GMM Estimation and Impulse Response Functions of Spatial Dynamic Panel Simultaneous Equations Models with Two Way Fixed Effects

Jiannan Mai* Yiguo Sun†

July 15, 2025

First draft (proofs and Monte Carlo simulations are in progress.)[1]

Abstract

This paper develops a Generalized Method of Moments (GMM) estimation for spatial dynamic panel simultaneous equations models with two-way fixed effects.This model specification permits each dependent variable to employ a distinct spatial weight matrix tailored to its unique transmission channel. We establish the asymptotic properties of the proposed GMM estimator under two asymptotic frameworks: one where both the cross-sectional dimension n and the time dimension T grow large, and another where n is large while T is fixed. Compared with the quasi-maximum likelihood (QML) approach, the GMM estimator is simpler to implement and does not rely heavily on a large T relative to n for consistency. Additionally, this paper investigates impulse response functions along with their associated confidence intervals and examines the estimation

*Department of Economics and Finance, University of Guelph, Guelph, ON, N1G2W1, Canada; Email: jmai@uoguelph.ca.

†Department of Economics and Finance, University of Guelph, Guelph, ON, N1G2W1, Canada; Email: yisun@uoguelph.ca

[1]This article is a preliminary version. It is not yet complete and may contain inaccuracies.

and inferential theory related to average direct, indirect, and total impacts. Monte Carlo simulations are conducted to evaluate the finite-sample performance of the proposed estimators and inferential procedures.

Keywords: Dynamic panels, Spatial simultaneous equations, Generalized method of moment, Impulse response analysis.

JEL Codes:

1 Introduction

Since the introduction of spatial autoregressive (SAR) models by Cliff and Ord (1973), spatial econometrics has gained significant attention and evolved into a fundamental framework for analyzing spatial dependencies in economic variables, both in cross-sectional and panel data settings. To estimate the SAR model, Kelejian and Prucha (1998) and Kelejian and Prucha (1999) developed generalized spatial two-stage least squares (2SLS) and moments-based methods. Theoretical foundations were advanced by Lee (2004) and Lee (2007), who established asymptotic properties of quasi-maximum likelihood (QML) estimators and generalized method of moments (GMM) approaches. Lin and Lee (2010) resolved heteroskedasticity concerns through robust GMM frameworks, enabling reliable inference in empirical applications. Further refining GMM efficiency, Liu et al. (2010) developed an efficient GMM estimator for SAR models with SAR disturbances, demonstrating that it matches ML estimator's efficiency under normality while outperforming it under non-normality.

The increasing availability of panel data has spurred significant advances in spatial panel modeling, which simultaneously captures cross-sectional dependence via spatial interactions and exploits time-varying information. Early developments in dynamic panel modeling established key theoretical results, including the within estimator Nickell (1981), GMM estimaiton Arellano and Bond (1991); Arellano and Bover (1995), and large-n, large-T asymptotics Alvarez and Arellano (2003); Hahn and Kuersteiner (2002). Building on these foundations, spatial panel models extend the framework by incorporating cross-sectional dependence through spatial interactions. The analysis of spatial interactions in panel data settings has evolved from its early developments in the works of Anselin (1988) and Elhorst (2003). Recent methodological advances have systematically addressed various dimensions of complexity in spatial panel modeling. Initial breakthroughs in static models came through Kapoor et al. (2007), who developed method of moments estimation to a spatial panel model with error components, while Baltagi et al. (2007) established critical testing procedures for spatial and serial dependence. The field subsequently progressed to dynamic specifications, with Yu et al. (2008) develops QML estimation for spatial dynamic panel data models (SDPD) with individual fixed effects under large-n and large-T setting. The study establishes asymptotic properties under different relative growth rates of n and T, and proposes a bias correction method that delivers centered inference when T grows sufficiently fast relative to n. Building on this work, Lee and Yu (2010b) extends the framework to incorporate two-way fixed effects. Further advancing the literature, Lee and Yu (2010a) generalizes the theoretical framework to SDPD models with SAR disturbances, addressing the additional complexities arising from spatial dependence in the error structure. Complementing these maximum likelihood approaches, Lee and Yu (2014) develop an efficient GMM framework for SDPD with multiple spatial lags. Their estimator maintains asymptotic consistency and normality when n is large and T is large but small relative to n, overcoming key limitations of QML estimation. Li (2017) proposes a QML estimator for SDPD models with high-order spatial-temporal lags, establishing impulse response analysis and formal inference for dynamic extensions of LeSage and Pace (2009)'s impact measures. Recent work on high-order dynamic spatial panel data models with interactive fixed effects, unknown heteroskedasticity, and endogenous time-varying spatial weights includes Kuersteiner and Prucha (2020), who propose a GMM estimator for short panels where the time dimension is fixed. In contrast, Lia and Yanga (2023) develops a likelihood-based approach that not only covers the fixed-time case but also provides new asymptotic results for settings where T is large but small relative to n.

While the preceding discussion has focused on single-equation spatial models for cross-sectional and panel data, most theoretical work in spatial econometrics relies on a restrictive assumption: that an agent's outcome depends only on its own characteristics and spatially lagged outcomes, while ignoring simultaneous interdependencies with other related activities. In reality, however, many economic phenomena empirically demonstrate such interdependencies—for example, in macroeconomics Elhost and Emili (2022), energy policy Tiba and Belaid (2021), corporate finance Gu (2024), and environmental economics Zhang et al. (2023), among others. Building on the empirical and theoretical motivations for system-equation frameworks, significant methodological advancements have emerged in recent decades, addressing the limitations of single-equation frameworks. Kelejian and Prucha (2004) pioneered the extension of SAR models to simultaneous equations systems, introducing instrumental variable (IV) estimators (including 2SLS and 3SLS) that achieve asymptotic efficiency while demonstrating computational advantages over maximum likelihood approaches. Liu (2014) develops identification and estimation for social network simultaneous equations, using Bonacich centrality as IVs with a many-IV bias correction. QML estimation witnessed parallel innovations. Yang and Lee (2017) established QML estimation for simultaneous SAR models with cross-equation disturbance correlations, proving efficiency gains over IV methods. This was extended to dynamic panel settings by Yang and Lee (2019), who integrated time-lagged effects with spatial and simultaneous dependencies, and by Yang and Lee (2021) through spatial vector autoregressions with cointegration analysis. Liu and Saraiva (2019) designed GMM estimators accommodating heteroskedasticity across equations, while Lütkepohl et al. (2020) developed partial identification frameworks for heteroskedastic systems. Model flexibility expanded along multiple dimensions: Cohen-Cole et al. (2018) formalized multivariate choice interdependencies, Wang et al. (2018) modeled strategic interactions in multi-network systems, and Drukker et al. (2023) incorporated higher-order spatial lags. Panel data extensions include Amba (2021) on three-way error components (individual/time/space) and Lu (2023) on common shock absorption in spatial panels. Collectively, these advances provide rigorously validated tools for modeling complex interdependencies that single-equation frameworks cannot capture.

In this paper, we shall consider the GMM estimation of spatial dynamic panel simultaneous equations models with two-way fixed effects. While maintaining the core structural framework of Yang and Lee (2019) featuring: spatial interdependence, dynamic time effects, space-time diffusion effects, and and cross-equation simultaneity. This model specification permits each dependent variable to employ a distinct spatial weight matrix tailored to its unique transmission channel. For instance, assuming homogeneous spatial structures for epidemiological processes (e.g., disease transmission via human mobility networks) and economic outcomes (e.g., GDP spillovers through trade linkages) would be empirically implausible, as their underlying transmission mechanisms operate on distinct spatial dimensions. Yang and Lee (2019) investigates a QML estimator that requires large T. When T is sufficient large than n, the estimators are consistent and asymptotically normal. When n is asymptotically proportional to T, the estimators are asymptotically biased due to the incidental parameters problem in Neyman and Scott (1948). When $n/T \to 0$, the estimators have a degenerate asymptotic distribution. Although they proposes a bias correction procedure, the method still requires the condition that $n^{1/3}/T \to 0$. To address the large or small T scenarios, they further develop IV-based estimators. This motivates our development of the GMM estimators that accommodate both large n, large T, and large n, fixed T asymptotic settings. We propose both single-equation (equation-by-equation) and system GMM estimators. Compare to QML estimation, the GMM approach possesses four key advantages: (i) asymptotic unbiasedness under small or fixed T scenarios. (ii) more efficient than QML the when the error distribution exhibits excess kurtosis or non-normality. (iii) Flexibility in handling non-row-normalized spatial weight matrices. (iv) Computational simplicity as it avoids compute the determinant of the Jacobian matrix in the likelihood function. In addition, compared to 3SLS, with the design of linear and quadratic moment conditions, the GMM estimator is asymptotically more efficient. Furthermore, we analyze the inference of impulse response function (IRF), establishing asymptotic properties, and study average direct impact (ADI), average indirect impact (AII), and average total effects (ATI).

The remainder of this paper is structured as follows. Section 2 details the model specification and develops the moment conditions. Identification under the GMM framework is established in Section 3. We present asymptotic properties of the GMM estimators, including consistency and limiting distributions in Section 4. Sections 5 and 6 derive IRF, ADI, ATI, and AII with associated asymptotic theory. Monte Carlo simulations are reported in Section 7, followed by concluding remarks in Section 8. Technical proofs and lemmas are compiled in Appendix.

For the convenience of readers, we adopt the following notation. **(i)** $\|\cdot\|$ denotes the Euclidean norm; $\|\cdot\|_2$ denotes the spectral norm; $\|\cdot\|_1$ and $\|\cdot\|_\infty$ denote the column and row sum matrix norms, respectively. **(ii)** Let $A = (a_{ij})_{i,j=1}^n$ be an $n \times n$ matrix. Then $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote its smallest and largest eigenvalues, respectively. **(iii)** Let $\rho(A)$ denote the spectral radius of matrix A. **(iv)** $\text{abs}(A)$ denotes the *element-wise absolute value* of matrix A, i.e., $[\text{abs}(A)]_{ij} = |A_{ij}|$ **(v)** Let A be an $n \times n$ matrix and B be an $m \times m$ matrix. Then $A \oplus B = A \otimes I_m + I_n \otimes B$, where $\otimes$ denotes the Kronecker product

2 The Model and Moment Conditions

2.1 The model

The spatial dynamic panel simultaneous model under investigation is:

$$\begin{aligned} \boldsymbol{Y}_{nm,t}\boldsymbol{\Gamma}_{m0} = {} & [W_{1n}Y_{n1,t}, W_{2n}Y_{n2,t}, \cdots, W_{mn}Y_{nm,t}]\,\boldsymbol{\Psi}_{ms0} + \boldsymbol{Y}_{nm,t-1}\boldsymbol{P}_{ms0} + \boldsymbol{X}_{n,t}\boldsymbol{\Pi}_{ms0} \\ & + [W_{1n}Y_{n1,t-1}, W_{2n}Y_{n2,t-1}, \cdots, W_{mn}Y_{nm,t-1}]\,\boldsymbol{\Phi}_{ms0} + \boldsymbol{C}_{nms0} + \alpha'_{ms0,t} \otimes l_n + \boldsymbol{U}_{nm,t} \end{aligned} \tag{1}$$

for $t = 1, \ldots, T$. This model consists of m equations representing endogenous variables for n spatial units over t time periods, giving a total of $n \times m \times t$ data points. The notation used in the model is as follows:

$\boldsymbol{Y}_{nm,t}$	$n \times m$ matrix of endogenous variables at time t.	$Y_{nl,t}$	$n \times 1$ vector of l-th endogenous variable.
$\boldsymbol{U}_{nm,t}$	$n \times m$ matrix of error terms, assumed to be $i.i.d.(0, \Sigma_{um})$.	$W_{1n}, \ldots, W_{mn}$	$n \times n$ spatial weight matrices.
$\boldsymbol{Y}_{nm,t-1}$	$n \times m$ matrix of time-lagged endogenous variables.	$W_{pn}Y_{np,t-1}$	$n \times 1$, space-time lag.
$\boldsymbol{X}_{n,t}$	$n \times k$ matrix of exogenous variables at time t.	$\boldsymbol{\Phi}_{\mathrm{ms0}}$	$m \times m$ matrix for spatial effects.
$\boldsymbol{\Gamma}_{\mathrm{m0}}$	$m \times m$ matrix with diagonal elements normalized to 1.	Σ_{um0}	Covariance matrix for error terms.
$\boldsymbol{P}_{\mathrm{ms0}}$	$m \times m$ matrix for dynamic time effects.	$\boldsymbol{\Psi}_{\mathrm{ms0}}$	$m \times m$ matrix for spatial diffusion effects.
$\boldsymbol{\Pi}_{\mathrm{ms0}}$	$k \times m$ coefficient matrix for regressors.	$\boldsymbol{C}_{\mathrm{nms0}}$	$n \times m$ matrix of individual effects.
$\alpha'_{\mathrm{ms0,t}} \otimes l_n$	$n \times m$ matrix for time effects.		

where $\boldsymbol{Y}_{nm,t} = [Y_{n1,t}, \ldots, Y_{nm,t}], \boldsymbol{U}_{nm,t} = [U_{n1,t}, \ldots, U_{nm,t}]$, and $\boldsymbol{U}_{nm,t}$ is assumed to be i.i.d.$(0, \Sigma_{um})$ across spatial units $i = 1, \ldots, n$ and time $t = 1, \ldots, T$, but they are allowed to be contemporaneously correlated across variables. Define $\mathcal{W}_{nm} = diag(W_{1n}, \ldots, W_{mn})$, a block diagonal matrix with dimension $nm \times nm$. By vectorization,

$$\begin{aligned}(\boldsymbol{\Gamma}'_{\mathrm{m0}} \otimes I_n)\mathrm{vec}(\boldsymbol{Y}_{nm,t}) &= (\boldsymbol{\Psi}'_{\mathrm{ms0}} \otimes I_n)\mathcal{W}_{nm}\mathrm{vec}(\boldsymbol{Y}_{nm,t}) + (\boldsymbol{P}'_{\mathrm{ms0}} \otimes I_n)\mathrm{vec}(\boldsymbol{Y}_{nm,t-1}) \\ &+ (\boldsymbol{\Pi}'_{\mathrm{ms0}} \otimes I_n)\mathrm{vec}(\boldsymbol{X}_{n,t}) + (\boldsymbol{\Phi}'_{\mathrm{ms0}} \otimes I_n)\mathcal{W}_{nm}\mathrm{vec}(\boldsymbol{Y}_{nm,t-1}) \\ &+ \mathrm{vec}(\boldsymbol{C}_{nms0}) + \alpha_{\mathrm{ms0,t}} \otimes l_n + \mathrm{vec}(\boldsymbol{U}_{nm,t})\end{aligned} \tag{2}$$

and the quasi reduced form model is:

$$\begin{aligned}\mathrm{vec}(\boldsymbol{Y}_{nm,t}) &= (\boldsymbol{\Psi}'_{\mathrm{m0}} \otimes I_n)\mathcal{W}_{nm}\mathrm{vec}(\boldsymbol{Y}_{nm,t}) + (\boldsymbol{P}'_{\mathrm{m0}} \otimes I_n)\mathrm{vec}(\boldsymbol{Y}_{nm,t-1}) \\ &+ (\boldsymbol{\Pi}'_{\mathrm{m0}} \otimes I_n)\mathrm{vec}(\boldsymbol{X}_{n,t}) + (\boldsymbol{\Phi}'_{\mathrm{m0}} \otimes I_n)\mathcal{W}_{nm}\mathrm{vec}(\boldsymbol{Y}_{nm,t-1}) \\ &+ \mathrm{vec}(\boldsymbol{C}_{nm0}) + \alpha_{\mathrm{m0,t}} \otimes l_n + \mathrm{vec}(\boldsymbol{V}_{nm,t})\end{aligned} \tag{3}$$

with $\boldsymbol{\Psi}_{m0} = \boldsymbol{\Psi}_{ms0}\boldsymbol{\Gamma}_{m0}^{-1}, \boldsymbol{\Phi}_{m0} = \boldsymbol{\Phi}_{ms0}\boldsymbol{\Gamma}_{m0}^{-1}, \boldsymbol{P}_{m0} = \boldsymbol{P}_{ms0}\boldsymbol{\Gamma}_{m0}^{-1}, \boldsymbol{\Pi}_{m0} = \boldsymbol{\Pi}_{ms0}\boldsymbol{\Gamma}_{m0}^{-1}, \boldsymbol{C}_{nm0} = \boldsymbol{C}_{nms0}\boldsymbol{\Gamma}_{m0}^{-1}, \alpha_{m0,t} = \alpha_{ms0,t}\boldsymbol{\Gamma}_{m0}^{-1}, \boldsymbol{V}_{nm,t} = \boldsymbol{U}_{nm,t}\boldsymbol{\Gamma}_{m0}^{-1}, \Sigma_{vm0} = \boldsymbol{\Gamma}_{m0}'^{-1}\Sigma_{um0}\boldsymbol{\Gamma}_{m0}^{-1}$. Define $\boldsymbol{S}_{nm0} \equiv \boldsymbol{S}_{nm}(\boldsymbol{\Psi}_{m0}) = I_{nm} - (\boldsymbol{\Psi}_{m0}' \otimes I_n)\mathcal{W}_{nm}, \boldsymbol{H}_{nm0} = \boldsymbol{S}_{nm}^{-1}\left[(\boldsymbol{P}_{m0}' \otimes I_n) + (\boldsymbol{\Phi}_{m0}' \otimes I_n)\mathcal{W}_{nm}\right]$, we can write the reduced form model in:

$$\operatorname{vec}(\boldsymbol{Y}_{nm,t}) = \boldsymbol{H}_{nm0}\operatorname{vec}(\boldsymbol{Y}_{nm,t-1}) + \boldsymbol{S}_{nm0}^{-1}\left[(\boldsymbol{\Pi}_{m0}' \otimes I_n)\operatorname{vec}(\boldsymbol{X}_{nk,t}) + \operatorname{vec}(\boldsymbol{C}_{nm0}) + \alpha_{m0,t} \otimes l_n + \operatorname{vec}(\boldsymbol{V}_{nm,t})\right] \tag{4}$$

Applying the recursive method then gives the moving average from of the model:

$$\operatorname{vec}(\boldsymbol{Y}_{nm,t}) = \sum_{h=0}^{\infty} \boldsymbol{H}_{nm0}^{h}\boldsymbol{S}_{nm0}^{-1}\left[(\boldsymbol{\Pi}_{\mathrm{m}0}' \otimes I_n)\operatorname{vec}(\boldsymbol{X}_{nk,t-h}) + \operatorname{vec}(\boldsymbol{C}_{nm0}) + \alpha_{m0,t-h} \otimes l_n + \operatorname{vec}(\boldsymbol{V}_{nm,t-h})\right] \tag{5}$$

In order to eliminate both individual and time fixed effects, we follow the method in Lee and Yu (2014), the forward orthogonal difference (FOD) transformation. In contrast to first-differencing, the FOD transformation introduced by Arellano and Bover (1995) preserves the orthogonality of disturbances when they are originally i.i.d., avoiding the serial correlation created by differencing. Let $J_T = I_T - \frac{1}{T}l_T l_T'$, where l_T is a $T \times 1$ vector of ones. The eigenvalues of J_T are $T-1$ ones and a single zero. Let $\begin{pmatrix} F_{T,T-1} & J_T/\sqrt{T} \end{pmatrix}$ be the orthonormal matrix of J_T, where $F_{T,T-1}$ is the $T \times (T-1)$ eigenvectors matrix corresponding to the eigenvalues of one, i.e., $J_T F_{T,T-1} = F_{T,T-1}, F_{T,T-1}'F_{T,T-1} = I_{T-1}, J_T l_T = 0, F_{T,T-1}' l_T = 0, F_{T,T-1}F_{T,T-1}' = J_T, F_{T,T-1}F_{T,T-1}' + \frac{1}{T}l_T l_T' = I_T$. Assumes $Y_{nm,0}$ is observable, then the individual fixed effects can be eliminated by the forward orthogonal difference (FOD) transformation.

$$\begin{bmatrix} \text{vec}(Y_{nm,1}) & \dots & \text{vec}(Y_{nm,T}) \end{bmatrix} F_{T,T-1} = \begin{bmatrix} \text{vec}(Y^*_{nm,1}) & \dots & \text{vec}(Y^*_{nm,T-1}) \end{bmatrix}$$

$$\begin{bmatrix} \text{vec}(Y_{nm,0}) & \dots & \text{vec}(Y_{nm,T-1}) \end{bmatrix} F_{T,T-1} = \begin{bmatrix} \text{vec}(Y^{(*,-1)}_{nm,0}) & \dots & \text{vec}(Y^{(*,-1)}_{nm,T-2}) \end{bmatrix}$$

$$\begin{bmatrix} \text{vec}(X_{n,1}) & \dots & \text{vec}(X_{n,T}) \end{bmatrix} F_{T,T-1} = \begin{bmatrix} \text{vec}(X^*_{n,1}) & \dots & \text{vec}(X^*_{n,T-1}) \end{bmatrix}$$

$$\begin{bmatrix} \text{vec}(U_{nm,1}) & \dots & \text{vec}(U_{nm,T}) \end{bmatrix} F_{T,T-1} = \begin{bmatrix} \text{vec}(U^*_{nm,1}) & \dots & \text{vec}(U^*_{nm,T-1}) \end{bmatrix}$$

