A New Perspective on Regional Tax Burden Differences in China

Wei XI

Institute of National Accounts, Beijing Normal University, Beijing 100875, China E-mail: juniux@126.com

Jun XU

Institute of National Accounts, Beijing Normal University, Beijing 100875, China E-mail: juniux@126.com

Abstract Regional tax burden difference is related to regional sustainable development and social interests balance. Based on the calculation of the tax burden of 31 provinces and cities in 1995—2013, this paper puts forward that the over-statement of GDP is an important reason that causes the deviation of tax share and economic share. By using spatial econometric method, the results show that the economy as a whole is overvalued; investment in fixed assets, the proportion of secondary industry and state-owned economy have positive effects on regional GDP overestimated; per capita income and R&D have significantly negative effects on regional tax growth. Besides, the regional tax burden has spatial autocorrelation characteristic. This paper explains the abnormal difference of regional tax burden from a new perspective.

Keywords tax burden; regional differences; GDP over-statement; spatial econometrics

1 Introduction

Tax burden has always been one of the most acute and sensitive issues in people's livelihood. Research in taxation from almost all schools of economics will inevitably involve tax burden issue. The moderation of tax burden in the economy, and the fairness of tax distribution in regions not only determine the efficiency of resource allocation, but also affect the sustainability of economic growth, as well as the robustness and coordination of social functioning.

Since the tax reform in 1994, the total tax revenue of China has shown a trend of rapid growth on the basis of rapid economic development. However, from perspective of regional equilibrium, the regional distributions of China's tax burden have showed great differences. In real terms, the maximum difference in regional GDP (ratio of the highest and lowest areas) is more than 87 times every year, and the highest one among all years is 101 times. However, the maximum difference in regional tax revenues is 169 times on average, with a maximum difference of 281 times in 2007.

Although the regional imbalance is a common phenomenon in the process of economic development in both developed and developing countries, the regional difference of tax burden

Received April 6, 2016, accepted October 11, 2016

Supported by "the Fundamental Research Funds for the Central Universities (SKZZY2013004, SKZZX2013011)", and "the National Social Science Fund Project (12CYJ004, 14BTJ015)"

in China is extremely prominent compared with other countries. Besides, the regional difference of tax burden is not only reflected in the difference determined by the economic performance, but also in the divergences between economic performance and tax revenue. Specifically, China's regional tax burden and economic level showed a negative correlation. It presents a unique phenomenon that the higher a region's economic level, the lower is its tax burden.

Table 1 Geographical distribution of tax burden and economic share (1995–2013) unit: %

Ranking	Region	Economic Share	Tax Burden		
Тор	East	Guangdong, Jiangsu, Shandong , Zhejiang, Hebei , Liaoning, Shanghai(7)	Beijing, Shanghai, Tianjin, Hainan, Jiangsu(5)		
	Middle	$\mathbf{Henan},\mathbf{Hubei}(2)$	Shanxi(1)		
	West	Sichuan(1)	Yunnan, Guizhou, Xinjiang(3)		
	East	Fujian, Beijing(2)	Guangdong, Zhejiang, Liaoning(3)		
Middle	Middle	Hunan, Anhui, Heilongjiang, Jiangxi, Jilin, Shanxi(6)	Heilongjiang, Jilin(2)		
	West	Guangxi, Shaanxi(2)	Gansu, Ningxia, Qinghai, Inner-Mongolia, Sichuan, Chongqing, Shaanxi(7)		
	East	Tianjin, Hainan(2)	Hebei, Fujian, Shandong(3)		
Bottom	Middle	(0)	Anhui, Hubei , Jiangxi, Hunan, Henan (5)		
	West	Inner-Mongolia, Chongqing, Yunnan, Xinjiang, Guizhou, Gansu, Ningxia, Qinghai, Tibet(9)	Guangxi, Tibet(2)		

Note: Figures in () represent the number of regions. Eastern regions includes Beijing, Tianjin, Shandong, Hebei, Liaoning, Shanghai, Guangdong, Jiangsu, Zhejiang, Fujian, Hainan; middle regions includes Heilongjiang, Jilin, Shanxi, Henan, Anhui, Jiangxi, Hubei, Hunan; western regions includes Shaanxi, Inner-Mongolia, Qinghai, Gansu, Ningxia, Xinjiang, Tibet, Sichuan, Chongqing, Yunnan, Guizhou, Guangxi.

Table 1 shows the sorted result of regional economic share and tax burden by size order and geographical distribution since 1995. From the order of economic share, eastern region ranks first, followed by middle region, and western region ranks at the bottom. From the order of tax burden, the sort of the three parts is more uniform. For some regions, there even shows a reverse order. For example, the economic shares in Shandong, Hebei, Henan and Hubei enter into the national top ten, but their tax burden rank at the bottom ten in the country. The economic shares of Yunnan, Guizhou, Xinjiang, Tianjin and Hainan rank at the bottom ten of the country, but their tax burden rank into the national top ten. For the other regions, they also showed this deviation phenomenon in some degree. Although there are reasons like special taxes (e.g., resource tax) and the centralized collection of tax in headquarters economy, it is only limited in few areas, and can not been fully explained in most regions.

Revenue comes from economy, that means tax revenue and economic growth are positively correlated. Regardless of the growth speed of regional economy, the tax burden should approximate a stable value in the same tax system. In other words, in the absence of regional tax legislative authority, there is no difference in regional tax burden, and regional tax share should be consistent with their GDP share. Obviously, the actual situation of China's regional tax burden does not comply with this law, which means there is a deviation mechanism between regional economic growth and tax revenue. So, what factors influence the abnormal distribution of regional tax burden in China is a subject worthy of further explore.

