ESSAYS IN APPLIED MICROECONOMICS

By

YING TAN

Bachelor of Arts in Economics
Xiangtan University
Xiangtan, China
2005

Master of Arts in Economics
Nationalities University of Yunnan
Kunming, China
2010

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
December, 2015

ESSAYS IN APPLIED MICROECONOMICS

Dissertation Approved:

Dr. Dan Rickman

Dissertation Adviser

Dr. John Winters

Dr. Lee Adkins

Dr. Chanjin Chung

ACKNOWLEDGEMENTS

I would like to thank my advisor Dan Rickman for his support and help. He has set up an example of diligence and excellence for me. His passion for research and work always inspires me and he is also my mentor. I would also like to thank my committees, Lee Adkins, John Winters and Chanjin Chung, for all their help.

I am deeply indebted to my parents, Tan Chunlin and Tian Shuizhen, for their support and love. I am very grateful to my lovely wife, Li Yu, for her encourage and support throughout all these years.

Finally, I would like to thank for the support from the Department of Economics and Legal studies in Business. Without all the support from the department, this research would not be possible.

Name: Ying Tan

Date of Degree: DECEMBER, 2015

Title of Study: DOCTOR OF PHILOSOPHY

Major Field: ECONOMICS

Abstract: In the first chapter, the hypothesis that workers are fully compensated in wages for differences in cost of living is tested for four groups of workers with different levels of educational attainment for the years of 2000, 2006, and 2011. Unlike previous studies, based on the 2SLS method, I find that the wage-price elasticity generally increases with educational attainment in the more recent years. The empirical results support the full compensation hypothesis for workers with educational attainment equal to or higher than some college in the more recent years, while workers whose educational attainment is lower than high-school are found to be incompletely compensated in all years. In the second chapter, the Panel Analysis of Nonstationarity in Idiosyncratic and Common components (PANIC) approach with recursive mean adjustment method (RMA) is applied to investigate regional economic convergence in China. Unlike previous research which finds evidence in favor of intra-region economic convergence, based on a panel set of real per capita GDP for 28 provincial unit in China from 1978-2012, after bias reduction, common factors are found to be nonstationary for China and its three sub-regions. Most of the idiosyncratic components are also found to have a unit root. My results show then that regional economic clubs do not exist in China; thus, reflecting the problem of provincial growth divergence in China. In the third chapter, a new G2SLS approach proposed by Lee (2007) and developed by Bramoullé et al. (2009) is first applied to investigate spillover effects of counties' employment growth, initial fiscal policy variables and initial employment density for the U.S. Unlike the conventional Spatial-Durbin IV model, according to the results of the G2SLS approach, no evidence of spillover effects of employment growth is found; positive spillover effects on employment growth is found for initial safety expenditures in 2000-2007 and 2000-2010, and for initial high-tech employment share in 2000-2010. Initial log of county employment density is found to lower its own employment growth rate. The Monte Carlo simulation results imply that, based on the group interaction structure in the sample, the G2SLS approach provides credible identification of the model parameters.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

Full Compensation Hypothesis: Does it Hold for All Workers?

1.1. Introduction

Spatial equilibrium theory (Roback 1982) suggests that workers should be fully compensated for differences in cost of living in equilibrium. More specifically, after controlling for observable individual characteristics and amenities across areas, the elasticity between wages and cost of livings should equal one. This is referred to as the full compensation hypothesis in the literature. With given levels of amenities, workers will require higher wages to work in a city with higher prices.

But, does the full compensation hypothesis really hold? A number of studies assumed that it held for all workers and used fully-adjusted wages for interarea wage differentials to measure the implicit prices of amenities across cities (e.g. Rosen, 1979; Roback, 1982; Gabriel et al., 2003; Gabriel and Rosenthal, 2004). But to date few studies have tested the full compensation hypothesis empirically.

The literature shows mixed results about the elasticity between wages and prices. Roback (1988) estimated a wage-price elasticity of 0.97, both with and without controls for amenities. She obtained this estimate by using a cost of living index produced by the Bureau of Labor Statistics (BLS), and dividing all workers into two different types: high income and low income.

Then she assigned different bundles of consumption goods for the two types of workers, and this results in different price levels for the two worker groups. Her results support the full compensation hypothesis. But her arbitrary assignment of different bundles of consumption goods may weaken the robustness of her empirical findings. Instead of using the now discontinued cost of living index by the BLS, DuMond et al. (1999) used the ACCRA Cost of Living Index to compose city price indexes. With controls for amenities, he estimates a wage-price elasticity of 0.46; while, without controlling for amenities, the estimated wage-price elasticity is 0.37.

Winters (2009) reexamines Roback (1988) and argues that her measurement of prices is inappropriate and biases her estimates. He finds that estimates of the wage-price elasticity are sensitive to whether housing costs are measured by housing values or rent payments and the estimation method. When housing prices are measured by housing values and using OLS, the elasticity between wages and the general price level is less than 0.5, when measuring housing prices by rents and using OLS, the wage-price elasticity is 0.7. Instrumenting the rent-based price index using rents for the previous year produces an elasticity nearly identical to one. He argues that the results from rent-based housing costs and IV estimation should be preferred, which lends support for the full compensation hypothesis.

Few studies have been done to study the validity of the full compensation hypothesis for different groups of workers in the US. This is because the mobility of workers changes with their educational attainment (Wozniak, 2010; Malamud and Wozniak, 2012). Workers who are more mobile are more likely to move to places where they could be fully compensated; while less mobile workers are more likely to be restricted in their current locations, although they are not fully compensated there. DuMond et al. (1999) divided workers into four groups according to their educational attainment. They find the wage-price elasticity to be decreasing with education, and the elasticity for all four groups is less than 0.5. Winters (2009)[1] also divided workers into four groups, and by using the 2006 Current Population Survey Outgoing Rotation Group samples, he finds the wage-price elasticity to be significantly larger than those found by DuMond et al. (1999) for all four groups. He finds the elasticity for high school dropouts to be significantly larger than one in all the cases, while the elasticity for college graduate is found to be less than one; he also finds the wage-price elasticity to be decreasing with education.

In this study, by applying Roback's (1982) framework and following Winters (2009), I develop a model which predicts that, after controlling for amenities, the elasticity between wages and the general price level should equal 1. But I use different samples than Winters (2009). My samples are American Community Survey's 2000 5% sample data and 1% sample data for 2006 and 2011. By using these samples, I investigate whether the full compensation hypothesis holds for workers with different levels of educational attainment. Unlike previous research, I assign both same and group-specific housing and non-housing goods weights to different groups of workers to calculate city level price indices for each group. This is because group-specific weights could eliminate any possible bias that may be caused by assigning incorrect weights to workers. Previous studies just looked at one year, but because the spatial equilibrium assumption may not be achieved in a particular year, I examine different years. I test the hypothesis for 2000, 2006, and 2011. I divide workers into four different groups according to their educational attainment: less than high-school, high-school graduate, some college, college and above. I test whether the full compensation hypothesis holds for each group.

My empirical results support the full compensation hypothesis for workers with higher educational attainment in the more recent years. Workers whose educational attainment is lower than high-school are found to be incompletely compensated, while workers whose educational

[1] See John Winters' dissertation, Essay I: "Wages and prices: Are Workers fully compensated for cost of living differences".

attainment are equal to or higher than some college are found to be fully compensated in 2006 and 2011. Workers who graduate from high school or have equivalent diploma are found to be incompletely compensated in most of the cases in the two years, and fully compensated only in year 2006 while group-specific price weights are used.

1.2. Model

The full compensation hypothesis suggests that workers should be fully compensated for different cost of living, while controlling for amenities. This implies that the elasticity between wage and price should be equal to one. Following Roback's (1982) framework, I include different groups of worker into Winters's (2009) model to develop a model of the equilibrium relationship between wages, prices, and amenities across cities for workers with different educational attainments. In any city, identical firms with constant returns to scale technology use different kinds of labor N, where $N = (N_1, N_2, ...N_n)$ is a vector that contains all kinds of labor that a firm uses in production, capital K and land L are used to produce housing H and other non-housing goods X given locational differences in productivity due to amenities A:

$$H = H(N, K, L; A)$$
$$X = X(N, K, L; A)$$

The marginal product of labor, capital, and land are all non-negative, but increases in amenities can either increase or decrease productivity. Wages, W, where $W = (W_1, W_2, ...W_n)$, and rent of land, R_L, are determined competitively in local markets, while the price of capital is determined exogenously in the world market and normalized to 1. Thus, the price of housing P_H and the price of non-housing goods P_x are the relative prices to the price of capital. In equilibrium, firms earn zero economic profits and the price of each good is equal to its average cost (C_H, C_X):

$$C_H(W, R_L; A) = P_H$$
$$C_X(W, R_L; A) = P_X$$

where C_H is the cost of producing housing services and C_X is the cost of producing non-housing goods.

Workers maximize utility subject to their total wages and their utility is a function of housing, non-housing goods and location-specific amenities:

$$U_i = U_i(H, X; A)$$

where i denotes different groups of workers.

