Dynamic regional expenditure competition in China: based on a SAR model with time varying spatial coefficients

Juncong Guo, Xi Qu

October 7, 2018

Abstract

We analyze county-level governments expenditure competition in China. To identify the dynamics of the strength of this competition, we construct a spatial autoregressive panel data model with time varying spatial coefficients. For estimation, we propose a two-stage least squares (2SLS) method and the maximum likelihood estimation (MLE). Finite sample properties of these two methods are investigated in a Monte Carlo simulation. After applying these steps to county-level expenditure data from 2001 to 2016, the empirical findings show that spatial coefficients reflecting the strength of competition gradually decrease until 2008 and then remain stable. Some hypothetical explanations are given to link this phenomenon to economic events, public policies and economic performance.

JEL classification: C23, E62, H72, H77, O43

Keywords: expenditure competition, spatial autoregressive, time varying spatial coefficients
1 Introduction

After the reform and opening from 1978, the world has witnessed the spectacular growth of China’s economy. The GDP of China increased to 74,059.9 billion (CNY) from 367.9 billion in 1979-2016, about 15% nominal growth rate and 9.6% real growth rate compared with 5.7% and 2.9% of the whole world in the same period. Corresponding to the “China miracle”, China’s institutions look ill-suited with many shortcomings which seems impossible to support the rapid development according to classical economic and political theory. That is called the “China puzzle”. For the past decades, many scholars try to crack the puzzle and have gradually made a consensus that it is the regionally decentralized authoritarian (RDA) system containing decentralized economic governance and centralized political governance that leads to the “China miracle” (Xu, 2011). Most of the existing literature focusing on explaining China’s growth is around these two aspects. One distinguished branch of research links China’s economic growth to fiscal decentralization like Qian et al. (1996, 1998). Another important branch ascribes economic performance to tournament competition among regional government leaders like Li and Zhou (2005) and Chen et al. (2005). These two branches which we will discuss in section 2 both regard economic growth as a result but driven by different incentives-maximizing fiscal revenues and seeking political promotion respectively. Even though different motivations, tax competition and expenditure competition among regional governments are the most commonly used approaches to achieve growth. Hence, many scholars contribute themselves to these two kinds of competition.

In this paper, we focus on regional expenditure competition in China, which is a form of decentralized economic governance and incentivised by centralized political governance, and is usually nested in the branch of fiscal decentralization. We will use spatial econometrics method to identify this kind of competition. However, different from past study by traditional spatial autoregressive (SAR) models like Yu et al. (2016), we construct a SAR model with
time varying spatial coefficients to identify regional governments expenditure competition in different time. Cheung and Coase (2008) think regional competition in China is at county-level because counties are the lowest-level administrative units which have the power to decide the use of lands. Therefore, we choose to analyze regional expenditure competition at county-level. Through data analysis and model estimation, we want to discover the dynamics of the regional competition in China, especially before and after the financial crisis in 2008, and try to give some explanations for its dynamic pattern.

The outline of the paper is as follows. Section 2 is the literature review. Literature on regional competition in China and our contribution are elaborated in subsection 2.1. Literature on spatial autoregressive panel data models and our contribution are presented in subsection 2.2. Section 2 give the model specification and estimation details for 2SLS and MLE. Monte Carlo simulation is also put in the last of this section. Section 4 is our empirical study, including the introduction of data in subsection 4.1, model setting in subsection 4.2, estimation results in subsection 4.3 and some further discussion in subsection 4.4. Section 5 concludes.

2 Literature review

2.1 Regional competition in China

The regional competition in China is based on the RDA regime which is characterized as a combination of economic regional decentralization and political centralization (Xu, 2011). For economic regional decentralization, fiscal decentralization is the most important aspect which has been studied by many scholars in the past decades. Some researchers believe that this fiscal decentralization can be seen as market-preserving fiscal federalism (Montinola et al., 1995; Qian and Weingast, 1996, 1997; Qian and Roland, 1998; Jin et al., 2005), while Cai
and Treisman (2004, 2005) think the fiscal federalism may not be market-preserving because it corrodes the capacities of central state. However, Zhang (2012) points out that the fiscal reform in 1994 should be considered when studying China’s fiscal decentralization. Before the reform, the central government and regional governments operated revenue-sharing contracts called “chengbao” which caused incentive incompatibility. Therefore, regional governments was motivated to conceal their fiscal revenues (Wong, 2005). After the reform, a new tax system was established, with tax revenues and tax sources distributed between the central and regional governments. The reform corrects the overly fiscal decentralization in the past and reaches incentive compatibility. Meanwhile, the reform relatively decreases revenues of regional governments but reserves their expenditure duties. That strengthens budget constraints of regional governments and make them have to pursue revenues maximization to ensure expenditure (Zhang, 2012). This is one reason that regional governments support economic growth to expand tax bases via two main ways-tax competition and expenditure competition. Lin and Liu (2000) conclude that fiscal decentralization is indeed good for economic growth via empirical study.

Political centralization is the key feature of the RDA regime of China, which is fundamentally different from federal systems in federal states. In China, regional leaders are appointed by upper-level governments not by regional elections. Upper-level governments have the power to decide the promotion of regional leaders by evaluating their performance based largely on regional economic growth. That gives regional leaders strong incentives to participate tournament competition and try to do better than other regional leaders in promoting economic growth (Li and Zhou, 2005; Chen et al., 2005). Zhang and Gao (2008) further discuss the relationship between economic growth and the regime in China. In this context, tax competition and expenditure competition become consensuses among regional governments again. Notice that, China’s particular political arrangement makes the mech-
anism of public expenditure different from Tiebout model (Tiebout, 1956) and Western yardstick competition (Besley and Case, 1992).

Therefore, there are two channels which lead regional governments to pursue economic growth by regional competition in China. Economic decentralization after the fiscal reform in 1994 makes regional governments try to increase their fiscal revenues, so they have to drive economic growth in order to expand tax bases. That is an indirect channel. Another channel from political centralization makes regional leaders pursue economic performance directly because it is about their political destinies. These two channels both make regional governments perform tax competition and expenditure competition to push economy.

Tax competition is “race to bottom” manipulating the effective tax rate via offering tax incentives (e.g., tax exemptions, tax breaks, and preferential tax rates) to attract more taxpayers especially foreign investors (Xu, 2011) in order to expand tax bases. Liu and Martinez-Vazquez (2014) provide empirical evidence of tax competition among provincial governments in China by using a panel of provincial-level data in 1993-2007. Long et al. (2014) find the existence of spatial tax competition among Chinese counties by using county-level panel data from 2000 to 2006. Regional governments also want to attract investors and companies by providing better public service, therefore, they tend to increase fiscal expenditure. Zhang et al. (2007) point out that regional governments prefer to invest fixed assets like infrastructure rather than spending money on medical and education because the former directly helps to attract investments. That explains why China enjoys better physical infrastructure. To our best knowledge, even though there are some empirical studies focusing on expenditure competition (e.g., Baicker, 2005; Solé-Ollé, 2006; Isen, 2014), few literature studies expenditure competition in China background, not to mention trying to identify its

---

1Jia et al. (2014) give some empirical evidences by using county-level dataset from 1997 to 2006.
2Li and Zhou (2009), Guo and Jia (2009), etc. published in Chinese journal give some evidence by using provincial dataset. Yu et al.(2016) focus on region strategic interaction using city-level total investment data.
change over time. We try to fill the gap in this paper by constructing a spatial autoregressive model (SAR) with time varying spatial coefficients and using county-level dataset.