After FOD transformation, the individual fixed effects are eliminated, the model becomes

$$\boldsymbol{Y}^*_{nm,t}\boldsymbol{\Gamma}_{\text{m0}} = \left[W_{1n}Y^*_{n1,t}, \dots, W_{mn}Y^*_{nm,t}\right]\boldsymbol{\Psi}_{\text{ms0}} + \boldsymbol{Y}^{(*,-1)}_{nm,t-1}\boldsymbol{P}_{\text{ms0}} + \left[W_{1n}Y^{(*,-1)}_{n1,t-1} \dots W_{mn}Y^{(*,-1)}_{nm,t-1}\right]\boldsymbol{\Phi}_{\text{ms0}} + \alpha^{*\prime}_{ms0,t} \otimes l_n + \boldsymbol{X}^*_{n,t}\boldsymbol{\Pi}_{\text{ms0}} + \boldsymbol{U}^*_{nm,t} \tag{6}$$

In vectorization form:

$$\begin{aligned} (\boldsymbol{\Gamma}'_{\text{m0}} \otimes I_n)\text{vec}\left(\boldsymbol{Y}^*_{nm,t}\right) =& (\boldsymbol{\Psi}'_{\text{ms0}} \otimes I_n)\mathcal{W}_{nm}\text{vec}\left(\boldsymbol{Y}^*_{nm,t}\right) + (\boldsymbol{P}'_{\text{ms0}} \otimes I_n)\text{vec}\left(\boldsymbol{Y}^{(*,-1)}_{nm,t-1}\right) + (\boldsymbol{\Phi}'_{\text{ms0}} \otimes I_n)\mathcal{W}_{nm}\text{vec}\left(\boldsymbol{Y}^{(*,-1)}_{nm,t-1}\right) \\ &+ (\boldsymbol{\Pi}'_{\text{ms0}} \otimes I_n)\text{vec}\left(\boldsymbol{X}^*_{n,t}\right) + \alpha^*_{ms0,t} \otimes l_n + \text{vec}\left(\boldsymbol{U}^*_{nm,t}\right) \end{aligned} \tag{7}$$

with $Y^*_{np,t} = \left(\frac{T-t}{T-t+1}\right)^{1/2}\left(Y_{np,t} - \frac{1}{T-t}\sum_{h=t+1}^{T} Y_{np,h}\right)$, $Y^{(*,-1)}_{np,t-1} = \left(\frac{T-t}{T-t+1}\right)^{1/2}\left(Y_{np,t-1} - \frac{1}{T-t}\sum_{h=t}^{T-1} Y_{np,h}\right)$, and $U^*_{np,t} = \left(\frac{T-t}{T-t+1}\right)^{1/2}\left(U_{np,t} - \frac{1}{T-t}\sum_{h=t+1}^{T} U_{np,h}\right)$. The time fixed effect can be further eliminated by multiplying the transformed model by $J_n = I_n - \frac{1}{n} l_n l_n'$, then we have

$$J_n \boldsymbol{Y}^*_{nm,t} \boldsymbol{\Gamma}_{\mathrm{m}0} = J_n \left[W_{1n} Y^*_{n1,t}, \ldots, W_{mn} Y^*_{nm,t} \right] \boldsymbol{\Psi}_{\mathrm{ms}0} + J_n \boldsymbol{Y}^{(*,-1)}_{nm,t-1} \boldsymbol{P}_{\mathrm{ms}0} + J_n \left[W_{1n} Y^{(*,-1)}_{n1,t-1}, \ldots W_{mn} Y^{(*,-1)}_{nm,t-1} \right] \boldsymbol{\Phi}_{\mathrm{ms}0} + J_n \boldsymbol{X}^*_{n,t} \boldsymbol{\Pi}_{\mathrm{ms}0} + J_n \boldsymbol{U}^*_{nm,t} \tag{8}$$

In vectorization form:

$$\begin{aligned}(\boldsymbol{\Gamma}'_{\mathrm{m}0} \otimes J_n)\mathrm{vec}\left(\boldsymbol{Y}^*_{nm,t}\right) =& (\boldsymbol{\Psi}'_{\mathrm{ms}0} \otimes J_n)\mathcal{W}_{nm}\mathrm{vec}\left(\boldsymbol{Y}^*_{nm,t}\right) + (\boldsymbol{P}'_{\mathrm{ms}0} \otimes J_n)\mathrm{vec}\left(\boldsymbol{Y}^{(*,-1)}_{nm,t-1}\right) + (\boldsymbol{\Phi}'_{\mathrm{ms}0} \otimes J_n)\mathcal{W}_{nm}\mathrm{vec}\left(\boldsymbol{Y}^{(*,-1)}_{nm,t-1}\right) \\ &+ (\boldsymbol{\Pi}'_{\mathrm{ms}0} \otimes J_n)\mathrm{vec}\left(\boldsymbol{X}^*_{n,t}\right) + (I_m \otimes J_n)\mathrm{vec}\left(\boldsymbol{U}^*_{nm,t}\right)\end{aligned} \tag{9}$$

We propose GMM estimation for the transformed model (9). One advantage is after eliminating individual and time fixed effects through FOD transformation, GMM circumvents the incidental parameter problem and handles endogenous regressors correlated with the error term. Unlike ML approach- which requires additional initial value specifications when T is small and may retain asymptotic bias even when T is large. Another advantage of GMM estimation is that it does not require row-normalization of the spatial weight matrix. In contrast, the ML approach loses its well-defined SAR structure for $J_n \boldsymbol{Y}^*_{nm,t}$ when the weight matrix is not row-normalized, since $J_n W_{jn} \neq J_n W_{jn} J_n$.

For estimation, we state the following assumptions regarding the DGP.

Assumption 1. *Let $U_{nm,i,t}$ denote the i-th row of $\boldsymbol{U}_{nm,t}$, a random vector of dimension m with zero mean and covariance matrix Σ_{um0}, and those vectors are i.i.d. for all i and t. The elements of disturbances satisfy the moment condition that* $\max_{1 \le k,l,p,q \le m} E\left[\left| u_{nk,i,t} u_{nl,i,t} u_{np,i,t} u_{nq,i,t} \right|^{1+\eta} \right]$ *for some positive constant $\eta > 0$ is bounded.*

Assumption 2. *The spatial weight matrix W_{jn}, $j = 1, \ldots, m$ is nonstochastic with zero diagonal. Row and column sums of W_{jn} in absolute value are uniformly bounded, uniformly in n. Also, $\boldsymbol{S}_{nm}(\boldsymbol{\Psi}_m)$ is invertible for all $\boldsymbol{\Psi}_m \in \boldsymbol{\Delta}_{\Psi}$, and uniform* *bounded.*

Assumption 3. *The parameter space of coefficients* $\boldsymbol{\theta} = [\boldsymbol{\Gamma}'_m, \boldsymbol{\Psi}'_{ms}, \boldsymbol{P}'_{ms}, \boldsymbol{\Phi}'_{ms}, \boldsymbol{\Pi}'_{ms}]'$ *is compact, the true parameter* $\boldsymbol{\theta}_0$ *is in the interior of the parameter space* $\boldsymbol{\Theta}$*. The covariance matrix* Σ_{um0} *lies on a compact space and is nonsingular. Also,* $\rho\left((\boldsymbol{\Psi}'_{ms0} \otimes I_n)\mathcal{W}_{nm}\right) < 1$*,* $\rho\left(\boldsymbol{H}_{nm0}\right) < 1$*, and* $\sum_{h=1}^{\infty} abs(\boldsymbol{H}^h_{nm0})$ *is bounded in row and column sum norms.*

Assumption 4. *The elements of* $\boldsymbol{X}_{n,t}$*,* $\boldsymbol{C}_{nms0}$*, and* $\alpha_{ms0,t}$ *are nonstochastic and uniformly bounded for all* n *and* t*. And,* $\lim_{n\to\infty} \frac{1}{n(T-1)} \boldsymbol{X}^{*\prime}_{n,t} J_n \boldsymbol{X}^*_{n,t}$ *exists and is nonsingular.*

Assumption (1) specifies regularity assumptions for the error term. Similar to classical simultaneous equations models, the error term is assumed to be i.i.d. across time and individual spatial units, however, they are allowed to be correlated across variables. Formally, let $U_{nm,j,t}$ represents the $n \times 1$ error vector for the j-th equation. Let $u_{j,i,t}$ be the i-th element of $U_{nm,j,t}$. According to Assumption (1), for a given $j, u_{j,i,t}$ is $i.i.d.$ in both index i and index t, and that $u_{j,i,t}$ and $u_{j',i,t}$ exhibit contemporaneous correlation. Specifically, $E(u^2_{j,i,t}) = \sigma^2_{jj}$ and $Cov(u_{j,i,t}, u_{j',i,t}) = \sigma_{jj'}$ for any $j \neq j'$. The off-diagonal block of Σ_{um0} is zero, but the off-diagonal elements of Ω_{um0} may not be 0.

$$\Sigma_{um0} = \begin{pmatrix} \sigma_{11} & \dots & \sigma_{1m} \\ \vdots & \ddots & \vdots \\ \sigma_{m1} & \dots & \sigma_{mm} \end{pmatrix} \otimes I_n = \Omega_{um0} \otimes I_n \tag{10}$$

Assumption (2) is standard in spatial econometrics, stems from Kelejian and Prucha (1998). The uniform boundedness condition ensures that no single unit can exert disproportionately large influence on the entire system. The zero diagonal excludes self-influence, which aids in interpreting spatial effects. The invertible condition of $\boldsymbol{S}_{nm}(\boldsymbol{\Psi}_m)$ ensures the reduced form model is valid. Assumption (3) points out the stability conditions of the system in both spatial and time dimensions. Following Yang and Lee (2019, 2021), spatial stability requires $\rho\left((\boldsymbol{\Psi}'_{\text{ms0}} \otimes I_n)\mathcal{W}_{nm}\right) < 1$, ensures bounded spatial spillover effects. This condition further enables the Neumann series expansion of the spatial multiplier matrix, which we utilize to construct IVs. Time stability requires $\rho\left(\boldsymbol{H}_{nm0}\right)<1$. The boundedness of $\sum_{h=1}^{\infty} abs(\boldsymbol{H}_{nm0}^{h})$ rules out unit roots and non-stationary dynamics while restricting the permissible degree of cross-sectional and serial dependence. The compactness of parameter spaces is a standard requirement for establishing consistency and deriving asymptotic properties of the estimators. Assumption (4) states exogeneity of the fixed effects $\boldsymbol{C}_{nms0}, \alpha_{ms0,t}$ and the regressor $\boldsymbol{X}_{n,t}$. The uniform boundedness condition serves to simplify asymptotic analysis but is non-essential, it may be replaced by appropriate finite moment conditions. The remaining assumptions stipulate that the regressors $\boldsymbol{X}_{n,t}^{*}$ are asymptotically linearly independent.

2.2 Moment conditions

2.2.1 Single equation moment conditions

The l-th equation derived from model (8) can be formulated as follows:

$$J_n Y_{nl,t}^{*} = -J_n Y_{l,nm,t}^{*} \Gamma_{l,m0} + J_n \ddot{Y}_{l,nm,t}^{*} \Psi_{l,ms0} + J_n Y_{l,nm,t-1}^{(*,-1)} P_{l,ms0} + J_n \ddot{Y}_{l,nm,t-1}^{(*,-1)} \Phi_{l,ms0} + J_n X_{l,nt}^{*} \Pi_{l,ms0} + J_n U_{nl,t}^{*} \tag{11}$$

In this equation, the term $Y_{nl,t}^{*}$ is a $n\times 1$ vector which represents the endogenous variable that is specifically associated with the l-th equation, and its coefficient is normalized to 1 for identification purposes. The matrix $Y_{l,nm,t}^{*}$ includes other endogenous variables that appear in the same equation, consisting of columns from the full matrix $\boldsymbol{Y}_{nm,t}^{*}$. Similarly, the matrices $\ddot{Y}_{l,nm,t}^{*}$, $Y_{l,nm,t-1}^{(*,-1)}$, $\ddot{Y}_{l,nm,t-1}^{(*,-1)}$, and $X_{l,nt}^{*}$ contain relevant columns from $\left[W_{1n}Y_{n1,t}^{*}, \ldots, W_{mn}Y_{nm,t}^{*}\right]$, $\boldsymbol{Y}_{nm,t-1}^{(*,-1)}$, $\left[W_{1n}Y_{n1,t-1}^{(*,-1)}, \ldots, W_{mn}Y_{nm,t-1}^{(*,-1)}\right]$, and $\boldsymbol{X}_{n,t}^{*}$, respectively. The corresponding coefficients $\Psi_{l,ms0}$, $P_{l,ms0}$, $\Phi_{l,ms0}$, and $\Pi_{l,ms0}$ are nonzero elements in $\boldsymbol{\Psi}_{ms0}$, $\boldsymbol{P}_{ms0}$, $\boldsymbol{\Phi}_{ms0}$, and $\boldsymbol{\Pi}_{ms0}$. Since $Y_{l,nm,t-1}^{(*,-1)}$ is correlated with $U_{nl,t}^{*}$, the model suffers from endogeneity. To address this issue, and given the simultaneous and spatial interactions in the system, we need to find instrumental variables (IVs) for regressors

$J_n\left[Y^*_{l,nm,t}, \ddot{Y}^*_{l,nm,t}, Y^{(*,-1)}_{l,nm,t-1}, \ddot{Y}^{(*,-1)}_{l,nm,t-1}\right]$. To construct the linear moment conditions, we begin by considering the following example to show the motivation of finding appropriate IVs, suppose that we have 2 endogenous variables ($m = 2$), and the first equation is

$$\begin{aligned}J_nY^*_{n1,t} = &-J_nY^*_{n2,t}\gamma_{21} + J_nW_{1n}Y^*_{n1,t}\psi_{11} + J_nW_{2n}Y^*_{n2,t}\psi_{21} + J_nY^{(*,-1)}_{n1,t-1}p_{11} + J_nY^{(*,-1)}_{n2,t-1}p_{21}\\ &+J_nW_{1n}Y^{(*,-1)}_{n1,t-1}\phi_{11} + J_nW_{2n}Y^{(*,-1)}_{n2,t-1}\phi_{21} + J_nX^*_{1,nt}\pi_1 + J_nU^*_{n1,t}\end{aligned}$$

If we derive the reduced form, we obtain the term $(I_n - W_{1n}\psi_{11})^{-1}X^*_{1,nt}\pi_1$. To find the IVs for $W_{1n}Y^*_{n1,t}$, we pre-multiply the reduced form by W_{1n} and observe the term $W_{1n}(I_n - W_{1n}\psi_{11})^{-1}X^*_{1,nt}\pi_1$. Given Assumption (3), we use the Neumann series expansion of $(I_n - W_{1n}\psi_{11})^{-1} = \sum_{d=0}^{D}(W_{1n}\psi_{11})^d$, which implies that the IVs for $W_{1n}Y^*_{n1,t}$ are given by $W_{1n}(I_n + W_{1n}\psi_{11} + W^2_{1n}\psi^2_{11} + \dots)X^*_{1,nt}$. Similarly, the IVs for $W_{2n}Y^*_{n2,t}$ are $W_{2n}(I_n + W_{1n}\psi_{11} + W^2_{1n}\psi^2_{11} + \dots)X^*_{1,nt}$. According to Alvarez and Arellano (2003), we can use the time lag variables before transformation to construct IVs for $Y^{(*,-1)}_{n1,t-1}, Y^{(*,-1)}_{n2,t-1}, W_{1n}Y^{(*,-1)}_{n1,t-1}, W_{2n}Y^{(*,-1)}_{n2,t-1}$, then IVs can be constructed as $W_{1n}(I_n + W_{1n}\psi_{11} + W^2_{1n}\psi^2_{11} + \dots)Y_{n1,t-1}$ and $W_{2n}(I_n + W_{1n}\psi_{11} + W^2_{1n}\psi^2_{11} + \dots)Y_{n2,t-1}$, respectively. Therefore, potential IVs for the l-th equation can be selected from the following set, provided that they are linearly independent.

$$
\left\{
\begin{array}{c}
\boldsymbol{X}^*_{n,t} \\
W_{1n}(I_n, W_{ln}, W^2_{ln}, \dots)\boldsymbol{X}^*_{n,t} = (W_{1n}, W_{1n}W_{ln}, W_{1n}W^2_{ln}, \dots)\boldsymbol{X}^*_{n,t} \\
\vdots \\
W_{mn}(I_n, W_{ln}, W^2_{ln}, \dots)\boldsymbol{X}^*_{n,t} = (W_{mn}, W_{mn}W_{ln}, W_{mn}W^2_{ln}, \dots)\boldsymbol{X}^*_{n,t} \\
\boldsymbol{Y}_{nm,t-1} \\
W_{1n}(I_n, W_{ln}, W^2_{ln}, \dots)Y_{n1,t-1} = (W_{1n}, W_{1n}W_{ln}, W_{1n}W^2_{ln}, \dots)Y_{n1,t-1} \\
\vdots \\
W_{mn}(I_n, W_{ln}, W^2_{ln}, \dots)Y_{nm,t-1} = (W_{mn}, W_{mn}W_{ln}, W_{mn}W^2_{ln}, \dots)Y_{nm,t-1}
\end{array}
\right\}
\tag{12}
$$

Define $\boldsymbol{W}_n Z_n = (W_{1n}Z_n, \dots, W_{mn}Z_n)\,, \boldsymbol{W}^2_{in} Z_n = (W_{1n}W_{in}, \dots, W_{mn}W_{in})Z_n, i = 1, \dots, m$. Then, an example of IVs for l-th equation is

$$
\begin{aligned}
Q_{nl,t} &= \big[\boldsymbol{X}^*_{n,t}, (W_{1n}, \dots, W_{mn})\boldsymbol{X}^*_{n,t}, (W_{1n}W_{ln}, \dots, W_{mn}W_{ln})\boldsymbol{X}^*_{n,t}, \\
&\qquad \boldsymbol{Y}_{nm,t-1}, (W_{1n}, \dots, W_{mn})\boldsymbol{Y}_{nm,t-1}, (W_{1n}W_{ln}, \dots, W_{mn}W_{ln})\boldsymbol{Y}_{nm,t-1}\big] \\
&= \big[\boldsymbol{X}^*_{n,t}, \boldsymbol{W}_n\boldsymbol{X}^*_{n,t}, \boldsymbol{W}^2_{ln}\boldsymbol{X}^*_{n,t}, \boldsymbol{Y}_{nm,t-1}, \boldsymbol{W}_n\boldsymbol{Y}_{nm,t-1}, \boldsymbol{W}^2_{ln}\boldsymbol{Y}_{nm,t-1}\big]
\end{aligned}
\tag{13}
$$

With dimension $n \times k_q$, in this example, $k_q = (m+k)(2m+1)$. Let $\boldsymbol{Q}_{nl,T-1} = \left[Q'_{nl,1}, \dots, Q'_{nl,T-1}\right]'_{[n(T-1)]\times k_q}$, $\boldsymbol{J}_{n,T-1} = I_{T-1} \otimes J_n$, and $\boldsymbol{U}^*_{nl,T-1}(\theta_l) = \left[U^{*'}_{nl,1}(\theta_l), \dots, U^{*'}_{nl,T-1}(\theta_l)\right]'_{[n(T-1)]\times 1}$, where $U^*_{nl,t}(\theta_l) = Y^*_{nl,t} + Y^*_{l,nm,t}\Gamma_{l,m0} - \ddot{Y}^*_{l,nm,t}\Psi_{l,ms0} - Y^{(*,-1)}_{l,nm,t-1}P_{l,ms0} - \ddot{Y}^{(*,-1)}_{l,nm,t-1}\Phi_{l,ms0} - X^*_{l,nt}\Pi_{l,ms0}$, and $\theta_l = (\Gamma'_{l,m0}, \Psi'_{l,ms0}, P'_{l,ms0}, \Phi'_{l,ms0}, \Pi'_{l,ms0})'$. The linear moment conditions for l-th equation are $E\left[\boldsymbol{Q}'_{nl,T-1}\boldsymbol{J}_{n,T-1}\boldsymbol{U}^*_{nl,T-1}\right] = 0$, and the empirical linear moment functions are $g^1_{nl,T}(\theta_l) = \boldsymbol{Q}'_{nl,T-1}\boldsymbol{J}_{n,T-1}\boldsymbol{U}^*_{nl.T-1}(\theta_l)$.Note that while equation (11) does not explicitly include time fixed effects, the spatial and space-time lag terms may implicitly retain time fixed effects if the spatial weight matrix is not row-normalized ($J_nW_{jn} \neq J_nW_{jn}J_n$). However, applying the within-transform J_n to the moment conditions effectively eliminates these residual time effects. In addition to the linear moment, Lee (2007) suggests using quadratic moment conditions to capture spatial correlations and potentially increase the efficiency of estimates. We consider the following quadratic form in $U^*_{nl,t}$, let P_{ln} be the $n \times n$ constant matrix, and taking the expectation: $E\left[(P_{ln}J_nU^*_{nl,t})'J_nU^*_{nl,t}\right] = \sigma^2_{ll}tr(P_{ln}J_n)$. For the quadratic moment conditions to be satisfied, it is required that $tr(P_{jn}J_n) = 0$. Again using the same example as above (first equation in the system), in the reduced from of this equation, in addition to the term $(I_n - W_{1n}\psi_{11})^{-1}X^*_{1,nt}\pi_1$, there is also the term $(I_n - W_{1n}\psi_{11})^{-1}U^*_{n1,t}$, from the series expansion, we have $(I_n + W_{1n}\psi_{11} + W^2_{1n}\psi^2_{11} + \dots)U^*_{n1,t}$, thus the constant matrices for $W_{1n}Y^*_{n1,t}$ could be $P_{1,s} = W^s_{1n} - \frac{tr(W^s_{1n}J_n)}{n-1}J_n, s = (1, \dots, S)$. In the notation $P_{1,s}$, the subscript 1 indicates that we only use W_{1n} in the constant matrix, and s indicates the power of the weight matrix. Constant matrices for $W_{2n}Y^*_{n2,t}$ could be $P_{21,d} = W_{2n}W^d_{1n} - \frac{tr(W_{2n}W^d_{1n}J_n)}{n-1}J_n, d = 0, \dots, D$. The subscript 21 means that we use information from weight matrix $W_{2n}W^d_{1n}$. For instance, when $d = 0, P_{21,0} = W_{2n} - \frac{tr(W_{2n}J_n)}{n-1}J_n$, with $tr(P_{21,0}J_n) = 0$. When $d = 1, P_{21,1} = W_{2n}W_{1n} - tr\left(\frac{tr(W_{2n}W_{1n})J_n}{n-1}J_n\right)$. The quadratic moments for the l-th equation are constructed similarly to those for the first equation. Specifically, for each $j = 1, \dots, m$, we define a family of constant matrices $P_{jl,sd}$ based on different choices of the integers s and d:

$$P_{jl,sd} = \begin{cases} W^s_{jn} - \dfrac{\mathrm{tr}(W^s_{jn}J_n)}{n-1}J_n, & \text{if } j = l, \quad \text{for } s = 1, \dots, S, \\ W_{jn}W^d_{ln} - \dfrac{\mathrm{tr}(W_{jn}W^d_{ln}J_n)}{n-1}J_n, & \text{if } j \neq l, \quad \text{for } d = 0, \dots, D. \end{cases} \tag{14}$$