The utility function is assumed to increase in H and X, while being concave in H and X. A is an exogenous parameter that shifts utility level. Workers are assumed to be mobile across cities. According to spatial equilibrium theory, because of perfect mobility the utility levels for identical workers are equal across areas in equilibrium. Thus, the indirect utility function for the different groups of workers is:

$$V_i = V_i(W_i, P_H, P_X; A) \qquad (1.1)$$

Taking the total differential of both sides of equation (1.1),

$$dV_i = \frac{\partial V_i}{\partial W_i} dW_i + \frac{\partial V_i}{\partial P_H} dP_H + \frac{\partial V_i}{\partial P_X} dP_X + \frac{\partial V_i}{\partial A} dA$$

setting $dV_i = 0$, rearranging, and employing Roy's Identity gives the relationship between wages and prices:

$$dW_i = H_i * dP_H + X_i * dP_X - P_A * dA \qquad (1.2)$$

Dividing both sides of equation (1.2) by W, and converting the equation into logarithmic form we have:

$$d\ln W_i = (P_H H_i / W_i) * d\ln P_H + (P_X X_i / W_i) * d\ln P_X - (P_A / W_i) * dA \qquad (1.3)$$

In equation (1.3), $P_H H_i / W_i$ is the share of wages spent on housing, and $P_X X_i / W$ is the share of wages spent on non-housing goods. Equation (1.3) implies that, ceteris paribus, if either housing costs or the price of non-housing goods increase one-percent, workers will demand wages to increase exactly the percentage of the share of the corresponding goods to keep the same utility. If individual workers maximize their utility we have $P_H H_i + P_X X_i = W_i$, which means $P_H H_i / W_i + P_X X_i / W_i = 1$ This implies that in order to keep the same utility level, wages should be increased by one percent if the price for all the goods increases one percent.

The assumption of perfect mobility may be true for workers with higher educational attainment, and less true for workers with lower educational attainment. Actually, Katz and Autor (1999), Lemieux (2008), Autor et al. (2008), and Moretti (2013) have showed that there has been real-wage inequality between skilled and unskilled workers since 1980. But, in order to see which group of workers is fully compensated and which is not, and how incompletely compensated a particular group is, we have to rely on the empirical results.

1.3. Empirical Method and Data

Based on the model, my basic empirical equation is a variation of equation (1.3):

$$\ln W_{eij} = \beta Y_{eij} + \theta \ln P_j + \gamma A_j + \phi Ind_e + \varphi Occ_e + \psi Job_e + \rho City_j + \tau Edu_{eij} + \varepsilon_{eij} \qquad (1.4)$$

where i denotes individual worker, e denotes the educational group that the worker belongs to, and j denotes city, the so-called city in this paper is either a Core Based statistical Area(CBSA) or a Combined Statistical Area(CSA)[2]. W_{eij} is the hourly wage of individual i belonging to group e in location j, Y_{eij} contains individual characteristic variables: dummy variables for race/ethnicity groups, sex, marriage status, employed part-time, enrolled part-time in school, educational attainment, naturalized citizen and non-citizen, and a quadratic specification for experience. P_j is the price level in city j. A_j is the vector that contains amenity variables for city j, these amenity variables are: mean January temperature, mean July temperature, mean July humidity, mean hours of sunlight, topology, and water percentage of the area. Ind_e contains 16 dummy variables denoting the industry that worker i works in, and Occ_e contains 25 dummy variables denoting worker i's occupation, Job_e are the dummy variables that denote whether worker i works in a non-profit organization, federal government, state government or local government. $City_j$ includes 7 dummy variables regarding the city j's size, and 9 dummy variables showing the region that city j locates. Edu_{eij} contains variables that control for the detailed educational attainment of each worker. ε_{eij} is the individual error term. This model is applied to three years: 2000, 2006 and 2011.

The dependent variable is the log of the hourly wage ($(W_{eij}(1-\tau_{eij}))$). Hourly wage is calculated by dividing total wage by the total hours worked (total hours worked equals weeks worked times usual hours worked per week[3]). Following Winters (2009), I adjust the wage for federal income tax by computing the federal tax rate (τ_{eij}) for each worker.[4] I use the usual weekly earnings times average number of weeks workers worked in the March CPS of each year to calculate the corresponding annual earnings. As a result, the average number of weeks worked

[2] Cities in this research are defined by the IPUMS metarea variable.

[3] For the year 2011, the exact number of weeks worked are not reported, and instead a interval of weeks worked is reported; the number of weeks worked is calculated as the mean value of the reported interval of weeks worked.

[4] For a detailed description of adjusting wages for federal tax, please see Winters (2009)

are 48 for year 2000, 48.3 for year 2006, and 48.6 for year 2011. Once the federal tax rate is calculated, the after-tax hourly wage is computed as $(W_{eij}(1-\tau_{eij}))$.

The city price level, P_j, is a weighted average of the housing price index and the non-housing price index, and weights are chosen based on the shares of housing and non-housing expenditures computed according to the Consumer Expenditure Survey of the corresponding years. As a result, the housing price is given a weight of 0.283 for year 2000, 0.292 for year 2006, and 0.286 for year 2011; non-housing price is given a weight of 0.717 for year 2000, 0.708 for year 2006, and 0.714 for year 2011.

Following Winters (2009), the housing price used in this paper is computed based on quality-adjusted rent. In order to calculate the quality-adjusted rent for each city in the sample for a given year, the first step is to regress log gross rent r, for each housing unit on a vector of housing characteristics, T_{hj}, and a vector of city-specific fixed effects, δ_j:

$$\ln r_{hj} = \beta T_{hj} + \delta_j + u_{hj} \qquad (1.5)$$

where h denotes individual house and j denote the city where the house is located.

The housing characteristics included are dummy variables for the number of bedrooms, the total number of rooms, the age of the structure, the number of units in the building, modern plumbing, modern kitchen facilities. Quality-adjusted rent is then predicted for each city based on the estimated parameters by using the sample mean of the housing characteristics. Then this predicted quality-adjusted rents for each city is divided by its mean across cities and times 100 to create the housing price index based on quality-adjusted rents.

1.4. Data

Non-housing price levels for cities are computed based on four sub-indices from the ACCRA cost of living index: groceries, transportation, healthcare, and miscellaneous goods and services[5]. These four sub-indices are given different weights each year based on their corresponding shares of total expenditure computed from CES[6]. Individual level data come from three samples of the Integrated Public Use Microdata Series (IPUMS). These samples are the American Community Survey's 2006 and 2011 1% sample and the 2000 5% sample. From these samples, I use individual observations aged 16 to 65, and individuals who didn't report income or rent are excluded from the sample[7]. I include part-time and part-year workers. The ACCRA cost of living index does not report price data for all areas in each year, and the cities included in the report also change year by year. Therefore, I have to exclude some cities from the analysis. The samples then contain 202 cities with 1,036,922 workers for 2000, 201 cities with 196,449 workers for 2006, and 202 cities with 213,904 workers for 2011. The workers in the sample cities account for 85.4%, 86.9%, and 84.9% of the total workers[8] in the sample for year 2000, 2006, and 2011, respectively. In this paper, I divided all the workers in each year into 4 groups by their educational attainment: less than high-school, high-school, some college and college and above. The less than high-school group contains workers who have never been to high school and high school dropouts. The high-school group contains workers who finished high school or have equivalent educational attainment, but did not go to college. The some college group contains workers who got an associate degree or have been to college but did not get a bachelor's degree.

[5] Utilities are not included because housing rent already contains a large part of utilities.

[6] The weight are: for 2000, groceries(0.124), transportation(0.272), healthcare(0.076), and miscellaneous goods and services(0.528); for 2006, groceries(0.131), transportation(0.248), healthcare(0.081), and miscellaneous goods and services(0.54); for 2011, groceries(0.123) transportation(0.234), healthcare(0.093), and miscellaneous goods and services(0.55)

[7] When choosing the sample, teenagers who live with their parents and don't pay rent are not included in this research. There is still a possibility that some teenagers live with their parents and pay lower rent than they should. Since the housing index in this research is a relative index to the mean of the whole sample, as long as those teenagers are randomly drawn from the population in all the cities included in the research, the bias of the housing index that may be caused by the inclusion of teenagers should not be worried too much about.

[8] The number of total workers here means the total number of workers in the data who are renters and have all the necessary information (rents, wages, and etc.).

The college-and-above group contains workers who got a bachelor's degree or higher. Because the gross rent for house owners are imputed rents[9], which is more subjective, so far in this research, only renters are included.

Table 1.1 contains the summary statistics for several price indexes for years 2000, 2006, and 2011.

1.5. Empirical Results

This section presents the empirical results. The reduced-form models in equation (1.4) are estimated with both Ordinary Least Squares (OLS) and Two Stage Least Squares (2SLS) with an instrument for prices to account for possible measurement errors in the ACCRA cost of living price index. In all the regressions, the individual characteristic variables, amenity variables, industry, occupation and job category dummy variables, detailed educational attainment variables and city size and region dummies are included. For both OLS and 2SLS, equation (1.4) is first estimated with all the observations and then for each education group.

1.5.1. Ordinary Least squares

Table 1.2 shows the results based on OLS for all workers for years 2000, 2006, and 2011. Based on the OLS results, we can see that the wage-price elasticity is significantly different from one at 1% level for all three years, which suggests that, for all the workers, the full compensation hypothesis generally does not hold in these years.

Table 1.3 contains the results based on OLS for different education groups for years 2000, 2006, and 2011.

In Table 1.3, Panel A contains the results for the year 2000. It shows that all four groups of workers are incompletely compensated, though workers with educational attainment equal to high school and some college are compensated roughly the same. Workers in the less-than-high-school

[9] Imputed rents are attained by asking house owners how much they would like to pay for renting their own house.

group are the most incompletely compensated. Results for year 2006 are in Panel B, the estimated θ generally increases with educational attainment. The wage-price elasticity for four groups of workers is significantly different from unity. Like year 2000, this suggests that all workers are incompletely compensated. The estimated wage-price elasticity for workers belonging to some-college and college-and-above groups is 0.874 and 0.85, respectively; while the wage-price elasticity for high-school graduates and less-than-high-school group is 0.755 and 0.597, respectively. The estimated elasticity is significantly smaller for workers belonging to the less-than-high-school and high-school groups, suggesting that they are more incompletely compensated. Panel C contains the results for year 2011. The empirical results for year 2011 suggest that the wage-price elasticity increases with educational attainment. Workers with higher educational attainment are more compensated than workers with lower educational attainment. The wage-price elasticity for all groups is significantly less than unity which implies that all workers are incompletely compensated.

