Spatial Econometrics: Models, Methods and Applications

Dissertation
Present in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By
Ji Tao, M.A.

The Ohio State University
2005

Dissertation Committee:
Lung-fei Lee, Adviser
Don Haurin
John Kagel

Approved by

Adviser
Economics Graduate Program
ABSTRACT

Three essays comprise this dissertation. The first chapter considers the Two-Stage Least-Square (2SLS) and Generalized Method of Moments (GMM) estimators for spatial and temporal autoregressive panel data models under the fixed-effects framework. The proposed 2SLS and GMM estimators are consistent, asymptotically normal and robust against unknown initial conditions. The classical 2SLS estimators are linear with closed-form expressions. The GMM estimators make full use of moment conditions under the standard assumptions and are efficient relative to the 2SLS estimators.

The second chapter considers the decision of local school spending in a dynamic game-theoretical model. I formulate a dynamic game model with strategic interaction and show that the local spending follows an exact spatial and temporal autoregressive path in general equilibrium. Empirical work uses local school expenditure data published by the Ohio State Department of Education. Without time effects, the current spending among local school districts are strategically correlated, but including time dummy effects makes such strategic interaction statistically insignificant. The latter contradict the empirical results in recent public finance literature that has a positive reaction slope. Furthermore, I have found that the current spending of the school districts is negatively related with the temporally lagged spending of their neighbors.
The third chapter considers specification test, identification and estimation of simultaneous system of spatially interrelated cross-sectional equations. The identification conditions of classical system are extended to the spatial system. Classical 2SLS and 3SLS estimators are derived for the spatial models. A simple Hausman specification test of exogeneity is provided.
Dedicated to my parents
ACKNOWLEDGMENTS

I thank my adviser, Lung-fei Lee for intellectual support, encouragement and patience, which made this thesis possible. I also thank him for devoting lots of time to correcting many mistakes in the paper. In the past several years, I have learned so much from him and really help me to grow both personally and professionally.

I thank Don Haurin and John Kagel for their patience, stimulating discussion and help.

I also thank Jihai Yu and Yong Yu, who helped me at various stages of the paper.

My thanks also go to my parents and my brother, who have given me great support for my whole life.
VITA

September 18, 1974 .......................................Born - Wuxi, P.R.China

2000................................................................M.A. The Ohio State University

1995................................................................B.A. Nankai University, P.R.C

FIELD OF STUDY

Econometrics

Public Finance

Empirical Industrial Organization
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>ii</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>v</td>
</tr>
<tr>
<td>VITA</td>
<td>vii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>x</td>
</tr>
<tr>
<td>CHAPTER 1</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.2 The Simple Model</td>
<td>4</td>
</tr>
<tr>
<td>1.2.1 Moment Conditions under Standard Assumptions</td>
<td>6</td>
</tr>
<tr>
<td>1.2.2 Identification Condition</td>
<td>10</td>
</tr>
<tr>
<td>1.2.3 Classical 2SLS Estimator</td>
<td>12</td>
</tr>
<tr>
<td>1.2.4 Efficient GMM Estimator</td>
<td>14</td>
</tr>
<tr>
<td>1.2.5 Asymptotic Relative Efficiency</td>
<td>15</td>
</tr>
<tr>
<td>1.3 Extension</td>
<td>17</td>
</tr>
<tr>
<td>1.3.1 Model with exogenous variables</td>
<td>17</td>
</tr>
<tr>
<td>1.3.2 Model with SAR disturbance</td>
<td>19</td>
</tr>
<tr>
<td>1.3.3 Model with higher-order SAR lags</td>
<td>21</td>
</tr>
<tr>
<td>1.3.4 Generalized Case</td>
<td>22</td>
</tr>
<tr>
<td>1.3.5 Initial consistent estimator</td>
<td>22</td>
</tr>
<tr>
<td>1.4 Monte Carlo Study</td>
<td>23</td>
</tr>
<tr>
<td>1.4.1 Monte Carlo Design</td>
<td>23</td>
</tr>
<tr>
<td>1.4.2 Monte Carlo Results</td>
<td>25</td>
</tr>
<tr>
<td>1.5 Conclusion</td>
<td>26</td>
</tr>
<tr>
<td>1.6 Appendix of Chapter 1</td>
<td>27</td>
</tr>
<tr>
<td>1.6.1 A: Moment Functions</td>
<td>28</td>
</tr>
<tr>
<td>1.6.2 B: Expectation</td>
<td>29</td>
</tr>
<tr>
<td>1.6.3 C: Score Vector</td>
<td>31</td>
</tr>
<tr>
<td>1.6.4 D: Variance of Moments</td>
<td>35</td>
</tr>
</tbody>
</table>
3.7 Extension ............................................................................................................. 106
  3.7.1 System with insufficient IVs ..................................................................... 106
  3.7.2 System with higher order spatial lags ....................................................... 109
  3.7.3 System with SAR disturbance ................................................................. 110

3.8 Conclusion .......................................................................................................... 113

3.9 Appendix of Chapter 3 ........................................................................................ 113
  3.9.1 A: Some useful Lemmas ........................................................................... 113
  3.9.2 B: Proofs on selected Theorems ............................................................... 115

APPENDIX OF TABLES ............................................................................................... 121

LIST OF REFERENCES ................................................................................................ 137
LIST OF TABLES

Table 1. Monte Carlo Experiments ................................................................. 121
Table 2. Monte Carlo Results ........................................................................ 122
Table 3. Descriptive Statistics of Ohio school expenditure data .................... 124
Table 4. Estimates of Nested Test Model ....................................................... 126
Table 5. G2SLS Estimates with geographic weights matrix ......................... 127
Table 6. G2SLS Estimates with secondary income weights matrix ............... 128
Table 7. G2SLS Estimates with secondary black weights matrix ................... 130
Table 8. Testing Nonlinear Restriction on Structure Parameters .................... 132
Table 9. G2SLS Estimates with geographic weights matrix and time effects ...... 133
Table 10. G2SLS Estimates with secondary income weights matrix and time effects ................................................................. 135
Table 11. G2SLS Estimates with secondary black weights matrix and time effects .... 136
CHAPTER 1

Estimation of Spatial Temporal Autoregressive Panel Data Models

1.1 Introduction

The spatial dynamic panel data model falls under a broad classification of spatiotemporal autoregressive (STAR) models. See Cressie (1993) has provided more details about STAR models in his book.

This paper considers the econometric estimation for the spatial dynamic panel. It is assumed that we have observations on a large number of cross-sectional dimensions, and a short time dimension. The explanatory variables are a simultaneous autoregressive (SAR) term, and weakly or strongly exogenous variables. The model disturbance contains a time-invariant individual effects as well as random noise. We study the asymptotic properties of estimators as \( n \) tends to infinity for fixed \( T \).

Both dynamic and spatial autocorrelations can exist in the same economic context. We have already been familiar with dynamic nature in economic relationships, for example, dynamic demand for addictive consumption goods like cigarettes. There exists an extensive literature on theoretic and empirical study of dynamic panel data models, see Baltagi (2001) for more applications in his book.

Topics on spatial autocorrelation are becoming more attractive in recent
theoretic and empirical research. Cross-sectional interdependence is a relevant issue in various economic fields including urban, real estate, regional, public, environmental, agricultural, finance and IO studies, see Anselin and Bera (1998) and Brock and Durlauf (2000) for recent survey. Social sciences other than economics have studies the role of interaction for a long time. The 1966 Coleman Report (Coleman, et al, 1966) argued that peer influences played an important role on academic performance of the disadvantaged students. If the sample data are randomly drawn from the population, we usually do not worry about the cross-sectional correlation. Two prominent examples of such panel data sets are the National Longitudinal Surveys of Labor Market Experience (NLS) and the University of Michigan’s Panel Study of Income Dynamics (PSID)\textsuperscript{1}. However, the social interaction generates spatial correlation between nearby locations. Hence, the aggregate data of census tracts, community areas, counties, states, regions, and countries are likely correlated across spatial units\textsuperscript{2}. There is a growing interest in empirical study of spatial panel data models during the past decade. See Case (1991, 1992), Driscoll and Kraay (1998) and Baltagi and Li (2001) for a few applications.

Until now, there are few theoretic and/or empirical literature that considers the co-existence of dynamic and spatial effects in panel data models.

\textsuperscript{1}See Hsiao (1986) for a concise description of NLS and PSID in the first page of the book.

\textsuperscript{2}Census tracts are fairly small units and a sub-division of larger units called Community Areas. For example, there are 75 such areas in Chicago, and each area contains 12 tracts on average. See Topa (2001) for more details.
In practice, cross-sectional units can be correlated spatially and temporally in the same context. For example, smoking cigarettes and/or marijuana is an addictive behavior and interrelated among students due to peer influences. This paper will consider the fixed-effects estimation of dynamic spatial panel data model. The fixed-effects model differs from the random effects model not in the assumption of the nature of the individual effects, but on whether to make inferences conditional or unconditional on the individual effects. Mundlak (1978) criticized the random-effects formulation about static panel model, because it assume there is no correlation between the individual effects and the explanatory variables. However, there may exist such correlation in many circumstances. Ignoring such correlation can lead to inconsistent random-effects estimator, while fixed-effects estimator remains consistent under the same scenario.

The plan of the paper is as follows. In Section 2, the simple spatial dynamic model without strongly exogenous variables is considered. All available moment conditions, linear and quadratic, are summarized under a standard set of assumptions. Identification conditions are discussed. Asymptotic properties of the efficient GMM estimator are derived. Analytical comparisons among various instrument-variable estimators is made. In Section 3, the GMM estimator is generalized to various cases. In Section 4, Monte Carlo study is conducted to investigate finite-sample properties. Conclusions are drawn in Section 5 and appendices are given in Section 6.

3See Hsiao’s book (1986) on simple panel data model in Chapter 3 for more explanations.
1.2 The simple model (Model I)

We consider the empirically relevant case as $n \to \infty$ for fixed $T$. At the $t$th period, the model equation in vector form is written as:

$$y_{n,t} = \lambda W_n y_{n,t} + \beta y_{n,t-1} + u_{n,t},$$

(1)

where $y_{n,t}$ is the $n$-dimensional vector of dependent variables, $W_n$ is an $n \times n$ constant matrix of spatial weights with a zero diagonal, $W_n y_{n,t}$ is the simultaneous autoregressive term, and $y_{n,t-1}$ is the lagged dependent variable. The parameter of interest contains simultaneous coefficient $\lambda$ and dynamic coefficient $\beta$ respectively. Denote the vector of structural parameters $\theta = (\lambda, \beta)'$. The disturbance $u_{n,t}$ has two error components:

$$u_{n,t} = \alpha_n + \varepsilon_{n,t},$$

(2)

where $\varepsilon_{n,t}$ are random noise with zero mean and finite variance, and $\alpha_{n,i}$ are time-invariant cross-sectional effects.

In scalar notation, we have

$$u_{n,it} = \alpha_{n,i} + \varepsilon_{n,it},$$

(3)

Ahn and Schmidt (1995) has summarized the following standard assumptions (SA) about $(\alpha_{n,i}, \varepsilon_{n,it})$:

**SA.1:** $\varepsilon_{n,t}$ are uncorrelated with $y_{n,0}$ for all $t$;

---

4We also assume that $\varepsilon$’s higher-order moment $E(|\varepsilon|^{4+2\delta})$ exists for some $\delta > 0$ for the Central Limit Theorem. See Lee LF (2002a) for more details.
**SA.2:** $\varepsilon_{n,t}$ are uncorrelated with $\alpha_n$ for all $t$;

**SA.3:** $\varepsilon_{n,it}$ are mutually uncorrelated for all $i$ and $t$.

The assumption about initial conditions $y_{n,0}$ plays an important role in the literature of dynamic panel models. We consider the fixed-effect GMM estimator under weak assumptions on $y_{n,0}$.

Some regularity conditions on spatial matrices (Lee, 2002a) will be assumed.

**Assumption 1:** The elements $w_{n,ij}$ of the weight matrix $W_n$ are of order $O\left(\frac{1}{h_n}\right)$ uniformly in all $i, j$, i.e., there exists a constant $c$ such that $|w_{n,ij}| \leq \frac{c}{n}$ for all $i, j,$ and $n$.

**Assumption 2:** The sequence of simultaneous autocorrelation $\{h_n\}$ is bounded.

In regional economics, this means each areas have finite neighborhoods. This assumption is necessary for GMM estimators to converge at the $\sqrt{n}$ rate if there is no dynamic effect. In other words, if the true parameter $\beta_0 = 0$, there is no valid exogenous instruments for classical 2SLS procedure. The identification of $\lambda$ will rely on the nonlinear moment conditions constructed

---

5The initial process $y_{n,0}$ has either been going on for a long time, or started from an unknown finite past. The former case requires vector stationary assumption, while the latter not (See Hsiao, et al., 2002). Our GMM estimator shall be robust to different data generating processes of $y_{n,0}$.

6If the dynamic effect does exist or other exogenous variables are relevant, the sequence $\{h_n\}$ can go to infinity but at a slower rate than $n$, i.e the ratio $\frac{h_n}{n} \to 0$ as $n$ goes to infinity. The GMM estimators converge at $\sqrt{n}$-rate, if the exogenous variables are relevant to the model.
Denote $S_n(\lambda) = I_n - \lambda W_n$ and $S_n = S_n(\lambda_0)$. Note that $S_n(\lambda) = S_n + (\lambda_0 - \lambda) W_n$.

**Assumption 3:** At the true parameter $\lambda_0$, $S_n$ is nonsingular.

Under the above assumption, we have the simultaneous equilibria at the true parameter vector $\theta_0 = (\lambda_0, \beta_0)'$.

$$y_{n,t} = S_n^{-1}y_{n,t-1}\beta_0 + S_n^{-1}u_{n,t}.$$

(4)

**Assumption 4:** The spatial matrices $\{W_n\}$ and $\{S_n^{-1}\}$ are uniformly bounded in both row and column sums.

The uniform boundedness of matrix is important in the asymptotic properties of estimators of SAR models. It limits the simultaneous autocorrelation among $y_n$ to a controllable degree. This assumption was initially suggested by Kelejian and Prucha (1998).

**1.2.1 Moment Conditions under SA**

Denote that $G_n = W_n S_n^{-1}$. Substituting the simultaneous equilibria into the model, we have that

$$y_{n,t} = \lambda (G_n y_{n,t-1} \beta_0 + G_n u_{n,t}) + \beta y_{n,t-1} + u_{n,t}.$$  

(5)

We eliminate the unobserved individual effects $\alpha_n$ by first-order difference and obtain that

$$\Delta y_{n,t} = \lambda (G_n \Delta y_{n,t-1} \beta_0 + G_n \Delta u_{n,t}) + \beta \Delta y_{n,t-1} + \Delta u_{n,t} = \lambda (G_n \Delta y_{n,t-1} \beta_0 + G_n \Delta \varepsilon_{n,t}) + \beta \Delta y_{n,t-1} + \Delta \varepsilon_{n,t}.$$  

(6)
Note that the first-order difference sign ∆ and constant matrix are interchangeable.

In their generalized 2SLS method for a mixed regressive, spatial autoregressive model (MRSAR), Kelejian and Prucha (1998) suggest using $W_nX_n, W_n^2X_n$ etc., together with $X_n$ itself as IVs. In the first-order differenced equation of our model, both $W_n\Delta y_{n,t}$ and $\beta\Delta y_{n,t-1}$ are correlated with $\Delta u_{n,t}$, the valid moments are $W_n y_{n,s}, W_n^2 y_{n,s}$ etc., together with $y_{n,s}$ ($s \leq t - 2$) as IVs for estimating $\theta$. Lee (2003) shows that a best nonlinear 2SLS IV can be constructed in MRSAR by choosing $E\left(\frac{\partial f(\theta_0)}{\partial \theta'}|X\right)$ (Amemiya 1985). The N2SLS IV corresponds to $G_n y_{n,s} \beta_0$ for $W_n \Delta y_{n,t}$ ($s \leq t - 2$) in our model. However, the nonlinear $G_n y_{n,s} \beta_0$ are no longer the best instrument variable in the panel data model.

Let $Q_{n,s}$ denote an $n \times k$ matrix of 2SLS IVs constructed from $W_n$ and $y_{n,s}$. For example, $Q_{n,s} = (W_n y_{n,s}, W_n^2 y_{n,s}, y_{n,s})$.

Under SA, the following 2SLS conditions hold:

$$E[Q_{n,s}' \Delta u_{n,t}] = 0, \quad s = 0, ..., T - 2; \quad t \geq s + 2. \quad (7)$$

There are $T \times (T - 1) \times k/2$ linear moment (LM) equations.

Under SA, Ahn and Schmidt (1995) find that there exist additional nonlinear moment conditions between the final period and preceding periods:

$$E[u_{n,T}' \Delta u_{n,t}] = 0, \quad t = 2, ..., T - 1. \quad (8)$$

There are $(T - 2)$ quadratic temporal moment (TM) equations.
Lee (2001c) suggests using $P_n u_n$ as possible instruments in the GMM framework for SAR models to estimate spatial coefficient $\lambda$.

**Assumption 5:** The constant square matrices $\{P_n\}$ are uniformly bounded in both row and column sums.

The constant matrices $\{P_n\}$ has either zero diagonal or $tr(P_n) = 0$ in general. Let $P_{1n}$ be a class of matrices with zero trace and $P_{2n}$ be the class of matrices with zero diagonals.\(^7\)

When $P_{2n}$ has a zero diagonal and $\varepsilon_{n,it}$ are mutually uncorrelated, we have

$$E[(P_{2n} u_{n,t})' \Delta u_{n,t}] = tr(P_{2n} E[\Delta u_{n,t} u_{n,t}']) = 0. \quad (9)$$

$$E[(P_{2n} u_{n,t-1})' \Delta u_{n,t}] = tr(P_{2n} E[\Delta u_{n,t} u_{n,t-1}']) = 0. \quad (10)$$

The $P_{2n} u_{n,t}$ and $P_{2n} u_{n,t-1}$ can be valid instruments for $G_n \Delta u_{n,t}$ if they are correlated with $G_n \Delta u_{n,t}$ but uncorrelated with $\Delta u_{n,t}$.

Under SA, we can not use $P_{1n} u_{n,t}$ and $P_{1n} u_{n,t-1}$ as possible instruments if $\varepsilon_{n,it}$’s are heteroskedastic cross-sectionally.

If we assume that $\varepsilon_{n,it}$ are homoskedastic for all $i$, $P_{1n}$ including $P_{2n}$ can be used to construct valid instruments.

**AA.4:** $var(\varepsilon_{n,t}) = \sigma^2 \varepsilon I_n$ for all $t$.

$$E[(P_{1n} u_{n,t})' \Delta u_{n,t}] = \sigma^2 \varepsilon tr(P_{1n}) = 0. \quad (11)$$

$$E[(P_{1n} u_{n,t-1})' \Delta u_{n,t}] = -\sigma^2 \varepsilon tr(P_{1n}) = 0. \quad (12)$$

By selecting matrices $P_{n1}, \cdots, P_{nm}$ from either $P_{1n}$ or $P_{2n}$, we have additional nonlinear moment conditions:

\(^7\)Note that $P_{2n}$ is a subclass of $P_{1n}$ because $tr(P_{2n}) = 0$.\(8\)
\[ E[(P_n u_{n,t})' \Delta u_{n,t}] = 0. \] (13)

\[ E[(P_n u_{n,t-1})' \Delta u_{n,t}] = 0. \] (14)

There are \(2(T-1) \times m\) quadratic spatial moment (SM) equations\(^8\).

In sum, under SA, we have \(2(T-1) \times m + (T-2) + T \times (T-1) \times k/2\) moment equations.

Note that the above moment equation implies that

\[ E[(P_n \Delta u_{n,t})' \Delta u_{n,t}] = 0. \] (15)

There are \((T-1) \times M\) moment equation. However, these two set of moment conditions are not equivalent, because \((P_n u_{n,t})' \Delta u_{n,t}\) is a proper subset of \((P_n u_{n,t})' u_{n,t}\).

**Assumption 6.** The regressors \(y_{n,t-1}\) is uniformly bounded and the limit of \(\lim_{n \to \infty} \frac{1}{n} \sum y_{n,t-1}^2\) exists and is nonsingular.

Assumption 6 will guarantee that the lagged dependent variables is vector-stationary. For simplicity, suppose that the spatial autoregressive parameter is zero, then this assumption is equivalent to say that the dynamic parameter is smaller than unit root and vector stationary process. Allowing the presence of spatial autocorrelation, this assumption excludes the case that \(\lambda + \beta \geq 1\) in which the vector process is non-stationary.

\(^8\)Note that AA 4 assumes only cross-sectional homoskedasiticity, i.e., \(\text{var}(\varepsilon_{n,t}) = \sigma^2_{e,t} I_n\) for fixed \(t\). But I suppress \(t\) in the subscripts for notational convenience. Hence, AA 4 does not generate the additional \(T-1\) moments otherwise.
1.2.2 Identification Condition

In the GMM framework, the identification condition requires that the moment equations $E[g_n(\theta)] = 0$ has a unique solution at the true parameter $\theta_0$ as $n$ tends to infinity (Hansen 1982).

$$E[g_n(\theta)] = 0 \implies \theta = \theta_0.$$

The latent variables $u_{n,t}(\theta)$ and $\Delta u_{n,t}(\theta)$ are functions of unknown parameter vector $\theta = (\lambda, \beta)$.

$$u_{n,t}(\theta) = S_n(\lambda)y_{n,t} - \beta y_{n,t-1},$$
$$\Delta u_{n,t}(\theta) = S_n(\lambda)\Delta y_{n,t} - \beta \Delta y_{n,t-1}. \quad (16)$$

Denote $d_{n,t-1}(\theta) = (\lambda_0 - \lambda) G_n y_{n,t-1} \beta_0 + (\beta_0 - \beta) y_{n,t-1}$. Then we can rewrite the latent variables as

$$u_{n,t}(\theta) = d_{n,t-1}(\theta) + S_n(\lambda)S_n^{-1} u_{n,t},$$
$$\Delta u_{n,t}(\theta) = \Delta d_{n,t-1}(\theta) + S_n(\lambda)S_n^{-1} \Delta u_{n,t}. \quad (17)$$

One of the linear moment equations corresponding to IV matrix $Q_{n,s}$ is

$$E[Q'_{n,s}\Delta u_{n,t}(\theta)]$$
$$= E[Q'_{n,s}\Delta d_{n,t-1}(\theta)]$$
$$= E \left[ \begin{bmatrix} Q'_{n,s}G_n\Delta y_{n,t-1}\beta_0 & Q'_{n,s}\Delta y_{n,t-1} \end{bmatrix} \begin{bmatrix} \lambda_0 - \lambda \\ (\beta_0 - \beta) \end{bmatrix} \right]$$
$$= E \left[ Q'_{n,s} \begin{bmatrix} G_n\Delta y_{n,t-1}\beta_0 & \Delta y_{n,t-1} \end{bmatrix} \begin{bmatrix} \lambda_0 - \lambda \\ (\beta_0 - \beta) \end{bmatrix} \right]. \quad (18)$$
If \((Q'_n, s \Delta y_{n,t-1}, Q'_n s G_n \Delta y_{n,t-1}/\beta_0)\) has a full column rank, the GMM moment vector will have a unique solution at \(\theta_0\). In other words, the sufficient rank condition for model identification is

\[
\text{rank} \left[ Q'_n s G_n \Delta y_{n,t-1}/\beta_0 \right] = 2.
\]

The sufficient rank condition requires the necessary rank condition that

\[
\text{rank}(Q_n) \geq 2, \quad \text{rank} \left[ G_n \Delta y_{n,t-1} / \beta_0 \Delta y_{n,t-1} \right] = 2. \tag{19}
\]

Under the sufficient rank condition, \(\theta_0\) can be identified at \(Q'_n, s \Delta d_{n,t-1}(\theta) = 0\). Thus, the structural parameter vector \(\theta\) can be estimated by the standard 2SLS instrument matrix \(Q_n\). The additional nonlinear moment equations can help improve the efficiency of the GMM estimators.

The necessary rank condition of \((\Delta y_{n,t}, G_n \Delta y_{n,t}/\beta_0)\) will not hold if \(G_n \Delta y_{n,t}/\beta_0\) is linearly dependent on \(\Delta y_{n,t}\). In particular, if \(\beta_0 = 0\), the model becomes a static spatial panel model. The classical 2SLS instruments are invalid since \(G_n \Delta y_{n,t}/\beta_0 = 0\). Intuitively, the Ahn-Schmidt-type moment conditions are also in question, because they originate from dynamic feature of the model.

The identification of \(\lambda_0\) relies only on the remaining nonlinear moments from the spatial structure. These moments are still valid at \(\beta_0 = 0\), and can identify the parameter \(\lambda\), if the model satisfies the regularity assumptions. \footnote{When \(\rho_0 = 0\), our model is a pure SAR process over short periods. Assumption 3a with other regularity will guarantee the \(\sqrt{n}\) consistency of \(\lambda\). (2001a).}

11
1.2.3 Classical 2SLS Estimator

The 2SLS approach is a special case of the GMM estimation framework. In particular, the 2SLS estimator is a linear estimator with closed-form expressions. It provides an initial consistent estimator in solving the nonlinear GMM function numerically.

The 2SLS moment conditions can be rewritten as

\[ E[Q' \Delta u_n] = E[Q' \Delta \varepsilon_n] = 0. \]

where \( \Delta u_n = (\Delta u_{n,2}', \cdots, \Delta u_{n,T}')' \) and \( \Delta \varepsilon_n = (\Delta \varepsilon_{n,2}', \cdots, \Delta \varepsilon_{n,T}')' \) and

\[
Q' = \begin{bmatrix}
[Q_{n,0}]' & 0 \\
[Q_{n,0}, Q_{n,1}]' & \cdots \\
0 & [Q_{n,0}, Q_{n,1}, \cdots, Q_{n,T-2}]'
\end{bmatrix}
\]

is the 2SLS IV matrix of \([T(T-1)k/2] \times [n(T-1)]\) dimensions.

The variance matrix of first-order differenced disturbance is

\[ E[\Delta u_n \Delta u_n'] = E[\Delta \varepsilon_n \Delta \varepsilon_n'] = \sigma_\varepsilon^2 (G_{T-1} \otimes I_n). \]
where $G_{T-1}$ is a square matrix of dimension $(T - 1)$ with the form

$$
\begin{bmatrix}
2 & -1 & \cdots & 0 \\
-1 & 2 & \cdots & \\
\vdots & \vdots & \ddots & \vdots \\
\vdots & \vdots & \ddots & 2 \\
0 & \cdots & -1 & 2
\end{bmatrix}
$$

Premultiplying the first-order differenced equation by $Q'$, we have that

$$
Q'\Delta y_n = \lambda Q'(I_{T-1} \otimes W_n) \Delta y_n + \beta Q' \Delta y_{n-1} + Q' \Delta u_n,
$$

$$
Q' \Delta y_n = Q' z_n \theta + Q' \Delta u_n,
$$

where $\Delta y_n = (\Delta y'_{n,2}, \ldots, \Delta y'_{n,T})'$, $\Delta y_{n-1} = (\Delta y'_{n,1}, \ldots, \Delta y'_{n,T-1})'$ and $z_n = [(I_{T-1} \otimes W_n) \Delta y_n, \Delta y_{n-1}]$.

The generalized 2SLS (G2SLS) estimator for $\theta = (\lambda, \beta)'$ has the following expression:

$$
\hat{\theta}_{2s} = \left[ z_n' Q (Q' (G_{T-1} \otimes I_n) Q)^{-1} Q' z_n \right]^{-1} \times \left[ z_n' Q (Q' (G_{T-1} \otimes I_n) Q)^{-1} Q' \Delta y_n \right]
= \theta_0 + \left[ z_n Q (Q' (G_{T-1} \otimes I_n) Q)^{-1} Q' z_n \right]^{-1} \times \left[ z_n Q (Q' (G_{T-1} \otimes I_n) Q)^{-1} Q' \Delta u_n \right].
$$

The asymptotic distribution of $\hat{\theta}_{2s}$ is derived as

$$
\sqrt{n} \left( \hat{\theta}_{2s} - \theta_0 \right) \xrightarrow{D} \mathcal{N} \left( 0, \sigma_v^2 \lim_{n \to \infty} \left[ z_n Q (Q' (G_{T-1} \otimes I_n) Q)^{-1} Q' z_n \right]^{-1} \right).
$$
1.2.4 Efficient GMM Estimator

The 2SLS estimators $\hat{\theta}_2$ is consistent and asymptotically normal, but not efficient. As noted in Subsection 2.1, the 2SLS approach does not make full use of all available moments under $\text{SA}$. There exist many nonlinear moment conditions.

The GMM estimator $\hat{\theta}_g$ can be derived from the optimization problem

$$
\min_{\theta \in \Theta} g_n'(\theta) c_n' c_n g_n(\theta)
$$

with the limiting distribution

$$
\sqrt{n} \left( \hat{\theta}_g - \theta_0 \right) \xrightarrow{D} N(0, \Sigma_n),
$$

The asymptotic variance has the following formula

$$
\Sigma_n = \lim_{n \to \infty} \left[ \frac{1}{n} D_n' c_n' c_n \left( \frac{1}{n} D_n \right)^{-1} \left[ \frac{1}{n} D_n' c_n' c_n \left( \frac{1}{n} \Omega_n \right) c_n' c_n \left( \frac{1}{n} D_n \right)^{-1} \right] \left[ \frac{1}{n} D_n' c_n' c_n \left( \frac{1}{n} D_n \right)^{-1} \right]^{-1}
$$

where the score vector $D_n'$ is given from

$$
\frac{1}{n} \frac{\partial g_n(\theta)}{\partial \theta} = -\frac{1}{n} D_n + o_P(1).
$$

The generalized Schwartz inequality implies that the optimal weighting matrix for $c_n' c_n$ is the inverse of moment variance $\Omega_n$.

$$
\min_{\theta \in \Theta} g_n'(\theta) \Omega_n^{-1} g_n(\theta)
$$

The asymptotic distribution of optimal GMM estimator $\hat{\theta}_{og}$ is
\[ \sqrt{n} \left( \hat{\theta}_{fog} - \theta_0 \right) \xrightarrow{D} N(0, \Sigma_{n,o}), \]  

(26)

where

\[ \Sigma_{n,o} = \lim_{n \to \infty} \left[ \frac{1}{n} \hat{D}'_n \hat{\Omega}_n^{-1} \hat{D}_n \right]^{-1}. \]  

(27)

The feasible optimal GMM estimation shall be formulated with a two-step approach:

1. Obtain an initial consistent estimate \( \hat{\theta}_1 \) of \( \theta_0 \) by IV method such generalized 2SLS approach.