These choices of (s, d) generate many distinct constant matrices. In practice, we often choose only a subset of such pairs to form a finite number of constant matrices. Denote by P total number of constant matrices to construct the quadratic moments, and let $p = 1, \dots, P$ to index these selected matrices. Let $\boldsymbol{P}_{jl,p,T-1}$ equal $I_{T-1} \otimes P_{jl,sd}$ for some (s, d), then the quadratic moment conditions for l-th equation are

$$E[\boldsymbol{U}^{*\prime}_{nl,T-1}\boldsymbol{J}_{n,T-1}\boldsymbol{P}_{jl,p,T-1}\boldsymbol{J}_{n,T-1}\boldsymbol{U}^*_{nl,T-1}] = 0 \quad \text{for } j = 1, \dots, m, p = 1, \dots, P \tag{15}$$

and the empirical quadratic moment functions are $g^2_{nl,T}(\theta_l) = \boldsymbol{U}^{*\prime}_{nl,T-1}(\theta_l)\boldsymbol{J}_{n,T-1}\boldsymbol{P}_{jl,p,T-1}\boldsymbol{J}_{n,T-1}\boldsymbol{U}^{*}_{nl,T-1}(\theta_l)$ for $j = 1, \ldots, m, p = 1, \ldots, P$. Combing both linear and quadratic moments, a candidate of empirical moment conditions could be

$$g_{nl,T}(\theta_l) = \begin{pmatrix} g^1_{nl,T}(\theta_l) \\ g^2_{nl,T}(\theta_l) \end{pmatrix} = \begin{pmatrix} \boldsymbol{Q}'_{nl,T-1}\boldsymbol{J}_{n,T-1}\boldsymbol{U}^{*}_{nl.T-1}(\theta_l) \\ \boldsymbol{U}^{*\prime}_{nl,T-1}(\theta_l)\boldsymbol{J}_{n,T-1}\boldsymbol{P}_{1l,1,T-1}\boldsymbol{J}_{n,T-1}\boldsymbol{U}^{*}_{nl,T-1}(\theta_l) \\ \vdots \\ \boldsymbol{U}^{*\prime}_{nl,T-1}(\theta_l)\boldsymbol{J}_{n,T-1}\boldsymbol{P}_{1l,P,T-1}\boldsymbol{J}_{n,T-1}\boldsymbol{U}^{*}_{nl,T-1}(\theta_l) \\ \vdots \\ \boldsymbol{U}^{*\prime}_{nl,T-1}(\theta_l)\boldsymbol{J}_{n,T-1}\boldsymbol{P}_{ml,1,T-1}\boldsymbol{J}_{n,T-1}\boldsymbol{U}^{*}_{nl,T-1}(\theta_l) \\ \vdots \\ \boldsymbol{U}^{*\prime}_{nl,T-1}(\theta_l)\boldsymbol{J}_{n,T-1}\boldsymbol{P}_{ml,P,T-1}\boldsymbol{J}_{n,T-1}\boldsymbol{U}^{*}_{nl,T-1}(\theta_l) \end{pmatrix} \tag{16}$$

and the single-equation GMM estimator is derived as $\hat{\theta}_{l,nT} = \arg\min\, g'_{nl,T}(\theta_l)a'_{nl,T}a_{nl,T}g_{nl,T}(\theta_l)$.

2.2.2 System equation moment condition

From (7), we derive the reduced form model:

$$\begin{aligned} \mathrm{vec}(\boldsymbol{Y}^*_{nm,t}) &= \left[I_{nm} - (\boldsymbol{\Gamma}'_{\mathrm{m0}} \otimes I_n)^{-1}(\boldsymbol{\Psi}'_{\mathrm{ms0}} \otimes I_n)\mathcal{W}_{nm}\right]^{-1}(\boldsymbol{\Gamma}'_{\mathrm{m0}} \otimes I_n)^{-1}(\boldsymbol{P}'_{\mathrm{ms0}} \otimes I_n)\mathrm{vec}(\boldsymbol{Y}^*_{nm,t-1}) \\ &+ \left[I_{nm} - (\boldsymbol{\Gamma}'_{\mathrm{m0}} \otimes I_n)^{-1}(\boldsymbol{\Psi}'_{\mathrm{ms0}} \otimes I_n)\mathcal{W}_{nm}\right]^{-1}(\boldsymbol{\Gamma}'_{\mathrm{m0}} \otimes I_n)^{-1}(\boldsymbol{\Pi}'_{ms0} \otimes I_n)\mathrm{vec}(\boldsymbol{X}^*_{n,t}) \\ &+ \left[I_{nm} - (\boldsymbol{\Gamma}'_{\mathrm{m0}} \otimes I_n)^{-1}(\boldsymbol{\Psi}'_{\mathrm{ms0}} \otimes I_n)\mathcal{W}_{nm}\right]^{-1}(\boldsymbol{\Gamma}'_{\mathrm{m0}} \otimes I_n)^{-1}(\boldsymbol{\Phi}'_{\mathrm{ms0}} \otimes I_n)\mathcal{W}_{nm}\mathrm{vec}(\boldsymbol{Y}^*_{nm,t-1}) \\ &+ \left[I_{nm} - (\boldsymbol{\Gamma}'_{\mathrm{m0}} \otimes I_n)^{-1}(\boldsymbol{\Psi}'_{\mathrm{ms0}} \otimes I_n)\mathcal{W}_{nm}\right]^{-1}(\boldsymbol{\Gamma}'_{\mathrm{m0}} \otimes I_n)^{-1}\left[\alpha_{\mathrm{ms0,t}} \otimes l_n + \mathrm{vec}(\boldsymbol{U}^*_{nm,t})\right] \end{aligned} \tag{17}$$

Denote $\boldsymbol{A}_{nm} = (\boldsymbol{\Gamma}'_{\mathrm{m0}} \otimes I_n)^{-1}(\boldsymbol{\Psi}'_{\mathrm{ms0}} \otimes I_n)\mathcal{W}_{nm}, \boldsymbol{B}_{nm} = (\boldsymbol{\Gamma}'_{\mathrm{m0}} \otimes I_n)^{-1}(\boldsymbol{\Pi}'_{ms0} \otimes I_n)$, then from the series expansion, we have $(I_{nm} -$

$\boldsymbol{A}_{nm})^{-1} = \sum_{d=0}^{\infty}(\boldsymbol{A}_{nm}\mathcal{W}_{nm})^d$, and IVs for $\mathcal{W}_{nm}\text{vec}(\boldsymbol{Y}^*_{nm,t})$ could be $\mathcal{W}_{nm}(I_{nm} + \boldsymbol{A}_{nm}\mathcal{W}_{nm} + \boldsymbol{A}^2_{nm}\mathcal{W}^2_{nm} + \dots)\boldsymbol{B}_{nmn}\text{vec}(\boldsymbol{X}^*_{n,t})$. If we take the first term from the series expansion, we have

$$\mathcal{W}_{nm}\boldsymbol{B}_{nm}\text{vec}(\boldsymbol{X}_{n,t}) = \mathcal{W}_{nm}\begin{pmatrix} B_{11} & \dots & B_{1k} \\ \vdots & \ddots & \vdots \\ B_{m1} & \dots & B_{mk} \end{pmatrix}\text{vec}(\boldsymbol{X}_{n,t}) = \begin{pmatrix} W_{1n}B_{11}X_{n1,t} + \dots + W_{1n}B_{1k}X_{nk,t} \\ \vdots \\ W_{mn}B_{m1}X_{n1,t} + \dots + W_{mn}B_{mk}X_{nk,t} \end{pmatrix}$$

If we take the second from the series expansion, we have

$$\mathcal{W}_{nm}\boldsymbol{A}_{nm}\mathcal{W}_{nm}\boldsymbol{B}_{nm}\text{vec}(\boldsymbol{X}_{n,t}) = \mathcal{W}_{nm}\begin{pmatrix} A_{11} & \dots & A_{1m} \\ \vdots & \ddots & \vdots \\ A_{m1} & \dots & A_{mm} \end{pmatrix}\mathcal{W}_{nm}\begin{pmatrix} B_{11} & \dots & B_{1k} \\ \vdots & \ddots & \vdots \\ B_{m1} & \dots & B_{mk} \end{pmatrix}\text{vec}(\boldsymbol{X}_{n,t})$$

$$= \begin{pmatrix} (W_{1n}A_{11}W_{1n}B_{11}X_{n1,t} + \dots + W_{1n}A_{1m}W_{mn}B_{m1}X_{n1,t}) + \dots + (W_{1n}A_{11}W_{1n}B_{1k}X_{nk,t} + \dots + W_{1n}A_{1m}W_{mn}B_{mk}X_{nk,t}) \\ \vdots \\ (W_{mn}A_{m1}W_{11}B_{m1}X_{n1,t} + \dots + W_{mn}A_{mm}W_{mn}B_{mk}X_{n1,t}) + \dots + (W_{mn}A_{m1}W_{1n}B_{1k}X_{n1,t} + \dots + W_{mn}A_{mm}W_{mn}B_{mk}X_{nk,t}) \end{pmatrix}$$

where $\boldsymbol{A}_{nm}$ and $\boldsymbol{B}_{nm}$ are unknown coefficients. A similar procedure can be applied to identify instruments for the other endogenous variables. Therefore, the system's potential instruments can be selected from the following set, provided they remain linearly independent.

$$\left\{\begin{array}{cc} \boldsymbol{X}^*_{n,t}, & \boldsymbol{Y}_{nm,t-1}, \\ (W_{1n},\ldots,W_{mn})\boldsymbol{X}^*_{n,t}, & (W_{1n},\ldots,W_{mn})\boldsymbol{Y}_{nm,t-1}, \\ (W_{1n}W_{1n},\ldots,W_{mn}W_{1n})\boldsymbol{X}^*_{n,t}, & (W_{1n}W_{1n},\ldots,W_{mn}W_{1n})\boldsymbol{Y}_{nm,t-1}, \\ \vdots & \vdots \\ (W_{1n}W_{mn},\ldots,W_{mn}W_{mn})\boldsymbol{X}^*_{n,t}, & (W_{1n}W_{mn},\ldots,W_{mn}W_{mn})\boldsymbol{Y}_{nm,t-1}, \end{array}\right\} = \left\{\begin{array}{c} \boldsymbol{X}^*_{n,t}, \boldsymbol{W}_n\boldsymbol{X}^*_{n,t}, (\boldsymbol{W}^2_{1n},\ldots,\boldsymbol{W}^2_{mn})\boldsymbol{X}^*_{n,t}, \\ \boldsymbol{Y}_{nm,t-1}, \boldsymbol{W}_n\boldsymbol{Y}_{nm,t-1}, (\boldsymbol{W}^2_{1n},\ldots,\boldsymbol{W}^2_{mn})\boldsymbol{Y}_{nm,t-1} \end{array}\right\} \tag{18}$$

As an example, we could use

$$Q_{n,t} = \left[\boldsymbol{X}^*_{n,t}, \boldsymbol{W}_n\boldsymbol{X}^*_{n,t}, (\boldsymbol{W}^2_{1n},\ldots,\boldsymbol{W}^2_{mn})\boldsymbol{X}^*_{n,t}, \boldsymbol{Y}_{nm,t-1}, \boldsymbol{W}_n\boldsymbol{Y}_{nm,t-1}, (\boldsymbol{W}^2_{1n},\ldots,\boldsymbol{W}^2_{mn})\boldsymbol{Y}_{nm,t-1}\right] \tag{19}$$

with dimension $n \times \boldsymbol{k}_q$, in this example, $\boldsymbol{k}_q = (m+k)(1+m^2+m^3)$. Let $\boldsymbol{Q}_{n,T-1} = [Q'_{n,1},\ldots,Q'_{n,T-1}]'_{[n(T-1)]\times\boldsymbol{k}_q}$, and $\boldsymbol{U}^*_{nm,T-1}(\boldsymbol{\theta}) = \left[\text{vec}\left(\boldsymbol{U}^*_{nm,1}(\boldsymbol{\theta})\right)',\ldots,\text{vec}\left(\boldsymbol{U}^*_{nm,T-1}(\boldsymbol{\theta})\right)'\right]'_{[nm(T-1)\times 1]}$, then the linear moment condition is $E[(I_m\otimes\boldsymbol{Q}_{n,T-1})'(I_m\otimes\boldsymbol{J}_{n,T-1})\boldsymbol{U}^*_{nm,T-1}] = 0$, where $\boldsymbol{\theta} = (\theta'_1,\ldots,\theta'_m)'$, and the empirical moment condition is $g^1_{nT}(\boldsymbol{\theta}) = (I_m\otimes\boldsymbol{Q}_{n,T-1})'(I_m\otimes\boldsymbol{J}_{n,T-1})\boldsymbol{U}^*_{nm,T-1}(\boldsymbol{\theta})$. For the quadratic moments, we need a constant matrix $(nm\times nm)$ so that

$$E\left[\text{vec}(\boldsymbol{U}^*_{nm,t})'(I_m\otimes J_n)\mathscr{P}_{nm,h}(I_m\otimes J_n)\text{vec}(\boldsymbol{U}^*_{nm,t})\right] = 0, \quad h = 1,\ldots,H$$

In our notation, the subscript h indexes the constant matrices used to construct the quadratic moment conditions for the system equation GMM estimation. For convenience, we drop the subscript d in the diagonal blocks, as these correspond to the case $i=j$. Similarly, we omit the subscripts s in the off-diagonal blocks to case $i\neq j$.

The constant matrix could be

$$\mathscr{P}_{nm,h} = \begin{pmatrix} P_{11,s} & \dots & P_{1m,d} \\ \vdots & \ddots & \vdots \\ P_{m1,d} & \dots & P_{mm,s} \end{pmatrix}_{nm \times nm}$$

for some (s, d), where $s = 1, \dots S$, and $d = 1, \dots, D$. A block diagonal matrix, where each $(n \times n)$ block is defined in the same way as in the single-equation case. Using different combinations of s and d, we can create a total number of H quadratic orthogonal moment conditions. And note that s and d for different blocks could be different. To gain a deeper insight:

$$\begin{aligned}
&\mathrm{vec}(\boldsymbol{U}^*_{nm,t})'(I_m \otimes J_n)\mathscr{P}_{nm,h}(I_m \otimes J_n)\mathrm{vec}(\boldsymbol{U}^*_{nm,t}) \\
=&\begin{pmatrix} U^{*\prime}_{n1,t} & \dots & U^{*\prime}_{nm,t} \end{pmatrix} \begin{pmatrix} J_n & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & J_n \end{pmatrix} \begin{pmatrix} P_{11,s} & \dots & P_{1m,d} \\ \vdots & \ddots & \vdots \\ P_{m1,d} & \dots & P_{mm,s} \end{pmatrix} \begin{pmatrix} J_n & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & J_n \end{pmatrix} \begin{pmatrix} U^*_{n1,t} \\ \vdots \\ U^*_{nm,t} \end{pmatrix} \\
=&\begin{pmatrix} U^{*\prime}_{n1,t}J_n & \dots & U^{*\prime}_{nm,t}J_n \end{pmatrix} \begin{pmatrix} P_{11,s} & \dots & P_{1m,d} \\ \vdots & \ddots & \vdots \\ P_{m1,d} & \dots & P_{mm,s} \end{pmatrix} \begin{pmatrix} J_nU^*_{n1,t} \\ \vdots \\ J_nU^*_{nm,t} \end{pmatrix} \\
=&\left[(U^{*\prime}_{n1,t}J_nP_{11,s} + \dots + U^{*\prime}_{nm,t}J_nP_{m1,d}) \quad \dots \quad (U^{*\prime}_{n1,t}J_nP_{1m,d} + \dots + U^{*\prime}_{nm,t}J_nP_{mm,s}) \right] \begin{pmatrix} J_nU^*_{n1,t} \\ \vdots \\ J_nU^*_{nm,t} \end{pmatrix} \\
=&U^{*\prime}_{n1,t}J_nP_{11,s}J_nU^*_{n1,t} + U^{*\prime}_{n2,t}J_nP_{21,d}J_nU^*_{n1,t} + \dots + U^{*\prime}_{nm,t}J_nP_{m1,d}J_nU^*_{n1,t} + \dots \\
&+U^{*\prime}_{n1,t}J_nP_{1m,d}J_nU^*_{nm,t} + U^{*\prime}_{n2,t}J_nP_{2m,d}J_nU^*_{nm,t} + \dots + U^{*\prime}_{nm,t}J_nP_{mm,s}J_nU^*_{nm,t}
\end{aligned}$$

As a result, $E\left[\mathrm{vec}(\boldsymbol{U}^*_{nm,t})'(I_m \otimes J_n)\mathscr{P}_{nm,h}(I_m \otimes J_n)\mathrm{vec}(\boldsymbol{U}^*_{nm,t})\right]$ equal to the expectation of each component which is 0. Taking $\boldsymbol{P}_{nm,h,T-1} = I_{T-1} \otimes \mathscr{P}_{nm,h}$, we then have the quadratic moment condition for the system equations:

$$E\left[\boldsymbol{U}^{*\prime}_{nm,T-1}(I_m \otimes \boldsymbol{J}_{n,T-1})\boldsymbol{P}_{nm,h,T-1}(I_m \otimes \boldsymbol{J}_{n,T-1})\boldsymbol{U}^*_{nm,T-1}\right] = 0$$

and the empirical quadratic moment functions are

$$g^2_{nT}(\boldsymbol{\theta}) = \boldsymbol{U}^{*\prime}_{nm,T-1}(\boldsymbol{\theta})(I_m \otimes \boldsymbol{J}_{n,T-1})\boldsymbol{P}_{nm,h,T-1}(I_m \otimes \boldsymbol{J}_{n,T-1})\boldsymbol{U}^*_{nm,T-1}(\boldsymbol{\theta})$$

Combing both linear and quadratic moment conditons, we have

$$g_{nT}(\boldsymbol{\theta}) = \begin{pmatrix} g^1_{nT}(\boldsymbol{\theta}) \\ g^2_{h,nT}(\boldsymbol{\theta}) \end{pmatrix} = \begin{pmatrix} (I_m \otimes \boldsymbol{Q}_{n,T-1})'(I_m \otimes \boldsymbol{J}_{n,T-1})\boldsymbol{U}^*_{nm,T-1}(\boldsymbol{\theta}) \\ \boldsymbol{U}^{*\prime}_{nm,T-1}(\boldsymbol{\theta})(I_m \otimes \boldsymbol{J}_{n,T-1})\boldsymbol{P}_{nm,1,T-1}(I_m \otimes \boldsymbol{J}_{n,T-1})\boldsymbol{U}^*_{nm,T-1}(\boldsymbol{\theta}) \\ \vdots \\ \boldsymbol{U}^{*\prime}_{nm,T-1}(\boldsymbol{\theta})(I_m \otimes \boldsymbol{J}_{n,T-1})\boldsymbol{P}_{nm,H,T-1}(I_m \otimes \boldsymbol{J}_{n,T-1})\boldsymbol{U}^*_{nm,T-1}(\boldsymbol{\theta}) \end{pmatrix} \tag{20}$$

and the system equation GMM estimator is derived as $\hat{\boldsymbol{\theta}}_{nT} = \arg\min g'_{nT}(\boldsymbol{\theta})a'_{nT}a_{nT}g_{nT}(\boldsymbol{\theta})$. where a_{nT} is a matrix with full row rank greater than or equal to $k + 2m + 1$, and is assumed to converge in probability to a constant matrix a_0. We make the following assumptions to impose the moment restrictions on the IV matrix and the uniform boundedness of the constant matrix. Let $\mathscr{L}_{t-1}$ denote the information set that contains all historical information up to time $t-1$.