Overall, the OLS results for all workers and the four education groups of workers suggest that full compensation hypothesis does not hold in 2000, 2006 and 2011. Although the wage-price elasticity for all workers is significantly less than one in all three years, workers with higher educational attainment are generally more fully compensated than workers with lower educational attainment.

1.5.2. Instrumental Variables Estimation

It is likely that the two components of the price index – the quality-adjusted rent and non-housing price based on the ACCRA cost of living index – have measurement error. Housing rents may be subject to some degree of sampling error and the ACCRA cost of living index may contain measurement errors that come from different sources. Given that the wage price elasticity is expected to be positive, the measurement error will cause the estimated wage-price elasticity based on OLS to be downwardly-biased towards zero. It is possible that OLS will give an estimated wage-price elasticity that is significantly less than unity for workers while those workers are actually full compensated. Instrumental variables should be used to address the measurement error in the price indices.

Winters (2009) argued that lagged components of the price index can be used as the instruments, and as long as the measurement error is random this will yield consistent estimate of the wage-price elasticity. However, if the measurement errors are serially correlated, the θ estimated by this method is still not consistent. Due to a number of reasons[10], the measurement error in the non-housing price based on ACCRA cost of living index is more likely to be serially correlated than the housing price index based on quality-adjusted rents. Thus, I choose the lagged housing price to be the instrument. Based on the data available, I use the housing price index based on quality-adjusted rents of years 1990, 2005, and 2010, as the instrument variable for price level in years 2000, 2006, and 2010, respectively. This causes the number of cities included in the sample to be reduced to 178 for year 2000[11], while it remains the same for the other two years.

Table 1.4 contains the results based on 2SLS for all workers for years 2000, 2006, and 2011. In all regressions, the test statistics reject the weak instrument hypothesis. As expected, the elasticity estimated by 2SLS using lagged housing price as the instrument is larger than the OLS estimate, confirming that an instrumental variable should indeed be used to take care of the downward-bias caused by measurement error, and the lagged housing price is a reasonable choice. The reason that the partial R^2 of the excluded instrument is considerably smaller for year 2000[12] may due to

---

[10] The ACCRA cost of living data for a given city is collected by people (in many cases the same people) who follow the same method in certain places in the city at different times. The American Community Survey data is collected based on random sampling of individual households. If there exists measurement error, the measurement error in ACCRA cost of living data is more likely to be serial correlated than that in the American Community Survey data.

[11] The total number of observations included in the model is reduced from 1,036,922 to 1,017,736 or a 1.85% decrease in sample size.

[12] This is also true in all the 2SLS models.

the 10-year time difference between 1990 and 2000. The other two cases have only 1-year time difference between the lagged housing price index and the rent-based price index.

Now the wage-elasticity for 2000, 2006, and 2011 are 0.796, 0.864, and 0.826, respectively, and they are still all significantly different from unity at the 1% level. This suggests that, overall, the full compensation hypothesis does not hold for all workers in these three years.

Table 1.5 contains the empirical results for education groups in years 2000, 2006, and 2011. The weak instrument hypothesis is rejected in all the regressions. Panel A contains the results for year 2000. With an estimated wage-price elasticity of 0.917, the some-college group is the only group that has a wage-price elasticity not significantly different from unity; while the wage-price elasticity for all the other groups is found to be significantly less than unity. As a result, in year 2000, workers belonging to some-college group are found to be fully compensated, while workers in other groups are found to be incompletely compensated. Panel B contains the results for year 2006. The wage-price elasticity for Some-college group and College-and-above group is 0.969 and 0.9, respectively, and both not significantly different from unity; while the wage-price elasticity for Less-than-high-school group and High-school group is 0.636 and 0.815 respectively, and both significantly less than one. These empirical results imply that, in 2006, workers with educational attainment equal to and higher than some college are fully compensated, while workers with educational attainment less than some college are incompletely compensated. Empirical results for year 2011 are shown in Panel C. Similar to year 2006, the most educated two groups: Some-college group and College-and-above group are found to be fully compensated, while their less educated counterparts: Less-than-high-school group and High-school group are found to be incompletely compensated.

Based on the above 2SLS results, the full compensation hypothesis holds for workers with educational attainment equal to or higher than some college, except for the college-and-above group in year 2000. On the contrary, the full compensation hypothesis does not hold for workers with educational attainment lower than some college in all three years. For the more recent years-2006 and 2011, the wage-price elasticity generally increases with educational attainment; the wage-price elasticity for workers in the two least educated groups is significantly less than unity, while the wage-price elasticity for workers in the two most educated groups are not significantly different from unity. This means that, in the more recent two years, the more educated a worker is, the more compensated he will be, and the full compensation hypothesis holds only for workers whose educational attainment is at least equal to some college.

The previous empirical results are based on using the same weights for different groups of workers in a given year. However, the weights for the housing index, non-housing index, and the components for non-housing index may differ not only across years, but also across different groups of workers.

From the following equation:

$$d\ln W_i = (P_H H_i / W_i) * d\ln P_H + (P_X X_i / W_i) * d\ln P_X - (P_A / W_i) * dA$$

where i denotes different groups of workers. "$P_H H_i / W_i$" and "$P_X X_i / W$" is the weight of housing index and non-housing index for the i^{th} group of workers, respectively. "$P_H H_i / W_i$" and "$P_X X_i / W$" may be different for different group of workers. A set of constant weights for all the workers in a given year may not be appropriate.

It is possible that conclusions about the compensation status of different workers may not be very accurate when we rely on empirical results based on the same weights for all workers in a given year. Group-specific weights for different groups of workers should be used for a robustness check. Based on CES data for the four groups of workers for years 2000, 2006, and

2011, I recalculate the weights for each group in every year[13] and use 2SLS to re-estimate the base model.

Table 1.6 contains the empirical results based on group-specific weights and 2SLS for all workers for years 2000, 2006, and 2011. Test statistics reject the hypothesis of weak instruments in all regressions. Now the wage-price elasticity for 2000, 2006, and 2011 are 0.823, 0.982, and 0.946, respectively. Unlike the empirical results based on the same weight, only the wage-price elasticity for all workers in year 2000 is found to be significantly less than one, while the wage-price elasticity for all workers in year 2006 and 2011 is found to be not significantly different from one. These results suggest that full compensation hypothesis holds for all workers in the more recent years-2006 and 2011.

Table 1.7 contains the empirical results based on group-specific weights and 2SLS method for education groups in years 2000, 2006, and 2011. The weak instrument hypothesis is rejected by test statistics in all models. Panel A contains the empirical results for year 2000. Now the Less-than-high-school and College-and-above groups are still found to be incompletely compensated, while the High-school and Some-college groups are found to be fully compensated. Similar to the results in Panel A of Table 1.5, the full compensation hypothesis does not hold for workers with educational attainment equal to or higher than college. Wage-price elasticity for workers with educational attainment equal to or higher than college is smaller than that for workers with educational attainment equal to high-school or some college. Panel B shows the empirical results for year 2006. Similar to Panel B of Table 1.5, workers belonging to the Less-than-high-school group are still found to be incompletely compensated, while workers with educational attainment equal to or higher than some college are found to be fully compensated. Unlike Panel B of Table 1.5, workers belonging to the high-school group are found to be fully compensated. The implications of the empirical results in Panel C of Table 1.7 are consistent with those based on

[13] Detailed weights for each price category for each groups of workers in different years are shown in Appendix table 1.8

Panel C of Table 1.5. Workers belonging to the two most educated groups are found to be fully compensated, while workers belonging to the two least educated groups are found to be incompletely compensated.

Although the weak instrument hypothesis is rejected in all the regressions in Table 1.7, the partial R-squares of excluded instruments in all the regressions in Table 1.7 are significantly smaller than the partial R-squares of excluded instruments in their corresponding regressions in Table 1.5. Unless we can solve the direct utility function for different groups of workers in all years, we can't find the "perfect" weights for all the sub-category indexes of the city price indexes for different groups of workers. By comparing the 2SLS empirical results based on same weights and group-specific weights, we can see that, in the more recent year-2006 and 2011, the wage-price elasticity generally increases with educational attainment. For 2006 and 2011, workers whose educational attainment is equal to or higher than some college are full compensated; workers whose educational attainment equal to high school may be incompletely compensated; and workers whose educational attainment is less than high school are incompletely compensated. A larger wage-price elasticity for more educated workers indicating that they are more compensated. More educated workers are usually more skilled and are very likely to be more mobile. It easier for them to move to places where they can be more compensated than their less educated counterparts. Studies have found evidence that worker mobility increases with educational attainment. Wozniak (2010) found that highly educated workers are better at locating in areas with high labor demand, and medium-term wage premium of entry labor market for college graduates equal or exceed those of less educated workers. Malamud and Wozniak (2012) found that there exists a causal impact of higher education on migration and additional years of college significantly increased the likelihood that men resided outside their birth states later in life.

1.6. Conclusion

In this paper, I examine whether the full compensation hypothesis holds for workers in different labor markets. Unlike previous research, I use both same and group-specific weights for the sub-category indexes for different groups of workers. Previous studies also just looked at one year, but because the spatial equilibrium assumption may not be achieved in a particular year, I examine three different years: 2000, 2006, and 2011. Empirical results from 2SLS (based on the two sets of price weights) do not suggest the exact same implications about the compensation status of different workers. But unlike previous studies, I still found that the wage-price elasticity generally increases with educational attainment, especially in 2006 and 2011. My empirical results support the full compensation hypothesis for workers with educational attainment equal to or higher than some college in 2006 and 2011. Workers whose educational attainment is lower than high-school are found to be incompletely compensated. Workers who graduated from high school or have an equivalent diploma are found to be incompletely compensated in most of the cases, and fully compensated only in 2000 and 2006 when group-specific price weights are used.