2.2 SAR panel data models

Spatial econometrics consists of econometric techniques dealing with spatial dependence (weak dependence) which can be compared with serial correlation and common factors (strong dependence). From the cross-sectional SAR model proposed by Cliff and Ord (1973) and classical quasi-maximum likelihood estimation (QMLE) method proposed by Lee (2004), spatial econometrics has been extended to panel data models (Elhorst, 2003; Baltagi et al., 2003). For the SAR panel data model, Lee and Yu (2010a) establish QMLE method for SAR panel data models with individual fixed effects:

\[
y_{nt} = \lambda W_n y_{nt} + X_{nt} \beta + u_n + \varepsilon_{nt},
\]

where \( y_{nt} \) is \( n \times 1 \)-dimension dependent variable at time \( t \); \( W_n \) is a \( n \times n \) spatial weights matrix; \( X_{nt} \) is a \( n \times k \) matrix containing \( k \) independent variables at time \( t \); \( u_n \) is a \( n \times 1 \) vector for individual fixed effects; \( \varepsilon_{nt} \) is a \( n \times 1 \) error term vector. After that, estimation methods for spatial dynamic panel data model with both time and individual effects (Lee and Yu, 2010b), for spatial dynamic panel data model with time varying spatial weights matrices (Lee and Yu, 2012) and for spatial dynamic panel data model with endogenous time varying spatial weights matrices (Qu et al., 2017) are be proposed successively. Recently, scholars try to make new explorations on spatial panel data models. Billé and Catania (2018) contribute themselves to discuss spatial weights matrices. Traditional ways to construct spatial weights matrices are using different well-known criteria\(^3\) based on the first law of geography (Tobler,

\(^3\)For example, contiguity criteria, k-nearest neighbors, geographic distances, economic distances, etc.)
1970), but these criteria are usually be criticized because of their arbitrary settings. Billé and Catania (2018) propose a new spatial panel data model without fixed effects but with time varying spatial weights matrices allowing for a general parameterization of the spatial matrix, such as inverse distances matrix with an unknown time varying distance decay parameter. Their model and estimation method are more suitable for the case with large time dimension and small cross-section dimension. Parameters of spatial matrices they get can be used to measure the strength of connections among different units under a fixed spatial coefficient. Blasques (2016) extend the static spatial panel data model by introducing time varying spatial coefficients also without fixed effects. In this case, they can sum up all the likelihood function in different time to get the aggregated likelihood function when using MLE. Our new exploration is also concentrated on spatial coefficients. We construct a static SAR panel data model with individual fixed effects and exogenous time varying spatial weights matrices but allowing for time varying spatial coefficients. When considering fixed effects, estimations will become complicated when we try to remove fixed effects terms via taking demean or difference. Contrast to Billé and Catania (2018), our model and estimation method are more suitable for the case with small time dimension and large cross-section dimension, and time varying spatial coefficients we get can be used to measure the strength of spillover effects among different units under given spatial weights matrices. In theory, the varying of spatial coefficients can be absorbed by time varying spatial weights matrices, then the model will degrade into Lee and Yu (2012)’s model, or a general form from Billé and Catania (2018). However, in empirical study, it is difficult to know how to set spatial weights matrices to achieve this absorption. Therefore, proposing this model and giving estimation methods have great empirical value.
3 Econometric model and estimation

To identify the varying spatial coefficients of county-level regional governments which reflect the strength of competition in different years, we build a spatial autoregressive (SAR) model with time varying spatial coefficients and spatial weights matrices.

\[ y_{nt} = \lambda_t W_{nt} y_{nt} + X_{nt} \beta + u_n + \varepsilon_{nt}, \]  

where \( y_{nt} \) is \( n \times 1 \)-dimension dependent variable at time \( t \); \( W_{nt} \) is a \( n \times n \) spatial weights matrix at time \( t \); \( X_{nt} \) is a \( n \times k \) matrix containing \( k \) independent variables at time \( t \); \( u_n \) is a \( n \times 1 \) vector for individual fixed effects; \( \varepsilon_{nt} \) is a \( n \times 1 \) error term vector. We focus on short panel data, so time fixed effects or the time trend can be controlled in \( X_{nt} \). Different from traditional spatial panel data model, we allow spatial coefficients \( \lambda_t \) and spatial weights matrices \( W_{nt} \) varying over time \( t \). The stationary requirement for spatial coefficients is that \( |\lambda_t| < 1 \) for row-normalized spatial weight matrices.

For model estimation, 2SLS and MLE are the two popular methods for spatial models. For our time varying spatial coefficients model, we also provide these two estimation methods which are different from methods for traditional spatial panel data models.

3.1 2SLS

To carry out 2SLS estimation conveniently, we rewrite equation (3.1) into individual level,

\[ y_{it} = \lambda_t z_{it} + x_{it} \beta + u_i + \varepsilon_{it}, \]  

where \( z_{it} = \sum_{j \neq i} w_{ijt} y_{jt} \). To eliminate the individual fixed effect \( u_i \), we use demeaned equation
\[
y_{it} - \overline{y}_i = \lambda_t z_{it} - \frac{1}{T} \sum_{t=1}^{T} \lambda_t z_{it} + (x_{it} - \overline{x}_i)\beta + \varepsilon_{it} - \overline{\varepsilon}_i,
\]  
(3.3)

where \( \overline{y}_i = \sum_{t=1}^{T} y_{it}/T \), \( \overline{x}_i \) and \( \overline{\varepsilon}_i \) similarly. Then, in the matrix form we can write it as

\[
\begin{pmatrix}
y_{11} - \overline{y}_1 \\
\vdots \\
y_{n1} - \overline{y}_n \\
y_{1T} - \overline{y}_1 \\
\vdots \\
y_{nT} - \overline{y}_n
\end{pmatrix}_{nT \times 1} = \frac{1}{T} \begin{pmatrix}
(T - 1)z_{11} & -z_{12} & \cdots & -z_{1T} \\
\vdots & \vdots & \ddots & \vdots \\
(T - 1)z_{n1} & -z_{n2} & \cdots & -z_{nT} \\
\vdots & \vdots & \ddots & \vdots \\
-z_{11} & -z_{12} & \cdots & (T - 1)z_{1T} \\
\vdots & \vdots & \ddots & \vdots \\
-z_{n1} & -z_{n2} & \cdots & (T - 1)z_{nT}
\end{pmatrix}_{nT \times T} \begin{pmatrix}
\lambda_1 \\
\lambda_2 \\
\vdots \\
\lambda_T
\end{pmatrix}_{T \times 1}
\]  
(3.4)

\[
+ \begin{pmatrix}
x_{11} - \overline{x}_1 \\
\vdots \\
x_{n1} - \overline{x}_n \\
x_{1T} - \overline{x}_1 \\
\vdots \\
x_{nT} - \overline{x}_n
\end{pmatrix}_{nT \times k} \begin{pmatrix}
\varepsilon_{11} - \overline{\varepsilon}_1 \\
\vdots \\
\varepsilon_{n1} - \overline{\varepsilon}_n \\
\varepsilon_{1T} - \overline{\varepsilon}_1 \\
\vdots \\
\varepsilon_{nT} - \overline{\varepsilon}_n
\end{pmatrix}_{nT \times 1}
\]