2. Then \( \frac{1}{n} \hat{\Omega}_n \) is a consistent estimate of \( \frac{1}{n} \Omega_n \) and can be calculated using the formula in appendix section.

The feasible optimal GMM estimator \( \hat{\theta}_{fog} \) derived from

\[ \min_{\theta \in \Theta} g'_n(\theta) \hat{\Omega}_n^{-1} g_n(\theta) \]  

(28)

has the asymptotic distribution

\[ \sqrt{n} \left( \hat{\theta}_{fog} - \theta_0 \right) \xrightarrow{D} N(0, \hat{\Sigma}_{n,o}). \]  

(29)

\[ \hat{\Sigma}_{n,o} = \left( \lim_{n \to \infty} \frac{1}{n} \hat{D}'_n \hat{\Omega}_n^{-1} \hat{D}_n \right)^{-1}. \]  

(30)

Furthermore, the overidentification test is formulated as

\[ g'_n(\hat{\theta}) \hat{\Omega}_n^{-1} g_n(\hat{\theta}) \xrightarrow{D} \chi^2 \left\{ \left[ 2 (T-1) \times m + (T-2) + T \times (T-1) \times k/2 \right] - 2 \right\}. \]  

(31)

1.2.5 Asymptotic Relative Efficiency
The generalized 2SLS estimator can be derived from the optimization problem:

$$\min_{\theta \in \Theta} \Delta u_n'(\theta) Q (Q' (G_{T-1} \otimes I_n) Q)^{-1} Q' \Delta u_n(\theta).$$  \hspace{1cm} (32)

Therefore, the 2SLS approach is a special case of the GMM estimation with the weighing matrix

$$c_n = \left(0, (Q' (G_{T-1} \otimes I_n) Q)^{-1/2}\right).$$  \hspace{1cm} (33)

It follows from the generalized Schwartz inequality, the optimal GMM estimator \(\hat{\theta}_{og}\) and its feasible counterpart shall be efficient relative to the G2SLS estimator \(\hat{\theta}_{2s}\).

For simplification, we assume that the spatial process is vector stationary in obtaining the score vector \(D_n\) and variance matrix \(\Omega_n\). The results are given in Appendix Section. It should be noted that our standard assumptions do not contain the vector stationary conditions. In fact, vector stationarity is not required to derive the consistency and asymptotic distribution of GMM estimator.

The asymptotic distribution of \(\hat{\theta}_{2s}\) has been derived as

$$\Sigma_{n,2s} = \sigma^2 \epsilon \lim_{n \to \infty} \left\{ \frac{1}{n} z_n Q (Q' (G_{T-1} \otimes I_n) Q)^{-1} Q' z_n \right\}^{-1}.$$  \hspace{1cm} (34)

The asymptotic distribution of optimal GMM estimator \(\hat{\theta}_{og}\) is

$$\Sigma_{n,og} = \left( \lim_{n \to \infty} \frac{1}{n} D_n' \Omega_n^{-1} D_n \right)^{-1}.$$  \hspace{1cm} (35)
where $D_n$ is the score vector of GMM moments, and the variance $\Omega_n = $

$$
\begin{pmatrix}
\text{var} (g_{n1}) & 0 & * \\
0 & \text{var} (g_{n2}) & * \\
* & * & \sigma^2 \varepsilon Q' (G_{T-1} \otimes I_n) Q
\end{pmatrix},
$$

(36)

where $*$ means that the corresponding blocks are NOT zero matrices. The complete expression of covariance matrix can be found in Appendix section.

**Result 1:** SM and TM are orthogonal to each other, and quadratic moments (SM and TM) are correlated with linear 2SLS moments.

$$
E(g_{n1} g_{n2}') = 0,
$$

$$
E(g_{n1} g_{n3}') \neq 0,
$$

$$
E(g_{n2} g_{n3}') \neq 0.
$$

**Result 2:** When the random noise $\varepsilon_n$ is symmetric ($\mu_3 = 0$), the GMM estimators are more efficient than the 2SLS estimator $\hat{\theta}_{2s}$.

$$
D_n' Q_n^{-1} D_n > z_n' Q (\sigma^2 \varepsilon Q' (G_{T-1} \otimes I_n) Q)^{-1} Q z_n.
$$

(37)

### 1.3 Extension

The GMM can be extended to the estimation of various models that contain high-order SAR lags, SAR disturbance, and other explanatory variables than spatial and temporal lags without additional computational complexity.

#### 1.3.1 Model II: model with exogenous variables
The model with exogenous variables (Model II) has the generic form:

\[ y_{n,t} = \lambda_W y_{n,t} + X_{n,t} \beta + u_{n,t} \]

\[ = \lambda_W y_{n,t} + X_{n,t}^w \beta_w + X_{n,t}^s \beta_s + u_{n,t} \]

where \( X_{n,t} \) contains a set of \( k_x \) exogenous variables. The regressors \( X_{n,t} \) can be partitioned into weakly exogenous variables \( X_{n,t}^w (n \times k_w) \), and strongly exogenous variables \( X_{n,t}^s (n \times k_s) \). For example, \( X_{n,t}^w = \begin{bmatrix} y_{n,t} - 1 \ W_{n} y_{n,t} - 1 \end{bmatrix} \).

At true parameter vector \( \theta_0 = (\lambda_0, \beta_0)' \), we obtain the general equilibria and substitute them into the model.

\[ y_{n,t} = S_n^{-1} X_{n,t} \beta_0 + S_n^{-1} u_{n,t} = S_n^{-1} X_{n,t}^w \beta_{w0} + S_n^{-1} X_{n,t}^s \beta_{s0} + S_n^{-1} u_{n,t} \]

\[ y_{n,t} = \lambda_W \left( S_n^{-1} X_{n,t}^w \beta_{w0} + S_n^{-1} X_{n,t}^s \beta_{s0} + S_n^{-1} u_{n,t} \right) + X_{n,t}^w \beta_w + X_{n,t}^s \beta_s + u_{n,t} \]

Taking the first-order difference, we have that

\[ \Delta y_{n,t} = \lambda \left( G_n \Delta X_{n,t}^w \beta_{w0} + G_n \Delta X_{n,t}^s \beta_{s0} + G_n \Delta u_{n,t} \right) + \Delta X_{n,t}^w \beta_w + \Delta X_{n,t}^s \beta_s + \Delta u_{n,t} \]

The 2SLS instrument matrix \( Q' \) in Subsection 2.3 is modified as:

\[
\begin{bmatrix}
[Q_{1n,1}; Q_{2n}]' \\
[0] \\
[Q_{1n,1}; Q_{1n,2}; Q_{2n}]' \\
\vdots \\
[0] \\
[Q_{1n,1}, Q_{1n,2}, \cdots, Q_{1n,T-1}; Q_{2n}]'
\end{bmatrix}
\]

\( ^{10} \)Note that \( x_t \) is strongly exogenous, if \( x_t \) is weakly exogenous and \( y_{t-1} \) does not Granger cause \( x_t \). (Greene, 1997 P. 714)

\( ^{11} \)Note that Time-invariant regressors in \( X_{n,t} \) do not exist in the differenced equation.
where \(Q_{2n} = [Q_{2n,1}; \cdots Q_{2n,t}; \cdots; Q_{2n,T}]\). We can consider the following 2SLS Instruments

\[
\begin{align*}
Q_{n,t} &= [W_n X_{nt}, W_n^2 X_{nt}, X_{nt}] , \\
Q_{n,t}^* &= [G_n X_{nt}, X_{nt}], \quad t = 1, 2, \cdots, T.
\end{align*}
\]

The matrix \(Q\) is of dimension \(\{T (T - 1) \left[\left(I_v + k_w\right)/2 + (I_v + k_s)\right]\} \times [n (T - 1)]\), where \(I_v\) is the number of instruments used for spatially lagged dependent variables. In particular, \(I_v = 2\) if \(Q_{n,t}\) is used; \(I_v = 1\), if \(Q_{n,t}^*\) is used. It is straightforward to formulate the GMM estimation vector for Model II with the replacement of \(g_{n3} (\theta)\) in subsection 3.1 by the linear moments \(Q' \Delta u_n (\theta)\).

### 1.3.2 Model III: model with SAR disturbance

The simple model with SAR disturbance (Model III) is

\[
\begin{align*}
y_{n,t} &= \lambda W_n y_{n,t} + \beta y_{n,t-1} + u_{n,t} \\
u_{n,t} &= \alpha_n + v_{n,t} \\
v_{n,t} &= \rho M_n v_{n,t} + \varepsilon_{n,t}
\end{align*}
\]

where the disturbance \(u_{n,t}\) contains individual effects \(\alpha_{n,i}\), correlated random shock \(v_{n,t}\), and idiosyncratic noise \(\varepsilon_{n,it}\) that is cross-sectionally uncorrelated. The parameter vector of interest is \(\theta = (\lambda, \beta, \rho)'\). Define some common
notations as follows.

\[ R_n(\rho) = I_n - \rho M_n \]

\[ F_n(\lambda, \rho) = R_n(\rho) S_n(\lambda) \]

\[ H_n(\rho) = M_n R_n^{-1}(\rho) \]

\[ u_{n,t}(\theta) = S_n(\lambda) y_{n,t} - \beta y_{n,t-1} \]

\[ \varepsilon_{n,t}(\theta) = R_n(\rho) v_{n,t}(\theta) \]

Note that \( u_{n,t}(\theta) \) is observable, but \( \varepsilon_{n,t}(\theta) \) is unobservable because of unobserved individual effects \( \alpha_n \). However, we can eliminate the time-invariant effects \( \alpha_n \) by first-order difference. Then, both \( \Delta u_{n,t}(\theta) \) and \( \Delta \varepsilon_{n,t}(\theta) \) are observable function of the data and unknown parameters.

\[ \Delta u_{n,t}(\theta) = S_n(\lambda) \Delta y_{n,t} - \beta \Delta y_{n,t-1} \]

\[ \Delta \varepsilon_{n,t}(\theta) = R_n(\rho) \Delta u_{n,t}(\theta) \]

\[ \Delta u_{n,t}(\theta) = R_n^{-1}(\rho) \Delta \varepsilon_{n,t}(\theta) \]

Let \( R_n = I_n - \rho_0 W_n \), which is invertible from Assumption 3 in Section 2. At the true parameter vector \( \theta_0 \), we have

\[ v_{n,t} = R_n^{-1} \varepsilon_{n,t} \]

\[ \Delta u_{n,t} = R_n^{-1} \Delta \varepsilon_{n,t} \]

\[ F_n = R_n S_n \]

\[ H_n = M_n R_n^{-1} \]
We have the reduced-form equation of the model
\[ y_{n,t} = S^{-1}_n y_{n,t-1} \beta_0 + S^{-1}_n u_{n,t} \]
\[ \Delta y_{n,t} = S^{-1}_n \Delta y_{n,t-1} \beta_0 + S^{-1}_n \Delta u_{n,t} = S^{-1}_n \Delta y_{n,t-1} \beta_0 + F^{-1}_n \Delta \varepsilon_{n,t} \]
\[ \Delta y_{n,t} = \lambda W_n \left( S^{-1}_n \Delta y_{n,t-1} \beta_0 + S^{-1}_n \Delta u_{n,t} \right) + \beta \Delta y_{n,t-1} + \Delta u_{n,t} \]
\[ = \lambda W_n \left( S^{-1}_n \Delta y_{n,t-1} \beta_0 + F^{-1}_n \Delta \varepsilon_{n,t} \right) + \beta \Delta y_{n,t-1} + R^{-1}_n (\rho) \Delta \varepsilon_{n,t} (\theta) \]

With the selected \( P_n \) matrices and 2SLS IV matrix \( Q \), we can form empirical GMM moments for each category:
\[
E \left[ \varepsilon'_{n,t-i} (\theta) P_n \Delta \varepsilon_{n,t} (\theta) \right] = E \left[ u'_{n,t-i} (\theta) R_n' (\rho) P_n R_n (\rho) \Delta u_{n,t} (\theta) \right] = 0 \quad i = 0, 1 \\
E \left[ u'_{n,T} (\theta) \Delta u_{n,s} (\theta) \right] = E \left[ u'_{n,T} (\theta) R_n' (\rho) R_n (\rho) \Delta u_{n,s} (\theta) \right] = 0 \quad s = 2, \ldots, T-1 \\
E \left[ Q'_{n,s} \Delta \varepsilon_{n,t} (\theta) \right] = E \left[ Q'_{n,s} R_n (\rho) \Delta u_{n,t} (\theta) \right] = 0 \quad t = 2, \ldots, T; \quad s \leq t - 2
\]

### 1.3.3 Model IV: model with higher-order SAR lags

The model with higher-order SAR lags (Model IV) is
\[ y_{n,t} = \sum_{j=1}^{p} \lambda_j W_{jn} y_{n,t} + \beta y_{n,t-1} + u_{n,t} \]
where \( W_{jn} \) are \( p \) distinct spatial weights matrices. For Model IV, define some common notations as follows
\[ \lambda = (\lambda_1, \ldots, \lambda_p)' \]
\[ \theta = (\lambda, \beta)' \]
\[ S_n (\lambda) = I_n - \sum_{j=1}^{p} \lambda_j W_{jn} \]
\[ S_n = S_n (\lambda_0) \]
\[ G_{jn} = W_{jn} S^{-1}_n \]
The empirical moments have the same form as \( g_n(\theta) \) in Subsection 2.1.

### 1.3.4 Model V: the generalized case

Putting all elements together, we have the generalized form

\[
y_{n,t} = \sum_{j=1}^{p} \lambda_j W_{jn} y_{n,t} + X_{n,t} \beta + u_{n,t} \quad t = 1, 2, \ldots, T
\]

\[
u_{n,t} = \alpha_n + v_{n,t} \quad v_{n,t} = \rho M_n v_{n,t} + \varepsilon_{n,t}
\]

The parameter vector of interest is \( \theta = (\lambda', \beta', \rho)' \). It is straightforward to formulate the GMM estimation vector (See the details in Appendix B).

### 1.3.5 Recipe: initial consistent estimator (GS2SLS)

The generalized spatial 2SLS estimation method has the following recipe.

**Step 1:** Estimate \((\lambda', \beta')\) by the IV method for the differenced model equation

\[
\Delta y_{n,t} = \sum_{j=1}^{p} \lambda_j W_{jn} \Delta y_{n,t} + \Delta X_{n,t} \beta + \Delta u_{n,t}.
\]

**Step 2:** Substitute the estimated \((\hat{\lambda}', \hat{\beta}')\) into the above equation and obtain the differenced disturbance.

\[
\Delta u_{n,t} = \Delta y_{n,t} - \sum_{j=1}^{p} \hat{\lambda}_j W_{jn} \Delta y_{n,t} - \Delta X_{n,t} \hat{\beta}.
\]

**Step 3:** Estimate \( \rho \) by MLE or GMM method for the spatial error process.

\[
\Delta u_{n,t} = \Delta v_{n,t} = \rho M_n \Delta v_{n,t} + \Delta \varepsilon_{n,t}.
\]

**Step 4:** Estimate \( \sigma^2_{\varepsilon} \) by the following formula.
\[
\hat{\sigma}_\varepsilon^2 = \frac{\Delta \hat{\varepsilon}'_n \Delta \hat{\varepsilon}_n}{2n(T-1)} = \frac{\Delta \hat{u}'_{n,t} (I_n - \hat{\rho}M'_n) (I_n - \hat{\rho}M_n) \Delta \hat{u}_{n,t}}{2n(T-1)}.
\]

Step 5: Obtain asymptotic variance of the IV estimator \(\hat{\theta}_{2s} = (\hat{\lambda}', \hat{\beta}')\).

Note that \(E[\Delta u_n \Delta u'_n] = \sigma^2 G_M\) with

\[
G_M = \left[I_{T-1} \otimes (I_n - \hat{\rho}M_n)^{-1}\right] \left[G_{T-1} \otimes I_n\right] \left[I_{T-1} \otimes (I_n - \hat{\rho}M'_n)^{-1}\right],
\]

\[
\text{var}(\hat{\theta}_{2s}) = \sigma^2 \left[z_n Q (Q' G_M Q)^{-1} Q' z_n\right]^{-1}.
\]

1.4 Monte Carlo Study

The computer is IBM-compatible equipped with Pentium 4 CPU 2.80 GHz and 512 MB of RAM that runs Matlab 6 Version software.

1.4.1 Monte Carlo Design

In this subsection, we describe the design of the simulation experiment that we shall conduct to assess the finite sample performance of the various estimators discussed in the previous sections.

The model of interest take the forms as follows:

\[
y_{n,t} = \lambda W_n y_{n,t} + \beta_1 y_{n,t-1} + \beta_2 W_n y_{n,t-1} + \beta_3 x_{n,t} + u_{n,t},
\]

(38)

where \(u_{n,t} = \alpha_n + \varepsilon_{n,t}\) defined in Section 1.2. In particular, \(W_n y_{n,t}\) is known as spatial autoregressive (SAR) lag term, \(y_{n,t-1}\) is temporal autoregressive (TAR) lag term and \(W_n y_{n,t-1}\) is spatio-temporal autoregressive (STAR) lag

\[\text{The problem of multicollinearity may be serious in equations (2) and (4).}\]
term. The above model includes both spatial panel and dynamic panel as its special cases.

In generating \( y_{n,t} \), we set \( y_{n,-s} = 0 \) and discard the first \( s \) observations \( (s = 50 \text{ in default}) \), and use the observations \( t = 0 \) through \( T \) for estimation.

The exogenous regressor \( X_{n,it} \) is generated according to

\[
x_{n,it} = \mu_i + gt + \xi_{it},
\]

where \( \mu_i \) and \( \xi_{it} \) has a standard normal distribution. The time trend coefficient \( g \) is set to 0, 0.5 or 1.

In spatial model, the weights matrices are given constant. Two kinds of spatial weights matrices shall be considered in experiments. One is from the empirical work: \( W_{49} \) used by Anselin in Columbus crime data; \( W_{612} \) used in Ohio school district data of the following Chapter 2. The other is artificially generated from regular lattice model: \( W_n \) is a block-diagonal matrix where each units has exactly four neighborhoods in diagonal blocks and zero elsewhere. It turns out that empirical or artificially spatial weights matrices has little effects in the finite-sample properties if they have the same dimensions.

Table 1 summarizes the different Monte Carlo designs that shall be conducted in our research. Each of the parameters studies has 3 different choices, which gives a maximum of \( 3^4 = 81 \) combinations. However, we will not run the experiments in which the chosen parameters will generate divergent process. For example, when both \( \lambda \) and \( \beta_1 \) are 0.7, the generated vector process \( y_{nt} \) is divergent to infinity. The process \( y_{nt} \) is vector stationary if \( \lambda = 0.7 \).
and $\beta_1 = 0.2$ and $\beta_2 = 0$. However, the process is divergent if $\lambda = 0.7$ and $\beta_1 = 0.2$ and $\beta_2 = 0.15$. As mentioned above, empirical or artificial spatial weights matrices generate little difference in experiments. It appears that $W_{49}$ used by Anselin is an ideal spatial weights matrix especially in small sample studies. To get medium to large sample size, we use the Kronecker product of $W_{49}$ and identity matrix to match the spatial weights with the desired sample size. For the panel data model, we consider the cases where the sample size $n$ is 49, 245 or 490; and the periods of time $T$ is 5 or 10.

### 1.4.2 Monte Carlo Results

This subsection reports the finite sample properties of estimators. We set the number of replication to 300 in simulation.

Both 2SLS and GMM (that sets optimal weighting matrix to identity in this stage) estimators are considered. The digit following the estimators denotes the number of $P_n$ instruments matrix used in practice. For example, 2SLS1 and GMM1 uses $W_n$ as instrument matrix; and 2SLS2 and GMM2 uses $W_n$ and $W_n^2$ (or $W_n^2 - diag(W_n^2)$) as instrument matrices.

In Table 2, we report the panel data models with the sample periods $T = 5$. The variance of individual effects and random disturbances are both 0.5. The parameter values given as: $\lambda = 0.7$, $\beta_1 = 0.2$, $\beta_2 = 0$ and $\beta_3 = 1$. The sample size ranges from 49 to 245 to 490.

First, as the sample size increases, all four estimators have better performance as the bias, deviation and root mean squared errors decrease almost monotonically.
Second, with regard to bias, GMM estimators performs much better than their corresponding 2SLS estimators in reducing the bias of spatial autoregressive parameters. When sample size is 49, the 2SLS estimators of SAR parameters has an upward bias about 0.2, four times of those for GMM estimators. When sample size is 245, the bias of 2SLS estimators decrease to about 0.12, while the bias of GMM estimator is about 0.03. When the sample size is 490, the bias of 2SLS estimator is about 0.08, still twice larger than those of GMM estimators. For the other parameter coefficients, 2SLS and GMM estimators have no large differences in their biases.

Third, with regard to the deviation, 2SLS estimators perform better than GMM estimators for both spatial autoregressive parameters and the other parameters in general.

Fourth, with regard to root mean squared error, both 2SLS and GMM estimators have shown similar performance. This is not surprising given that GMM estimators performs better in bias reduction and 2SLS estimators have smaller variance. Recall that mean squared error is the sum of deviation and squared bias. As a result, the root mean squared errors of 2SLS and GMM estimators have not much differences in general.

Finally, the selection of instruments in both 2SLS and GMM method yields similar performances in small sample properties of parameters.

Due to time and space constraint, we only report a few Monte Carlo designs. However, the findings above also holds for other Monte Carlo designs.

\section*{1.5 Conclusion}
In this paper, we have proposed GMM estimators for the estimation of dynamic spatial panel under a set of standard assumptions and regularity conditions. We have discussed about the construction of GMM moments and identification issues. Our GMM estimator is efficient among various instrument estimators, and can test the significance level of spatial effect where the lagged dependent variable is irrelevant. The GMM procedure can be extended to generalized models without additional computational complexity. The proposed 2SLS and GMM estimators are consistent, asymptotically normal and robust against unknown initial conditions. The classical 2SLS estimators are linear with closed-form expressions, which make it computationally simple. The GMM estimators make full use of moment conditions under the standard assumptions and are efficient relative to the 2SLS estimators.

1.6 Appendix of Chapter 1

The latent variables \( u_{n,t}(\theta) \) and \( \Delta u_{n,t}(\theta) \) are functions of unknown parameter vector \( \theta \).

\[
\begin{align*}
    u_{n,t}(\theta) &= S_n(\lambda) y_{n,t} - \beta y_{n,t-1} \\
    \Delta u_{n,t}(\theta) &= S_n(\lambda) \Delta y_{n,t} - \beta \Delta y_{n,t-1}
\end{align*}
\]

For notational convenience, we also assume that \( \varepsilon_{n,it} \) is homoskedastic for all \( i \) and \( t \). With the selected constant matrices \( P_n = (P_{n1}, \cdots, P_{nm}) \) and 2SLS instrument matrix \( Q_{n,s} \), the set of moment conditions forms a primary GMM estimation vector: \( g_n(\theta) = \left[ g'_{n1}(\theta) \quad g'_{n2}(\theta) \quad g'_{n3}(\theta) \right]' \), that
has $2 (T - 1) \times m + (T - 2) \times T (T - 1) \times k/2$ columns. The first category $g_{n1} (\theta)$ are the quadratic spatial moments; the middle category $g_{n2} (\theta)$ are the quadratic temporal moments; and the last category $g_{n3} (\theta)$ are linear 2SLS moments.

In this section, we will present the formula of moment functions, its expectation, score vector and moment variance that shall be used in empirical work.

A. Moment functions
The empirical GMM estimation vector for Model V is $g_n(\theta) =$

$$
\begin{pmatrix}
  g_{n,1}(\theta) \\
g_{n,2}(\theta) \\
g_{n,3}(\theta)
\end{pmatrix} = 
\begin{pmatrix}
  u'_{n,2}(\theta) R'_{n}(\rho) P_{n1} R_{n}(\rho) \Delta u_{n,2}(\theta) \\
  \vdots \\
  u'_{n,T}(\theta) R'_{n}(\rho) P_{n1} R_{n}(\rho) \Delta u_{n,T}(\theta) \\
  u'_{n,1}(\theta) R'_{n}(\rho) P_{n1} R_{n}(\rho) \Delta u_{n,2}(\theta) \\
  \vdots \\
  u'_{n,T-1}(\theta) R'_{n}(\rho) P_{n1} R_{n}(\rho) \Delta u_{n,T}(\theta) \\
  \vdots \\
  u'_{n,T-1}(\theta) R'_{n}(\rho) P_{nm} R_{n}(\rho) \Delta u_{n,T}(\theta) \\
  u'_{n,T}(\theta) R'_{n}(\rho) R_{n}(\rho) \Delta u_{n,2}(\theta) \\
  \vdots \\
  u'_{n,T}(\theta) R'_{n}(\rho) R_{n}(\rho) \Delta u_{n,T-1}(\theta) \\
  Q' (I_{T-1} \otimes R_{n}(\rho)) \Delta u_{n}(\theta)
\end{pmatrix}
$$

**B. Expectation**

For any $\theta$ in the parameter space, $E[g_n(\theta)] =$

29
\[
E \begin{bmatrix}
\Delta d'_{n,1}(\theta) P_{n1} \Delta d_{n,1}(\theta) \\
\vdots \\
\Delta d'_{n,T-1}(\theta) P_{n1} \Delta d_{n,T-1}(\theta) \\
\vdots \\
\Delta d'_{n,T-1}(\theta) P_{nm} \Delta d_{n,T-1}(\theta) \\
\vdots \\
d'_{n,T-1}(\theta) \Delta d_{n,1}(\theta) \\
\vdots \\
d'_{n,T-1}(\theta) \Delta d_{n,T-2}(\theta) \\
Q'_{n,0} \Delta d_{n,1}(\theta) \\
Q'_{n,0} \Delta d_{n,2}(\theta) \\
Q'_{n,1} \Delta d_{n,2}(\theta) \\
\vdots \\
Q'_{n,0} \Delta d_{n,T-1}(\theta) \\
Q'_{n,T-2} \Delta d_{n,T-1}(\theta)
\end{bmatrix}
+ \begin{bmatrix}
a_1 \sigma^2 - c_1 \sigma^2 \\
\vdots \\
a_m \sigma^2 - c_m \sigma^2 \\
\vdots \\
d(2) \sigma^2 \\
\vdots \\
d(T-1) \sigma^2 \\
0 \\
0 \\
0 \\
\vdots \\
0 \\
0
\end{bmatrix}
\]

where

\[
a_j = \text{tr}(S_{n-1} S_n^t(\lambda) P_{n,j} S_{n}(\lambda) S_{n-1}^{-1}),
\]

\[
c_j = \text{tr}(S_{n-1} H_n^t(\theta) P_{n,j} S_{n}(\lambda) S_{n-1}^{-1}),
\]

\[
d(t) = \rho_0^{-1-t} \text{tr}(S_n^{t-(T-t)} H_n^t(\theta) S_{n}(\lambda) S_{n-1}^{-1}) - \rho_0^{-t} \text{tr}(S_n^{t-(T-t+1)} H_n^t(\theta) S_{n}(\lambda) S_{n-1}^{-1}).
\]

At true parameter \(\theta_0 = (\lambda_0, \rho_0)'\), we can easily show that \(E[g_n(\theta_0)] = 0\).
C. Score Vector

For convenience, denote \( s = (a, c, d(t))' \), where

\[
a = \text{tr} (S_n^{t-1} S_n (\lambda) P_n^* S_n (\lambda) S_n^{-1}).
\]

\[
c = \text{tr} (S_n^{t-1} H_n' (\theta) P_n^* S_n (\lambda) S_n^{-1}).
\]

\[
d(t) = \rho_0^{T-t} \text{tr} (S_n^{t-(T-t)} H_n' (\theta) S_n (\lambda) S_n^{-1}) - \rho_0^{T-t} \text{tr} (S_n^{t-(T-t+1)} H_n' (\theta) S_n (\lambda) S_n^{-1}).
\]

At \( \theta_0 \), \(-\partial s (\theta_0) / \partial \theta' = \)

\[
\begin{pmatrix}
2 \text{tr} (P_n^* G_n) & 0 \\
\rho_0 \text{tr} (P_n^* G_n S_n^{-1}) & \text{tr} (P_n^* S_n^{-1}) \\
\rho_0^{T-t} \text{tr} (G_n S_n^{(T-t)}) - \rho_0^{T-t+1} \text{tr} (G_n S_n^{(T-t+1)}) & \rho_0^{T-t} \text{tr} S_n^{(T-t)} - \rho_0^{T-t} \text{tr} S_n^{(T-t+1)}
\end{pmatrix}.
\]

Recall that the score vector \( D_n \) is the probability limit of the first-order derivative of the moment functions as follows.