One important reason for the more compensated status of higher educated worker is they are more mobile. They are more likely to move to places where they can be fully compensated. In terms of policy suggestions, policies that can encourage mobility of less educated workers are needed. Government can provide better information on job opportunities, so that workers are more likely to either find jobs in the same place with higher wages or find jobs in new places where they can be more compensated for the cost of living. Government can improve relocation services, so it will cost less for workers to move to places where they can be more compensated for the cost of living. Post-school education and professional training are also important for less educated workers who never go to college. Government can start or support post-school high school education opportunities, and post-high school education and professional training opportunities. Workers whose educational attainment are equal to high-school can utilize the latter opportunities, and workers whose never go to high-school can utilize all the opportunities.

As these workers become more educated, they are more likely to move to places where they can be more compensated for the cost of living, and their individual welfares will be improved.

CHAPTER II

Regional divergence in China:
Evidence from the PANIC Approach with bias reduction

2.1. Introduction

The issue of regional economic convergence in China is re-investigated in this paper using the Panel Analysis of Nonstationarity in Idiosyncratic and Common (PANIC) components approach proposed by Bai and Ng (2004). Most of previous studies tested for a unit root in the provincial level per capita GDP to address this issue. So far, all these studies that applied time-series techniques assume cross-section independence across provinces. But the results from the cross-section independence test proposed by Pesaran (2007) suggest cross-section dependence among the provincial level per capita GDP data in China. It has been shown that when unit root tests assume cross section independence are applied to cross-section dependent data, the hypothesis of a unit root in the data is over rejected. So, panel unit root tests that incorporate cross-section dependence should be used to re-investigate regional economic convergence in China.

In this paper, I apply Bai and Ng's (2004) PANIC method to study the economic convergence in China for the "post-reform" period that starts since 1978. The PANIC approach allows me to test the null of a unit root in the common factor and idiosyncratic component separately. Thus, this gives a clearer picture about where the stationary or nonstationary in the income data comes from. So we can have a better understanding of the economic divergence or convergence in China. The common factor can capture the influence of the central government of China[14] on a province's economic development. If nonstationary is found in the common factor for China or the corresponding region, it means that the common factor is not leading to the economic convergence in China or that region. This is likely suggesting that economic policy and plans that are implemented by the central government are not helping to narrow the income gap among different provinces in China or that region. If stationarity is found in the idiosyncratic components, without the influence of the common factor, regional economic convergence is indicated. The empirical results show a unit root in all the common factors and most idiosyncratic components, which suggest regional economic divergence in China during the post-reform period. These findings have important implications for economic development policy.

Economic convergence in China is of particular interest. One very important slogan that characterizes China's "Reform and Open" policy is "let some people and places become rich first and then eventually make all people and places rich". Thus, one of the key facts in evaluating the success of China's astonishing economic development during the post-reform period is whether the benefit from the economic growth is shared by most of the people in China. Testing for economic convergence in China can help us understand more about how successful the great economic success has benefitted different regions of China, and what kind of economic development policy should the Chinese central and provincial government consider in the future.

Early empirical studies tried to use cross-section tests based on neoclassical growth models to find evidence for the convergence hypothesis. Baumol (1986), Dowrick and Nguyen (1989),

[14] China's central government plays a very important role in the economy. It makes the Five-year economic development plans, it decides the development priority of certain industries and how infrastructure investment is divided among all the provinces. Once the central government's economic policy is decided, provincial governments have to make their own policy according to it or change their policy to adapt to it, and these policies are always given the priorities in the province. Thus, except for the individual province's own economic policy and conditions, the economy of every individual province is influenced significantly by one common factor--central government economic policy and plans. As a result, province-level per capita incomes in China is likely to contain at least one common factor.

Wolff (1991), Barro and Xavier Salai-Martin (1991,1992), and Mankiw et al. (1992) all find evidence in favor of convergence by examining the cross-sectional relationship between the growth rate of output per capita over some time period and the initial level of per capita output. However, Evans and Karras (1996) showed that this conventional approach is valid under the assumption that economies must have identical first-order autoregressive dynamic structures and very strong assumptions about the cross-economy differences. These assumptions are usually violated in the data. Thus, they suggested using a time series approach that is valid under much less restrictive conditions than conventional cross-section tests. Bernard and Durlauf (1996) argued that time series tests have a stronger notion of convergence than cross section tests.

Most recent empirical studies of economic convergence rely on time series analysis. Under time series framework, the persistence of relative per capita outputs is evaluated and stochastic convergence occurs when the (log) per capita output of a country or region relative to a reference economy or region over a period is found to be stationary.

Time series empirical findings for the economic convergence of China are few and mixed. Zhang et al. (2001) and Weeks and Yao (2003) apply the augmented Dickey and Fuller (ADF) test, with and without trend, and find evidence of intrazonal economic convergence between provinces in the eastern, central and western regions of China, but divergence between regions. They conclude that club convergence exists in China, but there is no national convergence.

The conventional ADF test is known to have low power when the time period of the data is not long enough. Panel unit root tests which pool information across all the cross-sections together have better power. Pedroni and Yao (2006) employed the panel unit root test (IPS, MW) and find that per capita output of China of provinces converge before 1978, while diverge after it.

Perron (1989) and Im et al. (2005) showed that ignoring structural breaks in time series or panel data may lead to a substantial loss of power or serious size distortions in the ADF, IPS, and

MW tests. Liu et al. (2013) applied unit root tests with endogenously determined structural breaks to analyze the economic convergence in different regions in China from 1953 to 2007. They found that there are one or two structural breaks in some provinces, and for those provinces that have structural breaks, most of the breaks happened between 1967 and 1978. They found that economic convergence exists in regions but not between regions.

Although Liu et al. (2013) takes structural breaks into consideration, they still make the assumption that there is no cross-sectional dependence. Schwert (1989), Pantula (1991), and Ng and Perron (2001) argued and analyzed that the unit root test can be oversized and stationary tests will have no power when there exists a common factor in the time series data but is not separated from idiosyncratic components.

The PANIC approach proposed by Bai and Ng (2004) has been used in empirical research in many economic topics, such as unemployment, real interest rate, consumption and wealth relationship, interest parity, money demand, export and output and etc. (see, for example, Nagayasu, 2011; Cheng et al., 2012; Dreger and Reimers, 2012; Dobnik, 2013; Dreger and Zhang, 2014; Everaert, 2014). It is a very powerful tool to analyze cross-sectional dependent time series data. It is even referred to as "one of the single most popular and general second-generation approaches around."[15] In this study, the PANIC approach is applied to the provincial level per capita GDP time series of China (1978-2012) to estimate the common factor and idiosyncratic component. Then unit root tests are carried out for the common factors and idiosyncratic components. In order to reduce the downward bias in the unit root test for the common factor, the Recursive Mean Adjustment (RMA) unit root test that proposed by Shin and So (2001) is applied. Unlike previous studies which found evidence in favor of sub-region economic convergence in China, my empirical results imply that, in general, regional economic divergence prevails in

[15] See Westerlund, J. 2015. " The power of PANIC." Journal of Econometrics 185, 495-509.

China and its three sub-regions. Not only is there no economic convergence between different sub-regions, but there is also no economic convergence within the sub-regions.

2.2. Data

The data for this study are drawn from the China Compendium of Statistics (1949-1999) and the China Statistical Yearbook (2000-2012). This research is based on 28 provinces[16] that are in mainland China. Data are available for those 28 provinces from 1953-2012. But the research period in this paper is from 1978-2012. There are two main reasons: 1) Based on the research by Liu et al. (2013), for most of the cases, a structural break at the provincial level occurs between 1967-1978. So far, there is no PANIC with endogenous determined structural break approach. In order to avoid the bias that structural breaks in the provinces time series may cause, the period of 1978-2012 is chosen. 2) The Chinese government chooses very different economic policy, and openness to international trade pre- and post-year 1978. After 1978, the Chinese government adopted the "Market Mechanism" to substitute for a lot of central planning in the economy of China, and it has been made very clear that the Chinese government will continue the policy. So, testing the economic convergence before 1978 is interesting but may not have enough useful implications for understanding the current and future regional growth patterns in China.

Based on the basic geographic information for Chinese provinces, the 28 provinces are divided into three sub-regions: The east region, the central region, and the west region. The east region contains 10 provinces and municipalities that are on the east coast of China. The central region contains 9 provinces that are not on the coast but share the boundary with a coastal province. The West region contains 8 provinces that are neither on the coast nor have a common boundary with a coastal province. The 11 provinces in the east region are: Beijing, Tianjing, Hebei, Liaoling, Shanghai, Jiangsu, Zhejiang, Shandong, Fujian, Guangdong, and Guangxi. The 9 provinces in

[16] Provinces and equivalent contain provinces, autonomous regions, and municipalities directly under the central government. In China, these areas are all considered as provincial level areas. For convenient, they will all be referred to as provinces later in this paper.

the central region are: Shanxi, Inner Mongolia, Jilin, Heilongjiang, Anhui, Jiangxi, Henan, Hubei, and Hunan. The 8 provinces in the west region are: Sichuan, Guizhou, Yunnan, Shaaxi, Gansu, Qinghai, Ningxia, and XinJiang.

