td>0.934</td>
<td>0.940</td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>49</td>
<td>-0.3</td>
<td>0.6</td>
<td>Bias</td>
<td>0.00048</td>
<td>-0.00014</td>
<td>-0.00077</td>
<td>0.00030</td>
<td>0.00173</td>
<td>0.00034</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.948</td>
<td>0.952</td>
<td>0.948</td>
<td>0.946</td>
<td>0.941</td>
<td>0.947</td>
</tr>
</tbody>
</table>
6. 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). HECM is the predominant type of reverse mortgages in the United States, 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 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 investigate the extent of spatial spillovers. The HECM activity in a state may be contemporaneously affected by developments in neighboring states. We are also interested in understanding the factors that may give rise to the spatial spillovers. We investigate two types of spatial effects, spillovers from neighboring states and spillovers due to large lenders. We hypothesize that two types of lenders exist. Small lenders concentrate on local markets while large lenders serve multiple states. Two neighboring states that have high concentrations of large lenders might be more closely connected because large lenders may be more aware of the developments in neighboring states. Our data covers 50 states plus Washington, D.C. and 52 quarters from 2001 to 2013. Table C shows that the HECM market is fragmented with a handful of large lenders and a large number of small lenders.

Let $y_{it}$ denote the HECM origination rate, defined as the number of newly originated HECM loans in
## Table 6: Largest HECM Lenders by Volume (2001-2013)

<table>
<thead>
<tr>
<th>Name</th>
<th># Loans</th>
<th>Share of Total # Loans (2001-2013)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wells Fargo Bank</td>
<td>157,772</td>
<td>20.35%</td>
</tr>
<tr>
<td>Financial Freedom Senior Funding Corp</td>
<td>47,187</td>
<td>6.09%</td>
</tr>
<tr>
<td>Bank of America</td>
<td>25,343</td>
<td>3.27%</td>
</tr>
<tr>
<td>One Reverse Mortgage LLC</td>
<td>23,431</td>
<td>3.02%</td>
</tr>
<tr>
<td>MetLife Bank</td>
<td>20,179</td>
<td>2.60%</td>
</tr>
<tr>
<td>Liberty Home Equity Solutions Inc</td>
<td>17,997</td>
<td>2.32%</td>
</tr>
<tr>
<td>American Advisors Group</td>
<td>17,136</td>
<td>2.21%</td>
</tr>
<tr>
<td>Seattle Mortgage Company</td>
<td>13,344</td>
<td>1.72%</td>
</tr>
<tr>
<td>Urban Financial of America LLC</td>
<td>10,357</td>
<td>1.34%</td>
</tr>
<tr>
<td>Generation Mortgage Company</td>
<td>9,557</td>
<td>1.23%</td>
</tr>
<tr>
<td>World Alliance Financial Corp.</td>
<td>9,241</td>
<td>1.19%</td>
</tr>
<tr>
<td>Reverse Mortgage USA Inc</td>
<td>9,112</td>
<td>1.18%</td>
</tr>
<tr>
<td>Everbank Reverse Mortgage LLC</td>
<td>8,304</td>
<td>1.07%</td>
</tr>
<tr>
<td>Ditech Mortgage Corp</td>
<td>7,479</td>
<td>0.96%</td>
</tr>
<tr>
<td>M and T Bank</td>
<td>6,480</td>
<td>0.84%</td>
</tr>
<tr>
<td>Countrywide Bank FSB</td>
<td>6,391</td>
<td>0.82%</td>
</tr>
<tr>
<td>American Reverse Mortgage Corp</td>
<td>6,231</td>
<td>0.80%</td>
</tr>
<tr>
<td>Academy Mortgage LLC</td>
<td>4,719</td>
<td>0.61%</td>
</tr>
<tr>
<td>First Mariner Bank</td>
<td>4,341</td>
<td>0.56%</td>
</tr>
<tr>
<td>Reverse Mortgage Solutions Inc</td>
<td>4,200</td>
<td>0.54%</td>
</tr>
</tbody>
</table>

## Figure 3: Average House Price Deviations and Volatility by US Regions
state $i$ at quarter $t$ as a percentage of the senior population (age 65 plus) in state $i$ from the 2010 census. There are two $n \times n$ spatial weights matrices, $W_{1,nt} = (w_{1,ij})$ and $W_{2,nt} = (w_{2,ijjt})$. $w_{1,ij} = 1$ if states $i$ and $j$ share the same border and $w_{1,ij} = 0$ otherwise. $W_{2,nt}$ is time varying and captures a possibly different spillover effect from large lenders. $w_{2,ijjt} = w_{1,ij} \times z_{it} \times z_{jt}$, where $z_{it}$ is the share of the HECM loans originated by the large lenders (defined as being one of the top 10 largest lenders in Table 3) in state $i$ at quarter $t$. A larger weight is given to a state with more dominant large lenders. House price dynamic variables are constructed using the Federal Housing Finance Agency’s quarterly all-transactions house price indexes (HPI) deflated by the CPI, which start from 1975. Following [Haurin et al., 2014], the house price dynamic variables include percentage deviations from the previous 9 year averages ($hpi_{dev}^{y}$), standard deviations of house price changes in the previous 9 years ($hpi_{dev}^{v}$), 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 lender activity and origination rates are affected by some macroeconomic factors, which are captured by the time factors $f_{zt}$ and $f_{yt}$. It is also likely that macroeconomic factors have different impacts in different states, as captured by state-specific factor loadings. Note that factor loadings may be spatially correlated, capturing residual spatial effects not directly modeled. Interactive effects include additive individual and time effects as special cases, hence state time invariant variables are not included. The model consists of

$$z_{it} = hpi_{dev}^{y} \beta_{z1} + hpi_{dev}^{v} \beta_{z2} + (hpi_{dev}^{y} \times hpi_{dev}^{v}) \beta_{z3} + \gamma_{zt}^{y} f_{zt} + \epsilon_{it},$$

and

$$y_{it} = \lambda_{1} \sum_{j=1}^{n} w_{1,ij}y_{jt} + \lambda_{2} \sum_{j=1}^{n} w_{2,ijjt} y_{jt} + hpi_{dev}^{y} \beta_{y1} + hpi_{dev}^{v} \beta_{y2} + (hpi_{dev}^{y} \times hpi_{dev}^{v}) \beta_{y3} + \gamma_{yt}^{y} f_{yt} + v_{it},$$

where, respectively, $\epsilon_{it}$ and $v_{it}$ have variances $\sigma^{2}_{\epsilon}$ and $\sigma^{2}_{v}$ with correlation $\rho$. Denote $\alpha = \sigma^{-1}_{\epsilon}$, $\alpha_{v} = ((1 - \rho^{2}) \sigma^{-2}_{v})^{-1}$ and $\delta = \rho \sigma_{\epsilon} \sigma^{-1}_{v}$.

We assume that the upper bound of the number of factors is 10, so preliminary estimates use 10 factors. According to the eigenvalue ratio criterion (see Proposition 3), it indicates that $z_{it}$ has one unobserved factor and so is $y_{it}$, and the growth ratio criterion gives the same result. The same number of factors is selected if 5 factors are used for the preliminary estimates. Figure 4 shows a heat map of the estimated factor loadings ($\gamma_{zt}$), where darker colors indicate higher values. A higher factor loading indicates that the time factor has a larger impact. Because the factor loadings of all states are positive, the corresponding time factors can be interpreted as positively impacting the take-up rates. Figure 5 plots the estimated time factor ($f_{zt}$) with the quarterly percentage change in the real house price index. The time factors generate increasing HECM activity prior to 2009 and decreasing HECM activity post 2009, and they appear to move opposite to the change in house prices with a lag. As the common factor component can account for the impact of an economy wide shock, we explore possible links between the estimated time factor and some observable

---

3 Alternatively, we define a large lender as one of the top 20 largest lenders. The results are similar.

4 As in [Haurin et al., 2014], the choice of 9 years (or 36 quarters) is ad hoc. The time window needs to be long enough to capture long term norms, while short enough to reflect changing economic circumstances. We have assumed that households continuously update their beliefs on the long term norms.
economic variables. The credit available from a HECM loan depends on the interest rate, the principal limit factor set by HUD, age of the borrower and the property value. Before April 2009, only HECMs with adjustable interest rates were available. Since April 2009, fixed rate HECMs became widely available. We include the average interest rate of adjustable rate HECMs and the average spread between the interest rate on an adjustable rate HECM and a fixed rate HECM. There were two reductions to the principal limit factors in our sample period, corresponding to the two dummy variables “After 2009Q4” and “After 2010Q4”. We also consider households’ income fluctuations, as measured by the GDP growth rate.

Table 7 reports the results of the factor regression. More HECM credit is available with lower interest rates and higher principal limit factors, and as expected, we find that these are associated with higher HECM activity. In addition, more HECMs were closed in economic downturns, suggesting that households who face negative income shocks use HECMs to supplement their income. It is interesting to note that because the time factor is changing over time and its factor loadings are heterogeneous, the variation that has been attributed to the common factor component ($\gamma_{yi}f_{yt}$) can not be explained entirely by additive individual and time effects.

Table 8 reports the main estimation results with one unobserved factor for both definitions of large lenders, namely, top 10 largest lenders and top 20 largest lenders. As a comparison, we also report the estimates of panel regressions with additive state fixed effects and time effects in the “FE” column. The results via the z equation show that small lenders tend to be more active in states with more volatile house prices and current house prices above long term norms. Such states also have higher origination rates, consistent with the findings of Haurin et al. (2014). Comparing column FE with the spatial panel regression reveals that the coefficients of the panel regression are overestimated when spatial effects are ignored. It is also interesting to note that spatial effects work through different channels. Higher HECM activity in a state

---

5Because fixed rate HECMs were not available before 2009 April, the spread is coded as 0 before 2009Q2.
Table 7: Regression of Time Factors on Observable Economic Variables

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Adjustable Rate HECM Interest Rate</td>
<td>-0.0852***</td>
<td>0.0298</td>
</tr>
<tr>
<td>Spread Between Adjustable Rate HECM and Fixed Rate HECM</td>
<td>0.0634</td>
<td>0.0566</td>
</tr>
<tr>
<td>After 2009Q4</td>
<td>0.0258</td>
<td>0.0391</td>
</tr>
<tr>
<td>After 2010Q4</td>
<td>-0.1203**</td>
<td>0.0462</td>
</tr>
<tr>
<td>GDP Growth Rate</td>
<td>-0.0190***</td>
<td>0.0057</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.38</td>
<td></td>
</tr>
<tr>
<td># obs</td>
<td>52</td>
<td></td>
</tr>
</tbody>
</table>

The dependent variable is the estimated time factor. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. 
Table 8: Estimation of State-Level Origination Rates

<table>
<thead>
<tr>
<th></th>
<th>Top 10 Coefficient</th>
<th>Top 20 Coefficient</th>
<th>FE Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top Lender Share</td>
<td>−0.1779( ^{***} )</td>
<td>−0.0278 ( ^{**} )</td>
<td>0.0227 ( ^{*} )</td>
</tr>
<tr>
<td>House Price Deviation</td>
<td>−1.3030( ^{***} )</td>
<td>−1.4777( ^{***} )</td>
<td>−1.9331( ^{***} )</td>
</tr>
<tr>
<td>Deviation ( \times ) Volatility</td>
<td>−0.6965( ^{**} )</td>
<td>−1.4172( ^{***} )</td>
<td>−2.3791( ^{***} )</td>
</tr>
<tr>
<td>HECM Origination Rate</td>
<td>( \lambda_1 ) 0.0965( ^{***} )</td>
<td>0.1033( ^{***} )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>−0.0687( ^{***} )</td>
<td>0.0134 ( ^{**} )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>−0.0283( ^{***} )</td>
<td>−0.0272( ^{***} )</td>
<td>−0.0010 ( ^{**} )</td>
</tr>
<tr>
<td></td>
<td>0.1210( ^{**} )</td>
<td>0.1236( ^{**} )</td>
<td>0.2722( ^{**} )</td>
</tr>
<tr>
<td></td>
<td>0.8868( ^{**} )</td>
<td>0.8772( ^{**} )</td>
<td>0.9433( ^{**} )</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>7.5808( ^{***} )</td>
<td>7.1922( ^{***} )</td>
<td></td>
</tr>
<tr>
<td>( \alpha_x )</td>
<td>104.6246( ^{**} )</td>
<td>104.7460( ^{**} )</td>
<td></td>
</tr>
<tr>
<td>( \delta )</td>
<td>0.0036( ^{***} )</td>
<td>0.0035( ^{***} )</td>
<td></td>
</tr>
</tbody>
</table>

The sample size is \( n = 51 \) and \( T = 52 \). Bias corrected estimates are reported. \( ^{***} p < 0.01, ^{**} p < 0.05, ^{*} p < 0.1 \).

positively influences neighboring states, possibly due to higher awareness from increased media coverage. However, the spillover effect is smaller for states with more large lenders. This suggests that large lenders may shift resources towards states with higher demand, resulting in lower activity in neighboring states. The net spillover effect is still positive, indicating that adverse selection may be slightly mitigated. The correlation between the share of large lender and the origination rate is statistically significant, although its magnitude is small.

