Simultaneous Spatial Panel Data Models
with Common Shocks

Lina Lu

Simultaneous Spatial Panel Data Models with Common Shocks

Lina Lu [*]

This version: August 9, 2017

Abstract

I consider a simultaneous spatial panel data model, jointly modeling three effects: simultaneous effects, spatial effects and common shock effects. This joint modeling and consideration of cross-sectional heteroskedasticity result in a large number of incidental parameters. I propose two estimation approaches, a quasi-maximum likelihood (QML) method and an iterative generalized principal components (IGPC) method. I develop full inferential theories for the estimation approaches and study the trade-off between the model specifications and their respective asymptotic properties. I further investigate the finite sample performance of both methods using Monte Carlo simulations. I find that both methods perform well and that the simulation results corroborate the inferential theories. Some extensions of the model are considered. Finally, I apply the model to analyze the relationship between trade and GDP using a panel data over time and across countries.

Key Words: Panel data model; Spatial model; Simultaneous equations system; Common shocks; Simultaneous effects; Incidental parameters; Maximum likelihood estimation; Principal components; High dimensionality; Inferential theory.

JEL Classification: C13, C31, C33, C38, C51.

1 Introduction

In this paper, I consider a simultaneous spatial panel data model, jointly modeling three effects: simultaneous effects, spatial effects and common shock effects.[1] First, the simultaneous effect comes from the endogeneity of the dependent variables in a simultaneous equation system, and is important in many structural economic modeling. Second, the spatial effect is present in models where dependent variables are spatially interacted and spatial weights matrices are specified based on location and distance, in a geographic space or in more general economic, social or production network spaces. Third, common shocks stem from a common factor structure in panel data models, where the dependent variables' responses to shocks (i.e., factors) are heterogeneous and captured by the factor loadings.

That the model includes all three effects is useful in various fields. For example, this framework can be applied to analyze the relationship between trade volume and gross domestic product (GDP) within and across countries, a prominent research topic in international trade and macroeconomics. Within a country, trade volume is endogenously correlated with GDP, which can be regarded as a simultaneous effect. Across countries, a country's trade volume (or GDP) might be affected by other countries' trade volumes (or GDPs) through trade and financial linkages. This type of impact can be viewed as a spatial effect. Moreover, a global financial shock or a common energy shock might affect all countries' trade volumes and GDPs, which is referred to as a common shock effect. The model can also be applied in social network studies such as peer effects analysis in applied microeconomics, or in regional economic studies.[2]

In this paper, I consider the following simultaneous spatial panel data model, combining all three effects, with both a large time dimension T and a large cross-sectional dimension N:

$$
\begin{aligned}
y_{1it} &= \alpha_{1i} + \rho_1 \sum_{j=1}^{N} w_{1ij} y_{1jt} + \gamma_1 y_{2it} + \sum_{p=1}^{k_1} x_{1itp} \beta_{1p} + \lambda_i' f_t + e_{1it} \\
y_{2it} &= \alpha_{2i} + \rho_2 \sum_{j=1}^{N} w_{2ij} y_{2jt} + \gamma_2 y_{1it} + \sum_{q=1}^{k_2} x_{2itq} \beta_{2q} + \psi_i' f_t + e_{2it}
\end{aligned}
\tag{1.1}
$$

where $i = 1, 2, \ldots, N$; $t = 1, 2, \ldots, T$; y_{1it} and y_{2it} are the dependent variables for cross-section i at time t; $x_{1itp}, p = 1, 2, \ldots, k_1$ and $x_{2itq}, q = 1, 2, \ldots, k_2$ are explanatory variables, with their coefficients denoted as β_{1p} and β_{2q}, respectively; f_t is an r-dimensional vector of

---

[1] Related literature and studies of these three effects are provided in the end of the introduction section.

[2] For the peer effects studies as in Cohen-Cole et al. (2013) and Liu (2014), the common factor structure in my model can be used to capture unobservable individual characteristics which have time-varying impacts on individuals' decisions or choices. For these regional economic studies as in Jeanty et al. (2010), Baltagi and Bresson (2011), Gebremariam et al. (2011) and Hauptmeier et al. (2012), the common shocks can capture macroeconomic shocks which have heterogeneous impacts on local economies.

unobservable common shocks, termed the common factor; λ_i and ψ_i are the corresponding r-dimensional vectors of unobservable heterogeneous responses to the common shocks, termed the factor loadings; for $l = 1, 2$, $W_l = (w_{lij})_{N \times N}$ is a pre-specified spatial weights matrix whose diagonal elements w_{lii} are 0;[3] e_{1it} and e_{2it} are the idiosyncratic errors; and α_{1i} and α_{2i} are the intercepts. In model (1.1), taking the y_{1it} equation as an example, the term $(\gamma_1 y_{2it})$ captures the simultaneous effect from y_{2it} to y_{1it}, $(\rho_1 \sum_{j=1}^{N} w_{1ij} y_{1jt})$ captures the spatial effect, and $(\lambda_i' f_t)$ captures the common shock effect. The $(\lambda_i' f_t)$ part can also be viewed as an interactive fixed effect, which is more general than an additive fixed effect and provides a flexible way to model cross-sectional and serial correlations.[4] My interest is estimating the key coefficients $(\rho_1, \rho_2, \gamma_1, \gamma_2, \beta_1, \beta_2)$ and analyzing the asymptotic properties of their estimates.

In the econometrics literature, to the best of my knowledge, no existing paper jointly models these three effects. However, recently, a few papers consider two types of models combining two of these three effects. The first type is a spatial panel data model with common shocks in a single-equation system. Extending this type of model to a simultaneous equation system would make it applicable when multiple dependent variables are simultaneously interdependent, e.g., the above trade and GDP case. In estimating this type of model, Pesaran and Tosetti (2011) implement the same common correlated effects (CCE) estimation used in Pesaran (2006), while Bai and Li (2014b) propose a quasi-maximum likelihood (QML) method and Kuersteiner and Prucha (2015) use a generalized method of moments (GMM).[5] However, these estimation methods cannot be directly applied to my model due to the additional simultaneous structure of my framework. In addition, in the above trade and GDP example, if we use a single-equation system to study the effect of trade on GDP, their endogeneity would make the existing estimation methods in these papers inconsistent.

The second type is a spatial model in a simultaneous equation system but without the common shock effect. Two estimation methods have been studied for this type of model, instrumental variable (IV) methods (see Kelejian and Prucha (2004), Cohen-Cole et al. (2013), Baltagi and Deng (2015) and Liu (2014))[6] and QML methods (see Baltagi and

---

[3]More details of the weights can be found in Remark 2.1 on page 11 in this paper.

[4]The interactive fixed effects have been widely considered in the econometric literature, see Pesaran (2006), Bai (2009), Pesaran and Tosetti (2011), Bai and Li (2014a), and to name a few.

[5]The difference between the first two papers is that Pesaran and Tosetti (2011) specify the spatial interaction of the unobservable errors, while Bai and Li (2014b) specify the spatial interaction of the observable dependent variables. Thus, the CCE method cannot be applied to the model studied in Bai and Li (2014b), due to the endogeneity of the dependent variables. Kuersteiner and Prucha (2015) is based on a dynamic case where the dependent variable also depends on its previous value. In the estimation, Kuersteiner and Prucha (2015) first perform a quasi-transformation to eliminate the common shocks and then implement GMM.

[6]All of these papers focus on cross-sectional data, except Baltagi and Deng (2015), which is based on a panel data setting with random effects.

Bresson (2011), Wang et al. (2014) and Yang and Lee (2017)).[7] However, neither approach can be directly applied to my model due to the additional common shock effect. In all these papers, the errors are assumed to be idiosyncratic (i.e., uncorrelated over time and cross section), which is too strong in applications, and potential correlation of the errors would cause their estimation methods to be inconsistent. Augmenting this type of model with common shock effects can make it reasonable to assume that the new errors are idiosyncratic, since the common shocks would capture the correlations in the original errors, making the new errors idiosyncratic.[8]

In this paper, I focus on model (1.1). I present its estimation method and the corresponding asymptotic properties of the estimators. Under the joint presence of these three effects, there exist a large number of incidental parameters. In addition, I allow for cross-sectional heteroskedasticity in the errors, which is useful and important in spatial models[9] but gives rise to further incidental parameters due to the large number of variance parameters. To estimate the model, I propose two different approaches: a QML method and an iterative generalized principal components (IGPC) method. I show that both methods can effectively deal with the incidental parameters in model (1.1). For each method, I derive a full inferential theory for its estimators, which includes consistency, convergence rates and limiting distributions. To investigate finite sample performance, I conduct Monte Carlo simulations. I find that both methods perform well and that the simulation results corroborate the inferential theories derived in this paper. Furthermore, some extensions of the model are discussed. Finally, I apply the model to analyze the causal relationship between trade and GDP, taking into account spatial effects and global common shock effects.

Comparing the two approaches, I show that there is a trade-off between the model specification and the asymptotic property of the estimator. In the QML approach, I specify a model for the explanatory variables assuming that they are also affected by the common shocks and follow a common factor structure. The same specification of the explanatory variables has been considered in many papers; see Pesaran (2006), Bai and Li (2014a) and Castagnetti et al. (2006). Based on the fully specified model of the dependent and explanatory variables, I consider an objective function, which is the likelihood function if the factors and errors are assumed to be i.i.d normal distributed. Since the normality assumption is not required in this paper, this approach is referred to as the QML method. In computing its estimator (QMLE), the expectation maximization (EM) algorithm is implemented. Note

---

[7]Baltagi and Bresson (2011) propose a QML method to estimate a spatial seemingly unrelated regression panel data model with spatially correlated errors. Both Wang et al. (2014) and Yang and Lee (2017) are based on cross-sectional data with homoskedasticity. By comparison, Wang et al. (2014) implement a limited QML method without cross-equation correlation of the errors, while Yang and Lee (2017) consider a full information QML method allowing the errors to be correlated across equations.

[8]This common shock effect is an important feature to be implemented by various techniques, as noted in Pesaran (2006), Pesaran and Tosetti (2011), Bai (2009), Bai and Li (2014b) and Castagnetti et al. (2006).

[9]On inference, see Anselin (1988), Lin and Lee (2010), Kelejian and Prucha (2010), Bai and Li (2014b) and Baltagi and Deng (2015).

that I estimate the sample variance of the common factors instead of the factors themselves. The inferential theory shows that the QMLE is consistent, and its limiting distribution is unbiased (i.e., centered at zero) and has a smaller variance than that of the IGPC estimator. The gains of unbiasedness and more efficiency of the QMLE come at the cost of fully specifying the model of both the dependent and the explanatory variables.

In the IGPC approach, I do not specify the model for the explanatory variables but allow them to be arbitrarily correlated with the common factors and loadings, which is a more general approach than that used in QML. Unlike the treatment of the factors in the QML approach, I treat these as parameters and estimate them directly. In the estimation, I consider an objective function which is the likelihood function if errors are assumed to be i.i.d normal distributed, though such normality is not required in this paper. I then propose the IGPC method which is an iterative method based on the first-order conditions derived from the objective function. I call this estimation procedure the IGPC since one of the first-order conditions involves a generalized principal components method, and the word "generalized" stems from the heteroskedasticity assumption. The IGPC estimator (IGPCE) is consistent. Compared to the QMLE, the limiting distribution of the IGPCE is biased (i.e., not centered at zero) and has a larger variance. The cost of the bias and less efficiency of the IGPCE is offset by the gain of a more flexible model specification for the explanatory variables. In addition, based on the limiting distribution of the IGPCE, a bias-corrected IGPCE is obtained.

In Section 6, I apply the model to explore the relationship between trade volume and GDP using a panel data over time and across countries. My model is able to address endogeneity between trade and GDP, which is a well-known problem encountered in analyzing their relationship, as noted in Helpman (1988), Bradford et al. (1993), Rodrik (1995), Winters (2004) and Winters and Masters (2013). Thus far, economists have been making efforts to construct valid IVs for trade to tackle the endogeneity problem. For inferences, see Frankel and Romer (1999), Feyrer (2009), Felbermayr and Groschl (2013) and Ortega and Peri (2014). However, the validity of these IVs is still questionable. Unlike those papers that use a single-equation approach, I study the same type of question by modeling trade and GDP as a system of simultaneous equations and taking into account the endogeneity between them naturally. Moreover, despite their importance, global common shocks have not been well captured in the existing literature, whereas they can be captured using my model through a factor structure. Additionally, my model incorporates the spatial effect through international trade, which is implied from gravity theory as noted in Helpman (1987) and Anderson and van Wincoop (2003). In estimating the model, I implement the IGPC method, which does not need IVs. The empirical results show that all three effects emphasized in the model play important roles: 1) trade and GDP mutually and positively affect each other within a country (i.e., the simultaneous effect); 2) there exist spatial effects across countries for both trade and GDP (i.e., the spatial effect); and 3) global common shocks cannot be ignored.

The key finding is that the elasticity of GDP with respect to trade is approximately 0.1, while Feyrer (2009) finds an elasticity of approximately 0.5 using an IV approach.

Related literature. In both the empirical and theoretical literature, many papers consider the three effects separately. First, regarding spatial models, two estimation methods have been considered so far. One is the generalized method of moments (GMM) (see Kelejian and Prucha (1998, 1999, 2010), among others), and the other is the QML method (see Anselin (1988), Lee (2004a), Yu et al. (2008), Lee and Yu (2010a,b), Yu and Lee (2010), among others). Spatial models can be applied in many fields, such as spatial propogation of macroeconomic shocks in europe (Dewachter et al. (2012)), propogation of monetary policy shocks through production network (Ozdagli and Weber (2017)), international trade (Baltagi et al. (2008), Lawless (2009), and Rauch and Trindade (2002)), interregional trade (Keller and Shiue (2007)), banking and finance (Arezki et al. (2011) and Korte and Steffen (2014)), public economics (Egger et al. (2005)), transportation research (Frazier and Kockelman (2005)), good demand (Baltagi and Li (2006)), and agricultural economics (Druska and Horrace (2004)), among others.

Second, various methods have been studied for panel data models with common shocks. For instance, Pesaran (2006) propose CCE estimation; both Bai (2009) and Moon and Weidner (2017) consider a principle components (PC) method; Ahn et al. (2013) use GMM, and Bai and Li (2014b, 2015) implement QML. Regarding applications, common shocks models can be used in economic forecasting (Stock and Watson (2002a,b)), time trends modeling (Kneip et al. (2012)), analyzing spillovers in private returns to R&D (Eberhardt et al. (2013)), asset pricing (Bai and Ando (2014)), and so on.

Third, for simultaneous panel data models, IV approaches have been widely implemented; see Baltagi (1981), Balestra and Varadharajan-Krishnakumar (1987), Cornwell et al. (1992), Baltagi and Li (1992), among others. In practice, simultaneous panel data models can be applied to earnings studies (such as the income-schooling-ability simultaneous equations model considered in Chamberlain (1977a,b), Chamberlain and Griliches (1975), and Griliches (1979)), trade economics (Egger and Pfaffermayr (2004) and Serlenga and Shin (2007)), finance (Chen et al. (2006)) and operational management (Jain et al. (2013)).

In the application of spatial models, although many existing examples are based on a single-equation setup, spatial models with simultaneous equations have received more attention lately and have been widely used in various areas. For instance, these models have been in regional science studies of housing economics (Jeanty et al. (2010); Baltagi and Bresson (2011)); environmental and health economics (Ho and Hite (2008)); the determinants of local growth (interactions among migration, employment and income; see Gebremariam et al. (2011)); fiscal policy analysis (Hauptmeier et al. (2012) focus on fiscal competition over taxes and public input provisions, and Allers and Elhorst (2011) focus on the interactions between governments expenditures); and agricultural economics (Wu and Lin (2010)). Moreover, simultaneous spatial models can be applied in social network studies, such as the multi-choice

games in Cohen-Cole et al. (2013), Goldsmith-Pinkham and Imbens (2013) and Liu (2014). It would be potentially useful to apply my model to these areas by allowing common shocks to control for cross-sectional or serial correlations.

This paper proceeds as follows. I present the QML approach in Section 2 and the IGPC method in Section 3. In each section, I describe the model specification, assumptions, objective function, first-order conditions, inferential theory and computing algorithm. Then, in Section 4, I report the Monte Carlo simulation results for both approaches. Some extensions of the model are considered in Section 5, and an application is provided in Section 6. Finally, Section 7 concludes. Important notation is provided in Appendix A and B, and some proofs are presented in Appendix C. Other technical proofs and additional simulation results are provided in the supplementary material. Throughout the paper, $\|A\|$ is defined as the Frobenius norm of A, where $\|A\| = [\text{tr}(A'A)]^{1/2}$ for any $m \times n$ matrix A. In addition, $\dot{a}_t$ represents the de-meaned version of a column vector a_t, defined as $\dot{a}_t = a_t - \frac{1}{T}\sum_{t=1}^{T} a_t$, and M_{ab} is defined as $M_{ab} = \frac{1}{T}\sum_{t=1}^{T} \dot{a}_t\dot{b}_t'$ for any column vectors a_t and b_t.

2 First approach: the QML method

In the first approach, in addition to model (1.1), I specify a model of the explanatory variables by assuming that they are also affected by the common shocks and follow a factor structure. Such specification of the explanatory variables is applicable and widely considered, see Pesaran (2006), Bai and Li (2014a) and Castagnetti et al. (2006). Then, based on the fully specified model of both the dependent and explanatory variables, I consider the likelihood-based objective function and propose the QML method. In the estimation, I do not estimate the common factor f_t itself but its sample variance. Further, I develop a full inferential theory of its estimator and provide its computation algorithm. Some simulations results of this QML approach are presented in Section 4.

2.1 Model description and assumptions

In this section, in addition to model (1.1), I specify the model for the explanatory variables assuming that they are affected by the common shocks and following a factor structure of f_t, described as follows:

$$
\begin{aligned}
x_{1itp} &= \nu_{1ip} + \phi'_{1ip}f_t + v_{1itp}, \quad p = 1, 2, \ldots, k_1 \\
x_{2itq} &= \nu_{2iq} + \phi'_{2iq}f_t + v_{2itq}, \quad q = 1, 2, \ldots, k_2
\end{aligned}
\tag{2.1}
$$

where ϕ_{1ip} is an r-dimensional factor loading, representing the heterogeneous response of x_{1itp} to the common factor f_t; ϕ_{2iq} is defined in a similar way. Therefore, in the first approach, I consider a fully specified model of both dependent and explanatory variables, combining

(1.1) and (2.1).

Let $x_{1it} = (x_{1it1}, x_{1it2}, \ldots, x_{1itk_1})'$, $\beta_1 = (\beta_{11}, \beta_{12}, \ldots, \beta_{1k_1})'$, $\nu_1 = (\nu_{11}, \nu_{12}, \ldots, \nu_{1N})'$, $\phi_{1i} = (\phi_{1i1}, \phi_{1i2}, \ldots, \phi_{1ik_1})$, $v_{1it} = (v_{1it1}, v_{1it2}, \ldots, v_{1itk_1})'$, and define $x_{2it}, \beta_2, \nu_2, \phi_{2i}, v_{2it}$ in a similar way. Then, I can rewrite the model (1.1) and (2.1) as follows:

$$y_{1it} = \alpha_{1i} + \rho_1 \sum_{j=1}^{N} w_{1ij} y_{1jt} + \gamma_1 y_{2it} + x'_{1it}\beta_1 + \lambda'_i f_t + e_{1it}$$

$$y_{2it} = \alpha_{2i} + \rho_2 \sum_{j=1}^{N} w_{2ij} y_{2jt} + \gamma_2 y_{1it} + x'_{2it}\beta_2 + \psi'_i f_t + e_{2it} \tag{2.2}$$

$$x_{1it} = \nu_{1i} + \phi'_{1i} f_t + v_{1it}$$

$$x_{2it} = \nu_{2i} + \phi'_{2i} f_t + v_{2it}$$

Let $\mu_i = (\alpha_{1i}, \alpha_{2i}, \nu'_{1i}, \nu'_{2i})'$, $L_i = (\lambda_i, \psi_i, \phi_{1i}, \phi_{2i})$, and $\epsilon_{it} = (e_{1it}, e_{2it}, v'_{1it}, v'_{2it})'$. I can rewrite the above model as:

$$\begin{bmatrix} y_{1it} - \rho_1 \sum_{j=1}^{N} w_{1ij} y_{1jt} - \gamma_1 y_{2it} - x'_{1it}\beta_1 \\ y_{2it} - \rho_2 \sum_{j=1}^{N} w_{2ij} y_{2jt} - \gamma_2 y_{1it} - x'_{2it}\beta_2 \\ x_{1it} \\ x_{2it} \end{bmatrix} = \mu_i + L'_i f_t + \epsilon_{it}$$

Denote $\delta = (\rho_1, \rho_2, \gamma_1, \gamma_2, \beta'_1, \beta'_2)$, $k = k_1 + k_2$, and $\bar{k} = k + 2$. Let $D(\delta)$ be an $N\bar{k} \times N\bar{k}$ matrix whose (i,j) subblock, denoted by $D_{ij}(\delta)$, a $\bar{k} \times \bar{k}$ matrix, is equal to:

$$D_{ij}(\delta) = \begin{cases} \begin{bmatrix} 1 & -\gamma_1 & -\beta'_1 & 0 \\ -\gamma_2 & 1 & 0 & -\beta'_2 \\ 0 & 0 & I_{k_1} & 0 \\ 0 & 0 & 0 & I_{k_2} \end{bmatrix} & \text{if } i = j \\[2em] \begin{bmatrix} -\rho_1 w_{1ij} & 0 & 0 & 0 \\ 0 & -\rho_2 w_{2ij} & 0 & 0 \\ 0 & 0 & 0_{k_1} & 0 \\ 0 & 0 & 0 & 0_{k_2} \end{bmatrix} & \text{if } i \neq j \end{cases} \tag{2.3}$$

Now model (2.2) can be further transformed into the following matrix form (also a factor-structured model):

$$D(\delta)z_t = \mu + L f_t + \epsilon_t \tag{2.4}$$

where $z_t = (z_{1t}, z_{2t}, \ldots, z_{Nt})'$, with $z_{it} = (y_{1it}, y_{2it}, x'_{1it}, x'_{2it})'$, $L = (L_1, L_2, \ldots, L_N)'$, $\mu = (\mu'_1, \mu'_2, \ldots, \mu'_N)'$, and $\epsilon_t = (\epsilon'_{1t}, \epsilon'_{2t}, \ldots, \epsilon'_{Nt})'$. This matrix form will be used throughout the

8

first approach.

Throughout the paper, I assume that the number of factors r is fixed and known. In the simulation section, I propose a modified information criterion based on Bai and Ng (2002) to determine r for each of the two approaches.

2.1.1 Assumptions

To analyze model (2.2), I assume that there exists a constant $C > 0$ sufficiently large such that the following assumptions hold.

Assumption A: The factor f_t can be either fixed constants or random variables such that

A.1 Let $\dot{f}_t = f_t - \frac{1}{T}\sum_{t=1}^{T} f_t$, and $M_{ff} = T^{-1}\sum_{t=1}^{T} \dot{f}_t \dot{f}_t'$ be the sample variance of f_t. If f_t is fixed, I assume that $\|f_t\| \leq C$ for all t and $M_{ff} \to \Omega_F$. If f_t are random variables, I assume that $E(\|f_t\|^4) \leq C$ for all i and $M_{ff} \xrightarrow{p} \Omega_F$, where Ω_F is some positive definite matrix.

A.2 If f_t are random variables, I assume f_t to be independent of ϵ_{is} for all t and s.

Assumption B: The loading L_i can be either fixed constants or random variables such that

B.1 If L_i is fixed, I assume that $\|L_i\| \leq C$ for all i and $\frac{1}{N}L'\Sigma_{\epsilon\epsilon}^{-1}L \to \Omega_L$. If L_i are random variables, I assume that $E(\|L_i\|^4) \leq C$ for all i and $\frac{1}{N}L'\Sigma_{\epsilon\epsilon}^{-1}L \xrightarrow{p} \Omega_L$, where $\Sigma_{\epsilon\epsilon}$ is defined in Assumption B, and Ω_L is some positive definite matrix.

B.2 If L_i are random variables, I assume L_i to be independent of the idiosyncratic errors ϵ_{jt} for all i and j.

Assumptions A and B allow both the loadings and the common factors to be either fixed or random, which results in a model that is more general and applicable in various empirical studies.

Assumption C: The idiosyncratic errors $\epsilon_{it} = (e_{1it}, e_{2it}, v_{1it}', v_{2it}')'$ are such that

C.1 e_{lit} is independent and identically distributed over t and uncorrelated over i, with $E(e_{lit}) = 0$ and $E(e_{lit}^8) \leq \infty$ for all $l = 1, 2$, $i = 1, \cdots, N$ and $t = 1, \cdots, T$. Let σ_{li}^2 denote the variance of e_{lit}. I assume $C^{-1} \leq \sigma_{li}^2 \leq C$.

C.2 e_{1it} is independent of e_{2js} for all (i, j, t, s). Let Σ_{iie} denote the variance matrix of $e_{it} = (e_{1it}, e_{2it})'$, so I have $\Sigma_{iie} = \text{diag}(\sigma_{1i}^2, \sigma_{2i}^2)$, a diagonal 2×2 matrix. Let Σ_{ee} denote the variance matrix of $e_t = (e_{1t}', e_{2t}', \ldots, e_{Nt}')'$. Then, $\Sigma_{ee} = \text{diag}(\Sigma_{11e}, \Sigma_{22e}, \ldots, \Sigma_{NNe})$ is a diagonal $2N \times 2N$ matrix.

C.3 v_{lit} is independent and identically distributed over t and uncorrelated over i, with $E(v_{lit}) = 0$ and $E(\|v_{lit}\|^4) \leq \infty$ for all (l, i, t). Let Σ_{iivl} denote the variance matrix of v_{lit} and assume that all eigenvalues of Σ_{iivl} are uniformly bounded (UB) for all l and i. In addition, v_{1it} is independent of v_{2js} for all (i, j, t, s). Let $v_{it} = (v_{1it}, v_{2it})'$ and assume v_{it} is independent of e_{js} for all (i, j, t, s).

C.4 Let Σ_{ii} denote the variance matrix of ϵ_{it}, so I have $\Sigma_{ii} = \text{diag}(\sigma_{1i}^2, \sigma_{2i}^2, \Sigma_{iiv1}, \Sigma_{iiv2})$, a block-diagonal $\bar{k} \times \bar{k}$ matrix, where $\bar{k} = k_1 + k_2 + 2$. Let $\Sigma_{\epsilon\epsilon}$ denote the variance matrix of ϵ_t. Then, $\Sigma_{\epsilon\epsilon} = \text{diag}(\Sigma_{11}, \Sigma_{22}, \dots, \Sigma_{NN})$ is a block-diagonal $N\bar{k} \times N\bar{k}$ matrix.

Assumption C is that the variance of the idiosyncratic errors ϵ_t is a block-diagonal matrix, extending traditional factor analysis wherein a diagonal matrix is assumed instead. Moreover, Assumption C allows cross-sectional heteroskedasticity, which extends exiting studies with simultaneous spatial models where homoskedasticity is assumed, such as Kelejian and Prucha (2004), Baltagi and Bresson (2011), Wang et al. (2014), Baltagi and Deng (2015), Liu (2014) and Yang and Lee (2017). Note that neither Σ_{iiv1} nor Σ_{iiv2} need be diagonal, meaning that the k_1 components within the error v_{1it} can be correlated with each other. This is also the case for v_{2it}.

Assumption D: The underlying value $\delta = (\rho_1, \rho_2, \gamma_1, \gamma_2, \beta_1', \beta_2')'$ satisfies $\|\delta\| \leq C$.

Assumption E: Compactness of estimates.

E.1 The variances Σ_{ii} for all i and M_{ff} are estimated in a compact set, i.e., all the eigenvalues of $\hat{\Sigma}_{ii}$ and $\hat{M}_{ff}$ are in an interval $[C^{-1}, C]$.

E.2 The key parameters $\delta = (\rho_1, \rho_2, \gamma_1, \gamma_2, \beta_1', \beta_2')'$ are estimated in a compact set $\mathcal{A}_1 \times \mathcal{A}_2 \times \mathcal{A}_3 \times \mathcal{A}_4 \times \mathcal{A}_5 \times \mathcal{A}_6 \subset \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}^{k_1} \times \mathbb{R}^{k_2}$, where $\mathbb{R}$ is the set of real numbers.

Assumption E requires that the variance parameters are estimated in a compact set. Compactness is a condition for theoretical analysis, which is usually used when the objective function is highly nonlinear, for instance, in Newey and McFadden (1994), Jennrich (1969) and Wu (1981). I impose Assumption E here since the objective functions considered in both approaches presented in this paper are highly nonlinear. However, I do not require restrictions on the factor loading L_i.

Assumption F: Aassumptions about some important matrices.

F.1 The transformation matrix $D(\delta)$ is invertible.

F.2 W_1 and W_2 are constant $N \times N$ weights matrices with diagonal elements being zero.

F.3 Let $P_1 = (I_N - \rho_1 W_1)$ and $P_2 = (I_N - \rho_2 W_2)$. Then, I assume all matrices $P_1, P_2, (I - \gamma_1\gamma_2 P_1^{-1}P_2^{-1})$ and $(I - \gamma_1\gamma_2 P_2^{-1}P_1^{-1})$ are invertible.

F.4 Let $B_{12} = (I - \gamma_1\gamma_2 P_1^{-1}P_2^{-1})^{-1}$ and $B_{21} = (I - \gamma_1\gamma_2 P_2^{-1}P_1^{-1})^{-1}$. I assume that the row and column sums of matrices $W_1, W_2, P_1^{-1}, P_2^{-1}, B_{12}$ and B_{21} are all UB in absolute value.

Assumptions F.1–F.4 are standard in the spatial econometrics literature, for instance, Kelejian and Prucha (2004), Lee (2004a), Yu et al. (2008), Bai and Li (2014b) and Yang and Lee (2017). The invertibility of $D(\delta)$ (Assumption F.1) is standard in spatial models when using the QML method, which guarantees that the first-order conditions of δ exist

and the system has an equilibrium. Assumption F.2 is a standard normalization assumption for weights matrices. Assumption F.3 guarantees the invertibility of key matrices that will be used frequently in the theoretical analysis. The UB condition in Assumption F.4 keeps the degree of spatial correlation manageable and will be used in the theoretical analysis, especially in the consistency analysis.

Remark 2.1. In empirical applications, weights can be defined in many ways. Let w_{ij} be the entry of an $N \times N$ weights matrix W. The weight w_{ij} measures the presence and strength of an interaction between location i and j in a geographic space, or more generally, w_{ij} can be interpreted as the strength of a link between nodes i and j or between observations i and j in an economics or social network space. In applications, w_{ij} is usually a decreasing function of distance, as higher weights are assigned to closer observations than to distant observations. The most popular weighting scheme in practice is K-nearest neighbor weights, where location i is only affected by its K-nearest neighbors; more details about this scheme will be given in the simulation section. In the simplest case, when $K = 1$, the weights matrix is binary, where $w_{ij} = 1$ if i and j are neighbors (sharing a common boundary), and $w_{ij} = 0$ otherwise. The choice of weights matrix always depends on the empirical application. [10]

Remark 2.2. There is an alternative way to write the UB condition defined in Assumption F.4. First, an equivalent way to say that an m-by-n matrix A is UB in absolute row sum and column sum is to assume that $\limsup_{N \to \infty} \|A\|_\infty < \infty$ and $\limsup_{N \to \infty} \|A\|_1 < \infty$, where $\|A\|_\infty = \max_{1 \leq i \leq m} \sum_{j=1}^n |a_{ij}|$ represents the maximum absolute row-sum of A; $\|A\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^m |a_{ij}|$ represents the maximum absolute column sum of A, where a_{ij} is the (i,j)th element of A. Second, a set of three conditions $\limsup_{N \to \infty} \|W_1\|_\infty \leq 1$, $\limsup_{N \to \infty} \|W_1\|_1 \leq 1$ and $|\rho_1| < 1$, imply that P_1^{-1} is UB. This is because by definition of $P = (I_N - \rho_1 W_1)$,

$$\limsup_{N \to \infty} \|P^{-1}\|_\infty \leq \limsup_{N \to \infty} \sum_{j=0}^\infty (\|\rho_1 W_1\|_\infty)^j \leq \frac{1}{1 - \rho_1} < \infty$$

and

$$\limsup_{N \to \infty} \|P^{-1}\|_1 \leq \limsup_{N \to \infty} \sum_{j=0}^\infty (\|\rho_1 W_1\|_1)^j \leq \frac{1}{1 - \rho_1} < \infty$$

Further, a set of sufficient conditions for the assumption that B_{12} is UB can be that $\limsup_{N \to \infty} \|P_1^{-1} P_2^{-1}\|_\infty \leq 1$, $\limsup_{N \to \infty} \|P_1^{-1} P_2^{-1}\|_1 \leq 1$ and $|\gamma_1 \gamma_2| < 1$. Similar arguments can be made for P_2^{-1} and B_{21}.

---

[10] For example, in geographic spatial models, Ho and Hite (2008) uses a binary weights matrix, where the weight w_{ij} is nonzero only if i and j are neighbors. Jeanty et al. (2010) consider two choices of weights. One defines w_{ij} as a binary distance-based weight, as w_{ij} equals one only if the distance is smaller than a certain distance threshold and zero otherwise. The other defines the weight w_{ij} as an inverse distance function d_{ij}^{-a}, where d_{ij} measures the distance, and a is a dampening coefficient indicating how fast the weight decreases with distance. Furthermore, Cohen-Cole et al. (2013) and Liu (2014) consider the weights in the multi-choice game framework of a social network model.

Assumption G: Let $\eta = (\rho_1, \rho_2, \gamma_1, \gamma_2)$, for all $\eta^{\dagger} = (\rho_1^{\dagger}, \rho_2^{\dagger}, \gamma_1^{\dagger}, \gamma_2^{\dagger},) \in \mathcal{A}_1 \times \mathcal{A}_2 \times \mathcal{A}_3 \times \mathcal{A}_4$, with $\eta^{\dagger} \neq \eta$. One of the following two conditions holds:

G.1 For $\beta_1 \neq 0$ and $\beta_2 \neq 0$,

$$\liminf_{N \to \infty} \mathbb{M}_a > 0$$

where $\mathbb{M}_a = \begin{bmatrix} a_1 & 0 & b_1 & 0 \\ 0 & a_2 & 0 & b_2 \\ b_1 & 0 & c_1 & 0 \\ 0 & b_2 & 0 & c_2 \end{bmatrix}$, with $a_1, a_2, b_1, b_2, c_1, c_2$ being scalars and functions of $(\rho_1, \rho_2, \gamma_1, \gamma_2)$, depending on N, as defined in Table A2 in Appendix A.[11]

G.2

$$\liminf_{N \to \infty} \mathbb{M} > 0$$

where $\mathbb{M}$ is a 4×4 matrix, depending on $N, \eta, \eta^{\dagger}$ and variances $(\sigma_{1j}^2, \sigma_{2j}^2)$ $(j = 1, 2, \ldots, N)$. Its $(i,j)th$ entry is defined as $\mathbb{M}_{ij} = \frac{1}{N} \text{tr}(M_i M_j')$, where each M_l is an $N \times N$ matrix, for $l = 1, 2, 3, 4$, defined in Table A2 in Appendix A, and $\text{tr}(\cdot)$ is the trace operator.[12] The condition G.2 is equivalent to that, matrix $\mathbb{M}$ is positive definite for all N.