Since 1978, provinces in different sub-regions of China have experienced very different economic development speed, and the supports from central government's policy are also different among the sub-regions. Even in the same sub-region, the support that each province can get from the central government is still different, although the magnitude of difference is not as large as that for different regions. After more than 30 years, do people live in the 28 provinces enjoy a more evenly distributed benefit from the economic development? Do people live in different provinces in the same sub-region enjoy a closer standard of living than before? Testing for economic convergence in China would show whether there is economic convergence across different regions in China. Testing for economic convergence in the three sub-regions would show whether there exist "regional income clubs". The PANIC approach is applied to a panel that consists of all the provinces in China, and panels that include those corresponding provinces in its three sub-regions separately. For each sub-region, the whole process of PANIC is applied for provinces only in that region. The test results imply economic divergence/convergence: 1) for all the provinces in China, if stationary is found for the common factor, it suggests that central government policy is helping the economic convergence at the provincial level, and vice versa. If stationary is found for the panel of the idiosyncratic components, it suggests that, without other interventions, there exist economic convergence among the provinces, and vice versa. 2) for the panel of provinces in a sub-region of China, stationary for the common factor of that sub-region suggest that central government policy is helping the economy converge at the provincial level in that sub-region, and vice versa. Stationary for the panel of the idiosyncratic components suggests that, without other interventions, the economy converges among provinces in that region, and vice versa. 3) If stationary is found for the idiosyncratic component of an individual province, it suggests that the economy of that province converges to the mean of the region[17] it belongs to.

The Great Recession in the United States had a great impact on the economy of China. The Chinese Central government proposed a 4 trillion RMB fiscal stimulus package in 2008 and many other polices to help the economy keep growing. Since the infrastructure base and involvement in international trade are different across the 28 provinces, the impact on the Great Recession on the provinces varies a lot, and so does the benefits that provinces could get from the central government's stimulus package and other policies. All these mean that the Great Recession may have altered the pattern of regional economic growth in China. Thus, in this paper, the regional economic convergence is studied for two periods: 1978-2007 and 1978-2012.

Real per capita GDP for each province is computed using 1995 as the base year for the GDP deflator, and then I take the natural log of real per capita GDP to get the logged time series data that are needed for testing economic convergence.

2.3. Econometric model

Let $y_{ir,t}$ be real per capita GDP for province i in region r at period t and there are N provinces in region r. Define $Y_{ir,t} = \ln y_{ir,t} - \overline{\ln y_{r,t}}$, where $\overline{\ln y_{r,t}} = \frac{1}{N}\sum_{i=1}^{N} \ln y_{ir,t}$ is the cross sectional mean of the real per capita GDP for the n provinces in region r at time t. Thus, $Y_{ir,t}$ is the deviation of real per capita GDP (log) of province i from the mean real per capita GDP (log) of the region to which it belongs. According to Evans and Karras (1996), stochastic convergence exists if $Y_{ir,t}$ is found to be stationary. Thus, to test the economic convergence means to test the stationary of $Y_{ir,t}$.

[17] The region can be China or the sub-region the province belongs to, depending on which panel of provinces are studied by the PANIC approach.

First, whether there exists cross-sectional dependence among $Y_{ir,t}$ should be tested. In this paper the cross-sectional independence test proposed by Pesaran (2007) is applied for the period 1978-2012. Table 2.1 shows the results for the test.

The results of the cross-section independence test show that the null hypothesis of cross-section independence is rejected for all the regions, and thus implies cross-section dependence between $Y_{ir,t}$

2.3.1. PANIC Approach

Following Bai and Ng (2004), $Y_{ir,t}$ (i=1, ...N; t=1,...,T) is assumed to be generated by

$$Y_{ir,t} = c_i + \lambda_i' F_t + e_{ir,t} \qquad (2.1)$$

$$(I - L)F_t = C(L)u_t \qquad (2.2)$$

$$(1 - \rho_i L)e_{it} = D_i(L)\varepsilon_{it} \qquad (2.3)$$

where F_t is the common factor vector and e_{it} is the idiosyncratic component. L is the lag operator, $C(L) = \sum_{j=0}^{\infty} C_j L^j$ and $D_i(L) = \sum_{j=0}^{\infty} D_{ij} L^j$. $u_t \sim iid(0, \sum_u)$, for each i, $\varepsilon_{it} \sim iid(0, \delta_{\varepsilon i}^2)$. The idiosyncratic error $e_{ir,t}$ is I(1) if ρ_i=1, and is stationary if $|\rho_i|$<1.

To apply the PANIC approach, we must first decide the number of common factors. Bai and Ng (2002) propose the following way of determining the number of common factors. Assume that $Y_{ir,t}$ is generated by the data generating process (2.1), (2.2), and (2.3), let k be the number of common factors, let λ_{Nr}^k denote the factor loading for the k common factors of the N^{th} province in region r and define $\Lambda = (\lambda_{1r}^k, \lambda_{2r}^k, ..., \lambda_{Nr}^k)'$. The common factors are estimated by the method of principal components. Estimates of λ_{ir}^k and F_t^k are obtained by solving the optimization problem:

$$V(k) = \min_{\Lambda, F^k} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (Y_{ir,t} - \lambda_{1r}^k F_t^k)^2$$

The following two criteria[18] can be used to determine the number of common factors:

$$IC_1(K) = \ln(V(k, \hat{F}^k)) + k(\frac{N+T}{NT})\ln(\frac{NT}{N+T});$$
$$IC_2(K) = \ln(V(k, \hat{F}^k)) + k(\frac{N+T}{NT})\ln(\frac{NT}{N+T})\ln C_{NT}^2;$$

where $C_{NT}^2 = \min\{N, T\}$.

Table 2.2 shows the results for the information criteria.

Based on the results, for all the regions and the two different time periods (1978-2007 and 1978-2012), it is determined that one common factor exists. The existence of the common factor is likely caused by the prominent role of the Chinese central government in the economy of China.

Once the number of common factors has been determined to be one, the PANIC approach can be carried out as follows. Define $dy_{ir,t} = \Delta Y_{ir,t}$, $f_t = \Delta F_t$, and $\xi_{ir,t} = \Delta e_{ir,t}$.

Then the model in first-differenced form is

$$dy_{ir,t} = \lambda_i^{'} f_t + \xi_{ir,t}$$

After applying the method of principal component to $dy_{ir,t}$, we can get the estimated factors $\hat{f}_t$, the loadings $\hat{\lambda}_i$ and the estimated residuals, $\hat{\xi}_{ir,t} = dy_{ir,t} - \hat{\lambda}_i^{'} \hat{f}_t$. Then for t=2,...,T, define :

$$\hat{e}_{ir,t} = \sum_{s=2}^{t} \hat{\xi}_{ir,s} \qquad \text{(i=1,...,N)}$$

$$\hat{F}_t = \sum_{s=2}^{t} \hat{f}_s$$

Bai and Ng(2004) showed that $e_{ir,t}$ and F_t can be consistently estimated in this way. Then, according to Bai and Ng (2004), the ADF test with no deterministic terms and no trends can be

[18] Bai and Ng(2002) proposed three criterias, but the two that are used in this paper apply specifically to the principal components estimator.

applied to test the stationary of each idiosyncratic component and the ADF test with an intercept but with no trends can be applied to test the stationary of the common factor.

The ADF test for $\hat{e}_{ir,t}$ is:

$$\Delta \hat{e}_{ir,t} = \beta_{i0} \hat{e}_{ir,t-1} + \beta_{i1} \Delta \hat{e}_{ir,t-1} + ... + \beta_{ip} \Delta \hat{e}_{ir,t-1} + error \qquad (2.4)$$

The ADF test for $\hat{F}_t$ is:

$$\Delta \hat{F}_t = c + \gamma_0 \hat{F}_{t-1} + \gamma_1 \Delta \hat{F}_{t-1} + ... + \gamma_q \Delta \hat{F}_{t-q} + error \qquad (2.5)$$

where p and q are the optimal lags chosen by certain information criteria.

Two popular information criteria that have been used extensively to choose the optimal lags for ADF tests are AIC and SBIC.

For a regression $x_t = \alpha x_{t-1} + \sum_{j=1}^{p} \beta_j \Delta x_{t-j} + error$

$$SBIC = -2 * \ln(LL) + \ln(T) * k$$

where LL is the likehood, T is the length of the time period and k is the model degrees of freedom. In this research, SBIC[19] is used to choose the optimal lag.

Once the ADF test in Equation (2.4) has been applied to all the idiosyncratic components, a panel unit root test statistic for these idiosyncratic terms can be constructed as follows:

$$P_{\hat{e}} = \frac{-2\sum_{i=1}^{N} \log P_{\hat{e}}(i) - 2N}{\sqrt{4N}} \xrightarrow{d} N(0,1)$$

[19] I also use AIC to choose the optimal lag. For most of the cases, these two information criteria choose the same lags. Even when they choose different lags, the resulted significant level of the ADF test is the same.

where, $P_{\hat{e}}(i)$ is the p value for the ADF test for the ith idiosyncratic component and N is the number of cross sections.

2.3.2. RMA approach

Consider the following stochastic process for y_t :

$$y_t = c + \alpha y_{t-1} + \varepsilon_t \qquad (2.6)$$

where c is an intercept, α is the persistence parameter and ε_t is a white noise process. It is well-known that the least squares estimator for α is downwardly biased, and the bias is very severe when the data exhibits high persistence. The bias comes from the presence of the constant in Equation (2.6). To observe the bias, convert Equation (2.6) into its demeaned equivalent:

$$y_t - \overline{y} = \alpha(y_{t-1} - \overline{y}) + \mu_t$$

where $\overline{y} = \frac{1}{T}\sum_{s=1}^{T} y_s$, and $\mu_t = \varepsilon_t + c - (1-\alpha)\overline{y}$. y_s is correlated with ε_t for any $s \geq t$. Therefore, the demeaned regressor $(y_{t-1} - \overline{y})$ is correlated with μ_t and the OLS estimator is biased.

Instead of using $\overline{y}$ in the demeaned regression, the RMA estimator utilizes the recursively adjusted mean to reduce bias:

$$y_t - \overline{y}_{t-1} = \alpha(y_{t-1} - \overline{y}_{t-1}) + \nu_t$$

where $\overline{y}_{t-1} = \frac{1}{t-1}\sum_{s=1}^{t-1} y_s$. Now ε_t is orthogonal to the recursive mean adjusted regressor $(y_{t-1} - \overline{y}_{t-1})$, which results in a substantial bias reduction.