Assumption 5. *The $n \times k_q$ IV matrix Q_{nt} is predetermined such that $E[Q_{n,t}|\mathscr{L}_{t-1}] = Q_{n,t}$, where k_q is fixed. Let $q_{np,t}$ be an $n \times 1$ column vector from p-th column of Q_{nt}, $p = 1, \ldots, k_q$, $E[|q_{np,t,i}|^2]$ is bounded uniformly in all $i, n,$ and t. The $n \times n$ constant matrix $P_{jl,sd}$ are uniformly bounded in both row and column sums in absolute value.*

3 Identification

Following the identification strategy in Yang and Lee (2017, 2019) and Liu and Saraiva (2019), we first consider the identification of the quasi-reduced form model by the linear moments and quadratic moments, then discuss the identification of the structural parameters which is analogous to the identification for classical simultaneous equation models, as discussed in Schmidt (2020), relying on exclusive restrictions and normalization conditions. We summarize conditions in Assumption (6) and Assumption (7).

3.1 Identification of the Quasi-reduced parameters

From model (7), the reduced form is

$$\begin{aligned}\operatorname{vec}\left(\boldsymbol{Y}^*_{nm,t}\right) =&\boldsymbol{S}_{nm}^{-1}\left[(\boldsymbol{P}'_{m0}\otimes I_n)+(\boldsymbol{\Phi}'_{m0}\otimes I_n)\mathcal{W}_{nm}\right]\operatorname{vec}\left(\boldsymbol{Y}^{(*,-1)}_{nm,t-1}\right)+\boldsymbol{S}_{nm}^{-1}\left[(\boldsymbol{\Pi}'_{m0}\otimes I_n)\operatorname{vec}\left(\boldsymbol{X}^*_{nk,t}\right)+\alpha^*_{m0}\otimes l_n+\operatorname{vec}\left(\boldsymbol{V}^*_{nm,t}\right)\right]\\ =&\boldsymbol{H}_{nm}\operatorname{vec}\left(\boldsymbol{Y}^{(*,-1)}_{nm,t-1}\right)+\boldsymbol{S}_{nm}^{-1}\left[(\boldsymbol{\Pi}'_{m0}\otimes I_n)\operatorname{vec}\left(\boldsymbol{X}^*_{nk,t}\right)+\alpha^*_{m0}\otimes l_n+\operatorname{vec}\left(\boldsymbol{V}^*_{nm,t}\right)\right]\end{aligned} \tag{21}$$

Multiplying $\mathcal{W}_{nm}$, and let $\boldsymbol{G}_{nm}=\mathcal{W}_{nm}\boldsymbol{S}_{nm}^{-1}$, then

$$\begin{aligned}\mathcal{W}_{nm}\operatorname{vec}\left(\boldsymbol{Y}^*_{nm,t}\right) =&\mathcal{W}_{nm}\boldsymbol{S}_{nm}^{-1}\left[(\boldsymbol{P}'_{\mathrm{m0}}\otimes I_n)+(\boldsymbol{\Phi}'_{\mathrm{m0}}\otimes I_n)\mathcal{W}_{nm}\right]\operatorname{vec}\left(\boldsymbol{Y}^{(*,-1)}_{nm,t-1}\right)+\mathcal{W}_{nm}\boldsymbol{S}_{nm}^{-1}\left[(\boldsymbol{\Pi}'_{\mathrm{m0}}\otimes I_n)\operatorname{vec}\left(\boldsymbol{X}^*_{nk,t}\right)+\alpha^*_{m0}\otimes l_n+\operatorname{vec}\left(\boldsymbol{V}^*_{nm,t}\right)\right]\\ =&\boldsymbol{G}_{nm}\left[(\boldsymbol{P}'_{\mathrm{m0}}\otimes I_n)+(\boldsymbol{\Phi}'_{\mathrm{m0}}\otimes I_n)\mathcal{W}_{nm}\right]\operatorname{vec}\left(\boldsymbol{Y}^{(*,-1)}_{nm,t-1}\right)+\boldsymbol{G}_{nm}\left[(\boldsymbol{\Pi}'_{\mathrm{m0}}\otimes I_n)\operatorname{vec}\left(\boldsymbol{X}^*_{nk,t}\right)+\alpha^*_{m0}\otimes l_n+\operatorname{vec}\left(\boldsymbol{V}^*_{nm,t}\right)\right]\\ =&\boldsymbol{G}_{nm}\left[(\boldsymbol{P}'_{\mathrm{m0}}\otimes I_n)\operatorname{vec}\left(\boldsymbol{Y}^{(*,-1)}_{nm,t-1}\right)+(\boldsymbol{\Phi}'_{\mathrm{m0}}\otimes I_n)\mathcal{W}_{nm}\operatorname{vec}\left(\boldsymbol{Y}^{(*,-1)}_{nm,t-1}\right)+(\boldsymbol{\Pi}'_{\mathrm{m0}}\otimes I_n)\operatorname{vec}\left(\boldsymbol{X}^*_{nk,t}\right)+\alpha^*_{m0}\otimes l_n\right]+\boldsymbol{G}_{nm}\operatorname{vec}\left(\boldsymbol{V}^*_{nm,t}\right)\\ =&\boldsymbol{G}_{nm}((\boldsymbol{\vartheta}'_0\otimes I_n)\boldsymbol{Z}^*_t+\alpha^*_{m0}\otimes l_n)+\boldsymbol{G}_{nm}\operatorname{vec}\left(\boldsymbol{V}^*_{nm,t}\right)\\ =&\boldsymbol{L}^*_t+\boldsymbol{G}_{nm}\operatorname{vec}\left(\boldsymbol{V}^*_{nm,t}\right)\end{aligned} \tag{22}$$

with $\boldsymbol{L}^*_t=\boldsymbol{G}_{nm}((\boldsymbol{\vartheta}'_0\otimes I_n)\boldsymbol{Z}^*_t+\alpha^*_{m0}\otimes l_n)$, $\boldsymbol{\delta}=[\boldsymbol{\Psi}'_m,\boldsymbol{\vartheta}'_m]'$, $\boldsymbol{\vartheta}=\left[\boldsymbol{P}'_m,\boldsymbol{\Phi}'_m,\boldsymbol{\Pi}'_m\right]'$, $\boldsymbol{Z}^*_t=\left[\operatorname{vec}\left(\boldsymbol{Y}^{(*,-1)}_{nm,t-1}\right),\mathcal{W}_{nm}\operatorname{vec}\left(\boldsymbol{Y}^{(*,-1)}_{nm,t-1}\right),\operatorname{vec}\left(\boldsymbol{X}^*_{nk,t}\right)\right]$

We first consider how the true parameter vector $\boldsymbol{\delta}_0$ is identified by the linear moment conditions. The parameters in (21) can be identified by the moment conditions similar to (20), the linear moment conditions can be written as

$$f_{nT}^1(\boldsymbol{\delta}) = (I_m \otimes \boldsymbol{Q}_{n,T-1})'(I_m \otimes \boldsymbol{J}_{n,T-1})\boldsymbol{V}^*_{nm,T-1}(\boldsymbol{\delta})$$

With $\boldsymbol{V}^*_{nm,T-1}(\boldsymbol{\delta}) = [\text{vec}(\boldsymbol{V}^{*\prime}_{nm,1}(\boldsymbol{\delta})), \ldots, \text{vec}(\boldsymbol{V}^{*\prime}_{nm,T-1}(\boldsymbol{\delta}))]'$. In the GMM framework of Hansen (1982), identification requires that the linear moment condition $\text{plim}_{n\to\infty}\frac{1}{n(T-1)} f_{nT}^1(\boldsymbol{\delta}) = 0$ has a unique solution at $\boldsymbol{\delta} = \boldsymbol{\delta}_0$. Recall that $\text{vec}\left(\boldsymbol{Y}^*_{nm,t}\right) = \boldsymbol{S}_{nm}^{-1}\left[(\boldsymbol{\vartheta}_0' \otimes I_n)\boldsymbol{Z}_t^* + \alpha^*_{m0} \otimes l_n + \text{vec}(\boldsymbol{V}^*_{nm,t})\right]$,with $\boldsymbol{S}_{nm}(\boldsymbol{\Psi}_m)\boldsymbol{S}_{nm}^{-1} = I_{nm} - \left((\boldsymbol{\Psi}_m' - \boldsymbol{\Psi}_{m0}') \otimes I_n\right)\boldsymbol{G}_{nm}$, we can write the residual function in

$$\begin{aligned}
(I_m \otimes J_n)\text{vec}\left(\boldsymbol{V}^*_{nm,t}(\boldsymbol{\delta})\right) =&(I_m \otimes J_n)\left[\boldsymbol{S}_{nm}(\boldsymbol{\Psi}_m)\boldsymbol{S}_{nm}^{-1}\left((\boldsymbol{\vartheta}_0' \otimes I_n)\boldsymbol{Z}_t^* + \alpha^*_{m0} \otimes l_n\right) - (\boldsymbol{\vartheta}' \otimes I_n)\boldsymbol{Z}_t^* + \boldsymbol{S}_{nm}(\boldsymbol{\Psi}_m)\boldsymbol{S}_{nm}^{-1}\text{vec}(\boldsymbol{V}^*_{nm,t})\right] \\
=&(I_m \otimes J_n)\left[\left((\boldsymbol{\vartheta}_0' - \boldsymbol{\vartheta}') \otimes I_n\right)\boldsymbol{Z}_t^* + \left((\boldsymbol{\Psi}_{m0}' - \boldsymbol{\Psi}_m') \otimes I_n\right)\boldsymbol{L}_t^* + \boldsymbol{S}_{nm}(\boldsymbol{\Psi}_m)\boldsymbol{S}_{nm}^{-1}\text{vec}(\boldsymbol{V}^*_{nm,t})\right] \\
=&(I_m \otimes J_n)\boldsymbol{d}_{nm,t}(\boldsymbol{\delta}) + (I_m \otimes J_n)\boldsymbol{S}_{nm}(\boldsymbol{\Psi}_m)\boldsymbol{S}_{nm}^{-1}\left(\alpha^*_{m0} \otimes l_n\right) + (I_m \otimes J_n)\boldsymbol{S}_{nm}(\boldsymbol{\Psi}_m)\boldsymbol{S}_{nm}^{-1}\text{vec}(\boldsymbol{V}^*_{nm,t})
\end{aligned} \tag{23}$$

with $\boldsymbol{d}_{nm,t}(\boldsymbol{\delta}) = \boldsymbol{S}_{nm}(\boldsymbol{\Psi}_m)\boldsymbol{S}_{nm}^{-1}(\boldsymbol{\vartheta}_0' \otimes I_n)\boldsymbol{Z}_t^* - (\boldsymbol{\vartheta}' \otimes I_n)\boldsymbol{Z}_t^* = \left((\boldsymbol{\vartheta}_0 - \boldsymbol{\vartheta})' \otimes I_n\right)\boldsymbol{Z}_t^* - \left((\boldsymbol{\Psi}_m - \boldsymbol{\Psi}_{m0})' \otimes I_n\right)\boldsymbol{G}_{nm}(\boldsymbol{\vartheta}_0' \otimes I_n)\boldsymbol{Z}_t^*$. Let $\mathcal{Z}^*_{T-1} = [\boldsymbol{Z}_1^{*\prime}, \ldots, \boldsymbol{Z}_{T-1}^{*\prime}]', \mathcal{L}^*_{T-1} = [\boldsymbol{L}_1^{*\prime}, \ldots, \boldsymbol{L}_{T-1}^{*\prime}]', \mathcal{S}_{nm,T-1} = I_{T-1} \otimes \boldsymbol{S}_{nm}$, we have

$$f_{nT}^1(\boldsymbol{\delta}) = (I_m \otimes \boldsymbol{Q}_{n,T-1})'(I_m \otimes \boldsymbol{J}_{n,T-1})\left[\left((\boldsymbol{\vartheta}_0' - \boldsymbol{\vartheta}') \otimes I_{n(T-1)}\right)\mathcal{Z}^*_{T-1} + \left((\boldsymbol{\Psi}_{m0}' - \boldsymbol{\Psi}_m') \otimes I_{n(T-1)}\right)\mathcal{L}^*_{T-1} + \mathcal{S}_{nm,T-1}(\boldsymbol{\Psi}_m)\mathcal{S}_{nm,T-1}^{-1}\boldsymbol{V}^*_{nm,T-1}\right]$$

By Lemma (1)(i) $\text{plim}_{n\to\infty}\frac{1}{n(T-1)}\sum_{t=1}^{T-1} Q_{nt}' J_n S_n(\lambda) S_n^{-1} V_{nt}^* = 0$ uniodormly in $\theta \in \Theta$, the unique solution of $\text{plim}_{n\to\infty} f_{nT}^1(\boldsymbol{\delta}) = 0$ requires that $\text{plim}_{n\to\infty}\frac{1}{n(T-1)}(I_m \otimes \boldsymbol{Q}_{n,T-1})'(I_m \otimes \boldsymbol{J}_{n,T-1})\left(\mathcal{L}^*_{T-1}, \mathcal{Z}^*_T\right)\left(\left((\boldsymbol{\vartheta}_0' - \boldsymbol{\vartheta}') \otimes I_{n(T-1)}\right)', \left((\boldsymbol{\Psi}_{m0}' - \boldsymbol{\Psi}_m') \otimes I_{n(T-1)}\right)'\right)' = 0$ have a unique solution $\boldsymbol{\delta}_0$. It is obvious to see that when $\boldsymbol{\delta} = \boldsymbol{\delta}_0, \text{plim}_{n\to\infty}\frac{1}{n(T-1)} f_{nT}^1(\boldsymbol{\delta}) = 0$. To guarantee that is a unique solution, the sufficient condition is $\text{plim}_{n\to\infty}\frac{1}{n(T-1)}(I_m \otimes \boldsymbol{Q}_{n,T-1})'(I_m \otimes \boldsymbol{J}_{n,T-1})\left(\mathcal{L}^*_{T-1}, \mathcal{Z}^*_T\right)$ has full column rank. Since $\boldsymbol{Z}_t^*$

incorporates not only $\operatorname{vec}\left(\boldsymbol{X}^*_{nk,t}\right)$, but also the time-lagged and spatially-lagged terms $\operatorname{vec}\left(\boldsymbol{Y}^{(*,-1)}_{nm,t-1}\right), \mathcal{W}_{nm}\operatorname{vec}\left(\boldsymbol{Y}^{(*,-1)}_{nm,t-1}\right)$, the rank condition will typically be satisfied in most practical settings as long as $\boldsymbol{\delta}_0 \neq 0$. However, if this condition fails, identification can still be achieved via the quadratic moment conditions. The empirical quadratic moment conditions for the quasi-reduced form of model is

$$f^2_{nT}(\boldsymbol{\delta}) = \begin{pmatrix} \boldsymbol{V}^{*\prime}_{nm,T-1}(\boldsymbol{\delta})(I_m \otimes \boldsymbol{J}_{n,T-1})\boldsymbol{P}_{nm,1,T-1}(I_m \otimes \boldsymbol{J}_{n,T-1})\boldsymbol{V}^*_{nm,T-1}(\boldsymbol{\delta}) \\ \vdots \\ \boldsymbol{V}^{*\prime}_{nm,T-1}(\boldsymbol{\delta})(I_m \otimes \boldsymbol{J}_{n,T-1})\boldsymbol{P}_{nm,H,T-1}(I_m \otimes \boldsymbol{J}_{n,T-1})\boldsymbol{V}^*_{nm,T-1}(\boldsymbol{\delta}) \end{pmatrix}$$

The full column rank condition does not hold in linear moment conditions, if and only if $\boldsymbol{Z}^*_t$ and $\boldsymbol{G}_{nm}(\boldsymbol{\vartheta}'_0 \otimes I_n)\boldsymbol{Z}^*_t$ are linearly dependent, and there exists constant matrix C such that $C\boldsymbol{Z}^*_t = \boldsymbol{G}_{nm}(\boldsymbol{\vartheta}'_0 \otimes I_n)\boldsymbol{Z}^*_t$, then $\boldsymbol{d}_{nm,t}(\boldsymbol{\delta}) = \left[((\boldsymbol{\vartheta}_0 + C'\boldsymbol{\Psi}_{m0}) - (\boldsymbol{\vartheta} + C'\boldsymbol{\Psi}_m))' \otimes I_n\right]\boldsymbol{Z}^*_t$. When $(\boldsymbol{\vartheta}_0 + C'\boldsymbol{\Psi}_{m0}) = (\boldsymbol{\vartheta} + C'\boldsymbol{\Psi}_m)$, $\boldsymbol{d}_{nm,t}(\boldsymbol{\delta}) = 0$. The quadratic moment conditions becomes:

$$\begin{aligned}
& E\left[\operatorname{vec}\left(\boldsymbol{V}^*_{nm,t}(\boldsymbol{\delta})\right)'(I_m \otimes J_n)\mathscr{P}_{nm,h}(I_m \otimes J_n)\operatorname{vec}\left(\boldsymbol{V}^*_{nm,t}(\boldsymbol{\delta})\right)\right] \\
=& E\big\{\left[(I_m \otimes J_n)\boldsymbol{d}_{nm,t}(\boldsymbol{\delta}) + (I_m \otimes J_n)\boldsymbol{S}_{nm}(\boldsymbol{\Psi}_m)\boldsymbol{S}^{-1}_{nm}(\alpha^*_{m0} \otimes l_n) + (I_m \otimes J_n)\boldsymbol{S}_{nm}(\boldsymbol{\Psi}_m)\boldsymbol{S}^{-1}_{nm}\operatorname{vec}(\boldsymbol{V}^*_{nm,t})\right]'\mathscr{P}_{nm,h} \\
& \left[(I_m \otimes J_n)\boldsymbol{d}_{nm,t}(\boldsymbol{\delta}) + (I_m \otimes J_n)\boldsymbol{S}_{nm}(\boldsymbol{\Psi}_m)\boldsymbol{S}^{-1}_{nm}(\alpha^*_{m0} \otimes l_n) + (I_m \otimes J_n)\boldsymbol{S}_{nm}(\boldsymbol{\Psi}_m)\boldsymbol{S}^{-1}_{nm}\operatorname{vec}(\boldsymbol{V}^*_{nm,t})\right]\big\} \\
=& \Sigma_{vnm}\left[tr\left(\boldsymbol{S}^{-1\prime}_{nm}\boldsymbol{S}'_{nm}(\boldsymbol{\Psi}_m)(I_m \otimes J_n)'\mathscr{P}_{nm,h}(I_m \otimes J_n)\boldsymbol{S}_{nm}(\boldsymbol{\Psi}_m)\boldsymbol{S}^{-1}_{nm}\right)\right] \\
=& \Sigma_{vnm} tr\left[\left((\boldsymbol{\Psi}'_{m0} - \boldsymbol{\Psi}'_m) \otimes I_n\right)\boldsymbol{G}_{nm}(I_m \otimes J_n)\mathscr{P}^s_{nm,h}(I_m \otimes J_n)\right] \\
+& \Sigma_{vnm} tr\left[\left((\boldsymbol{\Psi}'_{m0} - \boldsymbol{\Psi}'_m) \otimes I_n\right)'(I_m \otimes J_n)\mathscr{P}_{nm,h}(I_m \otimes J_n)\left((\boldsymbol{\Psi}'_{m0} - \boldsymbol{\Psi}'_m) \otimes I_n\right)\boldsymbol{G}_{nm}\boldsymbol{G}'_{nm}\right]
\end{aligned}$$

Note that

$$\frac{\partial tr\left[\left((\boldsymbol{\Psi}'_{m0}-\boldsymbol{\Psi}'_{m})\otimes I_n\right)\boldsymbol{G}_{nm}(I_m\otimes J_n)\mathscr{P}^{s}_{nm,h}(I_m\otimes J_n)\right]}{\partial \text{vec}(\boldsymbol{\Psi}_m)'}=\text{vec}\left((I_m\otimes J_n)\mathscr{P}^{s\prime}_{nm,h}(I_m\otimes J_n)\boldsymbol{G}'_{nm}\right)'\frac{\partial \text{vec}\left((\boldsymbol{\Psi}'_{m0}-\boldsymbol{\Psi}'_{m})\otimes I_n\right)}{\partial \text{vec}(\boldsymbol{\Psi}_m)'}$$

$$\frac{\partial tr\left[\left((\boldsymbol{\Psi}'_{m0}-\boldsymbol{\Psi}'_{m})\otimes I_n\right)'(I_m\otimes J_n)\mathscr{P}_{nm,h}(I_m\otimes J_n)\left((\boldsymbol{\Psi}'_{m0}-\boldsymbol{\Psi}'_{m})\otimes I_n\right)\boldsymbol{G}_{nm}\boldsymbol{G}'_{nm}\right]}{\partial \text{vec}(\boldsymbol{\Psi}_m)'}$$

$$=\left(\frac{\partial \text{vec}\left((\boldsymbol{\Psi}'_{m0}-\boldsymbol{\Psi}'_{m})\otimes I_n\right)}{\text{vec}(\boldsymbol{\Psi}_m)'}\right)'\left[(\boldsymbol{G}_{nm}\boldsymbol{G}'_{nm})\otimes((I_m\otimes J_n)\mathscr{P}_{nm,h}(I_m\otimes J_n))+(\boldsymbol{G}_{nm}\boldsymbol{G}'_{nm})\otimes((I_m\otimes J_n)\mathscr{P}'_{nm,h}(I_m\otimes J_n))\right]'\frac{\partial \text{vec}\left((\boldsymbol{\Psi}'_{m0}-\boldsymbol{\Psi}'_{m})\otimes I_n\right)}{\text{vec}(\boldsymbol{\Psi}_m)'}$$

$$+\text{vec}\left((\boldsymbol{\Psi}'_{m0}-\boldsymbol{\Psi}'_{m})\otimes I_n\right)'\left[(\boldsymbol{G}_{nm}\boldsymbol{G}'_{nm})\otimes((I_m\otimes J_n)\mathscr{P}_{nm,h}(I_m\otimes J_n))+(\boldsymbol{G}_{nm}\boldsymbol{G}'_{nm})\otimes((I_m\otimes J_n)\mathscr{P}'_{nm,h}(I_m\otimes J_n))\right]\frac{\partial^2 \text{vec}\left((\boldsymbol{\Psi}'_{m0}-\boldsymbol{\Psi}'_{m})\otimes I_n\right)}{\text{vec}(\boldsymbol{\Psi}_m)\text{vec}(\boldsymbol{\Psi}_m)'}$$

and by the Taylor expansion, we have

$$\begin{aligned}
&E\left[\text{vec}\left(\boldsymbol{V}^{*}_{nm,t}(\boldsymbol{\delta})\right)'(I_m\otimes J_n)\mathscr{P}_{nm,h}(I_m\otimes J_n)\text{vec}\left(\boldsymbol{V}^{*}_{nm,t}(\boldsymbol{\delta})\right)\right]\\
=&\text{vec}\left((I_m\otimes J_n)\mathscr{P}^{s\prime}_{nm,h}(I_m\otimes J_n)\boldsymbol{G}'_{nm}\right)'\frac{\partial \text{vec}\left((\boldsymbol{\Psi}'_{m0}-\boldsymbol{\Psi}'_{m})\otimes I_n\right)}{\partial \text{vec}(\boldsymbol{\Psi}_m)'}\text{vec}(\boldsymbol{\Psi}_m-\boldsymbol{\Psi}_{m0})\\
+&\frac{1}{2}\text{vec}(\boldsymbol{\Psi}_m-\boldsymbol{\Psi}_{m0})'\left\{\left(\frac{\partial \text{vec}\left((\boldsymbol{\Psi}'_{m0}-\boldsymbol{\Psi}'_{m})\otimes I_n\right)}{\text{vec}(\boldsymbol{\Psi}_m)'}\right)'\left[(\boldsymbol{G}_{nm}\boldsymbol{G}'_{nm})\otimes((I_m\otimes J_n)\mathscr{P}_{nm,h}(I_m\otimes J_n))\right.\right.\\
+&\left.\left.(\boldsymbol{G}_{nm}\boldsymbol{G}'_{nm})\otimes((I_m\otimes J_n)\mathscr{P}'_{nm,h}(I_m\otimes J_n))\right]'\frac{\partial \text{vec}\left((\boldsymbol{\Psi}'_{m0}-\boldsymbol{\Psi}'_{m})\otimes I_n\right)}{\text{vec}(\boldsymbol{\Psi}_m)'}\right\}\text{vec}(\boldsymbol{\Psi}_m-\boldsymbol{\Psi}_{m0})
\end{aligned}$$

We then discuss the following four cases:

- $(I_m\otimes J_n)\mathscr{P}^{s}_{nm,h}(I_m\otimes J_n)\boldsymbol{G}'_{nm}=0,\quad (\boldsymbol{G}_{nm}\boldsymbol{G}'_{nm})\otimes((I_m\otimes J_n)\mathscr{P}_{nm,h}(I_m\otimes J_n))=0.$

  $\boldsymbol{\Psi}_{m0}$ is unidentifiable. All quadratic moment conditions vanish, leading to many roots.