Remark 2.3. Assumption G imposes identification conditions for the key coefficients $\delta = (\rho_1, \rho_2, \gamma_1, \gamma_2)'$. Specifically, Assumption G.1 depends on $\beta \neq 0$, while Assumption G.2 does not depend on β. Conditions (G.1) and (G.2) are related to Assumption 8 and 9 respectively in Lee (2004a), and the two conditions in Assumption 8 in Yu et al. (2008), but differ in that they impose homoskedasticity and use a single-equation setup, while I allow cross-sectional heteroskedasticity and focus on a simultaneous equation system together with common shocks in this paper. Assumption G is also related to Lemmas 2, 3 and 4 in Yang and Lee (2017), but the difference is that they consider a cross-sectional simultaneous spatial model without common shocks and assume homoskedasticity. As shown in Appendix C, condition G.2 is related to the unique solution to $\mathcal{T}_{1N}(\rho_1^{\dagger}, \rho_2^{\dagger}, \gamma_1^{\dagger}, \gamma_2^{\dagger}, \sigma_{11}^{\dagger 2}, \sigma_{21}^{\dagger 2}, \ldots, \sigma_{1N}^{\dagger 2}, \sigma_{2N}^{\dagger 2}) = 0$, with:

$$\mathcal{T}_{1N}(\rho_1^{\dagger}, \rho_2^{\dagger}, \gamma_1^{\dagger}, \gamma_2^{\dagger}, \sigma_{11}^{\dagger 2}, \sigma_{21}^{\dagger 2}, \ldots, \sigma_{1N}^{\dagger 2}, \sigma_{2N}^{\dagger 2}) = -\frac{1}{2N} \text{tr}[\mathcal{R}^{\dagger} \Sigma_{ee} \mathcal{R}^{\dagger\prime} \Sigma_{ee}^{\dagger -1}] + \frac{1}{2N} \ln[\mathcal{R}^{\dagger} \Sigma_{ee} \mathcal{R}^{\dagger\prime} \Sigma_{ee}^{\dagger -1}] + 1$$

where $\mathcal{R}^{\dagger} = \Upsilon(\eta^{\dagger}) \Upsilon(\eta)^{-1}$, with $\eta = (\rho_1, \rho_2, \gamma_1, \gamma_2)$ and $\eta^{\dagger} = (\rho_1^{\dagger}, \rho_2^{\dagger}, \gamma_1^{\dagger}, \gamma_2^{\dagger})$.

Remark 2.4. The intuition behind the above identification condition is that, if there are explanatory variables x_1 and x_2, the model can be identified based on condition G.1; if not, the model can still be identified if either spatial effect or cross-sectional heteroskedasticity exists, implied from condition G.2.

---

[11] All $a_1, a_2, b_1, b_2, c_1, c_2$ involve matrices $W_1, W_2, P_1, P_2, B_{12}, B_{21}$ and $\mathbb{G}_l$ $(l = 1, \ldots, 4)$. The $N \times N$ matrices $\mathbb{G}_l$ $(l = 1, \ldots, 4)$ are defined in Table A1 in Appendix A

[12] Compared to the definitions of $a_1, a_2, b_1, b_2, c_1, c_2$, the matrices M_i not only depend on $W_1, W_2, P_1, P_2, B_{12}, B_{21}, \mathbb{G}_l$, but also on the variances $(\sigma_{1j}^2, \sigma_{2j}^2)$ $(j = 1, 2, \ldots, N)$.

Remark 2.5. The matrix $\mathbb{M}_a$ in Assumption G.1 is positive-definite if and only if that

$$a_1 > 0, (a_1c_1 - b_1^2) > 0, a_2 > 0, (a_2c_2 - b_2^2) > 0.$$

Furthermore, a sufficient condition for $a_1 > 0$ is following denoted as **(GS.1.1)**: there exists a positive constant ε such that at least one of the following two conditions holds:

$$\frac{1}{N}\left[\text{tr}(\Sigma_{1ee}^{-1}\mathbb{G}_1\Sigma_{1ee}\mathbb{G}_1') + \text{tr}(\mathbb{G}_1^2) - 2\sum_{i=1}^{N}[\mathbb{G}_{1,ii}]^2\right] > \varepsilon$$

$$\frac{1}{N}\left[\text{tr}(\Sigma_{1ee}^{-1}W_1\mathbb{G}_4\Sigma_{1ee}\mathbb{G}_4'W_1') + \text{tr}[(W_1\mathbb{G}_4)^2] - 2\sum_{i=1}^{N}[(W_1\mathbb{G}_4)_{ii}]^2\right] > \varepsilon$$

where $\Sigma_{1ee} = \text{diag}(\sigma_{11}^2, \sigma_{12}^2, \ldots, \sigma_{1N}^2)$; the $N \times N$ matrices $\mathbb{G}_1$ and $\mathbb{G}_4$ are defined in Table A1 in Appendix A; $\mathbb{G}_{1,ii}$ denotes the (i,i)th entry of matrix $\mathbb{G}_1$, which is similar for $(W_1\mathbb{G}_1)_{ii}$. To see this, it can be shown that the above condition $\frac{1}{N}\Big[\text{tr}(\Sigma_{1ee}^{-1}\mathbb{G}_1\Sigma_{1ee}\mathbb{G}_1') + \text{tr}(\mathbb{G}_1^2) - 2\sum_{i=1}^{N}[\mathbb{G}_{1,ii}]^2\Big] > \varepsilon$ implies $\frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}[\mathbb{G}_{1,ij}]^2 > \varepsilon_1$ for some positive constant ε_1. Similarly, the above condition $\frac{1}{N}\Big[\text{tr}(\Sigma_{1ee}^{-1}W_1\mathbb{G}_4\Sigma_{1ee}\mathbb{G}_4'W_1') + \text{tr}[(W_1\mathbb{G}_4)^2] - 2\sum_{i=1}^{N}[(W_1\mathbb{G}_4)_{ii}]^2\Big] > \varepsilon$ implies that $\frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}[(W_1\mathbb{G}_4)_{ij}]^2 > \varepsilon_2$ for some positive constant ε_2. Then, summarizing the preceding analysis, together with the definition of $a_1 = \frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\Big[[\mathbb{G}_{1,ij}]^2 + [(W_1\mathbb{G}_4)_{ij}]^2\Big]$, it follows that condition **(GS.1.1)** implies $a_1 > 0$. Similar arguments can be made for the other conditions involved in the above sufficient condition for (G.1). More details can be found in the supplementary material.

2.1.2 Normalization conditions for factors and factor loadings

In the factor analysis literature, it is well known that the factors and corresponding loadings can only be identified up to a rotation. The model considered in this paper can be regarded as an extension of the factor model and has the same rotational indeterminacy problem. Thus, in this section, I introduce a set of normalization conditions (NC) for both factors and factor loadings in order to facilitate the inference analysis.

Model (2.4) can be alternatively written as follows:

$$
\begin{aligned}
D(\delta)z_t &= \mu + Lf_t + \epsilon_t & (2.5)\\
&= (\mu + L\bar{f}) + L(f_t - \bar{f}) + \epsilon_t\\
&= \underbrace{(\mu + L\bar{f})}_{\mu^\star} + \underbrace{(LM_{ff}^{1/2}R)}_{L^\star}\underbrace{\left(R'M_{ff}^{-1/2}(f_t - \bar{f})\right)}_{f_t^\star} + \epsilon_t
\end{aligned}
$$

where R is an orthogonal matrix that consists the eigenvectors of $M_{ff}L'\Sigma_{\epsilon\epsilon}^{-1}LM_{ff}$ arranged in descending order. Let $\mu^\star$, $L^\star$ and $f_t^\star$ be the new intercepts, new loadings and new factors, respectively, as defined in the above equation. Then, model (2.4) is equivalent to:

$$D(\delta)z_t = \mu^\star + L^\star f_t^\star + \epsilon_t$$

where $\frac{1}{T}\sum_{t=1}^T f_t^\star = 0$, $\frac{1}{T}\sum_{t=1}^T f_t^\star f_t^{\star\prime} = I_r$, and $\frac{1}{N}L^{\star\prime}\Sigma_{\epsilon\epsilon}^{-1}L^\star$ is a diagonal matrix. Therefore, without loss of generality, I can impose the following NC for the factors and factor loadings in model (2.4):

NC.1 $\bar{f} = \frac{1}{T}\sum_{t=1}^T f_t = 0$

NC.2 $M_{ff} = \frac{1}{T}\sum_{t=1}^T (f_t - \bar{f})(f_t - \bar{f})' = I_r$

NC.3 $\frac{1}{N}L'\Sigma_{\epsilon\epsilon}^{-1}L = Q_N$, where Q_N is a diagonal matrix with its distinct diagonal elements arranged in descending order.

Remark 2.6. As shown later, NC.3 is not needed for the QML estimation of the regression coefficients δ, but it is needed to identify the factors and factor loadings. Under this NC, the orthogonal matrix R in (2.5), which is associated with the rotational indeterminacy of factors and factor loadings, now can be uniquely determined up to a column sign change. In addition, NC.3 simplifies the asymptotic analysis of the QMLE of δ.

Remark 2.7. In the factor analysis literature, the above NC are commonly used in maximum likelihood estimation; see, for instance, Anderson (2003). There are other NC to deal with rotational indeterminacy; see Bai and Li (2012) and Bai and Ng (2013).[13] For the QML approach in this paper, different NC will induce different estimates of the sample variance of factors M_{ff} and loadings L, which are the nuisance parameters in this paper, but they will not change the estimates of the key parameters δ and $\Sigma_{\epsilon\epsilon}$.

2.2 Objective function and first-order conditions

Let $\theta_1 = (\delta, L, \Sigma_{\epsilon\epsilon})$ be the parameters to be estimated. In this approach, I consider the following objective function:

$$\mathcal{L}_1(\theta_1) = -\frac{1}{2N}\ln|\Sigma_{zz}| + \frac{1}{N}\ln|D| - \frac{1}{2N}\text{tr}[DM_{zz}D'\Sigma_{zz}^{-1}] \tag{2.6}$$

where $\Sigma_{zz} = LL' + \Sigma_{\epsilon\epsilon}$; $D = D(\delta)$ is given in equation (2.3); and $M_{zz} = \frac{1}{T}\sum_{t=1}^T \dot{z}_t\dot{z}_t'$ is the data matrix. The above objective function is the likelihood function if f_t and ϵ_t are assumed

---

[13]Bai and Li (2012) consider five different sets of identification conditions and derive the inferential theories of the the corresponding QMLEs. Bai and Ng (2013) discuss three different sets of identification conditions for static factors in the PC analysis.

to be i.i.d. normal. Without such assumption, function (2.6) is referred to as the quasi-likelihood function.[14] The QMLE denoted as $\hat{\theta} = (\hat{\delta}, \hat{L}, \hat{\Sigma}_{\epsilon\epsilon})$ is defined as the maximizer of the above objective function:

$$\hat{\theta} = \underset{\theta_1 \in \Theta_1}{\operatorname{argmax}} \ \mathcal{L}_1(\theta_1)$$

where Θ_1 is the parameter space specified by Assumptions E and G, NC.1, NC.2 and NC.3. By the definition of D, as shown in Lemma A.1, $\det(D) = \det(\Upsilon(\eta))$, where $\eta = (\rho_1, \rho_2, \gamma_1, \gamma_2)'$, and $\Upsilon(\eta)$ is a $2N \times 2N$ matrix, with its (i, j)th block, a 2×2 matrix, equal to:

$$\Upsilon_{ij}(\eta) = \begin{cases} \begin{bmatrix} 1 & -\gamma_1 \\ -\gamma_2 & 1 \end{bmatrix} & \text{if } i = j \\[12pt] \begin{bmatrix} -\rho_1 w_{1ij} & 0 \\ 0 & -\rho_2 w_{2ij} \end{bmatrix} & \text{if } i \neq j \end{cases} \tag{2.7}$$

Compared to $D = D(\delta)$, $\Upsilon(\eta)$ only depends on η, without involving β_1 and β_2. Replacing $\det(D)$ with $\det(\Upsilon(\eta))$ in (2.6) implies the following alternative objective function, which will simplify the derivation of the first-order conditions thereafter:

$$\mathcal{L}_1(\theta_1) = -\frac{1}{2N} \ln |\Sigma_{zz}| + \frac{1}{N} \ln |\Upsilon(\eta)| - \frac{1}{2N} \operatorname{tr}[DM_{zz}D'\Sigma_{zz}^{-1}] \tag{2.8}$$

where only the last part, $-\frac{1}{2N}\operatorname{tr}[DM_{zz}D'\Sigma_{zz}^{-1}]$, involves β_1, β_2. Based on the above expression, we can derive the following first-order conditions for θ_1.

The first-order condition for L is:

$$\hat{L}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}(\hat{D}M_{zz}\hat{D}' - \hat{\Sigma}_{zz}) = 0 \tag{2.9}$$

where $\hat{D} = D(\hat{\delta})$. The first-order condition for $\Sigma_{\epsilon\epsilon}$ is:

$$\hat{D}M_{zz}\hat{D}' - \hat{\Sigma}_{zz} = \mathbb{W} \tag{2.10}$$

where $\mathbb{W}$ is an $N\bar{k} \times N\bar{k}$ matrix ($\bar{k} = k_1 + k_2 + 2$) whose ith $\bar{k} \times \bar{k}$ diagonal subblock denoted as $\mathbb{W}_{ii}$ is such that the diagonal entries of the upper-left 2×2 are zeros. Regarding the lower-right $(k_1 + k_2) \times (k_1 + k_2)$ submatrix of $\mathbb{W}_{ii}$, all entries of the upper-left $k_1 \times k_1$ and the lower-right $k_2 \times k_2$ are zeros. The rest of the elements of $\mathbb{W}$ are unspecified. The unspecified elements of $\mathbb{W}$ correspond to the zero elements of $\Sigma_{\epsilon\epsilon}$.

The first-order condition for ρ_1 is:

---

[14]In this paper, such normality assumption of ϵ_t is not required, as shown in both the theoretical analysis and the simulation section, the QML method is robust for different underlying distributions of errors.

$$\frac{1}{N}\text{tr}\Big(\Upsilon(\hat{\eta})^{-1} \cdot \Upsilon_{\rho_1}\Big) + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\ddot{y}_{1it}}{\hat{\sigma}_{1i}^2}(\dot{y}_{1it} - \hat{\rho}_1\ddot{y}_{1it} - \hat{\gamma}_1\dot{y}_{2it} - \dot{x}_{1it}'\hat{\beta}_1)$$
$$- \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\ddot{y}_{1it}}{\hat{\sigma}_{1i}^2}\hat{\lambda}_i'\hat{G}\hat{L}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}\hat{D}\dot{z}_t = 0 \tag{2.11}$$

where $\ddot{y}_{pit} = \sum_{j=1}^{N}w_{pij}\dot{y}_{pjt}$ for $p = 1, 2$, $\hat{G} = (I_r + \hat{L}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}\hat{L})^{-1}$, and Υ_{ρ_1} is the partial derivative of $\Upsilon(\eta)$ with respect to ρ_1, which is a constant $2N \times 2N$ matrix dependent only on weights W_1. Specifically, the (i, j)th subblock of Υ_{ρ_1} is a 2×2 matrix denoted by $(\Upsilon_{\rho_1})_{ij}$, which equals $0_{2\times 2}$ if $i = j$ and $(-w_{1ij}, 0; 0, 0)$ otherwise.

The first-order condition for ρ_2 is:

$$\frac{1}{N}\text{tr}\Big(\Upsilon(\hat{\eta})^{-1} \cdot \Upsilon_{\rho_2}\Big) + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\ddot{y}_{2it}}{\hat{\sigma}_{2i}^2}(\dot{y}_{2it} - \hat{\rho}_2\ddot{y}_{2it} - \hat{\gamma}_2\dot{y}_{1it} - \dot{x}_{2it}'\hat{\beta}_2)$$
$$- \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\ddot{y}_{2it}}{\hat{\sigma}_{2i}^2}\hat{\psi}_i'\hat{G}\hat{L}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}\hat{D}\dot{z}_t = 0 \tag{2.12}$$

where Υ_{ρ_2} is the partial derivative of $\Upsilon(\eta)$ with respect to ρ_2, which is a constant $2N \times 2N$ matrix dependent only on weights W_2. Specifically, the (i, j)th subblock of Υ_{ρ_2} is a 2×2 matrix denoted by $(\Upsilon_{\rho_2})_{ij}$ and equal to $0_{2\times 2}$ if $i = j$ and $(0, 0; 0, -w_{2ij})$ otherwise.

The first-order condition for γ_1 is:

$$\frac{1}{N}\text{tr}\Big(\Upsilon(\hat{\eta})^{-1} \cdot \Upsilon_{\gamma_1}\Big) + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\dot{y}_{2it}}{\hat{\sigma}_{1i}^2}(\dot{y}_{1it} - \hat{\rho}_1\ddot{y}_{1it} - \hat{\gamma}_1\dot{y}_{2it} - \dot{x}_{1it}'\hat{\beta}_1)$$
$$- \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\dot{y}_{2it}}{\hat{\sigma}_{1i}^2}\hat{\lambda}_i'\hat{G}\hat{L}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}\hat{D}\dot{z}_t = 0 \tag{2.13}$$

where Υ_{γ_1} is the partial derivative of $\Upsilon(\eta)$ with respect to γ_1, which is a constant $2N \times 2N$ matrix. Specifically, the (i, j)th subblock of Υ_{γ_1} is a 2×2 matrix denoted by $(\Upsilon_{\gamma_1})_{ij}$ and equal to $0_{2\times 2}$ if $i \neq j$ and $(0, -1; 0, 0)$ otherwise.

The first-order condition for γ_2 is:

$$\frac{1}{N}\text{tr}\left(\Upsilon(\hat{\eta})^{-1} \cdot \Upsilon_{\gamma_2}\right) + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\dot{y}_{1it}}{\hat{\sigma}_{2i}^2}(\dot{y}_{2it} - \hat{\rho}_2\ddot{y}_{2it} - \hat{\gamma}_2\dot{y}_{1it} - \dot{x}_{2it}'\hat{\beta}_2)$$
$$-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\dot{y}_{1it}}{\hat{\sigma}_{2i}^2}\hat{\psi}_i'\hat{G}\hat{L}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}\hat{D}\dot{z}_t = 0 \tag{2.14}$$

where Υ_{γ_2} is the partial derivative of $\Upsilon(\eta)$ with respect to γ_2, which is a constant $2N \times 2N$ matrix. Specifically, the (i,j)th subblock of Υ_{γ_2} is a 2×2 matrix denoted by $(\Upsilon_{\gamma_2})_{ij}$ and equal to $0_{2\times2}$ if $i \neq j$ and $(0, 0; -1, 0)$ otherwise.

The first-order condition for β_1 is:

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_{1i}^2}\dot{x}_{1it}(\dot{y}_{1it} - \hat{\rho}_1\ddot{y}_{1it} - \hat{\gamma}_1\dot{y}_{2it} - \dot{x}_{1it}'\hat{\beta}_1) - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_{1i}^2}\dot{x}_{1it}\hat{\lambda}_i'\hat{G}\hat{L}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}\hat{D}\dot{z}_t = 0 \tag{2.15}$$

The first-order condition for β_2 is:

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_{2i}^2}\dot{x}_{2it}(\dot{y}_{2it} - \hat{\rho}_2\ddot{y}_{2it} - \hat{\gamma}_2\dot{y}_{1it} - \dot{x}_{2it}'\hat{\beta}_2) - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_{2i}^2}\dot{x}_{2it}\hat{\psi}_i'\hat{G}\hat{L}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}\hat{D}\dot{z}_t = 0 \tag{2.16}$$

The above first-order conditions are useful in the derivation of the asymptotic properties, including the convergence rate and limiting distributions of the QMLE $\hat{\delta}$. They are involved neither in the proof of consistency nor in the computation of the QMLE. The QMLEs are computed via the expectation maximization (EM) algorithm, which does not need to solve these first-order conditions, but the EM solutions satisfy these conditions (proof is provided in the supplementary material).

2.3 Asymptotic properties of the QMLE

In this section, I first show that the QMLE is consistent and then present its convergence rates. Further, I provide the asymptotic representation and limiting distributions of the QMLE.

Proposition 2.1. *(Consistency)* *Under Assumptions A–G, when $N, T \to \infty$, for $\delta = (\eta', \beta_1', \beta_2')'$, I have:*

$$\hat{\delta} - \delta = o_p(1)$$

$$\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii} - \Sigma_{ii}\|^2 = o_p(1)$$

17

In addition, if NC.1–NC.3 hold, I have:

$$\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}^{-1}\| \cdot \|\hat{L}_i - L_i\|^2 = o_p(1)$$

Remark 2.8. In order to derive asymptotic properties, I need to specify $\det(D)$, D^{-1} and DD^{-1}, where D is the high dimensional transformation matrix and makes the theoretical analysis complicated. The number of incidental parameters goes to infinity when $N, T \to \infty$ brings additional complex.

Based on the consistency result, I further derive the rates of convergence of the QMLE.

Theorem 2.1. *(**Convergence rates**) Under Assumptions A–G, when $N, T \to \infty$, I have:*

$$\hat{\delta} - \delta = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-3/2})$$

$$\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii} - \Sigma_{ii}\|^2 = O_p(T^{-1})$$

In addition, if NC.1–NC.3 hold, I have:

$$\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}^{-1}\| \cdot \|\hat{L}_i - L_i\|^2 = O_p(T^{-1})$$

Remark 2.9. From Theorem 2.1, it can be seen that the QMLE of δ is $\sqrt{T}$-consistent even when N is finite, implying that the QML method still works when N is finite. Under fixed N, however, the asymptotic representation and limiting distribution of the QMLE will change. Theorem 2.1 also implies that based on the result that $\hat{\delta} - \delta$ has a faster convergence rate, the limiting distributions of $\text{vec}(\hat{L}_i - L_i)$ and $\text{vech}(\hat{\Sigma}_{ii} - \Sigma_{ii})$ are not affected by the estimation of δ and are the same as those in the pure factor model without regressors. Thus, in the following, I provide only the asymptotic representation of $\hat{\delta}$, excluding the estimated loadings and variances.[15]

---

[15]Bai and Li (2012) provide asymptotic representations and limiting distributions of the QMLE of the loadings and variances.

In order to state the asymptotic representation of $\hat{\delta}$, I introduce the following notation:

$$\Omega = \begin{bmatrix} \Omega_{11} & \Omega_{12} & \ldots & \Omega_{16} \\ \Omega_{21} & \Omega_{22} & \ldots & \Omega_{26} \\ \ldots & \ldots & \ldots & \ldots \\ \Omega_{61} & \Omega_{62} & \ldots & \Omega_{66} \end{bmatrix}; \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{bmatrix}$$

where the details of each Ω and ε entry are given in Tables A4 and A5, respectively, of Appendix A. Then, I have the following theorem.

Theorem 2.2. *(**Asymptotic representation**) Under Assumptions A–G, when $N, T \to \infty$ and $\sqrt{N}/T \to 0$, I have:*

$$\sqrt{NT}(\hat{\delta} - \delta) = \Omega^{-1}\sqrt{NT}\varepsilon + o_p(1)$$

Remark 2.10. The above expression is equivalent to:

$$\sqrt{NT} \begin{bmatrix} \hat{\rho}_1 - \rho_1 \\ \hat{\rho}_2 - \rho_2 \\ \hat{\gamma}_1 - \gamma_1 \\ \hat{\gamma}_2 - \gamma_2 \\ \hat{\beta}_1 - \beta_1 \\ \hat{\beta}_2 - \beta_2 \end{bmatrix} = \begin{bmatrix} \Omega_{11} & \Omega_{12} & \ldots & \Omega_{16} \\ \Omega_{21} & \Omega_{22} & \ldots & \Omega_{26} \\ \ldots & \ldots & \ldots & \ldots \\ \Omega_{61} & \Omega_{62} & \ldots & \Omega_{66} \end{bmatrix}^{-1} \cdot \sqrt{NT} \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{bmatrix} + o_p(1)$$

From the calculation of Ω and $\mathrm{var}(\sqrt{NT}\varepsilon)$ in the supplementary material, I show that Ω is symmetric and that $\Omega = \mathrm{var}(\sqrt{NT}\varepsilon)$, implying the following corollary.

Corollary 2.1. *(**Limiting distribution**) Under the assumptions of Theorem 2.2, I have:*

$$\sqrt{NT}(\hat{\delta} - \delta) \xrightarrow{d} N(0, \Omega_{QML})$$

where $\Omega_{QML} = \lim\limits_{N \to \infty} \Omega^{-1}$.

Remark 2.11. To gain an intuitive understanding of the asymptotic expression in Theorem

(2.2), consider the following simultaneous spatial panel data model without common shocks:

$$
\begin{aligned}
y_{1it} &= \alpha_{1i} + \rho_1 \sum_{j=1}^{N} w_{1ij} y_{1jt} + \gamma_1 y_{2it} + v'_{1it}\beta_1 + e_{1it} \\
y_{2it} &= \alpha_{2i} + \rho_2 \sum_{j=1}^{N} w_{2ij} y_{2jt} + \gamma_2 y_{1it} + v'_{2it}\beta_2 + e_{2it}
\end{aligned}
\tag{2.17}
$$

where e_{1it}, e_{2it}, v_{1it} and v_{2it} satisfy the same conditions as in Assumption C, but v_{1it} and v_{2it} are assumed to be observable (kind of the regressors). Conditional on v_{1it} and v_{2it}, the quasi-likelihood function of the above model (2.17), assuming normality of the errors (after concentrating out α_{1i} and α_{2i}), is:

$$
\begin{aligned}
\mathcal{L}'(\theta) = &-\frac{1}{2N} \sum_{i=1}^{N} \ln \sigma_{1i}^2 - \frac{1}{2N} \sum_{i=1}^{N} \ln \sigma_{2i}^2 + \frac{1}{N} \ln |\Delta(\eta)| \\
&- \frac{1}{2N} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1}{\sigma_{1i}^2} (\dot{y}_{1it} - \rho_1 \ddot{y}_{1it} - \gamma_1 \dot{y}_{2it} - \dot{v}'_{1it}\beta_1)^2 \\
&- \frac{1}{2N} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1}{\sigma_{2i}^2} (\dot{y}_{2it} - \rho_2 \ddot{y}_{2it} - \gamma_1 \dot{y}_{1it} - \dot{v}'_{2it}\beta_2)^2
\end{aligned}
\tag{2.18}
$$

where $\Delta(\eta) = I_{2N} - P$, with $P = \begin{bmatrix} \rho_1 W_1 & \gamma_1 I_N \\ \gamma_2 I_N & \rho_2 W_2 \end{bmatrix}$, and $\ddot{y}_{pit}$ is defined as in (2.11). Let $\tilde{\theta} = (\tilde{\rho}_1, \tilde{\rho}_2, \tilde{\gamma}_1, \tilde{\gamma}_2, \tilde{\sigma}_{11}^2, \ldots, \tilde{\sigma}_{21}^2, \ldots, \tilde{\sigma}_{1N}^2, \tilde{\sigma}_{2N}^2)$ be the QMLE of the above likelihood function. It can be shown that $(\tilde{\delta} - \delta)$ has the same asymptotic representation as in Theorem (2.2), which implies that the QML method can help address the endogenous parts of x_{1it} and x_{2it}, as they are affected by the common factors.

Remark 2.12. From Corollary 2.1, it can be seen that the limiting variance of the QMLE is not of a sandwich form, indicating that the QMLE is asymptotically efficient for simultaneous spatial panel models under cross-sectional heteroskedasticity. However, the situation becomes different when homoskedasticity is imposed instead, where the limiting variance of the QMLE would have a sandwich form. More details follow.

Consider model (2.2) but assume homoskedasticity. Then, the asymptotic expression for the QMLE (estimating homoskedastic variances) becomes:

$$
\sqrt{NT}(\tilde{\delta} - \delta) = \tilde{\Omega}^{-1} \sqrt{NT} \tilde{\varepsilon} + o_p(1)
$$

where $\tilde{\delta}$ is the QMLE of δ under homoskedasticity; $\tilde{\Omega}^{-1}$ and $\tilde{\varepsilon}$ are defined in Table 6 and Table 7, respectively, of Appendix A. Note that $\tilde{\Omega}$ and $\tilde{\varepsilon}$ are different from Ω and ε in the heteroskedastic case. More importantly, $\tilde{\varepsilon}$ now involves e_{1it}^2 and e_{2it}^2, while ε does not, imply-

ing that the limiting variance of $\tilde{\delta} - \delta$ will depend on the kurtosis of e_{1it} and e_{2it}. However, $\tilde{\Omega}$ does not depend on such kurtosis, so the limiting variance of $\tilde{\delta} - \delta$ has a sandwich form, unless normality of the errors is assumed. As shown in Corollary 2.1, the limiting variance of the QMLE under the heteroskedasticity assumption is not of a sandwich form, regardless of normality. This is a meaningful finding, demonstrating two important advantages of imposing the heteroskedasticity assumption. First, it makes the limiting variance of the QMLE robust to the underlying distributions of the errors; second, it eliminates potential inconsistency when homoskedasticity is incorrectly imposed.

2.4   Computation of the QMLE

To compute the QMLE, I propose a computing algorithm for the QMLE of model (2.2) combining the usual maximization procedures with the EM algorithm. Let $\theta^{(s)} = (\eta^{(s)}, \beta_1^{(s)}, \beta_2^{(s)}, L^{(s)}, \Sigma_{\epsilon\epsilon}^{(s)})$ with $\eta^{(s)} = (\rho_1^{(s)}, \rho_2^{(s)}, \gamma_1^{(s)}, \gamma_2^{(s)})$ denote the estimated value at the sth iteration. My updating procedures consist of two steps. In the first step, I update $L, \Sigma_{\epsilon\epsilon}, \beta_1$ and β_2 according to the EM algorithm:

$$L^{(s+1)} = \Big[\frac{1}{T}\sum_{t=1}^{T} E(D\dot{z}_t f_t' | \theta^{(s)})\Big]\Big[\frac{1}{T}\sum_{t=1}^{T} E(f_t f_t' | \theta^{(s)})\Big]^{-1} \tag{2.19}$$

$$\begin{aligned}\Sigma_{\epsilon\epsilon}^{(s+1)} &= \mathrm{Dg}\Big[D^{(s)} M_{zz} D^{(s)\prime} - L^{(s+1)} L^{(s)\prime} (\Sigma_{zz}^{(s)})^{-1} D^{(s)} M_{zz} D^{(s)\prime}\Big] \\ &= \mathrm{Dg}\Big\{\Big[I_{N(k_1+k_2+2)} - L^{(s+1)} L^{(s)\prime} (\Sigma_{zz}^{(s)})^{-1}\Big] D^{(s)} M_{zz} D^{(s)\prime}\Big\}\end{aligned} \tag{2.20}$$

and

$$\begin{aligned}\beta_1^{(s+1)} &= \Big[\sum_{i=1}^{N}\sum_{t=1}^{T} \frac{1}{(\sigma_{1i}^{(s+1)})^2}\dot{x}_{1it}\dot{x}_{1it}'\Big]^{-1} \\ &\times \Big[\sum_{i=1}^{N}\sum_{t=1}^{T} \frac{1}{(\sigma_{1i}^{(s+1)})^2}\dot{x}_{1it}\Big(\dot{y}_{1it} - \rho_1^{(s)}\sum_{j=1}^{N} w_{1ij}\dot{y}_{1jt} - \gamma_1^{(s)}\dot{y}_{2it} - \lambda_i^{(s+1)\prime} f_t^{(s)}\Big)\Big]\end{aligned} \tag{2.21}$$

$$\begin{aligned}\beta_2^{(s+1)} &= \Big[\sum_{i=1}^{N}\sum_{t=1}^{T} \frac{1}{(\sigma_{2i}^{(s+1)})^2}\dot{x}_{2it}\dot{x}_{2it}'\Big]^{-1} \\ &\times \Big[\sum_{i=1}^{N}\sum_{t=1}^{T} \frac{1}{(\sigma_{2i}^{(s+1)})^2}\dot{x}_{2it}\Big(\dot{y}_{2it} - \rho_2^{(s)}\sum_{j=1}^{N} w_{2ij}\dot{y}_{2jt} - \gamma_2^{(s)}\dot{y}_{1it} - \psi_i^{(s+1)\prime} f_t^{(s)}\Big)\Big]\end{aligned} \tag{2.22}$$

where Dg is the operator that sets the entries of its argument to zero if their counterparts in $E(\epsilon_t \epsilon_t')$ are zeros; $(\sigma_{1i}^{(s+1)})^2$ is the $[(i-1)(k_1+k_2+2)+1]$th diagonal element of $\Sigma_{\epsilon\epsilon}^{(s+1)}$,

and $(\sigma_{2i}^{(s+1)})^2$ is the $[(i-1)(k_1 + k_2 + 2) + 2]$th diagonal element of $\Sigma_{\epsilon\epsilon}^{(s+1)}$; $\lambda_i^{(s+1)}$ is the transpose of the $[(i-1)(k_1 + k_2 + 2) + 1]$th row of $L^{(s+1)}$, and $\psi_i^{(s+1)}$ is the transpose of the $[(i-1)(k_1 + k_2 + 2) + 2]$th row of $L^{(s+1)}$. In addition:

$$\frac{1}{T}\sum_{t=1}^{T} E(D\dot{z}_t f_t'|\theta^{(s)}) = D^{(s)} M_{zz} D^{(s)\prime}(\Sigma_{zz}^{(s)})^{-1} L^{(s)} \tag{2.23}$$

$$\frac{1}{T}\sum_{t=1}^{T} E(f_t f_t'|\theta^{(s)}) = I_r - L^{(s)\prime}(\Sigma_{zz}^{(s)})^{-1} L^{(s)} + L^{(s)\prime}(\Sigma_{zz}^{(s)})^{-1} D^{(s)} M_{zz} D^{(s)\prime}(\Sigma_{zz}^{(s)})^{-1} L^{(s)} \tag{2.24}$$

and

$$f_t^{(s)} = L^{(s)\prime}(\Sigma_{zz}^{(s)})^{-1} D^{(s)} \dot{z}_t \tag{2.25}$$

In the second step, η is updated by maximizing (2.6) with respect to η at $\beta_1 = \beta_1^{(s+1)}, \beta_2 = \beta_1^{(s+1)}$, $L = L^{(s+1)}$ and $\Sigma_{\epsilon\epsilon} = \Sigma_{\epsilon\epsilon}^{(s+1)}$ with an initial value of η at $\eta^{(s)}$. The two-step procedure suggested above is a version of the Expectation/Conditional Maximization Either (ECME) procedure in Liu and Rubin (1994). Combining these two steps, I obtain $\theta^{(s+1)} = (\eta^{(s+1)}, \beta_1^{(s+1)}, \beta_2^{(s+1)}, L^{(s+1)}, \Sigma_{\epsilon\epsilon}^{(s+1)})$. The iteration continues until $\|\theta^{(s+1)} - \theta^{(s)}\|$ is smaller than a preset tolerance.