\[
\frac{1}{n} \frac{\partial g_n (\hat{\theta})}{\partial \theta'} = \frac{1}{n} D_n + o_P (1).
\]
The score vector $D_n =$

\[
\begin{pmatrix}
2\sigma^2 \varepsilon \text{tr} \left( P_{n1}^n G_n \right) & -\sigma^2 \rho_0 \varepsilon \text{tr} \left( P_{n1}^n G_n S_n^{-1} \right) \\
\vdots & \vdots \\
2\sigma^2 \varepsilon \text{tr} \left( P_{nm}^n G_n \right) & -\sigma^2 \rho_0 \varepsilon \text{tr} \left( P_{nm}^n G_n S_n^{-1} \right) \\
\vdots & \vdots \\
2\sigma^2 \varepsilon \text{tr} \left( P_{n1}^n G_n \right) & -\sigma^2 \rho_0 \varepsilon \text{tr} \left( P_{n1}^n G_n S_n^{-1} \right) \\
\vdots & \vdots \\
2\sigma^2 \varepsilon \text{tr} \left( P_{nm}^n G_n \right) & -\sigma^2 \rho_0 \varepsilon \text{tr} \left( P_{nm}^n G_n S_n^{-1} \right) \\
\sigma^2 \rho_0^{T-2} \text{tr} \left( G_n S_n^{-T+1} \right) & \sigma^2 \rho_0^{T-3} \left( \text{tr} S_n^{-T+2} \right) \\
\sigma^2 \rho_0^{T-1} \text{tr} \left( G_n S_n^{-T} \right) & \sigma^2 \rho_0^{T-2} \text{tr} \left( S_n^{-T+1} \right) \\
\vdots & \vdots \\
\sigma^2 \rho_0 \text{tr} \left( G_n S_n^{-1} \right) & \sigma^2 \rho_0^2 \text{tr} \left( G_n S_n^{-2} \right) \\
Q'_{n,0} \Delta G_n y_{n,1} & Q'_{n,0} \Delta y_{n,1} \\
Q'_{n,0} \Delta G_n y_{n,2} & Q'_{n,0} \Delta y_{n,2} \\
Q'_{n,1} \Delta G_n y_{n,2} & Q'_{n,1} \Delta y_{n,2} \\
\vdots & \vdots \\
Q'_{n,0} \Delta G_n y_{n,T-1} & Q'_{n,0} \Delta y_{n,T-1} \\
\vdots & \vdots \\
Q'_{n,T-2} \Delta G_n y_{n,T-1} & Q'_{n,T-2} \Delta y_{n,T-1} \\
\end{pmatrix}
\]

Note: It is interesting to see the special case when either $\lambda_0 = 0$ (pure dynamic panel) or $\rho_0 = 0$ (pure spatial panel).
Case 1: $\lambda_0 = 0$ pure dynamic panel: with $D_n =$

$$
\begin{pmatrix}
\sigma^2 \varepsilon (2 - \rho_0) tr (P_{n1}^s W_n) & 0 \\
\vdots & \vdots \\
\sigma^2 \varepsilon (2 - \rho_0) tr (P_{nm}^s W_n) & 0 \\
\vdots & \vdots \\
\sigma^2 \varepsilon (2 - \rho_0) tr (P_{n1}^s W_n) & 0 \\
\vdots & \vdots \\
\sigma^2 \varepsilon (2 - \rho_0) tr (P_{nm}^s W_n) & 0 \\
0 & n\sigma^2 \rho_0^{-3} - n\sigma^2 \rho_0^{-2} \\
\vdots & \vdots \\
0 & n\sigma^2 \rho_0^{-1} - n\sigma^2 \rho_0 \\
Q'_{n,0} \Delta W_n y_{n,1} \rho_0 & Q'_{n,0} \Delta y_{n,1} \\
Q'_{n,0} \Delta W_n y_{n,2} \rho_0 & Q'_{n,0} \Delta y_{n,2} \\
Q'_{n,1} \Delta W_n y_{n,2} \rho_0 & Q'_{n,1} \Delta y_{n,2} \\
\vdots & \vdots \\
Q'_{n,0} \Delta W_n y_{n,T-1} \rho_0 & Q'_{n,0} \Delta y_{n,T-1} \\
\vdots & \vdots \\
Q'_{n,T-2} \Delta W_n y_{n,T-1} \rho_0 & Q'_{n,T-2} \Delta y_{n,T-1}
\end{pmatrix}
$$
Case 2: $\rho_0 = 0$ pure spatial panel with $D_n =$

$$
\begin{pmatrix}
2\sigma_\varepsilon^2 \text{tr} (P_{n1}^s G_n) & -\sigma_\varepsilon^2 \text{tr} (P_{n1}^s S_n^{-1}) \\
: & : \\
2\sigma_\varepsilon^2 \text{tr} (P_{nm}^s G_n) & -\sigma_\varepsilon^2 \text{tr} (P_{nm}^s S_n^{-1}) \\
: & : \\
2\sigma_\varepsilon^2 \text{tr} (P_{n1}^s G_n) & -\sigma_\varepsilon^2 \text{tr} (P_{n1}^s S_n^{-1}) \\
: & : \\
2\sigma_\varepsilon^2 \text{tr} (P_{nm}^s G_n) & -\sigma_\varepsilon^2 \text{tr} (P_{nm}^s S_n^{-1}) \\
0 & 0 \\
: & : \\
0 & \sigma_\varepsilon^2 \text{tr} (S_n^{-1}) \\
0 & Q'_{n,0} \Delta y_{n,1} \\
0 & Q'_{n,0} \Delta y_{n,2} \\
0 & Q'_{n,1} \Delta y_{n,2} \\
: & : \\
0 & Q'_{n,0} \Delta y_{n,T-1} \\
: & : \\
0 & Q'_{n,T-2} \Delta y_{n,T-1}
\end{pmatrix}
$$
Case 3: $\theta_0 = 0$ pure static panel with $D_n =$

\[
\begin{pmatrix}
2\sigma^2 \text{tr} \left( P_{n1}^s W_n \right) & 0 \\
\vdots & \vdots \\
2\sigma^2 \text{tr} \left( P_{nm}^s W_n \right) & 0 \\
\vdots & \vdots \\
2\sigma^2 \text{tr} \left( P_{n1}^s W_n \right) & 0 \\
\vdots & \vdots \\
2\sigma^2 \text{tr} \left( P_{nm}^s W_n \right) & 0 \\
0 & 0 \\
\vdots & \vdots \\
0 & n\sigma^2 \\
0 & Q'_{n,0} \Delta y_{n,1} \\
0 & Q'_{n,0} \Delta y_{n,2} \\
0 & Q'_{n,1} \Delta y_{n,2} \\
\vdots & \vdots \\
0 & Q'_{n,0} \Delta y_{n,T-1} \\
\vdots & \vdots \\
0 & Q'_{n,T-2} \Delta y_{n,T-1}
\end{pmatrix}
\]

D. Variance of Moments

Denote $\Omega_n (\theta) = \text{var}[g_n (\theta)]$ with $\Omega_n = \Omega_n (\theta_0) = \text{var}[g_n (\theta_0)]$

At true parameter $\theta_0 = (\lambda_0, \rho_0)'$, the covariance-variance matrix is
\[ \Omega_n = \begin{pmatrix} B_{11} & B_{12} & B_{13} \\ B_{22} & B_{23} \\ B_{33} \end{pmatrix} \]

with each partitioned matrix

\[
B_{11} = \text{var}\left[g_{n1}\right], \\
B_{22} = \text{var}\left[g_{n2}\right], \\
B_{33} = \text{var}\left[g_{n3}\right], \\
B_{12} = E\left[g_{n1}g_{n2}'\right], \\
B_{13} = E\left[g_{n1}g_{n3}'\right], \\
B_{23} = E\left[g_{n2}g_{n3}'\right].
\]

Assume that \( \lim_{n \to \infty} \frac{1}{n}Q_n^tQ_n \) converges to finite number we have

\[
\frac{1}{n} \hat{\Omega}_n = \frac{1}{n} \Omega_n \hat{\theta} = \frac{1}{n} \Omega_n + o_p(1).
\]

The variance matrix \( \frac{1}{n} \Omega_n \) can be consistently estimated by \( \frac{1}{n} \hat{\Omega}_n \) where \( \hat{\theta} \) is a consistent estimator of \( \theta_0 \), and \( \frac{1}{n} \hat{\Omega}_n \) is a consistent estimate of \( \Omega_n \) and can be approximated by the average number of moment product. Hence, we can approximate the variance \( \frac{1}{n} \Omega_n \) by the average number of moment products.

The upper diagonal block \( B_{11} \) is the variance of moments \( g_{n1}(\theta_0) \) of di-
mension \((T - 1) \times M\).

\[
B_{11} = \sigma^4_\varepsilon G_{1,T-1} \otimes \Delta_{nm} + \left(\mu_4 - 3\sigma^4_\varepsilon\right) G_{0,T-1} \otimes \varpi'_n \varpi_{nm}
\]

\[
= \sigma^4_\varepsilon \begin{pmatrix}
4 & 1 & 0 \\
1 & 4 & \\
0 & 1 & 4
\end{pmatrix} \otimes \Delta_{nm} + \left(\mu_4 - 3\sigma^4_\varepsilon\right) \begin{pmatrix}
2 & 1 & 0 \\
1 & 2 & \\
0 & 1 & 2
\end{pmatrix} \otimes \varpi'_n \varpi_{nm}
\]

When random noises \(\varepsilon_{n,t}\) are normal, we have \(B_{22} = \sigma^4_\varepsilon G_{1,T-1} \otimes \Delta_{nm}\) due to \(\mu_4 = 3\sigma^4_\varepsilon\).

The middle diagonal block \(B_{22}\) is the variance of moments \(g_{n2}(\theta_0)\) of dimension \((T - 2)\).

\[
B_{22} = n\sigma^2_\varepsilon \left(\sigma^2_\alpha + \sigma^2_\varepsilon\right) G_{2,T-2}
\]

\[
= E \left[
\begin{array}{c}
u_{n,T}\Delta u_{n,2}u'_{n,T}\Delta u_{n,2} \\
u_{n,T}\Delta u_{n,3}u'_{n,T}\Delta u_{n,2} \\
u_{n,T}\Delta u_{n,4}u'_{n,T}\Delta u_{n,2} \\
\vdots \\
u_{n,T}\Delta u_{n,T-1}u'_{n,T}\Delta u_{n,2}u'_{n,T}\Delta u_{n,T-1}
\end{array}
\right]
\]

\[
= \left[n\sigma^2_\varepsilon \left(\sigma^2_\alpha + \sigma^2_\varepsilon\right)\right] \times \begin{pmatrix}
2 & -1 & 0 \\
-1 & 2 & \\
\vdots & \ddots & -1 \\
0 & \cdots & -1 & 2
\end{pmatrix}
\]

The group of GMM moments \(g_{n3}(\theta)\) are classical 2SLS moments.

\[
E[Q'_{n,s}\Delta u_{n,t}] = 0, \quad s = 0,...,T - 2; \ t \geq s + 2.
\]
Denote $\Delta u_n = (\Delta u'_{n,2}, \cdots, \Delta u'_{n,T})'$ and $\Delta \varepsilon_n = (\Delta \varepsilon'_{n,2}, \cdots, \Delta \varepsilon'_{n,T})'$ where $\Delta \varepsilon_{n,t}$ is unit-root process.

$$E[\Delta u_n \Delta u'_n] = E[\Delta \varepsilon_n \Delta \varepsilon'_n] = \sigma^2 \varepsilon (G_{1,T-1} \otimes I_n).$$

where $G_{3,T-1}$ is a square matrix of dimension $(T - 1)$ with the form

$$
\begin{bmatrix}
2 & -1 & \cdots & 0 \\
-1 & 2 & \cdots \\
\vdots & \vdots & \ddots & \vdots \\
\cdots & 2 & -1 \\
0 & \cdots & -1 & 2
\end{bmatrix}
$$

and $G_{3,T-1} \otimes I_n$ is of dimension $N \times (T - 1)$

$$
\begin{bmatrix}
2I_n & -1I_n & \cdots & 0 \\
-1I_n & 2I_n & \cdots \\
\vdots & \vdots & \ddots & \vdots \\
\cdots & 2I_n & -1I_n \\
0 & \cdots & -1I_n & 2I_n
\end{bmatrix}
$$

Denote the 2SLS instrument matrix $Q$ as

$$
\begin{bmatrix}
[Q_{n,0}] & 0 \\
[Q_{n,0}, Q_{n,1}] & \ldots \\
0 & [Q_{n,0}, Q_{n,1}, Q_{n,T-2}]
\end{bmatrix}
$$
which is a \((N \times (T - 1)) \times (T \times (T - 1) \times K/2)\) matrix. Then, the GMM moment vector \(g_{n3}(\theta)\) can be rewritten as \(E[Q'\Delta u_n] = 0\).

The lower diagonal block \(B_{33}\) is the variance of moments \(g_{n3}(\theta_0)\) of dimension \((T \times (T - 1) \times K/2)\).

\[
B_{33} = \sigma^2 \varepsilon E[Q' (G_{3,T-1} \otimes I_n) Q] = \begin{bmatrix}
2Q_{n,0}'Q_{n,0} & -Q_{n,0}'Q_{n,1} & \ldots & -Q_{n,0}'Q_{n,T-2} \\
-Q_{n,0}'Q_{n,0} & 2Q_{n,0}'Q_{n,0} & \ldots & \ldots \\
-Q_{n,0}'Q_{n,1} & 2Q_{n,0}'Q_{n,1} & \ldots & \ldots \\
0 & \ldots & \ldots & \ldots \\
\end{bmatrix}
\]

The upper-middle block matrix \(B_{12}\) is the covariance of moments \(g_{n1}(\theta_0)\) and \(g_{n2}(\theta_0)\) of dimension \([(T - 1) \times M] \times (T - 2)\).

\[
B_{12} = E \begin{bmatrix}
0_{m\times1} & \ldots & 0_{m\times1} \\
\ldots & \ldots & \ldots \\
0_{m\times1} & \ldots & 0_{m\times1} \\
0 & 0 & 0_{m\times1}
\end{bmatrix} = 0_{[(T-1)\times M] \times (T-2)}.
\]

The upper-right block matrix \(B_{13}\) is the covariance of moments \(g_{n1}(\theta_0)\)
and $g_{n3}(\theta_0)$ of dimension $[(T - 1) \times M] \times [T \times (T - 1) \times K/2]$.

$$
B_{13} = \mu_3 \begin{pmatrix}
0 & -1 & 0 \\
-1 & 0 & \ddots & \vdots \\
& \ddots & \ddots & 0 \\
& & 0 & -1
\end{pmatrix} \otimes \varpi'_{nm} E [Q]
$$

$$
= \mu_3 (G_{4,T-1} \otimes \varpi'_{nm}) E [Q].
$$

In particular, when random noises $\varepsilon_{n,it}$ are symmetric ($\mu_3 = 0$), the block is reduced to zero matrix.

The central-right block matrix $B_{23}$ is the covariance of moments $g_{n2}(\theta_0)$ and $g_{n3}(\theta_0)$ of dimension $(T - 2) \times [T \times (T - 1) \times K/2]$.

$$
B_{23} = \sigma^2 \varepsilon E \begin{bmatrix}
2\alpha'_{n} [Q_{n,0}] & -\alpha'_{n} [Q_{n,0}, Q_{n,1}] & 0 & 0 \\
-\alpha'_{n} Q_{n,0} & 2\alpha'_{n} [Q_{n,0}, Q_{n,1}] & \ddots & \ddots & 0 \\
& 0 & \ddots & \ddots & 0 \\
& & \ddots & -\alpha'_{n} [Q_{n,0}, \cdots, Q_{n,T-2}] & \ddots & \ddots & 0 \\
& & & \ddots & 0 & -1'_{T-1} & 2t'_{T-2} & -t'_{T-1}
\end{bmatrix}
$$

$$
= \sigma^2 \alpha^2 G_5 \otimes L_k,
$$

where $G_5$ is of dimension $(T - 2) \times T (T - 1)/2$, $\alpha_T$ is a $T$ columns vector of ones and $L_k$ is a row vector of $K$ columns. Note that $L_k = E [\alpha'_{n} Q_{n,s}]$
where \( s = 0, 1, \cdots T - 2 \). The equality is guaranteed by the assumption of stationary process.

At true parameter \( \theta_0 = (\lambda_0, \rho_0)' \), the full covariance-variance matrix is

\[
\Omega_n = \begin{pmatrix}
\sigma^4_\varepsilon G_{1,T-1} \otimes \Delta_{nm} + 0_{(|T-1| \times M) \times (T-2)} & \mu_3 (G_{4,T-1} \otimes \varpi_{nm}' E [Q]) \\
(\mu_4 - 3\sigma^4_\varepsilon) G_{0,T-1} \otimes \varpi_{nm}' \varpi_{nm} & n\sigma^2_\varepsilon (\sigma^2_\alpha + \sigma^2_\varepsilon) G_{2,T-2} & \sigma^2_\varepsilon^2 G_5 \otimes L_k \\
0 & \sigma^2_\varepsilon^2 E [Q'(G_{3,T-1} \otimes I_n) Q]
\end{pmatrix}
\]

In particular, when random noise \( \varepsilon_{n,ij} \) are normally distributed,

\[
\Omega_n = \begin{pmatrix}
\sigma^4_\varepsilon G_{1,T-1} \otimes \Delta_{nm} & 0 & 0 \\
0 & n\sigma^2_\varepsilon^2 (\sigma^2_\alpha + \sigma^2_\varepsilon) G_{2,T-2} & \sigma^2_\varepsilon^2 G_5 \otimes L_k \\
0 & 0 & \sigma^2_\varepsilon^2 E [Q'(G_{3,T-1} \otimes I_n) Q]
\end{pmatrix}.
\]

Then the inverse of \( \Omega_n \) has the form as follows.

\[
\Omega_n^{-1} = \begin{pmatrix}
\sigma^{-4}_\varepsilon (G_{1,T-1} \otimes \Delta_{nm})^{-1} & 0 \\
0 & \Gamma_n^{-1}
\end{pmatrix}
\]

\[
\Gamma_n = \begin{pmatrix}
\sigma^2_\varepsilon^2 (\sigma^2_\alpha + \sigma^2_\varepsilon) G_{2,T-2} & \sigma^2_\varepsilon^2 G_5 \otimes L_k \\
\sigma^2_\varepsilon^2 E [Q'(G_{3,T-1} \otimes I_n) Q]
\end{pmatrix}.
\]

Note that even if random noises \( \varepsilon_{n,il} \) are symmetric \( (\mu_3 = 0) \), \( B_{23} = \sigma^2_\varepsilon^2 \sigma^2_\alpha G_5 \otimes L_k \) is not zero in general. Hence, the variance matrix \( \Omega_n \) can not be easily inverted.
CHAPTER 2
Analysis of Local School Expenditures in a Dynamic Game

2.1 Introduction

This paper examines the role of strategic interaction and intertemporal considerations in deciding local school spending. Elementary and secondary education are the single most important component of local public expenditures, about 35 percent of local direct expenditures\(^\text{13}\). In general, local governments of the United States are responsible for deciding public school expenditures. For example, Ohio has a state-wide School Board Association, but local school boards seem to “exercise a healthy autonomy” in financing public schools [Rose (1974, p. 250)].

Strategic interaction among state and local governments has been a major focus of theoretical work, and is now a focus of a growing empirical work in public finance. See Brueckner (2003) for an overview of recent literature. Substantial theoretical literature posits the existence of budget spillovers, but only a few empirical studies examine its consequences. Weisbrod (1964, 1965) assumes that school districts that have lower expenditures will encounter higher rates of out-migration, and provides some evidence consistent with this view. In modern spillover model, a jurisdiction chooses its own spending level, but the jurisdiction is also directly affected by the spending level chosen

\(^{13}\text{U.S. Bureau of the Census (2000, p. 494).}\)
elsewhere. Case, Rosen and Hines (1993) assume that the utility levels of US state governments depend on their own spending, as well as on spending of other similar states in a static model. They find that one state’s per capita expenditure is “positively and significantly affected by the expenditure levels of its neighbors” (p. 285). In particular, a one-dollar increase in education spending of one state’s neighbors increases its own education spending by over 70 cents.

The intertemporal approach assumes that agents are forward-looking. In modern economics, agents make their intertemporal decisions between present and future consumptions or investments. States and local governments often face balanced-budget restrictions, but they can and do smooth their spending over time. Several studies by Holtz-Eakin and Rosen [Rosen (1997 [7], [8], [9])] on local durable goods indicate that the permanent income model can well characterize spending behavior for some communities. However, researchers in local public finance typically assume that state and local governments are backward-looking, ignoring the dynamic issue. Holtz-Eakin, Rosen and Tilly (1994) find that state and local spending on non-durable goods is determined primarily by current resources. They assume that policymakers are not strategic players.

In this paper, I propose a dynamic game-theoretical framework that allows for both strategic interaction and intertemporal considerations in deciding local school expenditures. I formalize and test the notion that a local policymaker decides the public spending in a forward-looking, strategic framework. I show that a local policymaker will react to the current spend-
ing of his neighbors and the previous spending of his own and his neighbors in general equilibrium. In particular, a nonlinear restriction equation exists among structural parameters. If a local policymaker react positively to the current spending of his neighbors and the previous spending of his own, he will react negatively to the previous spending of his neighbors due to the presence of an intertemporal resources constraint. The restriction condition implied by intertemporal resource constraints will make it possible to distinguish between the forward-looking and backward-looking behaviors in a game-theoretical framework. The local school spending path can be specified as a spatial temporal autoregressive (STAR) panel data model. I use a consistent instrumental-variable (IV) estimator to provide empirical results.

This paper makes several contributions to the local public finance literature. First, this paper is the first to study both strategic interaction and intertemporal considerations in public finance. No empirical work has so far attempted to study the decision of local school spending in a dynamic game-theoretical framework. Previous work either ignores the dynamic issue and focuses on the effects of strategic interaction in a static game, or explores the role of intertemporal considerations and posits no existence of spillover effects in a dynamic model. This paper propose a more general framework than previous work. I show that local policymakers follow a spatial temporal autoregressive spending path in a dynamic game.

The second contribution is that I formalize a nested test model on local public spending. Empirical studies use the Campbell and Mankiw (1990) method to measure the extent to which consumption and investment can be
explained by permanent income hypothesis. In contrast to households and firms, state and local governments could behave strategically. It follows that their spending would be affected by their neighbors’ spending in addition to their own revenues. This paper design a nested test that incorporate both the revenue effects and neighborhood effects on local public spending.

The third contribution is that I provide an identification criterion between the static game and dynamic game. There exist a nonlinear restriction condition among structural parameters of a dynamic game. Empirical work has shown that there exists a positive correlation between current and previous spending in a dynamic model, and a positive interaction between state and local governments in a static game. If a local policymaker is forward-looking and strategically minded, he will react negatively to the previous spending of his neighbors as a result of an intertemporal resources constraint. In a static game, a backward-looking local policymaker could show a passive response to the past neighborhood effect, which is likely to be positive or negligible. Hence, the sign of past neighborhood effects and nonlinear restriction can make it possible to differentiate the dynamic game from the static game.

Fourth, I propose a simple and consistent IV method for estimation. The derived local spending path belongs to a class of spatial temporal autoregressive panel data models. Without spatial autoregressive term (that is, current neighborhood effects), the model is a conventional dynamic panel data model that can be consistently estimated by IV method. Without dynamic terms (that is, the past own and neighborhood spending), the model is a static spatial panel data model that can be estimated by either Maximum-
Likelihood (ML) or IV method. Case et al (1993) use ML method to estimate their static spatial econometric equations. Following their work, I specify the neighborhood of local school districts by geographic, economic or demographic proximity. Case et al find that racial composition measured by the percent of the population that is black, performs significantly better than the other two. It is in question whether the fiscal dependence among local school districts mirrors what happens among US states. This paper will consider various combinations of neighborliness in spatial econometric models. The ML method is computationally infeasible to estimate the spatial models with two or more spatial autoregressive terms. The first chapter considers a generalized spatial two-stage least-square (GS2SLS) method to estimate the spatial temporal autoregressive panel data model. The GS2SLS estimator is simple, consistent and robust against unknown initial conditions. Further, the estimator can extend to higher-order spatial temporal autoregressive models without computational difficulty.

The plan of this paper is as follows. Section 2.2 considers a dynamic game that incorporates strategic interaction and intertemporal considerations in deciding local school spending. Identification issues between a dynamic game and a static game are addressed. Section 2.3 discusses empirical specifications implied by the theoretical framework and their econometric issues. A GS2SLS estimator is proposed. Section 2.4 describes the data. Section 2.5 presents the empirical results. Conclusion is drawn in section 2.6 and appendices is in section 2.7.
2.2 Theoretical consideration

The notations used throughout the paper are:

\( i = 1, 2, \cdots, n \) where \( n \) is the total number of communities;

\( t = 0, 1, \cdots, T - 1 \) where \( T \) is the ending period of time;

\( E_t \) = mathematical expectation conditional on all available information through the end of period \( t \);

\( U(\cdot) \) = period-specific, concave utility function;

\( \phi \) = constant rate of subjective time preference;

\( r \) = constant real interest rate;

\( Y_{it} \) = consumer’s income;

\( C_{it} \) = a private consumption good;

\( H_{it} \) = housing consumption;

\( G_{it} \) = a locally provided public good;

\( G_{-i,t} \) = a local public good provided by neighbors;

\( R_{it} \) = local government revenue;

\( A_{it} \) = net wealth stock at the end of period \( t \);

\( r_{it} \) = net-of-tax return on land at period \( t \);

\( t_{it} \) = local property tax rate levied at period \( t \).

2.2.1 A modern Tiebout model

This paper considers a modern Tiebout (1956) model where a local public good\(^{14}\) (that is, elementary and secondary education) exhibits externality.

\(^{14}\)A local public good (like school education or garbage collection) is not a pure public good, but stops at a community’s boundary.
among communities (that is, local school districts). The economy has $n$ heterogeneous communities. Consumers are fully mobile and will move to the community which best satisfies their demand for local public goods. Self-sorting leads to homogeneous preferences inside each community.

The goal of a forward-looking consumer is to maximize his discounted lifetime utility function under uncertainty, satisfying the present value of his permanent income constraint.

$$\max_{\{C_{i,t+s}\}} \ V_{it} = E_t \left\{ \sum_{s=0}^{\infty} \frac{1}{(1+\phi)^s} U (C_{i,t+s}) \right\}, \quad (39)$$

$$s.t. \ A_{it} + \sum \frac{1}{(1+r)^s} C_{i,t+s} = \sum \frac{1}{(1+r)^s} Y_{i,t+s}^{d}, \quad (40)$$

where disposable income $Y_{i,t+s}^{d} = (Y_{i,t+s} - R_{i,t+s})$ is the difference of gross income and local tax revenues.

A local policymaker, elected by his community residents, will act in the interests of the community to decide the optimal spending level of a local public good. A representative consumer residing in community $i$ is endowed with $Y_{it}$ units of stochastic goods and a fixed amount of residential land $L_i$. There is no other capital than land. Housing is produced by land only. The quantity of residential land is fixed, indicating perfectly inelastic housing supply. The community levies a property tax on housing consumption to finance the provision of a local public good. The net-of-tax return of land is $r_{it}$, and the property tax rate is $t_{it}$, both determined exogenously. The assumption of exogeneity does not affect the major results in the following sec-
gross price per unit of land is $r_{it} + t_{it}$.

Consumer $i$ rents land $L_i$ from himself as a landlord, and pays gross price $r_{it} + t_{it}$ for housing consumption $H_{it}$. The consumer and landowner receives $r_{it}L_i$ as land returns, and the community collects $t_{it}L_i$ as government revenues $R_{it}$ to provide a local public good $G_{it}$. Note that balanced-budget law require that $\sum_{s} \frac{1}{(1+r)^s} (R_{i,t+s} - G_{i,t+s}) = 0$. Housing market clearing condition requires housing consumption equal per capita land endowment: $H_{it} = (r_{it} + t_{it}) L_i$.

Substituting the balanced-budget constraint and housing market clearing condition into the following accounting identity,

$$A_{it} + \sum_{s} \frac{1}{(1+r)^s} \left[ G_{i,t+s} + H_{i,t+s} + G_{i,t+s} \right] = \sum_{s} \frac{1}{(1+r)^s} \left[ Y_{i,t+s} + (r_{i,t+s} + t_{i,t+s}) L_i \right],$$

we can derive the above permanent income constraint (40).

The disposable income $Y_t^{d}$ is stochastic, which is the only source of uncertainty in the consumer’s problem. If the consumer maximizes his expected utility that is quadratic, both marginal utility and consumption obey an exact dynamic regression path.

### 2.2.2 A dynamic game

A local policymaker is concerned with the public spending in the present period as well as in the infinite future. The goal of a forward-looking poli-
tions. Land return and property tax rate are potentially endogenous. But such consider-
ations are beyond the scope of current work, and will be of interest in future work.
cymaker is to maximize his discounted lifetime utility function under uncertainty, satisfying the present value of his permanent resources constraint.

\[
\max_{\{G_{i,t+s}\}} V_{it} = E_t \left\{ \sum \frac{1}{(1 + \phi)^s} U \left( G_{i,t+s} \right) \right\},
\]

(41)

\[
A_{it} + \sum \frac{1}{(1 + r)^s} G_{i,t+s} = \sum \frac{1}{(1 + r)^s} R_{i,t+s},
\]

(42)

where equation (42) is implied by the balanced-budget rule.

Strategic interaction can arise through budget spillovers and fiscal policy interdependence. In the spillover model, a community \( i \) chooses the spending level \( G_i \), but the community is also directly affected by the spending \( G_{-i} \) of other communities. The school expenditures of one community may influence the utility of residents elsewhere either because young people move outside the city where they once attended school, or because well-educated people in one community compete with people in other communities in labor markets [Case et al (1993, p.288)]. If educated people are inclined to move into the school districts that have higher education spending [Weisbrod (1964, 1965)], local communities are motivated to maintain the spending level and avoid falling behind others. Such notions can be formulated into the modern game-theoretical context where local policymakers respond strategically to their neighbors.