7. Conclusion

This paper presents a spatial panel data model with endogenous spatial weights matrices and common factors. The spatial weights matrices can be time varying. Endogeneity arises because the weights are constructed from variables that may correlate with disturbances in the main equation. By treating time factors and factor loadings as fixed effects, the proposed QML estimator is robust to the presence of unobserved time varying heterogeneity, which may allow to correlate with included regressors. We show that the QML estimator is consistent and asymptotically normal in panels with large \( n \) and \( T \). Because the approximate score vector is not centered for the spatial interaction parameter and the variances parameters, the limiting distribution of the estimator has a bias term in the order \( OP \left( \frac{1}{\sqrt{nt}} \right) \). An analytical bias correction procedure is then proposed to remove the leading order bias. Monte Carlo simulations demonstrate good finite sample performance of the bias corrected estimator. An empirical application of the spatial panel model with time factors to the study of effects of house prices on state-level origination rates of the Home Equity Conversion Mortgage reveals some interesting patterns.
References


Appendix

A1. Proof of Claim 2

Because elements of $X_{ntz}$ are uniformly bounded, $\|X_{ntz,k}\|_2 \leq \|X_{ntz,k}\|_F = O(\sqrt{nT})$. Similarly, $\|Z_{k,nT}^y\|_2 = O(\sqrt{nT})$ for $k = 1, \cdots, k_y$.

Next consider $Z_{k_y+1,nT}^y = \sigma_\xi^{-1}(G_{n1}(X_{nt1},\beta_0 + \varepsilon_{n1}\delta_0 + \tilde{\Gamma}_{n0}f_{\xi10}), \cdots, G_{nT}(X_{ntT},\beta_0 + \varepsilon_{nT}\delta_0 + \tilde{\Gamma}_{n0}f_{\xiT0})$.

For this paragraph, denote $\zeta^{n*}_n = (\zeta_{n1}^{*'}, \cdots, \zeta_{nT}^{*'})$ with $\zeta_{ni} = \sigma_\xi^{-1}(X_{nti},\beta_0 + \varepsilon_{ni}\delta_0 + \tilde{\Gamma}_{n0}f_{\xi0})$. Therefore we have

$$\text{tr}\left(Z_{k_y+1,nT}^y Z_{k_y+1,nT}^{y'}\right) = \sigma_\xi^{-2} \sum_{t=1}^T (X_{ntT}\beta_0 + \tilde{\varepsilon}_n\delta_0 + \tilde{\Gamma}_{n0}f_{\xi0})' G_{nT}^T G_{nT} (X_{ntT}\beta_0 + \tilde{\varepsilon}_n\delta_0 + \tilde{\Gamma}_{n0}f_{\xi0})$$

$$= \zeta_n^{n*'} G_{nT}^T G_{nT} \zeta_n^{n*}.$$

By Proposition 2, $\mathbb{E}\zeta_n^{n*'} G_{nT}^T G_{nT} \zeta_n^{n*} = O(nT)$. It follows that $\|Z_{k_y+1,nT}^y\|_2 \leq \|Z_{k_y+1,nT}^y\|_F = O_P(\sqrt{nT})$ by Markov’s inequality.

For $k = 1, \cdots, p$, $Z_{k_y+1+k,nT}^y = \sigma_\xi^{-1}\tilde{\varepsilon}_{k,nT}$ where $\tilde{\varepsilon}_{k,nT} = (\tilde{\varepsilon}_{n1,k}, \cdots, \tilde{\varepsilon}_{nT,k})$ with $\tilde{\varepsilon}_{nt,k}$ being the $k$-th column of the $n \times p$ matrix $\tilde{\varepsilon}_n = (e_{n1}, \cdots, e_{nT})'$. Because elements of $\tilde{\varepsilon}_{k,nT}$ are i.i.d. across $i$ and $t$ and have uniformly bounded 4-th moments, by Latala (2005), $\|\tilde{\varepsilon}_{k,nT}\|_2 = O_P(\sqrt{\max(n,T)})$ and therefore $\|Z_{k_y+1+k,nT}^y\|_2 = O_P(\sqrt{\max(n,T)})$.

A2. Proof of Claim 3

Claim 1 shows that $\|\Xi_{nt}\|_2 = O_P\left(\sqrt{\max(n,T)}\right)$. Now we show that $\|U_{nT}\|_2 = O_P\left(\sqrt{nT} \left(n^{-\frac{1}{2}} + T^{-\frac{1}{2}}\right)\right)$ with $U_{nT} = (u_{n1}, \cdots, u_{nT})$ and $u_{nt} = G_{nt}\xi_{nt}$. By definition,

$$\|U_{nT}\|_2^4 = (\mu_1(U_{nT}^TU_{nT}))^2 = \mu_1(U_{nT}^TU_{nT}U_{nT}^TU_{nT}) = \|U_{nT}^TU_{nT}\|_2^2$$

$$\leq \|U_{nT}^TU_{nT}\|_F^2 = \sum_{t=1}^T \sum_{i=1}^n \left(\sum_{j=1}^n \left|U_{nT}[i,U_{nT}][i]_T\right|^2\right)^2.$$

The $(i,t)$ element of $U_{nT}$ is $[U_{nT}][i]_T = \sum_{j=1}^n G_{nt,i,j}\xi_{nt,j}$. Then

$$\|U_{nT}\|_2^4 \leq \sum_{t,s=1}^T \sum_{i',j'=1}^n \sum_{i,j=1}^n G_{nt,i,j}G_{nt,i',j'}G_{ns,i,j}G_{ns,i',j'}\mathbb{E}\xi_{nt,j}\xi_{ns,j}\xi_{nt,j'}\xi_{ns,j'}.$$

Because $\xi_{nt,i}$ are independently distributed over $t$ and $i$, $\mathbb{E}\xi_{nt,i}|E_{nT}| = 0$, $\mathbb{E}\xi_{nt,i}|E_{nT}| = \sigma_\xi^2$, $\mathbb{E}\xi_{nt,j}|E_{nT}| = \mathbb{E}(\xi_{nt,j})$ and $\mathbb{E}(\xi_{nt,i}|E_{nT}) = \mathbb{E}(\xi_{nt,i})$,

$$\mathbb{E}\|U_{nT}\|_F^4$$

$$\leq \sum_{t,s=1}^T \sum_{i',j'=1}^n \sum_{i,j=1}^n \mathbb{E}\left(G_{nt,i,j}^4G_{nt,i',j'}^4G_{ns,i,j}^4G_{ns,i',j'}^4\right)\sigma_\xi^8 + \sum_{t=1}^T \sum_{i'=1}^n \sum_{j'=1}^n \mathbb{E}\left(G_{nt,i,j}^2G_{nt,i',j'}^2\right)\mathbb{E}(\xi_{nt,i})^2 \mathbb{E}(\xi_{nt,j})^2 \mathbb{E}(\xi_{nt,i'}^2) \mathbb{E}(\xi_{nt,j'}^2)$$

$$= O_P\left(\sqrt{nT} \left(n^{-\frac{3}{2}} + T^{-\frac{3}{2}}\right)\right).$$
Let $R_z$ and $R_y$ denote the number of factors specified in the model, and $R_{z0}$ and $R_{y0}$ the true numbers of factors. We have assumed that $R_z \geq R_{z0}$ and $R_y \geq R_{y0}$. In order to prove $\| \bar{\theta} - \theta_0 \|_2 \xrightarrow{P} 0$, by Lemma 1 of [Wu (1981)], it is sufficient to show that, for any $\tau > 0$, lim $n,T \to \infty P (\inf_{\theta \in \Theta} | \bar{\theta} - \theta_0 | \geq \tau (Q_{nT}(\theta_0) - Q_{nT}(\theta)) > 0) = 1$, where $Q_{nT}(\theta)$ is the concentrated likelihood in Eq. (6). The proof consists of four steps.

1. There exists a lower bound $Q_{nT}(\theta_0)$ at $\theta_0$, such that $Q_{nT}(\theta_0) \geq Q_{nT}(\theta_0)$. Because

$$\left( \sum_{\xi_0}^{n} I_n \right) (z_{nt} - X_{nt} \beta_{\theta_0}) - \Gamma_{nc,f_{\tau}} = \left( \sum_{\xi_0}^{n} I_n \right) e_{nt} + \Gamma_{nc,f_{\tau}} f_{\theta_0} - \Gamma_{nc,f_{\tau}}$$,
and

\[
\sigma_{\xi_0}^{-1} \left[ S_m y_{nt} - X_{n_0} \beta_0 + (\delta_0^t \otimes I_n) (z_{nt} - X_{nt} \beta_t) \right] - \Gamma_{ny} f_{yt}
\]

\[
= \sigma_{\xi_0}^{-1} \left[ (\delta_0^t \otimes I_n) \epsilon_{nt} + \Gamma_{ny} f_{yt} + \xi_{nt} - (\delta_0^t \otimes I_n) (\Gamma_{n_0} f_{yt} + \epsilon_{nt}) \right] - \Gamma_{ny} f_{yt}
\]

\[
= \sigma_{\xi_0}^{-1} \xi_{nt} + \sigma_{\xi_0}^{-1} \left( \Gamma_{ny} f_{yt} - (\delta_0^t \otimes I_n) \Gamma_{n_0} f_{yt} \right) - \Gamma_{ny} f_{yt} = \sigma_{\xi_0}^{-1} \xi_{nt} + \Gamma_{ny} f_{yt} - \Gamma_{ny} f_{yt}, \text{ hence,}
\]

\[
\log L_{nt}(\theta_0, \Gamma_{n_0}, \Gamma_{ny}, F_{Tz}, F_{Ty})
\]

\[= - \frac{p + 1}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma_{e0}| - \frac{1}{2} \log \sigma_{\xi_0}^2 + \frac{1}{nT} \sum_{t=1}^{T} \log |S_m|
\]

\[= - \frac{1}{2nT} \sum_{t=1}^{T} \left[ \left( \Sigma_{e0}^{-1} \otimes I_n \right) \epsilon_{nt} + \Gamma_{ny} f_{yt} - \Gamma_{nz} f_{zt} \right] \left[ \left( \Sigma_{e0}^{-1} \otimes I_n \right) \epsilon_{nt} + \Gamma_{ny} f_{yt} - \Gamma_{nz} f_{zt} \right] \]

\[= - \frac{1}{2nT} \sum_{t=1}^{T} \left[ \sigma_{\xi_0}^{-1} \xi_{nt} + \Gamma_{ny} f_{yt} - \Gamma_{ny} f_{yt} \right] \left[ \sigma_{\xi_0}^{-1} \xi_{nt} + \Gamma_{ny} f_{yt} - \Gamma_{ny} f_{yt} \right], \text{ and,}
\]

\[
\log L_{nt}(\theta_0, \Gamma_{n_0}, \Gamma_{ny}, F_{Tz}, F_{Ty})
\]

\[= - \frac{p + 1}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma_{e0}| - \frac{1}{2} \log \sigma_{\xi_0}^2 + \frac{1}{nT} \sum_{t=1}^{T} \log |S_m|
\]

\[= - \frac{1}{2nT} \sum_{t=1}^{T} \left( \Sigma_{e0}^{-1} \otimes I_n \right) \epsilon_{nt} - \frac{1}{2nT} \sigma_{\xi_0}^2 \sum_{t=1}^{T} \xi_{nt}^2 \xi_{nt}
\]

\[= - \frac{p + 1}{2} \log(2\pi) + 1 - \frac{1}{2} \log |\Sigma_{e0}| - \frac{1}{2} \log \sigma_{\xi_0}^2 + \frac{1}{nT} \sum_{t=1}^{T} \log |S_m| + O_P \left( \frac{1}{\sqrt{nT}} \right).
\]

Therefore,

\[
Q_{nt}(\theta_0) \geq L_{nt}(\theta_0, \Gamma_{n_0}, \Gamma_{ny}, F_{Tz}, F_{Ty})
\]

\[= - \frac{p + 1}{2} \log(2\pi) + 1 - \frac{1}{2} \log |\Sigma_{e0}| - \frac{1}{2} \log \sigma_{\xi_0}^2 + \frac{1}{nT} \sum_{t=1}^{T} \log |S_m| + O_P \left( \frac{1}{\sqrt{nT}} \right).
\]

(A.1)

Note that this lower bound may not hold if the number of factors is under-specified, in which case the variation due to factors may not all be accounted for.