This two-step iterative procedure guarantees that the value of the objective function (2.6) in each iteration does not decrease. This is because in the first step, letting $\eta = \eta^{(s)}$ be fixed and drawing on the standard theory of the EM algorithm, (for the inference, see Dempster et al. (1977) and McLachlan and Krishnan (1997)), I have the following inequality:

$$\mathcal{L}(\eta^{(s)}, \beta_1^{(s+1)}, \beta_2^{(s+1)}, \Phi^{(s+1)}, \Sigma_{\epsilon\epsilon}^{(s+1)}) \geq \mathcal{L}(\eta^{(s)}, \beta_1^{(s)}, \beta_2^{(s)}, \Phi^{(s)}, \Sigma_{\epsilon\epsilon}^{(s)}) \tag{2.26}$$

In the second step, by the definition of $\eta^{(s+1)}$, I have the following inequality:

$$\mathcal{L}(\eta^{(s+1)}, \beta_1^{(s+1)}, \beta_2^{(s+1)}, \Phi^{(s+1)}, \Sigma_{\epsilon\epsilon}^{(s+1)}) \geq \mathcal{L}(\eta^{(s)}, \beta_1^{(s+1)}, \beta_2^{(s+1)}, \Phi^{(s+1)}, \Sigma_{\epsilon\epsilon}^{(s+1)}) \tag{2.27}$$

In the supplementary material, I show that the limit of the iterated solution satisfies the first-order conditions (2.9)–(2.16) and hence possesses the local optimality property.

In the simulation results reported in the next section, I use the within-group estimator as the starting value for $\eta^{(1)}, \beta_1^{(1)}, \beta_2^{(1)}$, ignoring the endogeneity problem and the common shock effect. Then, let the initial values of $L^{(1)}$ and $\Sigma_{\epsilon\epsilon}^{(1)}$ be the maximizer of (2.6) given $\eta = \eta^{(1)}$, $\beta_1 = \beta_1^{(1)}$ and $\beta_2 = \beta_2^{(1)}$.

3   Second approach: the IGPC method

In the second approach, I do not specify the model for the explanatory variables but allow them to be arbitrarily correlated with the common factors, the factor loadings or both, which

is more general than the model specification considered in the first approach. Regarding the common factor f_t, I treat it as parameter and estimate it directly instead of estimating its sample variance as in the first approach. For this estimation, I propose an iterative approach based on a generalized principal components (GPC) method. Furthermore, I derive a full inferential theory of its estimator, the IGPCE, as in the first approach. Finally, I describe the computation of the IGPCE. The simulation results are provided in Section 4.

3.1 Model description and assumptions

In the second approach, I study model (1.1) without specifying the model for the explanatory variables. Using the same definitions of $x_{1it}, x_{2it}, \beta_1$ and β_2 given in (2.2), I can rewrite (1.1) as follows:

$$
\begin{aligned}
y_{1it} &= \alpha_{1i} + \rho_1 \sum_{j=1}^{N} w_{1ij} y_{1jt} + \gamma_1 y_{2it} + x_{1it}' \beta_1 + \lambda_i' f_t + e_{1it} \\
y_{2it} &= \alpha_{2i} + \rho_2 \sum_{j=1}^{N} w_{2ij} y_{2jt} + \gamma_2 y_{1it} + x_{2it}' \beta_2 + \psi_i' f_t + e_{2it}
\end{aligned}
\tag{3.1}
$$

Let $\alpha_i = (\alpha_{1i}, \alpha_{2i})'$, $x_{it} = \begin{bmatrix} x_{1it} & 0 \\ 0 & x_{2it} \end{bmatrix}$, $\beta = (\beta_1', \beta_2')'$, $\Gamma_i = (\lambda_i, \psi_i)$, and $e_{it} = (e_{1it}, e_{2it})'$. I can then rewrite model (3.1) as:

$$
\begin{bmatrix} y_{1it} - \rho_1 \sum_{j=1}^{N} w_{1ij} y_{1jt} - \gamma_1 y_{2it} \\ y_{2it} - \rho_2 \sum_{j=1}^{N} w_{2ij} y_{2jt} - \gamma_2 y_{1it} \end{bmatrix} = \alpha_i + x_{it}' \beta + \Gamma_i' f_t + e_{it}
$$

Using the same notation, $\eta = (\rho_1, \rho_2, \gamma_1, \gamma_2)$ and $\Upsilon(\eta)$, as in (2.7) in the first approach and letting $y_{it} = (y_{1it}, y_{2it})'$, model (3.1) can be transformed to:

$$
\sum_{j=1}^{N} \Upsilon_{ij}(\eta) y_{jt} = \alpha_i + x_{it}' \beta + \Gamma_i' f_t + e_{it}
\tag{3.2}
$$

Finally, let $Y_t = (y_{1t}', y_{2t}', \ldots, y_{Nt}')'$, $X_t = (x_{1t}, x_{2t}, \ldots, x_{Nt})'$, $\alpha = (\alpha_1', \alpha_2', \ldots, \alpha_N')'$, $\Gamma = (\Gamma_1, \Gamma_2, \ldots, \Gamma_N)'$, and $e_t = (e_{1t}', e_{2t}', \ldots, e_{Nt}')'$. I can then rewrite model (3.1) in the following matrix form:

$$
\Upsilon(\eta) Y_t = \alpha + X_t \beta + \Gamma f_t + e_t
\tag{3.3}
$$

3.1.1 Assumptions

In addition to the assumptions made in Section 2.1.1 in the first approach, I impose the following additional assumptions to facilitate the analysis in this second approach. Assume

that there is a sufficiently large constant $C > 0$ such that the following assumptions hold.

Assumption H: The explanatory variables x_{lit} can be either fixed constants or random variables for $l = 1, 2$. If x_{lit} is fixed, I assume $\|x_{lit}\| \leq C$ for all i and t. If x_{lit} are random variables, I assume $E(\|x_{lit}\|^4) \leq C$ for all i and t; in addition, x_{lit} is independent of the idiosyncratic error e_{mjs} for all (l, m, i, j, t, s).

Assumption H is newly imposed on the explanatory variables, since in this approach, I do not specify a model for them. To analyze model (3.1), I need to make the above assumption.

Assumption A$'$.2: If f_t are random variables, I assume f_t is independent of e_{is} for all t and s.

Assumption B$'$: The loading Γ_i can be either fixed constants or random variables such that

B$'$.1 If Γ_i is fixed, I assume that $\|\Gamma_i\| \leq C$ for all i and $\frac{1}{N}\Gamma'\Sigma_{ee}^{-1}\Gamma \to \Omega_\Gamma$. If Γ_i are random variables, I assume that $E(\|\Gamma_i\|^4) \leq C$ for all i and $\frac{1}{N}\Gamma'\Sigma_{ee}^{-1}\Gamma \xrightarrow{p} \Omega_\Gamma$, where Σ_{ee} is defined in Assumption C, and Ω_Γ is some positive definite matrix.

B$'$.2 If Γ_i are random variables, I assume that Γ_i is independent of the idiosyncratic errors e_{jt} for all i and j.

Assumption B$'$ is similar to Assumption B, but it is based on the new loading Γ, which is part of the loading L in the first approach. Since L contains Γ, Assumption B.2 implies Assumption B$'$.2 but not vice versa. However, Assumption B.1 cannot imply Assumption B$'$.1, and vice versa.

Assumption E$'$: Compactness of the estimates.

E$'$.1 The variances σ_{1i} and σ_{2i} for $i = 1, 2, \ldots, N$ are all estimated in compact sets, i.e., all variances σ_{1i} and σ_{2i} are estimated in an interval $[C^{-1}, C]$.

Assumption E.1 implies Assumption E$'$.1, since σ_{1i} and σ_{2i} are parts of Σ_{ii}. However, I do not need the compactness assumption of the estimate of M_{ff} here because in the second approach, I estimate the factor f_t itself instead of M_{ff}. Moreover, the compactness of the estimate of f_t is not required due to the nature of this estimation approach.

To state the following Assumption G$'$, let $\Im$ be the parameter space for Γ and Σ_{ee}, satisfying the assumptions and NC (which will be included in the following Section 3.1.2):

$$\Im = \left\{ \theta = (\Gamma, \Sigma_{ee}) \Big| C^{-1} \leq \sigma_{1i}^2 \leq C, C^{-1} \leq \sigma_{2i}^2 \leq C, \forall i; \frac{1}{N}\Gamma'\Sigma_{ee}^{-1}\Gamma = I_r \right\}$$

Assumption G$'$: One of the following two conditions holds:

G′.1 For $\beta_1 \neq 0$ and $\beta_2 \neq 0$, the matrix $\mathbb{D}_a = \begin{bmatrix} \mathbb{D}_b & \zeta \\ \zeta' & \mathbb{D}_c \end{bmatrix}$ is positive definite on $\Im$ for all N, where the $k \times k$ matrix $\mathbb{D}_b$, 4×4 matrix $\mathbb{D}_c$ and $k \times 4$ matrix ζ are all defined in Appendix B.

G′.2 For all $\eta^\dagger = (\rho_1^\dagger, \rho_2^\dagger, \gamma_1^\dagger, \gamma_2^\dagger,) \in \mathcal{A}_1 \times \mathcal{A}_2 \times \mathcal{A}_3 \times \mathcal{A}_4$, with $\eta^\dagger \neq \eta$, both $\mathbb{M}$ and $\mathbb{D}_b$ are positive definite on $\Im$ for all N, where the 4×4 matrix $\mathbb{M}$ is defined the same as in Assumption G.2.

Remark 3.1. The intuition behind Assumption G′ is similar to that behind Assumption G. In addition, the first part of condition G′.2 involving matrix $\mathbb{M}$ is same as Assumption G.2. However, Assumption G′.2 includes additional positive definite condition on $\mathbb{D}_b$ because now I estimate the factor f_t itself instead of its sample variance, which introduces more incidental parameters.

3.1.2 Normalization conditions for factors and factor loadings

As in Section 2.1.2, I introduce a set of NC to facilitate the inference analysis in the second approach. Note that model (3.3) can always be written as:

$$\Upsilon(\eta)Y_t = \underbrace{(\alpha + \Gamma \bar{f})}_{\alpha^\dagger} + X_t\beta + \underbrace{\Gamma Q^{-1/2}}_{\Gamma^\dagger} \underbrace{Q^{1/2}(f_t - \bar{f})}_{f_t^\dagger} + e_t$$

where $Q = \frac{1}{N}\Gamma'\Sigma_{ee}^{-1}\Gamma$, and $\bar{f} = \frac{1}{T}\sum_{t=1}^{T} f_t$. Using the definitions of $\alpha^\dagger$, $\Gamma^\dagger$ and $f_t^\dagger$ given in the above expression, I can treat them as the new intercept, new loading and new factor, respectively. Then, it can be seen that $\sum_{t=1}^{T} f_t^\dagger = 0$, and $\frac{1}{N}\Gamma^{\dagger\prime}\Sigma_{ee}^{-1}\Gamma^\dagger = I_r$. Thus, without loss of generality, in addition to NC.1 stated in Section 2.1.2, I can impose the following NC:

NC.4: $\frac{1}{N}\Gamma'\Sigma_{ee}^{-1}\Gamma = I_r$, where Σ_{ee} is defined in Assumption C.2.

3.2 Objective function and first-order conditions

In this approach, I allow the explanatory variables x_{1it} and x_{2it} to be arbitrarily correlated with the loading Γ_i and factor f_t. I treat both Γ_i and f_t as parameters and estimate them together.

Using the same definitions of $\delta, \eta, \Sigma_{ee}$ and $\Upsilon(\eta)$ as in the first approach, let $F = (f_1, f_2, \ldots, f_T)'$

and $\theta_2 = (\delta, \Gamma, \Sigma_{ee})$. I thus consider the following objective function in this approach:

$$\mathcal{L}_2^*(\theta_2, \alpha, F) = -\frac{1}{2NT}\sum_{t=1}^{T}\left(\Upsilon(\eta)Y_t - \alpha - X_t\beta - \Gamma f_t\right)'\Sigma_{ee}^{-1}\left(\Upsilon(\eta)Y_t - \alpha - X_t\beta - \Gamma f_t\right)$$
$$-\frac{1}{2N}\ln|\Sigma_{ee}| + \frac{1}{N}\ln|\Upsilon(\eta)|$$
(3.4)

The above expression can be viewed as the quasi-likelihood function by assuming the normality of e_{it}. Given δ, Γ and Σ_{ee}, it is easy to see that α and f_t maximize the above function $\mathcal{L}_2^*(\theta_2, \alpha, F)$ at:

$$\alpha = \Upsilon(\eta)\bar{Y} - \bar{X}\beta - \Gamma\bar{f} \tag{3.5}$$

and

$$f_t = (\Gamma'\Sigma_{ee}^{-1}\Gamma)^{-1}\Gamma'\Sigma_{ee}^{-1}(\Upsilon(\eta)\dot{Y}_t - \dot{X}_t\beta) \tag{3.6}$$

where $\bar{Y} = \frac{1}{T}\sum_{t=1}^{T}Y_t$, $\bar{X} = \frac{1}{T}\sum_{t=1}^{T}X_t$, $\bar{f} = \frac{1}{T}\sum_{t=1}^{T}f_t$, $\dot{Y}_t = Y_t - \bar{Y}$, and $\dot{X}_t = X_t - \bar{X}$. Substituting the above two formulas into $\mathcal{L}_2^*(\theta_2, \alpha, F)$ to concentrate out α and f_t, the objective function becomes:

$$\mathcal{L}_2(\theta_2) = -\frac{1}{2NT}\sum_{t=1}^{T}\left(\Upsilon(\eta)\dot{Y}_t - \dot{X}_t\beta\right)'\ddot{M}\left(\Upsilon(\eta)\dot{Y}_t - \dot{X}_t\beta\right) - \frac{1}{2N}\ln|\Sigma_{ee}| + \frac{1}{N}\ln|\Upsilon(\eta)| \quad (3.7)$$

where $\ddot{M} = \Sigma_{ee}^{-1} - \Sigma_{ee}^{-1}\Gamma(\Gamma'\Sigma_{ee}^{-1}\Gamma)^{-1}\Gamma'\Sigma_{ee}^{-1} = \Sigma_{ee}^{-1} - \frac{1}{N}\Sigma_{ee}^{-1}\Gamma\Gamma'\Sigma_{ee}^{-1}$ with the second equality due to NC.4.

Let $\tilde{\theta}_2 = (\tilde{\delta}, \tilde{\Gamma}, \tilde{\Sigma}_{ee})$ be the maximizer of the above objective function, defined as:

$$\tilde{\theta}_2 = \operatorname*{argmax}_{\theta_2 \in \Theta_2} \mathcal{L}_2(\theta_2)$$

where Θ_2 is the parameter space specified by Assumptions E.2, E'.1, G', NC.1 and NC.4.

Based on the above $\mathcal{L}_2(\theta_2)$, the first-order conditions for θ_2 can be derived as following (3.8)-(3.15). To compute $\tilde{\theta}_2$, I propose an iterative estimation procedure based on these first-order conditions. Since the first-order condition for loading Γ_i involves the generalized principal components (GPC) methodology, this estimation approach is referred to as an iterative generalized principal components method (IGPC) and its estimator (IGPCE) is denoted as $\check{\theta} = (\check{\delta}, \check{\Gamma}, \check{\Sigma}_{ee})$. The following are the first-order conditions for θ_2.

The first-order condition for Γ is:

$$\left[\frac{1}{NT}\sum_{t=1}^{T}\left(\Upsilon(\check{\eta})\dot{Y}_t - \dot{X}_t\check{\beta}\right)\left(\Upsilon(\check{\eta})\dot{Y}_t - \dot{X}_t\check{\beta}\right)'\right]\check{\Sigma}_{ee}^{-1}\check{\Gamma} = \check{\Gamma}\check{V} \tag{3.8}$$

where $\check{V}$ is a diagonal $r \times r$ matrix consisting of the first r largest eigenvalues of the $2N \times 2N$ matrix $\check{D}_\Gamma = \frac{1}{NT} \sum_{t=1}^{T} \left(\Upsilon(\check{\eta})\dot{Y}_t - \dot{X}_t\check{\beta} \right) \left(\Upsilon(\check{\eta})\dot{Y}_t - \dot{X}_t\check{\beta} \right)' \check{\Sigma}_{ee}^{-1}$. Here, $\check{\Gamma}$ contains the r eigenvectors associated with these r eigenvalues in $\check{V}$. Thus, the computation algorithm using the above equation is referred to as the GPC method, where the word "generalized" stems from the assumption of heteroskedasticity of the errors.

The first-order condition for σ_{1i}^2 is:

$$\check{\sigma}_{1i}^2 = \frac{1}{T} \sum_{t=1}^{T} \left(\dot{y}_{1it} - \check{\rho}_1 \ddot{y}_{1it} - \check{\gamma}_1 \dot{y}_{2it} - \dot{x}'_{1it}\check{\beta}_1 - \check{\lambda}'_i \check{f}_t \right)^2 \tag{3.9}$$

and the first-order condition for σ_{2i}^2 is:

$$\check{\sigma}_{2i}^2 = \frac{1}{T} \sum_{t=1}^{T} \left(\dot{y}_{2it} - \check{\rho}_2 \ddot{y}_{2it} - \check{\gamma}_2 \dot{y}_{1it} - \dot{x}'_{2it}\check{\beta}_2 - \check{\psi}'_i \check{f}_t \right)^2 \tag{3.10}$$

where $\dot{y}_{1it} = y_{1it} - \frac{1}{T}\sum_{s=1}^{T} y_{1is}$, $\dot{y}_{2it}$, $\dot{x}_{1it}$ and $\dot{x}_{2it}$ are defined in a similar way; $\ddot{y}_{1it} = \sum_{j=1}^{N} w_{1ij}\dot{y}_{1jt}$, and $\ddot{y}_{2it} = \sum_{j=1}^{N} w_{2ij}\dot{y}_{2jt}$; and

$$\check{f}_t = (\check{\Gamma}'\check{\Sigma}_{ee}^{-1}\check{\Gamma})^{-1}\check{\Gamma}'\check{\Sigma}_{ee}^{-1}\left(\Upsilon(\check{\eta})\dot{Y}_t - \dot{X}_t\check{\beta} \right) = \frac{1}{N}\check{\Gamma}'\check{\Sigma}_{ee}^{-1}\left(\Upsilon(\check{\eta})\dot{Y}_t - \dot{X}_t\check{\beta} \right)$$

The first-order condition for ρ_1 is:

$$-\frac{1}{NT} \sum_{t=1}^{T} \left(\Upsilon_{\rho_1} \cdot \dot{Y}_t \right)' \widehat{\check{M}} \left(\Upsilon(\check{\eta})\dot{Y}_t - \dot{X}_t\check{\beta} \right) + \frac{1}{N}\text{tr}\left\{ \Upsilon(\check{\eta})^{-1} \cdot \Upsilon_{\rho_1} \right\} = 0 \tag{3.11}$$

where Υ_{ρ_1} is defined in (2.11) and $\widehat{\check{M}} = \check{\Sigma}_{ee}^{-1} - \check{\Sigma}_{ee}^{-1}\check{L}(\check{L}'\check{\Sigma}_{ee}^{-1}\check{L})^{-1}\check{L}'\check{\Sigma}_{ee}^{-1} = \check{\Sigma}_{ee}^{-1} - \frac{1}{N}\check{\Sigma}_{ee}^{-1}\check{L}\check{L}'\check{\Sigma}_{ee}^{-1}$. The first-order condition for ρ_2 is:

$$-\frac{1}{NT} \sum_{t=1}^{T} \left(\Upsilon_{\rho_2} \cdot \dot{Y}_t \right)' \widehat{\check{M}} \left(\Upsilon(\check{\eta})\dot{Y}_t - \dot{X}_t\check{\beta} \right) + \frac{1}{N}\text{tr}\left\{ \Upsilon(\check{\eta})^{-1} \cdot \Upsilon_{\rho_2} \right\} = 0 \tag{3.12}$$

where Υ_{ρ_2} is defined in (2.12). The first-order condition for γ_1 is:

$$-\frac{1}{NT} \sum_{t=1}^{T} \left(\Upsilon_{\gamma_1} \cdot \dot{Y}_t \right)' \widehat{\check{M}} \left(\Upsilon(\check{\eta})\dot{Y}_t - \dot{X}_t\check{\beta} \right) + \frac{1}{N}\text{tr}\left\{ \Upsilon(\check{\eta})^{-1} \cdot \Upsilon_{\gamma_1} \right\} = 0 \tag{3.13}$$

where Υ_{γ_1} is defined in (2.13). The first-order condition for γ_2 is:

$$-\frac{1}{NT}\sum_{t=1}^{T}\left(\Upsilon_{\gamma_2}\cdot\dot{Y}_t\right)'\widehat{\ddot{M}}\left(\Upsilon(\check{\eta})\dot{Y}_t - \dot{X}_t\check{\beta}\right) + \frac{1}{N}\text{tr}\left\{\Upsilon(\check{\eta})^{-1}\cdot\Upsilon_{\gamma_2}\right\} = 0 \qquad (3.14)$$

where Υ_{γ_2} is defined in (2.14). The first-order condition for $\beta = (\beta_1', \beta_2')'$ is:

$$\frac{1}{NT}\sum_{t=1}^{T}\dot{X}_t'\widehat{\ddot{M}}\left(\Upsilon(\check{\eta})\dot{Y}_t - \dot{X}_t\check{\beta}\right) = 0 \qquad (3.15)$$

More details about the computation of the IGPCE are given in Section 3.4. These first-order conditions will be used in the derivation of the asymptotic properties of the IGPCE in Section 3.3.

3.3   Asymptotic properties of the IGPCE

In this section, I first show that the IGPCE is consistent and then derive its convergence rates, asymptotic representation and limiting distributions.

Proposition 3.1. (Consistency) *Under Assumptions A.1, A′.2, B′, C.1, C.2, D, E′.1, E.2, F, G′ and H, when $N, T \to \infty$, I have:*

$$\check{\delta} - \delta = o_p(1)$$

$$\frac{1}{N}\sum_{i=1}^{N}||\check{\Sigma}_{iie} - \Sigma_{iie}||^2 = o_p(1)$$

$$\frac{1}{N}\Gamma'\widehat{\ddot{M}}\Gamma = o_p(1)$$

where $\delta = (\eta', \beta')' = (\rho_1, \rho_2, \gamma_1, \gamma_2, \beta_1', \beta_2')'$, $\Sigma_{iie} = \text{diag}(\sigma_{1i}^2, \sigma_{2i}^2)$, and
$\widehat{\ddot{M}} = \check{\Sigma}_{ee}^{-1} - \check{\Sigma}_{ee}^{-1}\check{\Gamma}(\check{\Gamma}'\check{\Sigma}_{ee}^{-1}\check{\Gamma})^{-1}\check{\Gamma}'\check{\Sigma}_{ee}^{-1} = \check{\Sigma}_{ee}^{-1} - \frac{1}{N}\check{\Sigma}_{ee}^{-1}\check{\Gamma}\check{\Gamma}'\check{\Sigma}_{ee}^{-1}$.

Based on the consistency result, I derive the rates of convergence.

Theorem 3.1. (Convergence rates) *Let $H = \frac{1}{NT}\check{V}^{-1}(\check{\Gamma}'\check{\Sigma}_{ee}^{-1}\Gamma)(F'F)$. Under Assumptions A.1, A′.2, B′, C.1, C.2, D, E′.1, E.2, F, G′ and H, when $N, T \to \infty$, I have:*

$$\check{\delta} - \delta = O_p(N^{-1}) + O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-3/2})$$

$$\frac{1}{N}\sum_{i=1}^{N}||\check{\Sigma}_{iie} - \Sigma_{iie}||^2 = O_p(N^{-2}) + O_p(T^{-1})$$

$$\frac{1}{N}\sum_{i=1}^{N}||\check{\Gamma}_i - H\Gamma_i||^2 = O_p(N^{-2}) + O_p(T^{-1})$$

Remark 3.2. In the convergence rate for $\hat{\delta}$, note that there is a bias term of order $O_p(N^{-1})$, but there is no such bias term in the QMLE in the first approach. This bias term comes from the additional incidental parameters involved in the treatment of the common shocks. In this approach, I treat f_t as parameter and estimate it directly, whereas I estimate its sample variance instead in the first approach. Similarly, the extra term $O_p(N^{-2})$ included in the average convergence rates for $\check{\Sigma}_{iie}$ and $\check{\Gamma}_i$ occurs for the same reason. Because these extra terms depend only on N, the IGPCE is no longer consistent under fixed N, which is different from the QMLE. However, it is still true that since $\check{\delta} - \delta$ has a faster convergence rate, the limiting distributions of $\text{vec}(\check{\Gamma}_i - \Gamma_i)$ and $\text{vech}(\check{\Sigma}_{iie} - \Sigma_{iie})$ are the same as in the case of no regressors.

Theorem 3.2. *(Asymptotic representation)* *Under Assumptions A.1, A'.2, B', C.1, C.2, D, E'.1, E.2, F, G' and H, when $N, T \to \infty$ and $\sqrt{N}/T \to 0$, $\sqrt{T}/N \to 0$, I have:*

$$\sqrt{NT}(\check{\delta} - \delta + b) = \mathbb{D}^{-1}\sqrt{NT}\xi + o_p(1)$$

where $\delta = (\eta', \beta')' = (\rho_1, \rho_2, \gamma_1, \gamma_2, \beta_1', \beta_2')'$; $b, \mathbb{D}$ and ξ are defined as follows.

Let $k = k_1 + k_2$ and $\tilde{k} = k + 4$, and the $\tilde{k} \times 1$ vector ξ is defined as:

$$\xi = \frac{1}{NT} \begin{bmatrix} \sum_{t=1}^{T}(\dot{X}_t\beta + \Gamma f_t)'Q_1'\ddot{M}e_t - \sum_{t=1}^{T}\sum_{s=1}^{T}(\dot{X}_t\beta + \Gamma f_t)'Q_1'\ddot{M}e_s\pi_{st} + \varphi_1 \\ \sum_{t=1}^{T}(\dot{X}_t\beta + \Gamma f_t)'Q_2'\ddot{M}e_t - \sum_{t=1}^{T}\sum_{s=1}^{T}(\dot{X}_t\beta + \Gamma f_t)'Q_2'\ddot{M}e_s\pi_{st} + \varphi_2 \\ \sum_{t=1}^{T}(\dot{X}_t\beta + \Gamma f_t)'Q_3'\ddot{M}e_t - \sum_{t=1}^{T}\sum_{s=1}^{T}(\dot{X}_t\beta + \Gamma f_t)'Q_3'\ddot{M}e_s\pi_{st} + \varphi_3 \\ \sum_{t=1}^{T}(\dot{X}_t\beta + \Gamma f_t)'Q_4'\ddot{M}e_t - \sum_{t=1}^{T}\sum_{s=1}^{T}(\dot{X}_t\beta + \Gamma f_t)'Q_4'\ddot{M}e_s\pi_{st} + \varphi_4 \\ \sum_{t=1}^{T}\dot{X}_t'\ddot{M}e_t - \sum_{t=1}^{T}\sum_{s=1}^{T}\dot{X}_t'\ddot{M}e_s\pi_{st} \end{bmatrix}$$

where $\ddot{M} = \Sigma_{ee}^{-1} - \frac{1}{N}\Sigma_{ee}^{-1}\Gamma\Gamma'\Sigma_{ee}^{-1}$; $\pi_{st} = f_s'(F'F)^{-1}f_t$; $Q_1 = -\Upsilon_{\rho_1}\Upsilon(\eta)^{-1}$; $Q_2 = -\Upsilon_{\rho_2}\Upsilon(\eta)^{-1}$; $Q_3 = -\Upsilon_{\gamma_1}\Upsilon(\eta)^{-1}$; and $Q_4 = -\Upsilon_{\gamma_2}\Upsilon(\eta)^{-1}$, with all $\Upsilon_{\rho_1}, \Upsilon_{\rho_2}, \Upsilon_{\gamma_1}, \Upsilon_{\gamma_2}$ being constant matrices defined as in the first-order conditions (2.11)–(2.14) in Section 2.2. For $p = 1, 2, 3, 4$, the scalar φ_p is defined as:

$$\varphi_p = \frac{1}{NT}\sum_{t=1}^{T}e_t'Q_p^{o'}\Sigma_{ee}^{-1}e_t$$

where Q_p^o is an $2n \times 2N$ matrix that is obtained by setting all the diagonal elements of Q_p to zero. The $k \times 1$ vector b is defined as:

$$b = \mathbb{D}^{-1}\begin{bmatrix} \frac{1}{N}\text{tr}[\Gamma'Q_1^o\Sigma_{ee}^{-1}\Gamma(\Gamma'\Sigma_{ee}^{-1}\Gamma)^{-1}] \\ \frac{1}{N}\text{tr}[\Gamma'Q_2^o\Sigma_{ee}^{-1}\Gamma(\Gamma'\Sigma_{ee}^{-1}\Gamma)^{-1}] \\ \frac{1}{N}\text{tr}[\Gamma'Q_3^o\Sigma_{ee}^{-1}\Gamma(\Gamma'\Sigma_{ee}^{-1}\Gamma)^{-1}] \\ \frac{1}{N}\text{tr}[\Gamma'Q_4^o\Sigma_{ee}^{-1}\Gamma(\Gamma'\Sigma_{ee}^{-1}\Gamma)^{-1}] \\ 0_{k\times 1} \end{bmatrix} = \mathbb{D}^{-1}\begin{bmatrix} \frac{1}{N^2}\text{tr}[\Gamma'Q_1^o\Sigma_{ee}^{-1}\Gamma] \\ \frac{1}{N^2}\text{tr}[\Gamma'Q_2^o\Sigma_{ee}^{-1}\Gamma] \\ \frac{1}{N^2}\text{tr}[\Gamma'Q_3^o\Sigma_{ee}^{-1}\Gamma] \\ \frac{1}{N^2}\text{tr}[\Gamma'Q_4^o\Sigma_{ee}^{-1}\Gamma] \\ 0_{k\times 1} \end{bmatrix}$$

The $\tilde{k} \times \tilde{k}$ matrix $\mathbb{D}$ is defined as:

$$\mathbb{D} = \begin{bmatrix} \mathbb{D}_\eta + \Phi & \vartheta \\ \vartheta' & \mathbb{D}_\beta \end{bmatrix}$$

where the 4×4 matrices $\mathbb{D}_\eta$ and Φ, the $4 \times k$ matrix ϑ and the $k \times k$ matrix $\mathbb{D}_\beta$ are defined in Appendix B.

Remark 3.3. There is a bias term b of order $O_p(\frac{1}{N})$ in the IGPCE $\check{\delta}$ due to the treatment of the common shocks. The IGPC approach estimates f_t itself instead of its sample variance, which introduces more incidental parameters to the time dimension. As a comparison, the first approach estimates the sample variance of f_t instead, and its corresponding QMLE is unbiased in terms of limiting distribution.

Remark 3.4. There is some connection between matrix $\mathbb{D}$ in the above theorem and $\mathbb{D}_a$ in Assumption G'. Note that all matrices $\tilde{X}_l$ ($l = 1, 2, \ldots, \tilde{k}$) involved in $\mathbb{D}_a$ and $(\dot{X}_{1p}, \dot{X}_{2q})$ ($p = 1, 2, \ldots, k_1; q = 1, 2, \ldots, k_2$) involved in $\mathbb{D}$ are defined on the explanatory variables, different only in the ordering and presentation. Thus, I can simplify $\mathbb{D}_b$ (part of $\mathbb{D}_a$, which is associated with the identification of β) to $\mathbb{D}_\beta$ (part of $\mathbb{D}$, which is associated with the limiting variance of $\check{\beta}$). However, the matrix $\mathbb{D}_c$ (part of $\mathbb{D}_a$), which depends on the explanatory variables, does not equal $\mathbb{D}_\eta$ (part of $\mathbb{D}$), which depends on the dependent variables.

In the supplementary material, I show that $\mathbb{D}^{-1/2}\sqrt{NT}\xi \xrightarrow{d} N(0, I_{\tilde{k}})$ under the same conditions as in Theorem 3.2, which implies the following corollary.

Corollary 3.1. *(Limiting distribution)* *Under the assumptions of Theorem 3.2, when $N, T \to \infty$ and $T/N \to \kappa > 0$, I have:*

$$\sqrt{NT}(\hat{\delta} - \delta) \xrightarrow{d} N(-b^\diamond, \Omega_{IGPC})$$

where $\Omega_{IGPC} = \text{plim}_{N,T\to\infty}\mathbb{D}^{-1}$ and:

$$b^\diamond = \text{plim}_{N,T\to\infty} \left\{ \mathbb{D}^{-1} \begin{bmatrix} \kappa\frac{1}{N}\text{tr}[\Gamma'Q_1^o\Sigma_{ee}^{-1}\Gamma] \\ \kappa\frac{1}{N}\text{tr}[\Gamma'Q_2^o\Sigma_{ee}^{-1}\Gamma] \\ \kappa\frac{1}{N}\text{tr}[\Gamma'Q_3^o\Sigma_{ee}^{-1}\Gamma] \\ \kappa\frac{1}{N}\text{tr}[\Gamma'Q_4^o\Sigma_{ee}^{-1}\Gamma] \\ 0_{k\times 1} \end{bmatrix} \right\}$$

Remark 3.5. Regarding the limiting disribution, IGPCE has a bias term $b^\diamond$ while QMLE does not. In terms of efficiency, the limiting variance of the IGPCE is larger than that of the QMLE when the explanatory variables indeed follow the specification in QML, as shown in the supplementary material and such finding is also confirmed by the simulation results in

Section 4. The better asymptotic performance of the QMLE compared to the IGPCE comes at a cost: the QML approach restricts the model for the explanatory variables, whereas the IGPC approach does not.

Bias-corrected estimator: Using the asymptotic representation of $\check{\delta}$ stated in Theorem 3.2, I can construct a bias-corrected estimator for δ by substitution as follows:

$$\check{\delta}^{\star} = \check{\delta} + \check{b}$$

where $\check{b}$ is the estimator of b by replacing all the true parameters $\delta, \Gamma, F, \Sigma_{ee}$ with their IGPCEs $\check{\delta}, \check{\Gamma}, \check{F}, \check{\Sigma}_{ee}$. Then, I have the following limiting distribution for the bias-corrected estimator $\check{\delta}^{\star}$.

Theorem 3.3. *(Limiting distributions for bias-corrected estimators)* *Under the assumptions of Theorem 3.2, when $N, T \to \infty$ and $T/N \to \kappa > 0$, I have:*

$$\sqrt{NT}(\check{\delta}^{\star} - \delta) \xrightarrow{d} N(0, \Omega_{IGPC})$$

where Ω_{IGPC} is defined in Corollary 3.1.

3.4 Computation of the IGPCE

Computation of the IGPCE involves an iterative procedure based on the first-order conditions (3.8)–(3.15) in Section 3.2. In sth iteration, I update $\check{\theta}^{(s+1)} = (\check{\Gamma}^{(s+1)}, \check{\Sigma}_{ee}^{(s+1)}, \check{\beta}^{(s+1)}, \check{\eta}^{(s+1)})$, where $\check{\eta}^{(s+1)} = (\check{\rho}_1^{(s+1)}, \check{\rho}_2^{(s+1)}, \check{\gamma}_1^{(s+1)}, \check{\gamma}_2^{(s+1)})$ as follows.