The spending \( G_{-i} \) chosen by other communities enters the utility function of a strategic local policymaker \( i \) directly as follows

\[
U \left( G_{it} \right) = U \left( G_{it} - \lambda G_{-i,t} \right),
\]

50
where $\lambda$ denotes the interactive parameter, capturing the extent to which a local policymaker reacts to his neighbors.

Assume that there exists a unique equilibrium in the system. The dynamic game above can be solved in the two-step procedure.

Step 1. For the moment, a local policymaker, taking the spending of other communities $G_{-i}$ as given, makes his optimal intertemporal decision as usual. His optimal spending path $\{G_{it}\}$ is characterized by the system of Euler equations.

$$
E_t[U'(G_{i,t+1})] = \rho U'(G_{it}),
$$
where the dynamic parameter $\rho = [(1 + \phi)/(1 + r)] \leq 1^{16}$.

Step 2. Then all local policymakers choose their optimal spending levels simultaneously. Assume that the local policymaker has a quadratic utility function as follows.

$$
U(G_{it}) = -\frac{1}{2} (G_{it} - \lambda G_{-i,t} - \bar{G})^2
$$
where $\bar{G}$ is the bliss level of local spending. Substituting the utility function into the above Euler equation, I obtain the optimal local spending path.

$$
G_{i,t+1} = (1 - \rho) \bar{G} + \lambda G_{-i,t+1} + \rho G_{it} - \lambda \rho G_{-i,t} + \xi_{i,t+1},
$$
where $\xi_{i,t+1}$ is a rational error term with zero expectation conditional on $I_t$, a set of all available information through the end of period $t$. That is,

$$
E[\xi_{i,t+1}|I_t] = 0.
$$

Note that structure parameters $\phi$ and $r$ are unidentifiable in the econometric model.
Denote that \( \mu = (1 - \rho) \bar{G} \) and \( \delta = -\lambda \rho \). The local spending equation obeys an exact Spatial Temporal AutoRegressive path as follows.

\[
G_{i,t+1} = \mu + \lambda G_{-i,t+1} + \rho G_{it} + \delta G_{-i,t} + \xi_{i,t+1},
\]  

(43)

subject to a nonlinear restriction \( \delta + \lambda \rho = 0 \). For a generic utility function, the Euler equation can be approximated by equation (43) using a first-order Taylor series expansion.

The structure parameters \((\lambda, \rho, \delta)\) are of central interest in this paper. In general, theoretic literature cannot determine the sign of interactive parameter \( \lambda \) or dynamic parameter \( \rho \) a priori, which must be estimated empirically\textsuperscript{17}. But the nonlinear restriction says that if the signs of any two parameters are known, the sign of the third parameter must be opposite to their product. In particular, if parameters \( \lambda \) and \( \rho \) are positive, the previous neighborhood spending shall have a negative impact on current spending of community \( i \). Suppose that a local policymaker observes that his neighbors will raise their spending on average at period \( t \). If he reacts strategically and decides to increase his current spending at period \( t \), the policymaker shall cut his future spending at period \( t + 1 \) in order to maintain his intertemporal resources constraint. Hence, an increase in his neighbors’ current spending is negatively correlated with the future spending of a forward-looking policymaker in a game-theoretical framework.

\textbf{2.2.3 A static game}

\textsuperscript{17}Empirical results often find that both parameters are positive.
It is possible for a local policymaker to follow a Keynesian spending path. The goal of a backward-looking policymaker is to maximize his current utility function, $U(G_t)$, satisfying his current resources constraint, $G_t \leq R_t$. The history of aggregate and individual spending can affect a local policymaker’s utility that exhibits adaptive expectation or partial adjustment. Strategic interaction can also arise through budget spillovers in the utility function of a backward-looking but strategic local policymaker. Assume that his utility has a quadratic form as follows

$$U = -\frac{1}{2} \left( G_{i,t+1} - \lambda G_{-i,t+1} - \rho G_{it} - \delta G_{-i,t} \right)^2.$$

A first-order condition leads to a spatial temporal autoregressive panel data model, which has the same functional form as equation (43). However, the static model does not imply any functional relationship among the structural parameters $\lambda$, $\rho$ and $\delta$. If the interactive parameter $\lambda$ is zero, equation (43) reduces to a conventional dynamic panel data model\(^{18}\). If parameters $\rho$ and $\delta$ are zero, the equation is a static spatial panel data model, used by Case et al (1993) to study US state expenditure data. If current local spending is determined primarily by current local revenues, a local policymaker’s decision can be well characterized by a static spatial panel data models. If a strategic policymaker adjusts his spending in response to his neighbors’ decision, the previous neighborhood spending is very likely to have same effect

\(^{18}\)See Johnston (1972 p. 300-303) for a number of models that can produce an estimation equation with temporally lagged dependent variables on the right-hand side.
as the current neighborhood spending. That is, parameters $\lambda$ and $\delta$ would have the same (positive) signs. The adaptive expectation could justify the presence of lagged dependent variable as well as the previous neighborhood effects. This also imply that the parameter $\delta$ would have the same (positive) sign as dynamic parameter $\rho$. Hence, there would exist either negligible or positive previous neighborhood effects in a static game where parameters $\lambda$ and $\delta$ are positive.

2.2.4 Model Identification

Assuming the presence of strategic interaction, both the dynamic game and the static game can imply the same functional form, that is, spatial temporal autoregressive panel data model. But the dynamic game imposes a nonlinear restriction among the structural parameters, while the static game does not. In general, the interactive parameter $\lambda$ and the dynamic parameter $\rho$ are positive in empirical literature. Hence, the nonlinear restriction implies that the previous neighborhood spending has a negative effect on the current spending of a local government. In contrast, the previous neighborhood effects are either negligible or positive in a static game. This provides a simple sign test about the forward-looking assumption as follows.

$$H_0 : \delta \geq 0, \quad vs. \quad H_1 : \delta < 0.$$ 

If there exists an important negative previous neighborhood effect in his current spending equation, the local policymaker is likely to make an intertemporal consideration to balance his multi-years budget constraint. A
direct test on nonlinear restriction equation can further measure the extent to which the local decision making can be explained by the dynamic game. If the values of parameters $\lambda$ and $\rho$ are between zero and one, then the absolute value of parameter $\delta$ will be less than $\lambda$ or $\rho$. Say when $\lambda = 0.5$ and $\rho = 0.2$, $\delta = -\lambda \rho = -0.1$. Hence, the nonlinear restriction can not only predict the sign of $\delta$, but also its magnitude. In short, the nonlinear restriction makes it possible to differentiate between the forward- and backward-looking assumptions in a game-theoretical context.

So far our analysis of local policy decision has implicitly assumed the existence of strategic interaction. However, if there exists no strategic interdependence among local spending at first, the further identification between dynamic game and static game will be meaningless. The parameter signs and values are a priori unknown. We have to explore all meaningful cases that concern the behaviors of local policymakers.

First, $\lambda = 0$ implies that local policy makers are non-strategic players. The signs and magnitudes of dynamic parameters cannot justify the forward- or backward-looking assumptions of local policymakers.

Second, $\lambda > 0$ implies that local policy makers are strategic players. If dynamic parameters $\rho$ and $\delta$ are zero, the policymakers are static players.

Finally, assume that $\lambda > 0$ and the dynamic parameter $\rho > 0$. Hence, if $\delta < 0$, local policymakers are likely forward-looking strategic players; if $\delta \geq 0$, they are backward-looking policymakers.

Note that we have not exhausted all possible cases in above analysis. For example, $\lambda < 0$ or $\rho < 0$ or both. However, these cases does not have much
economic meanings and often empirically impossible.

2.3 Empirical Implementation

2.3.1 Econometric model

Theoretical framework implies that the spending path of a local school district follows a spatial temporal autoregressive pattern, depending on its own past spending and its neighbors’ current and past spending.

By definition, the sequence of information set \( \{I_t\} \) is monotone and increasing, i.e., \( I_t \subseteq I_{t+1} \). The law of conditional expectation implies that 
\[
E[\xi_{i,t+1} | I_t] = E[E(\xi_{i,t+1} | I_{t+1}) | I_t] = 0.
\]
Conditional on observed and unobserved characteristics at current period \( t+1 \), the rational error term has a linear projection as follows.

\[
\xi_{i,t+1} = X_{i,t+1} \beta + a_i + v_{i,t+1}, \tag{44}
\]

where \( X_{i,t+1} \) refers to (strongly) exogenous variables that affect the current spending of community \( i \), \( a_i \) is unobserved district-specific characteristics and \( v_{i,t+1} \) is a random noise with zero mean.

Substituting the linear projection into equation (43), the empirical specification has a conventional panel data model form.

\[
G_{i,t+1} = \mu + \lambda G_{-i,t+1} + \rho G_{it} + \delta G_{-i,t} + X_{i,t+1} \beta + u_{i,t+1}, \tag{45}
\]

where \( u_{i,t+1} \) contains two error components: district time-invariant effect \( a_i \) and random error \( v_{i,t+1} \). Case et al (1993) estimate a static case of the above equation that parameters \( \rho \) and \( \delta \) equal zero.
2.3.2 Specifying spatial weights matrix

The inclusion of $G_{-1}$ raises issues of specifying meaningful neighbors in the econometric model. An obvious candidate is geographical proximity. Local school districts often view their neighbors that share a common boundary as strategic opponents. Although geography may be relevant, it is not the only factor that can determine neighbors. Local school districts with similar economic or demographic characteristics may also influence each other within some distance. In the countryside of Ohio, neighboring rural school districts often have very similar economic or demographic characteristics. Geographic neighbors often coincide with economic or demographic neighbors. However, within the metropolitan areas, economic similarity may be more relevant than geographic closeness in specifying neighbors of urban school districts. For example, within Columbus Metropolitan Area, Upper Arlington school board may view both Columbus city and Bexley city as its strategic neighbors. Columbus city is a geographic neighbor of Upper Arlington city, but has a much lower average income. Although Bexley and Upper Arlington do not share a common border, their residents have similar income and education level and seem to have a higher demand for education.

Specifying neighbors is equivalent to specifying a spatial weights matrix in the spatial autoregressive model. The literature in spatial econometrics use $W_n$ to denote a generic spatial weights matrix. I construct spatial weights matrix $W_n$ based on geographic contiguity, per capita income, and percentage of the pupils that is black. These criteria for neighborliness has been explored
in specifying state neighbors by Case et al (1993). They use the following methods to construct different spatial weights matrices.

\( W^G_n \), neighbors with common borders. \( w_{ij} = 1/S_i \) if \( i \) and \( j \) share a border; \( w_{ij} = 0 \) otherwise; and \( S_i = \) the number of borders state \( i \) shares.

\( W^I_n \), neighbors with similar average income:

\[
w_{ij} = \frac{1}{|INCOME_i - INCOME_j|} / S_i,
\]

where \( INCOME_i \) is median per capita income in district \( i \) over sample period; and \( S_i \) is the sum of \( 1/|INCOME_i - INCOME_j| \) over all urban school districts other than \( i \) within the same metropolitan area.

\( W^B_n \), neighbors with similar proportion of black pupils:

\[
w_{ij} = \frac{1}{|BLACK_i - BLACK_j|} / S_i,
\]

where \( BLACK_i \) is median percentage of black pupils in district \( i \) over sample period, and \( S_i \) is defined similarly as above.

It is clear that all the specified spatial weights matrices have a zero diagonal. For a cross-sectional unit \( i \), its neighbors’ spending \( \sum_j w_{ij}G_{jt} \), a weighted average of neighboring values. The econometric equation (45) for all cross-sectional units in the period \( t + 1 \) can be written in matrix form.

\[
G_{n,t+1} = \mu + \sum_{k=1}^p \lambda_k W^k_n G_{n,t+1} + \rho G_{nt} + \sum_{k=1}^p \delta_k W^k_n G_{nt} + X_{n,t+1} \beta + a_n + v_{n,t+1}, \tag{46}
\]

where \( p \) is the total number of spatial weights matrices used in equation (46).

As a convention, \( W^k_n G_{n,t+1} \) is known as a \( k \)th order spatial autoregressive
(SAR) lag term, $G_{nt}$ is a temporal autoregressive (TAR) lag, and $W^k_n G_{nt}$ is a $k$-th order spatio-temporal autoregressive (STAR) lag.

Note that $W^G_n$ is statewide spatial weights matrix, but $W^I_n$ and $W^B_n$ are limited for urban school districts in metropolitan areas. Within the same metropolitan areas, the proportion of black pupils in schools is closely correlated with the average income in school districts. To avoid potential multicollinearity problem, I combine $W^I_n$ and $W^B_n$ separately with $W^G_n$ to estimate equation (46). In estimation, $W^G_n G_{n,t+1}$ is a first-order spatial autoregressive term in default, and $W^I_n G_{n,t+1}$ or $W^B_n G_{n,t+1}$ is a second-order spatial autoregressive term. The number of order $p$ is either 1 or 2 in equation (46).

To test the forward-looking assumption, the sign test and nonlinear restriction in equation (46) take the forms as $H_0$: $\delta_k \geq 0$ versus $H_1$: $\delta_k < 0$ and $\lambda_k \rho_k + \delta_k = 0$ for $k = 1$ or 2.

2.3.3 Econometric issues

Several econometric issues must be confronted with in the estimation of spatial temporal autoregressive panel data model, which has a large number of cross-sectional units but a short period of time in general. They are:

1. Possible correlation between observed $X_{nt}$ and the unobserved $a_n$;
2. Inclusion of temporally lagged spending $G_{nt}$ and $W_n G_{nt}$;
3. Endogeneity of the current neighborhood spending $W_n G_{n,t+1}$;
4. Possible spatial error dependence.

These issues\(^\text{19}\) are addressed in turn as follows.

\(^{19}\)Brueckner (2002) points out these problems (except for the second one) in his empirical research.
Correlation between observed and unobserved characteristics

There exists possible correlation between observed characteristics \( X_{nt} \) and unobserved effects \( a_n \). The random-effects assumption neglects such correlation between \( X_{nt} \) and \( a_n \), but the fixed-effects formulation recognizes such correlation and makes inference conditional on the effects in the sample data [Hsiao (2003)]. If included explanatory variables are indeed correlated with the error term, the fixed-effects estimator remains consistent for a short period, while the random-effects estimator will be inconsistent if the total period of time \( T \) is fixed.

The correlation between observed and unobserved characteristics could arise through omission of unobserved characteristics as well as endogenous sorting of residents across school districts. For instance, a large number of professors at The Ohio State University live in the Upper Arlington city, which is close to the Columbus campus. In general, well-educated professors (who are also well-paid) have a high demand for elementary and secondary education, thus ending up residing in a community with good public schools. Average income in local districts may be observed, while the degree of education in the community may not. Income has a positive effect on the spending level as dependent variables, and a positive correlation with the education, captured by the error term. A positive correlation between income and education will generate a positive correlation between income and error, leading to an inconsistent estimate of the income coefficient and potentially studies.
distorting others.

**Inclusion of temporally lagged spending**

Both fixed-effects and random-effects models incur serious issues if temporally lagged dependent variables (TAR and STAR lag terms) are included as explanatory variables. The maximum-likelihood estimator under the fixed-effects model is no longer consistent in a typical panel data set, involving only a short period of time. The first-order difference eliminates the time-invariant individual effects ($a_n$), but creates the correlation bias of order $(1/T)$ between the explanatory variables and the residuals in the transformed model. The random-effects model also confronts the serious problem raised by the initial conditions of a dynamic process, even allowing for the correlation between $X_{nt}$ and $a_n$. Different assumptions on initial conditions imply different likelihood functions, and incorrect choice of initial condition shall lead to inconsistent estimators in general. In a short panel data set, the consistency of maximum-likelihood estimators depends crucially on the assumption of initial observations, which turns more complicated in the existence of the spatial autocorrelation. Hence, an instrument-variables approach is used to provide a consistent estimator that is robust against various initial conditions for the spatial temporal autoregressive model.

**Endogeneity of current neighborhood spending**

Endogenous issue arises from simultaneous decisions of current spending by all individuals. The weighted average of current neighborhood spending
(SAR lag) is endogenous and correlated with the error term in the econometric equation (46). Such point can be understood in a simple first-order spatial autoregressive process: $G_n = \lambda W_n G_n + \epsilon_n$. Assume that the matrix $I_n - \lambda W_n$ is invertible, the simultaneous equilibrium of the spatial autoregressive process is $G_n = (I_n - \lambda W_n)^{-1} \epsilon_n$. It follows that the explanatory variable $W_n G_n$ equals $W_n (I_n - \lambda W_n)^{-1} \epsilon_n$. The SAR lag $W_n G_n$ is generally correlated with the error term $\epsilon_n$, even if $\epsilon_n$ is independently and identically distributed disturbance.

$$E [(W_n G_n)' \epsilon_n] = \sigma_0^2 \text{trace} [W_n (I_n - \lambda W_n)^{-1}] \neq 0,$$

where $\sigma_0^2$ is the variance of random noise $\epsilon$. The trace of the matrix $W_n (I_n - \lambda W_n)^{-1}$ is not zero in general. This result holds in generalized models including the spatial temporal autoregressive process. For consistent estimation of the spatial autoregressive parameter $\lambda$, valid instrumental variables shall be constructed so that they are correlated with the SAR lag but uncorrelated with the error term.

**Spatial error dependence**

Spatial error dependence arises when unobserved effects captured by the error terms are spatially correlated. Ignoring potential spatial error dependence, estimation of spatial autoregressive model could generate spuriously significant strategic interaction in equation (46). Allowing for possible spatial error correlation, I assume that the random shock $\nu_{nt}$ in equation (46)
follows a $k$th-order spatial autoregressive process.

$$v_{nt} = \sum_{k=1}^{p} \psi_k M_n^k v_{nt} + \epsilon_{nt},$$

(47)

where $M_n^k$ is a $k$th-order spatial weights matrix, $\psi$ is a spatial error parameter, and $\epsilon_{nt}$ is assumed to be independently and identically distributed\textsuperscript{20}. In general, the number of order $p$ and the spatial weights matrices $M_n$ in error process are the same as the order $p$ and matrices $W_n$ in the model equation (46).

### 2.3.4 A consistent G2SLS estimator

This paper considers an instrument-variables approach under the fixed-effects formulation about panel data model. The first chapter proposes a generalized spatial 2SLS (GS2SLS) estimator for the spatial temporal autoregressive panel data model. The G2SLS estimator has a linear closed-form expression. It is consistent and asymptotically normal as the number of cross-sectional units tend to be large. In principle, the 2SLS approach is less efficient than the Generalized Method of Moments (GMM) that exploits all available moment conditions (linear and nonlinear). Monte Carlo experiments show that if the exogenous variables are relevant, both 2SLS and GMM estimators have good small-sample properties and differ little in large sample properties. Furthermore, the 2SLS method is robust against the unknown form of error process, while efficient GMM estimator requires explicit assumption of error process.

\textsuperscript{20}Homoskedastic assumption can be relaxed in 2SLS estimation method.
A consistent GS2SLS estimator can be obtained by the following procedures.

Step 1. Eliminate the fixed effects $\alpha_n$ in equation (46) by first-order difference. The transformed model equation is given as follows.

$$\Delta G_{n,t+1} = \sum_{k=1}^{p} \lambda_k W_n^k \Delta G_{n,t+1} + \rho \Delta G_{nt} + \sum_{k=1}^{p} \delta_k W_n^k \Delta G_{nt} + \Delta X_{n,t+1} \beta + \Delta v_{n,t+1}$$

(48)

Step 2. Construct the valid IVs for the SAR, TAR and STAR lag terms. All these lag terms are correlated with the differenced error terms in equation (10). Literature in dynamic panel data model suggests that all temporal lags of level decision $G_{ns}$ and $W_n G_{ns}$ ($s \leq t - 1$) dated $t - 1$ and earlier are valid linear instruments for $\Delta G_{nt}$ and $W_n \Delta G_{nt}$. For the SAR lag, the simultaneous equilibrium solution of equation (48) is

$$\Delta G_{n,t+1} = \left( I_n - \sum_{k=1}^{p} \lambda_k W_n^k \right)^{-1} \left[ \rho \Delta G_{nt} + \sum_{k=1}^{p} \delta_k W_n^k \Delta G_{nt} + \Delta X_{n,t+1} \beta + \Delta v_{n,t+1} \right].$$

(49)

Let $P_n$ be a class of constant square matrix such as $W_n$, $W_n^2$, $W_n^3$, etc. Following Kelejian & Prucha (1998) and Lee (2003), I use $P_n G_{ns}$ ($s \leq t - 1$) and $P_n X_{nr}$ ($r = 1, 2, \cdots, T$) as instruments of a generic SAR lag term $W_n \Delta G_{n,t+1}$ in equation (48).

Step 3. Compute the consistent GS2SLS estimates using the formula in Appendix A.

Step 4. Compute the residuals $\Delta v_{n,t+1}$ in equation (48) and estimate the coefficient $\psi_k$ in equation (47) by a simple MOM method in Appendix B.
Step 5. Compute the standard errors of GS2SLS estimates using the updated variance matrix in Appendix A.

2.4 Data

I estimate the econometric equation (46) using annual data on the 612 school districts in the state of Ohio over the period 1994-1998. Total revenues for local school district is the sum of revenues from local, intermediate, state and federal sources, including local and state subsidies, state and federal funds, food service, lotteries and fees. Total current operating expenditures for local school district \( i \) in year \( t \), \( G_{it} \), are put on a dollar figure per pupil basis and adjusted from fiscal year 1993.

The following variables comprise of the regressor matrix \( X_{nt} \), which are considered strictly exogenous in the econometric equation.

1. Pupil density;
2. Real per tax return income;
3. Income squared;
4. Proportion of the pupils that are black;
5. Proportion of the population between 5-17 years old;
6. Proportion of the population at least 65 years old;
7. State grants;
8. Regional economic context.
9. Time effects

Pupil density, the total number of pupils divided by the area of school districts, captures the possible congestion effects and scale economies in the
provision of public education. Income measures the available resources for local spending, and the squares of income catch the potentially nonlinear effects of changing income. Income is put on a thousand-dollar figure per tax return provide by Ohio Department of Taxation and deflated using GDP deflator (The base year is 1994). Local school districts with different age and racial structures may have different preference for elementary and secondary education, which justifies the presence of the demographic variables. Proportion of pupils that are black is the non-white enrollment divided by the total enrollment in local school districts. The Census has no annual population data by age on local school districts. Only the county population data are provided and thus used as proxy for local population data. State grants are funds from state sources and put on a dollar figure per pupil basis. State grants are an attempt to raise funding in relatively poor school districts. The fiscal behaviors of one school district may vary in some way with regional economic situations. This is known as an exogenous “contextual effect” in the sociological literature. I use the product of geographic spatial weights matrix and local real income to stand for regional contextual effect. Time effects are used to capture some year-to-year change of state programs or macroeconomic variables that affects all school districts similarly. Both time dummy effects and time trend effects will be considered in empirical work.

21 However, even these data are not surveyed but produced by a method which is still in a developmental stage according to the Census. They may not be accurate for populations which are very small or have unusual age distributions. Hence the Bureau warns to use these figures with caution.
Table 3 presents descriptive statistics for the data.

I construct the geographic spatial weights matrix $W^n_G$ based on the GIS map provided by Center for Urban and Regional Analysis at The Ohio State University. Ohio has 8 Metropolitan Statistical Areas, centered in Akron, Canton, Cincinnati, Cleveland, Columbus, Dayton, Toledo and Youngstown. They are also the county seats of Summit, Stark, Hamilton, Cuyahoga, Franklin, Montgomery, Lucas and Mahoning. The urban school districts in these counties are used to construct $W^n_I$ and $W^n_B$ spatial weights matrices.

2.5 Results

2.5.1 Diagnostic tests

The Hall’s random-walk hypothesis implies that the change in household consumption is unpredictable under the permanent-income/life-cycle framework. No information available at time $t$ can be used to forecast the change in household consumption from $t$ to $t+1$. Campbell and Mankiw (1990) suggest an econometric test that allows one to measure the quantitative departure from the model. Holtz-Eakin et al (1994) use their test to understand the state and local government spending. The basic estimation equation is

$$\Delta \ln G_{n,t+1} = \alpha_0 + \alpha_1 \Delta \ln R_{n,t+1} + \epsilon_{n,t+1}, \quad (50)$$

where $\alpha_0$ is the intercept and slope coefficient $\alpha_1$ measures the quantitative departure from random-walk hypothesis. The equation nests both forward-looking ($\alpha_1 = 0$) and backward-looking ($\alpha_1 = 1$) decision making. Table 4 presents the OLS and IV estimates of the equation (50). OLS may be an
inappropriate estimation method, because the right-hand-side (RHS) variable \( \Delta \ln R_{n,t+1} \) are potentially correlated with error terms. Therefore, I present the results estimated by the IV method that uses the lagged value of revenues dated \( t - 1 \) and earlier. All estimates of \( \alpha_1 \) are positive and statistically significant. The OLS estimate of \( \alpha_1 \) is 0.38 with standard error 0.02; its IV estimate using differenced lags (\( \Delta \ln R_{n,t-1} \) and earlier) is 0.39 with standard error 0.14, its IV estimate using level lags (\( \ln R_{n,t-1} \) and earlier) is 0.67 with standard error 0.13. Thus, all the estimates suggest quantitatively large and statistically significant departures from the random-walk hypothesis: local school spending appears to increase by about 40 cents up to 67 cents in response to an anticipated one-dollar increase in local revenues, and the null hypothesis of no effect is strongly rejected. At the same time, the estimates of \( \alpha_1 \) are far below unity, suggesting the presence of smoothing spending by local governments.

When Holtz-Eakin et al (1993) apply the test of Campbell and Mankiw on state and local government spending, they ignore the potential presence of budget spillovers and fiscal policy interdependence among states and communities. Allowing for the neighborhood effects, I suggest a nested test equation that adds a new explanatory variable \( \Delta \ln W_n G_{n,t+1} \) on the RHS of the equation.

\[
\Delta \ln G_{n,t+1} = \alpha_0 + \alpha_1 \Delta \ln R_{n,t+1} + \alpha_2 \Delta \ln W_n G_{n,t+1} + \varepsilon_{n,t+1}.
\]  
(51)

If \( \alpha_2 = 0 \), equation (51) is the same as the one of Campbell and Mankiw. Table 4 also presents the OLS and IV estimates of equation (51). The OLS
estimate of $\alpha_1$ is 0.36 and statistically significant, but the IV estimates of $\alpha_1$ using differenced or level lags turn out to be negative and statistically indifferent from zero. The IV results reveal that the effects of revenue change found in either OLS or equation (50) are more spurious than real. All the estimates of $\alpha_2$ are positive and statistically significant, indicating the relevance of neighborhood change in school spending. The OLS estimate of $\alpha_2$ is 0.26 with standard error 0.04; its IV estimate using differenced lags($\Delta \ln W_n G_{n,t-1}$ and earlier) is 0.83 with standard error 0.20, its IV estimate using level lags ($\ln W_n G_{n,t-1}$ and earlier) is 0.90 with standard error 0.12. A local school district appears to increase its spending by about 83 cents up to 90 cents in response to an anticipated one-dollar increase in its neighborhood spending. One would easily reject the hypothesis of no neighborhood effects ($\alpha_2 = 0$), and note that the large magnitude of $\alpha_2$ implies the importance of neighborhood effects in deciding local school expenditures.

Next, I present a simple diagnostic test for the presence of spatial dependence based on IV residuals of a dynamic panel data model in Appendix C. Spurious spatial autocorrelation may arise if the functional form is misspecified. For example, a local policymaker can react to the lagged spending of his own and neighbors, but not to the current spending of his neighbors. Omission of lagged neighbors’ spending may produce a spurious strategic interaction among local school districts. Hence, the dynamic panel data model takes the following general form.

$$G_{n,t+1} = \rho G_{nt} + \delta W_n G_{nt} + X_{n,t+1} \beta + a_n + v_{n,t+1}.$$  \hspace{1cm} (52)
Equation (52) differs from equation (46) in the exclusion of SAR lag, and from conventional dynamic model in the inclusion of STAR lag. The parameters \((\rho, \delta, \beta')\) are estimated by the fixed-effects 2SLS estimator suggested by dynamic panel literature. Note that only the differenced residuals \(\Delta v_{n,t+1}\) can be obtained by substituting the 2SLS estimates into the first-differenced equation. The differenced 2SLS residuals are cross-sectionally independent for fixed period, but serially correlated between \(t\) and \(t+1\). Appendix C provides a simple Gradient test or Lagrangian multiplier (LM) test based on the differenced 2SLS residuals. The test statistic has a chi-square distribution with one degree of freedom. The value of LM test is 550, which suggests the strong presence of either spatial lag dependence or spatial error dependence in equation (52).

The spatial weights matrix used in this subsection is statewide \(W_n^{G}\) matrix. The urban weights matrix \(W_n^{I}\) or \(W_n^{B}\) together with \(W_n^{G}\) does not change the main results.

2.5.2 Interpreting the coefficients

Table 5-7 present the estimation results on local school spending that do not contain statewide time effects. Table 5 uses only the statewide geographic weights matrix; Table 6 and 7 add urban black or income weights matrix in the equations. Column (2) estimates a dynamic panel data model using equation (52); column (3) - (6) present consistent IV estimates of equation (46). To illustrate, I also provides inconsistent OLS estimates of equation (46) in column (1). Multicollinearity problem may exist in the RHS variables, in
particular, proportions of school children and elderly people. The IV estimates of these two demographic variables in column (2) have much larger values and standard errors than their corresponding OLS estimates, suggesting potential multicollinearity. The inclusion and deletion of multicollinear variables could cause major changes in the estimates of other regressors. To detect how severe multicollinearity exists in the equation, column (3) drops collinear demographic variables, and column (4) includes them in equation. A comparison between the results of column (3) and (4) reveals negligible changes in IV estimates and their standard errors, suggesting little concern over multicollinearity. To avoid potential omitted variable bias, I keep two demographic variables in column (5) and (6). Column (5) includes state grants that have potential important influences on local spending, and column (6) further includes regional economic situations to capture potential contextual effects on local spending.