- $(I_m\otimes J_n)\mathscr{P}^{s}_{nm,h}(I_m\otimes J_n)\boldsymbol{G}'_{nm}=0,\quad (\boldsymbol{G}_{nm}\boldsymbol{G}'_{nm})\otimes((I_m\otimes J_n)\mathscr{P}_{nm,h}(I_m\otimes J_n))\neq 0.$

$(I_m \otimes J_n)\mathscr{P}^s_{nm,h}(I_m \otimes J_n)\boldsymbol{G}'_{nm} = 0$ implies there is no linear correlation between IVs and endogenous variables.

- $(I_m \otimes J_n)\mathscr{P}^s_{nm,h}(I_m \otimes J_n)\boldsymbol{G}'_{nm} \neq 0, \quad (\boldsymbol{G}_{nm}\boldsymbol{G}'_{nm}) \otimes ((I_m \otimes J_n)\mathscr{P}_{nm,h}(I_m \otimes J_n)) = 0.$

  There is a single root if $(I_m \otimes J_n)\mathscr{P}^s_{nm,h}(I_m \otimes J_n)\boldsymbol{G}'_{nm}$ has full column rank. Then $\boldsymbol{\Psi}_{m0}$ could be uniquely identified by the quadratic moment conditions.

- $(I_m \otimes J_n)\mathscr{P}^s_{nm,h}(I_m \otimes J_n)\boldsymbol{G}'_{nm} \neq 0, \quad (\boldsymbol{G}_{nm}\boldsymbol{G}'_{nm}) \otimes ((I_m \otimes J_n)\mathscr{P}_{nm,h}(I_m \otimes J_n)) \neq 0.$

  If $[(I_m \otimes J_n)\mathscr{P}^s_{nm,1}(I_m \otimes J_n)\boldsymbol{G}'_{nm}, \ldots, (I_m \otimes J_n)\mathscr{P}^s_{nm,H}(I_m \otimes J_n)\boldsymbol{G}'_{nm}]'$ and $[(\boldsymbol{G}_{nm}\boldsymbol{G}'_{nm}) \otimes ((I_m \otimes J_n)\mathscr{P}_{nm,1,T-1}(I_m \otimes J_n)) \neq 0, \ldots, (\boldsymbol{G}_{nm}\boldsymbol{G}'_{nm}) \otimes ((I_m \otimes J_n)\mathscr{P}_{nm,H,T-1}(I_m \otimes J_n)) \neq 0]'$ are linearly independent, identification of the $\boldsymbol{\Psi}_{m0}$ can be established provided that distinctive moment conditions are imposed.

Assumption 6. *Either (i)* $plim_{n\to\infty}\frac{1}{n(T-1)}(I_m \otimes \boldsymbol{Q}_{n,T-1})'(I_m \otimes \boldsymbol{J}_{n,T-1})\left(\mathcal{L}^*_{T-1}, \mathcal{Z}^*_{T-1}\right)$ *exists and has full column rank, or (ii)* $lim_{n\to\infty}\frac{1}{n(T-1)}(I_m \otimes \boldsymbol{Q}_{n,T-1})'(I_m \otimes \boldsymbol{J}_{n,T-1})\mathcal{Z}^*_{T-1}$ *has full column rank ,* $lim_{n\to\infty}\frac{1}{n(T-1)}(I_m \otimes \boldsymbol{J}_{n,T-1})\mathscr{P}^s_{nm,h,T-1}(I_m \otimes \boldsymbol{J}_{n,T-1})\boldsymbol{G}'_{nm,T-1} \neq 0$ *for some* h*, and* $lim_{n\to\infty}\frac{1}{n(T-1)}[(I_m \otimes \boldsymbol{J}_{n,T-1})\mathscr{P}^s_{nm,1,T-1}(I_m \otimes \boldsymbol{J}_{n,T-1})\boldsymbol{G}'_{nm,T-1}, \ldots, (I_m \otimes \boldsymbol{J}_{n,T-1})\mathscr{P}^s_{nm,H,T-1}(I_m \otimes \boldsymbol{J}_{n,T-1})\boldsymbol{G}'_{nm,T-1}]'$ *and* $lim_{n\to\infty}\frac{1}{n(T-1)}[(\boldsymbol{G}_{nm,T-1}\boldsymbol{G}'_{nm,T-1}) \otimes ((I_m \otimes \boldsymbol{J}_{n,T-1})\mathscr{P}_{nm,1,T-1}(I_m \otimes \boldsymbol{J}_{n,T-1})), \cdots, (\boldsymbol{G}_{nm,T-1}\boldsymbol{G}'_{nm,T-1}) \otimes ((I_m \otimes \boldsymbol{J}_{n,T-1})\mathscr{P}_{nm,H,T-1}(I_m \otimes \boldsymbol{J}_{n,T-1}))]'$ *are linearly independent.*

3.2 Identification of the structural parameters

Once the quasi-reduced form parameters $\boldsymbol{\delta}_0$ are identified from the linear and quadratic moment conditions, the identification of the structural parameters $\boldsymbol{\theta}_0$ proceeds analogously to the identification strategy for classical simultaneous equation models, relying on exclusive restrictions, linear restrictions, and normalization conditions. Let R_j be a $r \times (4m+k)$ matrix representing all linear restrictions on the coefficients of the j-th equation such that $R_j\boldsymbol{\delta}_j = 0$. The necessary order condition is $r \geq m-1$. The necessary and sufficient rank condition is $rank(R_j\boldsymbol{\theta}_j) = m-1$. We impose the following assumption to ensure the identification of the structure form coefficients.

Assumption 7. *For $j = 1, \ldots, m$, $R_j \boldsymbol{\delta}_j = 0$ for some $r \times (4m + k)$ constant matrix R_j with $rank(R_j \boldsymbol{\delta}_j) = m - 1$*

4 Asymptotic properties of GMME

In this section, we analyze the asymptotic properties of the GMM estimator. Our focus is on the scenario where n is large while T may be either large or small. For the asymptotic distribution of GMME, and the optimal GMME, we first derive the variance matrix of the moment conditions. Denote $\boldsymbol{\Xi}_{H,T-1}$ be the variance matrix of quadratic moments, with the typical elements:

$$\begin{aligned}
\xi_{hh} =& Var\left(\boldsymbol{U}^{*\prime}_{nm,T-1}(I_m \otimes \boldsymbol{J}_{n,T-1})\boldsymbol{P}_{nm,h,T-1}(I_m \otimes \boldsymbol{J}_{n,T-1})\boldsymbol{U}^{*}_{nm,T-1}\right) \\
\xi_{hh'} =& Cov\left[\left(\boldsymbol{U}^{*\prime}_{nm,T-1}(I_m \otimes \boldsymbol{J}_{n,T-1})\boldsymbol{P}_{nm,h,T-1}(I_m \otimes \boldsymbol{J}_{n,T-1})\boldsymbol{U}^{*}_{nm,T-1}\right), \left(\boldsymbol{U}^{*\prime}_{nm,T-1}(I_m \otimes \boldsymbol{J}_{n,T-1})\boldsymbol{P}_{nm,h',T-1}(I_m \otimes \boldsymbol{J}_{n,T-1})\boldsymbol{U}^{*}_{nm,T-1}\right)\right]
\end{aligned}$$

The expression of ξ_{hh} and $\xi_{hh'}$ are shown in Appendix. The variance matrix of the moments in (20) is:

$$\boldsymbol{\Lambda}_{nT} = \begin{pmatrix} \boldsymbol{\Xi}_{H,T-1} & \boldsymbol{0}_{H\times k_q} \\ \boldsymbol{0}_{k_q \times H} & \Omega \otimes \left[\boldsymbol{Q}'_{n,T-1}\boldsymbol{J}_{n,T-1}\boldsymbol{Q}_{n,T-1}\right] \end{pmatrix} \tag{24}$$

Theorem 1. *Under Assumption 1-7, as $n \to \infty$, the GMME $\hat{\boldsymbol{\theta}}_{gmm}$, obtained by minimizing $g'_{nT}(\boldsymbol{\theta})a'_{nT}a_{nT}g_{nT}(\boldsymbol{\theta})$ is consistent and*

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{gmm} - \boldsymbol{\theta}_0) \xrightarrow{d} N\left(0, plim_{n\to\infty}\frac{1}{T-1}(\mathscr{D}'_{nT}a'_{nT}a_{nT}\mathscr{D}_{nT})^{-1}\mathscr{D}'_{nT}a'_{nT}a_{nT}\Sigma_{um0}a'_{nT}a_{nT}\mathscr{D}_{nT}(\mathscr{D}'_{nT}a'_{nT}a_{nT}\mathscr{D}_{nT})^{-1}\right)$$

when T is large,

$$\sqrt{nT}(\hat{\boldsymbol{\theta}}_{gmm} - \boldsymbol{\theta}_0) \xrightarrow{d} N\left(0, plim_{n\to\infty}(D'_{nT}a'_{nT}a_{nT}D_{nT})^{-1}D'_{nT}a'_{nT}a_{nT}\Sigma_{um0}a'_{nT}a_{nT}D_{nT}(D'_{nT}a'_{nT}a_{nT}D_{nT})^{-1}\right)$$

By the generalized Schwarz inequality, Σ_{um0}^{-1} is the optimal GMM weighting matrix, but it involves the true parameter, the optimal GMM weighting matrix is infeasible. However, it could be consistently estimated by Theorem (1).

Theorem 2. *Under Assumption 1-7, as $n \to \infty$, suppose that $\hat{\Sigma}_{um}^{-1} - \Sigma_{um0}^{-1} = o_p(1)$, the OGMME $\hat{\boldsymbol{\theta}}_{ogmm}$, obtained by minimizing $g'_{nT}(\boldsymbol{\theta})\hat{\Sigma}_{um}^{-1}g_{nT}(\boldsymbol{\theta})$ is consistent and*

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{ogmm} - \boldsymbol{\theta}_0) \xrightarrow{d} N\left(0, plim_{n\to\infty}\frac{1}{T-1}(\mathscr{D}'_{nT}\hat{\Sigma}_{um}^{-1}\mathscr{D}_{nT})^{-1}\right)$$

when T is large,

$$\sqrt{nT}(\hat{\boldsymbol{\theta}}_{ogmm} - \boldsymbol{\theta}_0) \xrightarrow{d} N\left(0, plim_{n\to\infty}(D'_{nT}\hat{\Sigma}_{um}^{-1}D)^{-1}\right)$$

As the covariance matrix of moments (24) is a block diagonal (the linear moment and quadratic moment conditions do not interact), we can improve the asymptotic efficiency by choose the "best" moments conditions. The best moment conditions involve individual fixed effects. When T is small, the individual fixed effects cannot be consistently estimated, however, when T is large, best moment conditions are available since we can consistently estimate those individual fixed effects.

Theorem 3. *Under Assumption 1-7, suppose that $\hat{\Sigma}_{um}^{-1} - \Sigma_{um0}^{-1} = o_p(1)$, and we use the best moment conditions. As both $n \to \infty$ and $T \to \infty$, the feasible best GMME (BGMME) is consistent and $\sqrt{n(T-1)}(\hat{\boldsymbol{\theta}}_{bgmm} - \boldsymbol{\theta}_0) \xrightarrow{d} N\left(0, \boldsymbol{\Sigma}_{um,b}^{-1}\right)$.*

5 Impulse response function and associated asymptotics

We first derive the moving average form from model (7)

$$\mathrm{vec}(\boldsymbol{Y}^*_{nm,t}) = \sum_{v=0}^{\infty} B_{0,v}(\boldsymbol{\Pi}'_{ms0} \otimes I_n)\mathrm{vec}(\boldsymbol{X}^*_{n,t}) + \sum_{v=0}^{\infty} B_{0,v}(\alpha^*_{ms0,t-v} \otimes l_n) + \sum_{v=0}^{\infty} B_{0,v}\mathrm{vec}(\boldsymbol{U}_{nm,t-v}) \tag{25}$$

by defining $\mathcal{D}_0 = (\boldsymbol{\Gamma}'_{m0} \otimes I_n) - (\boldsymbol{\Psi}'_{ms0} \otimes I_n)\mathcal{W}_{nm}, \mathcal{R}_0 = (\boldsymbol{P}'_{ms0} \otimes I_n) + (\boldsymbol{\Phi}'_{ms0} \otimes I_n)\mathcal{W}_{nm}, B_{0,\tau} = \mathcal{D}_0^{-1}\mathcal{R}_0 B_{0,\tau-1}, B_{0,0} = \mathcal{D}_0^{-1}$, and $B_{0,\tau} = 0$ if $\tau < 0$. Although the error terms are i.i.d. across time and spatial units, they may exhibit contemporaneous cross-variable correlation - that is, correlations among different variables within the same spatial unit at time t. This dependence structure is formally characterized by (10), where the off-diagonal blocks are nonzero. To orthogonalize the correlated errors for impulse response analysis, we decompose the covariance matrix using singular value decomposition (SVD) to Ω_{um0}. Let $T_{nm0} = \Omega_{um0}^{1/2} \otimes I_n = \begin{pmatrix} t_{11} & \dots & t_{1m} \\ \vdots & \ddots & \vdots \\ t_{m1} & \dots & t_{mm} \end{pmatrix} \otimes I_n$, $T_{nm0}^{-1} = \Omega_{um0}^{-\frac{1}{2}} \otimes I_n = \begin{pmatrix} \omega_{11} & \dots & \omega_{1m} \\ \dots & \ddots & \vdots \\ \omega_{m1} & \dots & \omega_{mm} \end{pmatrix} \otimes I_n$, and define $\mathrm{Vec}(\tilde{\boldsymbol{U}}^*_{nm,t}) = T_{nm0}^{-1}\mathrm{Vec}(\boldsymbol{U}^*_{nm,t})$, the model becomes

$$\mathrm{vec}(\boldsymbol{Y}^*_{nm,t}) = \sum_{v=0}^{\infty} B_{0,v}(\boldsymbol{\Pi}'_{ms0} \otimes I_n)\mathrm{vec}(\boldsymbol{X}^*_{n,t}) + \sum_{v=0}^{\infty} B_{0,v}(\alpha^*_{ms0,t-v} \otimes l_n) + \sum_{v=0}^{\infty} B_{0,v}T_{nm0}\mathrm{vec}(\boldsymbol{U}_{nm,t-v}) \tag{26}$$

To derive the response of the dependent variable at time $h+\tau$ to a standardized error shock at time h:

$$\frac{\partial \boldsymbol{Y}_{nm,h+\tau}}{\partial \tilde{\boldsymbol{U}}'_{nm,h}} = \frac{\partial (Y_{n1,h+\tau} \dots Y_{nm,h+\tau})'}{\partial (\tilde{U}_{n1,h} \dots \tilde{U}_{nm,h})} = \begin{pmatrix} \frac{\partial Y_{n1,h+\tau}}{\partial \tilde{U}'_{n1,h}} & \dots & \frac{\partial Y_{n1,h+\tau}}{\partial \tilde{U}'_{nm,h}} \\ \vdots & \ddots & \vdots \\ \frac{\partial Y_{nm,h+\tau}}{\partial \tilde{U}'_{n1,h}} & \dots & \frac{\partial Y_{nm,h+\tau}}{\partial \tilde{U}'_{nm,h}} \end{pmatrix}$$

with the typical block:

$$\frac{\partial Y_{ni,h+\tau}}{\partial \tilde{U}'_{nj,h}} = \underset{\mathbf{n\times n}}{\begin{pmatrix} \frac{\partial y_{1i,h+\tau}}{\partial \tilde{u}_{1j,h}} & \cdots & \frac{\partial y_{1i,h+\tau}}{\partial \tilde{u}_{nj,h}} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_{ni,h+\tau}}{\partial \tilde{u}_{1j,h}} & \cdots & \frac{\partial y_{ni,h+\tau}}{\partial \tilde{u}_{nj,h}} \end{pmatrix}} = B_{0\tau,i1}t_{1j} + \cdots + B_{0\tau,im}t_{mj} = \sum_{q=1}^{m} B_{0\tau,iq}t_{qj}$$

represents the response of i-th dependent variable at time $h+\tau$ to the shock in j-th equation's standardized error at time h, holding all else constant. The term $B_{0\tau,iq}$ is the (i,q)-th block of $B_{0\tau}$ which is an $n\times n$ matrix, and t_{qj} is a scalar. Then define the (l_1,l_2)-th element of $\sum_{q=1}^{m} B_{0\tau,iq}t_{qj}$ as $\sum_{q=1}^{m} B_{0\tau,iq,l_1l_2}t_{qj}$ with $(l_1,l_2)\in\{1,\cdots,n\}$. That is, we have $\frac{\partial y_{l_1,i,h+\tau}}{\partial \tilde{u}_{l_2,j,h}} = \sum_{q=1}^{m} B_{0\tau,iq,l_1l_2}t_{qj}$. Here $y_{l_1,i,h+\tau}$ is the dependent variable of the l_1-th individual from the ith equation observed at time $h+\tau$, and $\tilde{u}_{l_2,j,h}$ denotes the error term of the l_2-th individual from the jth equation observed at time h. The consistent estimators of $B_{0\tau,iq,l_1l_2}$ and t_{qj} are $\hat{B}_{0\tau,iq,l_1l_2}$ and $\hat{t}_{qj}$, respectively, where $\hat{B}_\tau = \hat{\mathcal{D}}^{-1}\hat{\mathcal{R}}\hat{B}_{\tau-1}$, $\hat{B}_0 = \hat{\mathcal{D}}^{-1}$, $\hat{B}_\tau = 0$ if $\tau<0$. $\hat{\mathcal{D}} = (\hat{\mathbf{\Gamma}}'_m \otimes I_n) - (\hat{\mathbf{\Psi}}'_{ms}\otimes I_n)\mathcal{W}_{nm}$, $\hat{\mathcal{R}} = (\hat{\boldsymbol{P}}'_{ms}\otimes I_n) + (\hat{\mathbf{\Phi}}'_{ms}\otimes I_n)\mathcal{W}_{nm}$ and $\hat{t}_{qj}$ is the (q,j)-th element of $\hat{\Omega}^{1/2}$. In addition to analyzing the responses from the error shock, we can compute the effects of changes in the regressor on the dependent variable.

$$\frac{\partial \boldsymbol{Y}_{nm,h+\tau}}{\partial \boldsymbol{X}'_{nk,h}} = \frac{\partial (Y_{n1,h+\tau}\ldots Y_{nm,h+\tau})'}{\partial (X_{n1,h}\ldots X_{nk,h})} = \begin{pmatrix} \frac{\partial Y_{n1,h+\tau}}{\partial X_{n1,h}} & \cdots & \frac{\partial Y_{n1,h+\tau}}{\partial X_{nk,h}} \\ \vdots & \ddots & \vdots \\ \frac{\partial Y_{nm,h+\tau}}{\partial X_{n1,h}} & \cdots & \frac{\partial Y_{nm,h+\tau}}{\partial X_{nk,h}} \end{pmatrix}$$

with the typical block:

$$\frac{\partial Y_{ni,h+\tau}}{\partial X'_{nl,h}} = \begin{pmatrix} \frac{\partial y_{1i,h+\tau}}{\partial x_{1l,h}} & \cdots & \frac{\partial y_{1i,h+\tau}}{\partial x_{nl,h}} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_{ni,h+\tau}}{\partial x_{1l,h}} & \cdots & \frac{\partial y_{ni,h+\tau}}{\partial x_{nl,h}} \end{pmatrix} = B_{0\tau,i1}\pi_{l1} + \cdots + B_{0\tau,im}\pi_{lm} = \sum_{q=1}^{m} B_{0\tau,iq}\pi_{lq}$$

represents the response of i-th dependent variable at time $h+\tau$ to an one-unit increase in the l-th regressor at time h, holding all else constant. Define the l_1, l_2 element of $\sum_{q=1}^{m} B_{0\tau,iq}\pi_{lq}$ as $\sum_{q=1}^{m} B_{0\tau,iq,l_1l_2}\pi_{lq}$ with $l_1, l_2 \in \{1,\dots,n\}$. That is, we have $\frac{\partial y_{l_1,i,h+\tau}}{\partial x_{l_2,l,h}} = \sum_{q=1}^{m} B_{0\tau,iq,l_1l_2}\pi_{lq}$ represents the response of l_1-th spatial unit in i-th dependent variable at time $h+\tau$ to a one-unit increase in l_2-th spatial unit in l-th regressor at time h, holding all else constant. And the consistent estimator of π_{lq} is $\hat{\pi}_{lq}$.