In the first step, $\check{\Gamma}^{(s+1)}$ is computed as the first r eigenvectors associated with the first r largest eigenvalues of the $2N \times 2N$ matrix $\check{D}_{\Gamma}^{(s)} = \frac{1}{NT} \sum_{t=1}^{T} \left(\Upsilon(\check{\eta}^{(s)}) \dot{Y}_t - \dot{X}_t \check{\beta}^{(s)} \right) \left(\Upsilon(\check{\eta}^{(s)}) \dot{Y}_t - \dot{X}_t \check{\beta}^{(s)} \right)' \left(\check{\Sigma}_{ee}^{(s)} \right)^{-1}$.

In the second step, update $\check{\Sigma}_{ee}^{(s+1)}$ according to:

$$(\check{\sigma}_{1i}^{(s+1)})^2 = \frac{1}{T} \sum_{t=1}^{T} \left(\dot{y}_{1it} - \check{\rho}_1^{(s)} \ddot{y}_{1it} - \check{\gamma}_1^{(s)} \dot{y}_{2it} - \dot{x}_{1it}' \check{\beta}_1^{(s)} - \check{\lambda}_i^{(s+1)\prime} \check{f}_t^{(s)} \right)^2$$

and

$$(\check{\sigma}_{2i}^{(s+1)})^2 = \frac{1}{T} \sum_{t=1}^{T} \left(\dot{y}_{2it} - \check{\rho}_2^{(s)} \ddot{y}_{2it} - \check{\gamma}_2^{(s)} \dot{y}_{1it} - \dot{x}_{2it}' \check{\beta}_2^{(s)} - \check{\psi}_i^{(s+1)\prime} \check{f}_t^{(s)} \right)^2$$

where

$$\check{f}_t^{(s)} = \frac{1}{N} \check{\Gamma}^{(s+1)\prime} (\check{\Sigma}_{ee}^{(s)})^{-1} \left(\Upsilon(\check{\eta}^{(s)}) \dot{Y}_t - \dot{X}_t \check{\beta}^{(s)} \right)$$

In the third step, update $\check{\beta}^{(s+1)}$ according to:

$$\check{\beta}^{(s+1)} = \left(\dot{X}_t' \widehat{\ddot{M}}^{(s+1)} \dot{X}_t\right)^{-1} \left(\dot{X}_t' \widehat{\ddot{M}}^{(s+1)} \Upsilon(\check{\eta}^{(s)}) \dot{Y}_t\right)^{-1}$$

where $\widehat{\ddot{M}}^{(s+1)} = (\check{\Sigma}_{ee}^{(s+1)})^{-1} - \frac{1}{N}(\check{\Sigma}_{ee}^{(s+1)})^{-1}\check{\Gamma}^{(s+1)}\check{\Gamma}^{(s+1)\prime}(\check{\Sigma}_{ee}^{(s+1)})^{-1}$.

In the final step, update $\check{\eta}^{(s+1)} = (\check{\rho}_1^{(s+1)}, \check{\rho}_2^{(s+1)}, \check{\gamma}_1^{(s+1)}, \check{\gamma}_2^{(s+1)})$ by directly maximizing the likelihood $\mathcal{L}_2(\theta_2)$ (3.7) with respect to η at $\Gamma = \check{\Gamma}^{(s+1)}, \Sigma_{ee} = \check{\Sigma}_{ee}^{(s+1)}$ and $\beta = \check{\beta}^{(s+1)}$. Combining these steps, I obtain $\check{\theta}^{(s+1)} = (\check{\Gamma}^{(s+1)}, \check{\Sigma}_{ee}^{(s+1)}, \check{\beta}^{(s+1)}, \check{\eta}^{(s+1)})$. The iteration continues until the distance $\|\check{\theta}^{(s+1)} - \check{\theta}^{(s)}\|$ is smaller than a preset tolerance.

In the simulation results reported in Section 4, similar to the QML approach, I use the within-group estimator as the starting value for $\check{\beta}^{(1)}$ and $\check{\eta}^{(1)}$. Then let $\check{\Gamma}^{(1)}, \check{\Sigma}_{ee}^{(1)}$ be the solution according to the above first and second steps, given $\check{\beta}^{(1)}, \check{\eta}^{(1)}$. The simulation results show that the IGPCE performs well in finite sample and corroborate its asymptotic properties, as derived in this paper.

4   Finite sample properties via simulations

In this section, I investigate the finite sample performance of both approaches by Monte Carlo simulation. The simulation results reported in the following sections show that both approaches work well and corroborate the inferential theories derived in this paper.

4.1   Data generating processes

I consider two different data generating processes (DGPs). Both DGPs follow the model of the dependent variables in (1.1), but they use different specifications of the explanatory variables. DGP1 generates the explanatory variables according to (2.1), while DGP2 does not follow (2.1).

Specifically, both DGP1 and DGP2 generate the dependent variables according to:

$$y_{1it} = \alpha_{1i} + \rho_1 \sum_{j=1}^{N} w_{1ij} y_{1jt} + \gamma_1 y_{2it} + x_{1it}\beta_1 + \lambda_i' f_t + e_{1it}$$

$$y_{2it} = \alpha_{2i} + \rho_2 \sum_{j=1}^{N} w_{2ij} y_{2jt} + \gamma_2 y_{1it} + x_{2it}\beta_2 + \psi_i' f_t + e_{2it}$$

$$((\rho_1, \rho_2, \gamma_1, \gamma_2, \beta_1, \beta_2) = (0.2, 0.2, 0.2, 0.2, 1, 2))$$

where $f_t = (f_{t1}, f_{t2})'$, $\lambda_i = (\lambda_{i1}, \lambda_{i2})'$, and $\psi_i = (\psi_{i1}, \psi_{i2})'$. The variables $\alpha_{1i}, \alpha_{2i}, \lambda_{il}, \psi_{il}$ and f_{tl} are all i.i.d. $N(0,1)$. I generate the errors e_{1it} and e_{2it} with cross-sectional heteroskedasticity. I set $e_t = \sqrt{\text{diag}(\Xi^\dagger)}\varepsilon_t^\dagger$, where $\varepsilon_t^\dagger$ is a $2N$ dimensional column vector with all the

elements being $(\chi_2^2 - 2)/2$ independently, where χ_2^2 denotes the chi-squared distribution with two degrees of freedom, and normalized to zero mean and unit variance. In addition, $\Xi^\dagger$ is a $2N$ dimensional column vector, whose mth element is set to:

$$\Xi_m^\dagger = 0.1 + \frac{1 - \eta_m}{\eta_m} \iota_m' \iota_m, \qquad m = 1, 2, \ldots, 2N$$

where η_m is drawn from U$[0.1, 0.9]$, and ι_m is the transpose of the mth row of Γ; the constant 0.1 keeps the variance away from zero.

The spatial weights matrices W_1 and W_2 are generated based on the "q ahead and q behind" framework, similar to that in Kelejian and Prucha (1999), Baltagi and Deng (2015), among others. In the "q ahead and q behind" framework, all the individuals are arranged in a circle, and each individual is affected only by the q individuals immediately in front and immediately behind it with equal weight. Then, the weight matrix is row normalized to ensure that the sum of each row is equal to 1. Thus, the non-zero weight equals $\frac{1}{2q}$.[16] I consider two setups of the spatial weights matrix, "1 ahead and 1 behind" and "5 ahead and 5 behind", for both the QML and the IGPC approaches.

DGP1 and DGP2 differ in the generation of the explanatory variables, as explained in the following.

4.1.1 DGP1

DGP1 generates them according to model (2.1) in the following specification:

$$x_{1it} = \nu_{1i} + a_1 \phi_{1i}' f_t + v_{1it}$$
$$x_{2it} = \nu_{2i} + a_2 \phi_{2i}' f_t + v_{2it}, \qquad (a_1 = a_2 = 1)$$

where f_t is the same as in the above model of y; $\phi_{1i} = (\phi_{1i1}, \phi_{1i2})'$, and $\phi_{2i} = (\phi_{2i1}, \phi_{2i2})'$, where $\phi_{1il} = \lambda_{il} + u_{1il}$ and $\phi_{2il} = \psi_{il} + u_{2il}$. All the variables $\nu_{1i}, \nu_{2i}, u_{1il}$ and u_{2il} are i.i.d. with $N(0, 1)$. I also generate the errors v_{1it} and v_{2it} with cross-sectional heteroskedasticity by setting $v_t = \sqrt{\text{diag}(\Xi^\star)} \varepsilon_t^\star$, where $\varepsilon_t^\star$ is a $2N$ dimensional vector generated as in $\varepsilon_t^\dagger$; $\Xi^\star$ is a $2N$ dimensional column vector generated similar to $\Xi^\dagger$ but Γ is replaced with $\Gamma^\star$, where $\Gamma^\star = (\Gamma_1^\star, \Gamma_2^\star, \ldots, \Gamma_N^\star)'$, with $\Gamma_i^\star = (\phi_{1i}, \phi_{2i})$.

---

[16]In the simulation, I set the cross-sectional dimension N to be larger than $2q$, so the non-zero weight under the "q ahead and q behind" framework is always $\frac{1}{2q}$. In practice, when $N < 2q$, then the weights matrix under the "q ahead and q behind" is defined as follows: all the diagonal elements are 0, and all the off-diagonal elements are $\frac{1}{(N-1)}$.

4.1.2 DGP2

DGP2 generates the explanatory variables based on DGP1 but with truncation (similarly to Moon et al. (2014)), specified as follows:

$$x_{1it} = \left[\nu_{1i} + a_1 \phi'_{1i} f_t + v_{1it}\right] \left[\nu_{1i} + a_1 \phi'_{1i} f_t + v_{1it} > -3.5\right]$$
$$x_{2it} = \left[\nu_{2i} + a_2 \phi'_{2i} f_t + v_{2it}\right] \left[\nu_{2i} + a_2 \phi'_{2i} f_t + v_{2it} > -3.5\right], \qquad (a_1 = a_2 = 1)$$

where the variables ν_{li}, ϕ_{li}, f_t and v_{1it} are the same as in DGP1. With truncation, the explanatory variables are no longer a factor structure of common shocks f_t as in model (2.1).

Remark 4.1. The errors in the above two DGPs are non-normal and skewed. I also consider the cases when the errors have normal or student's t distributions. The corresponding simulation results for both approaches are provided in the supplementary material.

4.2 Finite sample performance of the QMLE

In this section, I provide the simulation results of the QML approach based on the above two DGPs and both setups of the spatial weights matrix.

In addition, since the number of factors r is usually unknown in practice, I propose a likelihood-based information criterion following Bai and Ng (2002) to determine it in the QML approach. Specifically, r is determined by:

$$\hat{r} = \underset{0 \leq m \leq r_{max}}{\operatorname{argmin}} IC(m) \tag{4.1}$$

with

$$IC(m) = \frac{1}{2N\bar{k}} \ln \left|\hat{L}^m \hat{L}^{m\prime} + \hat{\Sigma}^m_{\epsilon\epsilon}\right| - \frac{1}{N\bar{k}} \ln |\Upsilon(\hat{\eta}^m)| + m \frac{N\bar{k} + T}{2N\bar{k}T} \ln[\min(N\bar{k}, T)]$$

where $\bar{k} = k_1 + k_2 + 2$, and $(\hat{\eta}^m, \hat{L}^m, \hat{\Sigma}^m_{\epsilon\epsilon})$ are the QMLE of $(\eta, L, \Sigma_{\epsilon\epsilon})$ when the number of factors is set to m. In the simulation, I take DGP1, for example, and report the percentage of r values that are correctly estimated by (4.1) (set $r_{\max} = 4$) based on 1000 repetitions in the third row of Tables 1 and 2. The results show that the percentage of correctly estimated r values is very high and equal to or close to 100% for different combinations of (N, T) and different setups of the weights matrix. I then conduct simulations for the QMLE by assuming that the true number of factors is known in both DGP1 and DGP2.

Tables 1–3 present the simulation results of the QMLE based on 1000 repetitions. Both biases and root mean square errors (RMSE) are reported. From the results for both DGP1 and DGP2, I find that the biases are small and the RMSE decrease as the sample increases, indicating that the QMLE performs well and is consistent. Moreover, the simulation re-

34

sults corroborate the asymptotic properties of the QMLE derived in this paper. Additional simulation results based on different distributions of errors (normal and student's t distributions) are reported in the supplementary material, which confirm that QMLE has good finite sample properties and is robust to different distributions of errors.

Table 1: The performance of QMLE under DGP1 & "1 ahead and 1 behind" weights matrix

	N	25	50	100	25	50	100
	T	50	50	50	100	100	100
	% $\hat{r} = r$	99.8	100.0	100.0	99.8	100.0	100.0
ρ_1	Bias	0.0005	0.0002	0.0002	0.0003	0.0001	0.0002
	RMSE	0.0070	0.0043	0.0027	0.0048	0.0030	0.0020
ρ_2	Bias	0.0001	0.0002	0.0001	0.0002	0.0000	0.0002
	RMSE	0.0043	0.0026	0.0017	0.0030	0.0019	0.0012
γ_1	Bias	0.0000	0.0000	0.0001	0.0000	0.0001	0.0001
	RMSE	0.0035	0.0020	0.0013	0.0023	0.0014	0.0009
γ_2	Bias	-0.0001	0.0005	0.0004	0.0002	0.0003	0.0003
	RMSE	0.0057	0.0033	0.0023	0.0038	0.0025	0.0016
β_1	Bias	0.0002	0.0000	0.0001	0.0000	0.0001	-0.0001
	RMSE	0.0090	0.0059	0.0035	0.0065	0.0039	0.0026
β_2	Bias	-0.0001	-0.0001	-0.0001	0.0001	-0.0003	-0.0002
	RMSE	0.0101	0.0058	0.0037	0.0071	0.0041	0.0027

Table 2: The performance of QMLE under DGP1 & "5 ahead and 5 behind" weights matrix

	N	25	50	100	25	50	100
	T	50	50	50	100	100	100
% $\hat{r} = r$		99.8	100.0	100.0	99.8	100.0	100.0
ρ_1	Bias	-0.0001	-0.0001	0.0003	0.0002	0.0000	0.0000
	RMSE	0.0150	0.0076	0.0050	0.0088	0.0056	0.0036
ρ_2	Bias	0.0002	0.0003	0.0001	0.0001	-0.0002	0.0002
	RMSE	0.0083	0.0050	0.0033	0.0059	0.0036	0.0022
γ_1	Bias	-0.0001	0.0000	0.0001	-0.0001	0.0001	0.0001
	RMSE	0.0036	0.0020	0.0013	0.0023	0.0014	0.0009
γ_2	Bias	-0.0002	0.0003	0.0003	0.0001	0.0002	0.0001
	RMSE	0.0057	0.0033	0.0023	0.0039	0.0026	0.0016
β_1	Bias	0.0003	0.0000	0.0001	0.0000	0.0001	-0.0001
	RMSE	0.0090	0.0058	0.0034	0.0065	0.0039	0.0026
β_2	Bias	0.0000	-0.0001	0.0000	0.0002	-0.0002	-0.0001
	RMSE	0.0102	0.0058	0.0037	0.0071	0.0041	0.0026

Table 3: The performance of QMLE under DGP2 & "1 ahead and 1 behind" weights matrix

	N	25	50	100	25	50	100
	T	50	50	50	100	100	100
ρ_1	Bias	0.0015	-0.0002	-0.0001	-0.0007	0.0002	0.0003
	RMSE	0.0098	0.0058	0.0038	0.0060	0.0041	0.0026
ρ_2	Bias	-0.0011	0.0003	-0.0001	0.0003	0.0003	0.0001
	RMSE	0.0070	0.0043	0.0030	0.0047	0.0034	0.0020
γ_1	Bias	0.0006	0.0001	0.0003	0.0000	-0.0005	0.0001
	RMSE	0.0058	0.0035	0.0023	0.0044	0.0024	0.0016
γ_2	Bias	-0.0001	0.0008	0.0000	0.0005	0.0001	0.0002
	RMSE	0.0060	0.0053	0.0031	0.0056	0.0037	0.0024
β_1	Bias	-0.0010	-0.0024	-0.0005	-0.0027	-0.0013	-0.0009
	RMSE	0.0154	0.0118	0.0062	0.0103	0.0062	0.0046
β_2	Bias	-0.0019	-0.0017	-0.0009	0.0004	-0.0018	-0.0014
	RMSE	0.0167	0.0098	0.0062	0.0110	0.0076	0.0052

4.3 Finite sample performance of the IGPCE

In this section, I present the simulation results of the IGPC approach based on the above two DGPs and both setups of the spatial weights matrix.

Similarly to the QML approach, I propose an information criterion adapting the ideas in Bai and Ng (2002) to determine the number of factors r for the IGPC approach, specified as:

$$\check{r} = \underset{0 \leq m \leq r_{max}}{\mathrm{argmin}} \ IC_2(m) \tag{4.2}$$

with

$$IC_2(m) = \frac{1}{2N\bar{k}} \sum_{i=1}^{N} \left(\ln \left[(\check{\sigma}_{1i}^m)^2 \right] + \ln \left[(\check{\sigma}_{2i}^m)^2 \right] \right) - \frac{1}{N\bar{k}} \ln |\Upsilon(\check{\eta}^m)| + m \frac{N\bar{k} + T}{2N\bar{k}T} \ln[\min(N\bar{k}, T)]$$

and

$$(\check{\sigma}_{1i}^m)^2 = \frac{1}{T} \sum_{t=1}^{T} \left(\dot{y}_{1it} - \check{\rho}_1^m \ddot{y}_{1it} - \check{\gamma}_1^m \dot{y}_{2it} - \dot{x}_{1it}' \check{\beta}_1^m - \check{\lambda}_i^{m\prime} \check{f}_t^m \right)^2$$

$$(\check{\sigma}_{2i}^m)^2 = \frac{1}{T} \sum_{t=1}^{T} \left(\dot{y}_{2it} - \check{\rho}_2^m \ddot{y}_{2it} - \check{\gamma}_2^m \dot{y}_{1it} - \dot{x}_{2it}' \check{\beta}_2^m - \check{\psi}_i^{m\prime} \check{f}_t^m \right)^2$$

where $\bar{k} = 2$; $(\check{\eta}^m, \check{\beta}_1^m, \check{\beta}_2^m, \check{\lambda}_i^m, \check{f}_t^m)$ are the IGPCEs of $(\eta, \beta_1, \beta_2, \lambda_i, f_t)$ when the number of factors is set to m. Again, I set $r_{\max} = 4$ and report the percentage of r values correctly estimated by (4.2) based on 1000 repetitions for DGP1. From the results shown in the third row of Tables 4 and 5, it can be seen that the percentage is high for most choices of (N, T) (except small (N, T)) and setups of the weights matrix. Although the percentage is slightly lower than that in the QML approach, it is a reasonably good choice in practice when researchers prefer to allow the explanatory variables to be arbitrarily correlated with the factors and loadings. Therefore, I assume that r is known in the simulation of the IGPCE under both DGPs.

Tables 4–6 state the simulation results of the IGPCE based on 1000 repetitions. The results show that both the bias and the RMSE of the IGPCE are small in terms of the sample size and the magnitude of the true underlying parameters across different combinations of (N, T) and different choices of the weights matrix. In addition, the RMSE of the IGPCE declines as sample becomes larger, indicating that the IGPCE is consistent. Thus, the simulation results indicate that the IGPCE works well in a finite sample.

Table 4: The performance of IGPCE under DGP1 & "1 ahead and 1 behind" weights matrix

	N	25	50	100	25	50	100
	T	50	50	50	100	100	100
% $\hat{r} = r$		78.60%	88.40%	88.00%	84.40%	86.80%	90.00%
ρ_1	Bias	-0.0025	-0.0017	-0.0011	-0.0024	-0.0019	-0.0014
	RMSE	0.0118	0.0073	0.0049	0.0091	0.0071	0.0058
ρ_2	Bias	-0.0011	-0.0006	-0.0005	-0.0009	-0.0006	-0.0006
	RMSE	0.0071	0.0046	0.0032	0.0057	0.0035	0.0028
γ_1	Bias	-0.0012	-0.0005	-0.0005	-0.0012	-0.0009	-0.0007
	RMSE	0.0066	0.0035	0.0026	0.0070	0.0037	0.0030
γ_2	Bias	-0.0064	-0.0030	-0.0016	-0.0052	-0.0033	-0.0027
	RMSE	0.0251	0.0155	0.0067	0.0215	0.0176	0.0108
β_1	Bias	0.0002	0.0006	0.0003	0.0006	0.0003	0.0003
	RMSE	0.0121	0.0073	0.0043	0.0080	0.0054	0.0033
β_2	Bias	0.0035	0.0014	0.0006	0.0024	0.0017	0.0011
	RMSE	0.0158	0.0097	0.0051	0.0130	0.0096	0.0055

Table 5: The performance of IGPCE under DGP1 & "5 ahead and 5 behind" weights matrix

	N	25	50	100	25	50	100
	T	50	50	50	100	100	100
% $\hat{r} = r$		77.40%	87.80%	86.00%	85.00%	88.40%	88.00%
ρ_1	Bias	-0.0021	-0.0012	-0.0010	-0.0023	-0.0014	-0.0007
	RMSE	0.0186	0.0110	0.0073	0.0142	0.0084	0.0053
ρ_2	Bias	-0.0010	-0.0005	-0.0002	-0.0006	-0.0003	-0.0004
	RMSE	0.0123	0.0073	0.0047	0.0092	0.0051	0.0033
γ_1	Bias	-0.0016	-0.0007	-0.0006	-0.0012	-0.0010	-0.0007
	RMSE	0.0099	0.0038	0.0030	0.0058	0.0039	0.0026
γ_2	Bias	-0.0061	-0.0027	-0.0017	-0.0059	-0.0030	-0.0024
	RMSE	0.0236	0.0133	0.0074	0.0238	0.0137	0.0091
β_1	Bias	0.0001	0.0004	0.0001	0.0004	0.0002	0.0002
	RMSE	0.0122	0.0074	0.0042	0.0080	0.0052	0.0031
β_2	Bias	0.0032	0.0012	0.0005	0.0024	0.0013	0.0010
	RMSE	0.0153	0.0090	0.0049	0.0146	0.0077	0.0048

Table 6: The performance of IGPCE under DGP2 & "1 ahead and 1 behind" weights matrix

	N	25	50	100	25	50	100
	T	50	50	50	100	100	100
ρ_1	Bias	-0.0025	-0.0009	-0.0018	-0.0011	-0.0027	-0.0017
	RMSE	0.0162	0.0126	0.0096	0.0106	0.0095	0.0068
ρ_2	Bias	-0.0012	-0.0014	-0.0001	-0.0006	-0.0008	-0.0003
	RMSE	0.0109	0.0086	0.0061	0.0099	0.0061	0.0043
γ_1	Bias	-0.0020	-0.0015	-0.0021	-0.0021	-0.0007	-0.0009
	RMSE	0.0096	0.0078	0.0055	0.0075	0.0046	0.0041
γ_2	Bias	-0.0065	-0.0046	-0.0038	-0.0036	-0.0014	-0.0026
	RMSE	0.0280	0.0150	0.0119	0.0231	0.0085	0.0101
β_1	Bias	0.0011	0.0028	0.0047	0.0022	0.0012	0.0012
	RMSE	0.0254	0.0173	0.0123	0.0151	0.0102	0.0076
β_2	Bias	0.0021	0.0022	0.0028	0.0014	0.0018	0.0002
	RMSE	0.0251	0.0166	0.0138	0.0185	0.0125	0.0102

4.4 Comparison of the performance of both approaches

A comparison of both approaches reveals that: (1), when the explanatory variables x are correctly specified, QML performs better than IGPC, otherwise the superiority of QML is weakened; (2), IGPC is robust for different underlying models of x.

Based on DGP1 when the model of explanatory variables is correctly specified in the QML approach, I find that the bias of IGPCE is relatively obvious compared to QMLE whose bias is close to zero. This finding is consistent with the inferential theory that IGPCE has a bias term in its limiting distribution while QMLE does not. At the same time, the RMSE of IGPCE is slightly larger than that of QMLE, implying that QMLE is more efficient than IGPCE.

Based on DGP2 when the explanatory variables do not satisfy the specification as in the QML approach but are still affected by the common shocks, QML performs slightly worse than in DGP1, while the performance of IGPC is similar to that in DGP1. This implies that IGPCE is robust to different underlying specifications of the explanatory variables, while QMLE is more sensitive. In addition, the superiority of QMLE is weakened when the explanatory variables are not correctly specified.

5 Some extensions

In this section, I discuss four important and useful extensions of model (1.1) with a brief summary as follows: (1) models with additional explanatory variables, denoted x_{3it}, which affect both dependent variables y_1 and y_2; (2) models with time-invariant and common regressors; (3) models with spatial autoregressive (SAR) errors; (4) models with additional

spatial lags: the dependent variable y_1 is affected not only by its own spatial lag but also by the spatial lag of y_2, and vice versa.[17]

5.1 Models with additional common explanatory variables

In model (1.1), I consider two different sets of explanatory variables x_{1it} and x_{2it} in the y_{1it} and y_{2it} equations, respectively. In this section, I augment model (1.1) with additional common explanatory variables, denoted x_{3it}, which affect both y_{1it} and y_{2it}, as follows:

$$
\begin{aligned}
y_{1it} &= \alpha_{1i} + \rho_1 \sum_{j=1}^{N} w_{1ij} y_{1jt} + \gamma_1 y_{2it} + \sum_{p=1}^{k_1} x_{1itp} \beta_{1p} + \sum_{l=1}^{k_3} x_{3itl} \beta_{3l} + \lambda_i' f_t + e_{1it} \\
y_{2it} &= \alpha_{2i} + \rho_2 \sum_{j=1}^{N} w_{2ij} y_{2jt} + \gamma_2 y_{1it} + \sum_{q=1}^{k_2} x_{2itq} \beta_{2q} + \sum_{l=1}^{k_3} x_{3itl} \beta_{4l} + \psi_i' f_t + e_{2it}
\end{aligned}
\tag{5.1}
$$

where $x_{3itl}, (l = 1, 2, \ldots, k_3)$ are additional explanatory variables. I then propose both QML and IGPC approaches for the above extension.

5.1.1 Extension 1 using QML approach

For model (5.1), I assume the additional explanatory variable x_{3it} is also affected by the common factor f_t and follows the same factor structure model as (2.1):

$$
x_{3itl} = \nu_{3il} + \phi_{3il}' f_t + v_{3itl}, \quad l = 1, 2, \ldots, k_3
\tag{5.2}
$$

Then, the extended model combining (5.1), (2.1) and (5.2) can be rewritten as:

$$
\begin{bmatrix}
y_{1it} - \rho_1 \sum_{j=1}^{N} w_{1ij} y_{1jt} - \gamma_1 y_{2it} - x_{1it}' \beta_1 - x_{3it}' \beta_3 \\
y_{2it} - \rho_2 \sum_{j=1}^{N} w_{2ij} y_{2jt} - \gamma_2 y_{1it} - x_{2it}' \beta_2 - x_{3it}' \beta_4 \\
x_{1it} \\
x_{2it} \\
x_{3it}
\end{bmatrix}
= \mu_i^\dagger + L_i^{\dagger\prime} f_t + \epsilon_{it}^\dagger
$$

where $x_{3it} = (x_{3it1}, x_{3it2}, \ldots, x_{3it,k_3})'$; $\beta_3 = (\beta_{31}, \beta_{32}, \ldots, \beta_{3,k_3})'$, which is similar for β_4 and v_{3it}; $\mu_i^\dagger = (\alpha_{1i}, \alpha_{2i}, \nu_{1i}', \nu_{2i}', \nu_{3i}')'$; $L_i^\dagger = (\lambda_i, \psi_i, \phi_{1i}, \phi_{2i}, \phi_{3i})$; and $\epsilon_{it}^\dagger = (e_{1it}, e_{2it}, v_{1it}', v_{2it}', v_{3it}')'$. Let $\delta^\dagger = (\rho_1, \rho_2, \gamma_1, \gamma_2, \beta_1', \beta_2', \beta_3', \beta_4')$, $k^\dagger = k_1 + k_2 + k_3$ and $\bar{k}^\dagger = k^\dagger + 2$. I can then rewrite the above model into the same framework used in (2.4):

$$
D^\dagger(\delta^\dagger) z_t^\dagger = \mu^\dagger + L^\dagger f_t + \epsilon_t^\dagger
\tag{5.3}
$$

---

[17]For each extension, I discuss both QML and IGPC estimation methods with modification. The large sample theory can be derived, but is much more involved.

where $z_t^{\dagger} = (z_{1t}^{\dagger}, z_{2t}^{\dagger}, \ldots, z_{Nt}^{\dagger})'$ with $z_{it}^{\dagger} = (y_{1it}, y_{2it}, x_{1it}', x_{2it}', x_{3it}')'$, $L^{\dagger} = (L_1^{\dagger}, L_2^{\dagger}, \ldots, L_N^{\dagger})'$, $\mu^{\dagger} = (\mu_1^{\dagger\prime}, \mu_2^{\dagger\prime}, \ldots, \mu_N^{\dagger\prime})'$ and $\epsilon_t^{\dagger} = (\epsilon_{1t}^{\dagger\prime}, \epsilon_{2t}^{\dagger\prime}, \ldots, \epsilon_{Nt}^{\dagger\prime})'$. The new transformation matrix $D^{\dagger}(\delta^{\dagger})$ is an $N\bar{k}^{\dagger} \times N\bar{k}^{\dagger}$ matrix whose (i,j) subblock, denoted by $D_{ij}^{\dagger}(\delta^{\dagger})$, a $\bar{k}^{\dagger} \times \bar{k}^{\dagger}$ matrix, equals:

$$
D_{ij}^{\dagger}(\delta^{\dagger}) = \begin{cases} \begin{bmatrix} 1 & -\gamma_1 & -\beta_1' & 0 & -\beta_3' \\ -\gamma_2 & 1 & 0 & -\beta_2' & -\beta_4' \\ 0 & 0 & I_{k_1} & 0 & 0 \\ 0 & 0 & 0 & I_{k_2} & 0 \\ 0 & 0 & 0 & 0 & I_{k_3} \end{bmatrix} & \text{if } i = j \\[2em] \begin{bmatrix} -\rho_1 w_{1ij} & 0 & 0 & 0 & 0 \\ 0 & -\rho_2 w_{2ij} & 0 & 0 & 0 \\ 0 & 0 & 0_{k_1} & 0 & 0 \\ 0 & 0 & 0 & 0_{k_2} & 0 \\ 0 & 0 & 0 & 0 & 0_{k_3} \end{bmatrix} & \text{if } i \neq j \end{cases}
\tag{5.4}
$$

Similarly, I propose the QML method to estimate this extended model. In order to derive the inferential theory of its QMLE, the key is to study the determinant and the inverse matrix of the transformation matrix $D^{\dagger}(\delta^{\dagger})$. Then, a similar analytical approach can be used in the analysis of the QMLE. Let η and $\Upsilon(\eta)$ be defined as in Section 2.2. Then, it can be verified that $\det(D^{\dagger}(\delta^{\dagger})) = \det(\Upsilon(\eta))$. Let $V^{\dagger}(\delta^{\dagger})$ denote the inverse matrix of $D^{\dagger}(\delta^{\dagger})$, which is an $N\bar{k}^{\dagger} \times N\bar{k}^{\dagger}$ matrix. Its (i,j)th subblock is denoted by $V_{ij}^{\dagger}$, a $\bar{k}^{\dagger} \times \bar{k}^{\dagger}$ matrix, and has the same expression as in Lemma A.2 of the supplementary material, with the same definition of F_{ij} but a different $\beta^{\dagger}$. Here, $\beta^{\dagger\prime} = \begin{bmatrix} \beta_1' & 0 & \beta_3' \\ 0 & \beta_2' & \beta_4' \end{bmatrix}$. Lemma A.3 of the supplementary material still holds but with the preceding definition of $\beta^{\dagger}$. Based on the preceding analysis, the inferential analysis for this extended model can be studied in a similar way as that for model (2.2).

5.1.2 Extension 1 using IGPC approach

In this approach, I consider model (5.1) without specifying the model for all explanatory variables and propose the IGPC method.

Let $x_{it}^{\dagger} = \begin{bmatrix} x_{1it}' & 0 & x_{3it}' & 0 \\ 0 & x_{2it}' & 0 & x_{3it}' \end{bmatrix}'$, and $\beta^{\dagger} = (\beta_1', \beta_2', \beta_3', \beta_4')'$. Then, model (5.1) can be rewritten as:

$$
\begin{bmatrix} y_{1it} - \rho_1 \sum_{j=1}^{N} w_{1ij} y_{1jt} - \gamma_1 y_{2it} \\ y_{2it} - \rho_2 \sum_{j=1}^{N} w_{2ij} y_{2jt} - \gamma_2 y_{1it} \end{bmatrix} = \alpha_i + x_{it}^{\dagger\prime} \beta^{\dagger} + \Gamma_i' f_t + e_{it}
\tag{5.5}
$$

where all α_i, Γ_i and e_{it} are defined as in Section 3.1. With the same η and $\Upsilon(\eta)$, I can rewrite

41

the above model into the form of (3.2) with the new $x_{it}^{\dagger}$ and $\beta^{\dagger}$. Then, I propose the same IGPC method for this extended model.

5.2 Models with time-invariant and common regressors

In applications, it is common to observe some time-invariant regressors (i.e., not varying with t, such as gender, race and education in microeconomic earnings studies) and some common regressors (i.e., not varying with individual i, such as unemployment rates, aggregate price index representing trends and other macroeconomic policy variables). Therefore, in this section, I extend model (1.1) to include some time-invariant and common regressors using both QML and IGPC approaches as follows.

5.2.1 Extension 2 using QML approach

In this approach, I allow the regression coefficients of the time-invariant regressors to be time varying and the coefficients of the common regressors to be individual dependent (varying with i). In addition, I allow both x_{1it} and x_{2it} to be affected by the time-invariant regressors r_i and the common regressors p_t using a factor structure specification. Specifically, I consider the following extended model:

$$
y_{1it} = \alpha_{1i} + \rho_1 \sum_{j=1}^{N} w_{1ij} y_{1jt} + \gamma_1 y_{2it} + \sum_{p=1}^{k_1} x_{1itp} \beta_{1p} + r_i' h_{1t} + \tau_{1i}' p_t + \lambda_i' f_t + e_{1it}
$$

$$
y_{2it} = \alpha_{2i} + \rho_2 \sum_{j=1}^{N} w_{2ij} y_{2jt} + \gamma_2 y_{1it} + \sum_{q=1}^{k_2} x_{2itq} \beta_{2q} + r_i' h_{2t} + \tau_{2i}' p_t + \psi_i' f_t + e_{2it} \qquad (5.6)
$$

$$
x_{1itp} = \nu_{1ip} + r_i' s_{1tp} + \eta_{1ip}' p_t + \phi_{1ip}' f_t + v_{1itp}, \quad p = 1, 2, \ldots, k_1
$$

$$
x_{2itq} = \nu_{2iq} + r_i' s_{2tq} + \eta_{2iq}' p_t + \phi_{2iq}' f_t + v_{2itq}, \quad q = 1, 2, \ldots, k_2
$$

where r_i represents a vector of observable time-invariant variables, and p_t represents a vector of observable common variables.