Column (2) reports the IV estimates of equation (52). Both TAR and STAR lag terms are positive and statistically significant. Equation (52) can be viewed as a reduced form of equation (46). So the positive previous neighborhood effects cannot account for the rejection of forward-looking assumption. The LM test in section 5.1 uses the computed 2SLS residuals of $\Delta v_{n,t+1}$ to detect the existence of spatial dependence. Assume that the error term $v_{n,t}$ follows the first-order spatial autoregressive process. In Table 5, the MOM estimate of SAR parameter $\psi$ is 0.36 with standard error 0.03, which suggests the presence of significant spatial autocorrelation and is consistent with the previous LM test value. Table 6 and 7 have similar results about
IV estimates and spatial error process in column (2).

Empirical results on the exogenous characteristics of column (3) - (6) in Table 5 - 7 confirm the previous work in local public finance. Column (5) employs the same set of characteristics as in the static model of Case et al (1993). The pupil density has a significant and negative effect on the local school spending, suggesting scale economies in the provision of local public goods. As expected, local average income has a significant and positive effect on the local school spending. The coefficient estimates of income squared are negative in all specifications but not always significant, suggesting somewhat nonlinear resources effects. The coefficient estimates of minority percentage are insignificant in column (3) and (4) that ignore the help of state funds; and they are significantly positive in column (5) and (6) that take into account the effects of state grants. This indicates the relevance of state subsidies in helping poor school districts. As expected, state grants and regional economic situation have significantly positive effects on local school spending in column (5) and (6).

The GS2SLS estimates of structure parameters $(\lambda, \rho, \delta)$ are of central interest in this paper. In Table 5, the coefficient estimates of SAR lag term range from 0.50 in column (6) to 0.69 in column (3) and are statistically significant in all specifications, suggesting the importance of strategic interaction among local school spending. The coefficient estimates of TAR lag term have significantly positive values, changing between 0.18 in column (6) to 0.35 in column (3). The most important finding is that coefficient estimates of STAR lag terms are statistically significant and negatively correlated with
the dependent variable. Their values change from -0.10 with standard error 0.04 in column (5) to -0.14 with standard error 0.05 in column (3) and thus reject the null hypothesis of sign test ($\delta_1 \geq 0$) at the conventional 5% level. The SAR and STAR lag terms in Table 5 are also the first-order SAR and STAR lag terms in Table 6 and 7 that add a secondary spatial autoregressive term. There is little change in empirical results of the first-order SAR, TAR and STAR lag terms between Table 5 and the others. In both Table 6 and 7, the second-order SAR lag terms are positive and STAR lag terms are negative. None of these lags terms are statistically significant because of large standard errors in data construction.

The effects of strategic interaction are mainly captured by parameters $\lambda_1$ and $\delta_1$, coefficients of first-order SAR lag and STAR lag term. The presence of strategic interaction changes the impact magnitudes of exogenous variables of column (2) as well as their interpretation. The ultimate effect of a change in a RHS exogenous variable on local school spending can be understood in the reduced form of equation (46). For exogenous variables, they not only have a direct impact on their own current expenditures, but an indirect impact on current expenditures of their neighbors and on future expenditures of their own and neighbors in a dynamic game-theoretical context.

Table 9 presents the estimation results on local school spending that include statewide time effects. For comparison, column (1) lists the estimation results without time effects, which is same as column (6) of Table 8. Column (2) uses time trend effects and column (3) uses time dummy effects. The most surprising results is that the magnitude of SAR lag parameter decrease
from 0.50 of column (1) to 0.24 of column (2) to 0.06 of column (3). The latter is statistically insignificant. In general, time dummy make more sense than time trend to capture statewide year-to-year change. Hence, the insignificance of SAR parameter suggests that there exist no strategic interaction. The TAR lag coefficients remain positive and significant. Surprisingly, the STAR lag coefficients are still negative and statistically significant. However, as the strategic assumption has been rejected, the interpretation of STAR lag coefficient signs cannot justify the assumption of forward-looking behaviors. Similar results also hold in models with secondary spatial weights matrices (See Table 10 and 11).

### 2.5.3 Test the nonlinear restriction

The simple sign test on parameters of the first-order STAR lag terms of Table 5-7 rejects the null hypothesis: \( \delta_1 \geq 0 \) at 5% significance level. Important and negative previous neighborhood effects imply that local policymakers are likely to make intertemporal considerations in the game-theoretical context. All the estimates of the second-order STAR lag terms are negative, but fails to reject the null hypothesis because of large standard errors.

A direct test on the nonlinear restriction \( (\lambda_k \rho_k + \delta_k = 0 \text{ for } k = 1 \text{ or } 2) \) provides a measure of the extent to which the local spending is explained by the dynamic game-theoretical framework. Appendix D presents a simple delta method to test the nonlinear restriction and Table 8 gives the results on test statistics. Consider the nonlinear restriction when \( k = 1 \). In column (3) and (4) that ignore the effects of state grants, the test values are ranging from
0.09 in Table 7 to about 0.10 in Table 5 and 6 with standard errors less than 0.05, rejecting the nonlinear restriction marginally at 5% significance level. In column (5) and (6) that include state grants on the RHS of equation (46), the largest absolute value of test statistics is trivially 0.026 with standard error about 0.04, failing to reject the nonlinear restriction at any important level. The test statistics on the second-order nonlinear restriction fail to reject the null because of large standard errors of the second-order SAR and STAR lag terms.

The test results on the signs of STAR lag terms of Table 5-7 and the nonlinear restrictions imply that the local policy making can be well characterized by the dynamic game-theoretical framework. In particular, the local spending path is well approximated by the first-order spatial temporal autoregressive panel data model with a set of exogenous characteristics in column (5) or (6).

However, as shown in Table 9, if time dummy effects are included, the SAR lag coefficients are statistically insignificant, indicating that the nonlinear restriction and further interpretation of negative STAR lag coefficient become more difficult.

2.6 Conclusion

This paper examines the role of strategic interaction and intertemporal considerations in deciding local school spending. I formulate a dynamic game-theoretical framework that allows a strategic policymaker to make his intertemporal optimization. The Euler equation implies that the local spend-
ing decision follows an exact spatial temporal autoregressive path. In par-

In particular, there exists a nonlinear restriction on the structural parameters of theoretical specifications. In positive response to the current neighborhood spending and his own previous spending, a forward-looking local policymaker will react negatively to the previous neighborhood spending as a result of an intertemporal resources constraint. This makes it possible to distinguish the forward-looking decision from the backward-looking decision in a game-

The first important finding is about effects of contemporaneous strategic interaction. In general, strategic interaction has a positive slope in most of recent public finance literature. However, the previous studies have employed a static spatial panel data model and ignored dynamic terms. Thus, their model and estimation results may suffer from serious specification problems. This paper provide a more general model form called as spatial dynamic panel data model. Using the Ohio school district data, I found that the SAR lag coefficients are positive and statistically significant if time effects is excludes, but these coefficients will be insignificant if time dummy effects is included. The latter results contradicts the empirical finding in recent literature.

The second important finding in this paper is that there exists an important and negative previous neighborhood effect on local school spending. If time effects are not included, the estimation results can be well explained by the dynamic game-theoretical framework. But since the SAR lag coefficients are insignificant, the interpretation of negative STAR lag coefficients become difficult.
2.7 Appendices of Chapter 2

A. a GS2SLS estimator

Appendix A summarizes the generalized spatial 2SLS estimation procedure.

For simplicity, I consider the estimation of first-order spatial temporal autoregressive panel data model. The generalization to higher order of spatial lags is straightforward but with considerable notational complications. The first-order STAR panel data model equation at period $t$ is given as follows.

$$G_{nt} = \lambda W \Delta G_{nt} + \rho G_{n,t-1} + \delta W_n G_{n,t-1} + X_{nt}\beta + u_{nt},$$

where $G_{n,t-1}$ and $W_n \Delta G_{n,t-1}$ is weakly exogenous and $X_{nt}$ is strongly exogenous by assumption. The model equation is first differenced as follows.

$$\Delta G_{nt} = \lambda W_n \Delta G_{nt} + \rho \Delta G_{n,t-1} + \delta W_n \Delta G_{n,t-1} + \Delta X_{nt}\beta + \Delta u_{nt}.$$

For the moment, assume that the spatial error coefficient $\psi = 0$. The 2SLS moment conditions can be summarized as follows.

$$E[Q_t' \Delta u_{nt}] = E[Q_t' \Delta \epsilon_{nt}] = 0 \quad t \geq 2$$

where $Q_t$ is the linear instrument-variables matrix for $t$-th period equation, consisting of $[Q_{1s}; Q_{2s}, Q_{3t}]$ ($s \leq t - 2$) as follows.

$$Q_{1s} = P_n [G_{n0}, \ldots, G_{ns}; X_{n1}, \ldots, X_{nT}],$$

$$Q_{2s} = [G_{n0}, \ldots, G_{ns}; W_n G_{n0}, \ldots, W_n G_{ns}],$$

$$Q_{3t} = [X_{n1}, X_{n2}, \ldots, X_{nT}].$$
where \( Q_1s \) is instruments of SAR lag \( W_n \Delta G_{nt} \), \( Q_2s \) is of temporary lag terms \( \Delta G_{n,t-1} \) and \( W_n \Delta G_{n,t-1} \); and \( Q_3t \) is of exogenous variables \( \Delta X_{nt} \). The matrix \( P_n \) is a class of constant square matrices such as \( W_n, W_n^2 \), etc.

Stacking all periods in vector form, the equation can be written as follows.

\[
\Delta G_n = \lambda (I_{T-1} \otimes W_n) \Delta G_n + \rho \Delta G_{n,-1} + \delta (I_{T-1} \otimes W_n) \Delta G_{n,-1} + \Delta X_n \beta + \Delta u_n.
\]

For notational convenience, denote \( \theta = (\lambda, \rho, \delta, \beta)' \), \( y_n = \Delta G_n \) and \( z_n = [(I_{T-1} \otimes W_n) \Delta G_n, \Delta G_{n,-1}, W_n \Delta G_{n,-1}, \Delta X_n] \). Then, the estimation equation has a simple compact form.

\[
y_n = z_n \theta + \Delta u_n.
\]

If \( \psi = 0 \), \( \Delta u_n = \Delta v_n = \Delta \epsilon_n \). The variance matrix of first-order differenced errors is

\[
E[\Delta u_n \Delta u_n'] = E[\Delta \epsilon_n \Delta \epsilon_n'] = \sigma_0^2 (G_{T-1} \otimes I_n).
\]

where \( \sigma_0^2 \) is the variance of random noise \( \epsilon_n \) and \( G_{T-1} \) is a square matrix of dimension \( (T - 1) \) with the form

\[
G_{T-1} = \begin{bmatrix}
2 & -1 & \cdots & 0 \\
-1 & 2 & \cdots \\
\vdots & \vdots & \ddots & \vdots \\
\cdots & \cdots & 2 & -1 \\
0 & \cdots & -1 & 2
\end{bmatrix}.
\]
Stacking all 2SLS moment conditions in a single equation lead to \( E [Q' \Delta u_n] = 0 \), where

\[
Q' = \begin{bmatrix}
Q_2' & 0 & 0 \\
0 & Q_3' & \cdots \\
\vdots & \ddots & \ddots \\
0 & \cdots & Q_T'
\end{bmatrix}.
\]

Premultiplying the compact form above by \( Q' \), the equation can be written as follows.

\[
Q' y_n = Q' z_n \theta + Q' \Delta u_n
\]

The generalized spatial 2SLS (GS2SLS) estimator of \( \theta \) has the following expression.

\[
\hat{\theta} = \left[ z'_n Q \left( Q' (G_{T-1} \otimes I_n) Q \right)^{-1} Q' z_n \right]^{-1} \times \left[ z'_n Q \left( Q' (G_{T-1} \otimes I_n) Q \right)^{-1} Q' y_n \right]
\]

\[
= \theta_0 + \left[ z'_n Q \left( Q' (G_{T-1} \otimes I_n) Q \right)^{-1} Q' z_n \right]^{-1} \times \left[ z'_n Q \left( Q' (G_{T-1} \otimes I_n) Q \right)^{-1} Q' \Delta u_n \right].
\]

The GS2SLS estimator \( \hat{\theta} \) of \( \theta \) has the following asymptotic distribution.

\[
\sqrt{n} \left( \hat{\theta} - \theta_0 \right) \overset{d}{\to} N \left( 0, \sigma_0^2 \lim_{n \to \infty} \left[ \frac{1}{n} z'_n Q \left( Q' (G_{T-1} \otimes I_n) Q \right)^{-1} Q' z_n \right]^{-1} \right).
\]

Now assume that \( \Delta u_{n,t} = \Delta v_{n,t} = \psi M_n \Delta v_{n,t} + \Delta \epsilon_{n,t} \). Given a consistent estimate \( \hat{\psi} \) of spatial error coefficient \( \psi \), the variance of difference error terms \( E [\Delta u_n \Delta u'_n] = \sigma_0^2 G_M \), where

\[
G_M = \left[ I_{T-1} \otimes \left( I_n - \hat{\psi} M_n \right)^{-1} \right] (G_{T-1} \otimes I_n) \left[ I_{T-1} \otimes \left( I_n - \hat{\psi} M'_n \right)^{-1} \right].
\]
Replacing $G_{T-1} \otimes I_n$ by $G_M$ in G2SLS estimator, I get the updated variance matrix of the GS2SLS estimator $\hat{\theta}$ as follows.

$$\text{var} \left( \hat{\theta} \right) = \sigma_0^2 \left[ z_n' Q (Q' G_M Q)^{-1} Q' z_n \right]^{-1}.$$

**B. a MOM estimator**

Appendix B presents a simple MOM estimator of spatial error coefficient $\psi$ in a first-order STAR panel data model.

Note that the fixed effects $\alpha_n$ are unobservable in general. But the differenced residuals of $\Delta u_{n,t}$ can be obtained if parameter vector $\theta$ can be consistently estimated. Assume that the model disturbance follows a first-order spatial autoregressive process as follows.

$$\Delta u_{nt} = \Delta v_{nt} = \psi M_n \Delta v_{nt} + \Delta \epsilon_{nt},$$

where the random noise $\Delta \epsilon_{nt}$ is independent cross-sectionally at fixed period $t$, but correlated between $t$ and $t + 1$. In general, the spatial weights matrix $M_n$ in the error process is the same as from the spatial weights matrix $W_n$ in the regression model.

If the model disturbance does not satisfy the i.i.d. assumption, the conventional Maximum-Likelihood (ML) estimator is inconsistent. Furthermore, the ML method is computationally impossible to deal with higher-order SAR process. The GMM or MOM estimator can be robust against unknown heteroskedasticity and extended to generalized cases without computational difficulty. A simple consistent estimator of $\psi$ is proposed to make use of the moment conditions in every single period.

80
For notational convenience, denote \( y_n = \Delta v_{nt} \) and \( \varepsilon_{n,t} = \Delta \varepsilon_{nt} \). At fixed period \( t \), the equation is written as follows.

\[
y_{nt} = \psi M_n y_{nt} + \varepsilon_{nt} \quad t = 2, 3, \cdots, T.
\]

Note that \( M_n y_{nt} \) is generally correlated with \( \varepsilon_{nt} \) because of the endogeneity. Valid IVs need to be constructed so that they are uncorrelated with \( \varepsilon_{nt} \) but correlated with \( M_n y_{nt} \). There are no exogenous variables in the system. Let \( P_n \) be a class of square constant matrices with either a zero diagonal or, more generally, \( \text{trace} (P_n) = 0 \). Denote \( S_n (\psi) = I_n - \psi M_n \). Lee (2001d) suggests \( P_n S_n (\psi) y_{nt} \) as an instrument of \( M_n y_{nt} \).

With a set of IV conditions, the MOM estimator \( \hat{\psi} \) with optimal weighting matrix \( \Omega^{-1}_n \) can be derived from the following objective function:

\[
\min_{\psi} Q_n (\psi) = g'_n (\psi) \Omega^{-1}_n g_n (\psi).
\]

In particular, with the IV matrix \( P_n = M_n \) in panel data models, the moment functions have the following expressions.

\[
g_n (\psi) = \begin{bmatrix}
y'_{n,2} S'_n (\psi) M_n S_n (\psi) y_{n,2} \\
y'_{n,3} S'_n (\psi) M_n S_n (\psi) y_{n,3} \\
\vdots \\
y'_{n,T} S'_n (\psi) M_n S_n (\psi) y_{n,T}
\end{bmatrix} = \begin{bmatrix}
\varepsilon'_{n,2} M_n \varepsilon_{n,2} \\
\varepsilon'_{n,3} M_n \varepsilon_{n,3} \\
\vdots \\
\varepsilon'_{n,T} M_n \varepsilon_{n,T}
\end{bmatrix},
\]

with the covariance-variance matrix \( \text{var} (g_n (\psi_0)) = \Omega_n \). Given that \( M_n \) has a zero diagonal, the MOM estimator of \( \psi \) has robust property against unknown heteroskedasticity.
Assuming that $\varepsilon_{n,t}$ is i.i.d. for fixed $t$ with variance $2\sigma_0^2$, the variance of moment conditions is

$$\Omega_n = \text{var}(g_n(\psi_0)) = \sigma_0^4 \text{trace}(M_n M_n^*) G_{T-1},$$

where $M_n^* = (M_n + M_n')/2$, a symmetric counterpart of $M_n$, and $G_{T-1}$ is a square matrix of dimension $(T-1)$ defined in Appendix A.

Note that computing the MOM estimate with $G_{T-1}^{-1}$ as the weighting matrix does not require a two-step procedure as in a typical optimal GMM estimation. The optimal MOM estimator $\hat{\psi}$ has the asymptotic distribution.

$$\sqrt{n} \left( \hat{\psi} - \psi_0 \right) \xrightarrow{d} N \left( 0, \frac{\text{trace}^2 \left( M_n^* M_n (I_n - \psi_0 M_n)^{-1} \right)}{\text{trace}(M_n M_n^*)} \iota_{T-1}' G_{T-1}^{-1} \iota_{T-1} \right)$$

where $\iota_{T-1}$ is a $(T-1)$ column vector of ones. Note that the asymptotic variance of $\hat{\psi}$ does not contain the error variance $\sigma_0^2$.

For a higher-order STAR panel data model, the spatial error dependence can follow either a first-order or higher-order SAR process. The above MOM estimator can be easily extended to $p$th-order process by replacing the instrumental matrix $P_n = M_n^1, M_n^2, \ldots, M_n^p$.

C. a Gradient test

Appendix C presents a simple diagnostic test for the presence of spatial dependence in the first-differenced errors of dynamic panel data model. Assume that the model disturbance follows a first-order spatial autoregressive process. The first-order derivative of MOM objective function defined in Appendix B is
\[
\frac{\partial Q_n(\psi)}{\partial \psi} = 2D'_n(\psi) \Omega_n^{-1} g_n(\psi),
\]
where the derivative \(D_n(\psi) = \frac{\partial g_n(\psi)}{\partial \psi}\). Evaluating at the restricted condition \(\hat{\psi} = 0\),

\[
g_n(\hat{\psi}) = \begin{bmatrix} y'_{n,2}M_{n}y_{n,2} \\ y'_{n,3}M_{n}y_{n,3} \\ \vdots \\ y'_{n,T}M_{n}y_{n,T} \end{bmatrix}
\]
and \(\hat{D}_n = -2 \begin{bmatrix} y'_{n,2}M'_{n}M_{n}y_{n,2} \\ y'_{n,3}M'_{n}M_{n}y_{n,3} \\ \vdots \\ y'_{n,T}M'_{n}M_{n}y_{n,T} \end{bmatrix}\).

Applying the GMM Gradient test to the MOM objective function defined in Appendix B, I obtain the test statistic with \(\chi^2(1)\) distribution as follows.

\[
G = g'_n(\hat{\psi}) \hat{\Omega}_n^{-1} \hat{D}_n \left( \hat{D}'_n \hat{\Omega}_n^{-1} \hat{D}_n \right)^{-1} \hat{D}'_n \hat{\Omega}_n^{-1} g'_n(\hat{\psi}) \sim \chi^2(1)
\]

The above gradient test can be easily extended to higher-order case by replacing the instrumental matrix \(P_n = M_{n,1}, M_{n,2}, \ldots, M_{n,p}\). The test statistics has \(p\) degrees of freedom for SAR\((p)\) process.

**D. a Delta method**

Appendix D presents a simple test on nonlinear restriction equation \(\delta + \lambda \rho = 0\). Let \(\pi = (\lambda, \rho, \delta)'\) and \(g(\pi) = \delta + \lambda \rho\). At null, \(g(\pi_0) = 0\). Appendix A provides a consistent GS2SLS estimator \(\hat{\pi}\) of \(\pi\). By delta method, I obtain

\[
\sqrt{n} [g(\hat{\pi}) - g(\pi_0)] \xrightarrow{d} N(0, J_0 \Omega J_0'),
\]
where the score vector \(J_0 = \partial g(\pi_0)/\partial \pi' = (\rho_0, \pi_0, 1)\) and the covariance matrix \(\Omega = var(\hat{\pi})\).
CHAPTER 3
Test, Identification and Estimation of a Simultaneous System of Spatially Interrelated Cross-Sectional Equations

3.1 Introduction

Spatial econometrics have recently attracted much interest in both theoretical and empirical econometric literature. Spatial autoregressive process, introduced by Cliff and Ord (1973, 1981), is the most widely used in the single equation model. The spatial autoregressive model can be estimated by the method of maximum likelihood or method of moments. In recent development, Kelejian and Prucha (1999) introduce a generalized moments (GM) method and Lee (2001c) proposes a generalized method of moments (GMM) framework to estimate the spatial autoregressive coefficients. When exogenous variables are relevant, Kelejian and Prucha (1998) and Lee (2001d) provide a consistent and computationally simple spatial two-stage least-square (S2SLS) estimator for estimation of spatial autoregressive, mixed regressive model. Lee (2003) also shows that there exists a best spatial 2SLS estimator that is asymptotically optimal instrumental-variable (IV) estimator.

Kelejian and Prucha (2004) extend their 2SLS estimation method of a single equation into a simultaneous system of spatially interrelated cross sectional equations. Assuming the equations or the system can be identified, they introduce both two- and three-stage least-square estimators for
the model parameters. In this paper, we consider the order and rank conditions for identifying the structural model parameters. In particular, if the spatial vector autoregressive process (SVAR) can be identified, the sufficient rank condition for model identification is equivalent to the consistent estimation of model parameters in linear GMM framework. We will show that the best 2SLS and 3SLS estimators are asymptotically optimal among the linear IV estimators.

The model is introduced and interpreted in Section 3.2. The identification conditions and estimation methods of its SVAR process are presented in Section 3.3. The identification of structure in terms of linear restrictions or exclusion constraints are discussed in Section 3.4. The limited and full information estimators are presented and their asymptotic properties are given in Section 3.5. Hausman specification test is discussed in Section 3.6. The identification and estimation results are extended to generalized spatial models without additional complexity in Section 3.7. Conclusions and suggestions for future work are made in Section 3.8. Some useful lemmas and proofs of selected theorems are collected in Appendix.

### 3.2 The Model

The spatial simultaneous equation model is an extension of the spatial single equation model originated from Whittle (1954), Mead (1967), Cliff and Ord (1973) and Ord (1975) to the simultaneous system of equations. In particular, we consider the following system of spatially interrelated cross sectional equations corresponding to \( n \) cross sectional units:
\[ Y_n B + W_n Y_n A + X_n \Gamma = U_n, \quad (53) \]

with

\[ Y_n = (y_{1n}, \ldots, y_{mn}) \quad X_n = (x_{1n}, \ldots, x_{kn}) \quad U_n = (u_{1n}, \ldots, u_{mn}) \]

where

- \( Y_n \) is a \( n \times m \) matrix of observed endogenous variables;
- \( X_n \) is a \( n \times k \) matrix of observed exogenous variables;
- \( U_n \) is a \( n \times m \) matrix of unobserved disturbances;
- \( y_{jn} \) is a \( n \times 1 \) vector of dependent variable in the \( j \)th equation;
- \( x_{ln} \) is a \( n \times 1 \) vector of the \( l \)th exogenous variable;
- \( u_{jn} \) is a \( n \times 1 \) vector of disturbance in the \( j \)th equation;
- \( W_n \) is a \( n \times n \) weights matrix of known constants,
- \( B \) and \( A \) are \( m \times m \) matrices of coefficients; and
- \( \Gamma \) is a \( k \times m \) matrix of coefficients.

We make the following assumptions on the model.

**Assumption 1:** The spatial weights matrix \( W_n \) has a zero diagonal.

**Assumption 2:** The matrix \( B \) is nonsingular.

Postmultiplying the matrix \( B \), we obtain the spatial vector autoregressive process (SVAR) as follows.

\[ Y_n = W_n Y_n A + X_n \Pi + V_n, \quad (54) \]

with
\[ \Lambda \equiv -AB^{-1}, \quad \Pi \equiv -\Gamma B^{-1}, \quad V_n = U_nB^{-1}. \]

We now express the structural model (and its SVAR process) in a form that will reveal its reduced form for the endogenous variables conveniently. Let

\[ y_n = \text{vec} (Y_n), \quad x_n = \text{vec} (X_n), \quad u_n = \text{vec} (U_n), \quad v_n = \text{vec} (V_n). \]

It follows that the structural model is

\[ (B' \otimes I_n) y_n + (A' \otimes W_n) y_n + (\Pi' \otimes I_n) x_n = u_n, \]  
(55)

and its SVAR form is

\[ y_n = (\Lambda' \otimes W_n) y_n + (\Pi' \otimes I_n) x_n + v_n. \]  
(56)

Remark: The parameter matrices \((\Lambda, \Pi)\) is identified in the SVAR process if the parameters matrices \((B, A, \Gamma)\) is identified in the structural model.

**Assumption 3:** The matrix \(I_{mn} - \Lambda' \otimes W_n\) is nonsingular at the true parameter matrix \(\Lambda_0\).

Denote \(S_n (\Lambda) = I_{mn} - \Lambda' \otimes W_n\) and \(S_n = S_n (\Lambda_0)\). It follows that \(S_n (\Lambda) = S_n + (\Lambda_0' - \Lambda') \otimes W_n\). Under assumption 3, we can get the reduced-form of the structural model as follows.

\[ y_n = S_n^{-1} (\Pi'_0 \otimes I_n) x_n + S_n^{-1} v_n. \]  
(57)
A sufficient condition for $S_n$ to be invertible is $\| \Lambda_0 \otimes W_n \| < 1$ for any matrix norm $\| \cdot \|$ (Horn and Johnson 1985, p301). If the elements of $W_n$ are nonnegative and row normalized, a sufficient condition is that $\| \Lambda_0 \|_\infty < 1$ where $\| \cdot \|_\infty$ is maximum row sum matrix norm by Lemma A1. However, such condition often fails in practice. In particular, if endogenous regressor $y_{jn}$ changes its scalar measure in $i$th equation, the nondiagonal elements of $(i,j)$ and $(j,i)$ entries can be either much larger or smaller than one in absolute values. Let $T$ be a collection of the $m$ eigenvectors of matrix $\Lambda_0$. Assume that the square matrix $T$ is invertible, the matrix $\Lambda_0$ can be decomposed as: $\Lambda_0 = TDT^{-1}$ where $D$ is a diagonal matrix consisting of $m$ distinct eigenvalues of matrix $\Lambda_0$. The decomposition requires the matrix $\Lambda_0$ to have $m$ linearly independent eigenvectors. This will be true when $\Lambda_0$ has $n$ distinct eigenvalues, and would still be true even if $\Lambda_0$ has some repeated eigenvalues.

We can rewrite the matrix $S_n$ as follows.

\[
S_n = I_{mn} - T^{-1}DT' \otimes W_n \\
= (T^{-1} \otimes I_n)(I_{mn} - D \otimes W_n)(T' \otimes I_n) .
\] (58)

It is clear that the matrix $S_n$ is invertible if and only if $I_{mn} - D \otimes W_n$ is invertible. It will be sufficient for $I_{mn} - D \otimes W_n$ and $S_n$ to be invertible if $\| D \otimes W_n \| < 1$ for any matrix norm. For the row-normalized matrix $W_n$, the sufficient condition is equivalent to that all of eigenvalues of matrix $\Lambda_0$ shall lie inside the interval $(-1, 1)$.

**Assumption 4:** The row and column sum of the matrices $W_n$ and $I_{mn} -$
\( \Lambda_0' \otimes W_n \) are uniformly bounded in absolute value.

If all of eigenvalues of matrix \( \Lambda_0 \) lie inside the interval \((-1, 1)\), \( S_n \) can be inverted into an exponentially converging power series of \( W_n \) with matrix coefficients \( \Lambda_0 \).

\[
S_n^{-1} = I_{mn} + \Lambda_0' \otimes W_n + \Lambda_0'^2 \otimes W_n^2 + \cdots
\]

\[
= (T^{-1} \otimes I_n) \left[ I_{mn} + D \otimes W_n + D^2 \otimes W_n^2 + \cdots \right] (T' \otimes I_n). \tag{59}
\]

It follows that \( S_n^{-1} \) will be uniformly bounded in row sums for row-normalized matrix \( W_n \).