For the ADF test with a constant for y_t, Shin and So (2001) proposed the following recursively mean adjusted regression equation:

$$\Delta y_t = \rho(y_{t-1} - \bar{y}_{t-1}) + \sum_{j=1}^{p} \theta_j \Delta y_{t-j} + error \qquad (2.7)$$

The RMA-based ADF t-statistic can be constructed for $\hat{\rho}$ for the null hypothesis of a unit-root.

The OLS estimate of ρ will be downwardly biased away from zero. As a result, the t-statistic for the ADF test will be larger in magnitude than its real value. Shin and So (2001) showed that the RMA-based unit root test is more powerful than the OLS-based test.

So for Equation (2.5), the RMA-based ADF test will be:

$$\Delta \hat{F}_t = \gamma_0(\hat{F}_{t-1} - \bar{\hat{F}}_{t-1}) + \gamma_1 \Delta \hat{F}_{t-1} + \ldots\ldots + \gamma_q \Delta \hat{F}_{t-q} + error \qquad (2.8)$$

where $\bar{\hat{F}}_{t-1} = \frac{1}{t-1}\sum_{s=1}^{t-1} \hat{F}_s$. This is the test for $\hat{F}_t$ that is utilized in this research. Because there is no constant in Equation (2.4), the RMA-based ADF test is not needed for the unit root test of the idiosyncratic components.

2.4. Empirical Results

2.4.1. Stationary of the Common Factor

Table 2.3 shows the unit root test results for the common factor for China and its three sub-regions for 1978-2007 and 1978-2012 under both the conventional ADF test and the RMA-based ADF test. As predicted, the t-statistic for the conventional ADF test is significantly larger in magnitude than that for the RMA-based ADF test. The common factor for China and the central region is found to be nonstationary for both the period even with the downwardly-biased ADF test. Based on the ADF test for the common factors, the common factors for the east and west regions are stationary for both periods. However, with the bias-reduced RM- based ADF test, the common factor for these two regions are found to be nonstationary in the two periods.

Overall, according to the RMA-based ADF test, common factors are not stationary in China and all three sub-regions, which suggest in general that a determinant factor above all the provinces is not eliminating the income dispersion between rich and poor provinces.

2.4.2. Stationary of the Idiosyncratic Component

Table 2.4 shows the unit root test results for the idiosyncratic component of 28 provinces when the PANIC approach is applied to the panel of all provinces in China. The panel unit root test statistics for idiosyncratic components $P_{\hat{e}}$ show that there exists a unit root in the panel of the 28 provinces for the two periods. The individual unit root test for each province finds that, only two provinces, Guangxi and Shanxi, have a stationary idiosyncratic component for the period 1978-2007.

For the period 1978-2012, only the idiosyncratic component for Guangxi is found to be stationary, and Shanxi in now found to have a unit root in its idiosyncratic component. For all the other provinces, the empirical results support the existence of nonstationarity in their idiosyncratic component for both periods.

The result for the idiosyncratic component taken together with that for the common factor implies that the time series $Y_{ir,t}$ for most provinces (except Guangxi and Shanxi in 1978-2007, and Guangxi in 1978-2012) consists of a nonstationary common factor and a nonstationary idiosyncratic component. This means that, overall, economic convergence does not exist at provincial level in China. This finding is consistent with previous studies which found that intra-regional economic convergence does not exist in China. However, previous studies usually found

$Y_{ir,t}$ to be stationary for more provinces and this may due to the fact that these studies assume cross-section independence among $Y_{ir,t}$ s.

Table 2.5 shows the unit root test results for the idiosyncratic component of 11 provinces when the PANIC approach is applied to the panel formed by the 11 provinces in the east region. The fact that $P_{\hat{e}}$ is not significant at the 10% level for both 1978-2007 and 1978-2012 suggests that there exists a unit root in the panel of the idiosyncratic component in east region. Only the idiosyncratic component of Guangxi is found to be stationary during 1978-2007, but it is found to have a unit root during 1978-2012. The idiosyncratic components for the other provinces are found to be nonstationary during both 1978-2007 and 1978-2012.

Unlike previous studies which found economic convergence for sub-regions in China, the empirical results for the idiosyncratic component of the provinces in east region implies that economic convergence does not exist in the east region.

Table 2.6 shows the unit root test results for the idiosyncratic component of 9 provinces while the PANIC approach is applied to the panel formed by the 9 provinces in the central region.

The result for the panel unit root test for the idiosyncratic component shows that there is a unit root in the panel. No idiosyncratic component is found be to stationary in the two periods. These results imply economic divergence among provinces in the central region.

Table 2.7 shows the unit root test results for the idiosyncratic component of 8 provinces while the PANIC approach is applied to the panel formed by the 8 provinces in the West region.

Like the other two sub-regions, the empirical results for $P_{\hat{e}}$ support a unit root in the panel. The idiosyncratic components for Gansu and Ningxia are found to be stationary during 1978-

2007, but nonstationary during 1978-2012. Again, economic divergence is supported by the empirical results in the west region.

2.4.3. Economic divergence in China

Unit root test results based on the PANIC approach for the 28 provinces of China suggest that there is no economic convergence among the 28 provinces during 1978-2007 and 1978-2012. Because the 28 provinces belong to the three sub-regions, no economic convergence among the 28 provinces also means no inter-region economic convergence. Unit root test results based on the PANIC for the corresponding provinces in each sub-region imply that there is no economic convergence among those provinces in the sub-region that they belong to. Economic divergence is found in China and its three sub-regions during 1978-2007 and 1978-2012. Unlike previous research which found economic divergence among regions, but found economic convergence for sub-regions, this paper found economic divergence both intra-region and inter-region. Empirical results for the panel of whole China tell us that there is no economic convergence between Beijing in the east region and Yunnan in the West region (or between Shanghai in the east region and Anhui in the central region). Empirical results for east region suggest there is no economic convergence between Beijing and Guangxi, even though they are in the same sub-region, and this economic divergence among provinces in the same sub-region is also true for the other two sub-regions.

The reason that previous research usually found economic convergence intra-region may due to: 1) they assume cross-section independence in the data. Once cross-section dependence among provinces is ignored, unit root tests will over rejected the hypothesis of economic divergence. 2) the downward bias caused by the constant in the ADF style regression. Conventional ADF, IPS, and LM tests with structural breaks are all based on AR(p) process, as long as constant is include, there will be downward bias in least square estimates and the bias can be very severe when there is high persistence. For Equation (2.5), the OLS estimate of γ_0 will be downwardly biased away from zero, leading to an over rejection of the null hypothesis of a unit root, and economic convergence is more likely to be supported. This is illustrated by the significant difference between the t-statistic for the conventional ADF test and the RMA-based ADF test for the common factors.

2.5. Conclusion

The Panel Analysis of Nonstationarity in Idiosyncratic and Common components (PANIC) approach is applied in this paper to investigate economic convergence in China and its three sub-regions. Under the structure of the PANIC approach, the recursive mean adjusted (RMA) method has been applied for the unit root test of the common factor to reduce bias. The common factor for China and its three sub-regions are found to be nonstationary, which implies pervasive nonstationarity in China's economy. For most of the time, the idiosyncratic component for individual provinces are found to be nonstationary, especially during 1978-2012, only Guangxi's idiosyncratic component to found to be stationary.

Unlike previous research which found intra-regional economic convergence in China. In this paper, the PANIC approach with bias reduction method finds evidence in favor of economic divergence both intra-region and inter-region. The unit root in the common factor possibly implies that central government policy is not eliminating income dispersion among provinces in China. The nonstationary in idiosyncratic components suggests that, even without the common factor, economic inequality between rich and poor provinces will not narrow with the current economic conditions. In order to make the economic development in China to be shared more evenly among people in different regions, at least province-based policy is needed. According to the findings of this research, central government policy that targets at a sub-region that includes some provinces does not seem to help the economy converge at the provincial level in that sub-region. The central government in China can make and implement province-based policy such as more investment in the infrastructure, lower taxes, better benefits for skilled workers, and more financial support for less developed provinces. The central government also can help provinces in central and west regions of China to build up more trade connections with other countries and regions in the world. The more open those provinces are to international trade, the faster economic growth they will enjoy. This paper shows the economic divergence in China during the post-reform period. Future studies can focus on the effectiveness of difference policies in reducing the economic inequality among provinces and regions in China.

CHAPTER III

Identification and the Reflection Problem in Spatial Econometrics: An Application using U.S. Counties

3.1. Introduction

Applied spatial economic research is known to suffer from identification problems. In many of the cases, empirical research based on spatial econometric models only provide correlations of the key variables among spatial units but no credible information about the causal relationships in the economic processes (Pinkse & Slade, 2010 and Gibbons & Overman, 2012). In this paper, I put identification and causality at the center, while studying how employment growth of a county is affected by employment growth, initial fiscal policy variables and certain initial conditions of other counties in the same commuting zone (CZ).

County employment growth is very important. For the welfare of individual residents in a county, losing a job has a very negative impact on individual residents' lives. Guaranteed county employment growth means guaranteeing the welfare of the residents in the county. The county is the basic geographic unit of the nation. County employment growth is the basis for national employment growth. Counties in the same commuting zone are thought to be in the same labor market[20]; studying the interaction mechanism among counties in the same commuting zone that affects county employment growth has important theoretical and policy implications. Previous studies have found spillover effects of employment growth and fiscal policies at the county level. Khan et al. (2001) and Desmet and Fafchamps (2005) found spatial spillovers among county employment levels. Shuai (2010) found substantial spillover of county fiscal benefits associated with job creation in Virginia. But this research is based on the conventional Spatial-Durbin model, which suffers from serious identification issues.