The above model can be rewritten as follows:

$$
\begin{bmatrix} y_{1it} - \rho_1 \sum_{j=1}^{N} w_{1ij} y_{1jt} - \gamma_1 y_{2it} - x_{1it}' \beta_1 \\ y_{2it} - \rho_2 \sum_{j=1}^{N} w_{2ij} y_{2jt} - \gamma_2 y_{1it} - x_{2it}' \beta_2 \\ x_{1it} \\ x_{2it} \end{bmatrix} = \mu_i + L_i^{\dagger\prime} f_t^{\dagger} + \epsilon_{it}
$$

where $f_t^\dagger = (h_{1t}', h_{2t}', p_t', s_{1t1}', \ldots, s_{1t,k_1}', s_{2t1}', \ldots, s_{2t,k_2}', f_t')'$, and

$$
L_i^{\dagger\prime} = \begin{bmatrix} r_i' & 0 & \tau_{1i}' & 0 & 0 & \lambda_i' \\ 0 & r_i' & \tau_{2i}' & 0 & 0 & \psi_i' \\ 0 & 0 & \eta_{1i}' & I_{k_1} \otimes r_i' & 0 & \phi_{1i}' \\ 0 & 0 & \eta_{2i}' & 0 & I_{k_2} \otimes r_i' & \phi_{2i}' \end{bmatrix}
$$

with $\eta_{1i} = (\eta_{1i1}, \ldots, \eta_{1i,k_1})$ and $\eta_{2i} = (\eta_{2i1}, \ldots, \eta_{2i,k_2})$.

The above model specification is similar to that in Section 2, with the difference that some components of the common factors $f_t^\dagger$ and some components of the factor loadings $L_i^\dagger$ are now observable. The QML method can still be implemented for this extension but with modifications (for the observable components of $f_t^\dagger$ and $L_i^\dagger$, the QMLE does not estimate them but fixes them at the observed value). The asymptotic properties of the QMLE can be analyzed in a similar way as for the basic model (2.2), with attention to the fact that some components of $f_t^\dagger$ and $L_i^\dagger$ are observable.

5.2.2 Extension 2 using IGPC approach

In this section, I do not specify a model for time-invariant or common regressors but allow them to be arbitrarily correlated with the common shocks. In addition, I treat them as explanatory variables with constant coefficients and specify the extended model as follows:

$$
\begin{aligned}
y_{1it} &= \alpha_{1i} + \rho_1 \sum_{j=1}^{N} w_{1ij} y_{1jt} + \gamma_1 y_{2it} + x_{1it}' \beta_1 + r_i' \beta_3 + p_t' \beta_5 + \lambda_i' f_t + e_{1it} \\
y_{2it} &= \alpha_{2i} + \rho_2 \sum_{j=1}^{N} w_{2ij} y_{2jt} + \gamma_2 y_{1it} + x_{2it}' \beta_2 + r_i' \beta_4 + p_t' \beta_6 + \psi_i' f_t + e_{2it}
\end{aligned}
\tag{5.7}
$$

where r_i is a vector of observable time-invariant variables, and p_t is a vector of observable common variables. The above model can be rewritten in the same form as (5.5) with different definitions of $x_{it}^\dagger$ and $\beta^\dagger$: $x_{it}^\dagger = \begin{bmatrix} x_{1it}' & 0 & r_i' & 0 & p_t' & 0 \\ 0 & x_{2it}' & 0 & r_i' & 0 & p_t' \end{bmatrix}$ and $\beta^\dagger = (\beta_1', \beta_2', \ldots, \beta_6')'$. Again, I propose the IGPC method for this extended model, and the corresponding inferential theory can be studied in a similar way as in Section 3.

5.3 Models with SAR disturbances

In the spatial econometric literature, SAR disturbances have received much attention and are considered an important part of spatial models. Based on model (1.1), which only considers spatial correlations in the dependent variables, I now develop a more general model by including additional spatial correlations on the errors (i.e., SAR errors) in the following

model specification:

$$
y_{1it} = \alpha_{1i} + \rho_1 \sum_{j=1}^{N} w_{1ij} y_{1jt} + \gamma_1 y_{2it} + x'_{1it}\beta_1 + \lambda'_i f_t + u_{1it}, \quad u_{1it} = \pi_1 \sum_{j=1}^{N} m_{1ij} u_{1jt} + e_{1it}
$$

$$
y_{2it} = \alpha_{2i} + \rho_2 \sum_{j=1}^{N} w_{2ij} y_{2jt} + \gamma_2 y_{1it} + x'_{2it}\beta_2 + \psi'_i f_t + u_{2it}, \quad u_{2it} = \pi_2 \sum_{j=1}^{N} m_{2ij} u_{2jt} + e_{2it}
$$

$$(5.8)$$

where m_{1ij} and m_{2ij} are spatial weights involved in the SAR disturbances. The above model can be rewritten as:

$$
Y_{1t} = \alpha_1 + \rho_1 W_1 Y_{1t} + \gamma_1 Y_{2t} + X_{1t}\beta_1 + \Lambda' f_t + U_{1t}, \quad U_{1t} = \pi_1 M_1 U_{1t} + e_{1t}
$$

$$
Y_{2t} = \alpha_2 + \rho_2 W_2 Y_{2t} + \gamma_2 Y_{1t} + X_{2t}\beta_2 + \Psi' f_t + U_{2t}, \quad U_{2t} = \pi_2 M_2 U_{2t} + e_{2t}
$$

$$(5.9)$$

where Y_{1t} is an $N \times 1$ vectors, defined as $Y_{1t} = (y_{11t}, y_{12t}, \ldots, y_{1Nt})'$, which is similar for $Y_{2t}, \alpha_1, \alpha_2, U_{1t}, U_{2t}, e_{1t}$ and e_{2t}; $X_{1t} = (x_{11t}, x_{12t}, \ldots, x_{1Nt})'$ is $N \times k_1$; $X_{2t} = (x_{21t}, x_{22t}, \ldots, x_{2Nt})'$ is $N \times k_2$; $\Lambda = (\lambda_1, \lambda_2, \ldots, \lambda_N)'$; and $\Psi = (\psi_1, \psi_2, \ldots, \psi_N)'$. Both W_1 and W_2 are $N \times N$ weights matrices associated with the spatial effects of the dependent variables, while M_1 and M_2 are $N \times N$ weights matrices representing the additional spatial correlations in the errors. Baltagi and Deng (2015) consider the above model (5.9) without the common shocks parts Λf_t and Ψf_t. Furthermore, they impose an error component specification instead of the SAR structure and assume cross-sectional homoskedasticity of the errors, while I allow cross-sectional heteroskedasticity here.[18]

To transform (5.9) into the framework of (1.1), premultiply $I_N - \pi_1 M_1$ on both sides of the Y_{1t} equation and premultiply $I_N - \pi_2 M_2$ on both sides of the Y_{2t} equation. Then, I obtain:

$$
\begin{aligned}
Y_{1t} =& (\alpha_1 - \pi_1 M_1 \alpha_1) + \rho_1 W_1 Y_{1t} + \pi_1 M_1 Y_{1t} - \rho_1 \pi_1 M_1 W_1 Y_{1t} + \gamma_1 Y_{2t} - \gamma_1 \pi_1 M_1 Y_{2t} \\
& + X_{1t}\beta_1 - \pi_1 M_1 X_{1t}\beta_1 + (\Lambda - \pi_1 M_1 \Lambda) f_t + e_{1t} \\
Y_{2t} =& (\alpha_2 - \pi_2 M_2 \alpha_2) + \rho_2 W_2 Y_{2t} + \pi_2 M_2 Y_{2t} - \rho_2 \pi_2 M_2 W_2 Y_{2t} + \gamma_2 Y_{1t} - \gamma_2 \pi_2 M_2 Y_{1t} \\
& + X_{2t}\beta_2 - \pi_2 M_2 X_{2t}\beta_2 + (\Psi - \pi_2 M_2 \Psi) f_t + e_{2t}
\end{aligned}
$$

$$(5.10)$$

Note that $(\alpha_1 - \pi_1 M_1 \alpha_1)$ is a free parameter, so I can treat it as a new α_1, Similarly, treat $(\Lambda - \pi_1 M_1 \Lambda)$ as a new Λ, $(\alpha_2 - \pi_2 M_2 \alpha_2)$ as a new α_2 and $(\Psi - \pi_2 M_2 \Psi)$ as a new Ψ. Then,

---

[18] The error component specification in Baltagi and Deng (2015) is described as $U_{lt} = \varphi_l + \varepsilon_{lt}$, where $\varphi_l \sim i.i.d.(0, \sigma^2_{\varphi ll} I_N)$ and $\varepsilon_{lt} \sim i.i.d.(0, \sigma^2_{\varepsilon ll} I_N)$, for $l = 1, 2$.

([5.10](#)) can be rewritten as:

$$
\begin{aligned}
Y_{1t} =&\, \alpha_1 + \rho_1 W_1 Y_{1t} + \pi_1 M_1 Y_{1t} - \rho_1 \pi_1 M_1 W_1 Y_{1t} + \gamma_1 Y_{2t} - \gamma_1 \pi_1 M_1 Y_{2t} \\
&+ X_{1t} \beta_1 - \pi_1 M_1 X_{1t} \beta_1 + \Lambda f_t + e_{1t} \\
Y_{2t} =&\, \alpha_2 + \rho_2 W_2 Y_{2t} + \pi_2 M_2 Y_{2t} - \rho_2 \pi_2 M_2 W_2 Y_{2t} + \gamma_2 Y_{1t} - \gamma_2 \pi_2 M_2 Y_{1t} \\
&+ X_{2t} \beta_2 - \pi_2 M_2 X_{2t} \beta_2 + \Psi f_t + e_{2t}
\end{aligned}
\tag{5.11}
$$

This can be further rewritten as:

$$
\begin{aligned}
y_{1it} =&\, \alpha_{1i} + \rho_1 \Big(\sum_{j=1}^{N} w_{1ij} y_{1jt} \Big) + \pi_1 \Big(\sum_{j=1}^{N} m_{1ij} y_{1jt} \Big) - \rho_1 \pi_1 \Big(\sum_{j=1}^{N} \sum_{l=1}^{N} m_{1ij} w_{1jl} y_{1lt} \Big) \\
&+ \gamma_1 y_{2it} - \gamma_1 \pi_1 \Big(\sum_{j=1}^{N} m_{1ij} y_{2jt} \Big) + x_{1it}' \beta_1 - \pi_1 \Big(\sum_{j=1}^{N} m_{1ij} x_{1jt}' \Big) \beta_1 + \lambda_i' f_t + e_{1it} \\
y_{2it} =&\, \alpha_{2i} + \rho_2 \Big(\sum_{j=1}^{N} w_{2ij} y_{2jt} \Big) + \pi_2 \Big(\sum_{j=1}^{N} m_{2ij} y_{2jt} \Big) - \rho_2 \pi_2 \Big(\sum_{j=1}^{N} \sum_{l=1}^{N} m_{2ij} w_{2jl} y_{2lt} \Big) \\
&+ \gamma_2 y_{1it} - \gamma_2 \pi_2 \Big(\sum_{j=1}^{N} m_{2ij} y_{1jt} \Big) + x_{2it}' \beta_2 - \pi_2 \Big(\sum_{j=1}^{N} m_{2ij} x_{2jt}' \Big) \beta_2 + \Psi_i' f_t + e_{2it}
\end{aligned}
\tag{5.12}
$$

Then, I analyze the above model using the following two approaches.

5.3.1 Extension 3 using QML approach

I assume that the explanatory variables follow the same model as ([2.1](#)). Then, combining ([5.12](#)) and ([2.1](#)), I can rewrite this extended model in the same framework as ([2.4](#)):

$$
D^{\dagger}(\delta^{\dagger}) z_t = \mu + L f_t + \epsilon_t
\tag{5.13}
$$

where $\delta^{\dagger} = (\rho_1, \rho_2, \gamma_1, \gamma_2, \beta_1', \beta_2', \pi_1, \pi_2)'$; z_t, μ, L and ϵ_t are defined in the same way as in Section 2; and $D^{\dagger}(\delta^{\dagger})$ is an $N\bar{k} \times N\bar{k}$ transformation matrix whose (i,j)th subblock, denoted

by $D_{ij}^{\dagger}(\delta^{\dagger})$, is defined as:

$$
D_{ij}^{\dagger}(\delta^{\dagger}) = \begin{cases} \begin{bmatrix} 1 + \rho_1\pi_1 m_{1i*}w_{1*i} & -\gamma_1 & -\beta_1' & 0 \\ -\gamma_2 & 1 + \rho_2\pi_2 m_{2i*}w_{2*i} & 0 & -\beta_2' \\ 0 & 0 & I_{k_1} & 0 \\ 0 & 0 & 0 & I_{k_2} \end{bmatrix} & \text{if } i = j \\[3em] \begin{bmatrix} \boldsymbol{d}_{1,ij} & \gamma_1\pi_1 m_{1ij} & \pi_1 m_{1ij}\beta_1' & 0 \\ \gamma_2\pi_2 m_{2ij} & \boldsymbol{d}_{2,ij} & 0 & \pi_2 m_{2ij}\beta_2' \\ 0 & 0 & 0_{k_1} & 0 \\ 0 & 0 & 0 & 0_{k_2} \end{bmatrix} & \text{if } i \neq j \end{cases} \tag{5.14}
$$

where $\boldsymbol{d}_{1,ij} = -\rho_1 w_{1ij} - \pi_1 m_{1ij} + \rho_1\pi_1 m_{1i*}w_{1*j}$ and $\boldsymbol{d}_{2,ij} = -\rho_2 w_{2ij} - \pi_2 m_{2ij} + \rho_2\pi_2 m_{2i*}w_{2*j}$, and m_{li*} is the ith row of matrix M_l and w_{l*j} is the jth column of matrix W_l, for $l = 1, 2$.

Note that model (5.13) is similar to model (2.4) but with a new $\delta^{\dagger}$ (including additional parameters π_1, π_2 due to the SAR errors) and a more complicated transformation matrix $D^{\dagger}(\delta^{\dagger})$. The QML method can be easily implemented in this extended model. To develop the inferential theory, similar to the derivation for model (2.2), the key is to specify the determinant and the inverse matrix of $D^{\dagger}(\delta^{\dagger})$. Let $\eta^{\dagger} = (\rho_1, \rho_2, \gamma_1, \gamma_2, \pi_1, \pi_2)'$ and $\Upsilon^{\dagger}(\eta^{\dagger})$ be a $2N \times 2N$ matrix whose (i, j)th subblock, denoted by $\Upsilon_{ij}^{\dagger}(\eta^{\dagger})$, is a 2×2 matrix that equals:

$$
\Upsilon_{ij}^{\dagger}(\eta^{\dagger}) = \begin{cases} \begin{bmatrix} 1 + \rho_1\pi_1 m_{1i*}w_{1*i} & -\gamma_1 \\ -\gamma_2 & 1 + \rho_2\pi_2 m_{2i*}w_{2*i} \end{bmatrix} & \text{if } i = j \\[2em] \begin{bmatrix} \boldsymbol{d}_{1,ij} & \gamma_1\pi_1 m_{1ij} \\ \gamma_2\pi_2 m_{2ij} & \boldsymbol{d}_{2,ij} \end{bmatrix} & \text{if } i \neq j \end{cases} \tag{5.15}
$$

where $\boldsymbol{d}_{1,ij}$ and $\boldsymbol{d}_{2,ij}$ are the same as in the definition of $D_{ij}^{\dagger}(\delta^{\dagger})$. Then, it can be verified that $\det(D^{\dagger}(\delta^{\dagger})) = \det(\Upsilon^{\dagger}(\eta^{\dagger}))$. Furthermore, let $V^{\dagger}(\delta^{\dagger})$ be the inverse matrix of $D^{\dagger}(\delta^{\dagger})$. Then, its (i, j)th block, a $(k + 2) \times (k + 2)$ matrix $(k = k_1 + k_2)$ denoted by $V_{ij}^{\dagger}(\delta)$ has a closed form, which is equal to:

$$
V_{ij}^{\dagger}(\delta^{\dagger}) = \begin{cases} \begin{bmatrix} F_{ii}^{\dagger} & F_{ii}^{\dagger}\beta' \\ 0 & I_k \end{bmatrix} & \text{if } i = j \\[2em] \begin{bmatrix} F_{ij}^{\dagger} & F_{ij}^{\dagger}\beta' \\ 0 & 0_{k \times k} \end{bmatrix} & \text{if } i \neq j \end{cases} \tag{5.16}
$$

where $\beta = \begin{bmatrix} \beta_1 & 0 \\ 0 & \beta_2 \end{bmatrix}$, and $F_{ij}^{\dagger}$ is the (i, j)th 2×2 block of the inverse matrix of $\Upsilon^{\dagger}(\eta)$. The

QML method can be used to estimate model (5.13), and the inferential theory can be studied similarly to basic model (2.2), together with the preceding results for the determinant and inverse of $D^\dagger(\eta^\dagger)$.

5.3.2 Extension 3 using IGPC approach

In this approach, I do not specify the model for the explanatory variables.

Using the same notation as in Section 5.3.1, let $\Upsilon^X(\delta^\dagger)$ be a $2N \times 2N$ matrix, with its (i,j)th subblock denoted as $\Upsilon_{ij}^X(\delta^\dagger)$, a 2×2 matrix that equals:

$$
\Upsilon_{ij}^X(\eta^\dagger) = \begin{cases} \begin{bmatrix} \beta_1' & 0 \\ 0 & \beta_2' \end{bmatrix} & \text{if } i = j \\[2em] \begin{bmatrix} -\pi_1 m_{1ij}\beta_1' & 0 \\ 0 & -\pi_2 m_{2ij}\beta_2' \end{bmatrix} & \text{if } i \neq j \end{cases} \tag{5.17}
$$

Then, the extended model (5.12) can be rewritten similarly to (3.2):

$$
\sum_{j=1}^N \Upsilon_{ij}^\dagger(\eta^\dagger) y_{jt} = \alpha_i + \sum_{j=1}^N \Upsilon_{ij}^X(\delta^\dagger) x_{jt} + \Gamma_i' f_t + e_{it} \tag{5.18}
$$

where $y_{jt}, x_{jt}, \alpha_i, \Gamma_i$ and e_{it} are defined as in (3.2); $\Upsilon_{ij}^\dagger(\eta^\dagger)$ is defined in Section 5.3.1. Then, based on the above expression, the IGPC method can be applied in this extension.

5.4 Models with additional spatial lags

In model (1.1), y_{1it} is affected by its own spatial lag only; likewise for y_{2it}. In this section, I enrich model (1.1) with additional spatial lags in both y_{1it} and y_{2it} equations as follows:

$$
\begin{aligned} y_{1it} &= \alpha_{1i} + \rho_1 \sum_{j=1}^N w_{1ij} y_{1jt} + \gamma_1 y_{2it} + \rho_3 \sum_{j=1}^N w_{3ij} y_{2jt} + x_{1it}' \beta_1 + \lambda_i' f_t + e_{1it} \\ y_{2it} &= \alpha_{2i} + \rho_2 \sum_{j=1}^N w_{2ij} y_{2jt} + \gamma_2 y_{1it} + \rho_4 \sum_{j=1}^N w_{4ij} y_{1jt} + x_{2it}' \beta_2 + \psi_i' f_t + e_{2it} \end{aligned} \tag{5.19}
$$

where y_{1it} is affected by the spatial lag of y_{2it}, and y_{2it} is affected by the spatial lag of y_{1it}, with ρ_3, ρ_4 being the additional parameters measuring the magnitudes of the spatial effects, and $W_3 = (w_{3ij})_{N \times N}$ and $W_4 = (w_{4ij})_{N \times N}$ being additional weights matrices. I analyze the above model using both the QML and IGPC approaches as follows.

5.4.1 Extension 4 using QML approach

I assume that the explanatory variables follow (2.1). Then, the extended model combining (5.19) and (2.1) can be rewritten as:

$$D^\dagger(\delta)z_t = \mu + Lf_t + \epsilon_t \tag{5.20}$$

in the same framework as (2.4) but with a different transformation matrix $D^\dagger(\delta)$ whose (i,j)th subblock is defined below:

$$D_{ij}^\dagger(\delta) = \begin{cases} \begin{bmatrix} 1 & -\gamma_1 & -\beta_1' & 0 \\ -\gamma_2 & 1 & 0 & -\beta_2' \\ 0 & 0 & I_{k_1} & 0 \\ 0 & 0 & 0 & I_{k_2} \end{bmatrix} & \text{if } i = j \\[40pt] \begin{bmatrix} -\rho_1 w_{1ij} & -\rho_3 w_{3ij} & 0 & 0 \\ -\rho_4 w_{4ij} & -\rho_2 w_{2ij} & 0 & 0 \\ 0 & 0 & 0_{k_1} & 0 \\ 0 & 0 & 0 & 0_{k_2} \end{bmatrix} & \text{if } i \neq j \end{cases} \tag{5.21}$$

Again, I implement the QML method for this extended model, and the inferential analysis of the QMLE can be derived in a similar way as for (2.2). In order to derive the inferential theory of the QMLE, the key is to study the determinant and the inverse matrix of the transformation matrix $D^\dagger(\delta)$. Unlike (2.7), I define a new $2N \times 2N$ matrix $\Upsilon^\dagger(\eta)$ whose (i,j)th block, a 2×2 matrix, is equal to:

$$\Upsilon_{ij}^\dagger(\eta) = \begin{cases} \begin{bmatrix} 1 & -\gamma_1 \\ -\gamma_2 & 1 \end{bmatrix} & \text{if } i = j \\[20pt] \begin{bmatrix} -\rho_1 w_{1ij} & -\rho_3 w_{3ij} \\ -\rho_4 w_{4ij} & -\rho_2 w_{2ij} \end{bmatrix} & \text{if } i \neq j \end{cases} \tag{5.22}$$

A preliminary step needed to conduct inferential analysis is to study the determinant and the inverse of the new transformation matrix $\Upsilon_{ij}^\dagger(\eta)$. With mathematical calculation, it can be verified that $\det(D^\dagger(\delta)) = \det(\Upsilon^\dagger(\eta))$. Let $V^\dagger(\delta)$ denote the inverse matrix of $D^\dagger(\delta)$, which is an $N\bar{k} \times N\bar{k}$ matrix. Its (i,j)th subblock, denoted by $V_{ij}^\dagger$, a $\bar{k} \times \bar{k}$ matrix, has the same expression as in Lemma A.2 of the supplementary material but with a different $F_{ij}^\dagger$, which now is the (i,j)th 2×2 block of the inverse matrix of $\Upsilon^\dagger(\eta)$.[19]

---

[19]In this extended model, Lemma A.3 needs modification due to the new $D^\dagger(\delta)$ and $\Upsilon^\dagger(\eta)$.

5.4.2 Extension 4 using IGPC approach

In this section, I apply the IGPC method to the extended model (5.19) without specifying the model for the explanatory variables. With the notation $\Upsilon^\dagger(\eta)$ defined in Section 5.4.1, I can rewrite (5.19) as in (3.2) (with a new $\Upsilon^\dagger(\eta)$):

$$\sum_{j=1}^{N} \Upsilon_{ij}^\dagger(\eta) y_{jt} = \alpha_i + x_{it}'\beta + \Gamma_i' f_t + e_{it} \tag{5.23}$$

Based on the above expression, the IGPC method can be implemented.

Remark 5.1. This paper considers a system of simultaneous equations with two dependent variables y_1 and y_2. It can be generalized to the multiple dependent variables case, i.e., $y_1, y_2, \ldots, y_p$. Both the estimation and the corresponding inferential analysis could be studied in a similar way but would require more mathematical calculation. This generalization warrants further study.

Remark 5.2. In this paper, I assume that the diagonal elements of the weights matrix w_{1ii} and w_{2ii} are zero for all i. This is a standard assumption implemented in the spatial modeling literature. However, in practice, there are cases where the diagonal elements of weights matrix are not all zero, for example, in an input-output matrix (in the production network by sector level as in Ozdagli and Weber (2017)). Then, one can slightly modify the definition of the transformation matrices $D(\delta)$ and $\Upsilon(\eta)$ and still apply the QML and IGPC estimation methods. The corresponding inferential theory needs modification according to the changes involved in $D(\delta)$ and $\Upsilon(\eta)$. The asymptotic analysis could be conducted as this paper with the same consistency and convergence rate but changes are needed for the limiting distribution. To avoid replication of the analysis in this paper, this is left for future research.

Remark 5.3. This paper considers a static case in the sense that the dependent variable y_t does not depend on its previous observation y_{t-1}. However, in practice, there might be cases when dynamic effects exist. In the trade and macroeconomics examples mentioned in the introduction, this year's GDP growth or trade growth might be affected by the previous year's values. Thus, it is potentially useful to study the dynamic case, where there are extra dynamic lags on the right-hand side of model (1.1). Taking the y_{1it} equation, for example, there would be a dynamic lag $y_{1i,t-1}$ on the right-hand side. The dynamic case of model (1.1) combines four effects: spatial effects, simultaneous effects, common shock effects and dynamic effects. Such a dynamic model would be useful for economic forecasting. Jointly modeling the first three effects is already difficult; the extra dynamic effect would make the analysis even more challenging. The dynamic case is studied in a work-in-progress paper.

6   Applications

In this section, I apply model (1.1) to explore the relationship between trade and GDP over time and across countries, taking into account spatial effects and global common shock effects.

In the literature on international trade, it has been difficult to establish a robust relationship between trade and GDP, due to the endogeneity issue between them. Many studies try to examine such a relationship using an IV approach. For instance, Frankel and Romer (1999) use a geographic instrument for trade and find a positive effect of trade on GDP in a cross-country setting. Such instruments are implemented by Irwin and Tervio (2002) and extended by Noguer and Siscart (2005), Felbermayr and Groschl (2013), Ortega and Peri (2014). Despite the importance of geographic instruments, as discussed in Winters and Masters (2013), they are time-invariant and thus preclude the use of panel data to analyze the effects of trade. Then, Feyrer (2009) makes progress by proposing a time-varying geographic instrument for trade. Based on a panel data model with simple additive individual and time fixed effects,[20] he finds that trade has a significant, positive impact on GDP, with an elasticity of approximately one-half. However, as mentioned in Feyrer (2009), his IV estimates are nearly identical to ordinary least squares estimates, indicating that his instrument is doubtful. Regardless of the contribution on the endogeneity issue, common shocks are not captured well by his model setup.

Instead, I apply framework (1.1) to model trade and GDP in a simultaneous equation system and incorporate common shock effects using a factor structure. Such a factor structure can be regarded as a form of interactive fixed effects, which provides a flexible way to control for potential serial and cross-sectional correlations. In addition, my model considers the spatial effect (i.e., spillover effects across countries for both trade and GDP) implied by gravity theory as noted in Helpman (1987) and Anderson and van Wincoop (2003).[21] In this application, the model is specified as follows:

$$Trade_{it} = \alpha_{1i} + \rho_1 \sum_{j=1}^{N} w_{ij} Trade_{jt} + \gamma_1 GDP_{it} + \beta_1 x_{1it} + \lambda_i' f_t + e_{1it}$$

$$GDP_{it} = \alpha_{2i} + \rho_2 \sum_{j=1}^{N} w_{ij} GDP_{jt} + \gamma_2 Trade_{it} + \beta_2 x_{2it} + \psi_i' f_t + e_{2it}$$

where $Trade_{it}$ and GDP_{it} are the log of total trade volume (export plus import) and the log

---

[20]Feyrer (2009) uses the real GDP per capita from the Penn World Tables, from 1950 to 1995 and cross 62 countries, with all estimation conducted on a panel with observations every 5 years.

[21]Similar spatial effects among firms due to cultural and social networks among firms as well as regional trade agreements have been studied in Baltagi et al. (2008), Lawless (2009), Rauch and Trindade (2002) and Defever et al. (2015).

Table 7: Estimation results using IGPC for the case without x ($**$ significant at 1%)

ρ_1	ρ_2	γ_1	γ_2
0.9597**	0.7751**	0.6326**	0.1074**
(0.0146)	(0.0195)	(0.0298)	(0.0047)

Table 8: Estimation results using IGPC for the case with x ($**$ significant at 1%)

ρ_1	ρ_2	γ_1	γ_2	β_1	β_2
0.9903**	0.5061**	0.5952**	0.0933**	0.1956**	0.9041**
(0.0332)	(0.0043)	(0.0853)	(0.0336)	(0.0118)	(0.0207)

of GDP, respectively, for country i in year t; weight w_{ij} is computed as $\frac{TotalTrade_{ij}}{\sum_{j=1}^{N} TotalTrade_{ij}}$ with $TotalTrade_{ij}$ being the total trade volume between country i and j.[22]

I investigate the above model in two ways, without explanatory variables and using population as an explanatory variable (i.e., $x_{1it} = x_{2it} = x_{it}$ denotes the log of population for country i in year t). Frankel and Romer (1999) adopt the same explanatory variable to control for country size.[23]

Without explanatory variables, I find one common factor based on the information criterion in (4.2). Given that $r = 1$, I estimate the model using the IGPC method; the results are presented in Table 7.[24] First, I find that trade and GDP are positively and significantly affected by each other. Specifically, the elasticity of trade with respect to GDP (γ_1) is approximately 0.6, while the elasticity of GDP with respect to trade (γ_2) is much smaller, approximately 0.1. By comparison, Feyrer (2009) identifies an elasticity of approximately one-half of GDP with respect to trade using an IV approach, which is much larger. The result in Feyrer (2009) might be less convincing, since the instrument is probably inappropriate, the spatial effect is not captured, and only additive individual and time fixed effects are controlled for. On the contrary, my model captures general interactive fixed effects through a factor structure. Moreover, his panel runs from 1950 to 1995 with observations every 5 years, while the application here uses annual data from 1961 to 2013. Second, I find that the trade volume of a country is positively affected by the trading parties' trade volumes; likewise for GDP. Specifically, the trade volume of a country can increase by almost 1% if

---

[22]Data source: Trade data and the weighting matrix (i.e. bilateral trade data) comes from IMF Directions of Trade Statistics. GDP and population data is obtained from Penn World Table. The sample period is from 1961 to 2013, with total 61 countries. In this case, $N = 61$ and $T = 53$. Weights are constructed using the bilateral trade data of the base year 1960, to avoid potential reversal causality. Both trade and GDP data are inflation adjusted.

[23]In both cases of without and with explanatory variables, the model is identified, by Assumption G$'$.

[24]Based on the IGPC results, I did panel unit root check for the errors by various tests, and overwhelming evidence rejects the hypothesis that the errors contain unit roots. Similarly for the case with control variable.

the average trade volume of the trading parties increases by 1%; similarly, the GDP of a country can increase by 0.77% if the average GDP of the trading parties increases by 1%.

With population included as a control variable, the estimation results and corresponding findings are similar. Again, the information criterion in (4.2) implies that there exists one common factor. The estimation results obtained by IGPC based on $r = 1$ are provided in Table 8. Similar estimated results are found for coefficients $(\rho_1, \rho_2, \gamma_1, \gamma_2)$. Additionally, the estimates of β_1 and β_2 indicate that country size in terms of population has positive impacts on both trade and GDP, with a larger impact on GDP: a 1% increase in population raises total trade volume by 0.2% whereas it increases GDP by 0.9%. By comparison, Frankel and Romer (1999) also find a positive impact of population on GDP but the magnitude is smaller: a 1% increase in population increases GDP by approximately 0.35% based on a cross-country study and an IV approach in a single equation setting.[25]

7   Conclusion

In this paper, I consider a simultaneous spatial panel data model, jointly modeling three important effects: spatial effects, common shock effects and simultaneous effects. Under joint modeling, there are many incidental parameters. Moreover, I take into account cross-sectional heteroskedasticity, which gives rise to additional incidental parameters. To estimate the model, I propose two different approaches, the QML method and the IGPC method. For each approach, I derive its identification condition and develop a full inferential theory for its estimators, including consistency, convergence rates and limiting distributions. The estimators from both methods are consistent. There is a trade-off between the model specification of the explanatory variable x and the asymptotic properties of the estimators for the two approaches. The QML method requires the model specification of x, but the gain is that its limiting distribution is unbiased (i.e., centered at zero) and more efficient (less variance than that of IGPC); the IGPC method does not require the specification of x, but the cost is that its limiting distribution contains a bias term and less efficient. Based on the limiting distribution of the IGPC estimator, the bias-corrected IGPC estimator is obtained. Then, I investigate the finite sample performance of both methods using Monte Carlo simulations. I find that both methods perform well and that the simulation results corroborate the inferential theories I derived in this paper. I also consider some extensions of the model. Finally, I apply the model to analyze the relationship between trade and GDP over time and cross countries, taking into account spatial effects and global common shock effects.

---

[25]The data used in Frankel and Romer (1999) are based on year 1985 only. In addition, Frankel and Romer (1999) use trade share (i.e., trade as a percentage of GDP) instead of trade itself, to study how openness affects GDP.

Appendix A   Notation I

In this Appendix, I define the important notation used in this paper. Table A1 includes the definitions of important matrices.

Table A1: Some important symbols used in the paper

$Y_{1t} = (y_{11t}, \ldots, y_{1Nt})'$	$Y_{2t} = (y_{21t}, \ldots, y_{2Nt})'$
$\ddot{Y}_{1t} = (\ddot{y}_{11t}, \ldots, \ddot{y}_{1Nt})'$	$\ddot{Y}_{2t} = (\ddot{y}_{21t}, \ldots, \ddot{y}_{2Nt})'$
$\Lambda = (\lambda_1, \lambda_2, \ldots, \lambda_N)'$	$\Psi = (\psi_1, \psi_2, \ldots, \psi_N)'$
$\Phi_{1p} = (\phi_{1p}, \phi_{2p}, \ldots, \phi_{Np})'$	$\Phi_{2p} = (\phi_{1p}, \phi_{2p}, \ldots, \phi_{Np})'$
$e_{1t} = (e_{11t}, \ldots, e_{1Nt})'$	$e_{2t} = (e_{21t}, \ldots, e_{2Nt})'$
$V_{1t} = (v_{11t}, \ldots, v_{1Nt})'$	$V_{2t} = (v_{21t}, \ldots, v_{2Nt})'$
$\Sigma_{1ee} = \operatorname{diag}(\sigma_{11}^2, \ldots, \sigma_{1N}^2)$	$\Sigma_{2ee} = \operatorname{diag}(\sigma_{21}^2, \ldots, \sigma_{2N}^2)$
$\Delta_{11} = \operatorname{diag}(\sigma_{11}^2 + \beta_1' \Sigma_{11v1}\beta_1, \ldots, \sigma_{1N}^2 + \beta_1' \Sigma_{NNv1}\beta_1)$	
$\Delta_{22} = \operatorname{diag}(\sigma_{21}^2 + \beta_2' \Sigma_{11v2}\beta_2, \ldots, \sigma_{2N}^2 + \beta_2' \Sigma_{NNv2}\beta_2)$	
$P_1 = I_N - \rho_1 W_1$	$P_2 = I_N - \rho_2 W_2$
$B_{12} = (I_N - \gamma_1\gamma_2 P_1^{-1}P_2^{-1})^{-1}$	$B_{21} = (I_N - \gamma_1\gamma_2 P_2^{-1}P_1^{-1})^{-1}$
$\mathbb{G}_1 = W_1 B_{12} P_1^{-1}$	$\mathbb{G}_2 = W_2 B_{21} P_2^{-1}$
$\mathbb{G}_3 = \gamma_2 B_{21} P_2^{-1} P_1^{-1}$	$\mathbb{G}_4 = \gamma_1 B_{12} P_1^{-1} P_2^{-1}$

Table A2 presents the definitions of scalars $a_1, a_2, b_1, b_2, c_1, c_2$ and matrices M_i $(i = 1, 2, 3, 4)$ used in Assumption G.