Denote equation (3.4) as

\[
\tilde{y} = \tilde{Z}\lambda + \tilde{X}\beta + \tilde{\varepsilon},
\]  
(3.5)
where \( \tilde{Z} \) can be seen as endogenous variables. If we can find a good instrumental variables–\( H \), 2SLS can be used to estimate the equation. Denote equation (3.5) above as

\[
\tilde{y} = \tilde{K} \delta + \tilde{\varepsilon},
\]

where \( \tilde{K} = (\tilde{Z}, \tilde{X}) \), \( \delta = (\lambda', \beta') \). The 2SLS estimator of \( \delta \) is

\[
\hat{\delta} = (K'PX)^{-1}K'Py,
\]

where \( P = H(H'H)^{-1}H' \). For the asymptotic distribution of \( \delta \), we have

\[
\sqrt{nT}(\hat{\delta} - \delta) \overset{d}{\to} N(0, \Sigma),
\]

where \( \Sigma \) can be estimated by

\[
\hat{\sigma}^2 \left[ \frac{1}{nT} (K'PK)^{-1}K'PVP'X(K'PK)^{-1} \right],
\]

and \( \hat{\sigma}^2 = \frac{1}{nT-n} (y - K\hat{\delta})'(y - K\hat{\delta}) \). Notice that \( \tilde{\varepsilon}_{it} \) is not independent identically distributed (i.i.d.) any more because we take demean for all variables at time level. Assume \( \varepsilon_{it} \) is i.i.d. \( (0, \sigma^2) \), then we have \( var(\tilde{\varepsilon}_{it}) = var(\varepsilon_{it} - \frac{1}{T} \sum_{t=1}^{T} \varepsilon_{it}) = (1 - \frac{1}{T})\sigma^2 \), \( cov(\tilde{\varepsilon}_{it}, \tilde{\varepsilon}_{is})_{|t \neq s} = -\frac{1}{T}\sigma^2 \) and \( cov(\tilde{\varepsilon}_{it}, \tilde{\varepsilon}_{js})_{|i \neq j} = 0 \). Hence, \( V \) can be calculated by \( \psi \otimes I_n \), where \( \psi \) is

\[
\begin{pmatrix}
(1 - \frac{1}{T})\sigma^2 & -\frac{1}{T}\sigma^2 & \cdots & -\frac{1}{T}\sigma^2 \\
-\frac{1}{T}\sigma^2 & (1 - \frac{1}{T})\sigma^2 & \cdots & -\frac{1}{T}\sigma^2 \\
\vdots & \vdots & \ddots & \vdots \\
-\frac{1}{T}\sigma^2 & -\frac{1}{T}\sigma^2 & \cdots & (1 - \frac{1}{T})\sigma^2
\end{pmatrix}_{T \times T}.
\]
Kelejian and Prucha (1998) suggest that the instrument set should be kept to a low order to avoid linear dependence and retain full column rank for the matrix of instruments and advise instrument set \([X, WX, W^2 X]\). In our model, instrumental variables–\(H\) can be constructed by \(\tilde{X}, W\tilde{X}, W^2\tilde{X}\), etc., where \(W\) can be gotten from stacking all \(W_{nt}\) in chronological order according to equation (3.5) or (3.6). Notice here, we need ensure that the column of \(H\) is no less than the column of \(\tilde{Z}\) which equals \(T\) to avoid under identification.

If the number of independent variables is larger than \(T\), \([\tilde{X}, W\tilde{X}]\) as instrument set is enough, else \(W^2\tilde{X}\) must be added into \(H\). Hence, the consistency of 2SLS estimators is depended on enough exogenous independent variables. That makes its scope of application limited.

### 3.2 MLE

For MLE, coming back to equation (3.1), assume errors are i.i.d. \(N(0, \sigma^2)\), then the likelihood function can be written as

\[
L_0 = -\frac{nT}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{t=1}^{T} (y_{nt} - \lambda_t W_{nt} y_{nt} - X_{nt} \beta - u_n)' (y_{nt} - \lambda_t W_{nt} y_{nt} - X_{nt} \beta - u_n) + \sum_{t=1}^{T} \ln |I_n - \lambda_t W_{nt}|.
\]  

(3.7)

However, individual fixed effects \(u_n\) in this likelihood function cannot be identified consistently, so we concentrate out \(u_n\) by \(\hat{u}_n = \frac{1}{T} \sum_{t=1}^{T} (y_{nt} - \lambda_t W_{nt} y_{nt} - X_{nt} \beta)\) and then get the concentrated likelihood function

\[
L_1 = -\frac{nT}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{t=1}^{T} (\tilde{y}_{nt} - \lambda_t \tilde{W}_{nt} \tilde{y}_{nt} - \tilde{X}_{nt} \beta)' (\tilde{y}_{nt} - \lambda_t \tilde{W}_{nt} \tilde{y}_{nt} - \tilde{X}_{nt} \beta) + \sum_{t=1}^{T} \ln |I_n - \lambda_t W_{nt}|,
\]  

(3.8)

where for any vector or matrix \(Z_{nt}\), define \(\tilde{Z}_{nt} = Z_{nt} - \frac{1}{T} \sum_{t=1}^{T} Z_{nt}\). Since first order conditions of the concentrated likelihood function does not have zero expectations at true values, we adjust likelihood function as equation (3.9) whose first order conditions with zero expecta-
\[ L(\theta) = -\frac{n(T-1)}{2} \ln \sigma^2 - \frac{1}{2 \sigma^2} \sum_{t=1}^{T} (\widetilde{y}_{nt} - W_{nt} \lambda_t y_{nt} - \bar{X}_{nt} \beta)' (\widetilde{y}_{nt} - W_{nt} \lambda_t y_{nt} - \bar{X}_{nt} \beta) + \frac{T-1}{T} \sum_{t=1}^{T} \ln |I_n - \lambda_t W_{nt}|, \]

where \( \theta = (\lambda', \beta', \sigma^2)' \), \( \lambda = (\lambda_1, \lambda_2, ..., \lambda_T)' \). Then we can get\(^4\)

\[ \sqrt{n(T-1)} (\hat{\theta} - \theta_0) \overset{d}{\rightarrow} N(0, H(\theta_0)^{-1}V(\theta_0)H(\theta_0)^{-1}), \]

where \( H(\theta_0) \) and \( V(\theta_0) \) can be estimated by \( H(\hat{\theta}) \) and \( V(\hat{\theta}) \).

### 3.3 Monte Carlo simulation

In this section, we evaluate the 2SLS and MLE methods of our model with randomly generated sparse spatial weights. The data generating process is

\[ y_{nt} = (I_n - \lambda_t W_{nt})^{-1}(a_0 x_0 + a_1 x_{1t} + a_2 x_{2t} + u + v_t), \]

where \( t = 1, 2, ..., T; \) \( x_0 \) is a unit column vector with dimension \( n; \) \( x_{1t}, x_{2t} \) and \( v_t \) are column vectors with dimensions \( n \) (their elements obey the standard normal distribution); \( u \) is a column vector with dimension \( n \) indicating individual fixed effect (its elements obey the uniform distribution between 0 and 1); \( I_n \) is a \( n \times n \) identity matrix. \( W_{nt} \) are sparse matrices that have 5-10 elements of ones, and all other elements are zeros at random in each row. Particularly, the diagonal elements of \( W_{nt} \) are zero. We perform simulations for 2SLS and MLE methods with 1000 replications for each setting. Notice here, in fixed effect model, \( a_0 \) and \( u \) cannot be estimated consistently, so we don’t report them. The following tables show simulation results of different parameter settings. In table 1 where \( T = 3, \) when