2. There exists a function \( \hat{Q}_{nt}(\cdot) \), such that \( Q_{nt}(\theta) \leq \hat{Q}_{nt}(\theta) \) for all \( \theta \in \Theta \).

\[
Q_{nt}(\theta)
\]

\[= - \frac{p + 1}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma_{e}| - \frac{1}{2} \log \sigma_{\xi}^2 + \frac{1}{nT} \sum_{t=1}^{T} \log |S_m(\lambda)|
\]

\[= - \frac{1}{2nT} \min_{\Gamma_{nt} \in \mathbb{R}^{R \times R}, F_{Tz} \in \mathbb{R}^{T \times R}, \lambda \in \mathbb{R}^{R \times R}} \sum_{t=1}^{T} \left[ \left( \Sigma_{e}^{-\frac{1}{2}} \otimes I_n \right) (z_{nt} - X_{nt} \beta_t) - \Gamma_{nz} f_{zt} \right] \left[ \left( \Sigma_{e}^{-\frac{1}{2}} \otimes I_n \right) (z_{nt} - X_{nt} \beta_t) - \Gamma_{nz} f_{zt} \right]
\]

(A.2)
Consider first Eq. (A.2).

\[
- \frac{1}{2nT} \min_{\Gamma_n \in \mathbb{R}^{p \times p}, F_f \in \mathbb{R}^{T \times k}} \sum_{t=1}^{T} \left[ (\sigma_e^{-1} (S_n (\lambda) y_{nt} - X_{nt}^T \beta_n - (\delta' \otimes I_n) (Z_{nt} - X_{nt}^T \beta_e)) - \Gamma_{nyf_t})' \right] \\
\cdot \left[ (\sigma_e^{-1} (S_n (\lambda) y_{nt} - X_{nt}^T \beta_n - (\delta' \otimes I_n) (Z_{nt} - X_{nt}^T \beta_e)) - \Gamma_{nyf_t}) \right].
\]

(A.3)

\[
\min_{\Gamma_n \in \mathbb{R}^{p \times p}, F_f \in \mathbb{R}^{T \times k}} \frac{1}{nT} \sum_{t=1}^{T} \left[ (\Sigma_e^{-1} \otimes I_n) (z_{nt} - X_{nt}^T \beta_z - \Gamma_{nzf_t})' \right] \\
\cdot \left[ (\Sigma_e^{-1} \otimes I_n) (z_{nt} - X_{nt}^T \beta_z - \Gamma_{nzf_t}) \right].
\]

(A.4)

\[
\geq b_2 \left\| \eta_c \right\|_2^2
\]

(A.5)

\[
+ \text{tr} (\Sigma_e^{-1} E_{e0}) - \frac{2R_c}{nT} \left\| \Sigma_e^{-1} \otimes I_n \right\|_2^2 ||E_{nt}E_{nt}'||_2 + O_p \left( \frac{1}{\sqrt{nT}} \right)
\]

(A.6)

\[
- \frac{2R_c}{nT} \left\| \Sigma_e^{-1} \otimes I_n \right\|_2^2 ||E_{nt}||_2^2 \sum_{k=1}^{k_c} \left\| \tilde{X}_{nt,k} \right\|_2 ||\beta_{0,k} - \beta_{e,k}||
\]

(A.7)

\[
= b_2 \left\| \eta_c \right\|_2^2 + \text{tr} (\Sigma_e^{-1} E_{e0}) + O_p \left( \left\| \eta_c \right\|_2 \left( n^{-\frac{1}{2}} + T^{-\frac{1}{2}} \right) \right) + O_p \left( \frac{1}{\sqrt{nT}} \right),
\]

(A.8)
where \( \eta_c = \beta_{c0} - \beta_c \). Eq. (A.5) is because
\[
\sum_{r=2R_{c1}+1}^{np} \mu_r \left( \frac{1}{nT} \left( \Sigma^{-1}_e \otimes I_n \right) (\eta_c \cdot Z_{nT}) \right) (\eta_c \cdot Z_{nT})' = \sum_{r=2R_{c1}+1}^{np} \mu_r \left( \frac{1}{nT} \left( \Sigma^{-1}_e \otimes I_n \right) (\eta_c \cdot Z_{nT}) \right) (\eta_c \cdot Z_{nT})' \]
(\text{A.9})
where the inequality in Eq. (A.9) can be found in Theorem 8.12 of Zhang (2011); the last inequality is due to Assumption [4] and \( b_c \) is some generic positive constant. In Eq. (A.6), we use the inequality that for a square matrix \( A \), \( |\text{tr}(A)| \leq \text{rank}(A) \|A\|_2 \) (see Zhang (2011)) and then \( \|E_{nT}\|_2 = O_p \left( \sqrt{\text{max}(n, T)} \right) \) from Claim 1. In Eq. (A.7), \( \|\tilde{X}_{nTz,k}\|_2 \leq O \left( \sqrt{nT} \right) \) from Claim 2. Next consider Eq. (A.3). Substituting in \( y_{nt} = \eta'_{nt} (X_{nt}(\beta_0 + (\delta'_n - \delta'_0) \otimes I_n)) + \tilde{\epsilon}_i \) and \( z_{nt} = X_{nt}(\beta_0 + \tilde{\epsilon}'_i) \), we have
\[
\min_{\Gamma_{ny} \in \mathbb{R}^{n \times k}, \tilde{\Gamma}_{ny} \in \mathbb{R}^{T \times k}} \frac{1}{nT} \sum_{t=1}^{T} \left\{ \left( S_{nt}(\lambda) y_{nt} - X_{nt}(\beta_0 - (\delta' \otimes I_n)) (z_{nt} - X_{nt}(\beta_c)) \right) - \Gamma_{ny} f_{yt} \right\} \]
\[
\geq \min_{\tilde{\Gamma}_{ny} \in \mathbb{R}^{n \times k}, \tilde{\Gamma}_{ny} \in \mathbb{R}^{T \times k}} \frac{1}{nT} \left\{ \left( S_{nt}(\lambda) y_{nt} - X_{nt}(\beta_0 - (\delta' \otimes I_n)) (z_{nt} - X_{nt}(\beta_c)) \right) - \tilde{\Gamma}_{ny} f_{yt} \right\} \]
where the inequality in Eq. (A.9) can be found in Theorem 8.12 of Zhang (2011); the last inequality is due to Assumption [4] and \( b_c \) is some generic positive constant. In Eq. (A.6), we use the inequality that for a square matrix \( A \), \( |\text{tr}(A)| \leq \text{rank}(A) \|A\|_2 \) (see Zhang (2011)) and then \( \|E_{nT}\|_2 = O_p \left( \sqrt{\text{max}(n, T)} \right) \) from Claim 1. In Eq. (A.7), \( \|\tilde{X}_{nTz,k}\|_2 \leq O \left( \sqrt{nT} \right) \) from Claim 2. Next consider Eq. (A.3). Substituting in \( y_{nt} = \eta'_{nt} (X_{nt}(\beta_0 + (\delta'_n - \delta'_0) \otimes I_n)) + \tilde{\epsilon}_i \) and \( z_{nt} = X_{nt}(\beta_0 + \tilde{\epsilon}'_i) \), we have
\[
\min_{\Gamma_{ny} \in \mathbb{R}^{n \times k}, \tilde{\Gamma}_{ny} \in \mathbb{R}^{T \times k}} \frac{1}{nT} \sum_{t=1}^{T} \left\{ \left( S_{nt}(\lambda) y_{nt} - X_{nt}(\beta_0 - (\delta' \otimes I_n)) (z_{nt} - X_{nt}(\beta_c)) \right) - \Gamma_{ny} f_{yt} \right\} \]
\[
\geq \min_{\tilde{\Gamma}_{ny} \in \mathbb{R}^{n \times k}, \tilde{\Gamma}_{ny} \in \mathbb{R}^{T \times k}} \frac{1}{nT} \left\{ \left( S_{nt}(\lambda) y_{nt} - X_{nt}(\beta_0 - (\delta' \otimes I_n)) (z_{nt} - X_{nt}(\beta_c)) \right) - \tilde{\Gamma}_{ny} f_{yt} \right\} \]
(\text{A.9})
where the inequality in Eq. (A.9) can be found in Theorem 8.12 of Zhang (2011); the last inequality is due to Assumption [4] and \( b_c \) is some generic positive constant. In Eq. (A.6), we use the inequality that for a square matrix \( A \), \( |\text{tr}(A)| \leq \text{rank}(A) \|A\|_2 \) (see Zhang (2011)) and then \( \|E_{nT}\|_2 = O_p \left( \sqrt{\text{max}(n, T)} \right) \) from Claim 1. In Eq. (A.7), \( \|\tilde{X}_{nTz,k}\|_2 \leq O \left( \sqrt{nT} \right) \) from Claim 2. Next consider Eq. (A.3). Substituting in \( y_{nt} = \eta'_{nt} (X_{nt}(\beta_0 + (\delta'_n - \delta'_0) \otimes I_n)) + \tilde{\epsilon}_i \) and \( z_{nt} = X_{nt}(\beta_0 + \tilde{\epsilon}'_i) \), we have
\[
\min_{\Gamma_{ny} \in \mathbb{R}^{n \times k}, \tilde{\Gamma}_{ny} \in \mathbb{R}^{T \times k}} \frac{1}{nT} \sum_{t=1}^{T} \left\{ \left( S_{nt}(\lambda) y_{nt} - X_{nt}(\beta_0 - (\delta' \otimes I_n)) (z_{nt} - X_{nt}(\beta_c)) \right) - \Gamma_{ny} f_{yt} \right\} \]
\[
\geq \min_{\tilde{\Gamma}_{ny} \in \mathbb{R}^{n \times k}, \tilde{\Gamma}_{ny} \in \mathbb{R}^{T \times k}} \frac{1}{nT} \left\{ \left( S_{nt}(\lambda) y_{nt} - X_{nt}(\beta_0 - (\delta' \otimes I_n)) (z_{nt} - X_{nt}(\beta_c)) \right) - \tilde{\Gamma}_{ny} f_{yt} \right\} \]
(\text{A.9})
where the inequality in Eq. (A.9) can be found in Theorem 8.12 of Zhang (2011); the last inequality is due to Assumption [4] and \( b_c \) is some generic positive constant. In Eq. (A.6), we use the inequality that for a square matrix \( A \), \( |\text{tr}(A)| \leq \text{rank}(A) \|A\|_2 \) (see Zhang (2011)) and then \( \|E_{nT}\|_2 = O_p \left( \sqrt{\text{max}(n, T)} \right) \) from Claim 1. In Eq. (A.7), \( \|\tilde{X}_{nTz,k}\|_2 \leq O \left( \sqrt{nT} \right) \) from Claim 2. Next consider Eq. (A.3). Substituting in \( y_{nt} = \eta'_{nt} (X_{nt}(\beta_0 + (\delta'_n - \delta'_0) \otimes I_n)) + \tilde{\epsilon}_i \) and \( z_{nt} = X_{nt}(\beta_0 + \tilde{\epsilon}'_i) \), we have
\[
\min_{\Gamma_{ny} \in \mathbb{R}^{n \times k}, \tilde{\Gamma}_{ny} \in \mathbb{R}^{T \times k}} \frac{1}{nT} \sum_{t=1}^{T} \left\{ \left( S_{nt}(\lambda) y_{nt} - X_{nt}(\beta_0 - (\delta' \otimes I_n)) (z_{nt} - X_{nt}(\beta_c)) \right) - \Gamma_{ny} f_{yt} \right\} \]
\[
\geq \min_{\tilde{\Gamma}_{ny} \in \mathbb{R}^{n \times k}, \tilde{\Gamma}_{ny} \in \mathbb{R}^{T \times k}} \frac{1}{nT} \left\{ \left( S_{nt}(\lambda) y_{nt} - X_{nt}(\beta_0 - (\delta' \otimes I_n)) (z_{nt} - X_{nt}(\beta_c)) \right) - \tilde{\Gamma}_{ny} f_{yt} \right\} \]
(\text{A.9})
\[
\begin{align*}
+ \min_{\tilde{\Gamma}_n^{\ast} \in \mathbb{R}^{2n \times 2n}} & \frac{1}{nT} \sum_{t=1}^{T} \left\{ \sigma^{-1}_\xi \left( (\delta^{\prime} \otimes I_n) X_{nTZ}(\beta_0 - \beta_c) \right) \right\}^\prime \tilde{\Gamma}_n^{\ast} \left\{ \sigma^{-1}_\xi \left( (\delta^{\prime} \otimes I_n) X_{nTZ}(\beta_0 - \beta_c) \right) \right\} \\
+ \min_{\tilde{\Gamma}_n^{\ast} \in \mathbb{R}^{2n \times 2n}} & \frac{2}{nT} \sum_{t=1}^{T} \left\{ \sigma^{-1}_\xi \left[ X_{nTY}(\beta_0 - \beta_c) + G_{nT}(X_{nTY}\beta_0 + \tilde{e}_{nT}\delta_0 + \tilde{\Gamma}_{n0}\tilde{f}_{n0}) (\lambda_0 - \lambda) + \tilde{e}_{nT}(\delta_0 - \delta) \right] \right\}^\prime \tilde{\Gamma}_n^{\ast} \left\{ \sigma^{-1}_\xi \left[ X_{nTY}(\beta_0 - \beta_c) + G_{nT}(X_{nTY}\beta_0 + \tilde{e}_{nT}\delta_0 + \tilde{\Gamma}_{n0}\tilde{f}_{n0}) (\lambda_0 - \lambda) + \tilde{e}_{nT}(\delta_0 - \delta) \right] \right\} \\
+ \min_{\tilde{\Gamma}_n^{\ast} \in \mathbb{R}^{2n \times 2n}} & \frac{1}{nT} \sum_{t=1}^{T} \left[ \sigma^{-1}_\xi S_n(\lambda) S^{-1}_n \xi_n \right]^\prime \tilde{\Gamma}_n^{\ast} \left[ \sigma^{-1}_\xi S_n(\lambda) S^{-1}_n \xi_n \right] \\
+ \min_{\tilde{\Gamma}_n^{\ast} \in \mathbb{R}^{2n \times 2n}} & \frac{2}{nT} \sum_{t=1}^{T} \left\{ \sigma^{-1}_\xi \left[ X_{nTY}(\beta_0 - \beta_c) + G_{nT}(X_{nTY}\beta_0 + \tilde{e}_{nT}\delta_0 + \tilde{\Gamma}_{n0}\tilde{f}_{n0}) (\lambda_0 - \lambda) + \tilde{e}_{nT}(\delta_0 - \delta) \right] \right\}^\prime \tilde{\Gamma}_n^{\ast} \left\{ \sigma^{-1}_\xi \left[ X_{nTY}(\beta_0 - \beta_c) + G_{nT}(X_{nTY}\beta_0 + \tilde{e}_{nT}\delta_0 + \tilde{\Gamma}_{n0}\tilde{f}_{n0}) (\lambda_0 - \lambda) + \tilde{e}_{nT}(\delta_0 - \delta) \right] \right\} \\
= & \sigma^{-2}_\xi \sum_{r=2R_{k+1}}^{n} \eta_r \left( \frac{1}{nT} \left( \eta_r \cdot Z_{nT}^r \right) \right) \left( \eta_r \cdot Z_{nT}^r \right) + O_P \left( \left\| \beta_0 - \beta_c \right\|_2^2 \right) + O_P \left( \left\| \beta_0 - \beta_c \right\|_2 \left\| \eta_r \right\|_2 \right) \\
+ & \frac{1}{nT} \sum_{t=1}^{T} \sigma^{-2}_\xi S_n(\lambda) S^{-1}_n \xi_n - \max_{\tilde{\Gamma}_n^{\ast} \in \mathbb{R}^{2n \times 2n}} \frac{1}{nT} \sum_{t=1}^{T} \left[ \sigma^{-2}_\xi S_n(\lambda) S^{-1}_n \xi_n \right]^\prime \tilde{\Gamma}_n^{\ast} \left[ \sigma^{-2}_\xi S_n(\lambda) S^{-1}_n \xi_n \right] (A.10) \\
+ & \frac{2}{nT} \sum_{t=1}^{T} \sigma^{-2}_\xi \left[ X_{nTY}(\beta_0 - \beta_c) + G_{nT}(X_{nTY}\beta_0 + \tilde{e}_{nT}\delta_0 + \tilde{\Gamma}_{n0}\tilde{f}_{n0}) (\lambda_0 - \lambda) \right] \\
+ & \tilde{e}_{nT}(\delta_0 - \delta) - (\delta^{\prime} \otimes I_n) X_{nTZ}(\beta_0 - \beta_c) \right\}^\prime S_n(\lambda) S^{-1}_n \xi_n (A.11) \\
- & \max_{\tilde{\Gamma}_n^{\ast} \in \mathbb{R}^{2n \times 2n}} \frac{1}{nT} \sum_{t=1}^{T} \left( S_n(\lambda) S^{-1}_n \xi_n \right) \left( X_{nTY}(\beta_0 - \beta_c) + G_{nT}(X_{nTY}\beta_0 + \tilde{e}_{nT}\delta_0 + \tilde{\Gamma}_{n0}\tilde{f}_{n0}) (\lambda_0 - \lambda) \right) \\
+ & \tilde{e}_{nT}(\delta_0 - \delta) - (\delta^{\prime} \otimes I_n) X_{nTZ}(\beta_0 - \beta_c) \right\}^\prime \tilde{\Gamma}_n^{\ast} \left( S_n(\lambda) S^{-1}_n \xi_n \right) (A.12) \\
\geq & b_y \left\| \eta_r \right\|_2^2 + O_P \left( \left\| \beta_0 - \beta_c \right\|_2^2 \right) + O_P \left( \left\| \beta_0 - \beta_c \right\|_2 \left\| \eta_r \right\|_2 \right) + \frac{\sigma^2_\xi}{\sigma^2_\xi} \frac{1}{nT} \sum_{t=1}^{T} \left( S_n^{-1}_n S_n(\lambda) S^{-1}_n \right) \\
+ & O_P \left( \left\| \eta_r \right\|_2 \left( n^{-\frac{1}{2}} + T^{-\frac{1}{2}} \right) \right) + O_P \left( \left( \left\| \eta_r \right\|_2 + \left\| \eta_r \right\|_2 \left\| \eta_r \right\|_2 \right) \left( n^{-\frac{1}{2}} + T^{-\frac{1}{2}} \right) \right) + O_P \left( \frac{1}{\sqrt{nT}} \right). (A.13)
\end{align*}
\]