Table A2: The definition of $a_1, a_2, b_1, b_2, c_1, c_2$ as in Section 2

$a_1 = \frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}$	$\left[\mathbb{G}_{1,ij}\right]^2 + \left[(W_1\mathbb{G}_4)_{ij}\right]^2$
$a_2 = \frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}$	$\left[(W_2\mathbb{G}_3)_{ij}\right]^2 + \left[\mathbb{G}_{2,ij}\right]^2$
$b_1 = \frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}$	$\mathbb{G}_{1,ij}\mathbb{G}_{3,ij} + (W_1\mathbb{G}_4)_{ij}(B_{21}P_2^{-1})_{ij}$
$b_2 = \frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}$	$(W_2\mathbb{G}_3)_{ij}(B_{12}P_1^{-1})_{ij} + \mathbb{G}_{2,ij}\mathbb{G}_{4,ij}$
$c_1 = \frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}$	$\left[\mathbb{G}_{3,ij}\right]^2 + \left[(B_{21}P_2^{-1})_{ij}\right]^2$
$c_2 = \frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}$	$\left[(B_{12}P_1^{-1})_{ij}\right]^2 + \left[\mathbb{G}_{4,ij}\right]^2$
$M_1 = -W_1\mathbb{G}_4\Sigma_{2ee} + \mathbb{G}_1\Sigma_{1ee}\mathbb{G}_3'W_2'(\rho_2^{\dagger} - \rho_2) + \mathbb{G}_1\Sigma_{1ee}P_1^{-1\prime}B_{12}'(\gamma_2^{\dagger} - \gamma_2)$	
$M_2 = -\Sigma_{1ee}\mathbb{G}_3 W_2' + W_1\mathbb{G}_4\Sigma_{2ee}\mathbb{G}_2'(\rho_1^{\dagger} - \rho_1) + B_{21}P_2^{-1}\Sigma_{2ee}\mathbb{G}_2'(\gamma_1^{\dagger} - \gamma_1)$	
$M_3 = -B_{21}P_2^{-1}\Sigma_{2ee} + \mathbb{G}_3\Sigma_{1ee}\mathbb{G}_3'W_2'(\rho_2^{\dagger} - \rho_2) + \mathbb{G}_3\Sigma_{1ee}P_1^{-1\prime}B_{12}'(\gamma_2^{\dagger} - \gamma_2)$	
$M_4 = -\Sigma_{1ee}P_1^{-1\prime}B_{12}' + W_1\mathbb{G}_4\Sigma_{2ee}\mathbb{G}_4'(\rho_1^{\dagger} - \rho_1) + B_{21}P_2^{-1}\Sigma_{2ee}\mathbb{G}_4'(\gamma_1^{\dagger} - \gamma_1)$	

Tables [A3](#) and [A4](#) list the definitions of Ω and ε, which are contained in Theorem [2.2](#).

Table A3: The detailed definition of Ω

$$\Omega_{11} = \frac{1}{N}\text{tr}\left[\Sigma_{1ee}^{-1}\left(\mathbb{G}_1\Delta_{11}\mathbb{G}_1' + W_1\mathbb{G}_4\Delta_{22}\mathbb{G}_4'W_1'\right)\right] - \left[\frac{2}{N}\sum_{i=1}^{N}(\mathbb{G}_{1,ii})^2\right] + \frac{1}{N}\text{tr}\left[\mathbb{G}_1^2\right]$$

$$\Omega_{13} = \frac{1}{N}\text{tr}\left[\Sigma_{1ee}^{-1}\left(\mathbb{G}_1\Delta_{11}\mathbb{G}_3' + W_1\mathbb{G}_4\Delta_{22}P_2^{-1'}B_{21}'\right)\right] - \left[\frac{2}{N}\sum_{i=1}^{N}\mathbb{G}_{1,ii}\mathbb{G}_{3,ii}\right] + \frac{1}{N}\text{tr}\left[\mathbb{G}_3\mathbb{G}_1\right]$$

$$\Omega_{12} = \frac{1}{N}\text{tr}\left[W_2\mathbb{G}_3W_1\mathbb{G}_4\right] \qquad \Omega_{14} = \frac{1}{N}\text{tr}\left[B_{12}P_1^{-1}W_1\mathbb{G}_4\right]$$

$$\Omega_{15} = \beta_1'\left[\frac{1}{N}\sum_{i=1}^{N}\mathbb{G}_{1,ii}\Sigma_{iiv1}/\sigma_{1i}^2\right] \qquad \Omega_{16} = 0_{1\times k_2}$$

$$\Omega_{22} = \frac{1}{N}\text{tr}\left[\Sigma_{2ee}^{-1}\left(W_2\mathbb{G}_3\Delta_{11}\mathbb{G}_3'W_2' + \mathbb{G}_2\Delta_{22}\mathbb{G}_2'\right)\right] - \left[\frac{2}{N}\sum_{i=1}^{N}(\mathbb{G}_{2,ii})^2\right] + \frac{1}{N}\text{tr}\left[\mathbb{G}_2^2\right]$$

$$\Omega_{24} = \frac{1}{N}\text{tr}\left[\Sigma_{2ee}^{-1}\left(W_2\mathbb{G}_3\Delta_{11}P_1^{-1'}B_{12}' + \mathbb{G}_2\Delta_{22}\mathbb{G}_4'\right)\right] - \left[\frac{2}{N}\sum_{i=1}^{N}\mathbb{G}_{2,ii}\mathbb{G}_{4,ii}\right] + \frac{1}{N}\text{tr}\left[\mathbb{G}_4\mathbb{G}_2\right]$$

$$\Omega_{21} = \frac{1}{N}\text{tr}\left[W_1\mathbb{G}_4W_2\mathbb{G}_3\right] \qquad \Omega_{23} = \frac{1}{N}\text{tr}\left[B_{21}P_2^{-1}W_2\mathbb{G}_3\right]$$

$$\Omega_{25} = 0_{1\times k_1} \qquad \Omega_{26} = \beta_2'\left[\frac{1}{N}\sum_{i=1}^{N}\mathbb{G}_{2,ii}\Sigma_{iiv2}/\sigma_{2i}^2\right]$$

$$\Omega_{31} = \frac{1}{N}\text{tr}\left[\Sigma_{1ee}^{-1'}\left(\mathbb{G}_3\Delta_{11}\mathbb{G}_1' + B_{21}P_2^{-1}\Delta_{22}\mathbb{G}_4'W_1'\right)\right] - \left[\frac{2}{N}\sum_{i=1}^{N}\mathbb{G}_{1,ii}\mathbb{G}_{3,ii}\right] + \frac{1}{N}\text{tr}\left[\mathbb{G}_1\mathbb{G}_3\right]$$

$$\Omega_{33} = \frac{1}{N}\text{tr}\left[\Sigma_{1ee}^{-1'}\left(\mathbb{G}_3\Delta_{11}\mathbb{G}_3' + B_{21}P_2^{-1}\Delta_{22}P_2^{-1'}B_{21}'\right)\right] - \left[\frac{2}{N}\sum_{i=1}^{N}(\mathbb{G}_{3,ii})^2\right] + \frac{1}{N}\text{tr}\left[\mathbb{G}_3^2\right]$$

$$\Omega_{32} = \frac{1}{N}\text{tr}\left[W_2\mathbb{G}_3B_{21}P_2^{-1}\right] \qquad \Omega_{34} = \frac{1}{N}\text{tr}\left[B_{12}P_1^{-1}B_{21}P_2^{-1}\right]$$

$$\Omega_{35} = \beta_1'\left[\frac{1}{N}\sum_{i=1}^{N}\mathbb{G}_{3,ii}\Sigma_{iiv1}/\sigma_{1i}^2\right] \qquad \Omega_{36} = 0_{1\times k_2}$$

$$\Omega_{42} = \frac{1}{N}\text{tr}\left[\Sigma_{2ee}^{-1}\left(W_2\mathbb{G}_3\Delta_{11}P_1^{-1'}B_{12}' + \mathbb{G}_2\Delta_{22}\mathbb{G}_4'\right)\right] - \left[\frac{2}{N}\sum_{i=1}^{N}\mathbb{G}_{2,ii}\mathbb{G}_{4,ii}\right] + \frac{1}{N}\text{tr}\left[\mathbb{G}_2\mathbb{G}_4\right]$$

$$\Omega_{44} = \frac{1}{N}\text{tr}\left[\Sigma_{2ee}^{-1}\left(B_{12}P_1^{-1}\Delta_{11}P_1^{-1'}B_{12}' + \mathbb{G}_4\Delta_{22}\mathbb{G}_4'\right)\right] - \left[\frac{2}{N}\sum_{i=1}^{N}(\mathbb{G}_{4,ii})^2\right] + \frac{1}{N}\text{tr}\left[\mathbb{G}_4^2\right]$$

$$\Omega_{41} = \frac{1}{N}\text{tr}\left[W_1\mathbb{G}_4B_{12}P_1^{-1}\right] \qquad \Omega_{43} = \frac{1}{N}\text{tr}\left[B_{21}P_2^{-1}B_{12}P_1^{-1}\right]$$

$$\Omega_{45} = 0_{1\times k_1} \qquad \Omega_{46} = \beta_2'\left[\frac{1}{N}\sum_{i=1}^{N}\mathbb{G}_{4,ii}\Sigma_{iiv2}/\sigma_{2i}^2\right]$$

$$\Omega_{51} = \left[\frac{1}{N}\sum_{i=1}^{N}\mathbb{G}_{1,ii}\Sigma_{iiv1}/\sigma_{1i}^2\right]\beta_1 \qquad \Omega_{52} = 0_{k_1\times 1}$$

$$\Omega_{53} = \left[\frac{1}{N}\sum_{i=1}^{N}\mathbb{G}_{3,ii}\Sigma_{iiv1}/\sigma_{1i}^2\right]\beta_1 \qquad \Omega_{54} = 0_{k_1\times 1}$$

$$\Omega_{55} = \frac{1}{N}\sum_{i=1}^{N}\Sigma_{iiv1}/\sigma_{1i}^2 \qquad \Omega_{56} = 0_{k_1\times k_2}$$

$$\Omega_{62} = \left[\frac{1}{N}\sum_{i=1}^{N}\mathbb{G}_{2,ii}\Sigma_{iiv2}/\sigma_{2i}^2\right]\beta_2 \qquad \Omega_{61} = 0_{k_2\times 1}$$

$$\Omega_{64} = \left[\frac{1}{N}\sum_{i=1}^{N}\mathbb{G}_{4,ii}\Sigma_{iiv2}/\sigma_{2i}^2\right]\beta_2 \qquad \Omega_{63} = 0_{k_2\times 1}$$

$$\Omega_{66} = \frac{1}{N}\sum_{i=1}^{N}\Sigma_{iiv2}/\sigma_{2i}^2 \qquad \Omega_{65} = 0_{k_2\times k_1}$$

Table A4: The detailed definition of $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_6)'$

$\varepsilon_1 = \frac{1}{NT}\sum_{t=1}^{T}\beta_1'V_{1t}'\mathbb{G}_1'\Sigma_{1ee}^{-1}e_{1t} + \frac{1}{NT}\sum_{t=1}^{T}\beta_2'V_{2t}'\mathbb{G}_4'W_1'\Sigma_{1ee}^{-1}e_{1t}$
$\quad + \frac{1}{NT}\sum_{t=1}^{T}e_{2t}'\mathbb{G}_4'W_1'\Sigma_{1ee}^{-1}e_{1t} + \frac{1}{NT}\sum_{t=1}^{T}e_{1t}'\mathbb{G}_1^{o\prime}\Sigma_{1ee}^{-1}e_{1t}$
$\varepsilon_2 = \frac{1}{NT}\sum_{t=1}^{T}\beta_2'V_{2t}'\mathbb{G}_2'\Sigma_{2ee}^{-1}e_{2t} + \frac{1}{NT}\sum_{t=1}^{T}\beta_1'V_{1t}'\mathbb{G}_3'W_2'\Sigma_{2ee}^{-1}e_{2t}$
$\quad + \frac{1}{NT}\sum_{t=1}^{T}e_{1t}'\mathbb{G}_3'W_2'\Sigma_{2ee}^{-1}e_{2t} + \frac{1}{NT}\sum_{t=1}^{T}e_{2t}'\mathbb{G}_2^{o\prime}\Sigma_{2ee}^{-1}e_{2t}$
$\varepsilon_3 = \frac{1}{NT}\sum_{t=1}^{T}\beta_2'V_{2t}'P_2^{-1\prime}B_{21}'\Sigma_{1ee}^{-1}e_{1t} + \frac{1}{NT}\sum_{t=1}^{T}\beta_1'V_{1t}'\mathbb{G}_3'\Sigma_{1ee}^{-1}e_{1t}$
$\quad + \frac{1}{NT}\sum_{t=1}^{T}e_{2t}'P_2^{-1\prime}B_{21}'\Sigma_{1ee}^{-1}e_{1t} + \frac{1}{NT}\sum_{t=1}^{T}e_{1t}'\mathbb{G}_3^{o\prime}\Sigma_{1ee}^{-1}e_{1t}$
$\varepsilon_4 = \frac{1}{NT}\sum_{t=1}^{T}\beta_1'V_{1t}'P_1^{-1\prime}B_{12}'\Sigma_{2ee}^{-1}e_{2t} + \frac{1}{NT}\sum_{t=1}^{T}\beta_2'V_{2t}'\mathbb{G}_4'\Sigma_{2ee}^{-1}e_{2t}$
$\quad + \frac{1}{NT}\sum_{t=1}^{T}e_{1t}'P_1^{-1\prime}B_{12}'\Sigma_{2ee}^{-1}e_{2t} + \frac{1}{NT}\sum_{t=1}^{T}e_{2t}'\mathbb{G}_4^{o\prime}\Sigma_{2ee}^{-1}e_{2t}$
$\varepsilon_5 = \frac{1}{NT}\sum_{t=1}^{T}V_{1t}'\Sigma_{1ee}^{-1}e_{1t}$
$\varepsilon_6 = \frac{1}{NT}\sum_{t=1}^{T}V_{2t}'\Sigma_{2ee}^{-1}e_{2t}$

where $\mathbb{G}_p^o = \mathbb{G}_p - \mathbb{G}_p^d$, with $\mathbb{G}_p^d$ being a diagonal matrix whose diagonal elements equal the diagonal elements of $\mathbb{G}_p$ for $p = 1, 2, 3, 4$.

Tables A6 and A5 provide the definitions of $\tilde{\Omega}$ and $\tilde{\varepsilon}$, which are involved in Remark 2.15.

Table A5: The detailed definition of $\tilde{\varepsilon} = (\tilde{\varepsilon}_1, \tilde{\varepsilon}_2, \ldots, \tilde{\varepsilon}_6)'$

$\tilde{\varepsilon}_1 = \frac{1}{NT\sigma_1^2}\sum_{t=1}^{T}\beta_1'V_{1t}'\mathbb{G}_1'e_{1t} + \frac{1}{NT\sigma_1^2}\sum_{t=1}^{T}\beta_2'V_{2t}'\mathbb{G}_4'W_1'e_{1t}$
$\quad + \frac{1}{NT\sigma_1^2}\sum_{t=1}^{T}e_{2t}'\mathbb{G}_4'W_1'e_{1t} + \frac{1}{NT\sigma_1^2}\left[\sum_{t=1}^{T}e_{1t}'\mathbb{G}_1^{o\prime}e_{1t} + \omega_1\right]$
$\tilde{\varepsilon}_2 = \frac{1}{NT\sigma_2^2}\sum_{t=1}^{T}\beta_2'V_{2t}'\mathbb{G}_2'e_{2t} + \frac{1}{NT\sigma_2^2}\sum_{t=1}^{T}\beta_1'V_{1t}'\mathbb{G}_3'W_2'e_{2t}$
$\quad + \frac{1}{NT\sigma_2^2}\sum_{t=1}^{T}e_{1t}'\mathbb{G}_3'W_2'e_{2t} + \frac{1}{NT\sigma_2^2}\left[\sum_{t=1}^{T}e_{2t}'\mathbb{G}_2^{o\prime}e_{2t} + \omega_2\right]$
$\tilde{\varepsilon}_3 = \frac{1}{NT\sigma_1^2}\sum_{t=1}^{T}\beta_2'V_{2t}'P_2^{-1\prime}B_{21}'e_{1t} + \frac{1}{NT\sigma_1^2}\sum_{t=1}^{T}\beta_1'V_{1t}'\mathbb{G}_3'e_{1t}$
$\quad + \frac{1}{NT\sigma_1^2}\sum_{t=1}^{T}e_{2t}'P_2^{-1\prime}B_{21}'e_{1t} + \frac{1}{NT\sigma_1^2}\left[\sum_{t=1}^{T}e_{1t}'\mathbb{G}_3^{o\prime}e_{1t} + \omega_3\right]$
$\tilde{\varepsilon}_4 = \frac{1}{NT\sigma_2^2}\sum_{t=1}^{T}\beta_1'V_{1t}'P_1^{-1\prime}B_{12}'e_{2t} + \frac{1}{NT\sigma_2^2}\sum_{t=1}^{T}\beta_2'V_{2t}'\mathbb{G}_4'e_{2t}$
$\quad + \frac{1}{NT\sigma_2^2}\sum_{t=1}^{T}e_{1t}'P_1^{-1\prime}B_{12}'e_{2t} + \frac{1}{NT\sigma_2^2}\left[\sum_{t=1}^{T}e_{2t}'\mathbb{G}_4^{o\prime}e_{2t} + \omega_4\right]$
$\tilde{\varepsilon}_5 = \frac{1}{NT}\sum_{t=1}^{T}V_{1t}'\Sigma_{1ee}^{-1}e_{1t}$
$\tilde{\varepsilon}_6 = \frac{1}{NT}\sum_{t=1}^{T}V_{2t}'\Sigma_{2ee}^{-1}e_{2t}$

where
$\omega_1 = \sum_{i=1}^{N}\sum_{t=1}^{T}\left[\mathbb{G}_{1,ii} - \frac{1}{N}\text{tr}(\mathbb{G}_1)\right](e_{1it}^2 - \sigma_1^2)$, $\omega_2 = \sum_{i=1}^{N}\sum_{t=1}^{T}\left[\mathbb{G}_{2,ii} - \frac{1}{N}\text{tr}(\mathbb{G}_2)\right](e_{2it}^2 - \sigma_2^2)$
$\omega_3 = \sum_{i=1}^{N}\sum_{t=1}^{T}\left[\mathbb{G}_{3,ii} - \frac{1}{N}\text{tr}(\mathbb{G}_3)\right](e_{1it}^2 - \sigma_1^2)$, $\omega_4 = \sum_{i=1}^{N}\sum_{t=1}^{T}\left[\mathbb{G}_{4,ii} - \frac{1}{N}\text{tr}(\mathbb{G}_4)\right](e_{2it}^2 - \sigma_2^2)$

Table A6: The detailed definition of $\tilde{\Omega}$

$\tilde{\Omega}_{11} = \frac{1}{N}\mathrm{tr}\left[\frac{1}{\sigma_1^2}\left(\mathbb{G}_1\Delta_{11}\mathbb{G}_1' + W_1\mathbb{G}_4\Delta_{22}\mathbb{G}_4'W_1'\right)\right] - \frac{2}{N^2}\left[\mathrm{tr}(\mathbb{G}_1)\right]^2 + \frac{1}{N}\mathrm{tr}\left[\mathbb{G}_1^2\right]$	
$\tilde{\Omega}_{13} = \frac{1}{N}\mathrm{tr}\left[\frac{1}{\sigma_1^2}\left(\mathbb{G}_1\Delta_{11}\mathbb{G}_3' + W_1\mathbb{G}_4\Delta_{22}P_2^{-1\prime}B_{21}'\right)\right] - \frac{2}{N^2}\left[\mathrm{tr}(\mathbb{G}_1)\mathrm{tr}(\mathbb{G}_3)\right] + \frac{1}{N}\mathrm{tr}\left[\mathbb{G}_3\mathbb{G}_1\right]$	
$\tilde{\Omega}_{12} = \frac{1}{N}\mathrm{tr}\left[W_2\mathbb{G}_3W_1\mathbb{G}_4\right]$	$\tilde{\Omega}_{14} = \frac{1}{N}\mathrm{tr}\left[B_{12}P_1^{-1}W_1\mathbb{G}_4\right]$
$\tilde{\Omega}_{15} = \beta_1'\left[\frac{1}{N}\sum_{i=1}^N \mathbb{G}_{1,ii}\Sigma_{iiv1}/\sigma_1^2\right]$	$\tilde{\Omega}_{16} = 0_{1\times k_2}$
$\tilde{\Omega}_{22} = \frac{1}{N}\mathrm{tr}\left[\frac{1}{\sigma_2^2}\left(W_2\mathbb{G}_3\Delta_{11}\mathbb{G}_3'W_2' + \mathbb{G}_2\Delta_{22}\mathbb{G}_2'\right)\right] - \frac{2}{N^2}\left[\mathrm{tr}(\mathbb{G}_2)\right]^2 + \frac{1}{N}\mathrm{tr}\left[\mathbb{G}_2^2\right]$	
$\tilde{\Omega}_{24} = \frac{1}{N}\mathrm{tr}\left[\frac{1}{\sigma_2^2}\left(W_2\mathbb{G}_3\Delta_{11}P_1^{-1\prime}B_{12}' + \mathbb{G}_2\Delta_{22}\mathbb{G}_4'\right)\right] - \frac{2}{N^2}\left[\mathrm{tr}(\mathbb{G}_2)\mathrm{tr}(\mathbb{G}_4)\right] + \frac{1}{N}\mathrm{tr}\left[\mathbb{G}_4\mathbb{G}_2\right]$	
$\tilde{\Omega}_{21} = \frac{1}{N}\mathrm{tr}\left[W_1\mathbb{G}_4W_2\mathbb{G}_3\right]$	$\tilde{\Omega}_{23} = \frac{1}{N}\mathrm{tr}\left[B_{21}P_2^{-1}W_2\mathbb{G}_3\right]$
$\tilde{\Omega}_{25} = 0_{1\times k_1}$	$\tilde{\Omega}_{26} = \beta_2'\left[\frac{1}{N}\sum_{i=1}^N \mathbb{G}_{2,ii}\Sigma_{iiv2}/\sigma_2^2\right]$
$\tilde{\Omega}_{31} = \frac{1}{N}\mathrm{tr}\left[\frac{1}{\sigma_1^2}\left(\mathbb{G}_3\Delta_{11}\mathbb{G}_1' + B_{21}P_2^{-1}\Delta_{22}\mathbb{G}_4'W_1'\right)\right] - \frac{2}{N^2}\left[\mathrm{tr}(\mathbb{G}_1)\mathrm{tr}(\mathbb{G}_3)\right] + \frac{1}{N}\mathrm{tr}\left[\mathbb{G}_1\mathbb{G}_3\right]$	
$\tilde{\Omega}_{33} = \frac{1}{N}\mathrm{tr}\left[\frac{1}{\sigma_1^2}\left(\mathbb{G}_3\Delta_{11}\mathbb{G}_3' + B_{21}P_2^{-1}\Delta_{22}P_2^{-1\prime}B_{21}'\right)\right] - \frac{2}{N^2}\left[\mathrm{tr}(\mathbb{G}_3)\right]^2 + \frac{1}{N}\mathrm{tr}\left[\mathbb{G}_3^2\right]$	
$\tilde{\Omega}_{32} = \frac{1}{N}\mathrm{tr}\left[W_2\mathbb{G}_3B_{21}P_2^{-1}\right]$	$\tilde{\Omega}_{34} = \frac{1}{N}\mathrm{tr}\left[B_{12}P_1^{-1}B_{21}P_2^{-1}\right]$
$\tilde{\Omega}_{35} = \beta_1'\left[\frac{1}{N}\sum_{i=1}^N \mathbb{G}_{3,ii}\Sigma_{iiv1}/\sigma_1^2\right]$	$\tilde{\Omega}_{36} = 0_{1\times k_2}$
$\tilde{\Omega}_{42} = \frac{1}{N}\mathrm{tr}\left[\frac{1}{\sigma_2^2}\left(W_2\mathbb{G}_3\Delta_{11}P_1^{-1\prime}B_{12}' + \mathbb{G}_2\Delta_{22}\mathbb{G}_4'\right)\right] - \frac{2}{N^2}\left[\mathrm{tr}(\mathbb{G}_2)\mathrm{tr}(\mathbb{G}_4)\right] + \frac{1}{N}\mathrm{tr}\left[\mathbb{G}_2\mathbb{G}_4\right]$	
$\tilde{\Omega}_{44} = \frac{1}{N}\mathrm{tr}\left[\frac{1}{\sigma_2^2}\left(B_{12}P_1^{-1}\Delta_{11}P_1^{-1\prime}B_{12}' + \mathbb{G}_4\Delta_{22}\mathbb{G}_4'\right)\right] - \frac{2}{N^2}\left[\mathrm{tr}(\mathbb{G}_4)\right]^2 + \frac{1}{N}\mathrm{tr}\left[\mathbb{G}_4^2\right]$	
$\tilde{\Omega}_{41} = \frac{1}{N}\mathrm{tr}\left[W_1\mathbb{G}_4B_{12}P_1^{-1}\right]$	$\tilde{\Omega}_{43} = \frac{1}{N}\mathrm{tr}\left[B_{21}P_2^{-1}B_{12}P_1^{-1}\right]$
$\tilde{\Omega}_{45} = 0_{1\times k_1}$	$\tilde{\Omega}_{46} = \beta_2'\left[\frac{1}{N}\sum_{i=1}^N \mathbb{G}_{4,ii}\Sigma_{iiv2}/\sigma_2^2\right]$
$\tilde{\Omega}_{51} = \left[\frac{1}{N}\sum_{i=1}^N \mathbb{G}_{1,ii}\Sigma_{iiv1}/\sigma_1^2\right]\beta_1$	$\tilde{\Omega}_{52} = 0_{k_1\times 1}$
$\tilde{\Omega}_{53} = \left[\frac{1}{N}\sum_{i=1}^N \mathbb{G}_{3,ii}\Sigma_{iiv1}/\sigma_1^2\right]\beta_1$	$\tilde{\Omega}_{54} = 0_{k_1\times 1}$
$\tilde{\Omega}_{55} = \frac{1}{N}\sum_{i=1}^N \Sigma_{iiv1}/\sigma_1^2$	$\tilde{\Omega}_{56} = 0_{k_1\times k_2}$
$\tilde{\Omega}_{62} = \left[\frac{1}{N}\sum_{i=1}^N \mathbb{G}_{2,ii}\Sigma_{iiv2}/\sigma_2^2\right]\beta_2$	$\tilde{\Omega}_{61} = 0_{k_2\times 1}$
$\tilde{\Omega}_{64} = \left[\frac{1}{N}\sum_{i=1}^N \mathbb{G}_{4,ii}\Sigma_{iiv2}/\sigma_2^2\right]\beta_2$	$\tilde{\Omega}_{63} = 0_{k_2\times 1}$
$\tilde{\Omega}_{66} = \frac{1}{N}\sum_{i=1}^N \Sigma_{iiv2}/\sigma_2^2$	$\tilde{\Omega}_{65} = 0_{k_2\times k_1}$

Table A7 provides the specifications of matrices Q_l ($l = 1, \ldots, 4$), which are involved in the IGPC approach as in Assumption G' and Theorem 3.2. Each Q_l is a $2N \times 2N$ matrix whose (i, j)th 2×2 subblock $Q_{l,ij}$ is defined as below.

Table A7: The specification of $Q_{l,ij}$ ($l = 1, \ldots, 4$) involved in the IGPC approach

$Q_{1,ij} = \begin{bmatrix} \mathbb{G}_{1,ij} & (W_1\mathbb{G}_4)_{ij} \\ 0 & 0 \end{bmatrix},$	$Q_{2,ij} = \begin{bmatrix} 0 & 0 \\ (W_2\mathbb{G}_3)_{ij} & \mathbb{G}_{2,ij} \end{bmatrix}$
$Q_{3,ij} = \begin{bmatrix} \mathbb{G}_{3,ij} & (B_{21}P_2^{-1})_{ij} \\ 0 & 0 \end{bmatrix},$	$Q_{4,ij} = \begin{bmatrix} 0 & 0 \\ (B_{12}P_1^{-1})_{ij} & \mathbb{G}_{4,ij} \end{bmatrix}$

where matrices $\mathbb{G}_l$ for $l = 1, 2, \ldots, 4$ are defined in Table A1.

Appendix B   Notation II

In this Appendix, I first introduce the definitions of $\mathbb{D}_b, \mathbb{D}_c$ and ζ, which are involved in Assumption G', and then the definitions of the $\mathbb{D}_\eta, \Phi, \vartheta$ and $\mathbb{D}_\beta$ matrices involved in Theorem 3.2. All these matrices are used in the IGPC approach.

Some matrices in Assumption G': $\mathbb{D}_b, \mathbb{D}_c$ and ζ.

For $p = 1, \ldots, k_1$, let $\widetilde{X}_p = (\widetilde{X}_{1p}, \ldots, \widetilde{X}_{Tp})$ be a $2N \times T$ matrix, where $\widetilde{X}_{tp} = (\dot{x}_{11t,p}, 0, \dot{x}_{12t,p}, 0, \ldots, \dot{x}_{1Nt,p}, 0)'$, with $\dot{x}_{1it,p}$ being the de-meaned version of $x_{1it,p}$ defined as $\dot{x}_{1it,p} = x_{1it,p} - \frac{1}{T}\sum_{s=1}^{T} x_{1is,p}$. For $p = k_1+1, \ldots, k$ with $k = k_1+k_2$, let $\widetilde{X}_p = (\widetilde{X}_{1p}, \ldots, \widetilde{X}_{Tp})$ be a $2N \times T$ matrix, where $\widetilde{X}_{tp} = (0, \dot{x}_{21t,(p-k_1)}, 0, \dot{x}_{22t,(p-k_1)}, \ldots, 0, \dot{x}_{2Nt,(p-k_1)})'$ with $\dot{x}_{2it,p}$ is defined as $\dot{x}_{1it,p}$. For $p = (k+1), \ldots, (k+4)$, let $\widetilde{X}_p = Q_{p-k}\left(\sum_{q=1}^{k} \widetilde{X}_q \tilde{\beta}_q\right)$, where $\tilde{\beta}_l = \beta_{1l}$ for $l = 1, \ldots, k_1$ and $\tilde{\beta}_l = \beta_{2,l-k_1}$ for $l = k_1 + 1, \ldots, k$. Each Q_l $(l = 1, \ldots, 4)$ is a $2N \times 2N$ matrix, defined as in Theorem 3.2, with its specification presented in Table A7 in Appendix A. Further, $\mathbb{D}_b$ is a $k \times k$ matrix defined as:

$$\mathbb{D}_b = \frac{1}{NT} \begin{bmatrix} \text{tr}(\widetilde{X}_1'\ddot{M}\widetilde{X}_1 M_F) & \text{tr}(\widetilde{X}_1'\ddot{M}\widetilde{X}_2 M_F) & \ldots & \text{tr}(\widetilde{X}_1'\ddot{M}\widetilde{X}_k M_F) \\ \text{tr}(\widetilde{X}_2'\ddot{M}\widetilde{X}_1 M_F) & \text{tr}(\widetilde{X}_2'\ddot{M}\widetilde{X}_2 M_F) & \ldots & \text{tr}(\widetilde{X}_2'\ddot{M}\widetilde{X}_k M_F) \\ \vdots & \vdots & \ddots & \vdots \\ \text{tr}(\widetilde{X}_k'\ddot{M}\widetilde{X}_1 M_F) & \text{tr}(\widetilde{X}_k'\ddot{M}\widetilde{X}_2 M_F) & \ldots & \text{tr}(\widetilde{X}_k'\ddot{M}\widetilde{X}_k M_F) \end{bmatrix}$$

$\mathbb{D}_c$ is a 4×4 matrix defined as:

$$\mathbb{D}_c = \frac{1}{NT} \begin{bmatrix} \text{tr}(\widetilde{X}_{k+1}'\ddot{M}\widetilde{X}_{k+1} M_F) & \text{tr}(\widetilde{X}_{k+1}'\ddot{M}\widetilde{X}_{k+2} M_F) & \ldots & \text{tr}(\widetilde{X}_{k+1}'\ddot{M}\widetilde{X}_{k+4} M_F) \\ \text{tr}(\widetilde{X}_{k+2}'\ddot{M}\widetilde{X}_{k+1} M_F) & \text{tr}(\widetilde{X}_{k+2}'\ddot{M}\widetilde{X}_{k+2} M_F) & \ldots & \text{tr}(\widetilde{X}_{k+2}'\ddot{M}\widetilde{X}_{k+4} M_F) \\ \vdots & \vdots & \ddots & \vdots \\ \text{tr}(\widetilde{X}_{k+4}'\ddot{M}\widetilde{X}_{k+1} M_F) & \text{tr}(\widetilde{X}_{k+4}'\ddot{M}\widetilde{X}_{k+2} M_F) & \ldots & \text{tr}(\widetilde{X}_{k+4}'\ddot{M}\widetilde{X}_{k+4} M_F) \end{bmatrix}$$

ζ is a $k \times 4$ matrix defined as:

$$\zeta = \frac{1}{NT} \begin{bmatrix} \text{tr}(\widetilde{X}_1'\ddot{M}\widetilde{X}_{k+1} M_F) & \text{tr}(\widetilde{X}_1'\ddot{M}\widetilde{X}_{k+2} M_F) & \ldots & \text{tr}(\widetilde{X}_1'\ddot{M}\widetilde{X}_{k+4} M_F) \\ \text{tr}(\widetilde{X}_2'\ddot{M}\widetilde{X}_{k+1} M_F) & \text{tr}(\widetilde{X}_2'\ddot{M}\widetilde{X}_{k+2} M_F) & \ldots & \text{tr}(\widetilde{X}_2'\ddot{M}\widetilde{X}_{k+4} M_F) \\ \vdots & \vdots & \ddots & \vdots \\ \text{tr}(\widetilde{X}_k'\ddot{M}\widetilde{X}_{k+1} M_F) & \text{tr}(\widetilde{X}_k'\ddot{M}\widetilde{X}_{k+2} M_F) & \ldots & \text{tr}(\widetilde{X}_k'\ddot{M}\widetilde{X}_{k+4} M_F) \end{bmatrix}$$

where $M_F = I_T - F(F'F)^{-1}F'$, and $\ddot{M} = \Sigma_{ee}^{-1} - \frac{1}{N}\Sigma_{ee}^{-1}\Gamma\Gamma'\Sigma_{ee}^{-1}$ with Σ_{ee} as defined in Assumption C.