2 Literature Review

On the formation mechanism of regional difference in tax burden, different scholars have different interpretations. Starting from relations between tax and economic, researchers find that economic growth is in an inverted U-shaped relationship with tax revenue, so different levels of economic level corresponds to different tax burden^[1-4]. Tanzi^[5] used social and economic factors as explanatory variables for regression analysis, and holds that the urbanization and government expenditure increase lead to the expansion in government debt, resulting in high tax burden; Ansari^[6] considered that high population density will increase the tax exemption and thereby reduce tax burden, because overpopulation is a symbol of underdevelopment; Stotsky and Woldemariam^[7] found that the economic structure has a significant impact on the whole tax burden, raising the proportion of agriculture will reduce tax burden, while raising the proportion of manufacturing sector will increase tax burden. Teera and Hudson^[8] found that per capita income, net exports, population density, scale of foreign debt and tax administration have an important influence on tax burden through empirical analysis. Garcia and Von Haldenwang^[9] used panel data of 131 countries in 1990–2008 and find that in addition to per capita income, demographic factors and economic structure, the type of political system also plays a role in tax burden. However, the findings of these studies usually support the conclusion that the more developed the economy is, the heavier its tax burden is. Besides, these studies mainly use the comparisons between countries, tax disparities within a country under the same tax system are few involved.

In view of the circumstances of China, domestic scholars Pan and Huang^[10] studied the issue of regional differences in tax burden in earlier time. They calculated China's regional tax burden by Theil Index, and attributed the differences to the ownership structure, extra budgetary revenue, income distribution, price indices and tax incentives. Subsequently, Wu^[11] broke down the Gini coefficient of total tax burden, and he believes that the ownership structure, industrial structure, income distribution, price indices, fiscal system and tax incentives are the main factors that cause the regional differences in tax burden. Zhang, Liu^[12] broke down the tax burden differences by regions, and consider the case of Beijing, Tianjin and Shanghai alone. They hold that tax burden has positively correlated with economic development, proportion of secondary industry, price indices and tax incentives, and has negatively correlated with extra-budgetary revenue. Zhang, Mo^[13] found that the tax burden has negatively correlated with proportion of primary industry, and that economic development and tax system are the determinants of tax burden. Lü and Guo^[14] believed that in addition to the factors as economic

growth, rising prices, industrial structure change, progressive taxation and taxation structure, the incentive function of tax decentralization and the focus releasing of revenue growth dividend are the main reasons that lead to the rapid growth of practical tax burden. Huang^[15] broke down Gini coefficient by tax categories, and found that indirect tax make a great contribution to the regional difference of tax burden, while the contribution of direct tax is increasingly important.

In the empirical analysis, Lin^[16] took empirical analysis with panel data in 1994–2006 and found that tax burden has a significant positive correlation with industrial structure, ownership structure and tax collection. Wang and Liu^[17] believed that the proportion of non-taxable GDP affects regional tax burden based on the framework of national accounting, they using input-output analysis found that industry structure has an important influence on tax burden, finance and insurance sector has the highest tax burden, while agriculture sector has the lowest tax burden. Huang, Hu and Liu^[18] found that per capita GDP, population density and total trade have significantly positive impacts on regional tax burden, and agriculture sector has the lowest tax capacity, while tax burden in industrial sector and service sector are relatively heavier. From the literature above, we find that most domestic studies focus on the comparison of tax burden difference (such as using statistical index to break total difference into within-group difference and between-group difference), as well as the qualitative analysis on its formation mechanism.

Furthermore, even if there is no conclusive support from quantitative test, the current studies also agree that the high tax burden comes from high economic growth, non-agricultural industries and other social economic factors. However, this does not fully explain the regional differences in tax burden showing in Table 1, especially the serious deviation between economic development and tax burden like Shandong, etc. In current studies, there is only one paper refers to the reverse distribution in regional tax burden^[15], but it does not give any qualitative or quantitative interpretation. This requires us to find a new outlet. One of the ideas is to retrospect the accuracy of statistical data, especially regional data of GDP. In general, tax data comes from micro-economic entities, and it will be deducted from the economic benefits of enterprises or residents. The possibility of exaggerating tax data is not very high when taxpayer's own economic benefits are considered, so the accuracy of tax data is relatively reliable. In contrast, under the current promotion model of local officials in China, which is termed as "promotion tournament model", economic indicators are often manipulated at each administration level^[19], thus the authenticity of GDP has always been questioned^[20–25].

In the long-time query of the authenticity of China's GDP data, the most intuitive and prominent one is the issue that regional data of GDP significantly differ from the national data. Since China established GDP accounting system formally, there has always been a gap between national data and regional data on GDP. We sum up regional GDP and compare it with national GDP in National Statistical Yearbook over years, find that the gap between them has been more than 1 trillion yuan since 2001, then the gap continued expansion. The gap then reached 6.12 trillion yuan by 2013, accounting for 10.75% of GDP. This gap weighted highest in 2004, which accounts for 19.26% of GDP, far exceeding the normal statistical error. Although, part of this gap is caused by survey system, in general it can be considered that there is an

overestimate tendency in regional GDP figures. In this paper we argue that it is the unbalanced overestimating of GDP in regions resulting in the abnormal distribution of tax burden.

Following the literature review above, we will have a theoretical analysis on the major factors and its mechanism affecting regional differences in tax burden. Then we will do the empirical analysis and robustness test on various hypotheses using spatial panel model. Lastly, we will give appropriate suggestions based on the results. Our study discussed the divergence of tax burden and economic development in view of the accuracy of GDP for the first time, which may enrich the empirical study on influence factors of regional tax burden.