The asymptotic properties of the above estimators depend on the asymptotic distribution of GMM estimator. When T is large, we can use Theorem (3), when T is small we use Theorem (2). o apply Delta Method, we first compute the derivative of the estimators with respect to $\boldsymbol{\theta}_0$. Let $\gamma_{i,j}$ denotes the coefficient of the i-th dependent variable in the j-th equation, similar to $\psi_{ij}, p_{ij}, \phi_{ij}, \pi_{ij}$. Denote $\nabla B_{0\tau,iq}^{\gamma_{ij}}$ represents the derivative of $B_{0\tau,iq}$ with respect to γ_{ij}, similar to $\nabla B_{0\tau,iq}^{p_{ij}}, \nabla B_{0\tau,iq}^{\phi_{ij}}, \nabla B_{0\tau,iq}^{\psi ij}, \nabla B_{0\tau,iq}^{\pi_{ij}}$. Then,

$$\frac{\partial \sum_{q=1}^{m} B_{0\tau,iq,l_1l_2}t_{qj}}{\partial \gamma_{ij}} = \sum_{q=1}^{m} \left(e'_{l1}\nabla B_{0\tau,iq}^{\gamma_{ij}} e_{l2}\right) t_{qj}, \quad \frac{\partial \sum_{q=1}^{m} B_{0\tau,iq,l_1l_2}t_{qj}}{\partial p_{ij}} = \sum_{q=1}^{m} \left(e'_{l1}\nabla B_{0\tau,iq}^{p_{ij}} e_{l2}\right) t_{qj}, \quad \frac{\partial \sum_{q=1}^{m} B_{0\tau,iq,l_1l_2}t_{qj}}{\partial \phi_{ij}} = \sum_{q=1}^{m} \left(e'_{l1}\nabla B_{0\tau,iq}^{\phi_{ij}} e_{l2}\right) t_{qj}$$

$$\frac{\partial \sum_{q=1}^{m} B_{0\tau,iq,l_1l_2}t_{qj}}{\partial \psi_{ij}} = \sum_{q=1}^{m} \left(e'_{l1}\nabla B_{0\tau,iq}^{\psi_{ij}} e_{l2}\right) t_{qj}, \quad \frac{\partial \sum_{q=1}^{m} B_{0\tau,iq,l_1l_2}t_{qj}}{\partial \pi_{lj}} = 0$$

Here, e_{l1} and e_{l2} are selection vectors, each of dimension $n \times 1$, with a 1 in the l_1-th and l_2-th positions, respectively. Let

$$\vartheta_{ij,\tau}^{\tilde{u}} = \left[\sum_{q=1}^{m} \left(e_{l1}' \nabla B_{0\tau,iq}^{\gamma_{ij}} e_{l2} \right) t_{qj}, \sum_{q=1}^{m} \left(e_{l1}' \nabla B_{0\tau,iq}^{p_{ij}} e_{l2} \right) t_{qj}, \sum_{q=1}^{m} \left(e_{l1}' \nabla B_{0\tau,iq}^{\phi_{ij}} e_{l2} \right) t_{qj}, \sum_{q=1}^{m} \left(e_{l1}' \nabla B_{0\tau,iq}^{\psi_{ij}} e_{l2} \right) t_{qj}, 0 \right]'$$

Similarly, the derivative of $\sum_{q=1}^{m} B_{0\tau,iq} \pi_{lq}$ with respect to parameters:

$$\vartheta_{ij,\tau}^{x} = \left[\sum_{q=1}^{m} \left(e_{l1}' \nabla B_{0\tau,iq}^{\gamma_{ij}} e_{l2} \right) \pi_{lq}, \sum_{q=1}^{m} \left(e_{l1}' \nabla B_{0\tau,iq}^{p_{ij}} e_{l2} \right) \pi_{lq}, \sum_{q=1}^{m} \left(e_{l1}' \nabla B_{0\tau,iq}^{\phi_{ij}} e_{l2} \right) \pi_{lq}, \sum_{q=1}^{m} \left(e_{l1}' \nabla B_{0\tau,iq}^{\psi_{ij}} e_{l2} \right) \pi_{lq}, B_{0\tau,iq,l_1 l_2} \delta_{qj} \right]'$$

with $\delta_{qj} = 1$ if $q = j$, 0 otherwise. Then we have $\frac{\partial \sum_{q=1}^{m} B_{0\tau,iq,l_1 l_2} t_{qj}}{\partial \theta'} = \vartheta_{ij,\tau}^{\tilde{u}}$, $\frac{\partial \sum_{q=1}^{m} B_{0\tau,iq,l_1 l_2} \pi_{lq}}{\partial \theta'} = \vartheta_{ij,\tau}^{x}$. Detailed derivations are provided in Appendix.

Theorem 4. *Under Assumption 1-7, as $n \to \infty$, suppose that $\hat{\Sigma}_{um}^{-1} - \Sigma_{um0}^{-1} = o_p(1)$, and use the OGMME from Theorem(2), then*

$$\sqrt{n} \left(\sum_{q=1}^{m} \hat{B}_{0\tau,iq,l_1 l_2} \hat{t}_{qj} - \sum_{q=1}^{m} B_{0\tau,iq,l_1 l_2} t_{qj} \right) \xrightarrow{d} N \left(0, plim_{n\to\infty} \frac{1}{T-1} \vartheta_{ij,\tau}^{\tilde{u}\prime} (\mathscr{D}_{nT}' \hat{\Sigma}_{um}^{-1} \mathscr{D}_{nT})^{-1} \vartheta_{ij,\tau}^{\tilde{u}} \right)$$

$$\sqrt{n} \left(\sum_{q=1}^{m} \hat{B}_{0\tau,iq,l_1 l_2} \hat{\pi}_{lq} - \sum_{q=1}^{m} B_{0\tau,iq,l_1 l_2} \pi_{lq} \right) \xrightarrow{d} N \left(0, plim_{n\to\infty} \frac{1}{T-1} \vartheta_{ij,\tau}^{x\prime} (\mathscr{D}_{nT}' \hat{\Sigma}_{um}^{-1} \mathscr{D}_{nT})^{-1} \vartheta_{ij,\tau}^{x} \right)$$

as $n, T \to \infty$,, and use the BGMME from Theorem (3), then

$$\sqrt{n(T-1)}\left(\sum_{q=1}^{m}\hat{B}_{0\tau,iq,l_1l_2}\hat{t}_{qj}-\sum_{q=1}^{m}B_{0\tau,iq,l_1l_2}t_{qj}\right)\xrightarrow{d}N\left(0,\vartheta_{ij,\tau}^{\tilde{u}\prime}\mathbf{\Sigma}_{um,b}^{-1}\vartheta_{ij,\tau}^{\tilde{u}}\right)$$

$$\sqrt{n(T-1)}\left(\sum_{q=1}^{m}\hat{B}_{0\tau,iq,l_1l_2}\hat{\pi}_{lq}-\sum_{q=1}^{m}B_{0\tau,iq,l_1l_2}\pi_{lq}\right)\xrightarrow{d}N\left(0,\vartheta_{ij,\tau}^{x\prime}\mathbf{\Sigma}_{um,b}^{-1}\vartheta_{ij,\tau}^{x}\right)$$

6 ADI, AII, ATI, and and associated asymptotics

Referencing to the definition of average direct impact (ADI), average indirect impact (AII), and average total impact (ATI) in LeSage and Pace (2009), we establish these three impacts within the framework of spatial simultaneous equations model. Specifically, the impacts of i-th dependent variable due to the j-th standardize error after τ periods are defined as

$$\mathrm{ADI}_{\tau,ij}^{\tilde{U}}=\frac{1}{N}\sum_{q=1}^{m}t_{qj}\mathrm{tr}(B_{0\tau,iq}),\ \mathrm{ATI}_{\tau,ij}^{\tilde{U}}\frac{1}{N}\sum_{q=1}^{m}t_{qj}(l_n'B_{0\tau,iq}l_n),\ \mathrm{AII}_{\tau,ij}^{\tilde{U}}\frac{1}{N}\sum_{q=1}^{m}t_{qj}(l_n'B_{0\tau,iq}l_n-\mathrm{tr}(B_{0\tau,iq}))$$

the impacts of i-th dependent variable due to the l-th regressor after τ periods are defined as

$$\mathrm{ADI}_{\tau,il}^{X}=\frac{1}{N}\sum_{q=1}^{m}\pi_{lq}\mathrm{tr}(B_{0\tau,iq}),\ \mathrm{ATI}_{\tau,il}^{X}=\frac{1}{N}\sum_{q=1}^{m}\pi_{lq}(l_n'B_{0\tau,iq}l_n),\ \mathrm{AII}_{\tau,il}^{X}=\frac{1}{N}\sum_{q=1}^{m}\pi_{lq}(l_n'B_{0\tau,iq}l_n-\mathrm{tr}(B_{\tau,iq}))$$

The consistent estimators are

$$\widehat{\text{ADI}}^{\tilde{U}}_{\tau,ij} = \frac{1}{N}\sum_{q=1}^{m}\hat{t}_{qj}\text{tr}(\hat{B}_{0\tau,iq}),\ \widehat{\text{ATI}}^{\tilde{U}}_{\tau,ij}\frac{1}{N}\sum_{q=1}^{m}\hat{t}_{qj}(l_n'\hat{B}_{0\tau,iq}l_n),\ \widehat{\text{AII}}^{\tilde{U}}_{\tau,ij}\frac{1}{N}\sum_{q=1}^{m}\hat{t}_{qj}(l_n'\hat{B}_{0\tau,iq}l_n - \text{tr}(\hat{B}_{0\tau,iq}))$$

$$\widehat{\text{ADI}}^{X}_{\tau,il} = \frac{1}{N}\sum_{q=1}^{m}\hat{\pi}_{lq}\text{tr}(\hat{B}_{0\tau,iq}),\ \widehat{\text{ATI}}^{X}_{\tau,il} = \frac{1}{N}\sum_{q=1}^{m}\hat{\pi}_{lq}(l_n'\hat{B}_{0\tau,iq}l_n),\ \widehat{\text{AII}}^{X}_{\tau,il} = \frac{1}{N}\sum_{q=1}^{m}\hat{\pi}_{lq}(l_n'\hat{B}_{0\tau,iq}l_n - \text{tr}(\hat{B}_{\tau,iq}))$$

To derive the asymptotic results for these estimated impacts, we introduce following notation. Denote $\mathbf{v}^d_\tau(X), \mathbf{v}^t_\tau(X), \mathbf{v}^i_\tau(X)$ represents the derivative of impacts with respect to parameters.

$$\begin{aligned}
\mathbf{v}^d_\tau(\tilde{U}) =& \frac{1}{N}\left[\sum_{q=1}^{m}t_{qj}\text{tr}(\nabla B^{\gamma_{ij}}_{0\tau,iq}), \sum_{q=1}^{m}t_{qj}\text{tr}(\nabla B^{p_{ij}}_{0\tau,iq}), \sum_{q=1}^{m}t_{qj}\text{tr}(\nabla B^{\phi_{ij}}_{0\tau,iq}), \sum_{q=1}^{m}t_{qj}\text{tr}(\nabla B^{\psi_{ij}}_{0\tau,iq}), 0\right]' \\
\mathbf{v}^t_\tau(\tilde{U}) =& \frac{1}{N}\left[\sum_{q=1}^{m}t_{qj}(l_n'\nabla B^{\gamma_{ij}}_{0\tau,iq}l_n), \sum_{q=1}^{m}t_{qj}(l_n'\nabla B^{p_{ij}}_{0\tau,iq}l_n), \sum_{q=1}^{m}t_{qj}(l_n'\nabla B^{\phi_{ij}}_{0\tau,iq}l_n), \sum_{q=1}^{m}t_{qj}(l_n'\nabla B^{\psi_{ij}}_{0\tau,iq}l_n), 0\right]' \\
\mathbf{v}^i_\tau(\tilde{U}) =& \mathbf{v}^t_\tau(\tilde{U}) - \mathbf{v}^d_\tau(\tilde{U}) \\
\mathbf{v}^d_\tau(X) =& \frac{1}{N}\left[\sum_{q=1}^{m}\pi_{lq}\text{tr}(\nabla B^{\gamma_{ij}}_{0\tau,iq}), \sum_{q=1}^{m}\pi_{lq}\text{tr}(\nabla B^{p_{ij}}_{0\tau,iq}), \sum_{q=1}^{m}\pi_{lq}\text{tr}(\nabla B^{\phi_{ij}}_{0\tau,iq}), \sum_{q=1}^{m}\pi_{lq}\text{tr}(\nabla B^{\psi_{ij}}_{0\tau,iq}), \sum_{q=1}^{m}\text{tr}(B_{0\tau,iq})\delta_{qj}\right]' \\
\mathbf{v}^t_\tau(X) =& \frac{1}{N}\left[\sum_{q=1}^{m}(l_n'\nabla B^{\gamma_{ij}}_{0\tau,iq}l_n), \sum_{q=1}^{m}(l_n'\nabla B^{p_{ij}}_{0\tau,iq}l_n), \sum_{q=1}^{m}(l_n'\nabla B^{\phi_{ij}}_{0\tau,iq}l_n), \sum_{q=1}^{m}(l_n'\nabla B^{\psi_{ij}}_{0\tau,iq}l_n), \sum_{q=1}^{m}(l_n'B_{0\tau,iq}l_n)\delta_{qj}, 0\right]' \\
\mathbf{v}^i_\tau(X) =& \mathbf{v}^t_\tau(X) - \mathbf{v}^d_\tau(X)
\end{aligned}$$

where the Kronecker delta is defined as $\delta_{qj} = 1$if $q = j$, 0 otherwise. Detailed derivations are provided in Appendix. Then we have

$$\frac{\partial \mathrm{ADI}^{\tilde{U}}_{\tau,ij}}{\partial\theta_0}=\mathbf{v}^d_\tau(\tilde{U}),\quad \frac{\partial \mathrm{ATI}^{\tilde{U}}_{\tau,ij}}{\partial\theta_0}=\mathbf{v}^t_\tau(\tilde{U}),\quad \frac{\partial \mathrm{AII}^{\tilde{U}}_{\tau,ij}}{\partial\theta_0}=\mathbf{v}^i_\tau(\tilde{U})\frac{\partial \mathrm{ADI}^{X}_{\tau,il}}{\partial\theta_0}=\mathbf{v}^d_\tau(X),\quad \frac{\partial \mathrm{ATI}^{X}_{\tau,il}}{\partial\theta_0}=\mathbf{v}^t_\tau(X),\quad \frac{\partial \mathrm{AII}^{X}_{\tau,il}}{\partial\theta_0}=\mathbf{v}^i_\tau(X)$$

Theorem 5. *Under Assumption 1-7, as $n\to\infty$, suppose that $\hat{\Sigma}^{-1}_{um}-\Sigma^{-1}_{um0}=o_p(1)$, and use the OGMME from Theorem(2), then*

$$\sqrt{n}\left(\widehat{ADI}^{\tilde{U}}_{\tau,ij}-ADI^{\tilde{U}}_{\tau,ij}\right)\xrightarrow{d}N\left(0,plim_{n\to\infty}\frac{1}{T-1}\bar{\mathbf{v}}^d_\tau(\tilde{U})'(\mathscr{D}'_{nT}\hat{\Sigma}^{-1}_{um}\mathscr{D}_{nT})^{-1}\bar{\mathbf{v}}^d_\tau(\tilde{U})\right)$$
$$\sqrt{n}\left(\widehat{ATI}^{\tilde{U}}_{\tau,ij}-ATI^{\tilde{U}}_{\tau,ij}\right)\xrightarrow{d}N\left(0,plim_{n\to\infty}\frac{1}{T-1}\bar{\mathbf{v}}^t_\tau(\tilde{U})'(\mathscr{D}'_{nT}\hat{\Sigma}^{-1}_{um}\mathscr{D}_{nT})^{-1}\bar{\mathbf{v}}^t_\tau(\tilde{U})\right)$$
$$\sqrt{n}\left(\widehat{AII}^{\tilde{U}}_{\tau,ij}-AII^{\tilde{U}}_{\tau,ij}\right)\xrightarrow{d}N\left(0,plim_{n\to\infty}\frac{1}{T-1}\bar{\mathbf{v}}^i_\tau(\tilde{U})'(\mathscr{D}'_{nT}\hat{\Sigma}^{-1}_{um}\mathscr{D}_{nT})^{-1}\bar{\mathbf{v}}^i_\tau(\tilde{U})\right)$$
$$\sqrt{n}\left(\widehat{ADI}^{X}_{\tau,il}-ADI^{X}_{\tau,il}\right)\xrightarrow{d}N\left(0,plim_{n\to\infty}\frac{1}{T-1}\bar{\mathbf{v}}^d_\tau(X)'(\mathscr{D}'_{nT}\hat{\Sigma}^{-1}_{um}\mathscr{D}_{nT})^{-1}\bar{\mathbf{v}}^d_\tau(X)\right)$$
$$\sqrt{n}\left(\widehat{ATI}^{X}_{\tau,il}-ATI^{X}_{\tau,il}\right)\xrightarrow{d}N\left(0,plim_{n\to\infty}\frac{1}{T-1}\bar{\mathbf{v}}^t_\tau(X)'(\mathscr{D}'_{nT}\hat{\Sigma}^{-1}_{um}\mathscr{D}_{nT})^{-1}\bar{\mathbf{v}}^t_\tau(X)\right)$$
$$\sqrt{n}\left(\widehat{AII}^{X}_{\tau,il}-AII^{X}_{\tau,il}\right)\xrightarrow{d}N\left(0,plim_{n\to\infty}\frac{1}{T-1}\bar{\mathbf{v}}^i_\tau(X)'(\mathscr{D}'_{nT}\hat{\Sigma}^{-1}_{um}\mathscr{D}_{nT})^{-1}\bar{\mathbf{v}}^i_\tau(X)\right)$$

as $n,T\to\infty$,, and use the BGMME from Theorem (3), then

$$\sqrt{n(T-1)}\left(\widehat{ADI}^{X}_{\tau,il}-ADI^{X}_{\tau,il}\right)\xrightarrow{d}N\left(0,\bar{\mathbf{v}}^d_\tau(X)'\mathbf{\Sigma}^{-1}_{um,b}\bar{\mathbf{v}}^d_\tau(X)\right)$$
$$\sqrt{n(T-1)}\left(\widehat{ATI}^{X}_{\tau,il}-ATI^{X}_{\tau,il}\right)\xrightarrow{d}N\left(0,\bar{\mathbf{v}}^t_\tau(X)'\mathbf{\Sigma}^{-1}_{um,b}\bar{\mathbf{v}}^t_\tau(X)\right)$$
$$\sqrt{n(T-1)}\left(\widehat{AII}^{X}_{\tau,il}-AII^{X}_{\tau,il}\right)\xrightarrow{d}N\left(0,\bar{\mathbf{v}}^i_\tau(X)'\mathbf{\Sigma}^{-1}_{um,b}\bar{\mathbf{v}}^i_\tau(X)\right)$$

where $\bar{\mathbf{v}}_\tau^a(\tilde{U}) = \lim_{n\to\infty} \mathbf{v}_\tau^a(\tilde{U})$ *with* $a = d, t$ *and* i.