Under Assumption 4, \( b_y > 0 \). If there are collinearity between \( G_n \left( X_{nTY}\beta_0 + \tilde{e}_{nT}\delta_0 + \tilde{\Gamma}_{n0}\tilde{f}_{n0} \right) \), \( X_{nTY} \) and \( \tilde{e}_{nT} \), then Assumption 4 will not be satisfied. However, with \( \lambda \) identified from the alternative Assumption 5(2), we can still have \( b_y > 0 \) from Assumption 5(1). In Eq. (A.10), \( \frac{1}{nT} \sum_{t=1}^{T} \sigma^{-2}_\xi S_n(\lambda) S^{-1}_n \xi_n \xi_n S_n(\lambda) S^{-1}_n = O_P \left( \frac{1}{\sqrt{nT}} \right) \) from Lemma 5(1), and

\[
\begin{align*}
\max_{\tilde{\Gamma}_n^{\ast} \in \mathbb{R}^{2n \times 2n}} & \frac{1}{nT} \sum_{t=1}^{T} \left( S_n^{-1}_n S_n(\lambda) S^{-1}_n \xi_n \xi_n S_n(\lambda) S^{-1}_n \xi_n \right) \left( S_n^{-1}_n S_n(\lambda) S^{-1}_n \xi_n \right) \left( S_n^{-1}_n S_n(\lambda) S^{-1}_n \xi_n \right) \\
\leq & \max_{\tilde{\Gamma}_n^{\ast} \in \mathbb{R}^{2n \times 2n}} \frac{1}{nT} \sum_{t=1}^{T} \left( \xi_n \xi_n P_{\tilde{\Gamma}_n^{\ast}} \right) + \max_{\tilde{\Gamma}_n^{\ast} \in \mathbb{R}^{2n \times 2n}} \frac{2}{nT} \left( \lambda_0 - \lambda \right) \sum_{t=1}^{T} \left( G_n \xi_n \xi_n P_{\tilde{\Gamma}_n^{\ast}} \right)
\end{align*}
\]

39
\[ Q_{nT}(\theta) \le -\frac{p+1}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma_x| - \frac{1}{2} \log \sigma_x^2 + \frac{1}{nT} \sum_{t=1}^{T} \log |S_{nt}(\lambda)| - \frac{1}{2} \frac{\sigma_x^2}{\sigma_\xi^2} \sum_{t=1}^{T} \text{tr}(S_{nt}^{-1} S_{nt}(\lambda)' S_{nt}(\lambda) S_{nt}^{-1}) \\
- \frac{1}{2} \left[ b_y \left( \eta_y, \Sigma_y^{-1} \eta_y \right) + O_p \left( \left( \eta_y, \Sigma_y^{-1} \eta_y \right)^2 \right) \right] + O_p \left( \left( \eta_y, \Sigma_y^{-1} \eta_y \right)^2 \right) + O_p \left( \left( \eta_y, \Sigma_y^{-1} \eta_y \right)^2 \right) \right] \\
+ O_p \left( n^{-1} + T^{-1} \right), \quad (A.16) \]

Notice that with a slight abuse of notation, we list \( Q_{nT}(\theta) \) separately from \( Q_{nT}(\theta) \) because the former terms come from

\[ \frac{1}{nT} \sum_{t=1}^{T} \left\{ \sigma_\xi^{-2} \left[ X_{nty}(\beta_0 - \beta_0) + G_{nt} \left( X_{nty}, \beta_0 + \xi_{nty} \delta_0 + \Gamma_{nty} f_{nty} \right)(\lambda_0 - \lambda) + \xi_{nty}(\delta_0 - \delta) - \left( \delta' \otimes I_n \right) X_{nty} (\beta_0 - \beta_0) \right] \right\}' M_{\xi}_{nty} \]

\[ \left[ X_{nty}(\beta_0 - \beta_0) + G_{nt} \left( X_{nty}, \beta_0 + \xi_{nty} \delta_0 + \Gamma_{nty} f_{nty} \right)(\lambda_0 - \lambda) + \xi_{nty}(\delta_0 - \delta) - \left( \delta' \otimes I_n \right) X_{nty} (\beta_0 - \beta_0) \right], \quad (A.17) \]

and \( M_{\xi}_{nty} \) is positive semi-definite, and therefore it is non-negative; while the later terms come from \( Q_{nT}(\theta) \) and \( Q_{nT}(\theta) \). In the following analysis, we can take \( Q_{nT}(\theta) \) from the right hand side of the preceding inequality for \( Q_{nT}(\theta) \).

3. \( \lim \inf_{n, T \to \infty} P \left( \inf_{\theta \in \Theta, \|\theta - \theta_0\|_2 \ge \tau} \left\{ Q_{nT}(\theta) - Q_{nT}(\theta) \right\} > 0 \right) = 1 \) for any \( \tau > 0 \). From \( Q_{nT}(\theta) \) and \( Q_{nT}(\theta) \),

\[ Q_{nT}(\theta) - Q_{nT}(\theta) \]
By the inequality of arithmetic and geometric means,

\[
\frac{1}{nT} \sum_{t=1}^{T} \text{tr} (S_n(\lambda)'S_n(\lambda)S_{nt}^{-1}S_{nt}' - 1) \geq \frac{1}{nT} \sum_{t=1}^{T} \left| \prod_{t=1}^{T} |S_n(\lambda)'S_n(\lambda)S_{nt}^{-1}S_{nt}'|^{\frac{1}{nT}} - 1 \right|
\]

(A.19)

\[
\frac{1}{p} \text{tr} (\Sigma_{\epsilon_0}\Sigma^{-1}_{\epsilon_0}) \geq \left| \Sigma_{\epsilon_0}\Sigma^{-1}_{\epsilon_0} \right|^\frac{1}{p}.
\]

(A.20)

Furthermore, because \( b_{\gamma} \| \eta_{\gamma} \|^2 \geq O_P \left( \| \eta_{\gamma} \|^2 \right) + O_P \left( \| \eta_{\gamma} \|_2 \| \eta_{\gamma} \|_2 \right) \geq 0 \) and \( \| \eta_{\gamma} \|_2 \) is bounded,

\[
Q_{nT}(\theta_0) - Q_{nT}(\theta) \geq \frac{1}{2} \left[ b_{\gamma} \| \eta_{\gamma} \|^2 + o_P(1) \right]
\]

(A.21)

\[
\frac{1}{2} \left( - \log \left[ \prod_{t=1}^{T} |S_n(\lambda)'S_n(\lambda)S_{nt}^{-1}S_{nt}'|^\frac{1}{nT} \right] \right) + \frac{1}{2} \left( - \log \left[ \prod_{t=1}^{T} |S_n(\lambda)'S_n(\lambda)S_{nt}^{-1}S_{nt}'|^\frac{1}{nT} \right] \right) - 1
\]

(A.22)

\[
\frac{1}{p} \text{tr} (\Sigma_{\epsilon_0}\Sigma^{-1}_{\epsilon_0}) \geq \left| \Sigma_{\epsilon_0}\Sigma^{-1}_{\epsilon_0} \right|^\frac{1}{p}.
\]

For any real number \( x > 0 \), \( -\log x + x - 1 \geq 0 \) with equality holds only if \( x = 1 \). Because \( \frac{1}{p} \text{tr} (\Sigma_{\epsilon_0}\Sigma^{-1}_{\epsilon_0}) > \left| \Sigma_{\epsilon_0}\Sigma^{-1}_{\epsilon_0} \right|^\frac{1}{p} \) unless \( \Sigma_{\epsilon} = c\Sigma_{\epsilon_0} \) for some scalar \( c \), and \( -\log \left| \Sigma_{\epsilon_0}\Sigma^{-1}_{\epsilon_0} \right|^\frac{1}{p} + \left| \Sigma_{\epsilon_0}\Sigma^{-1}_{\epsilon_0} \right|^\frac{1}{p} - 1 > 0 \) if \( c \neq 1 \), we have \( Q_{nT}(\theta_0) - Q_{nT}(\theta) > 0 \) if \( \Sigma_{\epsilon} \neq \Sigma_{\epsilon_0} \) with probability approaching (wpa) 1. Because the mapping from \( \Sigma_{\epsilon} \) to \( \alpha \) is one to one, \( Q_{nT}(\theta_0) - Q_{nT}(\theta) > 0 \) if \( \alpha \neq \alpha_0 \) wpa 1. For the other parameters, consider the case with Assumption \([n]\) first. If any of the following holds for some \( \tau > 0 \), \( \| \beta_0 - \beta_\xi \|_2 \geq \tau \), \( \| \beta_0 - \beta_\eta \|_2 \geq \tau \), \( \| \lambda_0 - \lambda \| \geq \tau \), \( \| \delta_0 - \delta \|_2 \geq \tau \), Eq. (A.21) and/or Eq. (A.22) is strictly greater than 0 wpa 1, and hence
\( Q_{nT}(\theta_0) - Q_{nT}(\theta) > 0 \) wpa 1. Alternatively, consider the case with Assumption 5. The inequality in Eq. (A.19) holds strictly for any \( \lambda \neq \lambda_0 \) wpa 1. On the other hand, with \( \lambda = \lambda_0 \), Eq. (A.21) and/or Eq. (A.22) is strictly positive wpa 1 if for some \( \tau > 0 \), either \( \|\beta_0 - \beta_*\|_2 \geq \tau \), \( \|\beta_0 - \beta_*\|_2 \geq \tau \), or \( \|\delta_0 - \delta_*\|_2 \geq \tau \) holds.