3 Research Design

In order to investigate the extent of deviation and influence factors of tax burden, the previous expression of tax burden that use regional tax revenue as the numerator and regional GDP as the denominator, is unable to measure whether the economy is overvalued or not. Here we use the deviation between tax share and economic share as the dependent variable, and its specific calculation equation is as follows:

$$Y_{it} = \frac{\text{Tax}_{it}}{\text{Tax}_t} - \frac{\text{GDP}_{it}}{\text{GDP}_t}.$$
 (1)

Tax_{it} represents the tax revenue of region i in year t, Tax_t represents total tax revenue of all the regions in year t, so the first term in right side of Equation 1 represents the tax share of region i in year t. Similarly, GDP_{it} represents the size of GDP of region i in year t, GDP_t represents total GDP of all the regions in year t, the second term in right side of Equation 1 represents GDP share of region i in year t. Using regional tax burden deviation as dependent variable, by share calculation, we can present regions whose tax burden are significantly different from the average level. If the regional tax deviation is zero, it indicates that the tax share of this region is in line with its GDP share, and its development is relatively compatible in both taxation and economy. If the deviation is positive, it indicates that the regional tax burden is higher than the average level of the country, and GDP figure in this region may be underestimated relatively. If the deviation is negative, it indicates the regional tax burden is below the average level, and its GDP figure may be overestimated relatively.

3.1 Theoretical Hypothesis

Besides GDP misestimating, other factors such as industrial structure, income per capita and ownership structure also have influence on tax burden according to previous studies. For the same factor, if there are two opposite mechanisms in tax deviation, we can test them by empirical analysis. In this paper, we use industrial structure, income per capita, ownership structure and fixed capital investment as explanatory variables. Meanwhile, taking real economic performance into consideration, we also introduce electricity consumption, railway freight traffic, medium and long-term loans, technology development as explanatory variables, then we make the following hypotheses.

3.1.1 Industrial Structure

Different industries have different tax ratio, and their ability of creating tax revenue is not the same, so industrial structure has an important impact on regional tax burden. On one hand, previous studies are generally accepted that there are many preferential tax policies for agriculture sector and service sector to encourage their development in a long time, therefore the tax contribute of these two sectors is less than industrial sector. On the other hand, China's private enterprises and SMEs (small and medium-sized enterprises) are mostly concentrated in service sector, their nature of self-financing makes them no incentive to overvalue output or added value, and even hide their earnings for the purpose of tax avoidance. Therefore the value-added in service sector is not easily overestimated, but in sectors with opposite conditions, such as industrial sector, their ability of tax contribution may be underestimated. When the negative effect of value-added overestimating is higher than the positive effect of non-exemption in tax, the dependent variable will be negatively correlated with the proportion of industrial sector.

H1: Regional Tax deviation (TGD) is negatively correlated with the proportion of industrial sector (IND).

3.1.2 Per Capita Income

Generally, regions with higher income levels are well developed, highly modernized and densely populated areas, that means adequate economic endowments and higher productivity, and it will produce abundant taxation resources and strong taxable capacity. It also means better performance of real economic situation and less inflation in economic data, so their tax burden is not easy to be diluted. Besides, in the current progressive tax system, the higher the income is, the higher the tax ratio is, and the heavier the tax burden is.

H2: Regional Tax deviation (TGD) is positively correlated with disposable income per capita (DIP).

3.1.3 Ownership Structure

In order to attract foreign investment, foreign-funded enterprises in China have been given a lot of tax incentives. The differentiated tax treatment in ownership leads to the different tax burden. Meanwhile, state-owned enterprises take more social responsibility on social security and environmental protection, and have more hidden tax. Therefore, many studies believe that the tax burden of state-owned enterprises is higher than that of collective enterprises and foreign-funded enterprises. On the other hand, administrator in state-owned enterprise is usually appointed by government, enterprise's business performance is closely linked with administrator's political future. For this reason, SOEs have stronger motivation to overestimate their performance than other type of enterprises, so the output in regions with higher proportion of state-owned economy are more likely overvalued and their tax burden tend to be diluted. When the negative effect of output overestimating in state-owned enterprises is higher than positive effect caused by implicit tax burden, the dependent variable will be negatively correlated with the proportion of state-owned economy.

H3: Regional Tax deviation (TGD) is negatively correlated with the proportion of state-owned economy (SOE).

3.1.4 Investment of Fixed Assets

For a long time, China applies a production-type VAT policy, which means VAT could not be deducted when enterprises purchase fixed assets. This will result in a double taxation, raising the tax burden of capital-intensive business imperceptibly. This situation improved after the VAT reform, but for the sectors which pay business tax, such as real estate industry and construction industry, there is still such a problem like double taxation. On the other hand, fixed assets investment represented by infrastructure is often dominated by government, which is also the primary driver of local economic growth. Under the current promotion mode of local officials, it is easier to stimulate economic growth by investment behavior, so their economic output will be correspondingly elevated, and the contribution of tax will be diluted. When the negative effect of GDP inflation stimulated by investment is higher than the positive effect of double taxation, dependent variable will be negatively correlated with fixed assets investment.

H4: Regional Tax deviation (TGD) is negatively correlated with fixed assets investment (FAI).

3.1.5 Real Economic Performance

In the current local official promotion mode, regional GDP tend to be deliberately overvalued. In order to remove the exaggeration part of statistical data, more and more scholars propose to use better indicators to reflect the true economic development, such as rail freight traffic, electricity consumption, medium and long-term loan, etc. In 2010 the famous British political and economic magazine called "The Economist" introduced the "Li Keqiang Index" based on the three indicators. Later, this index is accepted by Citibank and other international agencies, and it is considered to give a better reflection on reality situation of China's economy. In order to restore the true features of regional tax burden, we draw these three indicators into explanatory variables.