\(^4\)The derivation process is in the Appendix.
the dimensions increase but the number of ones remains constant in each row in $W_{nt}$, the accuracy of the estimations increases and the standard errors of estimators become smaller and smaller. When sample size is 400, both estimation methods give estimates very close to the true values. In terms of the variance of estimators, we find that there is no obvious difference between the estimated standard error based on an asymptotic variance-covariance matrix and the empirical standard deviation in all cases. This result means our methods are accurate for estimating variance when the sample size is in the hundreds. Compared with 2SLS, estimators from MLE have smaller variances, which is consistent with the theory that MLE is more efficient than 2SLS. However, as mentioned above, table 2 shows that when $T$ is larger but only two independent variables, estimators from 2SLS will be seriously biased because lack of effective instrumental variables. Estimators from MLE are still consistent. That reminds us if we have no enough independent variables in empirical study, 2SLS may not be a good estimation method. But if we have enough independent variables and lots of observations which makes a high dimension of $W_{nt}$, 2SLS may be better than MLE because it enjoys a faster computing speed.
Table 1: Simulation with T=3, $\lambda_1=0.3$, 0.4, 0.5

<table>
<thead>
<tr>
<th></th>
<th>2SLS</th>
<th></th>
<th>MLE</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$n=100$</td>
<td>$n=200$</td>
<td>$n=400$</td>
<td>$n=100$</td>
</tr>
<tr>
<td>$\lambda_1=0.3$</td>
<td>0.304</td>
<td>0.297</td>
<td>0.301</td>
<td>0.280</td>
</tr>
<tr>
<td></td>
<td>(0.171)</td>
<td>(0.138)</td>
<td>(0.103)</td>
<td>(0.084)</td>
</tr>
<tr>
<td></td>
<td>[0.156]</td>
<td>[0.133]</td>
<td>[0.102]</td>
<td>[0.084]</td>
</tr>
<tr>
<td>$\lambda_2=0.4$</td>
<td>0.395</td>
<td>0.404</td>
<td>0.401</td>
<td>0.378</td>
</tr>
<tr>
<td></td>
<td>(0.291)</td>
<td>(0.198)</td>
<td>(0.054)</td>
<td>(0.089)</td>
</tr>
<tr>
<td></td>
<td>[0.268]</td>
<td>[0.182]</td>
<td>[0.054]</td>
<td>[0.093]</td>
</tr>
<tr>
<td>$\lambda_3=0.5$</td>
<td>0.504</td>
<td>0.498</td>
<td>0.500</td>
<td>0.477</td>
</tr>
<tr>
<td></td>
<td>(0.232)</td>
<td>(0.100)</td>
<td>(0.116)</td>
<td>(0.056)</td>
</tr>
<tr>
<td></td>
<td>[0.212]</td>
<td>[0.094]</td>
<td>[0.121]</td>
<td>[0.055]</td>
</tr>
<tr>
<td>$a_1=1$</td>
<td>1.000</td>
<td>0.999</td>
<td>1.001</td>
<td>1.001</td>
</tr>
<tr>
<td></td>
<td>(0.078)</td>
<td>(0.055)</td>
<td>(0.037)</td>
<td>(0.072)</td>
</tr>
<tr>
<td></td>
<td>[0.072]</td>
<td>[0.053]</td>
<td>[0.037]</td>
<td>[0.072]</td>
</tr>
<tr>
<td>$a_2=2$</td>
<td>1.999</td>
<td>2.001</td>
<td>2.000</td>
<td>1.999</td>
</tr>
<tr>
<td></td>
<td>(0.083)</td>
<td>(0.055)</td>
<td>(0.035)</td>
<td>(0.075)</td>
</tr>
<tr>
<td></td>
<td>[0.077]</td>
<td>[0.052]</td>
<td>[0.035]</td>
<td>[0.075]</td>
</tr>
</tbody>
</table>

Note: Estimated standard error based on an asymptotic variance-covariance matrix is in parentheses; and empirical standard deviation is in brackets.
Table 2: Simulation with T=5, $\lambda_i = 0.2, 0.3, 0.4, 0.5, 0.6$

<table>
<thead>
<tr>
<th></th>
<th>(n=100)</th>
<th>(n=200)</th>
<th>(n=400)</th>
<th>(n=100)</th>
<th>(n=200)</th>
<th>(n=400)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\lambda_1 = 0.2)</td>
<td>-3.806</td>
<td>-0.746</td>
<td>-1.720</td>
<td>0.176</td>
<td>0.191</td>
<td>0.195</td>
</tr>
<tr>
<td></td>
<td>(9504.156)</td>
<td>(693.843)</td>
<td>(670.372)</td>
<td>(0.094)</td>
<td>(0.056)</td>
<td>(0.042)</td>
</tr>
<tr>
<td></td>
<td>[86.322]</td>
<td>[20.866]</td>
<td>[58.048]</td>
<td>[0.093]</td>
<td>[0.053]</td>
<td>[0.043]</td>
</tr>
<tr>
<td>(\lambda_2 = 0.3)</td>
<td>4.797</td>
<td>0.760</td>
<td>-0.999</td>
<td>0.280</td>
<td>0.290</td>
<td>0.295</td>
</tr>
<tr>
<td></td>
<td>(5946.600)</td>
<td>(445.152)</td>
<td>(256.823)</td>
<td>(0.078)</td>
<td>(0.055)</td>
<td>(0.039)</td>
</tr>
<tr>
<td></td>
<td>[70.834]</td>
<td>[24.324]</td>
<td>[47.186]</td>
<td>[0.078]</td>
<td>[0.055]</td>
<td>[0.040]</td>
</tr>
<tr>
<td>(\lambda_3 = 0.4)</td>
<td>-0.706</td>
<td>-1.292</td>
<td>4.343</td>
<td>0.385</td>
<td>0.391</td>
<td>0.398</td>
</tr>
<tr>
<td></td>
<td>(2881.169)</td>
<td>(278.356)</td>
<td>(1406.886)</td>
<td>(0.051)</td>
<td>(0.054)</td>
<td>(0.032)</td>
</tr>
<tr>
<td></td>
<td>[29.492]</td>
<td>[41.398]</td>
<td>[90.293]</td>
<td>[0.052]</td>
<td>[0.053]</td>
<td>[0.032]</td>
</tr>
<tr>
<td>(\lambda_3 = 0.5)</td>
<td>1.391</td>
<td>-1.127</td>
<td>-0.633</td>
<td>0.487</td>
<td>0.494</td>
<td>0.496</td>
</tr>
<tr>
<td></td>
<td>(3182.214)</td>
<td>(522.120)</td>
<td>(332.460)</td>
<td>(0.047)</td>
<td>(0.033)</td>
<td>(0.030)</td>
</tr>
<tr>
<td></td>
<td>[31.921]</td>
<td>[53.303]</td>
<td>[27.559]</td>
<td>[0.047]</td>
<td>[0.032]</td>
<td>[0.030]</td>
</tr>
<tr>
<td>(\lambda_3 = 0.6)</td>
<td>1.368</td>
<td>1.940</td>
<td>-0.360</td>
<td>0.586</td>
<td>0.595</td>
<td>0.598</td>
</tr>
<tr>
<td></td>
<td>(6165.977)</td>
<td>(761.818)</td>
<td>(270.346)</td>
<td>(0.048)</td>
<td>(0.028)</td>
<td>(0.020)</td>
</tr>
<tr>
<td></td>
<td>[50.919]</td>
<td>[51.014]</td>
<td>[23.636]</td>
<td>[0.048]</td>
<td>[0.028]</td>
<td>[0.020]</td>
</tr>
</tbody>
</table>

Note: Estimated standard error based on an asymptotic variance-covariance matrix is in parentheses; and empirical standard deviation is in brackets.
4 Empirical study