4. The rate of convergence. Let \( \hat{\theta} \) denote the QML estimator, \( \hat{\gamma}_T = \beta_0 - \beta_* \) and \( \hat{\gamma}_y = (\beta_0 - \beta_*, \lambda_0 - \lambda, \delta_0 - \delta)' \). Let \( c_{nT} = \min(n, T) \). By definition, \( Q_{nT}(\hat{\theta}) - Q_{nT}(\theta_0) \geq 0 \), which implies from Eq. (A.18),

\[
\begin{align*}
& b_y \|\hat{\gamma}_y\|_2^2 + O_P \left( \|\hat{\gamma}_T\|_2^\gamma \right) + O_P \left( \|\hat{\gamma}_2\|_T \right) \\
& + b_z \|\hat{\gamma}_z\|_2^2 + O_P \left( \|\hat{\gamma}_z\|_2 \|\hat{\gamma}_2\|_T \right) + O_P \left( \|\hat{\gamma}_y\|_2^\gamma \right) + O_P \left( \|\hat{\gamma}_z\|_2^\gamma \right) + O_P \left( c_{nT}^{-1} \right) \leq 0. \\
& \text{(A.23)}
\end{align*}
\]

Completing the squares, we have from (A.23),

\[
\begin{align*}
& \left( b_y \|\hat{\gamma}_y\|_2^2 + \frac{1}{2} a_{n_T} b_{n_T} \right) \left( \|\hat{\gamma}_2\|_T \right) \leq 0, \\
& \text{(A.24)}
\end{align*}
\]

with \( b_{n_T} = b_y + O_P \left( \|\beta_2\| \right), a_{n_T} = O_P \left( \|\beta_2\| \right) + O_P \left( \|\hat{\gamma}_2\|_2 \right) \) and \( d_{n_T} = O_P \left( \|\hat{\gamma}_2\|_2 \right) \) + \( b_z \|\hat{\gamma}_z\|_2^2 + O_P \left( \|\hat{\gamma}_z\|_2 \|\hat{\gamma}_2\|_T \right) \). Recall that \( b_y \|\hat{\gamma}_y\|_2^2 + O_P \left( \|\hat{\gamma}_2\|_2 \right) + O_P \left( \|\hat{\gamma}_z\|_2 \|\hat{\gamma}_2\|_T \right) \geq 0 \) (see (A.17)), then (A.23) implies that

\[
\begin{align*}
& b_z \|\hat{\gamma}_z\|_2^2 + O_P \left( \|\hat{\gamma}_z\|_2 \|\hat{\gamma}_2\|_T \right) + O_P \left( \|\hat{\gamma}_y\|_2^\gamma \right) + O_P \left( \|\hat{\gamma}_z\|_2 \|\hat{\gamma}_2\|_T \right) \leq 0. \\
& \text{(A.25)}
\end{align*}
\]

Similarly, by completing the squares, from (A.25),

\[
\begin{align*}
& \left( b_z \|\hat{\gamma}_z\|_2^2 + \frac{1}{2} a_{n_T} b_{n_T} \right) \left( \|\hat{\gamma}_2\|_T \right) \leq 0, \\
& \text{(A.26)}
\end{align*}
\]

Because the parameter spaces are bounded, \( d_{n_T} = O_P \left( \|\hat{\gamma}_2\|_2 \right) \), \( b_y \|\hat{\gamma}_y\|_2^2 + \frac{1}{2} a_{n_T} b_{n_T} \left( \|\hat{\gamma}_2\|_T \right) = O_P \left( \|\hat{\gamma}_2\|_2 \right) \). As \( a_{n_T} = O_P \left( \|\hat{\gamma}_2\|_2 \right) \), we then have \( \|\hat{\gamma}_2\|_2 = O_P \left( \|\hat{\gamma}_2\|_2 \right) \).

Now we proceed to show that \( \|\hat{\gamma}_2\|_2 = O_P \left( \|\hat{\gamma}_2\|_2 \right) \). Assuming that for some \( \frac{1}{2} \leq a < \frac{1}{2} \), \( \|\hat{\gamma}_2\|_2 = O_P \left( c_{n_T}^{-a} \right) \). Substituting \( \|\hat{\gamma}_2\|_2 = O_P \left( c_{n_T}^{-a} \right) \) into Eq. (A.24), we have \( d_{n_T} = O_P \left( c_{n_T}^{-2a} \right) \), \( a_{n_T} = O_P \left( c_{n_T}^{-a} \right) \). Therefore \( \|\hat{\gamma}_2\|_2 = O_P \left( c_{n_T}^{-a} \right) \). Substituting \( \|\hat{\gamma}_2\|_2 = O_P \left( c_{n_T}^{-a} \right) \) into Eq. (A.26), we have \( d_{n_T} = O_P \left( c_{n_T}^{-2a} \right) \) and \( \|\hat{\gamma}_2\|_2 = O_P \left( \max \left( c_{n_T}^{-1-a}, c_{n_T}^{-2a} \right) \right) \). Note that for \( a < \frac{1}{2} \), \( -a > \max \left( -a, -a, \frac{1}{2} - a \right) \), which gives \( \|\hat{\gamma}_2\|_2 = O_P \left( \max \left( c_{n_T}^{-1-a}, c_{n_T}^{-2a} \right) \right) \). Therefore, \( \|\hat{\gamma}_2\|_2 = O_P \left( c_{n_T}^{-1-a} \right) \).
Then \( \| \hat{\eta} \|_2 = O_P \left( \frac{c_{nI}}{nT} \right) \) follows from Eq. (A.24).

**Lemma 5.** Let \( \tilde{\xi}_{nT} = \sigma_{\xi}^{-1} (\tilde{\xi}_{n1}, \cdots, \tilde{\xi}_{nT})' \). Under Assumptions 3 and 6

\[
\frac{1}{nT} \sum_{t=1}^{T} \tilde{\xi}_{nt} S_{nt}^{-1} S_{nt}' (\lambda)' S_{nt} (\lambda) S_{nt}^{-1} \tilde{\xi}_{nt} = \frac{1}{nT} \sum_{t=1}^{T} \text{tr} \left( S_{nt}^{-1} S_{nt}' (\lambda)' S_{nt} (\lambda) S_{nt}^{-1} \right) + O_P \left( \frac{1}{\sqrt{nT}} \right).
\]

**Proof.** We have

\[
\frac{1}{nT} \sum_{t=1}^{T} \tilde{\xi}_{nt} S_{nt}^{-1} S_{nt}' (\lambda)' S_{nt} (\lambda) S_{nt}^{-1} \tilde{\xi}_{nt} = \frac{1}{nT} \tilde{\xi}_{nT}' (I_{nT} + (\lambda_0 - \lambda) (\tilde{G}_{nT} + \tilde{G}_{nT}) + (\lambda_0 - \lambda)^2 \tilde{G}_{nT}' \tilde{G}_{nT}) \tilde{\xi}_{nT} = \frac{1}{nT} \text{tr} \left( I_{nT} + (\lambda_0 - \lambda) (\tilde{G}_{nT} + \tilde{G}_{nT}) + (\lambda_0 - \lambda)^2 \tilde{G}_{nT}' \tilde{G}_{nT} \right) + O_P \left( \frac{1}{\sqrt{nT}} \right) \tag{A.27}
\]

\[
= \frac{1}{nT} \sum_{t=1}^{T} \text{tr} \left( S_{nt}^{-1} S_{nt}' (\lambda)' S_{nt} (\lambda) S_{nt}^{-1} \right) + O_P \left( \frac{1}{\sqrt{nT}} \right) \tag{A.28}
\]

where Eq. (A.27) is from Lemma 2. Let \( t_{k,nT} \) be an \( nT \times 1 \) vector with one in its \( k \)-th entry and zeros in its other entries. Because \( \text{tr} \left( \tilde{G}_{nT} \right) = \sum_{k=1}^{T} g_{k,nT} \) with \( g_{k,nT} = t_{k,nT} \tilde{G}_{nT} t_{k,nT} \) and \( g_{k,nT} \) satisfies the NED properties, Eq. (A.28) follows from Qu and Lee (2015).

**A4. Components of the Asymptotic Distribution**

As Eq. (11) shows, the asymptotic distribution depends on the approximate Hessian matrix \( \tilde{C}_{nT} \) and the approximate gradient vector \( \tilde{C}_{nT}^{(1)} \) whose components can be collected from Eq. (10) and further simplified using Lemmas 2 and 3. Recall that the parameter vector is \( \theta = (\beta_0, \beta_1, \lambda, \alpha_\xi, \alpha', \delta')' \).

(1) **Elements of \( \tilde{C}_{nT}^{(1)} \) (the Approximate Score Vector)**

The following lemma establishes that some of the second order terms in Eqs. (8) and (9) do not appear in the leading order term of the score vector.

**Lemma 6.** Under Assumptions 2 and 3

\[
\frac{1}{\sqrt{nT}} \text{tr} \left( M_{nT} V_k^{2} M_{F_T} E_{nT}' \left( \sum_{e_0}^{-\frac{1}{2}} \otimes I_n \right) P_{n,nT,F_T} E_{nT}' \left( \sum_{e_0}^{-\frac{1}{2}} \otimes I_n \right) \right) = o_P(1),
\]

\[
\frac{1}{\sqrt{nT}} \text{tr} \left( M_{nT} \left( \sum_{e_0}^{-\frac{1}{2}} \otimes I_n \right) E_{nT} M_{F_T} V_k^{2} P_{n,nT,F_T} E_{nT}' \left( \sum_{e_0}^{-\frac{1}{2}} \otimes I_n \right) \right) = o_P(1),
\]

\[
\frac{1}{\sqrt{nT}} \text{tr} \left( M_{nT} \left( \sum_{e_0}^{-\frac{1}{2}} \otimes I_n \right) E_{nT} M_{F_T} E_{nT}' \left( \sum_{e_0}^{-\frac{1}{2}} \otimes I_n \right) P_{n,nT,F_T} V_k^{2} \right) = o_P(1),
\]

43
for $k = 1, \cdots , k_z + J$, and $V_k^y$ are from Section 4.3.

\[
\frac{1}{\sqrt{nT}} \text{tr} \left( M_{Fy} V_k^y \Sigma_n M_{Fy} \Sigma_n' P_{Fy} \Sigma_n' \right) = \frac{1}{\sqrt{nT}} \text{tr} \left( M_{Fy} \Sigma_n M_{Fy} V_k^y P_{Fy} \Sigma_n' \right) = o_p(1),
\]

\[
\frac{1}{\sqrt{nT}} \text{tr} \left( M_{Fy} \Sigma_n M_{Fy} \Sigma_n' P_{Fy} V_k^y \right) = o_p(1),
\]

for $k = 1, \cdots , k_z + k_y + 2 + p$, and $V_k^y$ are from Section 4.3.

Proof. From Moon and Weidner [2014] (Lemma B.1), $\| P_{Fy} \|_2 = O_p \left( \frac{1}{\sqrt{nT}} \right)$ and $\| P_{Fy} \|_2 = O_p \left( \frac{1}{\sqrt{nT}} \right)$. They also show that the above terms are $o_p(1)$ if the regressor contains only exogenous components. Their result can be applied here except for the three terms, $- (\Omega_k \otimes I_n) E_n T$ in $V_k^y$ for $k = 1, \cdots , J$, $\sigma_{\xi_0}^{-1} (G_{n1} \xi_{n1}, \cdots , G_{nT} \xi_{nT})$ in $V_k^y$ for $k_z + k_y + 1$, and $- \Sigma_n$ in $V_k^y$. However, because these terms are $o_p \left( \sqrt{nT} \right)$, the relevant terms are $o_p(1)$ by using the inequality $|tr(A)| \leq \text{rank}(A) \|A\|_2$ for a square matrix $A$. \square

The followings are the components in the score vector. The $o_p(1)$ terms in the first lines are due to Lemma [5]

- $\beta_z$: for $k = 1, \cdots , k_z$,

\[
\tilde{C}_{nT,k}^{(1)} = \frac{1}{\sqrt{nT}} \text{tr} \left( M_{Fz} \tilde{\Sigma}_{nT,z,k} (\Sigma_{\xi_0}^{-1} \otimes I_n) M_{Fz} \Sigma_{nT} \right) - \frac{1}{\sqrt{nT}} \text{tr} \left( M_{Fz} \tilde{\Sigma}_{nT,z} (\delta_0 \otimes I_n) M_{Fz} \Sigma_{nT} \right) + o_p(1)
\]

\[
= \frac{1}{\sqrt{nT}} \left( a_{1nT}^{(k)} \tilde{e}_{nT} + b_{1nT}^{(k)} \tilde{\xi}_{nT} \right) + o_p(1),
\]

where $a_{1nT}^{(k)} = \text{vec} \left( M_{Fz} (\Sigma_{\xi_0}^{-1} \otimes I_n) \tilde{X}_{nT,z,k} M_{Fz} \right)$ and $b_{1nT}^{(k)} = - \sigma_{\xi_0}^{-1} \text{vec} \left( M_{Fy} (\delta_0 \otimes I_n) \tilde{X}_{nT,z,k} M_{Fy} \right)$.

- $\beta_y$: for $k = 1, \cdots , k_y$,

\[
\tilde{C}_{nT,k+k}^{(1)} = \frac{1}{\sqrt{nT}} \text{tr} \left( M_{Fy} \tilde{X}_{nT,y,k} M_{Fy} \Sigma_{nT} \right) + o_p(1) = \frac{1}{\sqrt{nT}} b_{2nT}^{(1)} \tilde{\xi}_{nT} + o_p(1),
\]

where $b_{2nT}^{(1)} = \sigma_{\xi_0}^{-1} \text{vec} \left( M_{Fy} \tilde{X}_{nT,y,k} M_{Fy} \right)$.