H5: Regional Tax deviation (TGD) is positively correlated with electricity consumption (EC), medium and long-term loans (MLL) and rail freight traffic (RFT).

3.1.6 Technology Advance

Since the three indicators in "Keqiang Index" have a limitation of overemphasizing on manufacturing sectors, in March 2015, the government work report for the National People's Congress introduced an additional indicator — The proportion of research & experimental expenses, and it is seen as a new version of the "Keqiang Index". Expensing more funding on science and technology also means companies have higher profitability and tax base, and these good performed companies are usually lack of the motive on exaggeration. Therefore we take this indicator as one of explanatory variables.

H6: Regional Tax deviation (TGD) has a positive correlation with the proportion of research & experimental expenses (R&D).

3.2 Model Selection

Traditional panel model assumes that variables in different regions are mutually independent from each other. However, tax in neighborhood regions may be dependent, one region's tax burden may be affected by other regions through economic development, industrial structure, resource endowments and other factors. If the impact of this spatial dependence is ignored, there will be a biased result from the estimates.

3.2.1 Spatial Autocorrelation Test

Spatial autocorrelation analysis can be used to quantitatively describe the tax burden dependence in space and its spatial variation, and it is the necessary condition for setting spatial econometric model correctly. To further explore the spatial distribution pattern and cluster status of tax burden in various regions, this paper use 31 regions of China as space units, and test the spatial autocorrelation of main variables by using global spatial correlation index called "Moran's I", which is defined as:

Moran's I =
$$\frac{\sum_{i=1}^{n} \sum_{j=1}^{n} W_{ij} (Y_i - \bar{Y}) (Y_j - \bar{Y})}{S^2 \sum_{j=1}^{n} W_{ij}}.$$
 (2)

Y represents variables, S represents standard deviation, n is the total number of regions, W represents the spatial weights matrix. Moran's I index is in the range of (-1,1). If its value is greater than zero, it indicates that there is a positive spatial autocorrelation between regions. If the value is less than zero, it indicates that there is a negative spatial autocorrelation between regions. If the value is zero, it indicates that there is no spatial autocorrelation between regions.

In this study, spatial weights are set to 1 when two regions are adjacent to each other, and spatial weights are set to zero in other cases. After a standardized process, the Moran's I Index of regional tax burden in Equation 2 is -0.19387, Moran's I Index's test statistic is -6.5207, which indicates regional tax burden has a significantly negative correlation in space. Here the results of conventional econometric models will be biased, and spatial econometric models will be more suitable.

3.2.2 Model Specification

According to the difference expression on space, spatial econometric models are mainly divided into two types (Anselin, 1995): Spatial Lag Model (SLM) and Spatial Error Model (SEM). The basic forms are as follows:

$$SLM: Y_i = \rho W Y_i + \beta X + \varepsilon, \tag{3}$$

SEM:
$$Y = \beta X + \mu, \quad \mu_i = \lambda W \mu_i + \varepsilon.$$
 (4)

In these equations, Y is the dependent variable, X represents explanatory variables, W is the spatial weights matrix, ρ is the spatial lag effect, λ is the spatial error effect, ε and μ are random errors. According to basic variables, combining the method of spatial econometric model, we construct the spatial panel model as follows:

Spatial Lag Model:

$$TGD_{it} = \beta_1 IND_{it} + \beta_2 DIP_{it} + \beta_3 SOE_{it} + \beta_4 FAI_{it} + \beta_5 EC_{it} + \beta_6 MLL_{it}$$
$$+ \beta_7 RFT_{it} + \beta_8 RD_{it} + \rho \sum_{j=1}^n W_{it} TGD_{jt} + \varepsilon_{it}.$$
(5)

Spatial Error Model:

$$TGD_{it} = \beta_1 IND_{it} + \beta_2 DIP_{it} + \beta_3 SOE_{it} + \beta_4 FAI_{it} + \beta_5 EC_{it} + \beta_6 MLL_{it}$$
$$+ \beta_7 RFT_{it} + \beta_8 RD_{it} + \lambda \sum_{i=1}^n W_{it} \mu_{jt} + \varepsilon_{it}.$$
(6)

3.3 Data Description

As mentioned above, for the dependent variable, we use the difference between tax share and GDP share in regions to represent the relative difference of regional tax burden. In general, there are three calibers of tax, which are tax revenue (narrow caliber), fiscal revenue (medium caliber) and government revenue (wide caliber). Government revenue include land-transferring fees, administrative fees, government funds revenue, township self financing, pooling funds, state-owned enterprises income, etc. The source of government revenue is very complicated, and it is difficult to track them, so it is usually detached from official fiscal statistics. Medium-caliber fiscal revenue includes non-tax revenue, fines and donations income, those parts contain little correlation with other variables. Therefore, we abandon the revenue in medium-caliber and wide-caliber, and apply the small caliber tax revenue into analysis.

Tax revenue is also divided into central tax and local tax, and there are two kinds of understanding for local tax. One is tax revenue coming from local administrative division, the other is tax revenue assigned to local government. To portray the deviation of tax and economy, we need to minimize the effect of administrative interference, so here we use the former concept of local tax, which is the tax revenue generated in local economy, namely the indicator which called "regional tax revenue organized by all the taxation departments" in China Taxation Yearbook.

Moreover, price level also play an important role in regional tax burden, so each variable is required to remove the effect of price fluctuations before analysis. Here we set 1995 as the base period by applying CPI in 1995–2013. At the same time, in order to eliminate the impact of dimension, we standardized each variable.