**Assumption 5:** The matrix of exogenous (and nonstochastic) variables \( X_n \) has full column rank. Furthermore, the elements of \( X_n \) are uniformly bounded in absolute value.

**Assumption 6:** The disturbances \( \{u_n\} \) are independently and identically distributed with \( E(\mathbb{u}_n) = 0 \) and \( \text{var}(\mathbb{u}_n) = \Sigma \otimes I_n \). Furthermore, their moments of order higher than the fourth exist.

The covariance matrix of SVAR disturbances \( \text{var}(v_n) = B^{-1} \Sigma B^{-1} \otimes I_n = \Omega \otimes I_n \).

Conditional on exogenous variables, the expectation of dependent variable can be written as follows.

\[
E(\mathbb{y}_n) = (\Pi_0' \otimes I_n) x_n + (\Lambda_0' \Pi_0' \otimes W_n) x_n + (\Lambda_0'^2 \Pi_0' \otimes W_n^2) x_n + \cdots,
\]

\[
E(\mathbb{y}_n) = X_n \Pi_0 + W_n X_n \Pi_0 \Lambda_0 + W_n^2 X_n \Pi_0 \Lambda_0^2 + \cdots.
\]
3.3 Identification and Estimation of SVAR process

For the spatial vector autoregression process, the identification is synonymous with the consistent estimation of parameters.

The spatial vector autoregressive process (SVAR) system can be written as follows.

\[ Y_n = (\overline{Y}_n, X_n) \Theta + V_n = Z_n \Theta + V_n. \]  
\[ y_n = [I_m \otimes (\overline{Y}_n, X_n)] \theta + v_n = [I_m \otimes Z_n] \theta + v_n. \]  

(60)

(61)

where \( \overline{Y}_n = W_n Y_n \), \( \Theta = (\Lambda', \Pi')' \), and \( \theta = vec(\Theta) \).

Denote \( v_n(\theta) = S_n(\Lambda) y_n - (\Pi' \otimes I_n) x_n = y_n - [I_m \otimes (\overline{Y}_n, X_n)] \theta \) for any possible value \( \theta \). The expectation of \( v_n(\theta) \) is

\[ E[v_n(\theta)] = S_n(\Lambda) E(y_n) - (\Pi' \otimes I_n) x_n \]
\[ = [(\Lambda'_0 - \Lambda') \otimes W_n] E(y_n) + [(\Pi'_0 - \Pi') \otimes I_n] x_n \]
\[ = [I_m \otimes (W_n E(Y_n), X_n)] (\theta - \theta_0). \]  

(62)

A 2SLS or linear GMM estimator will be considered for the estimation of the SVAR system. Let \( Q_n \) be an \( n \times q \) matrix of IVs constructed as function of \( X_n \) and \( W_n \) in a 2SLS approach.

Under the orthogonality condition of \( X_n \) and \( u_n \), we can establish the following linear moment conditions for the SVAR system

\[ E[(I_m \otimes Q'_n) v_n(\theta)] = 0. \]
In the GMM framework, the identification conditions require that the moment equations have a unique solution at true parameters $\theta_0$ as $n$ goes to infinity. The linear moment conditions corresponding the $j$th equation are

$$Q_n^t E(v_{jn}(\theta)) = Q_n^t [W_n E(Y_n), X_n] (\theta_j - \theta_{j0}) = 0,$$

(63)

where $\theta_j$ (and $\theta_{j0}$) is the $j$th column of parameter matrix $\Theta$ (and $\Theta_0$). They will have a unique solution at $\theta_0$ if $Q_n^t [W_n E(Y_n), X_n]$ has a full column rank, i.e., rank $(m + k)$. This sufficient rank identification condition implies the necessary rank condition that $[W_n E(Y_n), X_n]$ has a full column rank $(m + k)$. This rank condition will not hold, in particular, if $\Pi_0$ happens to be null or $E(Y_n)$ has a zero column. In other words, it is necessary for any of endogenous regressors $Y_n$ to have relevant exogenous variables directly or indirectly.

In addition, the 2SLS IV matrix $Q_n$ needs to have a rank at least as large as $(m + k)$ in order to have a sufficient rank condition. This requires that the number of linear independent IVs $q$ is at least equal to $(m + k)$, which is known as the necessary order condition. In addition to exogenous variables $X_n$, we need additional $m$ IVs for endogenous variables $Y_n$. The order condition can be satisfied if the number of exogenous variables $k$ is no less than the number of endogenous variables $m$ and $(W_n X_n, X_n)$ are linearly dependent. If the number of exogenous variables is less than the number of endogenous variables, then we can consider $(W_n X_n, W_n^2 X_n, \cdots, X_n)$, which has at least $m + k$ linearly independent columns. Although such linear inde-
dependence of \( (W_n, X_n, W_n^2 X_n, \cdots, X_n) \) can hold for most empirical weighting matrices, this is not the case in general. In particular, a weights matrix such that \( W_n^2 = c_0 I_n + c_1 W_n \) \((c_0 \text{ and } c_1 \text{ are nonzero constants}) \) can not provide enough independent IVs for \( \Sigma_n \) if the number of exogenous variables \( k \) is less than the number of endogenous variables \( m \). Consider the block-diagonal matrix \( W_{nr} = I_n \otimes \frac{1}{r-1} (I_r l_r' - I_r) \), which is commonly used in social interaction literature (Case, 1991, 1992) when interrelated individuals are from the same region or group. With this \( W_{nr} \),

\[
W_{nr}^2 = \left( \frac{r-2}{r-1} \right) W_{nr} + \frac{1}{(r-1)} I_{nr}.
\]

It follows that high-order \( W_{nr}^p \) including \( S_n \) is a linear combination of \( I_n \) and \( W_n \) for \( p \geq 2 \). There are at most \( m + k \) linear independent columns in IV matrix \( Q_n \). In order to identify the parameters, at least \( m - k \) a priori restrictions are required for a single equation to satisfy the necessary order conditions. If the structural system is fully identified, then the SVAR process is also fully identified, which implies there exists enough a priori restrictions. However, if the structural system is partially identified, then the SVAR process will not be full identified and it is possible that no single equation in the SVAR process is identified. For example, the first equation is identified and the rest is not in the structure, then it is possible for all the equation in the SVAR process remains unidentified. In other words, the system does not have sufficient IVs because the number of exogenous variable is less than the number of endogenous variables and the spatial weights matrix has special properties (that is, \( W_n^2 = c_0 I_n + c_1 W_n \)). The identification condition for the
structure with insufficient IVs will be discussed in Section 6.1.

We can summarize the identification conditions for unrestricted linear GMM or 2SLS estimation method as follow.

**Proposition 3.1** (order condition of the SVAR process)

If \( \theta_0 \) can be identified, then the number of instrumental-variables \( q \) is at least as large as the number of unknown parameters \( m + k \).

**Proposition 3.2** (rank condition of the SVAR process)

If \( Q'_n [W_n E(Y_n), X_n] \) has full column rank, then \( \theta_0 \) is identified.

Since moment condition are linear in parameters \( \theta_0 \), the rank condition is sufficient and necessary for global identification of linear GMM estimation framework.

Note that \( \text{rank} (I_m \otimes A) = m \times \text{rank} (A) \) for any matrix \( A \). Hence, the order and rank conditions for the single equation also hold for the SVAR system.

Denote \( Q^*_n = [W_n E(Y_n), X_n] \), the \( n \times (m + k) \) nonstochastic matrix, which belongs to a class of instrument matrix \( Q_n \) if \( \Theta_0 \) is known.

**Assumption 7**: The nonstochastic matrix \( Q_n \) (and \( Q^*_n \)) has full column rank. The elements of \( Q_n \) (and \( Q^*_n \)) are uniformly bounded in absolute value. Furthermore \( Q_n \) (and \( Q^*_n \)) has the following limiting properties:

(7.1) \( \lim_{n \to \infty} \frac{1}{n} Q'_n Q_n \) is a finite nonsingular matrix;

(7.2) \( \lim_{n \to \infty} \frac{1}{n} Q'_n Q^*_n \) is a finite matrix with full column rank.

Denote the projection matrix \( P_n = Q_n (Q'_n Q_n)^{-1} Q'_n \) (and \( P^*_n = Q^*_n (Q^*_n Q^*_n)^{-1} Q^*_n \)).

Then for the \( j \)th equation of SVAR process, the 2SLS estimator of \( \tilde{\theta}_{j,n} \) is given
by

\[ \tilde{\theta}_{j,n} = \left( \tilde{Z}_n^P Z_n \right)^{-1} \tilde{Z}_n^P y_{jn}. \]  

(64)

The 2SLS residuals are given by

\[ \tilde{\nu}_{jn} = y_n - Z_n \tilde{\theta}_{j,n}. \]  

(65)

**Theorem 3.1:** Suppose Assumptions 1-7 hold. Then the linear GMM estimator \( \tilde{\theta}_n \) is a consistent estimator of \( \theta_0 \) and \( \sqrt{n} \left( \tilde{\theta}_{j,n} - \theta_{j0} \right) \xrightarrow{d} N(0, \Omega_j) \) as \( n \to \infty \), where

\[ \Omega_j = \sigma^2_{\nu,jj} \left[ \lim_{n \to \infty} \frac{1}{n} Q_n^* P_n Q_n^* \right]^{-1}. \]  

(66)

The variance \( \sigma^2_{\nu,jj} \) is the \( j \)th diagonal element of covariance matrix \( \Omega \).

The following lemma establishes that \( \tilde{\Omega}_n \) is a consistent estimator for \( \Omega \).

**Lemma 3.1** Suppose Assumptions 1-7 hold. Then \( \text{plim}_{n \to \infty} \tilde{\Omega}_n = \Omega \).

### 3.4 Identification of the structure system

In this section, we derive the rank and order conditions for the identification of structural parameters when its SVAR process can be fully identified and consistently estimated.

In the classical linear system, the numbers of endogenous regressors and equations are in general equal. The number of unknown parameters in \( B \) and \( \Gamma \) is \( m^2 + km \); while Assumption 2, 5 and 6 are sufficient to identify \( \Pi \) in the reduced form. We therefore treat the elements of \( \Pi \) as known and derive the sufficient and necessary rank condition that can identify some or all of the elements of \( B \) and \( \Gamma \).
The spatial system has \( m \) equations, but up to \( 2m \) endogenous variables. We can approach the identification problem of the spatial system of equations similarly by considering its SVAR process instead of reduced form of the structural model. Suppose that the parameter matrices \( \Lambda \) and \( \Pi \) in the SVAR form can be identified and consistently estimated. The number of unknown parameters in \( B, A \) and \( \Gamma \) is \( 2m^2 + km \); and the number of known parameters in \( \Lambda \) and \( \Pi \) is \( m^2 + km \).

### 3.4.1 Linear Restrictions

Without a prior restriction on the structural parameters, the system remains unidentified since the number of unknown parameters is larger than the number of linear restrictions.

Without loss of generality, consider the identification of the first structural equation.

\[
Y_n\beta_{1.1} + \bar{V}_n\alpha_{.1} + X_n\gamma_{.1} = u_{1n}, \quad (67)
\]

where \( \beta_{1.1} \) and \( \alpha_{.1} \) are an \( m \)-dimensional column vector and \( \gamma_{.1} \) is a \( k \)-dimensional column vector. We can establish the following identities between structural parameters and SVAR parameters.

\[
\begin{align*}
\Lambda & \equiv -AB^{-1}, \quad \Lambda B + A = 0, \\
\Pi & \equiv -\Gamma B^{-1}, \quad \Pi B + \Gamma = 0.
\end{align*}
\]

The identities is equivalent to
\[ \begin{bmatrix} \Lambda & I_m & 0 \\ \Pi & 0 & I_k \end{bmatrix} \begin{bmatrix} B \\ A \\ \Gamma \end{bmatrix} = 0, \]  

(68)

which implies that

\[ \begin{bmatrix} \Lambda & I_m & 0 \\ \Pi & 0 & I_k \end{bmatrix} \begin{bmatrix} \beta_{1.1} \\ \alpha_{1.1} \\ \gamma_{1.1} \end{bmatrix} = 0. \]  

(69)

Suppose that there are a prior linear restrictions on the first structural parameters such that

\[ \Phi \begin{bmatrix} \beta_{1.1} \\ \alpha_{1.1} \\ \gamma_{1.1} \end{bmatrix} = 0, \]  

(70)

where \( \Phi \) is a \( r \times (2m + k) \) matrix with full rank \( r \) on the corresponding equation.

Stacking the above linear restrictions together, we have that

\[ \begin{bmatrix} \Lambda & I_m & 0 \\ \Pi & 0 & I_k \\ \Phi \end{bmatrix} \begin{bmatrix} \beta_{1.1} \\ \alpha_{1.1} \\ \gamma_{1.1} \end{bmatrix} = 0. \]  

(71)

With the normalization rule \( \beta_{11} = 1 \), this equation has a unique solution if and only if
\[
\begin{bmatrix}
\Lambda & I_m & 0 \\
\Pi & 0 & I_k \\
\Phi & & \\
\end{bmatrix}
\]
\[
\text{rank } = 2m + k - 1,
\]
(72)

which is known as the rank condition.

A necessary order condition for the existence of unique solution is

\[
m + k + r \geq 2m + k - 1,
\]
\[
r \geq m - 1.
\]
(73)

**Proposition 4.1** (order condition of the structure when its SVAR process is fully identified):

If the \( j \)th equation is identified, then the number of a prior restrictions should be at least as larger as the number of equations in the model less one.

The rank condition can be rewritten into a more appealing condition.

**Proposition 4.2** (rank condition of the structure when its SVAR process is fully identified):

The \( j \)th equation is identified if and only if

\[
\text{rank } \Phi \begin{pmatrix} B \\ A \\ \Gamma \end{pmatrix} = m - 1.
\]
(74)

The rank condition is necessary and sufficient for the identification of the structural model in linear GMM framework.
3.4.2 Exclusion

Suppose that the a priori restrictions are solely exclusion restrictions (i.e., omitted variables) in the first equation.

\[
y_{1n} = Y_{1n}\beta_1 + Y_{1n}^*\beta_1^* + Y_{1n}\alpha_1 + Y_{1n}^*\alpha_1^* + X_{1n}\gamma_1 + X_{1n}^*\gamma_1^* + u_{1n} \tag{75}
\]

where \(Y_{1n}\) is an \(n \times m_1\) matrix, \(\beta_1\) is a \(m_1\) column vector, \(Y_{1n}^*\) is an \(n \times m_1\) matrix, \(\alpha_1\) is an \(m_1\) column vector, \(X_{1n}\) is a \(n \times k_1\) matrix, \(\gamma_1\) is a \(k_1\) column vector, \(\beta_1^* = 0\), \(\alpha_1^* = 0\) and \(\gamma_1^* = 0\). Thus, in \(Y_{1n}\beta_1 + Y_{1n}^*\alpha_1 + X_{1n}\gamma_1 = u_{1n}\),

\[
\beta_1 = [1, -\beta_1', 0]', \quad \alpha_1 = [-\alpha_1', 0]', \quad \gamma_1 = [-\gamma_1', 0]'
\]

Write the SVAR process \(Y_n = Y_n\Lambda + X_n\Pi + V_n\) as

\[
[y_{1n}, Y_{1n}, Y_{1n}^*] = \left[ Y_{1n}, Y_{1n}^* \right] \begin{pmatrix} \lambda_1 & \Lambda_1 & \overline{\Lambda}_1 \\ \lambda_1^* & \Lambda_1^* & \overline{\Lambda}_1^* \end{pmatrix} + [X_{1n}, X_{1n}^*] \begin{pmatrix} \pi_1 & \Pi_1 & \bar{\Pi}_1 \\ \pi_1^* & \Pi_1^* & \bar{\Pi}_1^* \end{pmatrix} + [u_{1n}, V_{1n}, V_{1n}^*]. \tag{76}
\]

Since \(\Lambda\beta_1 + \alpha_1 = 0\) and \(\Pi\beta_1 + \gamma_1 = 0\), partition \(\Lambda\) and \(\Pi\) conformably leads to
\[
\begin{pmatrix}
\lambda_1 & \Delta_1 & \Lambda_1 \\
\lambda_1^* & \Delta_1^* & \Lambda_1^*
\end{pmatrix}
\begin{pmatrix}
1 \\
-\beta_1 \\
0
\end{pmatrix}
= 
\begin{pmatrix}
\alpha_1 \\
0
\end{pmatrix},
\]
\[
\begin{pmatrix}
\pi_1 & \Pi_1 & \Pi_1 \\
\pi_1^* & \Pi_1^* & \Pi_1^*
\end{pmatrix}
\begin{pmatrix}
1 \\
-\beta_1 \\
0
\end{pmatrix}
= 
\begin{pmatrix}
\gamma_1 \\
0
\end{pmatrix}.
\]

This implies that
\[
\lambda_1 - \Delta_1 \beta_1 = \alpha_1, \quad \lambda_1^* - \Delta_1^* \beta_1 = 0;
\]
\[
\pi_1 - \Pi_1 \beta_1 = \gamma_1, \quad \pi_1^* - \Pi_1^* \beta_1 = 0.
\]

So the problem is reduced to solve \(\beta_1\) uniquely from
\[
[\Lambda_1^*, \Pi_1^*] \beta_1 = [\lambda_1^*, \pi_1^*]. \tag{77}
\]

Let the number of excluded exogenous variables in the first equation be \(k_1^*\), the number of excluded spatially lagged endogenous variables be \(m_1^*\), and the number of excluded and included endogenous variables on the right hand (i.e., dimension of \(Y_{1n}^*\)) be \(m_1^*\). Then, \(r = m_1^* + m_1 + k_1^* = (m - (m_1 + 1)) + (m - m_1) + (k - k_1)\), where \(m_1\), \(m_1\) and \(k_1\) are the number of included endogenous variables, spatially lagged endogenous variables and exogenous variables respectively. It follows that \([\Lambda_1^*, \Pi_1^*]\) is a \((m_1^* + k_1^*) \times m_1\) matrix.
The necessary order condition can be rewritten as follows.

\[ r \geq m - 1 \]
\[ \iff k_1^* + m_1^* + m_1^* \geq m - 1 \]
\[ \iff k_1^* + m_1^* \geq m \]
\[ \iff k_1^* \geq m_1 + m_1 - m. \] (78)

The sufficient and necessary rank conditions is

\[ \text{rank} \left[ \Lambda_1^*, \Pi_1^* \right] = m_1. \] (79)

Partition the structural matrix into

\[
\begin{pmatrix}
B \\
A \\
\Gamma
\end{pmatrix}
= \begin{pmatrix}
1 & A_1 \\
-\beta_1 & A_2 \\
0 & A_3 \\
-\alpha_1 & A_4 \\
0 & A_5 \\
-\gamma_1 & A_6 \\
0 & A_7
\end{pmatrix},
\]

which gives an equivalent rank condition as

\[ \text{rank} \begin{pmatrix}
A_3 \\
A_5 \\
A_7
\end{pmatrix} = m - 1. \] (80)

3.5 Estimation of the structure system
In this section, we define the spatial 2SLS and 3SLS estimators for the structural model. Among the class of linear GMM estimators, best IV estimators are derived.

Without loss of generality, consider the estimation of the first equation with exclusion variables constraints only:

\[
y_{1n} = Y_{1n} \beta_1 + X_{1n} \gamma_1 + u_{1n} = Z_{1n} \delta_1 + u_{1n}, \quad (81)
\]

where \( Z_{1n} = (Y_{1n}, \bar{Y}_{1n}, X_{1n}) \) and \( \delta_1 = (\beta_1', \alpha_1', \gamma_1')' \).

Assume that its spatial vector autoregressive process (SVAR) is identified, the SVAR parameters can be consistently estimated as follows.

\[
Y_n = \bar{Y}_n \tilde{\Lambda} + X_n \tilde{\Pi} + \tilde{V}_n = Z_n \tilde{\Theta} + \tilde{V}_n \quad (82)
\]

Let \( J_1 \) be the selection matrix such that \((\bar{Y}_{1n}, X_{1n}) = (\bar{Y}_n, X_n) J_1 \). It follows that \([W_n E(Y_{1n}), X_{1n}] = (W_n E(Y_n), X_n) J_1 = Q_* J_1 \).

**Lemma 5.1:** Suppose Assumptions 1-7 hold. Then \( \lim_{n \to \infty} \frac{1}{n} Q_* E(Z_{jn}) \) is a finite nonsingular matrix, if the rank condition is satisfied for the \( j \)th structural equation.

The above lemma is equivalent to the Assumption 6(b) used by Kelejian and Prucha (2004) to identify the model equation. Our Assumption 7 is identical to their Assumption 6(a). As shown in Appendix, our Assumption 7 and rank condition will imply the nonsingularity of \( \lim_{n \to \infty} \frac{1}{n} Q_* E(Z_{jn}) \), which guarantee the consistency of the linear GMM estimators. Lemma 5.1 also implies that \( E(Z_{jn}) \) has full column rank if the rank condition holds.
Hence, the 2SLS estimator with IV matrix \( P_n^* = Q_n^* (Q_n^* Q_n^*)^{-1} Q_n^* \) can be formulated for the \( j \)th structural equation as follows.

\[
\hat{\delta}_{jn}^* - \delta_{jn} = (Z_{jn}' P_n^* Z_{jn})^{-1} Z_{jn}' P_n^* u_{jn}.
\] (83)

In general, if the \( j \)th equation satisfy the rank condition, a generic linear IV estimator can be formulated with idempotent matrix \( P_n = Q_n (Q_n' Q_n)^{-1} Q_n' \).

\[
\hat{\delta}_{jn} - \delta_{jn} = [Z_{jn}' P_n Z_{jn}]^{-1} Z_{jn}' P_n u_{jn}.
\] (84)

**Assumption 7.3:** \( \lim_{n \to \infty} \frac{1}{n} Q_n' E(Z_{jn}) \) exists and has full column rank if the rank condition is satisfied.

If above assumption holds, the IV estimator with \( P_n \) can be consistent. For the class of linear GMM estimators, the following theorems will establish their consistency and asymptotic normality, and derive the best one among them.

**Theorem 5.1:** Suppose Assumptions 1-7 and rank condition hold for the \( j \)th equation. Then, a linear GMM estimator \( \hat{\delta}_{jn} \) with \( P_n \) is a consistent estimator of \( \delta_j \) and \( \sqrt{n} \left( \hat{\delta}_{jn} - \delta_{j0} \right) \xrightarrow{d} N(0, \Sigma_{jn}) \) as \( n \to \infty \), where

\[
\Sigma_{jn} = \sigma_{u,jj}^2 \left[ \text{plim}_{n \to \infty} \frac{1}{n} Z_{jn}' P_n Z_{jn} \right]^{-1}
= \sigma_{u,jj}^2 \left[ \lim_{n \to \infty} \frac{1}{n} E(Z_{jn}') P_n E(Z_{jn}) \right]^{-1}.
\] (85)

The variance \( \sigma_{u,jj}^2 \) is the \( j \)th diagonal element of covariance matrix \( \Sigma \).

**Theorem 5.2:** Suppose Assumptions 1-7 and rank condition hold for the \( j \)th equation. Then, the 2SLS estimator \( \hat{\delta}_{jn}^* \) with \( P_n^* \) is a consistent estimator
of \( \delta_j \) and \( \sqrt{n} \left( \hat{\delta}_{jn} - \delta_{j0} \right) \overset{d}{\longrightarrow} N \left( 0, \Sigma_j^* \right) \) as \( n \to \infty \), where

\[
\Sigma_j^* = \sigma^2_{u,jj} \left[ \lim_{n \to \infty} \frac{1}{n} E \left( Z_{jn}' \right) P_n^* E \left( Z_{jn} \right) \right]^{-1}.
\] (86)

Furthermore, the 2SLS estimator \( \hat{\delta}_{jn}^* \) has minimum variance among the class of linear GMM estimator \( \hat{\delta}_{jn} \), which is given by the generalized Schwartz inequality.

Denote IV residuals \( \hat{u}_{jn} = y_{jn} - Z_{jn} \hat{\delta}_{jn} \), where \( \hat{\delta}_{jn} \) is a consistent estimator.

The following lemma establishes that \( \hat{\sigma}_{u,jj}^2 \) (and \( \hat{\Sigma} \)) is a consistent estimator for \( \sigma_{u,jj}^2 \) (and \( \Sigma \)).

**Lemma 5.2**: Suppose Assumptions 1-7 hold. Then \( \text{plim}_{n \to \infty} \hat{\sigma}_{u,jj}^2 = \sigma_{u,jj}^2 \) if the \( j \)th equation is identified, and \( \text{plim}_{n \to \infty} \hat{\Sigma} = \Sigma \) if the system is identified.

The spatial 2SLS estimator is an limited-information estimator in that it does not take into account the potential covariance information in the disturbance vector \( u_n \). To utilize the full information in the system, it is useful to stack the equation as follows. Suppose that the only restrictions are exclusion constraints in the system as follows:

\[
y_n = Z_n \delta + u_n, \tag{87}
\]

where \( Z_n = \text{diag}^m_{j=1} (Z_{jn}) \) and \( \delta = (\beta', \alpha', \gamma')' \).

Recall that \( E u_n = 0 \) and \( \text{var} u_n = \Sigma \otimes I_n \), a full information IV estimator \( \hat{\delta}_n \) of \( \delta \) would be

\[
\hat{\delta}_n = \left[ Z_n' P_n \left( \Sigma^{-1} \otimes I_n \right) Z_n \right]^{-1} Z_n' P_n \left( \Sigma^{-1} \otimes I_n \right) y_n. \tag{88}
\]
Theorem 5.3: Suppose Assumptions 1-7 and rank condition hold for the system of equations. Then, a linear GMM estimator $\hat{\delta}_n$ with $P_n$ is a consistent estimator of $\delta_0$ and $\sqrt{n} \left( \hat{\delta}_n - \delta_0 \right) \xrightarrow{d} N \left( 0, \Sigma_n \right)$ as $n \to \infty$, where

$$\Sigma_n = \left[ \lim_{n \to \infty} \frac{1}{n} E \left( Z_n' \right) P_n \left( \Sigma^{-1} \otimes I_n \right) P_n E \left( Z_n \right) \right]^{-1}. \quad (89)$$

A feasible counterpart of an 3SLS estimator is to replace the true unknown covariance matrix $\Sigma$ by its consistent estimate $\hat{\Sigma}$. Furthermore, a generalized Schwartz inequality implies that an 3SLS estimator with IV matrix $P_n^*$ has the minimum variance among the class of linear GMM estimators.

3.6 Hausman Specification Test

In this section, we consider the Hausman specification test of the simultaneous spatial systems. Without loss of generality, consider a single equation.

$$Y_n = \lambda W_n Y_n + \beta X_{1n} + U_n. \quad (90)$$

We maintain the same notation used in above sections for the system. However, since the above system is just a single equation, the parameters are just scalars. Note that $X_{1n}$ might be endogenous. Suppose there exists an exogenous IV $X_{2n}$ that correlates with $X_{1n}$, but not with $U_n$.

Denote that $Q_{1n} = \left[ W_n X_{1n}, X_{1n} \right]$ and $Q_{2n} = \left[ W_n X_{2n}, X_{2n} \right]$ and $Z_n = \left[ W_n Y_n, X_{1n} \right]$ and $Q_{0n} = \left[ W_n E \left( Y_n \right), X_{1n} \right]$.

In general, the Hausman specification will look for an efficient estimators under the null hypothesis. However, in the spatial model, the efficient estimator is nonlinear and has no closed-form expression. Instead, we will try to
give a linear consistent estimator under the null hypothesis. Under the null hypothesis that $X_{1n}$ is exogenous, the suggested 2SLS estimator with linear instrument $P_{1n} = Q_{1n} (Q'_{1n} Q_{1n})^{-1} Q'_{1n}$

$$\hat{\theta}_{1n} = (Q'_{1n} Z_n)^{-1} Q'_{1n} Y_n = (Z'_{n} P_{1n} Z_n)^{-1} Z'_{n} P_{1n} Y_n. \quad (91)$$

However, if the null hypothesis is false, the above IV estimator is inconsistent. The consistent S2SLS estimator with valid instrument $P_{2n} = Q_{2n} (Q'_{2n} Q_{2n})^{-1} Q'_{2n}$

$$\hat{\theta}_{2n} = (Q'_{2n} Z_n)^{-1} Q'_{2n} Y_n = (Z'_{n} P_{2n} Z_n)^{-1} Z'_{n} P_{2n} Y_n. \quad (92)$$

The variance-covariance matrix of $\hat{\theta}_{1n}$ and $\hat{\theta}_{2n}$ is

$$\text{var} \left[ \begin{array}{c}
\hat{\theta}_{1n} \\
\hat{\theta}_{2n}
\end{array} \right] = \sigma^2 \begin{bmatrix}
(Q'_{0n} P_{1n} Q_{0n})^{-1} & (Z'_{n} P_{1n} Z_n)^{-1} Z'_{n} P_{1n} P'_{2n} Z_n (Z'_{n} P_{2n} Z_n)^{-1} \\
* & (Q'_{0n} P_{2n} Q_{0n})^{-1}
\end{bmatrix}, \quad (93)
$$

where the asterik is transpose of the upper-right block matrix.

The Hausman-flavored test is given as

$$\hat{\theta}_{2n} - \hat{\theta}_{1n} = (Z'_{n} P_{2n} Z_n)^{-1} Z'_{n} P_{2n} Y_n - (Z'_{n} P_{1n} Z_n)^{-1} Z'_{n} P_{1n} Y_n$$

$$= (Z'_{n} P_{2n} Z_n)^{-1} Z'_{n} P_{2n} \left[ I_n - Z_n (Z'_{n} P_{1n} Z_n)^{-1} Z'_{n} P_{1n} \right] Y_n$$

$$= (Q'_{2n} Z_n)^{-1} Q'_{2n} \left[ I_n - Z_n (Q'_{1n} Z_n)^{-1} Q'_{1n} \right] Y_n. \quad (94)$$

Note that the suggest Hausman test is a linear function of dependent variable. So it is not difficult to obtain its variance term and then to form the test statistics.
3.7 Extension

In this section, we extend the identification conditions and estimation methods into generalized spatial models without sufficient instrumental variables, with higher order spatial lags and with spatial error dependence.