4.1 Data

Taking the suggestion of Cheung and Coase (2008), we focus on county-level governments expenditure competition. The dependent variable is governments fiscal expenditure in one year. Referring to Baicker (2005) and based on the information we can get, we choose 8 control variables shown in table 3. The data source is Support System for China Statistics Application\(^5\) which collects data from China Statistical Yearbook (County-level). Our dataset is from 2001 to 2016 since we care about the new change of the regional expenditure competition in China. We delete district administrative units\(^6\) inside prefecture-level cities and some counties because of data missing. Those counties which have been merged in the target period are also eliminated. Then, there are 1996 counties in our sample with 49.1\% of GDP, 71.1\% of population and 90.3\% of area compared with the whole China in 2016. Table 4 shows the descriptive statistics of variables in 2001 and 2016 including means, standard deviation and average growth rate from 2001 to 2016. The last column reports the share of our sample compared with the whole state in 2001 and 2016 respectively. From the table, we can find that the fiscal expenditure is much larger than the fiscal income in regional governments and the growth rate of expenditure is larger than the growth rate of income, which means regional governments may be overspending and have increasing government debts under expenditure competition. The absolutely and relatively decreasing primary school students and middle school students contrast to the expanding total population reflects the phenomenon that regions in our sample are suffering from aging problem and the loss of young people. Figure 1 gives the quantile maps of county-level expenditure in 2001 and 2016. As shown in the figure, expenditure enjoys agglomeration with high-high and low-low

\(^{5}\)http://info.acmr.cn/index.aspx

\(^{6}\)They are also county-level.
patterns, which indicates the existence of regional competition.

Table 3: Variables and units

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E_t$</td>
<td>County-level governments fiscal expenditure in year $t$</td>
<td>100 million CNY</td>
</tr>
<tr>
<td>$Inc_t$</td>
<td>County-level governments fiscal expenditure in year $t$</td>
<td>100 million CNY</td>
</tr>
<tr>
<td>$Fst_t$</td>
<td>Gross domestic product value of primary industry in year $t$</td>
<td>100 million CNY</td>
</tr>
<tr>
<td>$Snd_t$</td>
<td>Gross domestic product value of secondary industry in year $t$</td>
<td>100 million CNY</td>
</tr>
<tr>
<td>$Trd_t$</td>
<td>Gross domestic product value of tertiary industry in year $t$</td>
<td>100 million CNY</td>
</tr>
<tr>
<td>$Pop_t$</td>
<td>Total registered population in year $t$</td>
<td>10 thousand</td>
</tr>
<tr>
<td>$Pst_t$</td>
<td>The number of primary school students in year $t$</td>
<td>10 thousand</td>
</tr>
<tr>
<td>$Mst_t$</td>
<td>The number of middle school students in year $t$</td>
<td>10 thousand</td>
</tr>
<tr>
<td>$Te_t$</td>
<td>Total governments fiscal expenditure in the whole country in year $t$</td>
<td>1 trillion CNY</td>
</tr>
</tbody>
</table>

Table 4: Descriptive statistics

<table>
<thead>
<tr>
<th>Variables</th>
<th>2001 S.D.</th>
<th>2016 S.D.</th>
<th>Growth rate</th>
<th>Share to the whole state</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E_t$</td>
<td>1.860</td>
<td>1.339</td>
<td>20.87%</td>
<td>19.64%, 33.93%</td>
</tr>
<tr>
<td>$Inc_t$</td>
<td>0.967</td>
<td>1.247</td>
<td>19.93%</td>
<td>11.78%, 18.48%</td>
</tr>
<tr>
<td>$Fst_t$</td>
<td>6.621</td>
<td>6.003</td>
<td>9.54%</td>
<td>85.26%, 81.46%</td>
</tr>
<tr>
<td>$Snd_t$</td>
<td>10.399</td>
<td>17.055</td>
<td>15.19%</td>
<td>41.55%, 58.06%</td>
</tr>
<tr>
<td>$Trd_t$</td>
<td>8.117</td>
<td>10.888</td>
<td>15.43%</td>
<td>35.45%, 36.38%</td>
</tr>
<tr>
<td>$Pop_t$</td>
<td>45.203</td>
<td>33.524</td>
<td>0.58%</td>
<td>70.70%, 71.13%</td>
</tr>
<tr>
<td>$Pst_t$</td>
<td>4.743</td>
<td>3.846</td>
<td>-2.34%</td>
<td>75.49%, 66.92%</td>
</tr>
<tr>
<td>$Mst_t$</td>
<td>2.764</td>
<td>2.243</td>
<td>-1.63%</td>
<td>69.66%, 64.42%</td>
</tr>
</tbody>
</table>
4.2 Model setting

We build a SAR model with time varying spatial coefficients like equation (3.1). The model is

\[ E_t = \lambda_t W E_t + X_{t-1} \beta_1 + Z_t \beta_2 + u + \varepsilon_t, \]

(4.1)

where \( u \) denotes individual fixed effects; \( t \) is from 1 to \( T \) (=16), representing year 2001 to 2016 respectively. \( X_{t-1} \) contains \( \text{Inc}_{t-1}, \text{Fst}_{t-1}, \text{Snd}_{t-1}, \text{Trd}_{t-1} \) and their square. Using lag terms is to avoid endogeneity problem. \( Z_t \) contains \( \text{Pop}_t, \text{Pst}_t, \text{Mst}_t \) and \( \text{Te}_t \), where adding \( \text{Te}_t \) which reflects the macroeconomic trend into regressors as time trend term is to avoid using too many dummies for controlling time fixed effects. In our empirical study, we use geographic information to construct spatial weights matrices-\( W \), which is constant over time. Compared with the general model in subsection 3.2, our empirical model is a special case.
For spatial weights matrices, we construct four matrices to do robustness check, denoting them as $W_1$, $W_2$, $W_3$ and $W_4$.\(^7\) $W_1$ is inverse distance matrix. Elements in $W_1$ are defined as $w_{ij} = 1/d_{ij}^2$ if $d_{ij} < 500\text{km}$, otherwise, $w_{ij} = 0$. $W_2$ and $W_3$ are based on contiguity. To be more specific, $W_2$ is first order adjacency matrix with $w_{ij} = 1$ if county $i$ and county $j$ are neighbors; 0, otherwise. $W_3$ is second order adjacency matrix with $w_{ij} = 1$ if county $i$ and county $j$ are neighbors or second neighbors (neighbors’ neighbors); 0, otherwise. $W_4$ is built according to administrative affiliation. Its elements defined as following:

$$w_{ij} = \begin{cases} 
3 & \text{county } i \text{ and county } j \text{ are in the same prefecture-level city}, \\
1 & \text{county } i \text{ and county } j \text{ are in different prefecture-level cities but in the same province}, \\
0 & \text{county } i \text{ and county } j \text{ are in different provinces}.
\end{cases}$$

As discuss in section 3, when the number of independent variables is much smaller than $T$, estimators of 2SLS will be seriously biased because lack of enough effective instruments. Therefore, in our empirical study, we choose the MLE methods to do estimation.

### 4.3 Results

Table 5 presents the results from MLE methods under different spatial weights matrices settings. For comparison, the first column reports the estimation results of equation (2.1) with a fixed spatial coefficient by using Lee and Yu (2010a)’s method under $W_1$. The sixteen significant spatial coefficients under every matrices setting are reported in figure 3. We also estimate spatial coefficients year by year via classical cross-sectional SAR model and report them in figure 2.

Table 5 shows that regional fiscal income, GDP of primary and tertiary industries and

\(^7\)For counties with enclaves or exclaves, we consider their main areas in which administrative centers locate.
the number of middle school students all have inverted “U” shape relationship with fiscal expenditure. However, considering maximal values of parabolas, these variables actually have positive relationship with fiscal expenditure. GDP of the secondary industry has no significant relationship with fiscal expenditure under most settings, but its square term shows positive pattern. Similarly, it seems that regional population has “U” shape relationship with fiscal expenditure, but when we consider minimal values of parabolas, population actually has positive effect on expenditure. These relationships above are all consistent with economic intuition. However, it is weird that the number of primary school students negatively affects county-level government expenditure considering its “U” shape and parabolas. This might be because some unobservable factors are correlated to the number of students\(^8\), but we cannot control these factors due to data constraint. To our relief, regional governments are not willing to spend much money on education because it cannot promote economic growth in a short run (Zhang et al., 2007). They just fulfill the minimum spending responsibility on education, so the education expenditure just accounts for a small ratio of the total expenditure\(^9\). Hence, we think the negative term will not significantly affect the estimation of regional total expenditure competition which is what we care most about.