Table 2 shows the description and calculation methods of each variable required in analysis. For the need of robustness analysis, this paper selects fiscal revenue (medium caliber), average wages in urban units, energy consumption, proportion of the number of state-owned enterprises, freight ton-kilometers and regional patent applications as alternative indicators.

Table 2 Main variables description

No.	Variables	Code	Unit	Calculation Method
1	Tax deviation from GDP	TGD	%	Regional Tax share minus its GDP share
2	Industrial sector proportion	IND	%	Regional industrial added-value/regional GDP
3	Per capita disposable income	DIP	yuan	Per capita disposable income of urban residents
4	State-owned economy proportion	SOE	%	Main business income of State-owned enterprises/main business income of enterprises above designated size
5	Fixed assets investment	FAI	yuan	Annual regional total investment in fixed assets
6	Electricity consumption	EC	kwh	Annual regional actual electricity consumption
7	Medium and long-term loans	MLL	yuan	Change in annual balance of medium and long-term loans
8	Railway freight traffic	RFT	tons	Actual weight of Railway transport goods
9	R&D expenses proportion	RD	%	Regional R&D expenditure/regional GDP

4 Empirical Analysis

4.1 Regression Results

Spatial sample data used in this paper covers the thirty one provinces, autonomous regions and municipalities in mainland China, not including Taiwan, Hong Kong and Macao. Sample period is from 1995 to 2013. Data mainly come from "China Statistical Yearbook", "China Taxation Yearbook", "China Industry Economy Statistical Yearbook" and "Almanac of China Finance and Banking", which are published by the National Bureau of statistics of China. According to the spatial econometric model established previously, we employ time period fixed effect, spatial fixed effect, spatial and time period fixed effect, random effect for both SLM and SEM models, as shown in Table 3. By the comparison of log-likelihood value, we think the result of spatial and time period fixed effect SLM model is the best.

According to the test results of spatial panel models, we have the following findings:

1) The proportion of industrial sector has a positive effect on GDP overestimation

Theoretically, there are two types effect (positive and negative) of industry proportion on TGD. Positive effect refers to the non-preferential policies effect on industrial sector, which leads to the increase of tax burden in this sector. Negative effect refers to the phenomenon that output-overestimation in industry is more common than in other sectors. It makes regional GDP share exceeds its tax share, and regional tax burden exceeded by other regions. According to Table 3, IND coefficients are between -0.32 to -0.22, which confirmed the conclusion that regional tax deviation is associated with industry proportion. Both spatial lag model and

Table 3 Spatial regression result of regional tax deviation

Table 3 Spatial regression result of regional tax deviation										
	SLM					SEM				
	Time Fixed	Spatial Fixed	Spatial& Time Fixed	Random Effect	Time Fixed	Spatial Fixed	Spatial& Time Fixed	Random Effect		
IND	-0.2848 *** (-6.562)	-0.2381 *** (-3.465)	-0.2843 *** (-6.532)	-0.2178 *** (-3.539)	-0.3187 *** (-7.764)	-0.2597 *** (-3.815)	-0.3186 *** (-7.749)	-0.2215 *** (-3.568)		
DIP	0.5509 *** (12.233)	0.2316 *** (3.548)	0.5499 *** (12.178)	0.3027 *** (5.436)	0.5016 *** (11.962)	0.2033 *** (3.250)	0.5009 *** (11.925)	0.2749 *** (5.102)		
SOE	-0.0561 *** (-3.260)	0.0185 (0.927)	-0.0562 *** (-3.253)	0.0194 (1.017)	-0.0287 * (-1.907)	0.0183 (0.938)	-0.0288 ** (-1.915)	0.0181 (0.966)		
FAI	-0.1064 ** (-2.090)	0.1707 *** (4.332)	-0.1059 ** (-2.076)	0.1431 *** (3.713)	-0.0600 (-1.323)	0.1733 *** (4.498)	-0.0602 (-1.325)	0.1517 *** (3.995)		
EC	0.01824 (0.320)	-0.1110 * (-1.901)	0.0177 (0.309)	-0.1126 * (-1.926)	-0.0021 (-0.039)	-0.1078 * (-1.857)	0.0006 (0.011)	-0.1178 ** (-2.011)		
MLL	0.0592 *** (3.137)	0.0502 *** (4.052)	0.0596 *** (3.150)	0.0506 *** (4.019)	0.0733 *** (4.086)	0.0530 *** (4.294)	0.0737 *** (4.099)	0.0542 *** (4.302)		
RFT	0.0350 *** (3.059)	0.0076 (0.423)	0.0352 *** (3.065)	0.0156 (0.932)	0.0495 *** (4.621)	0.0060 (0.332)	0.0494 *** (4.604)	0.0109 (0.634)		
RD	0.2326 *** (12.771)	0.1989 *** (13.643)	0.2325 *** (12.733)	0.2031 *** (13.774)	0.2145 *** (12.372)	0.1955 *** (13.453)	0.2145 *** (12.348)	0.1994 *** (13.487)		
ρ	-0.3260 *** (-6.792)	-0.0998 ** (-1.989)	-0.3219 *** (-6.655)	-0.1249 ** (-2.502)	_	_	_	-		
λ	-	-	-	-	-0.4479 *** (-7.495)	-0.1270 ** (-2.136)	-0.4470 *** (-7.476)	-0.1263 ** (-2.069)		
teta	-	-	-	0.1514 *** (5.625)	-	-	-	2.9532 *** (4.262)		
$Adj - R^2$	0.6018	0.8587	0.5997	0.8506	0.5463	0.8567	0.5447	0.8505		
Log-L	2360.029	2046.034	2363.417	2284.794	2359.952	2053.773	2363.386	2280.488		
obs	589	589	589	589	589	589	589	589		

Note: numbers in () represents T statistics of regression coefficient, ***, **, * represent the significant level at 1%, 5% and 10%.

spatial error model, whether random effects or fixed effects, we all find that tax burden and industry proportion are negatively correlated. It means the negative effect exceeds the positive effect. GDP overestimation is not just an existing issue in industrial sector, but also plays an important role in tax share deviation.