3.7.1 System with insufficient IVs

As discussed in Section 3, the system and its SVAR process does not have sufficient instrumental-variables, because the number of exogenous variables is less than the number of endogenous variables and the spatial weights matrix has special properties such that $W_n^2 = c_0 I_n + c_1 W_n$, where $c_0$ and $c_1$ are nonzero constants. In general, $c_0$ and $c_1$ are known if $W_n$ is given.

The reduced-form of the structural model is

$$y_n = (I_{mn} - \Lambda' \otimes W_n)^{-1} (\Pi' \otimes I_n)x_n + S_n^{-1}v_n$$

We can establish the following equalities.

$$\Pi_0 - c_0 \Pi_1 \Lambda - \Pi = 0,$$

$$\Pi_1 - c_1 \Pi_1 \Lambda - \Pi_0 \Lambda = 0.$$
\[ \Pi_0 B + c_0 \Pi_1 A + \Gamma = 0 \]
\[ \Pi_1 B + (\Pi_0 + c_1 \Pi_1) A = 0 \]

which is equivalent to

\[
\begin{bmatrix}
\Pi_0 & c_0 \Pi_1 & I_k \\
\Pi_1 & \Pi_0 + c_1 \Pi_1 & 0
\end{bmatrix}
\begin{pmatrix}
B \\
A \\
\Gamma
\end{pmatrix} = 0
\]

For the first structural equation, this implies that

\[
\begin{bmatrix}
\Pi_0 & c_0 \Pi_1 & I_k \\
\Pi_1 & \Pi_0 + c_1 \Pi_1 & 0
\end{bmatrix}
\begin{pmatrix}
\beta_1 \\
\alpha_1 \\
\gamma_1
\end{pmatrix} = 0
\]

Suppose that there are a priori linear restrictions on the structural parameters such that

\[
\Phi \begin{pmatrix}
\beta_1 \\
\alpha_1 \\
\gamma_1
\end{pmatrix} = 0
\]

where \( \Phi \) is a \( r \times (2m + k) \) matrix with full rank \( r \) that impose restriction on the corresponding equation.

Stacking the above linear restrictions together, we have that
\[
\begin{bmatrix}
\Pi_0 & c_0 \Pi_1 & I_k \\
\Pi_1 & \Pi_0 + c_1 \Pi_1 & 0 \\
\Phi & & \\
\end{bmatrix}
\begin{pmatrix}
\beta_1 \\
\alpha_1 \\
\gamma_1 \\
\end{pmatrix} = 0.
\]

where normalization rule sets $\beta_{11} = 1$.

**Proposition 6.1** (order condition of the structure with insufficient IVs):

If the $j$th equation is identified, then the number of a prior restrictions should be at least as larger as $2m - k - 1$.

In other words, additional $(m - k)$ restrictions are required in addition to $(m - 1)$ conditions if the structure has insufficient IVs. The inequality $r \geq 2m - k - 1$ is derived from $2k + r \geq 2m + k - 1$.

**Proposition 6.2** (rank condition of the structure with insufficient IVs):

The $j$th equation is identified if and only if

\[
\text{rank}
\begin{bmatrix}
\Pi_0 & c_0 \Pi_1 & I_k \\
\Pi_1 & \Pi_0 + c_1 \Pi_1 & 0 \\
\Phi & & \\
\end{bmatrix}
= 2m + k - 1.
\]

In particular, if $[\Pi_0, \Pi_1]$ has full row rank, then rank condition is equivalent to

\[
\text{rank}\Phi \begin{pmatrix} B \\ A \\ \Gamma \end{pmatrix} = 2m - k - 1.
\]

The full rank assumption on $[W_n E(Y_n), X_n]$ implies that at least one of exogenous variables in $[W_n X_n, X_n]$ is relevant to each dependent variable.
This is equivalent to say that there is no zero row in \([\Pi_0, \Pi_1]\), which is necessary for the full rank of reduced-form parameter matrix.

If Assumption 1-7 and the rank condition hold, the 2SLS IV estimator with \(Q_n = [W_n X_n, X_n]\) is consistent, asymptotically normal and best among linear IV estimators.

### 3.7.2 System with higher order spatial lags

The identification conditions in terms of linear restriction and linear GMM estimation method can be easily extended to the identification and estimation of a spatial system with high-order spatial lags, assuming that its SVAR process can be identification and estimated by linear IV method. Consider the identification and estimation of a spatial system with \(p\)-order spatial lags:

\[
Y_n B + \sum_{s=1}^{p} W_{sn} Y_n A_s + X_n \Gamma = U_n
\]

where \(W_{sn}\) are \(p\) distinct spatial weights matrices.

Suppose that Assumption 2 holds, its \(p\)-order spatial vector autoregressive process is given as follows.

\[
Y_n = \sum_{s=1}^{p} W_{sn} Y_n \Lambda_s + X_n \Pi + V_n
\]

with \(\Lambda_s \equiv -A_s B^{-1}\) for \(s = 1, \cdots, p\). For this system, let \(A = [A'_1, \cdots, A'_p]'\) and \(\Lambda = [\Lambda'_1, \cdots, \Lambda'_p]'\). It follows that \(\Theta = (\Lambda', \Pi')'\). Denote that \(S_n(\Lambda) = I_{mn} - \sum_{s=1}^{p} \Lambda_s \otimes W_{sn}\) and \(S_n = S_n(\Lambda_0)\). Assume that \(S_n\) is invertible and
uniformly bounded in row and column sums. Analogous to Proposition 3.2, if $Q_n' \left[ \sum_{s=1}^{p} W_{sn} E (Y_n), X_n \right]$ has full column rank, then $\theta_0$ is identified in $p$-order spatial vector autoregressive process.

For the identification of the $j$th structural equation, it shall become evident that the rank and order conditions in Proposition 4.1 and 4.2 holds for the system of higher order spatial lags. If the rank condition is satisfied, we can derive limited and full information estimators under similar assumptions in the previous sections, which are consistent and asymptotically normal. Further, it can be shown that the best 2SLS and 3SLS estimators use $\left[ \sum_{s=1}^{p} W_{sn} E (Y_n), X_n \right]$ as IVs and are asymptotically optimal among linear IV estimators.

3.7.3 System with SAR disturbance

In addition to spatial lagged dependence in the system, the spatial correlation can arise from the error correlation. Following the specification of Kelejian and Prucha (2004), we assume that the model disturbances are generated by the following spatial autoregressive process.

$$U_n = \sum_{s=1}^{p} (R_s \otimes W_{sn}) U_n + E_n,$$

where $R_s = \text{diag}^m_{j=1} (\rho_{js})$ and $E_n = (\varepsilon_{1n}, \cdots, \varepsilon_{mn})$, a $n \times m$ matrix of idiosyncratic noises. In general, the order of spatial lags in the error process coincides with the order of spatial lags in the model equation.

Let $\varepsilon_n = \text{vec}(E_n)$ be independently and identically distributed with $E \varepsilon_n = 0$ and $\text{var} \varepsilon_n = \Sigma \otimes I_n$ in this section. Denote that $R_n (\rho) = I_{mn} -$
\[ \sum_{s=1}^{p} (R_s \otimes W_{sn}) \text{ and } R_n = R_n (\rho_0). \] Assume that \( R_n \) is invertible and uniformly bounded in row and column sums. It follows that \( \varepsilon_n = R_n (\rho) u_n \) and \( u_n = R_n^{-1} \varepsilon_n \) in general equilibrium. This gives \( \text{var} u_n = R_n^{-1} (\Sigma \otimes I_n) R_n^{-1'} \) in the system and \( \text{var} u_{jn} = \sigma_{\varepsilon,jj}^2 R_{jn}^{-1} R_{jn}^{-1'} \) with \( R_{jn} = I_n - \sum_{s=1}^{p} \rho_{0,js} W_{sn} \) in the \( j \)th equation.

The identification and estimation of the structural model parameters treat error parameters in \( R_s \) and \( \Xi \) as unrestricted in the previous sections. So the 2SLS and 3SLS estimators remains consistent despite the potential spatial error correlation.

Assuming that the error parameters are known, we can apply a Cochrane-Orcutt-type transformation to the system equations:

\[ R_n y_n = R_n Z_n \delta + \varepsilon_n, \]

and for the \( j \)th equation, this corresponds to:

\[ R_{jn} y_{jn} = R_{jn} Z_{jn} \delta_j + \varepsilon_{jn}. \]

If the \( j \)th equation satisfies the rank condition, the generalized spatial 2SLS estimator \( \hat{\delta}_{jn} \) of \( \delta_j \) with IV matrix \( P_n \) is given as follows:

\[ \hat{\delta}_{jn} = [Z_{jn}' R_{jn}' P_n R_{jn} Z_{jn}]^{-1} Z_{jn}' R_{jn}' P_n R_{jn} y_{jn}. \]

Under similar assumptions in the previous sections, the asymptotic properties of \( \hat{\delta}_{jn} \) can be derived as follows.
\[
\sqrt{n} \left( \hat{\delta}_n - \delta_0 \right) \overset{d}{\longrightarrow} N \left( 0, \sigma^2 \left[ \lim_{n \to \infty} \frac{1}{n} \left( Z'_{jn} R'_{jn} P_n R_{jn} Z_{jn} \right) \right]^{-1} \right)
\]

Similarly, the generalized spatial 3SLS estimators \( \hat{\delta}_n \) of \( \delta \) with IV matrix \( P_n \) of the system are given as follows.

\[
\hat{\delta}_n = \left[ Z'_n R'_{jn} \left( \Sigma^{-1} \otimes I_n \right) R_n Z_n \right]^{-1} Z'_n R'_{jn} P_n \left( \Sigma^{-1} \otimes I_n \right) R_n y_n,
\]

and its asymptotic property is:

\[
\sqrt{n} \left( \hat{\delta}_n - \delta_0 \right) \overset{d}{\longrightarrow} N \left( 0, \left[ \lim_{n \to \infty} \frac{1}{n} \left( Z'_n R'_{jn} \left( \Sigma^{-1} \otimes P_n \right) R_n Z_n \right) \right]^{-1} \right)
\]

Corresponding to the GS2SLS and GS3SLS estimators (\( \hat{\delta}_n \) and \( \hat{\delta}_n \)), we can define the feasible GS2SLS and GS3SLS estimators (\( \hat{\delta}^F_{jn} \) and \( \hat{\delta}^F_n \)) as follows.

\[
\hat{\delta}^F_{jn} = \left[ Z'_{jn} R'_{jn} \left( \hat{\rho}_n \right) P_n R_{jn} \left( \hat{\rho}_n \right) Z_{jn} \right]^{-1} Z'_{jn} R'_{jn} \left( \hat{\rho}_n \right) P_n R_{jn} \left( \hat{\rho}_n \right) y_{jn},
\]

\[
\hat{\delta}^F_n = \left[ Z'_n R'_{jn} \left( \hat{\rho}_n \right) P_n \left( \hat{\Sigma}^{-1} \otimes I_n \right) R_n \left( \hat{\rho}_n \right) Z_n \right]^{-1} \\
\times Z'_n R'_{jn} \left( \hat{\rho}_n \right) P_n \left( \hat{\Sigma}^{-1} \otimes I_n \right) R_n \left( \hat{\rho}_n \right) y_n,
\]

where \( \hat{\rho}_n \) and \( \hat{\Sigma}^{-1} \) are any consistent estimator for true value \( \rho_0 \) and \( \Sigma \). In general, the spatial error coefficients \( \rho \) can be consistently estimated by Kelejían and Prucha’s the method of moments or Lee’s GMM procedures.
Lemma 5.2 shows that $\lim_{n\to\infty} \hat{\Sigma} = \Sigma$ under relevant assumptions. Theorem 4 of Kelejian and Prucha (2004, p39) shows that the true and feasible G2SLS and G3SLS estimators have the same asymptotic distribution. The above asymptotic results of true GS2SLS and GS3SLS estimators shall hold if nuisance parameters $\rho_0$ and $\Sigma$ are replaced by any of their consistent estimators. Furthermore, the feasible GS2SLS and GS3SLS estimators utilizing best IV $P_n^*$ are asymptotically optimal among linear IV estimators.

3.8 Conclusion

This paper develops the identification conditions and estimation methods for a simultaneous system of spatially interrelated cross-sectional equations. We show that if the spatial vector autoregressive process of the system can be identified, the structural model can be identified under sufficient and necessary rank conditions analogous to the classical linear system of equations. Then we introduce the limited information and full information estimators (i.e., FG2SLS and FG3SLS estimators) and derives their asymptotic properties. Among the class of linear IV estimators, there exist best 2SLS and 3SLS estimators that are asymptotically optimal. Hausman test statistics is also provided. The identification conditions and estimation results have been extended to generalized spatial models with insufficient IVs, higher order spatial lags and spatial error dependence.

3.9 Appendix of Chapter 3

A: Some Useful Lemmas
Lemma A.1. Let $A_1$, $A_2$, $B_1$ and $B_2$ be conformable matrices, then

$$(A_1 \otimes B_1) (A_2 \otimes B_2) = A_1 A_2 \otimes B_1 B_2.$$ 

Lemma A.2. If $A_1$ and $A_2$ are conformable matrices, then

$$vec(A_1 A_2) = (I \otimes A_1) vec(A_2) = (A'_2 \otimes I) vec(A_1).$$ 

Lemma A.3. Let $A$ be $m \times m$ and $B$ be $n \times n$. Then $\|A \otimes B\| = \|A\| \|B\|$ where $\|\cdot\|$ denotes maximum column (or row) sum matrix norm.

Proof. Without loss of generality, consider the maximum row sum matrix norm at first.

$$\|A \otimes B\| = \max_{1 \leq j \leq m} \sum_{i=1}^{m} \|a_{ij}B\| = \|B\| \max_{1 \leq j \leq m} \sum_{i=1}^{m} |a_{ij}| = \|A\| \|B\|$$

The above result also holds for maximum column sum matrix norm.

Lemma A.4. Suppose Assumptions 1-6 holds:

$$\text{plim}_{n \to \infty} \frac{1}{n} Q'_n \overline{Z}_n = \lim_{n \to \infty} \frac{1}{n} Q'_n [W_n E(Y_n), X_n] = \lim_{n \to \infty} \frac{1}{n} Q'_n Q^*_n.$$

Proof. Recall that $\overline{Z}_n = [W_n Y_n, X_n] = [\overline{Y}_n, X_n]$ and $X_n$ is nonstochastic. Hence, to prove (A.4), it suffices to show that

$$\text{plim}_{n \to \infty} \frac{1}{n} Q'_n \overline{Y}_n = \lim_{n \to \infty} \frac{1}{n} Q'_n [W_n E(Y_n), X_n].$$

Observe that $y_n = S^{-1}_n (\Pi'_0 \otimes I_n) x_n + S^{-1}_n v_n$. It follows that

$$\overline{y}_n = G_n (\Pi'_0 \otimes I_n) x_n + G_n v_n,$$
with $G_n = (I_m \otimes W_n) S_n^{-1}$. Thus,

$$\frac{1}{n} vec (Q_n' \overline{Y}_n) = \frac{1}{n} vec (I_m \otimes Q_n') \overline{y}_n$$

$$= \frac{1}{n} (I_m \otimes Q_n') E(\overline{y}_n) + \frac{1}{n} (I_m \otimes Q_n') G_n v_n$$

$$= \frac{1}{n} vec (Q_n' W_n E(Y_n)) + \frac{1}{n} (I_m \otimes Q_n') G_n v_n.$$ 

By Assumptions 4 and 5, the elements of $(I_m \otimes Q_n') G_n$ are bounded uniformly in absolute value. Since $v_n$ is a vector of i.i.d. random variable with zero mean and finite variance matrix by Assumption 6, it follows that

$$\frac{1}{n} (I_m \otimes Q_n') G_n v_n = o_p(1).$$

**B. Proofs on selected Theorems**

**Proof of Proposition 4.2:** First we note that

$$\left[ \begin{array}{cc}
\Lambda & I_m \\
\Pi & 0
\end{array} \right] \left[ \begin{array}{c}
B \\
A
\end{array} \right] = 0 \implies \det \left[ \begin{array}{ccc}
\Lambda' & \Pi' & B \\
I_m & 0 & A \\
0 & I_k & \Gamma
\end{array} \right] \neq 0.$$ 

The matrix above is of dimension $(2m + k) \times (2m + k)$ and $\Phi$ is a $(2m + k) \times r$ matrix, there exists constant matrices $C_1$, $C_2$ and $C_3$ such that

$$\Phi' = \left[ \begin{array}{c}
\Lambda' \\
I_m \\
0
\end{array} \right] C_1 + \left[ \begin{array}{c}
\Pi' \\
0 \\
I_k
\end{array} \right] C_2 + \left[ \begin{array}{c}
B \\
A \\
\Gamma
\end{array} \right] C_3$$

Since
\[
\text{rank} \left( \begin{array}{ccc}
\Lambda & I_m & 0 \\
\Pi & 0 & I_k \\
\Phi & & 
\end{array} \right)
\]

\[=
\text{rank} \left( \begin{array}{ccc}
\Lambda' & \Pi' & I_m \\
I_m & 0 & 0 \\
0 & I_k & 0 \\
\end{array} \right) \left( \begin{array}{c}
\Lambda' \\
I_m \\
I_k \\
\end{array} \right) C_1 + \left( \begin{array}{c}
\Pi' \\
0 \\
I_k \\
\end{array} \right) C_2 + \left( \begin{array}{c}
B \\
0 \\
\Gamma \\
\end{array} \right) C_3 \right)
\]

\[=
\text{rank} \left( \begin{array}{ccc}
\Lambda' & \Pi' & B \\
I_m & 0 & A \\
0 & I_k & \Gamma \\
\end{array} \right) C_3 \right)
\]

and

\[
\text{rank} \left( \begin{array}{c}
B \\
A \\
\Gamma \\
\end{array} \right) C_3 = \text{rank} \left( \begin{array}{ccc}
\Lambda' & \Pi' & B \\
I_m & 0 & A \\
0 & I_k & \Gamma \\
\end{array} \right) \left( \begin{array}{c}
0 \\
0 \\
C_3 \\
\end{array} \right)
\]

\[=
\text{rank} \left( \begin{array}{c}
0 \\
0 \\
C_3 \\
\end{array} \right) = \text{rank} C_3
\]

It follows that for the rank condition is equivalent to that
\begin{align*}
\iff \quad \text{rank} \begin{pmatrix}
\Lambda & I_m & 0 \\
\Pi & 0 & I_k \\
\Phi
\end{pmatrix} = 2m + k - 1 \\
\iff \quad \text{rank} \begin{pmatrix}
\Lambda' & II' \\
I_m & 0 \\
0 & I_k
\end{pmatrix} + \text{rank} \begin{pmatrix}
B \\
A \\
\Gamma
\end{pmatrix} C_3 = 2m + k - 1 \\
\iff \quad \text{rank} \begin{pmatrix}
B \\
A \\
\Gamma
\end{pmatrix} C_3 = m - 1 \\
\iff \quad \text{rank} C_3 = m - 1.
\end{align*}

On the other hand, substituting the span of the restriction matrix

\[
\Phi' = \begin{pmatrix}
\Lambda' \\
I_m \\
0
\end{pmatrix} C_1 + \begin{pmatrix}
II' \\
0 \\
I_k
\end{pmatrix} C_2 + \begin{pmatrix}
B \\
A \\
\Gamma
\end{pmatrix} C_3
\]

into
\[ \Phi \begin{pmatrix} B \\ A \\ \Gamma \end{pmatrix} = C_1' \begin{pmatrix} \Lambda & I_m & 0 \\ A \\ \Gamma \end{pmatrix} + C_2' \begin{pmatrix} \Pi & 0 & I_k \\ A \\ \Gamma \end{pmatrix} + C_3' \begin{pmatrix} B' & A' & \Gamma' \\ A \\ \Gamma \end{pmatrix} = C_3' (B'B + A'A + \Gamma'\Gamma), \]

which implies that

\[ \text{rank} \Phi \begin{pmatrix} B \\ A \\ \Gamma \end{pmatrix} = \text{rank} C_3. \]

Therefore, the rank condition can be rewritten as

\[ \text{rank} \Phi \begin{pmatrix} B \\ A \\ \Gamma \end{pmatrix} = m - 1. \]
Proof of Lemma 5.1: Without loss of generality, consider the first structural equation with 2SLS IV matrix $Q_n^* (Q_n^* Q_n^*)^{-1} Q_n^*$.

\[
\begin{align*}
\text{plim}_{n \to \infty} & \quad \frac{1}{n} \left[ Z_{1n}^\prime Q_n^* (Q_n^* Q_n^*)^{-1} Q_n^* Z_{1n} \right] \\
= & \quad \text{plim}_{n \to \infty} \frac{1}{n} \begin{pmatrix}
\tilde{\Theta}_1^\prime Q_n^* Q_n^* \tilde{\Theta}_1 & \tilde{\Theta}_1^\prime Q_n^* J_1 \\
J_1^\prime Q_n^* Q_n^* \tilde{\Theta}_1 & J_1^\prime Q_n^* Q_n^* J_1
\end{pmatrix} \\
= & \quad \begin{pmatrix}
\Theta_1^\prime \\
J_1^\prime
\end{pmatrix} \left[ \lim_{n \to \infty} \frac{1}{n} Q_n^* Q_n^* \right] (\Theta_1, J_1)
\end{align*}
\]

where the second equality is implied by Lemma A.4 and $\Theta_1$ (and $\tilde{\Theta}_1$) are corresponding parameter columns (and theirs estimators) of endogenous variables $Y_{1n}$ in the first equation.

Recall that $Y_n B + Y_n A + X_n \Gamma = U_n$ implies that $\Lambda B + A \equiv 0$ and $\Pi B + \Gamma \equiv 0$. For the first structural equation, $y_{1n} = Y_{1n} \beta_1 + Y_{1n} \alpha_1 + X_{1n} \gamma_1 + u_{1n}$, the identity equations implies that

\[
\begin{align*}
\Lambda \begin{pmatrix} 1 & -\beta_1^\prime & 0 \end{pmatrix} &= \begin{pmatrix} \alpha_1^\prime & 0 \end{pmatrix}, \\
\Pi \begin{pmatrix} 1 & -\beta_1^\prime & 0 \end{pmatrix} &= \begin{pmatrix} \gamma_1^\prime & 0 \end{pmatrix}.
\end{align*}
\]

Partition $\Lambda$ and $\Pi$ conformably as

\[
\Lambda = \begin{pmatrix}
\Lambda_{m_1} & \Lambda_{m_1 m_1} & \Lambda_{m_1 m_1}^* \\
\Lambda_{m_1}^* & \Lambda_{m_1 m_1} & \Lambda_{m_1 m_1}^*
\end{pmatrix} \quad \text{and} \quad \Pi = \begin{pmatrix}
\Pi_{k_1} & \Pi_{k_1 m_1} & \Pi_{k_1 m_1}^* \\
\Pi_{k_1}^* & \Pi_{k_1 m_1} & \Pi_{k_1 m_1}^*
\end{pmatrix},
\]

119
where \( \overline{m}_1 + \overline{m}_1^* = m \), \( k_1 + k_1^* = k \) and \( m_1 + m_1^* + 1 = m \). It follows that

\[
\begin{align*}
\Lambda_{\overline{m}_1 m_1} \beta_1 + \alpha_1 &= \Lambda_{\overline{m}_1 m_1}, \\
\Lambda_{\overline{m}_1^* m_1^*} - \Lambda_{\overline{m}_1 m_1} \beta_1 &= 0; \\
\Pi_{k_1 m_1} \beta_1 + \gamma_1 &= \Pi_{k_1}, \\
\Pi_{k_1^* m_1^*} - \Pi_{k_1 m_1} \beta_1 &= 0.
\end{align*}
\]

The rank condition is equivalent to

\[
\text{rank} \left[ \Lambda_{\overline{m}_1 m_1}, \Pi_{k_1^* m_1} \right] = m_1.
\]

It follows that

\[
(\Theta_1, J_1) = \begin{pmatrix}
\Lambda_{\overline{m}_1 m_1} & \Pi_{k_1 m_1} & I_{\overline{m}_1} & I_{k_1} \\
\Lambda_{\overline{m}_1 m_1}^* & \Pi_{k_1^* m_1} & 0 & 0
\end{pmatrix}
\]

has full column rank.

If Assumption 7 holds, then we have that

\[
\left( \lim_{n \to \infty} \frac{1}{n} Q_n^* Q_n^* \right) (\Theta_1, J_1)
\]

is nonsingular.
## APPENDIX OF TABLES

### Parameters Design

<table>
<thead>
<tr>
<th>Name</th>
<th>Explanations</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda$ (0.2, 0.4, 0.7)</td>
<td>Coefficient of Spatial Autoregressive Lag</td>
</tr>
<tr>
<td>$\beta_1$ (0.2, 0.4, 0.7)</td>
<td>Coefficient of Temporal Autoregressive Lag</td>
</tr>
<tr>
<td>$\beta_2$ (-0.15, 0, 0.15)</td>
<td>Coefficient of Spatial-Temporal Autoregressive Lag</td>
</tr>
<tr>
<td>$\beta_3$ (0, 0.5, 1)</td>
<td>Coefficient of Strictly Exogenous Regressor</td>
</tr>
</tbody>
</table>

### Panel Design

<table>
<thead>
<tr>
<th>Name</th>
<th>Explanations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size (n = 49, 245, 490)</td>
<td>Number of cross-sectional units</td>
</tr>
<tr>
<td>Period (T= 5, 10)</td>
<td>Number of time periods covered in the model</td>
</tr>
<tr>
<td>Spatial Weights Matrix (Wn)</td>
<td>The matrix, of dimension 49×49, has been used in Columbus crime data by Anselin. It will be Kronecker-producted by 1, 5 and 10 when size is 49, 245 and 490 respectively.</td>
</tr>
</tbody>
</table>

### Disturbance Design

<table>
<thead>
<tr>
<th>Name</th>
<th>Explanations</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma^2_a$ (0.5, 1, 2)</td>
<td>Variance of individual effects in the model.</td>
</tr>
<tr>
<td>$\sigma^2_\epsilon$ (0.5, 1, 2)</td>
<td>Variance of random disturbance in the model.</td>
</tr>
</tbody>
</table>

**Table 1: Monte Carlo Experiment Design**
Note that the following table reports the bias, deviation and root mean squared error of parameters vertically in cells. The numbers in parentheses are deviations and the numbers in brackets are root mean squared errors. The selected model design parameters are $T = 5$, $\sigma^2_\alpha = 0.5$ and $\sigma^2_\varepsilon = 0.5$. The Anselin spatial weights matrix are used in the spatial models.