Some matrices in Theorem 3.2: $\mathbb{D}_\eta, \Phi, \vartheta$ **and** $\mathbb{D}_\beta$. First, $\mathbb{D}_\eta$ is defined as:

$$
\mathbb{D}_\eta = \frac{1}{NT}
\begin{bmatrix}
\text{tr}[\ddot{Y}_1'\ddot{M}_\Lambda\ddot{Y}_1 M_F] & \text{tr}[\ddot{Y}_1'\ddot{M}_{\Lambda\Psi}\ddot{Y}_2 M_F] & \text{tr}[\ddot{Y}_1'\ddot{M}_\Lambda\dot{Y}_2 M_F] & \text{tr}[\ddot{Y}_1'\ddot{M}_{\Lambda\Psi}\dot{Y}_1 M_F] \\
\text{tr}[\ddot{Y}_2'\ddot{M}_{\Psi\Lambda}\ddot{Y}_1 M_F] & \text{tr}[\ddot{Y}_2'\ddot{M}_\Psi\ddot{Y}_2 M_F] & \text{tr}[\ddot{Y}_2'\ddot{M}_{\Psi\Lambda}\dot{Y}_2 M_F] & \text{tr}[\ddot{Y}_2'\ddot{M}_\Psi\dot{Y}_1 M_F] \\
\text{tr}[\dot{Y}_2'\ddot{M}_\Lambda\ddot{Y}_1 M_F] & \text{tr}[\dot{Y}_2'\ddot{M}_{\Lambda\Psi}\ddot{Y}_2 M_F] & \text{tr}[\dot{Y}_2'\ddot{M}_\Lambda\dot{Y}_2 M_F] & \text{tr}[\dot{Y}_2'\ddot{M}_{\Lambda\Psi}\dot{Y}_1 M_F] \\
\text{tr}[\dot{Y}_1'\ddot{M}_{\Psi\Lambda}\ddot{Y}_1 M_F] & \text{tr}[\dot{Y}_1'\ddot{M}_\Psi\ddot{Y}_2 M_F] & \text{tr}[\dot{Y}_1'\ddot{M}_{\Psi\Lambda}\dot{Y}_2 M_F] & \text{tr}[\dot{Y}_1'\ddot{M}_\Psi\dot{Y}_1 M_F]
\end{bmatrix}
$$

where $\Lambda = (\lambda_1, \lambda_2, \ldots, \lambda_N)'$; $\Psi = (\psi_1, \psi_2, \ldots, \psi_N)'$; $\ddot{M}_\Lambda = \Sigma_{1ee}^{-1} - \frac{1}{N}\Sigma_{1ee}^{-1}\Lambda\Lambda'\Sigma_{1ee}^{-1}$; $\ddot{M}_\Psi = \Sigma_{2ee}^{-1} - \frac{1}{N}\Sigma_{2ee}^{-1}\Psi\Psi'\Sigma_{2ee}^{-1}$; $\ddot{M}_{\Lambda\Psi} = -\frac{1}{N}\Sigma_{1ee}^{-1}\Lambda\Psi'\Sigma_{2ee}^{-1}$; and $\ddot{M}_{\Psi\Lambda} = -\frac{1}{N}\Sigma_{2ee}^{-1}\Psi\Lambda'\Sigma_{1ee}^{-1}$, with both Σ_{1ee} and Σ_{2ee} being $N \times N$ matrices defined as $\Sigma_{1ee} = \text{diag}(\sigma_{11}^2, \sigma_{12}^2, \ldots, \sigma_{1N}^2)$ and $\Sigma_{2ee} = \text{diag}(\sigma_{21}^2, \sigma_{22}^2, \ldots, \sigma_{2N}^2)$. In addition, $\ddot{Y}_1$ is an $N \times T$ matrix whose (i,t)th entry is $\ddot{y}_{1it}$; the $N \times T$ matrices $\ddot{Y}_2, \dot{Y}_1, \dot{Y}_2$ are defined similarly.

Then, Φ is a 4×4 symmetric matrix whose diagonal elements and upper diagonal elements are denoted by Φ_{ij} defined as follows:

$$
\Phi_{pp} = -\Big[\frac{2}{N}\sum_{i=1}^N (\mathbb{G}_{p,ii})^2\Big] + \frac{1}{N}\text{tr}\Big[\mathbb{G}_p^2\Big], \qquad \text{for} \quad p = 1, 2, 3, 4
$$

$$
\Phi_{12} = \frac{1}{N}\text{tr}[W_2\mathbb{G}_3 W_1\mathbb{G}_4], \qquad \Phi_{13} = -\frac{2}{N}\sum_{i=1}^N(\mathbb{G}_{1,ii}\mathbb{G}_{3,ii}) + \frac{1}{N}\text{tr}[\mathbb{G}_3\mathbb{G}_1]
$$

$$
\Phi_{14} = \frac{1}{N}\text{tr}[B_{12}P_1^{-1}W_1\mathbb{G}_4], \qquad \Phi_{24} = -\frac{2}{N}\sum_{i=1}^N(\mathbb{G}_{2,ii}\mathbb{G}_{4,ii}) + \frac{1}{N}\text{tr}[\mathbb{G}_4\mathbb{G}_2]
$$

$$
\Phi_{23} = \frac{1}{N}\text{tr}[B_{21}P_2^{-1}W_2\mathbb{G}_3], \qquad \Phi_{34} = \frac{1}{N}\text{tr}[B_{12}P_1^{-1}B_{21}P_2^{-1}]
$$

where those $N\times N$ matrices $\mathbb{G}_p$ for $p = 1, 2, 3, 4$ are defined in Table A1 in Notation Appendix I, with its (i,j)th entry being denoted by $\mathbb{G}_{p,ij}$; those $N \times N$ matrices P_1, P_2, B_{12}, B_{21} are defined in Assumption F.

The $4 \times k$ matrix ϑ is defined as:

$$
\vartheta = [\vartheta_a \quad \vartheta_b]
$$

with ϑ_a being a $4 \times k_1$ matrix:

$$
\vartheta_a = \frac{1}{NT}
\begin{bmatrix}
\text{tr}[\ddot{Y}_1'\ddot{M}_\Lambda\dot{X}_{11}M_F] & \text{tr}[\ddot{Y}_1'\ddot{M}_\Lambda\dot{X}_{12}M_F] & \ldots & \text{tr}[\ddot{Y}_1'\ddot{M}_\Lambda\dot{X}_{1k_1}M_F] \\
\text{tr}[\ddot{Y}_2'\ddot{M}_{\Psi\Lambda}\dot{X}_{11}M_F] & \text{tr}[\ddot{Y}_2'\ddot{M}_{\Psi\Lambda}\dot{X}_{12}M_F] & \ldots & \text{tr}[\ddot{Y}_2'\ddot{M}_{\Psi\Lambda}\dot{X}_{1k_1}M_F] \\
\text{tr}[\dot{Y}_2'\ddot{M}_\Lambda\dot{X}_{11}M_F] & \text{tr}[\dot{Y}_2'\ddot{M}_\Lambda\dot{X}_{12}M_F] & \ldots & \text{tr}[\dot{Y}_2'\ddot{M}_\Lambda\dot{X}_{1k_1}M_F] \\
\text{tr}[\dot{Y}_1'\ddot{M}_{\Psi\Lambda}\dot{X}_{11}M_F] & \text{tr}[\dot{Y}_1'\ddot{M}_{\Psi\Lambda}\dot{X}_{12}M_F] & \ldots & \text{tr}[\dot{Y}_1'\ddot{M}_{\Psi\Lambda}\dot{X}_{1k_1}M_F]
\end{bmatrix}
$$

and ϑ_b being a $4 \times k_2$ matrix:

$$\vartheta_b = \frac{1}{NT} \begin{bmatrix} \mathrm{tr}[\ddot{Y}_1' \ddot{M}_{\Lambda\Psi} \dot{X}_{21} M_F] & \mathrm{tr}[\ddot{Y}_1' \ddot{M}_{\Lambda\Psi} \dot{X}_{22} M_F] & \ldots & \mathrm{tr}[\ddot{Y}_1' \ddot{M}_{\Lambda\Psi} \dot{X}_{2K_2} M_F] \\ \mathrm{tr}[\ddot{Y}_2' \ddot{M}_{\Psi} \dot{X}_{21} M_F] & \mathrm{tr}[\ddot{Y}_2' \ddot{M}_{\Psi} \dot{X}_{22} M_F] & \ldots & \mathrm{tr}[\ddot{Y}_2' \ddot{M}_{\Psi} \dot{X}_{2k_2} M_F] \\ \mathrm{tr}[\dot{Y}_2' \ddot{M}_{\Lambda\Psi} \dot{X}_{21} M_F] & \mathrm{tr}[\dot{Y}_2' \ddot{M}_{\Lambda\Psi} \dot{X}_{22} M_F] & \ldots & \mathrm{tr}[\dot{Y}_2' \ddot{M}_{\Lambda\Psi} \dot{X}_{2k_2} M_F] \\ \mathrm{tr}[\dot{Y}_1' \ddot{M}_{\Psi} \dot{X}_{21} M_F] & \mathrm{tr}[\dot{Y}_1' \ddot{M}_{\Psi} \dot{X}_{22} M_F] & \ldots & \mathrm{tr}[\dot{Y}_1' \ddot{M}_{\Psi} \dot{X}_{2k_2} M_F] \end{bmatrix}$$

where $\dot{X}_{1p}$ is a $N \times T$ matrix whose (i,t)th entry is $\dot{x}_{1it,p}$ for $p = 1, 2, \ldots, k_1$; $\dot{X}_{2q}$ is defined similarly for $q = 1, 2, \ldots, k_2$.

The $k \times k$ matrix $\mathbb{D}_\beta$ is defined as

$$\mathbb{D}_\beta = \begin{bmatrix} \mathbb{D}_\beta^a & \mathbb{D}_\beta^b \\ \mathbb{D}_\beta^{b'} & \mathbb{D}_\beta^c \end{bmatrix}$$

where the $k_1 \times k_1$ matrix $\mathbb{D}_\beta^a$ is defined as

$$\mathbb{D}_\beta^a = \frac{1}{NT} \begin{bmatrix} \mathrm{tr}[\dot{X}_{11}' \ddot{M}_\Lambda \dot{X}_{11} M_F] & \mathrm{tr}[\dot{X}_{11}' \ddot{M}_\Lambda \dot{X}_{12} M_F] & \ldots & \mathrm{tr}[\dot{X}_{11}' \ddot{M}_\Lambda \dot{X}_{1k_1} M_F] \\ \mathrm{tr}[\dot{X}_{12}' \ddot{M}_\Lambda \dot{X}_{11} M_F] & \mathrm{tr}[\dot{X}_{12}' \ddot{M}_\Lambda \dot{X}_{12} M_F] & \ldots & \mathrm{tr}[\dot{X}_{12}' \ddot{M}_\Lambda \dot{X}_{1k_1} M_F] \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{tr}[\dot{X}_{1k_1}' \ddot{M}_\Lambda \dot{X}_{11} M_F] & \mathrm{tr}[\dot{X}_{1k_1}' \ddot{M}_\Lambda \dot{X}_{12} M_F] & \ldots & \mathrm{tr}[\dot{X}_{1k_1}' \ddot{M}_\Lambda \dot{X}_{1k_1} M_F] \end{bmatrix}$$

and the $k_1 \times k_2$ matrix $\mathbb{D}_\beta^b$ is defined as

$$\mathbb{D}_\beta^b = \frac{1}{NT} \begin{bmatrix} \mathrm{tr}[\dot{X}_{11}' \ddot{M}_{\Lambda\Psi} \dot{X}_{21} M_F] & \mathrm{tr}[\dot{X}_{11}' \ddot{M}_{\Lambda\Psi} \dot{X}_{22} M_F] & \ldots & \mathrm{tr}[\dot{X}_{11}' \ddot{M}_{\Lambda\Psi} \dot{X}_{2k_2} M_F] \\ \mathrm{tr}[\dot{X}_{12}' \ddot{M}_{\Lambda\Psi} \dot{X}_{21} M_F] & \mathrm{tr}[\dot{X}_{12}' \ddot{M}_{\Lambda\Psi} \dot{X}_{22} M_F] & \ldots & \mathrm{tr}[\dot{X}_{12}' \ddot{M}_{\Lambda\Psi} \dot{X}_{2k_2} M_F] \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{tr}[\dot{X}_{1k_1}' \ddot{M}_{\Lambda\Psi} \dot{X}_{21} M_F] & \mathrm{tr}[\dot{X}_{1k_1}' \ddot{M}_{\Lambda\Psi} \dot{X}_{22} M_F] & \ldots & \mathrm{tr}[\dot{X}_{1k_1}' \ddot{M}_{\Lambda\Psi} \dot{X}_{2k_2} M_F] \end{bmatrix}$$

and the $k_2 \times k_2$ matrix $\mathbb{D}_\beta^c$ is defined as

$$\mathbb{D}_\beta^c = \frac{1}{NT} \begin{bmatrix} \mathrm{tr}[\dot{X}_{21}' \ddot{M}_\Psi \dot{X}_{21} M_F] & \mathrm{tr}[\dot{X}_{21}' \ddot{M}_\Psi \dot{X}_{22} M_F] & \ldots & \mathrm{tr}[\dot{X}_{21}' \ddot{M}_\Psi \dot{X}_{2k_2} M_F] \\ \mathrm{tr}[\dot{X}_{22}' \ddot{M}_\Psi \dot{X}_{21} M_F] & \mathrm{tr}[\dot{X}_{22}' \ddot{M}_\Psi \dot{X}_{22} M_F] & \ldots & \mathrm{tr}[\dot{X}_{22}' \ddot{M}_\Psi \dot{X}_{2k_2} M_F] \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{tr}[\dot{X}_{2k_2}' \ddot{M}_\Psi \dot{X}_{21} M_F] & \mathrm{tr}[\dot{X}_{2k_2}' \ddot{M}_\Psi \dot{X}_{22} M_F] & \ldots & \mathrm{tr}[\dot{X}_{2k_2}' \ddot{M}_\Psi \dot{X}_{2k_2} M_F] \end{bmatrix}$$

Appendix C   Some proofs

This appendix only includes the proof of consistency in Proposition 2.1 of QML approach. The proofs of other propositions and theorems in this paper are provided in the Supplementary Material. The symbols introduced in Table A1 and the following table will be used throughout proofs.

Table A8: More symbols

$\hat{H} = (\hat{L}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}\hat{L})^{-1}$	$\hat{H}_N = N \cdot \hat{H}$
$\hat{G} = (I_r + \hat{L}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}\hat{L})^{-1}$	$\hat{G}_N = N \cdot \hat{G}$
$\hat{J} = \Upsilon(\eta) - \Upsilon(\hat{\eta})$	$S_N = \Upsilon(\eta)^{-1}$

From $(A+B)^{-1} = A^{-1} - A^{-1}B(A+B)^{-1}$, I have $\hat{H} = \hat{G}(I_r - \hat{G})^{-1}$ and $\hat{H} + \hat{G} = \hat{H}\hat{G} = \hat{G}\hat{H}$. From $\Sigma_{zz} = LL' + \Sigma_{\epsilon\epsilon}$, I have

$$\Sigma_{zz}^{-1} = \Sigma_{\epsilon\epsilon}^{-1} - \Sigma_{\epsilon\epsilon}^{-1}L(I_r + L'\Sigma_{\epsilon\epsilon}^{-1}L)^{-1}L'\Sigma_{\epsilon\epsilon}^{-1} \tag{C.1}$$

The above formulas will be used frequently throughout the appendix.

While in the main text, I use $(\delta, L, \Sigma_{\epsilon\epsilon})$ to denote the true value of the coefficients. For proving consistency, I shall use a superscript "*" to denote the true values of parameters; the variables without "*" denote the input variables of the likelihood function. This notation is only used in Appendix A. Proofs of all the following lemmas are provided in the Supplementary Material.

Lemma A.1. *Let $\eta = (\rho_1, \rho_2, \gamma_1, \gamma_2)$ and $\Upsilon(\eta)$ be a $2N \times 2N$ matrix, with its (i, j)th block, a 2×2 matrix, equal to*

$$\Upsilon_{ij}(\eta) = \begin{cases} \begin{bmatrix} 1 & -\gamma_1 \\ -\gamma_2 & 1 \end{bmatrix} & \text{if } i = j \\ \begin{bmatrix} -\rho_1 w_{1ij} & 0 \\ 0 & -\rho_2 w_{2ij} \end{bmatrix} & \text{if } i \neq j \end{cases} \tag{A.1}$$

Then I have $\det(D(\delta)) = \det(\Upsilon(\eta))$.

Lemma A.2. *Let $V(\delta)$ be the inverse matrix of $D(\delta)$, then its (i, j)th block, a $(k+2) \times (k+2)$*

matrix, denoted as $V_{ij}(\delta)$ has a closed form, which is equal to

$$
V_{ij}(\delta) = \begin{cases} \begin{bmatrix} F_{ii} & F_{ii}\beta' \\ 0 & I_k \end{bmatrix} & \text{if } i = j \\[2em] \begin{bmatrix} F_{ij} & F_{ij}\beta' \\ 0 & 0_{k \times k} \end{bmatrix} & \text{if } i \neq j \end{cases} \tag{A.2}
$$

where $\beta = \begin{bmatrix} \beta_1 & 0 \\ 0 & \beta_2 \end{bmatrix}$ and F_{ij} is the (i,j)th 2×2 block of the inverse matrix of $\Upsilon(\eta)$.

Lemma A.3. *Let $R = \Upsilon(\eta)(\Upsilon(\eta^*))^{-1}$, which is a $2N \times 2N$ matrix, then I can specify its (i,j)th block, a 2×2 matrix, denoted as R_{ij}, as following. For $i = j$,*

$$
R_{ii} = I_2 - \begin{bmatrix} (\rho_1 - \rho_1^*)\mathbb{G}_{1,ii} + (\gamma_1 - \gamma_1^*)\mathbb{G}_{3,ii} & (\rho_1 - \rho_1^*)(W_1\mathbb{G}_4)_{ii} + (\gamma_1 - \gamma_1^*)(B_{21}P_2^{-1})_{ii} \\ (\rho_2 - \rho_2^*)(W_2\mathbb{G}_3)_{ii} + (\gamma_2 - \gamma_2^*)(B_{12}P_1^{-1})_{ii} & (\rho_2 - \rho_2^*)\mathbb{G}_{2,ii} + (\gamma_2 - \gamma_2^*)\mathbb{G}_{4,ii} \end{bmatrix}
$$

For $i \neq j$,

$$
R_{ij} = - \begin{bmatrix} (\rho_1 - \rho_1^*)\mathbb{G}_{1,ij} + (\gamma_1 - \gamma_1^*)\mathbb{G}_{3,ij} & (\rho_1 - \rho_1^*)(W_1\mathbb{G}_4)_{ij} + (\gamma_1 - \gamma_1^*)(B_{21}P_2^{-1})_{ij} \\ (\rho_2 - \rho_2^*)(W_2\mathbb{G}_3)_{ij} + (\gamma_2 - \gamma_2^*)(B_{12}P_1^{-1})_{ij} & (\rho_2 - \rho_2^*)\mathbb{G}_{2,ij} + (\gamma_2 - \gamma_2^*)\mathbb{G}_{4,ij} \end{bmatrix}
$$

where W_1 and W_2 are the weights matrices defined in Assumption E, and $\mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_3, \mathbb{G}_4, P_1, P_2, B_{12}$ and B_{21} are defined in Table 1.

Furthermore, let $\mathbb{D} = DD^{*-1}$ with $D^* = D(\delta^*)$, I have

$$
\mathbb{D}_{ij} = \begin{cases} \begin{bmatrix} R_{ii} & R_{ii}\beta^{*\prime} - \beta' \\ 0 & I_k \end{bmatrix} & \text{if } i = j \\[2em] \begin{bmatrix} R_{ij} & R_{ij}\beta^{*\prime} \\ 0 & 0_{k \times k} \end{bmatrix} & \text{if } i \neq j \end{cases}
$$

where $\beta = \begin{bmatrix} \beta_1 & 0 \\ 0 & \beta_2 \end{bmatrix}$ and $\beta^* = \begin{bmatrix} \beta_1^* & 0 \\ 0 & \beta_2^* \end{bmatrix}$ as defined in Lemma A.2; $\mathbb{D}_{ij}$ is the (i,j)th $(k+2) \times (k+2)$ subblock of $\mathbb{D}$ and R_{ij} is the (i,j)th 2×2 subblock of R defined as above.

Lemma A.4. *Let $(\rho_1, \rho_2, \gamma_1, \gamma_2, \beta_1', \beta_2')' \in \mathcal{A}_1 \times \mathcal{A}_2 \times \mathcal{A}_3 \times \mathcal{A}_4 \times \mathcal{A}_5 \times \mathcal{A}_6$, where $\mathcal{A}_l$ is a compact set for all $l = 1, \ldots, 6$. Under Assumptions A-F, uniformly on $\mathcal{A}_1 \times \mathcal{A}_2 \times \mathcal{A}_3 \times \mathcal{A}_4 \times \mathcal{A}_5 \times \mathcal{A}_6$,*

I have

(a) $\left\| \sum_{j=1}^{N} (\varphi_i^* + \varphi_i^* \beta^*) R_{ij}' - \phi_i^* \beta \right\| \leq C, \quad$ *for all i*;

(b) $\dfrac{1}{N} \sum_{i=1}^{N} \left\| \dfrac{1}{T} \sum_{t=1}^{T} \left(\sum_{j=1}^{N} R_{ij} (e_{jt} + \beta^{*\prime} v_{jt}) - \beta' v_{it} \right) f_t' \right\|^2 = O_p(T^{-1})$,

(c) $\dfrac{1}{N} \sum_{i=1}^{N} \left| \dfrac{1}{T} \sum_{t=1}^{T} [\tilde{e}_{pit}^2 - E(\tilde{e}_{pit}^2)] \right|^2 = O_p(T^{-1}), \quad$ *for $p = 1, 2, 3, 4$*;

(d) $\dfrac{1}{N} \sum_{i=1}^{N} \left| \dfrac{1}{T} \sum_{t=1}^{T} [\tilde{e}_{pit} e_{qit} - E(\tilde{e}_{pit} e_{qit})] \right|^2 = O_p(T^{-1}), \quad$ *for $(p, q) = (1,1), (3,1), (2,2), (4,2)$*;

(e) $\dfrac{1}{N} \sum_{i=1}^{N} \left\| \dfrac{1}{T} \sum_{t=1}^{T} [\tilde{e}_{pit} v_{qit} - E(\tilde{e}_{pit} v_{qit})] \right\|^2 = O_p(T^{-1}), \quad$ *for $(p, q) = (1,1), (3,1), (2,2), (4,2)$*;

(f) $\dfrac{1}{N} \sum_{i=1}^{N} \left| \dfrac{1}{T} \sum_{t=1}^{T} [\tilde{e}_{pit} \tilde{e}_{qit} - E(\tilde{e}_{pit} \tilde{e}_{qit})] \right|^2 = O_p(T^{-1}), \quad$ *for $(p, q) = (1,3), (2,4)$*;

(g) $\dfrac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left| \dfrac{1}{T} \sum_{t=1}^{T} [\tilde{e}_{pit} \tilde{e}_{qjt} - E(\tilde{e}_{pit} \tilde{e}_{qjt})] \right|^2 = O_p(T^{-1}), \quad$ *for $p, q = 1, 2, 3, 4$*;

(h) $\dfrac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left| \dfrac{1}{T} \sum_{t=1}^{T} [\tilde{e}_{pit} e_{qjt} - E(\tilde{e}_{pit} e_{qjt})] \right|^2 = O_p(T^{-1}), \quad$ *for $p = 1, 2, 3, 4$ and $q = 1, 2$*;

(i) $\dfrac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left| \dfrac{1}{T} \sum_{t=1}^{T} [\tilde{e}_{pit} v_{qjt} - E(\tilde{e}_{pit} v_{qjt})] \right|^2 = O_p(T^{-1}), \quad$ *for $p = 1, 2, 3, 4$ and $q = 1, 2$*;

(j) $\dfrac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left| \dfrac{1}{T} \sum_{t=1}^{T} [\tilde{e}_{pit} v_{qjtl} - E(\tilde{e}_{pit} v_{qjtl})] \right|^2 = O_p(T^{-1})$,

$$\text{for } p = 1, 2, 3, 4, \ q = 1, 2 \text{ and } l = 1, 2, \ldots, k_q;$$

where β, β^* are the same as defined in Lemma A.3; $\varphi_i^* = (\lambda_i^*, \psi_i^*)$, $\phi_i^* = (\phi_{1i}^*, \phi_{2i}^*)$; $\tilde{e}_{1it} = \sum_{j=1}^{N} \left[\mathbb{G}_{1,ij}(e_{1jt} + \beta_1^{*\prime} v_{1jt}) + (W_1 \mathbb{G}_4)_{ij}(e_{2jt} + \beta_2^{*\prime} v_{2jt}) \right]$, $\tilde{e}_{2it} = \sum_{j=1}^{N} \left[(W_2 \mathbb{G}_3)_{ij}(e_{1jt} + \beta_1^{*\prime} v_{1jt}) + \mathbb{G}_{2,ij}(e_{2jt} + \beta_2^{*\prime} v_{2jt}) \right]$, $\tilde{e}_{3it} = \sum_{j=1}^{N} \left[\mathbb{G}_{3,ij}(e_{1jt} + \beta_1^{*\prime} v_{1jt}) + (B_{21} P_2^{-1})_{ij}(e_{2jt} + \beta_2^{*\prime} v_{2jt}) \right]$, and $\tilde{e}_{4it} = \sum_{j=1}^{N} \left[(B_{12} P_1^{-1})_{ij}(e_{1jt} + \beta_1^{*\prime} v_{1jt}) + \mathbb{G}_{4,ij}(e_{2jt} + \beta_2^{*\prime} v_{2jt}) \right]$, where the matrices $\mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_3, \mathbb{G}_4, P_1, P_2, B_{12}, B_{21}$ are defined in Table 1.

Lemma A.5. *Under Assumptions A-F,*

$$(a) \quad \sup_{\theta \in \Theta} \left| \frac{1}{N} \text{tr} \left[\mathbb{D} L^* \left(\frac{1}{T} \sum_{t=1}^{T} f_t \epsilon_t' \right) \mathbb{D}' \Sigma_{zz}^{-1} \right] \right| = o_p(1)$$

$$(b) \quad \sup_{\theta \in \Theta} \left| \frac{1}{N} \text{tr} \left[\mathbb{D} \frac{1}{T} \sum_{t=1}^{T} (\epsilon_t \epsilon_t' - \Sigma_{\epsilon\epsilon}^*) \mathbb{D}' \Sigma_{zz}^{-1} \right] \right| = o_p(1)$$

$$(c) \quad \sup_{\theta \in \Theta} \left| \frac{1}{N} \text{tr} \left[\mathbb{D} \bar{\epsilon} \bar{\epsilon}' \mathbb{D}' \Sigma_{zz}^{-1} \right] \right| = o_p(1)$$

where $\mathbb{D} = DD^{*-1}$ *and* $\Sigma_{zz} = LL' + \Sigma_{\epsilon\epsilon}$.

Lemma A.6. *Let* $P(\eta) = \begin{bmatrix} \rho_1 W_1 & \gamma_1 I_N \\ \gamma_2 I_N & \rho_2 W_2 \end{bmatrix}$, $\Delta(\eta) = I_{2N} - P(\eta)$ *and* $\mathcal{R} = \Delta(\eta) \left(\Delta(\eta^*) \right)^{-1}$. *Then I have*

$$\text{tr} \left((R\Sigma_{ee}^* R' - \Sigma_{ee})(R\Sigma_{ee}^* R' - \Sigma_{ee})' \right) = \text{tr} \left((\mathcal{R}\Omega_{ee}^* \mathcal{R}' - \Omega_{ee})(\mathcal{R}\Omega_{ee}^* \mathcal{R}' - \Omega_{ee})' \right)$$

where R is defined in Lemma A.3; $\Omega_{ee} = \begin{bmatrix} \Omega_{1ee} & 0 \\ 0 & \Omega_{2ee} \end{bmatrix}$ *with* $\Omega_{1ee} = \text{diag}(\sigma_{11}^2, \sigma_{12}^2, \dots, \sigma_{1N}^2)$

and $\Omega_{2ee} = \text{diag}(\sigma_{21}^2, \sigma_{22}^2, \dots, \sigma_{2N}^2)$; Ω_{ee}^* *equals* Ω_{ee} *with parameters evaluated at their true values. In addition,*

$$\mathcal{R} = I_{2N} - \begin{bmatrix} (\rho_1 - \rho_1^*)\mathbb{G}_1 + (\gamma_1 - \gamma_1^*)\mathbb{G}_3, & (\rho_1 - \rho_1^*)W_1\mathbb{G}_4 + (\gamma_1 - \gamma_1^*)B_{21}P_2^{-1} \\ (\rho_2 - \rho_2^*)W_2\mathbb{G}_3 + (\gamma_2 - \gamma_2^*)B_{12}P_1^{-1}, & (\rho_2 - \rho_2^*)\mathbb{G}_2 + (\gamma_2 - \gamma_2^*)\mathbb{G}_4 \end{bmatrix}$$

where matrices P_1, P_2, B_{12}, B_{21} *and* $\mathbb{G}_l$ $(l = 1, 2, 3, 4)$ *are defined in Table A1 with parameters evaluated at true values.*

Proof of Proposition 2.1: Consider the following centered objective function:

$$\begin{aligned} \mathcal{L}(\theta) = \quad &- \frac{1}{2N} \ln |\Sigma_{zz}| + \frac{1}{N} \ln |D| - \frac{1}{2N} \text{tr}[DM_{zz}D'\Sigma_{zz}^{-1}] \\ &+ \frac{1}{2N} \ln |\Sigma_{zz}^*| - \frac{1}{N} \ln |D^*| + \frac{k+2}{2}. \end{aligned} \quad \text{(A.3)}$$

Note that the term $(\frac{1}{2N} \ln |\Sigma_{zz}^*| - \frac{1}{N} \ln |D^*| + \frac{k+2}{2})$ is a constant as it does not depend on any unknown parameters and is for the purpose of centering. By $D^* \dot{z}_t = \Phi^* f_t + \dot{\epsilon}_t$ and the identification condition $\bar{f} = 0$, I have

$$D^* M_{zz} D^{*'} = \Sigma_{zz}^* + L^* \left(\frac{1}{T} \sum_{t=1}^{T} f_t \epsilon_t' \right) + \left(\frac{1}{T} \sum_{t=1}^{T} \epsilon_t f_t' \right) L^{*'} + \frac{1}{T} \sum_{t=1}^{T} (\epsilon_t \epsilon_t' - \Sigma_{\epsilon\epsilon}^*) - \bar{\epsilon}\bar{\epsilon}',$$

where $\Sigma_{zz}^* = L^* L^{*\prime} + \Sigma_{\epsilon\epsilon}^*$ and $\bar{\epsilon} = \frac{1}{T}\sum_{t=1}^T \epsilon_t$. Then I get

$$M_{zz} = D^{*-1}\Sigma_{zz}^* D^{*\prime-1} + \mathbb{R}, \tag{A.4}$$

where

$$\mathbb{R} = D^{*-1}L^*\Big(\frac{1}{T}\sum_{t=1}^T f_t\epsilon_t'\Big)D^{*\prime-1} + D^{*-1}\Big(\frac{1}{T}\sum_{t=1}^T \epsilon_t f_t'\Big)L^{*\prime}D^{*\prime-1}$$
$$+ D^{*-1}\frac{1}{T}\sum_{t=1}^T(\epsilon_t\epsilon_t' - \Sigma_{\epsilon\epsilon}^*)D^{*\prime-1} - D^{*-1}\bar{\epsilon}\bar{\epsilon}'D^{*\prime-1}.$$

Substituting (A.4) into (A.3),

$$\mathcal{L}(\theta) = \mathcal{L}_1(\theta) + \mathcal{L}_2(\theta), \tag{A.5}$$

where

$$\mathcal{L}_1(\theta) = -\frac{1}{2N}\ln|\Sigma_{zz}| + \frac{1}{N}\ln|D| - \frac{1}{2N}\mathrm{tr}[DD^{*-1}\Sigma_{zz}^* D^{*-1\prime}D'\Sigma_{zz}^{-1}]$$
$$+ \frac{1}{2N}\ln|\Sigma_{zz}^*| - \frac{1}{N}\ln|D^*| + \frac{k+2}{2}$$

and

$$\mathcal{L}_2(\theta) = -\frac{1}{2N}\mathrm{tr}[D\mathbb{R}D'\Sigma_{zz}^{-1}]. \tag{A.6}$$

Since $\hat{\theta}$ maximizes $\mathcal{L}(\theta)$, I have $\mathcal{L}(\hat{\theta}) \geq \mathcal{L}(\theta^*)$, implying $\mathcal{L}_1(\hat{\theta}) \geq \mathcal{L}_1(\theta^*) + \mathcal{L}_2(\theta^*) - \mathcal{L}_2(\hat{\theta})$. By Lemma A.5, I have $\sup_{\theta\in\Theta}|\mathcal{L}_2(\theta)| = o_p(1)$, and then $|\mathcal{L}_2(\theta^*) - \mathcal{L}_2(\hat{\theta})| \geq -2\sup_{\theta\in\Theta}|\mathcal{L}_2(\theta)| = -|o_p(1)|$. Given this result, together with $\mathcal{L}_1(\theta^*) = 0$, I have

$$\mathcal{L}_1(\hat{\theta}) \geq -|o_p(1)|. \tag{A.7}$$

With notation $\hat{\mathbb{D}} = \hat{D}D^{*-1}$, I rewrite $\mathcal{L}_1(\hat{\theta})$ as

$$\mathcal{L}_1(\hat{\theta}) = -\frac{1}{2N}\ln|\hat{\Sigma}_{zz}| + \frac{1}{N}\ln|\hat{D}| - \frac{1}{2N}\mathrm{tr}[\hat{\mathbb{D}}\Sigma_{zz}^*\hat{\mathbb{D}}'\hat{\Sigma}_{zz}^{-1}]$$
$$+ \frac{1}{2N}\ln|\Sigma_{zz}^*| - \frac{1}{N}\ln|D^*| + \frac{k+2}{2}$$

With the definition $\hat{\Sigma}_{zz} = \hat{L}\hat{L}' + \hat{\Sigma}_{\epsilon\epsilon}$, I have $|\hat{\Sigma}_{zz}| = |\hat{\Sigma}_{\epsilon\epsilon}| \cdot |I_r + \hat{L}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}\hat{L}|$. Thus,

$$\ln|\hat{\Sigma}_{zz}| = \ln|\hat{\Sigma}_{\epsilon\epsilon}| + \ln|I_r + \hat{L}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}\hat{L}| = \sum_{i=1}^N(\ln|\hat{\Sigma}_{iie}| + \ln|\hat{\Sigma}_{iiv}|) + \ln|I_r + \hat{L}'\hat{\Sigma}_{\epsilon\epsilon}^{-1\prime}\hat{L}|$$

Similarly $\ln|\Sigma_{zz}^*| = \sum_{i=1}^{N}(\ln|\Sigma_{iie}^*|+\ln|\Sigma_{iiv}^*|)+\ln|I_r+L^{*\prime}\Sigma_{\epsilon\epsilon}^{*-1}L^*|$. Notice that $|I_r+L^{*\prime}\Sigma_{\epsilon\epsilon}^{*-1}L^*| = O_p(N)$, so uniformly on Θ,

$$-\frac{1}{2N}\ln|\hat{\Sigma}_{zz}| + \frac{1}{2N}\ln|\Sigma_{zz}^*| = -\frac{1}{2N}\sum_{i=1}^{N}(\ln|\hat{\Sigma}_{iie}| + \ln|\hat{\Sigma}_{iiv}|) + \frac{1}{2N}\sum_{i=1}^{N}(\ln|\Sigma_{iie}^*| + \ln|\Sigma_{iiv}^*|)$$

$$-\frac{1}{2N}\ln|I_r + \hat{L}'\hat{\Sigma}_{\epsilon\epsilon}^{-1\prime}\hat{L}| + O_p(\frac{\ln(N)}{N})$$

Next consider the term $\frac{1}{2N}\mathrm{tr}[\hat{\mathbb{D}}\Sigma_{zz}^*\hat{\mathbb{D}}'\hat{\Sigma}_{zz}^{-1}]$, which can be written as, in view of $\Sigma_{zz}^* = L^*L^{*\prime} + \Sigma_{\epsilon\epsilon}^*$,

$$\frac{1}{2N}\mathrm{tr}[\hat{\mathbb{D}}\Sigma_{zz}^*\hat{\mathbb{D}}'\hat{\Sigma}_{zz}^{-1}] = \frac{1}{2N}\mathrm{tr}[\hat{\mathbb{D}}\Sigma_{\epsilon\epsilon}^*\hat{\mathbb{D}}'\hat{\Sigma}_{zz}^{-1}] + \frac{1}{2N}\mathrm{tr}[\hat{\mathbb{D}}L^*L^{*\prime}\hat{\mathbb{D}}'\hat{\Sigma}_{zz}^{-1}] \triangleq i_1 + i_2, \quad \text{say.}$$

By the Woodbury formula $\hat{\Sigma}_{zz}^{-1} = \hat{\Sigma}_{\epsilon\epsilon}^{-1} - \hat{\Sigma}_{\epsilon\epsilon}^{-1}\hat{L}\hat{G}\hat{L}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}$, where $\hat{G} = (I_r + \hat{L}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}\hat{L})^{-1}$, i_1 can be written as

$$i_1 = \frac{1}{2N}\mathrm{tr}[\hat{\mathbb{D}}\Sigma_{\epsilon\epsilon}^*\hat{\mathbb{D}}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}] - \frac{1}{2N}\mathrm{tr}[\hat{\mathbb{D}}\Sigma_{\epsilon\epsilon}^*\hat{\mathbb{D}}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}\hat{L}\hat{G}\hat{L}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}] \triangleq i_3 - i_4, \quad \text{say}$$

With the definition of $\mathbb{D}$ and calculation, I get

$$i_3 = \frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathrm{tr}\left(\hat{R}_{ij}\Sigma_{jje}^*\hat{R}_{ij}'\hat{\Sigma}_{iie}^{-1}\right) + \frac{1}{2N}\sum_{i=1}^{N}\mathrm{tr}\left((\hat{R}_{ii}\beta^{*\prime} - \hat{\beta}')\Sigma_{iiv}^*(\hat{R}_{ii}\beta^{*\prime} - \hat{\beta}')'\hat{\Sigma}_{iiv}^{-1}\right)$$

$$+ \frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\mathrm{tr}\left(\hat{R}_{ij}\beta^{*\prime}\Sigma_{jjv}^*\beta^*\hat{R}_{ij}'\hat{\Sigma}_{iie}^{-1}\right) + \frac{1}{2N}\mathrm{tr}\left[\sum_{i=1}^{N}\Sigma_{iiv}^*\hat{\Sigma}_{iiv}^{-1}\right]$$

where $\hat{R}_{ij}$ is the (i,j)-th subblock $\hat{R}$. Now we show $i_4 = o_p(1)$ uniformly on Θ. To see this, by the boundedness of $\hat{\Sigma}_{ii}$ and Σ_{ii}^*, $\hat{\mathbb{D}}\Sigma_{\epsilon\epsilon}^*\hat{\mathbb{D}}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}$ is less than[26] $C_1\hat{\mathbb{D}}\hat{\mathbb{D}}'$ for some C_1, which is further less than $C_1 C_2 I_{N(k+2)}$ for some constant C_2, as shown in the proof of Lemma A.5 (c). This result leads to $i_4 \leq C_1 C_2 \frac{1}{2N}\mathrm{tr}[\hat{L}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}\hat{L}\hat{G}] = O_p(N^{-1})$.