2) The per capita income has a positive correlation with regional tax burden

The results of various regression models show the impact of per capita income is significantly positive. DIP coefficient in time period fixed effects SLM model is reached 0.5499, indicating H2 is confirmed. Regions with higher per capita income in China mostly are eastern and coastal areas, due to their good economic base, their tax paying ability are strong, and relatively lack of the politically motivation to fraud statistics. Affected by this positive influence, regional share of tax burden rarely deviate negatively. Besides, due to the progressive mechanism of income tax, enterprises and households in these regions will undertake more tax burden, so the positive effect of per capita income is further enhanced.

3) State-owned economy has a positive effect on GDP overestimation

The effects of state-owned economy on regional tax burden present both positive and negative mechanisms. The positive mechanism refers that non-preferential tax policies and social responsibility lead to higher tax burden in state-owned enterprises. Negative mechanism refers that the responsive person in state-owned enterprise has GDP overestimation motivation, resulting in a lighter tax burden in state-owned enterprises. The regression result shows that SOE coefficient is -0.056, indicating that the ownership structure has a slight negative influence on regional tax burden. It confirms the GDP overestimation of state-owned enterprises to a certain extent, and this negative influence exceeds the positive effect. Furthermore, since the state-owned enterprises are more concentrated in industrial sector, its impact on TGD may be more explained by industrial sector proportion and only show a slight negative effect.

4) Fixed assets investment is one cause of GDP bloat

On the basis of previous analysis, spontaneous investment behaviors of enterprises undertake double tax burden due to the breakage of tax deduction chain. It will draw a positive shift on TGD. On the other hand, in order to improve output performance, local governments usually overly rely on the boosting impact of investment on economic growth. The long-standing "local investment hunger" and "overcapacity of production" have fully illustrated this issue. It will draw a negative shift on TGD. According to the estimation result in Table 3, it shows that there is a significant negative correlation between TGD and FAI. Regression coefficient is -0.1059, and H4 is confirmed. In the formation mechanism of TGD, the negative effect of fixed assets investment exceeds its positive effect and plays a leading role. This confirms the over-reliance on fixed assets investment of local economy. As a result, regional tax share is less than its economy share, and mitigates the regional tax burden.

5) The real economic indicators confirmed the exaggeration in overall GDP

In this paper, dependent variable is designed to take average tax burden of the country as a reference implicitly. As long as one region's GDP is overestimated, the national average tax burden will be pulled down, and TGD in regions with true GDP figures will turn to positive. Therefore, if coefficients of real economic indicators are positive, it means at least one region's GDP figure is overestimated. The empirical result shows that MLL and RFT are both

significantly positive correlated with the dependent variable, which means the overall economy is over estimated. EC has not been confirmed, because it may be correlated with IND, and the latter has a negative effect on the dependent variable.

6) Technology advance has a positive correlation with regional tax burden

According to regression results in Table 3, it shows that the proportion of R&D expenses has a positive correlation with regional tax burden, coefficients in all kinds of regression models are above 0.2, so H6 is confirmed. It indicates that if a region has a relatively higher technology level, its tax share will be higher its GDP share, technology advance improves the profitability of enterprises and carrying capacity of tax burden. On the other hand, it also indicates that enterprises with highly research is lack of the motivation to overvalue its output, GDP data is relatively real.

7) Regional tax burden deviation has a negative spillover effects

Table 3 shows that the spatial autoregressive coefficient in SLM model is negative, and it passed significance test, it means that regional tax deviation is not only affected by regional factors of itself, but also by the negative spillover effect of surrounding regions. Intuitively, the sum of all regions TGD is zero. One region's overestimation of GDP will result in a negative deviation of its own TGD and a positive deviation of TGD in other regions. Table 3 also shows that the spatial error autocorrelation coefficient in SEM models is negative, and passed the significance test. It indicates that regional TGD is also affected by the inhibitory effect of unobserved factors from surrounding areas, such as tax competition. Although, there is no independent taxation legislative authority of local government in China, they may carry out an actual tax competition by the fiscal decentralization system. In order to attract capital and other production factors, local governments may reduce its effective tax rate by improper resort. It will create a magnetic effect for production factors in other regions, and strengthen the competitiveness of their own, as well as weaken the competitiveness of other regions.

From the entire regression results, the six research hypotheses are basically confirmed.

4.2 Robust Tests

To further confirm the reliability of above conclusions, we carry out some robust tests from the following aspects. The results do not change original conclusions on the whole.

1) Ordinary least squares model

Before spatial econometric models are carried out, we also used ordinary least squares regression model for analysis, the results of which are showing in the last column of Table 4. In general, the determination coefficient of OLS model turns out to be 0.5438, a little weaker than the fitting results of spatial econometric models. According to regression coefficients, there is no directional change for all the explanatory variables except EC. Indicators of IND, SOE and FAI have a positive effect on regional GDP overestimation, and indicators of DIP, MLL, RFT and R&D are still true reflections of economic growth. As for the scale of regression coefficients, except FAI, all of them showed no significantly change with spatial econometric models, demonstrated a good stability of parameters. The scale change of FAI's coefficient can be attributed to the direction change of EC, which partly weakening the influence of overinvestment on GDP overestimation.