<table>
<thead>
<tr>
<th>n = 49</th>
<th>2SLS1</th>
<th>2SLS2</th>
<th>GMM1</th>
<th>GMM2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda = 0.7$</td>
<td>0.2039 (0.0587)</td>
<td>0.2150 (0.0521)</td>
<td>0.0503 (0.1032)</td>
<td>0.0512 (0.1023)</td>
</tr>
<tr>
<td></td>
<td>[0.3167]</td>
<td>[0.3136]</td>
<td>[0.3252]</td>
<td>[0.3240]</td>
</tr>
<tr>
<td>$\beta_1 = 0.2$</td>
<td>-0.1004 (0.0813)</td>
<td>-0.1321 (0.0769)</td>
<td>-0.1023 (0.0902)</td>
<td>-0.1030 (0.0947)</td>
</tr>
<tr>
<td></td>
<td>[0.3023]</td>
<td>[0.3073]</td>
<td>[0.3172]</td>
<td>[0.3245]</td>
</tr>
<tr>
<td>$\beta_2 = 0$</td>
<td>-0.0462 (0.1008)</td>
<td>-0.0195 (0.0933)</td>
<td>-0.0320 (0.1254)</td>
<td>-0.0282 (0.1263)</td>
</tr>
<tr>
<td></td>
<td>[0.3208]</td>
<td>[0.3061]</td>
<td>[0.3555]</td>
<td>[0.3565]</td>
</tr>
<tr>
<td>$\beta_3 = 1$</td>
<td>-0.0133 (0.1351)</td>
<td>-0.0054 (0.1350)</td>
<td>-0.0120 (0.2240)</td>
<td>-0.0086 (0.2240)</td>
</tr>
<tr>
<td></td>
<td>[0.3678]</td>
<td>[0.3675]</td>
<td>[0.4735]</td>
<td>[0.4734]</td>
</tr>
</tbody>
</table>

(Continued)
Table 2: Continued

<table>
<thead>
<tr>
<th>n = 245</th>
<th>2SLS1</th>
<th>2SLS2</th>
<th>GMM1</th>
<th>GMM2</th>
</tr>
</thead>
<tbody>
<tr>
<td>λ = 0.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1141</td>
<td>(0.0427)</td>
<td>0.1271</td>
<td>(0.0359)</td>
<td>0.0325</td>
</tr>
<tr>
<td>[0.2381]</td>
<td></td>
<td>[0.2281]</td>
<td></td>
<td>[0.2891]</td>
</tr>
<tr>
<td>β1 = 0.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.0206</td>
<td>(0.0397)</td>
<td>-0.0288</td>
<td>(0.0390)</td>
<td>-0.0287</td>
</tr>
<tr>
<td>[0.2002]</td>
<td></td>
<td>[0.1994]</td>
<td></td>
<td>[0.2438]</td>
</tr>
<tr>
<td>β2 = 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.0690</td>
<td>(0.0557)</td>
<td>-0.0720</td>
<td>(0.0501)</td>
<td>-0.0056</td>
</tr>
<tr>
<td>[0.2068]</td>
<td></td>
<td>[0.2352]</td>
<td></td>
<td>[0.3149]</td>
</tr>
<tr>
<td>β3 = 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.0217</td>
<td>(0.0608)</td>
<td>-0.0222</td>
<td>(0.0600)</td>
<td>-0.0151</td>
</tr>
<tr>
<td>[0.2476]</td>
<td></td>
<td>[0.2459]</td>
<td></td>
<td>[0.4082]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>n = 490</th>
<th>2SLS1</th>
<th>2SLS2</th>
<th>GMM1</th>
<th>GMM2</th>
</tr>
</thead>
<tbody>
<tr>
<td>λ = 0.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0769</td>
<td>(0.0333)</td>
<td>0.0774</td>
<td>(0.0310)</td>
<td>0.0295</td>
</tr>
<tr>
<td>[0.1980]</td>
<td></td>
<td>[0.1923]</td>
<td></td>
<td>[0.2566]</td>
</tr>
<tr>
<td>β1 = 0.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.0103</td>
<td>(0.0278)</td>
<td>-0.0120</td>
<td>(0.0275)</td>
<td>-0.0182</td>
</tr>
<tr>
<td>[0.1670]</td>
<td></td>
<td>[0.1662]</td>
<td></td>
<td>[0.2026]</td>
</tr>
<tr>
<td>β2 = 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.0512</td>
<td>(0.0401)</td>
<td>-0.0521</td>
<td>(0.0395)</td>
<td>-0.0098</td>
</tr>
<tr>
<td>[0.2068]</td>
<td></td>
<td>[0.2055]</td>
<td></td>
<td>[0.3167]</td>
</tr>
<tr>
<td>β3 = 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.0172</td>
<td>(0.0414)</td>
<td>-0.0168</td>
<td>(0.0403)</td>
<td>-0.0068</td>
</tr>
<tr>
<td>[0.2043]</td>
<td></td>
<td>[0.2014]</td>
<td></td>
<td>[0.3113]</td>
</tr>
</tbody>
</table>
### Table 3: Mean of Variables in analysis of local school expenditures

(Standard Deviations in parentheses)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Revenue</td>
<td>5012</td>
<td>5199</td>
<td>5408</td>
<td>5822</td>
<td>6170</td>
</tr>
<tr>
<td></td>
<td>(1229)</td>
<td>(1385)</td>
<td>(1550)</td>
<td>(1729)</td>
<td>(1435)</td>
</tr>
<tr>
<td>Expenditure</td>
<td>4982</td>
<td>5152</td>
<td>5238</td>
<td>5602</td>
<td>5858</td>
</tr>
<tr>
<td></td>
<td>(1054)</td>
<td>(1112)</td>
<td>(1395)</td>
<td>(1250)</td>
<td>(1238)</td>
</tr>
<tr>
<td>State Grants</td>
<td>2106</td>
<td>2163</td>
<td>2357</td>
<td>2459</td>
<td>2603</td>
</tr>
<tr>
<td></td>
<td>(600.2)</td>
<td>(780.0)</td>
<td>(831.7)</td>
<td>(1312)</td>
<td>(933.6)</td>
</tr>
<tr>
<td>Income</td>
<td>30.55</td>
<td>30.15</td>
<td>30.83</td>
<td>31.41</td>
<td>32.36</td>
</tr>
<tr>
<td></td>
<td>(10.96)</td>
<td>(10.87)</td>
<td>(10.86)</td>
<td>(11.92)</td>
<td>(13.17)</td>
</tr>
<tr>
<td>Pupil Density</td>
<td>114.7</td>
<td>115.5</td>
<td>115.9</td>
<td>116.2</td>
<td>115.3</td>
</tr>
<tr>
<td></td>
<td>(190.0)</td>
<td>(191.3)</td>
<td>(190.6)</td>
<td>(190.2)</td>
<td>(186.8)</td>
</tr>
<tr>
<td></td>
<td>(12.50)</td>
<td>(12.88)</td>
<td>(13.23)</td>
<td>(13.37)</td>
<td>(13.72)</td>
</tr>
<tr>
<td>Prop. School Age</td>
<td>19.31</td>
<td>19.33</td>
<td>19.33</td>
<td>19.40</td>
<td>19.32</td>
</tr>
<tr>
<td></td>
<td>(1.54)</td>
<td>(1.54)</td>
<td>(1.53)</td>
<td>(1.51)</td>
<td>(1.49)</td>
</tr>
<tr>
<td></td>
<td>(2.11)</td>
<td>(2.11)</td>
<td>(2.09)</td>
<td>(2.09)</td>
<td>(2.09)</td>
</tr>
</tbody>
</table>
Table 3: Continued

Sources:
1. Revenue, expenditure, state grant, income and pupil data are from the Ohio Department of Education (ODE: http://www.ode.state.oh.us/data/).
2. Demographic characteristics of the population are from the Bureau of Census.
3. School district areas used in computing pupil density are from the Center for Urban and Regional Analysis (CURA) at Ohio State University.

Notes:
1. All dollar figures are in real per pupil or per tax return term. Pupil density is in units of pupils per square mile. Income is measured in thousands of dollars. Proportion of the pupils that are black, proportion of the population between 5 to 17 and at least 65 years old are percent number.
2. Expenditure (and revenue) is the sum of local and state subsidies, state and federal funds, food service, lotteries and fees.
3. State grants are an attempt to subsidize relatively poor school districts.
4. No annual observations are reported for one school district in the ODE data set.
<table>
<thead>
<tr>
<th>Model Variable</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ΔlnR_{nt}</td>
<td>0.376</td>
<td>0.360</td>
<td>0.394</td>
<td>−0.006</td>
<td>0.673</td>
<td>−0.038</td>
</tr>
<tr>
<td></td>
<td>(0.021)</td>
<td>(0.021)</td>
<td>(0.138)</td>
<td>(0.164)</td>
<td>(0.132)</td>
<td>(0.152)</td>
</tr>
<tr>
<td>ΔlnW_{nt}^{geo}G_{nt}</td>
<td>0.260</td>
<td>0.828</td>
<td>0.895</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.040)</td>
<td>(0.195)</td>
<td>(0.123)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>0.033</td>
<td>0.085</td>
<td>0.031</td>
<td>0.011</td>
<td>0.012</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.003)</td>
<td>(0.010)</td>
<td>(0.012)</td>
<td>(0.009)</td>
<td>(0.008)</td>
</tr>
</tbody>
</table>

Table 4: Estimates of Nested Test Model
(Standard errors in parentheses)

Note:
1. In all nested test equations, dependent variable is ΔlnGn,t+1, growth rate of local school spending; explanatory variables are ΔlnR_{nt}, growth rate of local revenue, and ΔlnW_{nt}^{geo}G_{nt}, growth rate of neighborhood spending, where spatial weights matrix W_{nt}^{geo} uses geographic proximity.
2. Model (1) – (2) use OLS method; Model (3) – (6) use IV method.
3. Model (3) - (4) use differenced lags as IVs; model (5) - (6) use level lags as IVs.
<table>
<thead>
<tr>
<th>Model</th>
<th>Variable</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAR Lag</td>
<td>OLS</td>
<td>0.227</td>
<td>0.689</td>
<td>0.674</td>
<td>0.515</td>
<td>0.498</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.032)</td>
<td>(0.055)</td>
<td>(0.051)</td>
<td>(0.041)</td>
<td>(0.042)</td>
<td></td>
</tr>
<tr>
<td>TAR Lag</td>
<td>IV</td>
<td>0.906</td>
<td>0.391</td>
<td>0.350</td>
<td>0.335</td>
<td>0.186</td>
<td>0.183</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.009)</td>
<td>(0.034)</td>
<td>(0.023)</td>
<td>(0.023)</td>
<td>(0.024)</td>
<td>(0.024)</td>
</tr>
<tr>
<td>STAR Lag</td>
<td>IV</td>
<td>-0.180</td>
<td>0.643</td>
<td>-0.143</td>
<td>-0.117</td>
<td>-0.102</td>
<td>-0.118</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.034)</td>
<td>(0.060)</td>
<td>(0.051)</td>
<td>(0.047)</td>
<td>(0.040)</td>
<td>(0.041)</td>
</tr>
<tr>
<td>Pupil Density</td>
<td></td>
<td>-0.119</td>
<td>-3.01</td>
<td>-2.55</td>
<td>-2.63</td>
<td>-2.60</td>
<td>-2.68</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.063)</td>
<td>(1.27)</td>
<td>(1.04)</td>
<td>(1.05)</td>
<td>(1.06)</td>
<td>(1.06)</td>
</tr>
<tr>
<td>Income</td>
<td></td>
<td>13.35</td>
<td>5.83</td>
<td>20.10</td>
<td>21.80</td>
<td>35.90</td>
<td>28.34</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2.65)</td>
<td>(12.85)</td>
<td>(11.20)</td>
<td>(11.15)</td>
<td>(10.58)</td>
<td>(10.55)</td>
</tr>
<tr>
<td>Income Squared</td>
<td></td>
<td>-0.058</td>
<td>-0.005</td>
<td>-0.046</td>
<td>-0.053</td>
<td>-0.086</td>
<td>-0.069</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.018)</td>
<td>(0.056)</td>
<td>(0.048)</td>
<td>(0.048)</td>
<td>(0.047)</td>
<td>(0.042)</td>
</tr>
<tr>
<td>Prop. Black pupils</td>
<td></td>
<td>3.97</td>
<td>-8.69</td>
<td>3.93</td>
<td>3.50</td>
<td>57.81</td>
<td>55.83</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.88)</td>
<td>(13.17)</td>
<td>(11.24)</td>
<td>(11.25)</td>
<td>(11.26)</td>
<td>(11.26)</td>
</tr>
<tr>
<td>Prop. School age</td>
<td></td>
<td>-13.90</td>
<td>162.01</td>
<td>32.28</td>
<td>-32.67</td>
<td>-46.44</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(7.09)</td>
<td>(98.89)</td>
<td>(83.74)</td>
<td>(77.18)</td>
<td>(77.30)</td>
<td></td>
</tr>
<tr>
<td>Prop. Elderly</td>
<td></td>
<td>-0.16</td>
<td>-176.45</td>
<td>-35.42</td>
<td>17.34</td>
<td>13.84</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(4.70)</td>
<td>(88.21)</td>
<td>(72.76)</td>
<td>(65.25)</td>
<td>(65.11)</td>
<td></td>
</tr>
<tr>
<td>State Grants</td>
<td></td>
<td>0.140</td>
<td></td>
<td></td>
<td>0.364</td>
<td>0.360</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.011)</td>
<td></td>
<td></td>
<td>(0.024)</td>
<td>(0.024)</td>
<td></td>
</tr>
<tr>
<td>Reg. Economy</td>
<td></td>
<td>0.760</td>
<td></td>
<td></td>
<td>20.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.79)</td>
<td></td>
<td></td>
<td>(9.81)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAR Error</td>
<td></td>
<td>0.360</td>
<td>-0.646</td>
<td>-0.618</td>
<td>-0.360</td>
<td>-0.355</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.030)</td>
<td>(0.025)</td>
<td>(0.024)</td>
<td>(0.024)</td>
<td>(0.024)</td>
<td></td>
</tr>
</tbody>
</table>

Table 5: G2SLS Estimates with geographic weights matrix

(Standard errors in parentheses)
<table>
<thead>
<tr>
<th>Model Variable</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAR 1ST Lag</td>
<td>0.219</td>
<td>0.670</td>
<td>0.667</td>
<td>0.508</td>
<td>0.492</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.032)</td>
<td>(0.055)</td>
<td>(0.051)</td>
<td>(0.041)</td>
<td>(0.042)</td>
<td></td>
</tr>
<tr>
<td>SAR 2nd Lag</td>
<td>0.164</td>
<td>0.186</td>
<td>0.198</td>
<td>0.139</td>
<td>0.124</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.103)</td>
<td>(0.196)</td>
<td>(0.208)</td>
<td>(0.162)</td>
<td>(0.141)</td>
<td></td>
</tr>
<tr>
<td>TAR Lag</td>
<td>0.906</td>
<td>0.388</td>
<td>0.331</td>
<td>0.314</td>
<td>0.174</td>
<td>0.173</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.036)</td>
<td>(0.023)</td>
<td>(0.022)</td>
<td>(0.023)</td>
<td>(0.023)</td>
</tr>
<tr>
<td>STAR 1ST Lag</td>
<td>-0.172</td>
<td>0.645</td>
<td>-0.125</td>
<td>-0.110</td>
<td>-0.094</td>
<td>-0.106</td>
</tr>
<tr>
<td></td>
<td>(0.035)</td>
<td>(0.069)</td>
<td>(0.050)</td>
<td>(0.047)</td>
<td>(0.040)</td>
<td>(0.040)</td>
</tr>
<tr>
<td>STAR 2nd Lag</td>
<td>-0.165</td>
<td>-0.009</td>
<td>-0.103</td>
<td>-0.115</td>
<td>-0.106</td>
<td>-0.132</td>
</tr>
<tr>
<td></td>
<td>(0.107)</td>
<td>(0.104)</td>
<td>(0.214)</td>
<td>(0.228)</td>
<td>(0.178)</td>
<td>(0.155)</td>
</tr>
<tr>
<td>Pupil Density</td>
<td>-0.139</td>
<td>-2.48</td>
<td>-2.77</td>
<td>-2.83</td>
<td>-2.72</td>
<td>-2.80</td>
</tr>
<tr>
<td></td>
<td>(0.067)</td>
<td>(1.44)</td>
<td>(1.15)</td>
<td>(1.15)</td>
<td>(1.11)</td>
<td>(1.08)</td>
</tr>
<tr>
<td>Income</td>
<td>13.17</td>
<td>7.54</td>
<td>20.09</td>
<td>21.44</td>
<td>35.60</td>
<td>25.65</td>
</tr>
<tr>
<td></td>
<td>(2.67)</td>
<td>(13.81)</td>
<td>(12.09)</td>
<td>(12.23)</td>
<td>(11.21)</td>
<td>(10.71)</td>
</tr>
<tr>
<td>Income Squared</td>
<td>-0.057</td>
<td>0.0123</td>
<td>-0.052</td>
<td>-0.056</td>
<td>-0.087</td>
<td>-0.070</td>
</tr>
<tr>
<td></td>
<td>(0.018)</td>
<td>(0.059)</td>
<td>(0.049)</td>
<td>(0.050)</td>
<td>(0.048)</td>
<td>(0.047)</td>
</tr>
<tr>
<td>Prop. Black pupils</td>
<td>3.85</td>
<td>-7.14</td>
<td>-0.869</td>
<td>-0.594</td>
<td>55.66</td>
<td>55.03</td>
</tr>
<tr>
<td></td>
<td>(0.89)</td>
<td>(13.77)</td>
<td>(13.69)</td>
<td>(13.79)</td>
<td>(13.18)</td>
<td>(12.32)</td>
</tr>
<tr>
<td>Prop. School age</td>
<td>-11.56</td>
<td>189.82</td>
<td>-10.52</td>
<td>-54.15</td>
<td>-54.97</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(7.50)</td>
<td>(104.34)</td>
<td>(96.68)</td>
<td>(85.75)</td>
<td>(83.17)</td>
<td></td>
</tr>
<tr>
<td>Prop. Elderly</td>
<td>-1.13</td>
<td>-178.19</td>
<td>-17.05</td>
<td>26.92</td>
<td>18.26</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.88)</td>
<td>(93.90)</td>
<td>(80.14)</td>
<td>(69.68)</td>
<td>(68.05)</td>
<td></td>
</tr>
</tbody>
</table>

(Continued)

Table 6: G2SLS Estimates with Secondary Income Weights Matrix
(Standard errors in parentheses)
Table 6: Continued

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.141</td>
<td>0.366</td>
<td>0.363</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.025)</td>
<td>(0.025)</td>
</tr>
<tr>
<td>Reg. Economy</td>
<td>0.392</td>
<td>20.26</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.83)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAR 1\text{st} Error</td>
<td>0.400</td>
<td>-0.580</td>
<td>-0.576</td>
</tr>
<tr>
<td></td>
<td>(0.027)</td>
<td>(0.025)</td>
<td>(0.027)</td>
</tr>
<tr>
<td>SAR 2\text{nd} Error</td>
<td>-0.580</td>
<td>-1.29</td>
<td>-1.31</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.010)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>Model</td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>-----------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>Variable</td>
<td>OLS</td>
<td>IV</td>
<td>IV</td>
</tr>
<tr>
<td>SAR 1\textsuperscript{st} Lag</td>
<td>0.229</td>
<td>0.682</td>
<td>0.676</td>
</tr>
<tr>
<td></td>
<td>(0.032)</td>
<td>(0.054)</td>
<td>(0.051)</td>
</tr>
<tr>
<td>SAR 2\textsuperscript{nd} Lag</td>
<td>-0.042</td>
<td>0.215</td>
<td>0.194</td>
</tr>
<tr>
<td></td>
<td>(0.102)</td>
<td>(0.203)</td>
<td>(0.204)</td>
</tr>
<tr>
<td>TAR Lag</td>
<td>0.906</td>
<td>0.369</td>
<td>0.334</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.036)</td>
<td>(0.022)</td>
</tr>
<tr>
<td>STAR 1\textsuperscript{st} Lag</td>
<td>-0.183</td>
<td>0.645</td>
<td>-0.141</td>
</tr>
<tr>
<td></td>
<td>(0.035)</td>
<td>(0.067)</td>
<td>(0.050)</td>
</tr>
<tr>
<td>STAR 2\textsuperscript{nd} Lag</td>
<td>0.051</td>
<td>0.071</td>
<td>-0.097</td>
</tr>
<tr>
<td></td>
<td>(0.105)</td>
<td>(0.102)</td>
<td>(0.218)</td>
</tr>
<tr>
<td>Pupil Density</td>
<td>-0.145</td>
<td>-2.90</td>
<td>-2.74</td>
</tr>
<tr>
<td></td>
<td>(0.066)</td>
<td>(1.75)</td>
<td>(1.30)</td>
</tr>
<tr>
<td>Income</td>
<td>12.95</td>
<td>7.71</td>
<td>18.72</td>
</tr>
<tr>
<td></td>
<td>(2.67)</td>
<td>(14.53)</td>
<td>(11.51)</td>
</tr>
<tr>
<td>Income Squared</td>
<td>-0.056</td>
<td>-0.002</td>
<td>-0.052</td>
</tr>
<tr>
<td></td>
<td>(0.018)</td>
<td>(0.064)</td>
<td>(0.050)</td>
</tr>
<tr>
<td>Prop. Black pupils</td>
<td>3.84</td>
<td>-7.25</td>
<td>-2.38</td>
</tr>
<tr>
<td></td>
<td>(0.89)</td>
<td>(13.43)</td>
<td>(13.41)</td>
</tr>
<tr>
<td>Prop. School age</td>
<td>-10.59</td>
<td>175.15</td>
<td>-27.76</td>
</tr>
<tr>
<td></td>
<td>(7.50)</td>
<td>(102.3)</td>
<td>(96.72)</td>
</tr>
</tbody>
</table>

(Continued)

Table 7: G2SLS Estimates with Secondary Black Weights Matrix
(Standard errors in parentheses)
<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-1.95</td>
<td>-161.8</td>
<td>-9.40</td>
<td>27.10</td>
<td>18.14</td>
</tr>
<tr>
<td></td>
<td>(4.87)</td>
<td>(91.77)</td>
<td>(80.60)</td>
<td>(72.07)</td>
<td>(70.15)</td>
</tr>
<tr>
<td>State Grants</td>
<td>0.141</td>
<td></td>
<td>0.366</td>
<td>0.362</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td></td>
<td>(0.025)</td>
<td>(0.025)</td>
<td></td>
</tr>
<tr>
<td>Reg. Economy</td>
<td>0.259</td>
<td></td>
<td>19.58</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.83)</td>
<td></td>
<td>(10.79)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAR 1st Error</td>
<td>0.407</td>
<td>-0.622</td>
<td>-0.602</td>
<td>-0.319</td>
<td>-0.331</td>
</tr>
<tr>
<td></td>
<td>(0.027)</td>
<td>(0.020)</td>
<td>(0.015)</td>
<td>(0.017)</td>
<td>(0.023)</td>
</tr>
<tr>
<td>SAR 2nd Error</td>
<td>-0.740</td>
<td>-1.54</td>
<td>-1.49</td>
<td>-1.24</td>
<td>-0.811</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.002)</td>
<td>(0.014)</td>
<td>(0.014)</td>
<td>(0.009)</td>
</tr>
<tr>
<td>Model Equation</td>
<td>SAR $\lambda_1$</td>
<td>TAR $\rho_1$</td>
<td>STAR $\delta_1$</td>
<td>Test Value</td>
<td>SAR $\lambda_2$</td>
</tr>
<tr>
<td>----------------</td>
<td>----------------</td>
<td>--------------</td>
<td>-----------------</td>
<td>------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Table 3.1 Col (3)</td>
<td>0.689</td>
<td>0.350</td>
<td>-0.143</td>
<td>0.099</td>
<td>(0.055)</td>
</tr>
<tr>
<td>Table 3.1 Col (4)</td>
<td>0.674</td>
<td>0.335</td>
<td>-0.117</td>
<td>0.109</td>
<td>(0.051)</td>
</tr>
<tr>
<td>Table 3.1 Col (5)</td>
<td>0.515</td>
<td>0.186</td>
<td>-0.102</td>
<td>-0.006</td>
<td>(0.041)</td>
</tr>
<tr>
<td>Table 3.1 Col (6)</td>
<td>0.498</td>
<td>0.183</td>
<td>-0.118</td>
<td>-0.026</td>
<td>(0.042)</td>
</tr>
<tr>
<td>Table 3.2 Col (3)</td>
<td>0.670</td>
<td>0.331</td>
<td>-0.125</td>
<td>0.097</td>
<td>(0.055)</td>
</tr>
<tr>
<td>Table 3.2 Col (4)</td>
<td>0.667</td>
<td>0.314</td>
<td>-0.110</td>
<td>0.099</td>
<td>(0.051)</td>
</tr>
<tr>
<td>Table 3.2 Col (5)</td>
<td>0.508</td>
<td>0.174</td>
<td>-0.094</td>
<td>-0.005</td>
<td>(0.041)</td>
</tr>
<tr>
<td>Table 3.2 Col (6)</td>
<td>0.492</td>
<td>0.173</td>
<td>-0.106</td>
<td>-0.021</td>
<td>(0.042)</td>
</tr>
<tr>
<td>Table 3.3 Col (3)</td>
<td>0.682</td>
<td>0.334</td>
<td>-0.141</td>
<td>0.087</td>
<td>(0.054)</td>
</tr>
<tr>
<td>Table 3.3 Col (4)</td>
<td>0.676</td>
<td>0.311</td>
<td>-0.123</td>
<td>0.087</td>
<td>(0.051)</td>
</tr>
<tr>
<td>Table 3.3 Col (5)</td>
<td>0.513</td>
<td>0.174</td>
<td>-0.096</td>
<td>-0.006</td>
<td>(0.041)</td>
</tr>
<tr>
<td>Table 3.3 Col (6)</td>
<td>0.498</td>
<td>0.173</td>
<td>-0.108</td>
<td>-0.022</td>
<td>(0.042)</td>
</tr>
</tbody>
</table>

Table 8: Testing Nonlinear Restriction on Structural Parameters
(Standard errors in parentheses)
<table>
<thead>
<tr>
<th>Model Variable</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAR Lag</td>
<td>0.498</td>
<td>0.240</td>
<td>0.058</td>
</tr>
<tr>
<td></td>
<td>(0.042)</td>
<td>(0.050)</td>
<td>(0.051)</td>
</tr>
<tr>
<td>TAR Lag</td>
<td>0.183</td>
<td>0.138</td>
<td>0.121</td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.025)</td>
<td>(0.024)</td>
</tr>
<tr>
<td>STAR Lag</td>
<td>-0.118</td>
<td>-0.184</td>
<td>-0.157</td>
</tr>
<tr>
<td></td>
<td>(0.041)</td>
<td>(0.040)</td>
<td>(0.040)</td>
</tr>
<tr>
<td>Pupil Density</td>
<td>-2.68</td>
<td>-1.97</td>
<td>-1.663</td>
</tr>
<tr>
<td></td>
<td>(1.06)</td>
<td>(1.08)</td>
<td>(1.05)</td>
</tr>
<tr>
<td>Income</td>
<td>28.34</td>
<td>5.33</td>
<td>7.76</td>
</tr>
<tr>
<td></td>
<td>(10.55)</td>
<td>(10.89)</td>
<td>(10.62)</td>
</tr>
<tr>
<td>Income Squared</td>
<td>-0.069</td>
<td>0.017</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>(0.042)</td>
<td>(0.048)</td>
<td>(0.046)</td>
</tr>
<tr>
<td>Prop. Black pupils</td>
<td>55.83</td>
<td>41.47</td>
<td>43.28</td>
</tr>
<tr>
<td></td>
<td>(11.26)</td>
<td>(11.10)</td>
<td>(10.67)</td>
</tr>
<tr>
<td>Prop. School age</td>
<td>-46.44</td>
<td>-62.88</td>
<td>-88.44</td>
</tr>
<tr>
<td></td>
<td>(77.30)</td>
<td>(71.90)</td>
<td>(70.11)</td>
</tr>
<tr>
<td>Prop. Elderly</td>
<td>13.84</td>
<td>42.53</td>
<td>65.84</td>
</tr>
<tr>
<td></td>
<td>(65.11)</td>
<td>(58.78)</td>
<td>(54.87)</td>
</tr>
<tr>
<td>State Grants</td>
<td>0.360</td>
<td>0.306</td>
<td>0.318</td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.025)</td>
<td>(0.024)</td>
</tr>
<tr>
<td>Reg. Economy</td>
<td>20.22</td>
<td>-3.62</td>
<td>-5.92</td>
</tr>
<tr>
<td></td>
<td>(9.81)</td>
<td>(9.84)</td>
<td>(9.44)</td>
</tr>
</tbody>
</table>

(Continued)

Table 9: G2SLS Estimates with first-order geographic weights matrix
(Standard errors in parentheses)
Table 9: Continued

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>143.37</td>
<td>16.57</td>
</tr>
<tr>
<td></td>
<td>(18.59)</td>
<td>(23.60)</td>
</tr>
<tr>
<td>Time Dummy2</td>
<td>308.86</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(25.89)</td>
<td></td>
</tr>
<tr>
<td>Time Dummy3</td>
<td>188.77</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(30.84)</td>
<td></td>
</tr>
<tr>
<td>SAR Error</td>
<td>-0.355</td>
<td>-0.144</td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.025)</td>
</tr>
</tbody>
</table>
Table 10: G2SLS Estimates with Secondary Income Weights Matrix  
(Standard errors in parentheses)

<table>
<thead>
<tr>
<th>Variable</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>IV</td>
<td>Time</td>
<td>Time</td>
</tr>
<tr>
<td>SAR 1\textsuperscript{st} Lag</td>
<td>0.492</td>
<td>0.239</td>
<td>0.0813</td>
</tr>
<tr>
<td></td>
<td>(0.042)</td>
<td>(0.049)</td>
<td>(0.050)</td>
</tr>
<tr>
<td>SAR 2\textsuperscript{nd} Lag</td>
<td>0.124</td>
<td>0.153</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td>(0.141)</td>
<td>(0.0142)</td>
<td>(0.0142)</td>
</tr>
<tr>
<td>TAR Lag</td>
<td>0.173</td>
<td>0.131</td>
<td>0.114</td>
</tr>
<tr>
<td></td>
<td>(0.023)</td>
<td>(0.025)</td>
<td>(0.024)</td>
</tr>
<tr>
<td>STAR 1\textsuperscript{st} Lag</td>
<td>-0.106</td>
<td>-0.204</td>
<td>-0.155</td>
</tr>
<tr>
<td></td>
<td>(0.040)</td>
<td>(0.043)</td>
<td>(0.044)</td>
</tr>
<tr>
<td>STAR 2\textsuperscript{nd} Lag</td>
<td>-0.132</td>
<td>-0.349</td>
<td>-0.033</td>
</tr>
<tr>
<td></td>
<td>(0.155)</td>
<td>(0.0156)</td>
<td>(0.0156)</td>
</tr>
<tr>
<td>Model Variable</td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>----------------</td>
<td>---------</td>
<td>------------</td>
<td>------------</td>
</tr>
<tr>
<td>IV</td>
<td>0.498</td>
<td>0.256</td>
<td>0.097</td>
</tr>
<tr>
<td>Time Trend</td>
<td>0.256</td>
<td>0.155</td>
<td>0.1172</td>
</tr>
<tr>
<td>Time Dummy</td>
<td>0.097</td>
<td>0.027</td>
<td>(0.024)</td>
</tr>
<tr>
<td>SAR 1st Lag</td>
<td>0.498</td>
<td>0.256</td>
<td>0.097</td>
</tr>
<tr>
<td>SAR 2nd Lag</td>
<td>0.153</td>
<td>0.155</td>
<td>0.027</td>
</tr>
<tr>
<td>TAR Lag</td>
<td>0.173</td>
<td>0.133</td>
<td>0.1172</td>
</tr>
<tr>
<td>STAR 1st Lag</td>
<td>-0.108</td>
<td>-0.173</td>
<td>-0.154</td>
</tr>
<tr>
<td>STAR 2nd Lag</td>
<td>-0.166</td>
<td>-0.168</td>
<td>-0.047</td>
</tr>
</tbody>
</table>

Table 11: G2SLS Estimates with Secondary Income Weights Matrix
(Standard errors in parentheses)
LIST OF REFERENCES


138