\(^8\)The total expenditure including the education is increasing along economic growth but the number of students is decreasing because aging problem. That causes the negative relationship even though we have controlled the macroeconomic trend using a time trend term \(T_e\).

\(^9\)Unfortunately, we don’t have education expenditure data at county-level. Using expenditure of all subnational governments in 2012 as an example, the general education expenditure account for about 13% according to the State Statistical Bureau.
Table 5: MLE results under different spatial weights matrices settings

<table>
<thead>
<tr>
<th></th>
<th>$\hat{W}_1$</th>
<th>$W_1$</th>
<th>$W_2$</th>
<th>$W_3$</th>
<th>$W_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda$</td>
<td>0.208***</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Inc$</td>
<td>0.672***</td>
<td>0.667***</td>
<td>0.607***</td>
<td>0.618***</td>
<td>0.580***</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.011)</td>
<td>(0.011)</td>
<td>(0.011)</td>
<td>(0.011)</td>
</tr>
<tr>
<td>$Inc^2$</td>
<td>-0.001***</td>
<td>-0.001***</td>
<td>-0.001***</td>
<td>-4.794e-4***</td>
<td>-3.380e-4***</td>
</tr>
<tr>
<td></td>
<td>(5.584e-5)</td>
<td>(1.014e-4)</td>
<td>(6.962e-5)</td>
<td>(8.876e-5)</td>
<td>(9.835e-5)</td>
</tr>
<tr>
<td>$Fst$</td>
<td>0.261***</td>
<td>0.280***</td>
<td>0.239***</td>
<td>0.245***</td>
<td>0.238***</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.009)</td>
<td>(0.009)</td>
<td>(0.009)</td>
<td>(0.009)</td>
</tr>
<tr>
<td>$Fst^2$</td>
<td>-6.837e-4***</td>
<td>-0.001***</td>
<td>-4.986e-4***</td>
<td>-0.001***</td>
<td>-4.506e-4***</td>
</tr>
<tr>
<td>$Snd$</td>
<td>-0.001</td>
<td>0.001</td>
<td>-0.003*</td>
<td>-4.335-4</td>
<td>0.004***</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>$Snd^2$</td>
<td>1.254e-5***</td>
<td>1.126e-5***</td>
<td>1.714e-5***</td>
<td>1.632e-5***</td>
<td>1.648e-5***</td>
</tr>
<tr>
<td></td>
<td>(1.083e-6)</td>
<td>(1.963e-6)</td>
<td>(1.897e-6)</td>
<td>(1.914e-6)</td>
<td>(1.903e-6)</td>
</tr>
<tr>
<td>$Trd$</td>
<td>0.046***</td>
<td>0.045***</td>
<td>0.038***</td>
<td>0.042***</td>
<td>0.037***</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>$Trd^2$</td>
<td>-1.425e-5***</td>
<td>-1.294e-5***</td>
<td>-1.771e-5***</td>
<td>-2.328e-5***</td>
<td>-2.220e-5***</td>
</tr>
<tr>
<td></td>
<td>(3.822e-6)</td>
<td>(4.040e-6)</td>
<td>(3.871e-6)</td>
<td>(3.942e-6)</td>
<td>(3.915e-6)</td>
</tr>
<tr>
<td>$Pop$</td>
<td>0.126***</td>
<td>0.150***</td>
<td>0.120***</td>
<td>0.138***</td>
<td>0.147***</td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td>(0.017)</td>
<td>(0.016)</td>
<td>(0.017)</td>
<td>(0.017)</td>
</tr>
<tr>
<td>$Pop^2$</td>
<td>0.001***</td>
<td>0.001***</td>
<td>0.001***</td>
<td>0.001***</td>
<td>0.001***</td>
</tr>
<tr>
<td></td>
<td>(7.195e-5)</td>
<td>(7.453e-5)</td>
<td>(7.105e-5)</td>
<td>(7.239e-5)</td>
<td>(7.208e-5)</td>
</tr>
<tr>
<td>$Pst$</td>
<td>-0.691***</td>
<td>-0.785***</td>
<td>-0.603***</td>
<td>-0.636***</td>
<td>-0.718***</td>
</tr>
<tr>
<td></td>
<td>(0.025)</td>
<td>(0.027)</td>
<td>(0.004)</td>
<td>(0.026)</td>
<td>(0.026)</td>
</tr>
<tr>
<td>$Pst^2$</td>
<td>0.004***</td>
<td>0.005***</td>
<td>0.004***</td>
<td>0.004***</td>
<td>0.005***</td>
</tr>
<tr>
<td></td>
<td>(2.138e-4)</td>
<td>(2.243e-4)</td>
<td>(2.144e-4)</td>
<td>(2.185e-4)</td>
<td>(2.178e-4)</td>
</tr>
<tr>
<td>$Mst$</td>
<td>1.196***</td>
<td>1.254***</td>
<td>1.192***</td>
<td>1.257***</td>
<td>1.231***</td>
</tr>
<tr>
<td></td>
<td>(0.066)</td>
<td>(0.069)</td>
<td>(0.066)</td>
<td>(0.067)</td>
<td>(0.067)</td>
</tr>
<tr>
<td>$Mst^2$</td>
<td>-0.102***</td>
<td>-0.103***</td>
<td>-0.095***</td>
<td>-0.100***</td>
<td>-0.094***</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.005)</td>
<td>(0.005)</td>
<td>(0.005)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>$Te$</td>
<td>0.487***</td>
<td>0.476***</td>
<td>0.469***</td>
<td>0.480***</td>
<td>0.399***</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.012)</td>
<td>(0.009)</td>
<td>(0.009)</td>
<td>(0.010)</td>
</tr>
<tr>
<td>Likelihood</td>
<td>-8.484e+4</td>
<td>-5.281e+4</td>
<td>-5.162e+4</td>
<td>-5.210e+4</td>
<td>-5.186e+4</td>
</tr>
</tbody>
</table>

Note: ***significant at 1% level, **significant at 5% level, *significant at 10% level.
Figure 2 shows spatial coefficients estimated year by year via classical cross-sectional SAR model. Under different spatial weights matrices settings, the dynamic trends of regional expenditure competition have the same pattern. The strength of competition shows decline trend in 2001-2007 and 2009-2016, but there is dramatic rise in 2008. In 2008, a serious financial crisis happened to the world and then China government launched a four trillion fiscal stimulus plan. That maybe explain the sharp change of regional competition in 2008. Under financial crisis and fiscal stimulus, regional governments have stronger motivation to give impetus to economic growth via raising government spending competitively. However, when we use a SAR panel data model with time vary coefficients and consider fixed effects, the pattern become different as shown in figure 3. That means our model is effective and fixed effects cannot be overlooked. According to the figure, we find that the strength of regional competition goes down until 2008, and then remains stable around 0.3. There is no extreme rise in 2008.