7 Monte Carlo Simulations

We conduct Monte Carlo simulations to investigate the finite sample performance of the GMM estimator, IRF. We use the same data generating process (DGP) as Yang and Lee (2019):

$$\boldsymbol{Y}_{nm,t}\boldsymbol{\Gamma}_{m0} = \begin{pmatrix} W_{1n}Y_{n1,t} & W_{2n}Y_{n2,t} \end{pmatrix} \boldsymbol{\Psi}_{ms0} + \boldsymbol{Y}_{nm,t-1}\boldsymbol{P}_{ms0} + \begin{pmatrix} W_{1n}Y_{n1,t-1} & W_{2n}Y_{n2,t-1} \end{pmatrix} \boldsymbol{\Phi}_{ms0} + \boldsymbol{X}_{n2,t}\boldsymbol{\Pi}_{ms0} + \boldsymbol{C}_{nms} + \alpha'_{ms0,t} \otimes l_n + \boldsymbol{U}_{nm,t}$$

where

$$\begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -0.2 & 1 \end{pmatrix}, \begin{pmatrix} \psi_{11} & \psi_{12} \\ \psi_{21} & \psi_{22} \end{pmatrix} = \begin{pmatrix} 0.5 & 0.3 \\ 0 & -0.5 \end{pmatrix}, \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} = \begin{pmatrix} 0.3 & 0.3 \\ 0.3 & -0.3 \end{pmatrix}$$
$$\begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix} = \begin{pmatrix} -0.2 & -0.2 \\ -0.2 & -0.2 \end{pmatrix}, \begin{pmatrix} \pi_1 & 0 \\ 0 & \pi_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \mathcal{W} = \begin{pmatrix} W_1 & 0 \\ 0 & W_2 \end{pmatrix}$$

then the DGP model is explicitly

$$y_{1,t} - 0.2y_{2,t} = 0.5W_1^* y_{1,t} + 0.3y_{1,t-1} + 0.3y_{2,t-1} - 0.2W_1^* y_{1,t-1} - 0.2W_2^* y_{2,t-1} + x_{1,t} + c_1 + d_{1,t} + v_{1,t},$$
$$y_{2,t} = 0.3W_1^* y_{1,t} - 0.5W_2^* y_{2,t} + 0.3y_{1,t-1} - 0.3y_{2,t-1} - 0.2W_1^* y_{1,t-1} - 0.2W_2^* y_{2,t-1} + x_{2,t} + c_2 + d_{2,t} + v_{2,t}$$

The spatial weight matrices are constructed as follows: W_1 is generated randomly under the queen contiguity criterion, while W_2 follows the rook contiguity rule. Both matrices are row-standardized. The exogenous variables $(x_{1,t}, x_{2,t})$, individual effects

(c_1, c_2), and time effects $(d_{1,t}, d_{2,t})$ are independently drawn from $\mathcal{U}[0,1]$. The error terms are randomly generated from the multivariate normal distribution $N(0, \Sigma)$, with $\Sigma = \begin{pmatrix} 1 & 0.8 \\ 0.8 & 1 \end{pmatrix} \otimes I_n$. In this experiment, we use sample size $(n, T) = (100, 10)$ to simulate the case large$n$, small T, use $(n, T) = (30, 30)$ to simulate the case large$n$, small T, where $n/T \to M$. We construct 300 repetitions for each sample size, and report the empirical bias and standard deviation. The estimation methods considered are

1. Single equation GMM
2. System equation GMM
3. System equation OGMM
4. System equation BGMM
5. QMLE from Yang and Lee (2019), and only use sample size $(n, T) = (30, 30)$ to compare.

We allow each dependent variable may have different or same spatial weight matrices to fit more empirical studies.

8 Conclusion

This paper develops the GMM estimation framework for spatial dynamic panel simultaneous equations models with two-way fixed effects. To address the fixed effects, we employ FOD transformations and incorporate demeaning matrices directly into our moment conditions, effectively avoiding the incidental parameters problem. Our framework accommodates both asymptotic regimes where where $n, T \to \infty$ and $n \to \infty$ with fixed T. The proposed estimators achieve $\sqrt{nT}$ or $\sqrt{n}$ consistency and asymptotically normal. We further study impulse response functions in this model and derive the limiting distribution.

9 Appendix

9.1 Some Lemmas

Lemma 1. *Under Assumption (1), for any $nm \times nm$ nonstochastic matrix $\boldsymbol{B}_{nm}$ and B_n satisfying row and column sums uniformly bounded in absolute value. Then,*

(i) $\frac{1}{n(T-1)}\sum_{t=1}^{T-1} Q'_{n,t} B_n U^*_{nl,t} = O_p\left(\frac{1}{\sqrt{nT}}\right)$

(ii) $E\left(Vec(\boldsymbol{U}^*_{nm,t})'\boldsymbol{B}_{nm}\, Vec(\boldsymbol{U}^*_{nm,t})\right) = tr(\boldsymbol{B}_{nm}\Sigma_{um0}) = O(n)$

$Var\left(Vec(\boldsymbol{U}^*_{nm,t})'\boldsymbol{B}_{nm}\, Vec(\boldsymbol{U}^*_{nm,t})\right) = O(n)$

$Vec(\boldsymbol{U}^*_{nm,t})'\boldsymbol{B}_{nm}\, Vec(\boldsymbol{U}^*_{nm,t}) = O_p(n)$

$\frac{1}{n(T-1)}\sum_{t=1}^{T-1} Vec(\boldsymbol{U}^*_{nm,t})'\boldsymbol{B}_{nm}\, Vec(\boldsymbol{U}^*_{nm,t}) - \frac{1}{n}tr(\boldsymbol{B}_{nm}\Sigma_{um0}) = O_p\left(\frac{1}{\sqrt{nT}}\right) = o_p(1)$

(iii) $E\frac{1}{n(T-1)}\boldsymbol{Y}^{(*,-1)\prime}_{nm,T-1}\boldsymbol{B}_{nm,T-1}\boldsymbol{U}^*_{nm,T-1} = \frac{1}{n(T-1)}tr\left[\sum_{h=1}^{T-1}\left(1-\frac{h}{T}\right)\boldsymbol{H}^{h-1}_{nm}\boldsymbol{S}^{-1}_{nm}(\boldsymbol{\Gamma}^{-1\prime}_{m0}\otimes I_n)\Sigma_{um0}B'_{nm}\right]$

Proof. Let M denote a positive constant that can take different values at different occurrences. Denote $C_{Tt} = \left(\frac{T-t}{T-t+1}\right)^{\frac{1}{2}}$, which has the order $O(1)$.

(i) From Assumption (5), $Q_{n,t}$ is predetermined such that $E[Q_{n,t}|\mathscr{L}_{t-1}] = Q_{n,t}$, and from the orthogonal condition, $E(U^*_{nl,t}|Q_{n,t}) = 0$ for all t, where $\mathscr{L}_{t-1}$ denotes the set containing all historical information up to time $t-1$. Let $Q_{np,t}$ denote the p-th column of $Q_{nt}, p = 1, \ldots, k_q$, since $Q_{n,t} \subseteq \mathscr{L}_{t-1}$, then

$$E(U^*_{nl,t}|\mathscr{L}_{t-1}) = E[E(U^*_{nl,t}|\mathscr{L}_{t-1}, Q_{n,t})|\mathscr{L}_{t-1}] = E[E(U^*_{nl,t}|Q_{n,t})|\mathscr{L}_{t-1}] = 0$$

and

$$E[Q'_{nj,t}B_nU^*_{nl,t}] = E[E(Q'_{nj,t}B_nU^*_{nl,t}|\mathscr{L}_{t-1})] = E[Q'_{nj,t}B_nE(U^*_{nl,t}|\mathscr{L}_{t-1})] = 0$$

For the variance:

$$\operatorname{Var}\left(\sum_{t=1}^{T-1} Q'_{np,t}B_nU^*_{nl,t}\right) = \sum_{t=1}^{T-1}\sum_{s=1}^{T-1} \operatorname{Cov}(Q'_{np,t}B_nU^*_{nl,t}, Q'_{np,s}B_nU^*_{nl,s})$$

where

$$\operatorname{Cov}(Q'_{np,t}B_nU^*_{nl,t}, Q'_{np,s}B_nU^*_{nl,s}) = E(Q'_{np,t}B_nU^*_{nl,t}U^{*\prime}_{nl,s}B'_nQ_{np,s})$$

when $t = s$:

$$E(Q'_{np,t}B_nU^*_{nl,t}U^{*\prime}_{nl,t}B'_nQ_{np,t}) = \sigma_{ll}E(Q'_{np,t}B_nQ_{np,t})$$

when $t \neq s$:

$$E(Q'_{np,t}B_nU^*_{nl,t}U^{*\prime}_{nl,s}B'_nQ_{np,s}) = 0$$

because if $t > s$:

$$E(Q'_{np,t}B_nU^*_{nl,t}U^{*\prime}_{nl,s}B'_nQ_{np,s}) = E[E(Q'_{np,t}B_nU^*_{nl,t}U^{*\prime}_{nl,s}B'_nQ_{np,s}|\mathscr{L}_{t-1})] = E[Q'_{np,t}B_nE(U^*_{nl,t}|\mathscr{L}_{t-1})U^{*\prime}_{nl,s}B'_nQ_{np,s}]$$

Similar to the case if $s < t$, hence

$$\text{Var}\left(\sum_{t=1}^{T-1} Q'_{np,t} B_n U^*_{nl,t}\right) = \sigma_{ll} \sum_{t=1}^{T-1} E(Q'_{np,t} B_n B'_n Q_{np,t})$$

where

$$\begin{aligned} E(Q'_{np,t} B_n B'_n Q_{np,t}) \leq & \lambda_{max}(B_n B'_n) E(\|Q_{np,t}\|^2) \leq \|B_n B'_n\|_2 E(\|Q_{np,t}\|^2) \leq \|B_n\|_2^2 E(\|Q_{np,t}\|^2) \\ \leq & \|B_n\|_1 \|B_n\|_\infty E(\|Q_{np,t}\|^2) \leq E\left(\sum_{i=1}^{n} Q^2_{np,t,i}\right) = nM = O(n) \end{aligned}$$

by Assumption (5), $E[|Q_{np,t,i}|^2]$ is bounded uniformly in all $i, n,$ and t. Thus $\text{Var}\left(\sum_{t=1}^{T-1} Q'_{np,t} B_n U^*_{nl,t}\right) = O(nT)$. And Chebyshev inequality gives

$$\frac{1}{n(T-1)} \sum_{t=1}^{T-1} Q'_{np,t} B_n U^*_{nl,t} = O_p\left(\frac{1}{\sqrt{nT}}\right) = o_p(1)$$

(ii)

$$\mathrm{Vec}(\boldsymbol{U}^*_{nm,t})'\boldsymbol{B}_{nm}\mathrm{Vec}(\boldsymbol{U}^*_{nm,t}) = \begin{bmatrix} U^*_{n1,t} & \dots & U^*_{nm,t} \end{bmatrix} \begin{bmatrix} B_{11} & \dots & B_{1m} \\ \vdots & \ddots & \vdots \\ B_{m1} & \dots & B_{mm} \end{bmatrix} \begin{bmatrix} U^*_{n1,t} \\ \vdots \\ U^*_{nm,t} \end{bmatrix}$$

$$= (U^{*\prime}_{n1,t}B_{11}U^*_{n1,t} + \cdots + U^{*\prime}_{nm,t}B_{m1}U^*_{n1,t}) + \cdots + (U^{*\prime}_{n1,t}B_{1m}U^*_{nm,t} + \cdots + U^{*\prime}_{nm,t}B_{mm}U^*_{nm,t})$$

$$= \sum_{k=1}^{m}\sum_{l=1}^{m} U^{*\prime}_{nk,t}B_{kl}U^*_{nl,t}$$

$$\mathrm{Vec}(\boldsymbol{U}^*_{nm,t})'\boldsymbol{B}_{nm}\mathrm{Vec}(\boldsymbol{U}^*_{nm,t})\mathrm{Vec}(\boldsymbol{U}^*_{nm,t})'\boldsymbol{B}_{nm}\mathrm{Vec}(\boldsymbol{U}^*_{nm,t})$$

$$= \left(U^{*\prime}_{n1,t}B_{11}U^*_{n1,t}U^{*\prime}_{n1,t}B_{11}U^*_{n1,t} + \cdots + U^{*\prime}_{nm,t}B_{m1}U^*_{n1,t}U^{*\prime}_{nm,t}B_{m1}U^*_{n1,t}\right)$$

$$+ \cdots + \left(U^{*\prime}_{n1,t}B_{1m}U^*_{nm,t}U^{*\prime}_{n1,t}B_{1m}U^*_{nm,t} + \cdots + U^{*\prime}_{nm,t}B_{mm}U^*_{nm,t}U^{*\prime}_{nm,t}B_{mm}U^*_{nm,t}\right)$$

$$+ 2U^{*\prime}_{n1,t}B_{11}U^*_{n1,t}U^{*\prime}_{n2,t}B_{21}U^*_{n1,t} + \cdots + 2U^{*\prime}_{n1,t}B_{11}U^*_{n1,t}U^{*\prime}_{nm,t}B_{mm}U^*_{nm,t}$$

$$+ \cdots + 2U^{*\prime}_{n(m-1),t}B_{(m-1)m}U^*_{nm,t}U^{*\prime}_{nm,t}B_{mm}U^*_{nm,t}$$

$$= \sum_{k=1}^{m}\sum_{l=1}^{m}\sum_{p=1}^{m}\sum_{q=1}^{m} U^{*\prime}_{nk,t}B_{kl}U^*_{nl,t}U^{*\prime}_{np,t}B_{pq}U^*_{nq,t}$$

To derive the variance of $\left(\mathrm{Vec}(\boldsymbol{U}^*_{nm,t})'\boldsymbol{B}_{nm}\mathrm{Vec}(\boldsymbol{U}^*_{nm,t})\right)$, we first compute the expectation, then its square, and finally the expectation of the squared term.

$$\begin{aligned}
E\left[\mathrm{Vec}(\boldsymbol{U}^*_{nm,t})'\boldsymbol{B}_{nm}\mathrm{Vec}(\boldsymbol{U}^*_{nm,t})\right] =& E\left[(U^{*\prime}_{n1,t}B_{11}U^*_{n1,t}+\cdots+U^{*\prime}_{nm,t}B_{m1}U^*_{n1,t})+\cdots+(U^{*\prime}_{n1,t}B_{1m}U^*_{nm,t}+\cdots+U^{*\prime}_{nm,t}B_{mm}U^*_{nm,t})\right]\\
=&\left[(\sigma_{11}tr(B_{11})+\cdots+\sigma_{1m}tr(B_{m1}))+\cdots+(\sigma_{1m}tr(B_{1m})+\cdots+\sigma_{mm}tr(B_{mm}))\right]\\
=&tr(\boldsymbol{B}_{nm}\Sigma_{um0})\\
\left(E\left[\mathrm{Vec}(\boldsymbol{U}^*_{nm,t})'\boldsymbol{B}_{nm}\mathrm{Vec}(\boldsymbol{U}^*_{nm,t})\right]\right)^2 =&\Big[\left(\sigma^2_{11}tr^2(B_{11})+\cdots+\sigma^2_{1m}tr^2(B_{m1})\right)+\cdots+\left(\sigma^2_{1m}tr^2(B_{1m})+\cdots+\sigma^2_{mm}tr^2(B_{mm})\right)\\
&+2\sigma_{11}\sigma_{12}tr(B_{11})tr(B_{21})+\cdots+2\sigma_{11}\sigma_{1m}tr(B_{11})tr(B_{m1})+\cdots+2\sigma_{(m-1)m}\sigma_{mm}tr(B_{(m-1)m})tr(B_{mm})\Big]\\
=&tr^2(\boldsymbol{B}_{nm}\Sigma_{um0})\\
=&O(n^2)
\end{aligned}$$

For the the expectation of the squared term, we first look at the square term of own equation (l-th equation):

$$E\left[U^{*\prime}_{nl,t}B_{ll}U^*_{nl,t}U^{*\prime}_{nl,t}B_{ll}U^*_{nl,i,t}\right]=C^4_{Tt}\left(1+\frac{1}{(T-t)^3}\right)\sum_{i=1}^{n}B^2_{ll,ii}\left(E(u_{nl,i,t})^4-3\sigma^2_{ll}\right)+\sigma^2_{ll}[tr^2(B_{ll})+tr(B_{ll}B'_{ll})+tr(B_{ll})^2]=O(n^2),\ l=1,\ldots,m$$

The leading term is $\sigma_{ll}^2 tr^2(B_{ll}) = O(n^2)$. Then, look at the square term of cross equation (l-th and l'-th equations):

$$
\begin{aligned}
E[U_{nl,t}^{*\prime} B_{ll'} U_{nl',t}^{*} U_{nl,t}^{*\prime} B_{ll'} U_{nl',t}^{*}] =& E\left(\sum_{i=1}^{n}\sum_{j=1}^{n} u_{nl,i,t}^{*} u_{nl',j,t}^{*} B_{ll',ij}\right)^2 \\
=& E\left(\sum_{i=1}^{n} u_{nl,i,t}^{*} A_i\right)^2 \text{ with } A_i = \sum_{j=1}^{n} u_{nl',j,t}^{*\prime} B_{ll',ij} \\
=& E\left(\sum_{i=1}^{n} u_{nl,i,t}^{*2} A_i^2\right) + E\left(\sum_{i=1}^{n}\sum_{j\neq i} u_{nl,i,t}^{*} u_{nl,j,t}^{*} A_i A_j\right)
\end{aligned}
$$

where

$$
\begin{aligned}
E\left(\sum_{i=1}^{n} u_{nl,i,t}^{*2} A_i^2\right) =& E\left[\sum_{i=1}^{n} u_{nl,i,t}^{*2}\left(\sum_{j=1}^{n} u_{nl',j,t}^{*2} B_{ll',ij}^2 + \sum_{p\neq j}^{n} u_{nl',j,t}^{*} u_{nl',p,t}^{*} B_{ll',ij} B_{ll',ip}\right)\right] \\
=& \sum_{i=1}^{n}\sum_{j=1}^{n} E\left(u_{nl,i,t}^{*2} u_{nl',j,t}^{*2}\right) B_{ll',ij}^2 \\
=& \sum_{i=1}^{n} E\left(u_{nl,i,t}^{*2} u_{nl',i,t}^{*2}\right) B_{ll',ii}^2 + \sum_{i=1}^{n}\sum_{j\neq i}^{n} E\left(u_{nl,i,t}^{*2} u_{nl',j,t}^{*2}\right) B_{ll',ij}^2 \\
=& \left[C_{Tt}^4\left(1 + \frac{1}{(T-t)^3}\right)\sum_{i=1}^{n} B_{ll',ii}^2\left[E\left(u_{nl,i,t}^2 u_{nl',i,t}^2\right) - \sigma_{ll}\sigma_{l'l'}\right] + \sigma_{ll}\sigma_{l'l'} tr(B_{ll'} B_{ll'}')\right]
\end{aligned}
$$

and

$$
\begin{aligned}
&E\left(\sum_{i=1}^{n}\sum_{j\neq i}^{n} u^*_{nl,i,t}u^*_{nl,j,t}A_iA_j\right)\\
=&E\left[\sum_{i=1}^{n}\sum_{j\neq i}^{n} u^*_{nl,i,t}u^*_{nl,j,t}\left(u^*_{nl',i,t}B_{ll',ii}+\sum_{p\neq i}^{n}u^*_{nl',p,t}B_{ll',ip}\right)\left(u^*_{nl',j,t}B_{ll',jj}+\sum_{q\neq i}^{n}u^*_{nj',q,t}B_{ll',jq}\right)\right]\\
=&\sum_{i=1}^{n}\sum_{j\neq i}^{n}E[u^*_{nl,i,t}u^*_{nl,j,t}u^*_{nl',i,t}u^*_{nl',j,t}]B_{ll',ii}B_{ll',jj}+\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{q\neq i}^{n}E[u^*_{nl,i,t}u^*_{nl,j,t}u^*_{nl',i,t}u^*_{nj',q,t}]B_{ll',ii}B_{ll',jq}\\
+&\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{p\neq i}^{n}E[u^*_{nl,i,t}u^*_{nl,j,t}u^*_{nl',j,t}u^*_{nl',p,t}]B_{ll',ip}B_{ll',jj}+\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{p\neq i}^{n}\sum_{q\neq i}^{n}E[u^*_{nl,i,t}u^*_{nl,j,t}u^*_{nl',p,t}u^*_{nj',q,t}]B_{ll',ip}B_{ll',jq}\\
=&\sum_{i=1}^{n}\sum_{j\neq i}^{n}E[u^*_{nl,i,t}u^*_{nl,j,t}u^*_{nl',i,t}u^*_{nl',j,t}]B_{ll',ii}B_{ll',jj}+\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{p\neq i}^{n}\sum_{q\neq i}^{n}E[u^*_{nl,i,t}u^*_{nl,j,t}u^*_{nl',p,t}u^*_{nj',q,t}]B_{ll',ip}B_{ll',jq}\\
=&\sigma^2_{ll'}\left[tr^2(B_{ll'})+tr(B^2_{ll'})\right]-2C^4_{Tt}\left(1+\frac{1}{(T-t)^3}\right)\sum_{i=1}^{n}B^2_{ll',ii}
\end{aligned}
$$

thus

$$
E[U^{*\prime}_{nl,t}B_{ll'}U^*_{nl',t}U^{*\prime}_{nl,t}B_{ll'}U^*_{nl',t}]=\left[C^4_{Tt}\left(1+\frac{1}{(T-t)^3}\right)\sum_{i=1}^{n}B^2_{ll',ii}\left(E\left(u^2_{nl,i,t}u^2_{nl',i,t}\right)-\sigma_{ll}\sigma_{l'l'}-2\sigma^2_{ll'}\right)+\sigma_{ll}\sigma_{l'l'}tr(B_{ll'}B'_{ll'})+\sigma^2_{ll'}\left[tr^2(B_{ll'})+tr(B^2_{ll'})\right]\right]
$$