- $\lambda$:

\[
\tilde{C}_{nT,k+k+1}^{(1)} = - \frac{1}{\sqrt{nT}} \text{tr} \left( \tilde{G}_{nT} \right) + \frac{1}{\sqrt{nT}} \Sigma_{\xi_0}^{-1} \tilde{e}_{nT} (M_{Fy} \otimes M_{Fy}) \tilde{G}_{nT} \left( \tilde{X}_{nT,y} \beta_0 + (I_T \otimes \tilde{I}_y) \tilde{F}_{Ty} + \tilde{e}_{nT} \Sigma_{\xi_0}^{-1} \delta_0 \right)
\]

\[
+ \frac{1}{\sqrt{nT}} \tilde{G}_{nT} \tilde{X}_{nT,y} \tilde{\xi}_{nT} - \frac{1}{\sqrt{nT}} \tilde{e}_{nT} (P_{y} \otimes I_n) \tilde{G}_{nT} \tilde{\xi}_{nT} - \frac{1}{\sqrt{nT}} \tilde{X}_{nT,y} (I_T \otimes P_{y}) \tilde{G}_{nT} \tilde{\xi}_{nT} + o_p(1)
\]

\[
= \frac{1}{\sqrt{nT}} \left( b_{3nT}^{(1)} \tilde{\xi}_{nT} + b_{3nT}^{(1)} B_{1nT} \tilde{\xi}_{nT} - \text{tr}(B_{1nT}) \right) - \frac{1}{\sqrt{nT}} \text{tr} \left( (P_{y} \otimes I_n + I_T \otimes P_{y}) \tilde{G}_{nT} \right) + o_p(1),
\]

where $\tilde{G}_{nT} = \text{diag} (G_{n1}, \cdots , G_{nT})$ is an $nT \times nT$ matrix, $\tilde{X}_{nT,y} = \left( \tilde{X}_{nT,y}^1, \cdots , \tilde{X}_{nT,y}^j \right)'$ is $nT \times k_y$, $\tilde{F}_{Ty} = \tilde{F}_{Ty}^1, \cdots , \tilde{F}_{Ty}^j$.
\[
\left( \tilde{f}_{y1}, \cdots, \tilde{f}_{yT} \right)^T \text{ is } TR_{y0} \times 1, \quad \tilde{e}_{nT} = \begin{pmatrix} e_{111} & e_{112} & \cdots & e_{11p} \\ \vdots & \vdots & \ddots & \vdots \\ e_{n11} & e_{n12} & \cdots & e_{n1p} \\ e_{nT1} & e_{nT2} & \cdots & e_{nTp} \end{pmatrix} \text{ is } nT \times p,
\]

\[
b_{3nT} = \sigma_{\xi0}^{-1} \left( M_{F_T} \otimes M_{\Gamma_{y_0}} \right) \tilde{G}_{nT} \left( \tilde{X}_{nT} \beta_{y_0} + (I_T \otimes \tilde{\Gamma}_{y_0}) \tilde{F}_{y_T} + \tilde{\xi}_{nT} \Sigma_{Y0}^{-1} \delta_0 \right), \text{ and}
\]

\[
B_{1nT} = \frac{1}{2} \left( I_{nT} - P_{F_T} \otimes I_n - I_T \otimes P_{\Gamma_{y_0}} \right) \tilde{G}_{nT} + \frac{1}{2} \tilde{G}_{nT}' \left( I_{nT} - P_{F_T} \otimes I_n - I_T \otimes P_{\Gamma_{y_0}} \right). 
\]

- \(\alpha_5\): 
  \[
  \tilde{c}_{nT,k}^{(1)} = - \frac{\sigma_{\xi0} \tilde{\xi}'_{nT} \left( M_{F_T} \otimes M_{\Gamma_{y_0}} \right) \tilde{\xi}_{nT} + \sqrt{nT} \sigma_{\xi0} + O(1)}{nT^{1/2}} = \frac{1}{nT} \left( \tilde{\xi}'_{nT} B_{2nT} \tilde{\xi}_{nT} - \text{tr}(B_{2nT}) \right) + \left( \sqrt{nT} + \sqrt{\frac{T}{n}} \right) R_{00} \sigma_{\xi0} + O(1),
  \]
  where \(B_{2nT} = - \sigma_{\xi0} M_{F_T} \otimes M_{\Gamma_{y_0}}\).

- \(\alpha\): for \(k = 1, \cdots, J\), by using the relation \(\text{tr}(B'_1 B_2 B_3 B_4) = \text{vec}(B_1)' (B_4' \otimes B_2) \text{ vec}(B_3)\) for any conformable matrices \(B_1, \cdots, B_4\),
  \[
  \tilde{c}_{nT,k}^{(1)} = \frac{1}{\sqrt{nT} \sigma_{\xi0}^{2}} \text{tr} \left[ M_{\Gamma_{y_0}} \tilde{X}_{nT} M_{F_T} \left( \tilde{\Gamma}_{nT}' \tilde{F}_{Tz} + E_{nT} \right)' (e_k \otimes I_n) \right] + O(1)
  \]
  \[
  = \frac{1}{\sqrt{nT} \sigma_{\xi0}^{2}} \text{tr} \left[ \tilde{b}_{4nT}^{(k)'} \tilde{\xi}_{nT} + \tilde{e}'_{nT} D_{nT}^{(k)} \tilde{\xi}_{nT} \right] + o(1),
  \]
  where \(e_k\) be a \(p \times 1\) vector with one in its \(k\)-th entry and zeros in its other entries, 
  \(b_{4nT}^{(k)} = \sigma_{\xi0}^{-1} \text{vec} \left( M_{\Gamma_{y_0}} (e_k' \otimes I_n) \tilde{\Gamma}_{nT}' \tilde{F}_{Tz} M_{F_T} \right)\) and 
  \(D_{nT}^{(k)} = \sigma_{\xi0}^{-1} M_{F_T} \otimes \left( \Sigma_{Y0}^{-1} e_k' \otimes I_n \right) M_{\Gamma_{y_0}}\).
(2) Variance of $\tilde{c}^{(1)}_{nT} - \varphi_{nT}$

$\varphi_{nT}$ is defined in Proposition 2 and $\mathbb{E} \left( \tilde{c}^{(1)}_{nT} - \varphi_{nT} \right) = 0$. The variance of $\tilde{c}^{(1)}_{nT} - \varphi_{nT}$ can be straightforwardly shown to be $\Sigma_{nT} + \Pi_{nT}$ with terms given below:

$$
\Sigma_{nT} = 
\begin{pmatrix}
\Sigma_{\beta,\beta}^{nT} & \Sigma_{\beta,\lambda}^{nT} & \Sigma_{\beta,\alpha}^{nT} & 0 \\
\Sigma_{\lambda,\beta}^{nT} & \Sigma_{\lambda,\lambda}^{nT} & \Sigma_{\lambda,\alpha}^{nT} & 0 \\
\Sigma_{\alpha,\beta}^{nT} & \Sigma_{\alpha,\lambda}^{nT} & \Sigma_{\alpha,\alpha}^{nT} & 0 \\
0 & 0 & 0 & 0
\end{pmatrix},
\quad
\Pi_{nT} = 
\begin{pmatrix}
0 & 0 & \Pi_{\beta,\lambda}^{nT} & \Pi_{\beta,\alpha}^{nT} \\
0 & 0 & \Pi_{\lambda,\lambda}^{nT} & \Pi_{\lambda,\alpha}^{nT} \\
0 & 0 & \Pi_{\alpha,\alpha}^{nT} & 0 \\
0 & 0 & \Pi_{\alpha,\alpha}^{nT} & 0
\end{pmatrix},
$$

(A.30)

with

$$
\begin{align*}
\left[ \Sigma_{\beta,\beta}^{nT} \right]_{k_1k_2} &= \frac{1}{nT} \text{tr} \left( M_{F_T} \tilde{X}'_{nT,z,k_1} \left( \Sigma_{\epsilon \theta} \otimes I_n \right) M_{T_{\alpha}} \left( \Sigma_{\epsilon \theta} \otimes I_n \right) \tilde{X}_{nT,z,k_2} \right) \\
&\quad + \frac{1}{nT} \sigma_{\epsilon \theta} \text{tr} \left( M_{F_T} \tilde{X}'_{nT,z,k_1} \left( \delta_0 \otimes I_n \right) M_{T_{\alpha}} \left( \delta_0 \otimes I_n \right) \tilde{X}_{nT,z,k_2} \right), \\
\left[ \Sigma_{\beta,\lambda}^{nT} \right]_{k_1k_2} &= - \frac{1}{nT} \sigma_{\epsilon \theta} \text{tr} \left( M_{F_T} \tilde{X}'_{nT,z,k_1} \left( \delta_0 \otimes I_n \right) M_{T_{\alpha}} \tilde{X}_{nT,y,k_2} \right), \\
\left[ \Sigma_{\lambda,\beta}^{nT} \right]_{k} &= \frac{1}{nT} \mathbb{E} \left( b_{1nT}' b_{3nT} \right) \\
\left[ \Sigma_{\lambda,\lambda}^{nT} \right]_{k_1k_2} &= \frac{1}{nT} \mathbb{E} \left( b_{1nT}' r_{3nT} \right) \\
\left[ \Sigma_{\alpha,\beta}^{nT} \right]_{k_1k_2} &= \frac{1}{nT} \text{tr} \left( M_{F_T} \tilde{X}'_{nT,y,k_1} M_{T_{\alpha}} \tilde{X}_{nT,y,k_2} \right), \\
\left[ \Sigma_{\alpha,\alpha}^{nT} \right]_{k} &= \frac{1}{nT} \mathbb{E} \left( b_{2nT}' b_{3nT} \right) \\
\left[ \Sigma_{\alpha,\alpha}^{nT} \right]_{k_1k_2} &= \frac{1}{nT} \mathbb{E} \left( b_{2nT}' r_{3nT} \right) \\
\left[ \Sigma_{\beta,\beta}^{nT} \right]_{k} &= \frac{1}{nT} \text{tr} \left( M_{F_T} \tilde{X}'_{nT,y,k_1} M_{T_{\alpha}} \left( \epsilon_0 \otimes I_n \right) \tilde{X}_{nT,y,k_2} \right) \\
\left[ \Sigma_{\beta,\beta}^{nT} \right]_{k} &= \frac{1}{nT} \text{tr} \left( M_{F_T} \tilde{X}'_{nT,y,k_1} M_{T_{\alpha}} \left( \epsilon_0 \otimes I_n \right) \tilde{X}_{nT,y,k_2} \right) \\
\Sigma_{\lambda,\lambda}^{nT} &= \frac{1}{nT} \text{tr} \left( B_{1nT} B_{2nT} \right) \\
\Sigma_{\alpha,\alpha}^{nT} &= \frac{1}{nT} \text{tr} \left( A_{nT}^T A_{nT} \right) \\
\Sigma_{\alpha,\alpha}^{nT} &= \frac{1}{nT} \text{tr} \left( B_{nT} A_{nT} \right)
\end{align*}
$$
\[
= \frac{1}{nT \sigma_{\xi_0}^2} \text{tr} \left( M_{\mathcal{F}_T} F_{\mathcal{F}_T} \tilde{\Gamma}_{nz} (e_k \otimes I_n) M_{\mathcal{F}_T} \left( e_k \otimes I_n \right) \tilde{\Gamma}_{nz} F_{\mathcal{F}_T} M_{\mathcal{F}_T} \right) + \sigma_{\xi_0}^2 \text{tr} \left( e_k^\prime \Sigma \varepsilon e_k \right) + o(1),
\]
and

\[
\begin{align*}
\left[ \Pi_{\beta \lambda}^{nT} \right]_k &= \frac{1}{nT} \mathbb{E} \left( \sum_{i=1}^{nT} b_{1nT,ij}^{(k)} B_{1nT,ij} \tilde{\varepsilon}_{nT,i}^3 \right) \\
\left[ \Pi_{\beta \alpha}^{nT} \right]_k &= \frac{1}{nT} \mathbb{E} \left( \sum_{i=1}^{nT} b_{2nT,ij}^{(k)} B_{2nT,ij} \tilde{\varepsilon}_{nT,i}^3 \right) \\
\left[ \Pi_{\alpha \delta}^{nT} \right]_k &= \frac{1}{nT} \mathbb{E} \left( \sum_{i=1}^{nT} b_{3nT,ij}^{(k)} B_{3nT,ij} \tilde{\varepsilon}_{nT,i}^3 \right)
\end{align*}
\]

\[
\Pi_{\alpha \delta}^{nT} = \frac{1}{nT} \mathbb{E} \left( \sum_{i=1}^{nT} b_{2nT,ij}^{(k)} B_{2nT,ij} \tilde{\varepsilon}_{nT,i}^3 \right)
\]

(3) Elements of \(\tilde{\mathcal{C}}_{nT}\) (the Approximate Hessian Matrix)

\[
\tilde{\mathcal{C}}_{nT} = \begin{pmatrix}
\tilde{c}_{\beta \beta}^{nT} & \tilde{c}_{\beta \beta}^{nT} & \tilde{c}_{\beta \lambda}^{nT} & \tilde{c}_{\beta \alpha}^{nT} & \tilde{c}_{\beta \alpha}^{nT} & \tilde{c}_{\beta \delta}^{nT} \\
\tilde{c}_{\beta \beta}^{nT} & \tilde{c}_{\beta \beta}^{nT} & \tilde{c}_{\beta \lambda}^{nT} & 0 & \tilde{c}_{\beta \alpha}^{nT} & \tilde{c}_{\beta \delta}^{nT} \\
\tilde{c}_{\beta \lambda}^{nT} & \tilde{c}_{\beta \lambda}^{nT} & \tilde{c}_{\lambda \lambda}^{nT} & 0 & \tilde{c}_{\lambda \alpha}^{nT} & \tilde{c}_{\lambda \delta}^{nT} \\
\tilde{c}_{\beta \alpha}^{nT} & \tilde{c}_{\beta \alpha}^{nT} & \tilde{c}_{\lambda \alpha}^{nT} & 0 & \tilde{c}_{\alpha \alpha}^{nT} & 0 \\
\tilde{c}_{\beta \delta}^{nT} & \tilde{c}_{\beta \delta}^{nT} & \tilde{c}_{\lambda \delta}^{nT} & 0 & \tilde{c}_{\alpha \alpha}^{nT} & 0 \\
\tilde{c}_{\delta \delta}^{nT} & \tilde{c}_{\delta \delta}^{nT} & \tilde{c}_{\alpha \alpha}^{nT} & 0 & \tilde{c}_{\alpha \alpha}^{nT} & 0
\end{pmatrix}
\]