Figure 2: Spatial coefficients estimated year by year
4.4 Further discussion

We estimate the dynamics of regional governments competition and find an interesting pattern. The strength of regional competition does not go down any more and start to become stable in 2008. How to explain this pattern? We propose three possible explanations. One is the financial crisis in 2008 which leads to strong pressure of economy. The second is fiscal stimulus in 2008, coming with the financial crisis. The third explanation may be the implementation of the new Labor Law in 2007 which is criticized by Cheung and Coase (2008). And the decreasing competition may explain why the growth rate of China economy become slow in recently years in some extent. Unfortunately, due to data constraints, it is hard to prove these possible explanations by econometric analysis. We can just see some correlations rather than causalities. On the other hand, the RDA system may be the cause of the “China miracle”, but drawbacks it arouses are discussed gradually, such as government debts (Li and Lin, 2011; Lu and Sun, 2013; Zhong and Lu, 2015) and market segmentation (Lu and Chen, 2009; Zhong and Lu, 2015). How to remain this regional competition to promote economy and overcome its shortcomings via appropriate reform is the new challenge of China
development.

5 Conclusion

In this paper, we focus on the dynamic change of county-level governments expenditure competition in China. We explain incentives of regional competition by introducing the mechanism of the RDA regime of China. Decentralized economic governance and centralized political governance are the key two parts in China’s particular institution. The fiscal reform in 1994 makes a incentive compatible contract between the central government and regional governments, which motivate regional governments to promote economic growth for increasing fiscal income. At the same time, the assessment mechanism for regional leaders attaches importance to leaders’ performance on regional GDP, which provides a strong incentive for leaders to promote economic development in order to improve political achievements. Based on these two aspects, competitive fiscal expenditure for pushing economy becomes one of regional governments’ common choices.

To identify the dynamics of the strength of expenditure competition, we propose a spatial autoregressive panel data model with fixed effects to estimate spatial coefficients. In this model, we allow spatial coefficients and spatial weights matrices to vary over time. We give two estimation methods-2SLS and MLE. The effect of 2SLS depends on the number of independent variables which are used to construct instruments. Inadequate instruments can leads to seriously biased estimators of 2SLS. MLE performs well even independent variables is not enough, but its computing speed is lower.

We empirically find an interesting dynamic pattern of the strength of expenditure competition. Before 2008, the strength shows decreasing trend, but after 2008 it become stable. We give some hypothetical explanations for this phenomenon and try to link it to the lower
GDP growth rate in recent years. However, due to data constraints especially the length of time series, it is hard to test these hypotheses via rigorous causal analyses. Nevertheless, the correlations may also give us some enlightenments for future reform.

References

Yingyi Qian, Chenggang Xu, Li’an Zhou, Steven N.S. Cheung, Ye’an Zhou


Qian, Y., Weingast, B.R., 1996. Chinas transition to markets: market-preserving feder-


Economic development and cultural change, 49(1), pp.1-21.


Appendix

The first-order derivatives are

\[ \frac{\partial L}{\partial \lambda_s} = \frac{1}{\sigma^2} (W_{ns}y_{ns})'\widetilde{\varepsilon}_{ns}(\theta) - \frac{T-1}{T} \text{tr}[(I - \lambda_sW_{ns})^{-1}W_{ns}], \]
\[ \frac{\partial L}{\partial \beta} = \frac{1}{\sigma^2} \sum_{t=1}^{T} \widetilde{X}'_{nt}\varepsilon_{nt}(\theta), \]
\[ \frac{\partial L}{\partial \sigma^2} = -\frac{n(T-1)}{2\sigma^4} + \frac{1}{2\sigma^4} \sum_{t=1}^{T} \varepsilon'_{nt}(\theta)\varepsilon_{nt}(\theta). \]

The second-order derivatives are

\[ \frac{\partial^2 L}{\partial \lambda_s^2} = -\frac{T-1}{\sigma^2 T} (W_{ns}y_{ns})'(W_{ns}y_{ns}) - \frac{T-1}{T} \text{tr}[(I - \lambda_sW_{ns})^{-1}W_{ns}(I - \lambda_sW_{ns})^{-1}W_{ns}], \]
\[ \frac{\partial^2 L}{\partial \lambda_s \partial \lambda_m} |_{s \neq m} = \frac{1}{\sigma^2 T} (W_{ns}y_{ns})'(W_{nm}y_{nm}), \]
\[ \frac{\partial^2 L}{\partial \beta^2} = -\frac{1}{\sigma^2} \sum_{t=1}^{T} \widetilde{X}'_{nt}\widetilde{X}_{nt}, \]
\[ \frac{\partial^2 L}{\partial \sigma^4} = n(T-1) - \frac{1}{\sigma^6} \sum_{t=1}^{T} \varepsilon'_{nt}(\theta)\varepsilon_{nt}(\theta), \]
\[ \frac{\partial^2 L}{\partial \lambda_s \partial \beta} = -\frac{1}{\sigma^2} (W_{ns}y_{ns})'\widetilde{X}_{ns}, \]
\[ \frac{\partial^2 L}{\partial \lambda_s \partial \sigma^2} = -\frac{1}{\sigma^4} (W_{ns}y_{ns})'\widetilde{\varepsilon}_{ns}(\theta), \]
\[ \frac{\partial^2 L}{\partial \beta \partial \sigma^2} = -\frac{1}{\sigma^4} \sum_{t=1}^{T} \widetilde{X}'_{nt}\varepsilon_{nt}(\theta). \]

Denote \((I - \lambda_tW_{nt})^{-1}W_{nt} = G_t(\lambda_t)\) and \((I - \lambda^0_tW_{nt})^{-1}W_{nt} = G_t\), where \(\lambda^0_t\) is the true value of \(\lambda_t\). We have
E\left( \frac{\partial^2 L(\theta_0)}{\partial \lambda_s^2} \right) = - \frac{T-1}{T} \text{tr}(G'_s G_s) - \frac{T-1}{\sigma_0^2 T} (X_{ns} \beta_0 + u_n)' G'_s G_s (X_{ns} \beta_0 + u_n),

E\left( \frac{\partial^2 L(\theta_0)}{\partial \lambda_s \partial \lambda_m} \right)_{s \neq m} = \frac{1}{\sigma_0^2 T} (X_{ns} \beta_0 + u_n)' G'_s G_m (X_{nm} \beta_0 + u_n),

E\left( \frac{\partial^2 L(\theta_0)}{\partial \beta^2} \right) = - \frac{1}{\sigma_0^2} \sum_{t=1}^T \widetilde{X}'_{nt} \widetilde{X}_{nt},

E\left( \frac{\partial^2 L(\theta_0)}{\partial \sigma^4} \right) = - \frac{n(T-1)}{2 \sigma_0^4},

E\left( \frac{\partial^2 L(\theta_0)}{\partial \lambda_s \partial \beta} \right) = - \frac{1}{\sigma_0^2} (X_{ns} \beta_0 + u_n)' G'_s \widetilde{X}_{ns},

E\left( \frac{\partial^2 L(\theta_0)}{\partial \lambda_s \partial \sigma^2} \right) = - \frac{T-1}{\sigma_0^2 T} \text{tr}(G_s),