and the leading term is $\sigma_{ll'}^2 tr^2(B_{ll'}) = O(n^2)$. Finally, the interaction term:

$$\begin{aligned}
E[2U_{nl,t}^{*\prime}B_{ll}U_{nl,t}^{*}U_{nl',t}^{*\prime}B_{l'l}U_{nl,t}^{*}] =& 2C_{Tt}^4\left(1+\frac{1}{(T-t)^3}\right)\sum_{i=1}^{n}B_{ll,ii}B_{l'l,ii}\left[E(u_{nl,i,t}^3u_{nl',i,t})-3\sigma_{ll}\sigma_{ll'}\right]\\
&+2\sigma_{ll}\sigma_{ll'}\left[tr(B_{ll})tr(B_{l'l})+tr(B_{ll}B_{l'l}')+tr(B_{ll}B_{l'l})\right]\\
E[2U_{nl,t}^{*\prime}B_{ll}U_{nl,t}^{*}U_{nl',t}^{*\prime}B_{l'l'}U_{nl',t}^{*}] =& 2C_{Tt}^4\left(1+\frac{1}{(T-t)^3}\right)\sum_{i=1}^{n}B_{ll,ii}B_{l'l',ii}\left[E(u_{nl,i,t}^2u_{nl',i,t}^2)-\sigma_{ll}\sigma_{l'l'}-2\sigma_{ll'}^2\right]\\
&+2\sigma_{ll}\sigma_{l'l'}tr(B_{ll})tr(B_{l'l'})+2\sigma_{ll'}^2[tr(B_{ll}B_{l'l'}')+tr(B_{ll}B_{l'l'})]\\
E[2U_{nl,t}^{*\prime}B_{ll}U_{nl,t}^{*}U_{nl',t}^{*\prime}B_{l'k}U_{nk,i,t}^{*}] =& 2C_{Tt}^4\left(1+\frac{1}{(T-t)^3}\right)\sum_{i=1}^{n}B_{ll,ii}B_{l'k,ii}\left[E(u_{nl,i,t}^2u_{nl',i,t}u_{nk,i,t})-\sigma_{ll}\sigma_{l'k}-2\sigma_{ll'}\sigma_{lk}\right]\\
&+2\sigma_{ll}\sigma_{l'k}tr(B_{ll})tr(B_{l'k})+2\sigma_{ll'}\sigma_{lk}[tr(B_{ll}B_{l'k}')+tr(B_{ll}B_{l'k})]\\
E[2U_{nl,t}^{*\prime}B_{ll'}U_{nl',t}^{*}U_{nk,t}^{*\prime}B_{kk'}U_{nk',t}^{*}] =& 2C_{Tt}^4\left(1+\frac{1}{(T-t)^3}\right)\sum_{i=1}^{n}B_{ll',ii}B_{kk',ii}\left[E(u_{nl,i,t}u_{nl',i,t}u_{nk,i,t}u_{nk',i,t})-\sigma_{ll'}\sigma_{kk'}-\sigma_{lk}\sigma_{l'k'}-\sigma_{lk'}\sigma_{l'k}\right]\\
&+2\left[\sigma_{ll'}\sigma_{kk'}tr(B_{ll'})tr(B_{kk'})+\sigma_{lk}\sigma_{l'k'}tr(B_{ll'}B_{kk'}')+\sigma_{lk'}\sigma_{l'k}tr(B_{ll'}B_{kk'})\right]
\end{aligned}$$

and the leading term is $2\sigma_{ll}\sigma_{ll'}tr(B_{ll})tr(B_{l'l}) = O(n^2), 2\sigma_{ll}\sigma_{l'l'}tr(B_{ll})tr(B_{l'l'}) = O(n^2), 2\sigma_{ll}\sigma_{l'k}tr(B_{ll})tr(B_{l'k}) = O(n^2), 2\sigma_{ll'}\sigma_{kk'}tr(B_{ll'})tr(B_{kk'}) = O(n^2)$ respectively. We then apply the variance formula:

$$\text{Var}\left(\text{Vec}(\boldsymbol{U}_{nm,t}^*)'\boldsymbol{B}_{nm}\text{Vec}(\boldsymbol{U}_{nm,t}^*)\right) = E\left(\text{Vec}(\boldsymbol{U}_{nm,t}^*)'\boldsymbol{B}_{nm}\text{Vec}(\boldsymbol{U}_{nm,t}^*)\right)^2 - \left[E\left(\text{Vec}(\boldsymbol{U}_{nm,t}^*)'\boldsymbol{B}_{nm}\text{Vec}(\boldsymbol{U}_{nm,t}^*)\right)\right]^2$$

Note that the expansion of $E\left(\text{Vec}(\boldsymbol{U}_{nm,t}^*)'\boldsymbol{B}_{nm}\text{Vec}(\boldsymbol{U}_{nm,t}^*)\right)^2$ contains dominant terms of order $O(n^2)$ that are identical to those in $\left[E\left(\text{Vec}(\boldsymbol{U}_{nm,t}^*)'\boldsymbol{B}_{nm}\text{Vec}(\boldsymbol{U}_{nm,t}^*)\right)\right]^2$, these $O(n^2)$ terms cancel exactly in the variance formula. Hence, we can conclude that $\text{Var}\left(\text{Vec}(\boldsymbol{U}_{nm,t}^*)'\boldsymbol{B}_{nm}\text{Vec}(\boldsymbol{U}_{nm,t}^*)\right) = O(n)$, $\text{Var}\left(\frac{1}{n(T-1)}\sum_{t=1}^{T-1}\text{Vec}(\boldsymbol{U}_{nm,t}^*)'\boldsymbol{B}_{nm}\text{Vec}(\boldsymbol{U}_{nm,t}^*)\right) = O\left(\frac{1}{nT}\right)$. As

$E\left(\mathrm{Vec}(\boldsymbol{U}^*_{nm,t})'\boldsymbol{B}_{nm}\mathrm{Vec}(\boldsymbol{U}^*_{nm,t})\right)^2 = O(n^2)$, the generalized Chebyshev inequality implies $\mathrm{Vec}(\boldsymbol{U}^*_{nm,t})'\boldsymbol{B}_{nm}\mathrm{Vec}(\boldsymbol{U}^*_{nm,t}) = O_p(n)$ and $\frac{1}{n(T-1)}\sum_{t=1}^{T-1}\mathrm{Vec}(\boldsymbol{U}^*_{nm,t})'\boldsymbol{B}_{nm}\mathrm{Vec}(\boldsymbol{U}^*_{nm,t}) - \frac{1}{n(T-1)}\sum_{t=1}^{T-1}E\left[\mathrm{Vec}(\boldsymbol{U}^*_{nm,t})'\boldsymbol{B}_{nm}\mathrm{Vec}(\boldsymbol{U}^*_{nm,t})\right] = o_p(1)$

(iii) Define $\boldsymbol{Y}^{(-1)}_{nm,T} = (\mathrm{vec}(Y_{nm,0})', \ldots, \mathrm{vec}(Y_{nm,T-1})')'$, note that

$$
\begin{aligned}
&\boldsymbol{Y}^{(*,-1)'}_{nm,T-1}\boldsymbol{B}_{nm,T-1}\boldsymbol{U}^*_{nm,T-1}\\
=&\boldsymbol{Y}^{(-1)'}_{nm,T}(F'_{T,T-1}\otimes I_{nm})'(I_{T-1}\otimes B_{nm})(F'_{T,T-1}\otimes I_{nm})\boldsymbol{U}_{nm,T}\\
=&\boldsymbol{Y}^{(-1)'}_{nm,T}(F_{T,T-1}\otimes I_{nm})(F'_{T,T-1}\otimes B_{nm})\boldsymbol{U}_{nm,T}\\
=&\boldsymbol{Y}^{(-1)'}_{nm,T}(F_{T,T-1}F'_{T,T-1})\otimes B_{nm}\boldsymbol{U}_{nm,T}\\
=&\boldsymbol{Y}^{(-1)'}_{nm,T}(J_T\otimes B_{nm})\boldsymbol{U}_{nm,T}\\
=&\boldsymbol{Y}^{(-1)'}_{nm,T}\left[\left(I_T-\frac{1}{T}l_Tl'_T\right)\otimes B_{nm}\right]\boldsymbol{U}_{nm,T}\\
=&\boldsymbol{Y}^{(-1)'}_{nm,T}(I_T\otimes B_{nm})\boldsymbol{U}_{nm,T}-\frac{1}{T}\boldsymbol{Y}^{(-1)'}_{nm,T}(l_Tl'_T\otimes B_{nm})\boldsymbol{U}_{nm,T}
\end{aligned}
$$

then, we have

$$
\begin{aligned}
&E\frac{1}{n(T-1)}\boldsymbol{Y}^{(*,-1)\prime}_{nm,T-1}\boldsymbol{B}_{nm,T-1}\boldsymbol{U}^*_{nm,T-1}\\
=&E\frac{1}{n(T-1)}\sum_{t=1}^{T}\left(\mathrm{vec}(\boldsymbol{Y}_{nm,t-1})-\frac{1}{T}\sum_{s=0}^{T-1}\mathrm{vec}(\boldsymbol{Y}_{nm,s})\right)'B_{nm}\left(\mathrm{vec}(\boldsymbol{U}_{nm,t})-\frac{1}{T}\sum_{s=1}^{T}\mathrm{vec}(\boldsymbol{U}_{nm,s})\right)\\
=&E\frac{1}{nT(T-1)}[\mathrm{vec}(\boldsymbol{Y}_{nm,0}),\ldots,\mathrm{vec}(\boldsymbol{Y}_{nm,T-1})]'B_{nm}[\mathrm{vec}(\boldsymbol{U}_{nm,1}),\ldots,\mathrm{vec}(\boldsymbol{U}_{nm,T})]
\end{aligned}
$$

From (5), we have

$$\begin{aligned}\operatorname{vec}\left(\boldsymbol{Y}_{nm,t}\right) &= \sum_{h=0}^{\infty} \boldsymbol{H}_{nm}^{h} \boldsymbol{S}_{nm}^{-1} \left[(\boldsymbol{\Pi}_{\mathrm{m}0}' \otimes I_n)\operatorname{vec}\left(\boldsymbol{X}_{nk,t-h}^{*}\right) + \operatorname{vec}(\boldsymbol{C}_{nm0}) + \alpha_{m0,t-h}^{*} \otimes l_n + \operatorname{vec}\left(\boldsymbol{V}_{nm,t-h}\right)\right] \\ &= \cdots + \sum_{h=0}^{\infty} \boldsymbol{H}_{nm}^{h} \boldsymbol{S}_{nm}^{-1} \operatorname{vec}\left(\boldsymbol{V}_{nm,t-h}\right) \\ &= \cdots + \boldsymbol{S}_{nm}^{-1} \operatorname{vec}\left(\boldsymbol{V}_{nm,t}\right) + \boldsymbol{H}_{nm}^{1} \boldsymbol{S}_{nm}^{-1} \operatorname{vec}\left(\boldsymbol{V}_{nm,t-1}\right) + \boldsymbol{H}_{nm}^{2} \boldsymbol{S}_{nm}^{-1} \operatorname{vec}\left(\boldsymbol{V}_{nm,t-2}\right) + \ldots\end{aligned}$$

and

$$\begin{aligned}\operatorname{vec}\left(\boldsymbol{Y}_{nm,1}\right) &= \cdots + \boldsymbol{S}_{nm}^{-1}(\boldsymbol{\Gamma}_{m0}'^{-1} \otimes I_n)\operatorname{vec}\left(\boldsymbol{U}_{nm,1}\right) + \ldots \\ \operatorname{vec}\left(\boldsymbol{Y}_{nm,2}\right) &= \cdots + \boldsymbol{S}_{nm}^{-1}(\boldsymbol{\Gamma}_{m0}'^{-1} \otimes I_n)\operatorname{vec}\left(\boldsymbol{U}_{nm,2}\right) + \boldsymbol{H}_{nm}^{1}\boldsymbol{S}_{nm}^{-1}(\boldsymbol{\Gamma}_{m0}'^{-1} \otimes I_n)\operatorname{vec}\left(\boldsymbol{U}_{nm,1}\right) + \ldots \\ \operatorname{vec}\left(\boldsymbol{Y}_{nm,3}\right) &= \cdots + \boldsymbol{S}_{nm}^{-1}(\boldsymbol{\Gamma}_{m0}'^{-1} \otimes I_n)\operatorname{vec}\left(\boldsymbol{U}_{nm,3}\right) + \boldsymbol{H}_{nm}^{1}\boldsymbol{S}_{nm}^{-1}(\boldsymbol{\Gamma}_{m0}'^{-1} \otimes I_n)\operatorname{vec}\left(\boldsymbol{U}_{nm,2}\right) + \boldsymbol{H}_{nm}^{2}\boldsymbol{S}_{nm}^{-1}(\boldsymbol{\Gamma}_{m0}'^{-1} \otimes I_n)\operatorname{vec}\left(\boldsymbol{U}_{nm,1}\right) + \ldots\end{aligned}$$

thus

$$
\begin{aligned}
&E\frac{1}{nT(T-1)}[\mathrm{vec}(\boldsymbol{Y}_{nm,0}),\dots,\mathrm{vec}(\boldsymbol{Y}_{nm,T-1})]'B_{nm}[\mathrm{vec}(\boldsymbol{U}_{nm,1}),\dots,\mathrm{vec}(\boldsymbol{U}_{nm,T})]\\
=&\frac{1}{nT(T-1)}tr\begin{pmatrix}(\boldsymbol{\Gamma}_{m0}^{-1}\otimes I_n)\boldsymbol{S}_{nm}^{-1\prime}B_{nm}\Sigma_{um0}\\
+(\boldsymbol{\Gamma}_{m0}^{-1}\otimes I_n)\boldsymbol{S}_{nm}^{-1\prime}B_{nm}\Sigma_{um0}+(\boldsymbol{\Gamma}_{m0}^{-1}\otimes I_n)\boldsymbol{S}_{nm}^{-1\prime}B_{nm}\boldsymbol{H}_{nm}^{1\prime}\Sigma_{um0}\\
+\dots\\
+(\boldsymbol{\Gamma}_{m0}^{-1}\otimes I_n)\boldsymbol{S}_{nm}^{-1\prime}B_{nm}\Sigma_{um0}+(\boldsymbol{\Gamma}_{m0}^{-1}\otimes I_n)\boldsymbol{S}_{nm}^{-1\prime}B_{nm}\boldsymbol{H}_{nm}^{1\prime}\Sigma_{um0}+\cdots+(\boldsymbol{\Gamma}_{m0}^{-1}\otimes I_n)\boldsymbol{S}_{nm}^{-1\prime}B_{nm}\boldsymbol{H}_{nm}^{T-2\prime}\Sigma_{um0}\end{pmatrix}\\
=&\frac{1}{nT(T-1)}tr\left[\Sigma_{um0}(\boldsymbol{\Gamma}_{m0}^{-1}\otimes I_n)\boldsymbol{S}_{nm}^{-1\prime}\left((T-1)+(T-2)\boldsymbol{H}_{nm}^{1\prime}+\cdots+\boldsymbol{H}_{nm}^{T-2\prime}\right)B_{nm}\right]\\
=&\frac{1}{n(T-1)}tr\left[\Sigma_{um0}(\boldsymbol{\Gamma}_{m0}^{-1}\otimes I_n)\boldsymbol{S}_{nm}^{-1\prime}\left(\frac{(T-1)}{T}+\frac{(T-2)}{T}\boldsymbol{H}_{nm}^{1\prime}+\cdots+\frac{1}{T}\boldsymbol{H}_{nm}^{T-2\prime}\right)B_{nm}\right]\\
=&\frac{1}{n(T-1)}tr\left[\Sigma_{um0}(\boldsymbol{\Gamma}_{m0}^{-1}\otimes I_n)\boldsymbol{S}_{nm}^{-1\prime}\sum_{h=1}^{T-1}\frac{T-h}{T}\boldsymbol{H}_{nm}^{h-1\prime}B_{nm}\right]\\
=&\frac{1}{n(T-1)}tr\left[\Sigma_{um0}(\boldsymbol{\Gamma}_{m0}^{-1}\otimes I_n)\boldsymbol{S}_{nm}^{-1\prime}\sum_{h=1}^{T-1}\left(1-\frac{h}{T}\right)\boldsymbol{H}_{nm}^{h-1\prime}B_{nm}\right]\\
=&\frac{1}{n(T-1)}tr\left[\sum_{h=1}^{T-1}\left(1-\frac{h}{T}\right)\boldsymbol{H}_{nm}^{h-1}\boldsymbol{S}_{nm}^{-1}(\boldsymbol{\Gamma}_{m0}^{-1\prime}\otimes I_n)\Sigma_{um0}B_{nm}'\right]
\end{aligned}
$$

■

9.2 Proofs for theorems

References

Alvarez, J. and Arellano, M. (2003). The time series and cross-section asymptotics of dynamic panel data estimators. *Econometrica*, 71(4):1121–1159.

Amba, M. C. (2021). Simultaneous equations with three way error components. *Journal of Quantitative Economics*, 19(3):583–596.

Anselin, L. (1988). *Spatial econometrics: methods and models*, volume 4. Springer Science & Business Media.

Arellano, M. and Bond, S. (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The review of economic studies*, 58(2):277–297.

Arellano, M. and Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of econometrics*, 68(1):29–51.

Baltagi, B. H., Song, S. H., Jung, B. C., and Koh, W. (2007). Testing for serial correlation, spatial autocorrelation and random effects using panel data. *Journal of econometrics*, 140(1):5–51.

Cliff, A. D. and Ord, J. K. (1973). Spatial autocorrelation. *(No Title)*.

Cohen-Cole, E., Liu, X., and Zenou, Y. (2018). Multivariate choices and identification of social interactions. *Journal of Applied Econometrics*, 33(2):165–178.

Drukker, D., Egger, P., and Prucha, I. (2023). Simultaneous equations models with higher-order spatial or social network interactions. *Econometric Theory*, 39:1154–1201.

Elhorst, J. P. (2003). Specification and estimation of spatial panel data models. *International regional science review*, 26(3):244–268.

Elhost, J. and Emili, S. (2022). A spatial econometric multivariate model of okun's law. *Regional Science and Urban Economics*, 93:103756.

Gu, J. (2024). Digitalization, spillover and environmental, social, and governance performance: evidence from china. *The Journal of Environment & Development*, 33(2):286–311.

Hahn, J. and Kuersteiner, G. (2002). Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and t are large. *Econometrica*, 70(4):1639–1657.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the econometric society*, pages 1029–1054.

Kapoor, M., Kelejian, H. H., and Prucha, I. R. (2007). Panel data models with spatially correlated error components. *Journal of econometrics*, 140(1):97–130.

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*, 17:99–121.

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(1-2):27–50.

Kuersteiner, G. M. and Prucha, I. R. (2020). Dynamic spatial panel models: Networks, common shocks, and sequential exogeneity. *Econometrica*, 88(5):2109–2146.

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

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

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

Lee, L.-f. and Yu, J. (2010b). A spatial dynamic panel data model with both time and individual fixed effects. *Econometric Theory*, 26(2):564–597.

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

LeSage, J. and Pace, R. K. (2009). *Introduction to spatial econometrics*. Chapman and Hall/CRC.

Li, K. (2017). Fixed-effects dynamic spatial panel data models and impulse response analysis. *Journal of Econometrics*, 198(1):102–121.

Lia, L. and Yanga, Z. (2023). Dynamic spatial panel data models with interactive fixed effects: M-estimation and inference under fixed or relatively small t. *Available at SSRN 4815863*.

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 & Economic Statistics*, 32(4):516–536.

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

Liu, X. and Saraiva, P. (2019). Gmm estimation of spatial autoregressive models in a system of simultaneous equations with heteroskedasticity. *Econometric Reviews*, 38(4):359–385.

Lu, L. (2023). Simultaneous spatial panel data models with common shocks. *Journal of Business & Economic Statistics*, 41(2):608–623.

Lütkepohl, H., Milunovich, G., and Yang, M. (2020). Inference in partially identified heteroskedastic simultaneous equations models. *Journal of Econometrics*, 218(2):317–345.

Neyman, J. and Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica: journal of the Econometric Society*, pages 1–32.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica: Journal of the econometric society*, pages 1417–1426.

Schmidt, P. (2020). *Econometrics*. CRC Press.

Tiba, S. and Belaid, F. (2021). Modeling the nexus between sustainable development and renewable energy: the african perspectives. *Journal of Economic Surveys*, 35(1):307–329.

Wang, W., Lee, L.-F., and Bao, Y. (2018). Gmm estimation of the spatial autoregressive model in a system of interrelated networks. *Regional Science and Urban Economics*, 69:167–198.

Yang, K. and Lee, L.-f. (2017). Identification and qml estimation of multivariate and simultaneous equations spatial autoregressive models. *Journal of Econometrics*, 196(1):196–214.

Yang, K. and Lee, L.-f. (2019). Identification and estimation of spatial dynamic panel simultaneous equations models. *Regional Science and Urban Economics*, 76:32–46.

Yang, K. and Lee, L.-f. (2021). Estimation of dynamic panel spatial vector autoregression: Stability and spatial multivariate cointegration. *Journal of Econometrics*, 221(2):337–367.

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

Zhang, J., Zhang, P., Wang, R., Liu, Y., and Lu, S. (2023). Identifying the coupling coordination relationship between urbanization and forest ecological security and its impact mechanism: Case study of the yangtze river economic belt, china. *Journal of Environmental Management*, 342:118327.