\[
\left[ \tilde{c}_{\beta \beta}^{nT} \right]_{k_1 k_2} = \frac{1}{nT} \text{tr} \left( \tilde{M}_{\mathcal{F}_T} \tilde{X}_{nz,k_1} \left( \Sigma_{e_0}^{-\frac{1}{2}} \otimes I_n \right) M_{\mathcal{F}_T} \left( \Sigma_{e_0}^{-\frac{1}{2}} \otimes I_n \right) \tilde{X}_{nz,k_2} \right)
\]
\[
\begin{align*}
\mathcal{C}_{\mathcal{H}_{\mathcal{H}}^T}^{\beta,\beta}_{k_1k_2} &= -\frac{1}{nT\sigma_{\xi_0}^2} \text{tr} \left( M_{\mathcal{F}_{\mathcal{Y}}} \tilde{X}_{\mathcal{H}_{\mathcal{H}},k_1} \left( \delta_0 \otimes I_n \right) M_{\mathcal{F}_{\mathcal{Y}}} \left( \delta_0 \otimes I_n \right) \tilde{X}_{\mathcal{H}_{\mathcal{H}},k_2} \right) \\
\mathcal{C}_{\mathcal{H}_{\mathcal{H}}^T}^{\beta,\lambda}_{\kappa} &= \frac{1}{nT} \frac{b_{\kappa}^{(k)}}{b_{\kappa}^{(k)}} - \frac{1}{nT\sigma_{\xi_0}^2} \text{vec} \left( M_{\mathcal{F}_{\mathcal{Y}}} \left( \delta_0 \otimes I_n \right) \tilde{X}_{\mathcal{H}_{\mathcal{H}},\kappa} \right) \left( M_{\mathcal{F}_{\mathcal{Y}}} \otimes M_{\mathcal{F}_{\mathcal{Y}}} \right) \tilde{G}_{\mathcal{H}_{\mathcal{H}}^T} \tilde{\xi}_{\mathcal{H}_{\mathcal{H}}^T} \\
\mathcal{C}_{\mathcal{H}_{\mathcal{H}}^T}^{\beta,\alpha}_{\kappa} &= \frac{1}{nT\sigma_{\xi_0}^2} \text{tr} \left( M_{\mathcal{F}_{\mathcal{Y}}} \tilde{X}_{\mathcal{H}_{\mathcal{H}},\kappa} \left( \delta_0 \otimes I_n \right) M_{\mathcal{F}_{\mathcal{Y}}} \tilde{X}_{\mathcal{H}_{\mathcal{H}},\kappa} \right) = o_p(1) \\
\mathcal{C}_{\mathcal{H}_{\mathcal{H}}^T}^{\beta,\delta}_{\kappa} &= \frac{1}{nT\sigma_{\xi_0}^2} \text{tr} \left( M_{\mathcal{F}_{\mathcal{Y}}} \tilde{X}_{\mathcal{H}_{\mathcal{H}},\kappa} \left( \delta_0 \otimes I_n \right) M_{\mathcal{F}_{\mathcal{Y}}} \left( \delta_0 \otimes I_n \right) \tilde{X}_{\mathcal{H}_{\mathcal{H}},\kappa} \right) = o_p(1) \\
\mathcal{C}_{\mathcal{H}_{\mathcal{H}}^T}^{\beta,\beta}_{k_1k_2} &= \frac{1}{nT\sigma_{\xi_0}^2} \text{tr} \left( M_{\mathcal{F}_{\mathcal{Y}}} \tilde{X}_{\mathcal{H}_{\mathcal{H}},k_1} \left( \delta_0 \otimes I_n \right) M_{\mathcal{F}_{\mathcal{Y}}} \tilde{X}_{\mathcal{H}_{\mathcal{H}},k_2} \right) + \frac{1}{nT} \text{tr} \left( \tilde{G}_{\mathcal{H}_{\mathcal{H}}^T} \tilde{\xi}_{\mathcal{H}_{\mathcal{H}}^T} \right) \\
\mathcal{C}_{\mathcal{H}_{\mathcal{H}}^T}^{\beta,\alpha}_{\kappa} &= \frac{1}{nT\sigma_{\xi_0}^2} \text{tr} \left( M_{\mathcal{F}_{\mathcal{Y}}} \tilde{X}_{\mathcal{H}_{\mathcal{H}},\kappa} \left( \delta_0 \otimes I_n \right) M_{\mathcal{F}_{\mathcal{Y}}} \left( \delta_0 \otimes I_n \right) \tilde{X}_{\mathcal{H}_{\mathcal{H}},\kappa} \right) = o_p(1) \\
\mathcal{C}_{\mathcal{H}_{\mathcal{H}}^T}^{\beta,\delta}_{\kappa} &= \frac{1}{nT\sigma_{\xi_0}^2} \text{tr} \left( M_{\mathcal{F}_{\mathcal{Y}}} \tilde{X}_{\mathcal{H}_{\mathcal{H}},\kappa} \left( \delta_0 \otimes I_n \right) M_{\mathcal{F}_{\mathcal{Y}}} \left( \delta_0 \otimes I_n \right) \tilde{X}_{\mathcal{H}_{\mathcal{H}},\kappa} \right) = o_p(1) \\
\mathcal{C}_{\mathcal{H}_{\mathcal{H}}^T}^{\alpha,\beta}_{\kappa} &= \frac{1}{nT\sigma_{\xi_0}^2} \text{tr} \left( M_{\mathcal{F}_{\mathcal{Y}}} \tilde{X}_{\mathcal{H}_{\mathcal{H}},\kappa} \left( \delta_0 \otimes I_n \right) M_{\mathcal{F}_{\mathcal{Y}}} \left( \delta_0 \otimes I_n \right) \tilde{X}_{\mathcal{H}_{\mathcal{H}},\kappa} \right) = o_p(1) \\
\mathcal{C}_{\mathcal{H}_{\mathcal{H}}^T}^{\delta,\delta}_{\kappa} &= \frac{1}{nT\sigma_{\xi_0}^2} \text{tr} \left( M_{\mathcal{F}_{\mathcal{Y}}} \tilde{X}_{\mathcal{H}_{\mathcal{H}},\kappa} \left( \delta_0 \otimes I_n \right) M_{\mathcal{F}_{\mathcal{Y}}} \left( \delta_0 \otimes I_n \right) \tilde{X}_{\mathcal{H}_{\mathcal{H}},\kappa} \right) = o_p(1) \\
\mathcal{C}_{\mathcal{H}_{\mathcal{H}}^T}^{\lambda}_{\kappa} &= \frac{1}{nT\sigma_{\xi_0}^2} \text{tr} \left( M_{\mathcal{F}_{\mathcal{Y}}} \tilde{X}_{\mathcal{H}_{\mathcal{H}},\kappa} \left( \delta_0 \otimes I_n \right) M_{\mathcal{F}_{\mathcal{Y}}} \left( \delta_0 \otimes I_n \right) \tilde{X}_{\mathcal{H}_{\mathcal{H}},\kappa} \right) + \frac{1}{nT} \text{tr} \left( \tilde{G}_{\mathcal{H}_{\mathcal{H}}^T} \tilde{\xi}_{\mathcal{H}_{\mathcal{H}}^T} \right) \\
\mathcal{C}_{\mathcal{H}_{\mathcal{H}}^T}^{\alpha,\alpha}_{\kappa} &= \frac{1}{nT\sigma_{\xi_0}^2} \text{tr} \left( M_{\mathcal{F}_{\mathcal{Y}}} \tilde{X}_{\mathcal{H}_{\mathcal{H}},\kappa} \left( \delta_0 \otimes I_n \right) M_{\mathcal{F}_{\mathcal{Y}}} \left( \delta_0 \otimes I_n \right) \tilde{X}_{\mathcal{H}_{\mathcal{H}},\kappa} \right) + \frac{1}{nT} \text{tr} \left( \tilde{G}_{\mathcal{H}_{\mathcal{H}}^T} \tilde{\xi}_{\mathcal{H}_{\mathcal{H}}^T} \right) \\
\mathcal{C}_{\mathcal{H}_{\mathcal{H}}^T}^{\alpha,\delta}_{\kappa} &= \frac{1}{nT\sigma_{\xi_0}^2} \text{tr} \left( M_{\mathcal{F}_{\mathcal{Y}}} \tilde{X}_{\mathcal{H}_{\mathcal{H}},\kappa} \left( \delta_0 \otimes I_n \right) M_{\mathcal{F}_{\mathcal{Y}}} \left( \delta_0 \otimes I_n \right) \tilde{X}_{\mathcal{H}_{\mathcal{H}},\kappa} \right) + \frac{1}{nT} \text{tr} \left( \tilde{G}_{\mathcal{H}_{\mathcal{H}}^T} \tilde{\xi}_{\mathcal{H}_{\mathcal{H}}^T} \right)
\end{align*}
\]
\[
\frac{\sigma^2_{\xi_0}}{nT} \mathbb{E}_{nT} \left( M_{F_T} \otimes M_{F_T} \right) \hat{\xi}_{nT} + \sigma^2_{\hat{\xi}_0} = 2\sigma^2_{\xi_0} + o_p(1)
\]

A.5. Proof of Proposition [3]

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 \( \| \hat{\theta} - \theta_0 \|_2 = o_p \left( n^{-\frac{1}{2}} \right) \), we have the following results.

1. \( \left\| \left( \Sigma^{\frac{1}{2}}_{\xi_0} \otimes I_n \right) E_{nT} + \hat{E}_{nT} \left( \hat{\theta} \right) \right\|_2 = O_p(\sqrt{n}) \), because \( \| E_{nT} \|_2 = O_p(\sqrt{n}) \) by Assumption 2 and \( \| \hat{E}_{nT} \left( \hat{\theta} \right) \|_2 \leq \sum_{k=1}^{k+J} \| \hat{\eta}_{k_1} \|_2 + \sum_{k_1, k_2 = 1}^{k+J} \| \hat{\eta}_{k_1} \|_2 \| \hat{V}_{k_2} \|_2 = o_p(\sqrt{n}) \).

2. \( \mu_{\text{up}} \left( \frac{1}{n} \left( \left( \Sigma^{\frac{1}{2}}_{\xi_0} \otimes I_n \right) E_{nT} + \hat{E}_{nT} \left( \hat{\theta} \right) \right) \left( \left( \Sigma^{\frac{1}{2}}_{\xi_0} \otimes I_n \right) E_{nT} + \hat{E}_{nT} \left( \hat{\theta} \right) \right)' \right) \geq c + o_p(1) \) for some positive constant \( c \). This is because

\[
\mu_{\text{up}} \left( \frac{1}{n} \left( \left( \Sigma^{\frac{1}{2}}_{\xi_0} \otimes I_n \right) E_{nT} \right) \left( \left( \Sigma^{\frac{1}{2}}_{\xi_0} \otimes I_n \right) E_{nT} \right)' + \mu_{\text{up}} \left( \frac{1}{n} \hat{E}_{nT} \left( \hat{\theta} \right) \hat{E}_{nT} \left( \hat{\theta} \right)' \right) \\
\geq \mu_{\text{up}} \left( \frac{1}{n} \left( \left( \Sigma^{\frac{1}{2}}_{\xi_0} \otimes I_n \right) E_{nT} \hat{E}_{nT} \left( \hat{\theta} \right)' + \frac{1}{n} \hat{E}_{nT} \left( \hat{\theta} \right) E_{nT} \left( \Sigma^{\frac{1}{2}}_{\xi_0} \otimes I_n \right) \right) \right) + \mu_{\text{up}} \left( \frac{1}{n} \left( \left( \Sigma^{\frac{1}{2}}_{\xi_0} \otimes I_n \right) E_{nT} \hat{E}_{nT} \left( \hat{\theta} \right)' \right) \right) - \frac{1}{n} \left\| \hat{E}_{nT} \left( \hat{\theta} \right) \hat{E}_{nT} \left( \hat{\theta} \right)' \right\|_2 \] (A.31)
\[-\frac{1}{n} \left\| \left( \Sigma_{\varepsilon_0}^{-\frac{1}{2}} \otimes I_n \right) E_{nT} \tilde{\Theta}_{nT} (\tilde{\Theta})' + \tilde{E}_{nT} (\tilde{\Theta}) E_{nT}' \left( \Sigma_{\varepsilon_0}^{-\frac{1}{2}} \otimes I_n \right) \right\|_2 \]
\[
= n_{op} \left( \frac{1}{n} \left( \Sigma_{\varepsilon_0}^{-1} \otimes I_n \right) E_{nT} E_{nT}' \right) + o_P(1) \geq c + o_P(1),
\]

where the inequality in Eq. (A.31) can be found in Theorem 8.12 of [Zhang (2011)], and the last inequality is from Lemma A.1 of [Ahn and Horenstein (2013)] (due to [Bai and Yin (1993)]).