E\left( \frac{\partial^2 L(\theta_0)}{\partial \beta \partial \sigma^2} \right) = 0,

where \theta_0 is the true value of \theta. Hence, the Hessian matrix is

\[ H(\theta_0) = \frac{1}{n(T-1)} E \begin{pmatrix} \frac{\partial^2 \ln L(\theta_0)}{\partial \lambda \partial \lambda'} & \frac{\partial^2 \ln L(\theta_0)}{\partial \lambda \partial \beta'} & \frac{\partial^2 \ln L(\theta_0)}{\partial \lambda \partial \sigma^2} \\ \frac{\partial^2 \ln L(\theta_0)}{\partial \lambda \partial \beta} & \frac{\partial^2 \ln L(\theta_0)}{\partial \beta \partial \beta'} & \frac{\partial^2 \ln L(\theta_0)}{\partial \beta \partial \sigma^2} \\ \frac{\partial^2 \ln L(\theta_0)}{\partial \lambda \partial \sigma^2} & \frac{\partial^2 \ln L(\theta_0)}{\partial \beta \partial \sigma^2} & \frac{\partial^2 \ln L(\theta_0)}{\partial \sigma^2 \partial \sigma^2} \end{pmatrix}. \]
To get the variance-covariance matrix, we calculate:

\[
\frac{\partial L}{\partial \lambda_s} \frac{\partial L}{\partial \lambda_s} = \frac{1}{\sigma^4} (W_{nsys})' \varepsilon_{ns}(\theta) \varepsilon_{ns}(\theta) W_{nsys} - \frac{2(T - 1)}{\sigma^2 T} (W_{nsys})' \varepsilon_{ns}(\theta) tr [(I - \lambda_s W_{ns})^{-1} W_{ns}]
\]
\[+ \frac{(T - 1)^2}{T^2} tr [(I - \lambda_s W_{ns})^{-1} W_{ns}] tr [(I - \lambda_s W_{ns})^{-1} W_{ns}],
\]

\[
\frac{\partial L}{\partial \lambda_s} \frac{\partial L}{\partial \lambda_m} = \frac{1}{\sigma^4} (W_{nsys})' \varepsilon_{ns}(\theta) \varepsilon_{nm}(\theta) W_{nmynm} - \frac{T - 1}{\sigma^2 T} (W_{nsys})' \varepsilon_{ns}(\theta) tr [(I - \lambda_m W_{nm})^{-1} W_{nm}]
\]
\[+ \frac{T - 1}{\sigma^2 T} (W_{nmynm})' \varepsilon_{nm}(\theta) tr [(I - \lambda_s W_{ns})^{-1} W_{ns}]
\]
\[+ \frac{(T - 1)^2}{T^2} tr [(I - \lambda_s W_{ns})^{-1} W_{ns}] tr [(I - \lambda_m W_{nm})^{-1} W_{nm}],
\]

\[
\frac{\partial L}{\partial \beta} \frac{\partial L}{\partial \beta}' = \frac{1}{\sigma^4} \sum_{t=1}^{T} \tilde{X}'_{nt} \varepsilon_{nt}(\theta) \sum_{t=1}^{T} \tilde{X}'_{nt} \varepsilon_{nt}(\theta)',
\]

\[
\frac{\partial L}{\partial \sigma^2} \frac{\partial L}{\partial \sigma^2} = \frac{n^2(T - 1)^2}{4\sigma^4} + \frac{1}{4\sigma^8} \sum_{t=1}^{T} \varepsilon'_{nt}(\theta) \varepsilon_{nt}(\theta) \sum_{t=1}^{T} \varepsilon'_{nt}(\theta) \varepsilon_{nt}(\theta) - \frac{n(T - 1)}{2\sigma^6} \sum_{t=1}^{T} \varepsilon'_{nt}(\theta) \varepsilon_{nt}(\theta),
\]

\[
\frac{\partial L}{\partial \lambda_s} \frac{\partial L}{\partial \beta}' = \frac{1}{\sigma^4} (W_{nsys})' \varepsilon_{ns}(\theta) \sum_{t=1}^{T} \varepsilon'_{nt}(\theta) X_{nt} - \frac{T - 1}{\sigma^2 T} tr [(I - \lambda_s W_{ns})^{-1} W_{ns}] \sum_{t=1}^{T} \varepsilon'_{nt}(\theta) X_{nt},
\]

\[
\frac{\partial L}{\partial \lambda_s} \frac{\partial L}{\partial \sigma^2} = -\frac{n(T - 1)}{2\sigma^4} (W_{nsys})' \varepsilon_{ns}(\theta) + \frac{n(T - 1)^2}{2\sigma^2 T} tr [(I - \lambda_s W_{ns})^{-1} W_{ns}]
\]
\[+ \frac{1}{2\sigma^6} (W_{nsys})' \varepsilon_{ns}(\theta) \sum_{t=1}^{T} \varepsilon'_{nt}(\theta) \varepsilon_{nt}(\theta) - \frac{T - 1}{2\sigma^4 T} tr [(I - \lambda_s W_{ns})^{-1} W_{ns}] \sum_{t=1}^{T} \varepsilon'_{nt}(\theta) \varepsilon_{nt}(\theta),
\]

\[
\frac{\partial L}{\partial \beta} \frac{\partial L}{\partial \sigma^2} = -\frac{n(T - 1)}{2\sigma^4} \sum_{t=1}^{T} \tilde{X}'_{nt} \varepsilon_{nt}(\theta) + \frac{1}{2\sigma^6} \sum_{t=1}^{T} \tilde{X}'_{nt} \varepsilon_{nt}(\theta) \sum_{t=1}^{T} \varepsilon'_{nt}(\theta) \varepsilon_{nt}(\theta).
\]
Denote $g_{ij}^s$ is the element of matrix $G_s$ at row $i$ and column $j$,

$$E\left( \frac{\partial L(\theta_0)}{\partial \lambda_s} \frac{\partial L(\theta_0)}{\partial \lambda_s} \right) = \frac{(T - 1)^2}{\sigma_0^2 T^2} (\mu_4 - 3\sigma_0^4) \sum_{i=1}^n (g_{ii}^s)^2 + \frac{T - 1}{T} tr(G_s'G_s) + \frac{(T - 1)^2}{T^2} tr(G_s^2)$$

$$+ \frac{T - 1}{\sigma_0^2 T} (X_{ns}\beta_0 + u_n)'G_s'G_s(X_{ns}\beta_0 + u_n),$$

$$E\left( \frac{\partial L(\theta_0)}{\partial \lambda_s} \frac{\partial L(\theta_0)}{\partial \lambda_m} \right) = \frac{1}{T^2} tr(G_s G_m) - \frac{1}{\sigma_0^2 T} (X_{ns}\beta_0 + u_n)'G_s'G_m(X_{nm}\beta_0 + u_n),$$

$$E\left( \frac{\partial L(\theta_0)}{\partial \beta} \frac{\partial L(\theta_0)}{\partial \beta} \right)' = \frac{1}{\sigma_0^2} \sum_{t=1}^T \tilde{X}_{nt}' \tilde{X}_{nt} - \frac{1}{\sigma_0^2 T} \sum_{t=1}^T \sum_{s=1}^T \tilde{X}_{nt}' \tilde{X}_{ns},$$

$$E\left( \frac{\partial L(\theta_0)}{\partial \sigma^2} \frac{\partial L(\theta_0)}{\partial \sigma^2} \right) = \frac{n(T - 1)^2}{4\sigma_0^8 T} (\mu_4 - 3\sigma_0^4) + \frac{n(T - 1)}{2\sigma_0^4},$$

$$E\left( \frac{\partial L(\theta_0)}{\partial \beta} \frac{\partial L(\theta_0)}{\partial \lambda_s} \right) = \frac{1}{\sigma_0^2} (X_{ns}\beta_0 + u_n)'G_s' \tilde{X}_{ns},$$

$$E\left( \frac{\partial L(\theta_0)}{\partial \sigma^2} \frac{\partial L(\theta_0)}{\partial \sigma^2} \right) = \frac{(T - 1)^2}{2\sigma_0^6 T^2} (\mu_4 - 3\sigma_0^4) + \frac{T - 1}{\sigma_0^2 T} tr(G_s),$$

$$E\left( \frac{\partial L(\theta_0)}{\partial \beta} \frac{\partial L(\theta_0)}{\partial \beta} \right) = 0.$$