2) The replacement of dependent variable

For the dependent variable, this paper replaces tax revenue with fiscal revenue to recalculate and estimate the deviation of regional tax burden, and the results of which are presented in column 4, 5, 6 of Table 4. According to the determination coefficient and likelihood value, the estimation result of SLM model is still better than the SEM model. The spatial autoregressive coefficient and spatial error autocorrelation coefficient are significantly negative, indicating that the rising up of one region's economic share by GDP overestimation will relatively undervalue other region's economic share at the same time. For the regression parameters, all the coefficients in regression results are very close to the original models except EC. Although the coefficient of EC turn to negative, it does not pass the significant test, and its impact on dependent variable is less than 0.004, which can be ignored.

3) The replacement of explanatory variables

In order to test the robustness of explanatory variables, we intends to replace per capita income with urban wages, replace state-owned economy proportion of main business income with the proportion of staff and workers number, replace railway freight traffic with freight ton-kilometers, replace R&D expenses with the number of patent applications. Descriptive statistical analysis found that there is a strong correlation between original indicators and surrogate indicators, most of the replacements have little difference in terms of trends and volatility. So for explanatory variables, here are two major changes. One is to replace the RFT with freight ton-kilometers, the other is to remove RFT from explanatory variables. The results are presents in the first four columns of Table 4. The results show that there is little change after the replacements of explanatory variables. Overall, the estimation results in SLM model are still better than that of SEM model. The coefficients of IND, SOE, FAI and spatial spillover effects are negative, while the coefficients of DIP, EC, MLL, FT and RD are positive. Comparatively, the estimated effect in the replacement of explanatory variables is weaker than that in the original indicators, while it is better than the replacement of dependent variable and OLS model.

5 Conclusions and Implications

Tax burden is related to the sustainable development of economic as well as the Equilibrium and stability of social interests, therefore the issues of regional differences in tax burden have been always concerned by researchers. By computing the tax share deviation of all regions in China, the phenomenon of regional GDP overestimate is found an important cause of the regional differences in tax burden, applying the space panel data model to analyze the determinant of area tax share deviation from 1995–2013, this study found that:

First, long-term loans, electricity and rail freight, these three variables that characterize real economic development shows that the aggregate figures of GDP in China is overestimated. In the GDP-centered governance structure of China's local governments, which is termed as "promotion tournament model", to get praise from the performance evaluation, local government officials often deliberately cater to the GDP measurement that deviate from the original purpose of the evaluation. We can now consider taking the real economic development indicators into the official assessment system to avoid the inaccuracy of regional GDP figures.

However, in humans' economic and social life, the interaction between measure behavior and measure objects is inevitable. So in the long run, we also need to introduce a variety of assessment methods and index, checks and balances the measure behavior, then to form a healthy competition ultimately.

Table 4 Robust tests of regional tax burden deviation

	Table 4 Robust tests of regional tax burden deviation					OT C	
	Freight Ton-kilometers Without RFT Fiscal Revenue		OLS				
	SLM	SEM	SLM	SEM	SLM	SEM	
	-0.2667	-0.2665	-0.2542	-0.2633	-0.3032	-0.3215	-0.2489
IND	***	***	***	***	***	***	***
	(-6.342)	(-6.603)	(-5.987)	(-6.500)	(-7.129)	(-7.836)	(-5.353)
	0.4850	0.4360	0.5341	0.4694	0.5728	0.520	0.4911
DIP	***	***	***	***	***	***	***
	(10.534)	(10.018)	(11.834)	(11.061)	(12.944)	(12.401)	(10.307)
	-0.0305	0.0144	-0.0529	-0.0322	-0.0785	-0.0507	-0.0618
SOE	*	-0.0164	***	**	***	***	***
	(-1.697)	(-1.003)	(-3.053)	(-2.081)	(-4.637)	(-3.324)	(-3.365)
	-0.0966	0.0490	-0.1049	0.0000	-0.1073	0.0500	0.0500
FAI	**	-0.0439	**	-0.0393	**	-0.0509	-0.0590
	(-1.906)	(-0.948)	(-2.042)	(-0.845)	(-2.154)	(-1.106)	(-1.089)
	0.0004	0.0977	0.0935	0.1237	0.0027	0.0050	0.0000
EC	0.0824	**	*	**	-0.0037	-0.0250	-0.0209
	(1.602)	(2.006)	(1.797)	(2.565)	(-0.0667)	(-0.461)	(-0.342)
	0.0618	0.0677	0.0570	0.0680	0.0746	0.0899	0.0756
MLL	***	***	***	***	***	***	***
	(3.293)	(3.704)	(3.002)	(3.706)	(4.035)	(5.015)	(3.761)
	0.0463	0.0369			0.0511	0.0612	0.0483
RFT	***	***	-	-	***	***	***
	(4.004)	(3.133)			(4.5616)	(5.740)	(4.006)
	0.2255	0.2003	0.2234	0.2006	0.2137	0.2038	0.2314
RD	***	***	***	***	***	***	***
	(12.578)	(11.516)	(12.317)	(11.477)	(11.971)	(11.742)	(11.866)
	-0.3439		-0.3470		-0.2480		
ho	***	_	***	-	***	-	-
	(-7.240)		(-7.303)		(-5.109)		
		-0.4069		-0.422		-0.3340	
λ	-	***	-	***	-	***	-
		(-6.779)		(-7.039)		(-5.539)	
$Adj - R^2$	0.5434	0.6056	0.5334	0.5951	0.5659	0.5993	0.5438
Log-L	2359.935	2047.089	2280.288	2041.803	2238.360	2066.74	-
obs	589	589	589	589	589	589	589

Note: Numbers in () represents T statistics of regression coefficient, ***, **, * represent the significant level at 1%, 5% and 10%.