The conventional Spatial-Durbin model and related models are used in many empirical studies, and those models are essentially parallel with linear-in-mean neighborhood effect models (Lee 2004, 2007). Manski (1993) has already showed that this type of model has the "reflection problem" which means that the model cannot provide identification of the causal effect of the average behavior or characteristics of a group on the behavior of the individual agent that belongs to that group. Unfortunately, this identification problem does not draw enough attention among empirical studies based on the Spatial-Durbin model and its variations. There are few studies that discuss the identification issue in spatial econometric models and how to carry out valid spatial econometric research. Pinkse and Slade (2010) discussed the problems of spatial econometrics that are yet unsolved and advise that new spatial econometric theory should be inspired by actual empirical applications, and not be directed by what appears to be the most obvious extension of what is currently available. Gibbons and Overman (2012) discussed the characteristics and identification problems of different types of spatial econometric models and argue that, in many cases, conventional spatial econometric approach is uninformative about the causal economic process at work. They suggest that future spatial econometric research should focus on "providing

[20] Commuting zones are defined by the United State Department of Agriculture to combine counties into units intended to more closely reflect the geographic interrelationships between employers and labor supply. For more details, see http://www.ers.usda.gov.

credible estimates of causal processes that can guide understanding of our world, and guide policy makers on how to change it."[21]

In the spirit of "new spatial econometric theory that should be inspired by actual empirical applications" and "providing credible estimates of causal processes", in this research, I use the G2SLS approach proposed by Lee (2007) and developed by Bramoullé et al. (2009) to investigate spillover effects of employment growth among counties in the same commuting zone and the spillover effects of certain initial fiscal policy variables and initial county conditions on the employment growth of their neighboring counties in the same commuting zone for both 2000-2007 and 2000-2010. No spillover effects are found among employment growth of counties in the same commuting zone. County safety expenditures in year 1997 is found to have positive spillover effects on employment growth of neighboring counties in the same commuting zone. The county initial level of the high-tech employment share of year 2000 is found to have positive spillover effects on the employment growth of neighboring counties in the same commuting zone for the period 2000-2010. The county's own initial log employment density is found to have negative effect on its employment growth for both periods.

3.2. Identification strategy

Following Lee (2007), assume that N counties form R commuting zones (CZs) and counties in the same CZ interact with each other[22]. Let m_r denote group size, then m_r is the number of counties in the r^{th} CZ, and $\sum_{i=1}^{R} m_i = N$. For an individual county i in the r^{th} CZ, i interacts with the other ($m_r - 1$) counties in the CZ. Define M_{ri} as the peer group for county i, then M_{ri}

[21] Gibbons and Overman, 2012.
[22]A county does not have to interact with all other counties in the same CZ. For simplicity of illustrating the identification strategy, here it just assumes that counties interact with each other in the same CZ.

includes all the counties in the r^{th} CZ but i. Consider the following spatial autoregressive (SAR) model:

$$Y_r = l_{m_r}\alpha_r + W_r Y_r \beta + X_{r1}\gamma + W_r X_{r2}\delta + \varepsilon_r \quad (3.1)$$

where Y_r is a $m_r \times 1$ vector of outcomes for the m_r counties. W_r is an $m_r \times m_r$ interaction matrix with $W_{ij} = \frac{1}{m_r - 1}$ if $i \neq j$, and W_{ij}=0 if i=j. In application, W_r does not need to be in such a structure so that counties in the CZ end up having the same weight in interactions. Any W_r matrix that is a row-normalized matrix with zeroes as the elements on its diagonal is fine. X_{r1} and X_{r2} contain the characteristic variables of each county in the CZ. X_{r1} and X_{r2} can be the same or different[23]. l_{mr} is a $m_r \times 1$ vector of ones, α_r represents the unobservables of the r^{th} CZ. Because it is possible that those unobservables may correlate with X_{r1} or X_{r2}, α_r the so called " correlated effect" are treated as fixed effects. $\varepsilon_r \sim iid(0, \sigma_0^2 I)$ is the error term. β is the so called "endogenous effects", it is the structural interaction effect, it captures how the outcome of the agent i is influenced by the outcomes of other units in the same group. γ is the so-called "individual effect", and it captures how an individual county's outcome is affected by its own characteristics. δ is the so called "contextual effect", and it captures the effects of certain characteristics of other counties in the r^{th} CZ on the outcome of county i.

Define $X_r = [X_{r1}, X_{r2}]$ [24], and Equation (3.1) can be expressed equivalently in terms of individual county i in CZ r,

[23] In application, X_{r2} can be a set of variable that are included in X_{r1}

[24] If there are common elements in X_{r1} and X_{r2}, those common elements are include in X_r without duplication.

$$y_{ri} = \alpha_r + \frac{\sum_{j\in M_{ri}} y_{rj}}{m_r - 1}\beta + x_{ri,1}\gamma + \frac{\sum_{j\in M_{ri}} x_{rj,2}}{m_r - 1}\delta + \varepsilon_{ri}, \quad E(\varepsilon_{ri} \mid X_r, m_r, \alpha_r) = 0 \quad (3.2)$$

The exogeneity assumption $E(\varepsilon_{ri} \mid X_r, m_r, \alpha_r) = 0$ can accommodate situations where the numbers of counties in CZs are endogenous. The number of counties in a given CZ may depend on unobserved common characteristics of the CZ, which means $E(\varepsilon_{ri} \mid X_r, m_r) \neq 0$. Commuting zones were developed to be a spatial measure of the local labor market and the central objective of CZs was to develop such a geographic unit that better captures the economic and social diversity of non-metro areas. So when commuting zones are constructed, the size of the commuting zone—how many counties should be included in a given commuting zone—may depend on certain unobservable characteristics of the CZ. Model (2) allows for this type of correlation. Conditional on these common characteristics, the number of counties in the CZ is assumed to be independent of the idiosyncratic unobserved characteristics of individual county i: $E(\varepsilon_{ri} \mid X_r, m_r, \alpha_r) = 0$. This assumption is maintained throughout this study.

Similar to those of a panel data regression, Equation (3.2) can have a "*between"* transformation and a "*within"* transformation. Equation (3.2.1) and Equation (3.2.2) represent the *between* and *within* transformation of Equation (3.2), respectively.

$$(1-\beta)\bar{y}_r = \alpha_r + \bar{x}_{r1}\gamma + \bar{x}_{r2}\delta + \bar{\varepsilon}_r \qquad (3.2.1)$$

$$(y_{ri} - \bar{y}_{Mri})(1 + \frac{\beta}{m_r - 1}) = (x_{ri,1} - \bar{x}_{Mri1})\gamma - \frac{1}{m_r - 1}(x_{ri,2} - \bar{x}_{Mri2})\delta + (\varepsilon_{ri} - \bar{\varepsilon}_{Mri}) \qquad (3.2.2)$$

where $\bar{y}_{Mri} = \frac{1}{m_r - 1}\sum_{j\in M_{ri}} y_{rj}$, $\bar{x}_{Mri1} = \frac{1}{m_r - 1}\sum_{j\in M_{ri}} x_{rj,1}$, and $\bar{x}_{Mri2} = \frac{1}{m_r - 1}\sum_{j\in M_{ri}} X_{rj,2}$ are means for county i's peer group.

From Equation (3.2.1) we have:

$$\bar{y}_r = \frac{\alpha_r}{(1-\beta)} + \bar{x}_{r1}\frac{\gamma}{(1-\beta)} + \bar{x}_{r2}\frac{\delta}{(1-\beta)} + \frac{\bar{\varepsilon}_r}{(1-\beta)} \qquad (3.2.3)$$

From Equation (3.2.2) we have:

$$(y_{ri} - \bar{y}_{Mri}) = (x_{ri,1} - \bar{x}_{Mri1})\frac{\gamma(m_r - 1)}{(m_r - 1 + \beta)} - (x_{ri,2} - \bar{x}_{Mri2})\frac{\delta}{(m_r - 1 + \beta)} + (\varepsilon_{ri} - \bar{\varepsilon}_{Mri})\frac{(m_r - 1)}{(m_r - 1 + \beta)} \qquad (3.2.4)$$

Obviously, it is not possible to separate the CZ fixed effect α_r , individual effect γ and contextual effect δ from the endogenous effect β based on the regression Equation (3.2.3).

From Equation (3.2.4), when $\gamma\beta + \delta \neq 0$ [25], only two composite parameters: $\frac{\gamma(m_r - 1)}{(m_r - 1 + \beta)}$ and $\frac{\delta}{(m_r - 1 + \beta)}$ can be recovered from this form for each group size m_r. In order to identify the structural parameters β , γ and δ from the two composite parameters, we need at least three different group sizes. So variation in group size is the key to identify β , γ and δ . In this research, variation in group size means variation in the number of counties that are included in different CZs. When the outcome of a county is influenced by its own characteristics and the outcome and characteristics of other counties in the same CZ, different numbers of counties in different CZs means the number of the spillover channels are different in those CZs, and this leads to the variation in the overall spillover effects. For example, let's assume spillover effects exists and assume two counties: A and B. County A is in commuting zone CZ1 that contains 4 counties, and county B is in commuting zone CZ2 that contains 10 counties. As a result, the outcome of county

[25] When $\gamma\beta + \delta = 0$, if there are common elements in X_{r1} and X_{r2} , then only the individual effect of those common elements can be identified.

A is influenced by the outcomes and characteristics of the other three counties in CZ1, while the outcome of county B is influenced by the outcomes and characteristics of the other nine counties in CZ2. The structural parameters β and δ that denote the spillover effects are the same for county A and B. But B receives endogenous and contextual effects from nine counties while A does only from three counties. As a result, the overall endogenous and contextual effects that B in CZ2 receives are different from those received by A in CZ1. The variation in the overall spillover effects leads to the identification of β, γ and δ.