Given the above results, together with the fact that $\ln|\hat{D}| - \ln|D^*| = \ln|\hat{D}D^{*-1}| = \ln|\hat{\mathbb{D}}| = \ln|\hat{R}|$, I can rewrite the $\mathcal{L}_1(\theta)$ as

$$\mathcal{L}_1(\hat{\theta}) = -\frac{1}{2N}\sum_{i=1}^{N}(\ln|\hat{\Sigma}_{iie}| + \ln|\hat{\Sigma}_{iiv}|) + \frac{1}{2N}\sum_{i=1}^{N}(\ln|\Sigma_{iie}^*| + \ln|\Sigma_{iiv}^*|) - \frac{1}{2N}\ln|I_r + \hat{L}'\hat{\Sigma}_{\epsilon\epsilon}^{-1\prime}\hat{L}|$$

$$-\frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathrm{tr}\left(\hat{R}_{ij}\Sigma_{jje}^*\hat{R}_{ij}'\hat{\Sigma}_{iie}^{-1}\right) - \frac{1}{2N}\sum_{i=1}^{N}\mathrm{tr}\left((\hat{R}_{ii}\beta^{*\prime} - \hat{\beta}')\Sigma_{iiv}^*(\hat{R}_{ii}\beta^{*\prime} - \hat{\beta}')'\hat{\Sigma}_{iiv}^{-1}\right)$$

---

[26]For matrices A and B, I say $A \leq B$ if $B - A$ is a semi-definite positive matrix.

65

$$-\frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\text{tr}\left(\hat{R}_{ij}\beta^{*\prime}\Sigma^*_{jjv}\beta^*\hat{R}'_{ij}\hat{\Sigma}^{-1}_{iie}\right) - \frac{1}{2N}\text{tr}\left[\sum_{i=1}^{N}\Sigma^*_{iiv}\hat{\Sigma}^{-1}_{iiv}\right]$$

$$-\frac{1}{2N}\text{tr}[\hat{\mathbb{D}}L^*L^{*\prime}\hat{\mathbb{D}}'\hat{\Sigma}^{-1}_{zz}] + \frac{1}{N}\ln|\hat{R}| + \frac{k+2}{2} \geq -|o_p(1)| \tag{A.8}$$

Notice that the term $-\frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1}^{N}\text{tr}\left(\hat{R}_{ij}\Sigma^*_{jje}\hat{R}'_{ij}\hat{\Sigma}^{-1}_{iie}\right) + \frac{1}{N}\ln|\hat{R}|$ is equivalent to

$$-\frac{1}{2N}\text{tr}[\hat{R}\Sigma^*_{ee}\hat{R}'\hat{\Sigma}^{-1}_{ee}] + \frac{1}{2N}\ln|\hat{R}\Sigma^*_{ee}\hat{R}'\hat{\Sigma}^{-1}_{ee}| + \frac{1}{2N}\sum_{i=1}^{N}(\ln|\hat{\Sigma}_{iie}| - \ln|\Sigma^*_{iie}|) \tag{A.9}$$

Substituting (A.9) into (A.8), I can rewrite $\mathcal{L}_1(\hat{\theta})$ as

$$\mathcal{L}_1(\hat{\theta}) = -\left\{\frac{1}{2N}\text{tr}[\hat{R}\Sigma^*_{ee}\hat{R}'\hat{\Sigma}^{-1}_{ee}] - \frac{1}{2N}\ln|\hat{R}\Sigma^*_{ee}\hat{R}'\hat{\Sigma}^{-1}_{ee}| - 1\right\} - \left\{\frac{1}{2N}\text{tr}[\hat{\mathbb{D}}L^*L^{*\prime}\hat{\mathbb{D}}'\hat{\Sigma}^{-1}_{zz}]\right\}$$

$$-\left\{\frac{1}{2N}\sum_{i=1}^{N}\text{tr}\left((\hat{R}_{ii}\beta^{*\prime} - \hat{\beta}')\Sigma^*_{iiv}(\hat{R}_{ii}\beta^{*\prime} - \hat{\beta}')'\hat{\Sigma}^{-1}_{iiv}\right)\right\} - \left\{\frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\text{tr}\left(\hat{R}_{ij}\beta^{*\prime}\Sigma^*_{jjv}\beta^*\hat{R}'_{ij}\hat{\Sigma}^{-1}_{iie}\right)\right\}$$

$$-\left\{\frac{1}{2N}\sum_{i=1}^{N}\left(\text{tr}[\Sigma^*_{iiv}\hat{\Sigma}^{-1}_{iiv}] - \ln|\Sigma^*_{iiv}\hat{\Sigma}^{-1}_{iiv}| - k\right)\right\} - \left\{\frac{1}{2N}\ln|I_r + \hat{L}'\hat{\Sigma}^{-1}_{\epsilon\epsilon}\hat{L}|\right\} \geq -|o_p(1)|$$

In the above equation, all the expressions in the braces are non-negative, so each expression must be $o_p(1)$. From the first five expressions, I have

$$\frac{1}{2N}\text{tr}[\hat{R}\Sigma^*_{ee}\hat{R}'\hat{\Sigma}^{-1}_{ee}] - \frac{1}{2N}\ln|\hat{R}\Sigma^*_{ee}\hat{R}'\hat{\Sigma}^{-1}_{ee}| - 1 = o_p(1) \tag{A.10}$$

$$\frac{1}{2N}\sum_{i=1}^{N}\text{tr}\left((\hat{R}_{ii}\beta^{*\prime} - \hat{\beta}')\Sigma^*_{iiv}(\hat{R}_{ii}\beta^{*\prime} - \hat{\beta}')'\hat{\Sigma}^{-1}_{iiv}\right) = o_p(1) \tag{A.11}$$

$$\frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\text{tr}\left(\hat{R}_{ij}\beta^{*\prime}\Sigma^*_{jjv}\beta^*\hat{R}'_{ij}\hat{\Sigma}^{-1}_{iie}\right) = o_p(1) \tag{A.12}$$

$$\frac{1}{2N}\sum_{i=1}^{N}\left(\text{tr}[\Sigma^*_{iiv}\hat{\Sigma}^{-1}_{iiv}] - \ln|\Sigma^*_{iiv}\hat{\Sigma}^{-1}_{iiv}| - k\right) = o_p(1) \tag{A.13}$$

$$\frac{1}{2N}\text{tr}[\hat{\mathbb{D}}L^*L^{*\prime}\hat{\mathbb{D}}'\hat{\Sigma}^{-1}_{zz}] = o_p(1) \tag{A.14}$$

First consider (A.12). For $\beta_1^* \neq 0$ and $\beta_2^* \neq 0$, by definition of $\beta^* = (\beta_1', \beta_2')'$ and the boundedness of Σ^*_{iiv} and $\hat{\Sigma}_{iie}$, there exists a positive constant c such that $\beta^{*\prime}\Sigma^*_{jjv}\beta^* > c$ and

$\hat{\Sigma}_{iie}^{-1} > cI_2$. Then together with (A.12), I have

$$o_p(1) = \frac{1}{2N} \sum_{i=1}^{N} \sum_{j=1,j\neq i}^{N} \text{tr}\Big(\hat{R}_{ij}\beta^{*\prime}\Sigma_{jjv}^{*}\beta^{*}\hat{R}_{ij}'\hat{\Sigma}_{iie}^{-1}\Big) > \frac{c^2}{2N} \sum_{i=1}^{N} \sum_{j=1,j\neq i}^{N} \text{tr}\Big(\hat{R}_{ij}\hat{R}_{ij}'\Big) > 0$$

which implies that

$$\frac{1}{2N} \sum_{i=1}^{N} \sum_{j=1,j\neq i}^{N} \text{tr}\Big(\hat{R}_{ij}\hat{R}_{ij}'\Big) = o_p(1) \tag{A.15}$$

By the expressions of R_{ij} in Lemma A.3, I have

$$\begin{aligned}
o_p(1) =& \frac{1}{2N} \sum_{i=1}^{N} \sum_{j=1,j\neq i}^{N} \text{tr}\Big(\hat{R}_{ij}\hat{R}_{ij}'\Big) \\
=& (\hat{\rho}_1 - \rho_1^*)^2 a_1 + 2(\hat{\rho}_1 - \rho_1^*)(\hat{\gamma}_1 - \gamma_1^*)b_1 + (\hat{\gamma}_1 - \gamma_1^*)^2 c_1 \\
& + (\hat{\rho}_2 - \rho_2^*)^2 a_2 + 2(\hat{\rho}_2 - \rho_2^*)(\hat{\gamma}_2 - \gamma_2^*)b_2 + (\hat{\gamma}_2 - \gamma_2^*)^2 c_2 \\
=& (\hat{\eta} - \eta^*)'\mathbb{M}_a(\hat{\eta} - \eta^*)
\end{aligned} \tag{A.16}$$

where the 4×4 matrix $\mathbb{M}_a$ is defined as in Assumption G.1, and a_p, b_p, c_p for $(p = 1, 2)$ are all scalars, and their definitions are given in Table A2. Based on the above equation, together with Assumption (G.1), I have the consistency $\hat{\eta} - \eta^* = o_p(1)$.

Next consider (A.10), which can be written as

$$\frac{1}{2N}\text{tr}[\hat{\Sigma}_{ee}^{-1/2}\hat{R}\Sigma_{ee}^{*}\hat{R}'\hat{\Sigma}_{ee}^{-1/2}] - \frac{1}{2N}\ln|\hat{\Sigma}_{ee}^{-1/2}\hat{R}\Sigma_{ee}^{*}\hat{R}'\hat{\Sigma}_{ee}^{-1/2}| - 1 = o_p(1)$$

Let l_i $(i = 1, 2, \ldots, 2N)$ denote the eigenvalues of the $2N \times 2N$ matrix $\hat{\Sigma}_{ee}^{-1/2}\hat{R}\Sigma_{ee}^{*}\hat{R}'\hat{\Sigma}_{ee}^{-1/2}$. By the boundedness of $\hat{\eta}$ and $\hat{\Sigma}_{iie}$, there exits some large constant C such that $l_i \in [C^{-1}, C]$ for all i. Together with the fact that $x - \ln x - 1 \geq \frac{1}{4C^2}(x-1)^2$ for all $x \in [C^{-1}, C]$, I have

$$\begin{aligned}
o_p(1) =& \frac{1}{2N}\text{tr}[\hat{\Sigma}_{ee}^{-1/2}\hat{R}\Sigma_{ee}^{*}\hat{R}'\hat{\Sigma}_{ee}^{-1/2}] - \frac{1}{2N}\ln|\hat{\Sigma}_{ee}^{-1/2}\hat{R}\Sigma_{ee}^{*}\hat{R}'\hat{\Sigma}_{ee}^{-1/2}| - 1 \\
=& \frac{1}{2N}\sum_{i=1}^{2N}(l_i - \ln l_i - 1) \geq \frac{1}{4C^2}\frac{1}{2N}\sum_{i=1}^{2N}(l_i - 1)^2 = \frac{1}{4C^2}\frac{1}{2N}\|\hat{\Sigma}_{ee}^{-1/2}\hat{R}\Sigma_{ee}^{*}\hat{R}'\hat{\Sigma}_{ee}^{-1/2} - I_{2N}\|^2
\end{aligned}$$

implying

$$\frac{1}{2N}\|\hat{\Sigma}_{ee}^{-1/2}\hat{R}\Sigma_{ee}^{*}\hat{R}'\hat{\Sigma}_{ee}^{-1/2} - I_{2N}\|^2 = o_p(1)$$

which is equivalent to

$$\frac{1}{2N}\text{tr}\Big[(\hat{\Sigma}_{ee}^{-1/2}\hat{R}\Sigma_{ee}^{*}\hat{R}'\hat{\Sigma}_{ee}^{-1/2} - I_{2N})(\hat{\Sigma}_{ee}^{-1/2}\hat{R}\Sigma_{ee}^{*}\hat{R}'\hat{\Sigma}_{ee}^{-1/2} - I_{2N})'\Big] = o_p(1)$$

can be further written as

$$\frac{1}{2N}\text{tr}\left[\hat{\Sigma}_{ee}^{-1/2}(\hat{R}\Sigma_{ee}^*\hat{R}' - \hat{\Sigma}_{ee})\hat{\Sigma}_{ee}^{-1}(\hat{R}\Sigma_{ee}^*\hat{R}' - \hat{\Sigma}_{ee})\hat{\Sigma}_{ee}^{-1/2}\right] = o_p(1)$$

By the boundedness of $\hat{\Sigma}_{iie}$, there exists come constant c such that $\hat{\Sigma}_{ee}^{-1/2} \geq cI_{2N}$. Then

$$o_p(1) = \frac{1}{2N}\text{tr}\left[\hat{\Sigma}_{ee}^{-1/2}(\hat{R}\Sigma_{ee}^*\hat{R}' - \hat{\Sigma}_{ee})\hat{\Sigma}_{ee}^{-1}(\hat{R}\Sigma_{ee}^*\hat{R}' - \hat{\Sigma}_{ee})\hat{\Sigma}_{ee}^{-1/2}\right]$$

$$\geq c^4\frac{1}{2N}\text{tr}\left[(\hat{R}\Sigma_{ee}^*\hat{R}' - \hat{\Sigma}_{ee})(\hat{R}\Sigma_{ee}^*\hat{R}' - \hat{\Sigma}_{ee})\right] = c^4\frac{1}{2N}\|\hat{R}\Sigma_{ee}^*\hat{R}' - \hat{\Sigma}_{ee}\|^2 > 0$$

implying

$$\frac{1}{2N}\|\hat{R}\Sigma_{ee}^*\hat{R}' - \hat{\Sigma}_{ee}\|^2 = o_p(1) \tag{A.17}$$

By Lemma A.6, the above equation is equivalent to

$$\frac{1}{2N}\|\hat{\mathcal{R}}\Omega_{ee}^*\hat{\mathcal{R}}' - \hat{\Omega}_{ee}\|^2 = o_p(1) \tag{A.18}$$

Let $\mathcal{U} = \hat{\mathcal{R}}\Omega_{ee}^*\hat{\mathcal{R}}' - \hat{\Omega}_{ee}$, together with Lemma A.6, I have

$$\mathcal{U} = \Omega_{ee}^* - \hat{\Omega}_{ee} + \mathbb{H}$$

where $\mathbb{H} = -\hat{H}\Omega_{ee}^* - \Omega_{ee}^*\hat{H}' + \hat{H}\Omega_{ee}^*\hat{H}'$, and

$$\hat{H} = \begin{bmatrix} (\hat{\rho}_1 - \rho_1^*)\mathbb{G}_1 + (\hat{\gamma}_1 - \gamma_1^*)\mathbb{G}_3, & (\hat{\rho}_1 - \rho_1^*)W_1\mathbb{G}_4 + (\hat{\gamma}_1 - \gamma_1^*)B_{21}P_2^{-1} \\ (\hat{\rho}_2 - \rho_2^*)W_2\mathbb{G}_3 + (\hat{\gamma}_2 - \gamma_2^*)B_{12}P_1^{-1}, & (\hat{\rho}_2 - \rho_2^*)\mathbb{G}_2 + (\hat{\gamma}_2 - \gamma_2^*)\mathbb{G}_4 \end{bmatrix}$$

where matrices P_1, P_2, B_{12}, B_{21} and $\mathbb{G}_l$ $(l = 1, 2, 3, 4)$ are all evaluated at true parameters. Then (A.18) can be further rewritten as

$$o_p(1) = \text{tr}\left[\left(\Omega_{ee}^* - \hat{\Omega}_{ee} + \text{diag}(\mathbb{H})\right)^2\right] + (\hat{\eta} - \eta^*)'\text{tr}(\mathbb{M})(\hat{\eta} - \eta^*) = m_1 + m_2, \text{say}$$

where the 4×4 matrix $\mathbb{M}$ is defined in Assumption G. Note that both m_1 and m_2 are non-negative, therefore I have $m_1 = o_p(1)$ and $m_2 = o_p(1)$. Combining the result $m_2 = o_p(1)$ and Assumption (G.2) implies that $\hat{\eta} - \eta^* = o_p(1)$, which further implies $\mathbb{H} = o_p(1)$. Plugging these results into $m_1 = o_p(1)$, together with the boundedness of variances, I get the average

consistency for the variances:

$$\frac{1}{N} \sum_{i=1}^{N} (\sigma_{1i}^{*2} - \hat{\sigma}_{1i}^2)^2 = o_p(1)$$

$$\frac{1}{N} \sum_{i=1}^{N} (\sigma_{2i}^{*2} - \hat{\sigma}_{2i}^2)^2 = o_p(1)$$

(A.19)

The consistency of $\hat{\eta} = (\hat{\rho}_1, \hat{\rho}_2, \gamma_1, \gamma_2)$ implies that the 2×2 subblock $\hat{R}_{ii} \to I_2$ for all i. Plug this result into (A.11), I prove the consistency that $\hat{\beta}_1 \to \beta_1^*$ and $\hat{\beta}_2 \to \beta_2^*$. Now consider (A.13) and (A.14), which are the results corresponding to the pure factor structure part. Using a similar way as in Bai and Li (2014a), I can show from (A.13) that

$$\frac{1}{N} \sum_{i=1}^{N} \|\hat{\Sigma}_{iiv} - \Sigma_{iiv}^*\|^2 = o_p(1)$$

(A.20)

Combining the results (A.19) and (A.20), I have

$$\frac{1}{N} \sum_{i=1}^{N} \|\hat{\Sigma}_{ii} - \Sigma_{ii}^*\|^2 = o_p(1)$$

(A.21)

The last claim of Proposition (2.1) can be proved from (A.14) together with NC.1–NC.3 and first order condition of L, using a similar approach as in Bai and Li (2014a). This completes the proof for Proposition 2.1. $\square$

References

Ahn, S. G., Lee, Y. H. and Schmidt, P. (2013): Panel data models with multiple time-varying effects. *Journal of Econometrics*, 174, 1–14.

Anderson, T.W. (2003). *Introduction to Multivariate Statistical Analysis.* Hoboken, New Jersey: John Wiley & Sons.

Allers, M.A. and Elhorst, J.P. (2011) A simultaneous equations model of fiscal policy interactions. *Journal of Regional Science*, 51(2), 271–291.

Anderson, J.E. and van Wincoop, E. (2003) Gravity with Gravitas: a solution to the border puzzle. *American Economic Review*, 93(1), 170–192.

Anselin, L. (1988). *Spatial econometrics: methods and models.* The Netherlands: Kluwer Academic Publishers.

Anselin, L., Le Gallo, J., and Jayet, H. (2008) Spatial Panel Econometrics. Ch. 19 in L. Matyas and P. Sevestre, eds., The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice, Springer-Verlag, Berlin, 625–660.

Arezki, R., Candelon, B., and Sy, A.N.R. (2011) Sovereign rating news and financial markets spillovers: evidence from the European debt crisis. *manuscript.*

Bai, J. (2009) Panel data models with interactive fixed effects. *Econometrica*, 77(4), 1229–1279.

Bai, J. and Ando, T. (2014) Asset pricing with a general multifactor structure. *Journal of Financial Econometrics*, Vol.0(No.0), 1–49.

Bai, J. and Li, K. (2012) Statistical analysis of factor models of high dimension. *The Annals of Statistics*, 40(1), 436–465.

Bai, J. and Li, K. (2014a) Theory and methods of panel data models with interactive effects. *The Annals of Statistics*, 42(1), 142–170.

Bai, J. and Li, K. (2014b) Spatial panel data models with common shocks, *manuscript.*

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

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1), 191–221.

Bai, J. and Ng, S. (2013). Principal components estimation and identification of static factors. *Journal of Econometrics*, 176, 18–29.

Balestra, P. and Varadharajan-Krishnakumar, J. (1987) Full information estimations of a system of simultaneous equations with error component structure, *Econometric Theory*, 3(2), 223–246.

Baltagi, B.H. (1981) Simultaneous equations with error components, *Journal of Econometrics*, 17(2), 603–624.

Baltagi, B.H. and Li, Q. (1992) A note on the estimation of simultaneous equations with error components. *Econometric Theory*, 8(1), 113–119.

Baltagi, B.H. and Li, D. (2006) Prediction in the panel data model with spatial correlation: the case of liquor. *Spatial Economic Analysis*, 1, 175–185.

Baltagi, B.H., Egger, P. and Pfaffermayr, M. (2008) Estimating regional trade agreement effects on FDI in an interdependent world. *Journal of Econometrics*, 145(1), 194–208.

Baltagi, B.H. and Bresson, G. (2011) Maximum likelihood estimation and Lagrange multiplier tests for panel seemingly unrelated regressions with spatial lag and spatial errors: An application to hedonic housing prices in Paris. *Journal of Urban Economics*, 69(1), 24–42.

Baltagi, B.H. (2011) *Spatial Panels.* Chapter 15 in The Handbook of Empirical Economics and Finance, Aman Ullah and David E.A. Giles, editors, Chapman and Hall, pp.435–454.

Baltagi, B.H. and Deng, Y. (2015) EC3SLS Estimator for a Simultaneous System of Spatial Autoregressive Equations with Random Effects. *Econometric Reviews*, 34(6-10), 659–694.

Bradford, C.J. and Chakwin, N. (1993) Alternative explanations of the trade-output correlation in East Asian economies. *OECD* Development Centre Technical Paper No.87, August 1993.

Brock, W.A. and Durlauf, S.N. (2001) Growth empirics and reality. *The World Bank Economic Review*, 15(2), 229–272.

Castagnetti, C., Rossi, E., and Trapani, L. (2015) Inference on factor structures in heterogeneous panels. *Journal of Econometrics*, 184, 145–157.

Chamberlain, G. (1977a) Education, income and ability revisited, in *Latent Variables in Socio-Economic Models*, edited by D.J. Aigner and A.S. Goldberger, pp. 143–161. Amsterdam; North-Holland.

Chamberlain, G. (1977b) An instrumental variable interpretation of identification in variance-components and MIMIC models, in *Kinometrics: Determinants of Social-Economic Success within and between Families*, edited by P.Taubman, pp. 235–254. Amsterdam; North-Holland.

Chamberlain, G. and Griliches, Z. (1975) Unobservables with a variance-components structure: ability, schooling and the economic success of brothers. *International Economic Review*, 16, 422–450.

Chen, C.R., Steiner, T.L., and Whyte, A.M. (2006) Does stock option-based executive compensation induce risk-taking? An analysis of the banking industry. *Journal of Banking and Finance*, 30, 915–945.

Cohen-Cole, E., Liu, X. and Zenou, Y. (2013) Multivariate choice and identification of social interactions. CEPR Discussion Papers 9159, C.E.P.R. Discussion Papers.

Cornwell, C., Schmidt, P., and Wyhowski, D. (1992) Simultaneous equations and panel data. *Journal of Econometrics*, 51, 151–181.

Defever, F., Heid, B., and Larch, M. (2015) Spatial Exporters. *Journal of International Economics*, 95(1), 145–156.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1–38.

Dewachter, H., Houssa, R., and Toffano, P. (2012) Spatial propagation of macroeconomic shocks in Europe. *Review of World Economics*, 148(2), 377–402.

Druska, V. and Horrace, W.C. (2004) Generalized moments estimation for spatial panel data: Indonesian rice farming. *American Journal of Agricultural Economics*, 86, 185–198.

Eberhardt, M., Helmers, C., and Strauss, H. (2013) Do spillovers matter when estimating private returns to R&D?. *Review of Economics and Statistics*, 95(2), 436–448.

Egger, P. and Pfaffermayr, M. (2004) Distance, trade and FDI: a Hausman-Taylor SUR approach. *Journal of Applied Econometrics*, 19, 227–246.

Egger, P., Pfaffermayr, M. and Winner, H. (2005) An unbalanced spatial panel data approach to US state tax competition. *Economics Letters*, 88, 329–335.

Elhorst, P.J. (2003) Specification and estimation of spatial panel data models. *International Regional Science Review*, 26(3), 244–268.

Feyrer, J. (2009) Trade and income –exploiting time series in geography. *NBER*, working paper 14910.

Felbermayr, G. and Groschl, J. (2013) Natural disasters and the effect of trade on income: a new panel approach. *European Economic Review*, 58, 18–30.

Frazier, C. and Kockelman, K.M. (2005) Spatial econometric models for panel data: incorporating spatial and temporal data. *Transportation Research Record: Journal of the Transportation Research Board*, 1902 (2005), 80–90.

Frankel, J. and Romer, D. (1999) Does trade cause growth? *American Economic Review*, 89(3), 379–399.

Gebremariam, G.H., Gebremedhin T.G. and Schaeffer, P.V. (2011) Employment, income and migration in appalachia: a spatial simultaneous equations approach. *Journal of Regional Science*, 51(1), 102–120.

Glaeser, E. L., Sacerdote, B., and Scheinkman, J. A. (1996) Crime and social interactions. *The Quarterly Journal of Economics*, 111(2), 507–548.

Goldsmith-Pinkham, P. and Imbens, W.G. (2013) Social networks and the identification of peer effects. *Journal of Business and Economic Statistics*, 31(3), 253–64.

Griliches, Z. (1979) Sibling models and data in economics: beginning of a survey. *Journal of Political Economy*, 87(Supplement 2), S37–S64.

Hauptmeier, S., Mittermaier, F. and Rincke, J. (2012) Fiscal competition over taxes and public inputs. *Regional Science and Urban Economics*, 42(3), 407–419.

Helpman, E. (1987) Imperfect competition and international trade: evidence from fourteen industrial countries. *Journal of the Japanese and International Economies*, 1, 62–81.

Helpman, E. (1988) Growth, technological progress, and trade. *NBER* (Cambridge, MA), reprint no. 1145.

Ho, C., and Hite, D. (2008) The benefit of environmental improvement in the southeastern United States: Evidence from a simultaneous model of cancer mortality, toxic chemical releases and house values. *Papers in Regional Science*, 87(4), 589–604.

Irwin, D.A., and Tervio, M. (2002) Does trade raise income? Evidence from the twentieth century. *Journal of International Economics*, 58(1), 1–18.

Jain, N., Girotra, K., and Netessine, S. (2013) Managing global sourcing: inventory performance. *Management Science*, 60(5), 1202–1222.

Jeanty, P.W., Partridge, M., and Irwin, E. (2010) Estimation of a spatial simultaneous equation model of population migration and housing price dynamics. *Regional Science and Urban Economics*, 40, 343–352.

Jennrich, R. I. (1969) Asymptotic properties of non-linear least squares estimators. *The Annals of Mathematical Statistics*, 40(2), 633–643.

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

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

Kelejian, H.H. and Prucha I.R. (2004). Estimation of simultaneous systems of spatially interrelated cross sectional equations. *Journal of Econometrics*, 118, 27–50.

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

Keller, W. and Shiue, C. (2007). The origins of spatial interaction: evidence from Chinese rice markets, 1742–1795. *Journal of Econometrics*, 140(1), 304–332.

Kneip, A., Sickles, R.C., and Song, W. (2012). A new panel data treatment for heterogeneity in time trends. *Econometric Theory*, 28, 590–628.

Korte, J., and Steffen, S. (2014). Zero risk contagion -banks' sovereign exposure and sovereign risk spillovers. *manuscript*.

Kuersteiner, G.M., and Prucha, I.R. (2015). Dynamic spatial panel models: networks, common shocks, and sequential exogeneity. *CESIFO* working paper no. 5445.

Lawless, M. (2009) Firm export dynamics and the geography of trade. *Journal of International Economics*, 77(2), 245–254.

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

Lee, L. (2004b) A supplement to "Asymptotic distribution of quasi-maximum likelihood estiators for spatial augoregressive models", *Manuscipt.*

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

Lee, L. and Yu, J. (2010b) A spatial dynamic panel data model with both time and individual fixed effects. *Econometric Theory*, 26, 564–597.

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

Liu, X. (2014) Identification and efficient estimation of simultaneous equations network models. *Journal of Business and Economic Statistics*, 32(4), 516–536.

Liu, C., and Rubin, D. B. (1994) The ECME algorithm: a simple extension of EM and ECM with faster monotone convergence. *Biometrika*, 81(4), 633–648.

McLachlan, G. J., and Krishnan, T. *The EM Algorithm and Extensions.* 1997. New York: Wiley

Moon, H., Shum, M., and Weidner, M. (2014) Interactive Fixed Effects in the BLP Random Coefficients Demand Model, *R&R Journal of Econometrics.*

Moon, H. and Weidner, M. (2009) Likelihood expansion for panel regression models with factors, *manuscipt, University of Southern California.*

Moon, H. and Weidner, M. (2017) Dynamic linear panel regression models with interactive fixed effects, *Econometric Theory*, 33(1), 158–195.

Newey, W. and McFadden, D. (1994) Large Sample Estimation and Hypothesis Testing, in Engle, R.F. and D. McFadden (eds.) *Handbook of Econometrics,* North Holland.

Noguer, M. and Siscart, M. (2005) Trade raises income a precise and robust result. *Journal of International Economics*, 65(2), 447–460.

Ortega, F. and Peri, G. (2014) Openness and income: The roles of trade and migration. *Journal of International Economics*, 92, 231–251.

Ozdagli, A. and Weber, M. (2017) Monetary policy through production networks: evidence from the stock market. *manuscript.*

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

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

Rauch, J.E., and Trindade, V. (2002) Ethnic chinese networks in international trade. *Review of Economics and Statistics*, 84(1), 116–130.

Rodrik, D. (1995) Getting interventions right: how south korea and taiwan grew rich. *Economic Policy*, April, 20, 53–97.

Rodriguez, F., and Rodrik, D. (2001) Trade policy and economic growth: a skeptic's guide to the cross-national evidence. In (B.Bernanke, and K.S.Rogoff, eds.), *Macroeconomics Annual 2000*, 261–324, Cambridge, MA: MIT Press for NBER.

Serlenga, L. and Shin, Y. (2007). Gravity models of intra-EU trade: application of the CCEP-HT estimation in heterogeneous panels with unobserved common time-specific factors. *Journal of Applied Econometrics*, 22, 361–381.

Stock, J.H. and Watson, M.W (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97, 1167–1179.

Stock, J.H. and Watson, M.W (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics*, 20, 147–162.

Wang, L., Li, K., and Wang, Z. (2014) Quasi maximum likelihood estimation for simultaneous spatial autoregressive models, *manuscript*.

Winters, L.A. (2004) Trade liberalisation and economic performance: an overview. *The Economic Journal*, 114, F4–F21.

Winters, L.A. and Masters, A. (2013) Openness and growth: still an open question? *Journal of International Development*, 25, 1061–1070.

Wu, C. (1981) Asymptotic theory of nonlinear least squares estimation. *The Annals of Statistics*, 9(3), 501–513.

Wu, J. and Lin, H. (2010) The effect of the conservation reserve program on land values. *Land Economics*, 86, 1–21.

Yang, K. and Lung-fei Lee (2017) Identification and QML estimation of multivariate and simultaneous spatial autoregressive models, *Journal of Econometrics*, 196(1), 196–214.

Yu, J. and Lee, L. (2010) Estimation of unit root spatial dynamic panel data models. *Econometric Theory*, 26, 1332–1362.

Yu, J., de Jong, R., and Lee, L. (2008). Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large. *Journal of Econometrics*, 146(1), 118–134.