Secondly, per capita income and R&D expenditure are positively related with the original tax burden. In the areas of higher per capita income and more adequate R&D expenditure, enterprises have strong profitability, and the whole area has a good economic performance, therefore the motivation of GDP figures fraud is relatively weak, per capita income and R&D expenditure are real factors for the differences in regional tax burden. Moreover, this difference in tax burden is an important manifestation of taxation function in income distribution adjustment. The crucial policy orientation is not to reduce the tax burden, but rather to build a scientific and rational system of local fiscal expenditure, so that the fiscal expenditure will be more inclined to the people's livelihood, especially improving the living standards of low-income groups who live in these areas, improving social security, regulate the underlying people's feeling to pay taxes.

Besides, the regional spatial effect presents negative coefficients, which can be explained from two aspects. On the one hand, the negative deviation is brought by the share shift of region economy. From the formula of Yit, there is a tradeoff between regions. On the other hand, due to the difference in production elements and institutional factors, different regions have become independent economic space units, which means there are negative dynamic competitive relationship with each other. For example, although there is no independent taxation legislative authority for the local governments, but they may attract production elements from the surrounding area to promote their own economic performance through tax credits, government subsidies and other economic tools, while the surrounding areas may be impaired due to the loss of production elements. Thus, by building a scientific and rational revenue distribution system, encouraging interregional cooperation, and establishing a unified market, we can promote the balanced development among regions.

Finally, investment in fixed assets, proportion of secondary industry and proportion of stateowned economic have a positive impact on regional GDP overestimated. Previous studies have generally considered that these three factors play an important role in pushing up the local tax burden, but it is under the assumption that local economic has not been overestimated. Negative regression coefficients in empirical analysis confirmed the existence of GDP overestimated phenomenon, and their negative impact on regional tax burden have surpass their positive effects under the theoretical analysis. The extensive economic growth pattern over depend on investment is deficient not only in revenue contribution, but also in enterprises' sustainable development. The rapid development of local economy needs layout at a higher level on industry structure and industrial patterns.

References

- [1] Marsden K. Links between taxes and economic growth: Some empirical evidence. Journal of Economic Growth, 1986, 4(1): 3–16.
- [2] Barro R. Government Spending in a simple model of endogenous growth. Journal of Political Economy, 1990, 98(5): 103–125.
- [3] Scully G W. Tax rate, tax revenue and economic growth. Policy Analysis, Dallas, 1991.
- [4] Nir J, Sergio R. Non-linear effects of taxation on growth. NBER Working Paper Series, 2012.
- [5] Tanzi V. Comparing international tax burdens: A suggested method. The Journal of Political Economy, 1968, 76(5): 1078–1084.

- [6] Ansari M. Determinants of tax ratio: A cross-country analysis. Economic and Political Weekly, 1982(19): 1035–1042.
- [7] Stotsky J G, Woldemariam A. Tax effort in Sub-Saharan Africa. IMF Working Paper. IMF: Washington, 1997.
- [8] Teera J M, Hudson J, Tax performance: A comparative study. Journal of International Development, 2004(16): 785–802.
- [9] Garcia M M, Haldenwang C. Do Democracies tax more? Political regime type and taxation. Journal of International Development, 2016, 28(4): 458–506.
- [10] Pan X Z, Huang Y J. An empirical analysis of tax burden differences and its influence factors in China. Economic Research Journal, 1998(11): 69–73.
- [11] Wu L H. An empirical analysis on area tax burden's disparity in China. Journal of Finance and Economics, 2000, 26(10): 40–47.
- [12] Zhang W C, Liu Y. Macro tax burden: Analysis of difference across the regions of our country. Economic Theory and Business Management, 2001(3): 11–16.
- [13] Zhang B, Mo E X. Comparison of regional differences of tax burden. Taxation Research, 2005(5): 24-26.
- [14] Lü B Y, Guo Q W. Why China's tax revenue is likely to maintain its rapid growth: An explanation within the framework of tax capacity and tax effort. Social Sciences in China, 2011(2): 76–90.
- [15] Huang X L. Tax differences between regions and tax structures. Reform of Economic System, 2012(3): 112–116.
- [16] Lin Y. Empirical analysis of the reasons for regional differences in tax burden. Taxation Research, 2009(8): 54–58.
- [17] Wang J P, Liu Q Y. Judgment on the level of macroeconomic tax rate in China. Finance & Trade Economics, 2005(8): 34–40.
- [18] Huang X L, Hu Z Q, Liu Y. Tax capacity, tax effort and regional differences in tax burden. Economic Science, 2012(4): 80–90.
- [19] Zhou L A. Governing China's local officials: An analysis of promotion tournament model. Economic Research Journal, 2007(7): 36–50.
- [20] Meng L, Wang X L. An estimation of the reliability of statistic data on China's economic growth. Economic Research Journal, 2000(10): 3–14.
- [21] Ren R E. How fast has China's economy grown? A comment on two recent assessment. China Economic Quarterly, 2002, 2(1): 37–52.
- [22] Rawski T G. Measuring China's recent GDP growth: Where do we standChina Economic Quarterly, 2002, 2(1): 53–62.
- [23] Zhang X, Jiang D C. China's economic growth: Reliability of growth statistics and micro foundation of growth. China Economic Quarterly, 2002, 2(1): 1–22.
- [24] Que L, Zhong X H. A regional level test on the authenticity of China's GDP growth statistics. The Journal of Quantitative & Technical Economics, 2005(4): 3–12.
- [25] Liu H, Chang X Y. Chinese GDP data accuracy assessment based on TFP. Statistical Research, 2011, 28(2): 81–86.