3. For any \( r \times q \) matrix \( A = (a_1, \cdots, a_q) \) such that \( A'A = T I_q \),
\[
\frac{1}{n^3} \left| \text{tr} \left( A' F_T \Gamma_{n} (\left( \Sigma_{\varepsilon_0}^{-\frac{1}{2}} \otimes I_n \right) E_{nT} + \tilde{E}_{nT} (\tilde{\Theta}) \right) A \right) \leq R_0 \frac{n^2}{n^3} \frac{\| AA' \|_2 \| F_T \|_2 \| \left( \Sigma_{\varepsilon_0}^{-\frac{1}{2}} \otimes I_n \right) E_{nT} + \tilde{E}_{nT} (\tilde{\Theta}) \|_2}{\| \left( \Sigma_{\varepsilon_0}^{-\frac{1}{2}} \otimes I_n \right) E_{nT} + \tilde{E}_{nT} (\tilde{\Theta}) \|_2} = O_P \left( n^{-\frac{1}{2}} \right).
\]

Similarly, we have
\[
\frac{1}{n^3} \left| \text{tr} \left( A' \left( \Sigma_{\varepsilon_0}^{-\frac{1}{2}} \otimes I_n \right) E_{nT} + \tilde{E}_{nT} (\tilde{\Theta}) \right) \Gamma_{n} (\Gamma_{n}^{-} \Gamma_{n}^{-})^{-1} \Gamma_{n} (\left( \Sigma_{\varepsilon_0}^{-\frac{1}{2}} \otimes I_n \right) E_{nT} + \tilde{E}_{nT} (\tilde{\Theta}) A \right) \leq R_0 \frac{n^2}{n^3} \frac{\| AA' \|_2 \left( \Sigma_{\varepsilon_0}^{-\frac{1}{2}} \otimes I_n \right) E_{nT} + \tilde{E}_{nT} (\tilde{\Theta}) \|_2^2 \| \Gamma_{n} (\Gamma_{n}^{-} \Gamma_{n}^{-})^{-1} \Gamma_{n} \|_2}{\| \left( \Sigma_{\varepsilon_0}^{-\frac{1}{2}} \otimes I_n \right) E_{nT} + \tilde{E}_{nT} (\tilde{\Theta}) \|_2} = O_P \left( n^{-1} \right).
\]

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 \( \Xi \) and \( \Xi \) above. Their Assumption B is used to prove Lemma A.10, which is satisfied by \( \Xi \) and their Assumption A is satisfied by our Assumption \( \Xi \). It is straightforward to check that \( \tilde{G}_{nT} \) also satisfies the relevant conditions and similar methods can be applied to determine the number of factors in the \( y \) equation. In summary, [Ahn and Horenstein (2013)]’s result applies here.

**A6. Additional Monte Carlo Results**

**Estimation with Redundant Factors**

Tables 9 and 10 provide the Monte Carlo results where the number of factors in the outcome equation is over-specified by 2 and 3.
Table 9: Performance of the Bias-Corrected QML Estimator $\hat{\theta}_{nT}$ when $R_y = R_x + 2$.

<table>
<thead>
<tr>
<th>n</th>
<th>$T$</th>
<th>$\lambda_0$</th>
<th>$\rho_0$</th>
<th>$\hat{\beta}_0$</th>
<th>$\hat{\beta}_1$</th>
<th>$\hat{\lambda}$</th>
<th>$\hat{\alpha}_0$</th>
<th>$\hat{\alpha}_1$</th>
<th>$\hat{\delta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>25</td>
<td>0.2</td>
<td>0.2</td>
<td>0.00252</td>
<td>-0.00255</td>
<td>-0.00276</td>
<td>0.07660</td>
<td>0.00197</td>
<td>0.00071</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.952</td>
<td>0.879</td>
<td>0.858</td>
<td>0.228</td>
<td>0.945</td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>0.2</td>
<td>0.2</td>
<td>0.00013</td>
<td>0.00165</td>
<td>-0.00099</td>
<td>0.05266</td>
<td>0.00086</td>
<td>0.00175</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.945</td>
<td>0.909</td>
<td>0.902</td>
<td>0.226</td>
<td>0.943</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.2</td>
<td>0.2</td>
<td>-0.00115</td>
<td>-0.00006</td>
<td>-0.00106</td>
<td>0.05187</td>
<td>0.00135</td>
<td>0.00151</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.943</td>
<td>0.924</td>
<td>0.891</td>
<td>0.242</td>
<td>0.940</td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>0.2</td>
<td>0.2</td>
<td>0.00028</td>
<td>0.00171</td>
<td>-0.00078</td>
<td>0.03550</td>
<td>0.00063</td>
<td>0.00040</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.948</td>
<td>0.932</td>
<td>0.916</td>
<td>0.242</td>
<td>0.945</td>
</tr>
<tr>
<td>(2)</td>
<td>25</td>
<td>0.2</td>
<td>0.6</td>
<td>0.00108</td>
<td>-0.00353</td>
<td>-0.00086</td>
<td>0.09293</td>
<td>0.00189</td>
<td>-0.00009</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.949</td>
<td>0.898</td>
<td>0.867</td>
<td>0.224</td>
<td>0.950</td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>0.2</td>
<td>0.6</td>
<td>0.00020</td>
<td>0.00097</td>
<td>-0.00077</td>
<td>0.06438</td>
<td>0.00047</td>
<td>0.00032</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.948</td>
<td>0.923</td>
<td>0.909</td>
<td>0.224</td>
<td>0.949</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.2</td>
<td>0.6</td>
<td>-0.00119</td>
<td>-0.00040</td>
<td>-0.00028</td>
<td>0.06389</td>
<td>0.00073</td>
<td>0.00068</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.947</td>
<td>0.931</td>
<td>0.896</td>
<td>0.227</td>
<td>0.938</td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>0.2</td>
<td>0.6</td>
<td>0.00047</td>
<td>0.00144</td>
<td>-0.00110</td>
<td>0.04330</td>
<td>0.00051</td>
<td>0.00092</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.942</td>
<td>0.924</td>
<td>0.912</td>
<td>0.249</td>
<td>0.950</td>
</tr>
<tr>
<td>(3)</td>
<td>25</td>
<td>0.6</td>
<td>0.2</td>
<td>0.00252</td>
<td>-0.00148</td>
<td>-0.00313</td>
<td>0.07603</td>
<td>0.00197</td>
<td>0.00166</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.952</td>
<td>0.878</td>
<td>0.865</td>
<td>0.244</td>
<td>0.945</td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>0.6</td>
<td>0.2</td>
<td>0.00013</td>
<td>0.00203</td>
<td>-0.00128</td>
<td>0.05237</td>
<td>0.00086</td>
<td>0.00186</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.945</td>
<td>0.906</td>
<td>0.903</td>
<td>0.245</td>
<td>0.943</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.6</td>
<td>0.2</td>
<td>-0.00115</td>
<td>0.00051</td>
<td>-0.00157</td>
<td>0.05158</td>
<td>0.00135</td>
<td>0.00228</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.943</td>
<td>0.927</td>
<td>0.901</td>
<td>0.277</td>
<td>0.940</td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>0.6</td>
<td>0.2</td>
<td>0.00028</td>
<td>0.00200</td>
<td>-0.00091</td>
<td>0.03533</td>
<td>0.00063</td>
<td>0.00046</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.948</td>
<td>0.931</td>
<td>0.918</td>
<td>0.272</td>
<td>0.945</td>
</tr>
<tr>
<td>(4)</td>
<td>25</td>
<td>0.6</td>
<td>0.6</td>
<td>0.00108</td>
<td>-0.00268</td>
<td>-0.00195</td>
<td>0.09242</td>
<td>0.00189</td>
<td>0.00134</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.948</td>
<td>0.901</td>
<td>0.866</td>
<td>0.244</td>
<td>0.950</td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>0.6</td>
<td>0.6</td>
<td>0.00020</td>
<td>0.00127</td>
<td>-0.00096</td>
<td>0.06413</td>
<td>0.00047</td>
<td>0.00055</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.948</td>
<td>0.927</td>
<td>0.915</td>
<td>0.239</td>
<td>0.949</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.6</td>
<td>0.6</td>
<td>-0.00119</td>
<td>-0.00001</td>
<td>-0.00078</td>
<td>0.06368</td>
<td>0.00073</td>
<td>0.00171</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.948</td>
<td>0.936</td>
<td>0.901</td>
<td>0.252</td>
<td>0.938</td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>0.6</td>
<td>0.6</td>
<td>0.00047</td>
<td>0.00170</td>
<td>-0.00100</td>
<td>0.04309</td>
<td>0.00051</td>
<td>0.00106</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td>0.942</td>
<td>0.927</td>
<td>0.914</td>
<td>0.266</td>
<td>0.950</td>
</tr>
</tbody>
</table>

The true number of factor is 1 in the $z$ equation and 2 in the $y$ equation. In the estimation, the number of factors is set at 1 for the $z$ equation and 4 for the $y$ equation. True parameter values: $\beta_0 = 1$, $\beta_1 = 1$, $\sigma_0 = \sqrt{\frac{1}{2}}$, $\sigma_0 = \sqrt{\frac{1}{2}}$ (hence $\alpha_0 = \left(\frac{1}{4}\right)^{-\frac{1}{2}}$). In the low endogeneity case, $\rho_0 = 0.2$, $\alpha_0 = 0.88$ and $\delta_0 = 0.2$. In the high endogeneity case, $\rho_0 = 0.6$, $\alpha_0 = 1.08$ and $\delta_0 = 0.6$. $\hat{\theta}_{nT}$ is the bias-corrected QML estimator. The coverage probabilities (CP) are calculated using the theoretical standard deviations obtained from the diagonal elements of $\frac{1}{nT} \sum_{t=1}^{T} \hat{\Sigma}_{nT}$. 

51
Table 10: Performance of the Bias-Corrected QML Estimator $\hat{\theta}_{nT}$ when $R_z = R_{y0} + 3$.

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>T</th>
<th>$\lambda_0$</th>
<th>$\rho_0$</th>
<th>$\hat{\beta}_z$</th>
<th>$\hat{\beta}_y$</th>
<th>$\hat{\lambda}$</th>
<th>$\hat{\alpha}_z$</th>
<th>$\hat{\alpha}$</th>
<th>$\hat{\delta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25</td>
<td>25</td>
<td>0.2</td>
<td>0.2</td>
<td>0.00252</td>
<td>-0.00269</td>
<td>-0.00357</td>
<td>0.11187</td>
<td>0.00197</td>
<td>0.00152</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>25</td>
<td>0.2</td>
<td>0.2</td>
<td>0.00013</td>
<td>0.00168</td>
<td>-0.00179</td>
<td>0.07701</td>
<td>0.00086</td>
<td>0.00143</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>25</td>
<td>49</td>
<td>0.2</td>
<td>0.2</td>
<td>-0.00494</td>
<td>-0.00049</td>
<td>-0.00181</td>
<td>0.07649</td>
<td>0.00135</td>
<td>0.00150</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>49</td>
<td>0.2</td>
<td>0.2</td>
<td>0.00028</td>
<td>0.00196</td>
<td>-0.00107</td>
<td>0.05199</td>
<td>0.00063</td>
<td>0.00048</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>25</td>
<td>0.2</td>
<td>0.6</td>
<td>0.00108</td>
<td>-0.00473</td>
<td>-0.00100</td>
<td>0.13613</td>
<td>0.00189</td>
<td>-0.00024</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>25</td>
<td>0.2</td>
<td>0.6</td>
<td>0.00020</td>
<td>0.00119</td>
<td>-0.00093</td>
<td>0.09410</td>
<td>0.00047</td>
<td>0.00005</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>25</td>
<td>49</td>
<td>0.2</td>
<td>0.6</td>
<td>-0.00119</td>
<td>-0.00035</td>
<td>-0.00034</td>
<td>0.09390</td>
<td>0.00073</td>
<td>0.00050</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>49</td>
<td>0.2</td>
<td>0.6</td>
<td>0.00047</td>
<td>0.00151</td>
<td>-0.00121</td>
<td>0.06342</td>
<td>0.00051</td>
<td>0.00058</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>25</td>
<td>25</td>
<td>0.6</td>
<td>0.2</td>
<td>0.00252</td>
<td>-0.00120</td>
<td>-0.00463</td>
<td>0.11097</td>
<td>0.00197</td>
<td>0.00268</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>25</td>
<td>0.6</td>
<td>0.2</td>
<td>0.00013</td>
<td>0.00229</td>
<td>-0.00204</td>
<td>0.07658</td>
<td>0.00086</td>
<td>0.00156</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>49</td>
<td>0.6</td>
<td>0.2</td>
<td>-0.00115</td>
<td>0.00049</td>
<td>-0.00286</td>
<td>0.07594</td>
<td>0.00135</td>
<td>0.00238</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>49</td>
<td>0.6</td>
<td>0.2</td>
<td>0.00028</td>
<td>0.00227</td>
<td>-0.00113</td>
<td>0.05178</td>
<td>0.00063</td>
<td>0.00056</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>25</td>
<td>25</td>
<td>0.6</td>
<td>0.6</td>
<td>0.00108</td>
<td>-0.00365</td>
<td>-0.00247</td>
<td>0.13543</td>
<td>0.00189</td>
<td>0.00134</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>25</td>
<td>0.6</td>
<td>0.6</td>
<td>0.00020</td>
<td>0.00151</td>
<td>-0.00102</td>
<td>0.09383</td>
<td>0.00047</td>
<td>0.00023</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>49</td>
<td>0.6</td>
<td>0.6</td>
<td>-0.00119</td>
<td>0.00020</td>
<td>-0.00117</td>
<td>0.09356</td>
<td>0.00073</td>
<td>0.00173</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>49</td>
<td>49</td>
<td>0.6</td>
<td>0.6</td>
<td>0.00047</td>
<td>0.00178</td>
<td>-0.00110</td>
<td>0.06319</td>
<td>0.00051</td>
<td>0.00073</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The true number of factor is 1 in the $z$ equation and 2 in the $y$ equation. In the estimation, the number of factors is set at 1 for the $z$ equation and 5 for the $y$ equation. True parameter values: $\beta_0 = 1$, $\beta_0 = 1$, $\sigma_{z0} = \sqrt{\frac{1}{2}}$, $\sigma_{y0} = \sqrt{\frac{1}{3}}$ (hence $\alpha_0 = (\frac{3}{4})^{-\frac{1}{2}}$). In the low endogeneity case, $\rho_0 = 0.2$, $\alpha_{z0} = 0.88$ and $\delta_0 = 0.2$. In the high endogeneity case, $\rho_0 = 0.6$, $\alpha_{z0} = 1.08$ and $\delta_0 = 0.6$. $\hat{\theta}_{nT}$ is the bias-corrected QML estimator. The coverage probabilities (CP) are calculated using the theoretical standard deviations obtained from the diagonal elements of $\Sigma^{-1}$.