3.3. DATA

In this paper, my sample includes counties in the 48 US continental states. Commuting zones with only one county are not included. This results in 3,039 counties in 656 Commuting zones[26].

My empirical model is:

$$Y_{ri} = l_{mr}\alpha_s + W_r Y_r \beta + X_r \gamma + W_r X_{r1} \delta + \varepsilon_r \qquad (3.3.1)$$

where y_{ri} is the employment growth rate of county i in the r^{th} commuting zone in the given period. α_s is the state fixed effect. The state fixed effect is the same for counties in the same state, while for most commuting zones, counties in the commuting zone come from the same state, and even for those few commuting zones where not all counties come from the same state, still there are some counties from the same state, so in practice[27], after the within transformation,

[26] Independent cities in Virginia are combined with neighboring cites and/or counties to form a "new" county.

[27] As it will become more clear late, In practice, both side of (3.1) will time $(I - W_r)$ to get the within transformation, and the weight in the W matrix is build upon the proportion of work flows. For those few commuting zones that consist of counties from different states, the weight of county i in state s for a county j in state t in most of the time is relative small compare to the weight of county i for those counties also in state s, thus after the transformation $(I - W_r) * l_{mr}\alpha_s$ is a vector with most of its element equal to zero and those few doesn't equal to zero has a value that is close to zero.

the fixed effect is eliminated. x_{r1} includes 5 county fiscal variables and 2 variables to control for knowledge spillovers. x_{r} includes all the variables in x_{r1}, 9 demographic variable, 4 education variables, 6 amenity variables and 2 additional control variables. The employment data are from BEA, county fiscal data is from US Census of Government in 1997, demographic data and education data are from US 2000 census. Amenity data and Rural-Urban Continuum Codes are from USDA ERS. For a detailed description and summary statistics of the variables, please see Table 3.1.

The empirical model is applied to two periods: 2000-2007, and 2000-2010. I want to see if the Great Recession has changed the pattern of interactions between counties at the commuting zone level. The W_{r} matrix is constructed in the following way: for a county i in commuting zone r, suppose that there are m_i counties in i's peer group[28]. Let wf_{ij} denote the work flow from county i to j in year 2000. Define $wf_i = \sum_{j=1}^{m_i} wf_{ij}$, then $W_{ij} = \frac{wf_{ij}}{wf}$, if $i \neq j$ and j is in i's peer group in commuting zone r, and W_{ij}=0 otherwise.

As a result, W_{r} is a row-normalized matrix with zero as the element on its diagonal. Because the variation of the group size is the key to identification and Lee (2007) showed that smaller group size tends to help the identification. The characteristic of the group size in this research should satisfy these requirements. Table 3.2 shows the summary statistics of the group sizes.

The mean group size is 4.63 while the standard deviation is 2.26. So there is enough variation in the group size and the average group size is relatively small. The group structure of the sample in this research should be able to help identify the endogenous and contextual effects of interest.

[28] m_i does not have to equal the number of counties in the commuting zone minus one, which means county i does not have to be connected with all other counties in the same commuting zone.

3.4. Estimation Method

Multiply both sides of Equation (3.3.1) with (I-W), where I is a $N \times N$ identical matrix:

$$(I-W)y = (I-W)Wy\beta + (I-W)x\gamma + (I-W)Wx_1\delta + \varepsilon^* \qquad (3.4.1)$$

Where $\varepsilon^* = (I-W)\varepsilon$.

Lee (2007) proposed two ways to consistently estimate the structural parameters in Equation (3.4.1): conditional maximum likelihood estimation (CML) and generalized two stage least square (G2SLS). The two methods all give consistent estimates of the structural parameters; the generalized two stage least square method imposes less restrictions on the error term. But because of the way that the weight matrix W is constructed in this paper, the CMLE is not applicable for this research. G2SLS is used in this research to consistently estimate the parameters. The detailed explanation about the inapplicability of CMLE in this paper is in the Appendix.

Bramoullé et al. (2009) showed that, for the model in Equation (3.4.1), as long as $E[(I-W)Wy(\theta) \mid x, W]$ is not perfectly collinear with the regressor $((I-W)x, (I-W)Wx_1)$, then $(I-W)W^2x_1$ can be used as instruments for $(I-W)Wy$. If we multiply each side of Equation (3.4.1) by the weighting matrix W, and take the expectation with respect to x (x includes all the variables that are in x_1), we can have:

$$E[(I-W)Wy(\theta) \mid x, W] = W(I-\beta W)^{-1}[\gamma(I-W)x + \delta(I-W)Wx_1)] \qquad (3.4.2)^{29}$$

In Equation (3.4.2), the W matrix is a block-diagonal matrix with elements on the diagonal equal to zero and its non-zero elements equal to the corresponding relative size of the work flows

[29] Because of the way W matrix is formed in this research, It can be proved that $(I-W)W = W(I-W)$.

between two counties in the same commuting zone. It is clear that $E[(I-W)Wy(\theta)\mid x,W]$ are not perfectly collinear with $(I-W)x$ and $(I-W)Wx_1$. But the non-perfectly collinearity among $E[(I-W)Wy(\theta)\mid x,W]$, $(I-W)x$ and $(I-W)Wx_1$ does not guarantee the exogeneity of $(I-W)W^2x_1$. If $(I-W)W^2x_1$ is not exogenous, then it cannot be used as the excluded instrument for $(I-W)Wy$. Denote $x_1^* = (I-W)W^2x_1$ and $y^* = (I-W)y$. For any given observation in x_1^*, it is the linear combination of the observations of the corresponding characteristic variable of all the counties in that particular commuting zone, as a result, each observation in x_1^* contains the information of the corresponding characteristic variable of all the counties in that commuting zone. For any given observation in y^*, based on the same logic as that for x_1^*, it contains the information of the dependent variable of all the counties in that commuting zone. As a result, if we have enough reasons to believe that x_1 is exogenous to y, we can believe in the exogeneity of x_1^* to y^*. The dependent variable y in this research is the employment growth rate of a given county for the period 2000-2007 or 2000-2010. The county characteristic variables that are included in x_1 are 5 county fiscal variables of year 1997 which are pre-determined, and natural log of county employment density of year 2000 and the employment share of high-tech industry at the county level in year 2000 which are the variables to control for the initial conditions. There may exist commuting zone level characteristic variable in the error term in Equation (3.3.1) that correlate with both the employment growth of counties in the commuting zone and the above 7 county level characteristic variables. Equation (3.4.2) is actually the deviation-from-peer-group version of Equation (3.3.1). Because the W matrix is a row-normalized matrix, any omitted variable in the error of Equation (3.3.1) that are at the commuting zone level no longer exists in the error term of Equation (3.4.2). These features of the variables that are included in x_1 and the error term of Equation (3.4.2) gives confidence about the exogeneity of x_1, and as a result, confidence in the exogeneity of the excluded instrument: $(I-W)W^2x_1$.

Based on Lee (2007), Bramoullé et al. (2009) proposed a generalized Two-Stage Least Square (G2SLS) procedure consists of two rounds of Two-Stage Least Square (2SLS) estimation to consistently estimate the parameters in Equation (3.4.2). In this research, the procedure proposed by Bramoullé et al. (2009) is used to identify the parameters of the empirical model. The following are the two rounds of this procedure:

1st round: In the first round 2SLS, $(I-W)W^2x_1$ is used as the excluded instrument for $(I-W)Wy$, and we get estimates of the parameters of the model. Denote $\theta=(\beta,\gamma,\delta)$, and let's call the estimate of θ that is obtained from this first round $\hat{\theta}^{2sls}$.

2nd round: The excluded instruments used in the second round 2SLS is based on $\hat{\theta}^{2sls}$. The excluded instrument for $(I-W)Wy$ in the second round 2SLS is $\hat{E}[(I-W)Wy(\theta)\,|\,x,W]$, the estimate of $E[(I-W)Wy(\theta)\,|\,x,W]$ based on $\hat{\theta}^{2sls}$. $\hat{E}[(I-W)Wy(\theta)\,|\,x,W]$ is obtained following Equation (3.4.2): $\hat{E}[(I-W)Wy(\theta)\,|\,x,W]=W(I-\hat{\beta}W)^{-1}[(I-W)(\hat{\gamma}x+\hat{\delta}Wx_1)]$. Denote the estimate of θ obtained from this round $\hat{\theta}^{Lee}$, and Lee (2007) showed that $\hat{\theta}^{Lee}$ is a consistent estimate of θ.

The exogeneity of the excluded instrument, $(I-W)W^2x_1$, in the above 1st round has already been discussed. How about the exogeneity of the excluded instrument, $\hat{E}[(I-W)Wy(\theta)\,|\,x,W]$, in 2nd round? Because $\hat{E}[(I-W)Wy(\theta)\,|\,x,W]=W(I-\hat{\beta}W)^{-1}[(I-W)(\hat{\gamma}x+\hat{\delta}Wx_1)]$, based on the same logic as that for $(I-W)W^2x_1$, There should be enough reasons to believe that those variable in x and x_1 are exogenous in order for me to be confident about the exogenous of $\hat{E}[(I-W)Wy(\theta)\,|\,x,W]$. Exogeneity of variables in x_1 had already been discussed. x includes all 7 variable of x_1, 9 demographic variables and 4 education variables controlling for initial conditions and 6 amenity variables and 2 additional control variables[30]. The characteristics of these variables give confidence about the exogeneity of $\hat{E}[(I-W)Wy(\theta)\,|\,x,W]$. From the above two steps, it is clear that whether $\hat{\theta}^{Lee}$ is a consistent estimate of θ relies heavily on how well $\hat{E}[(I-W)Wy(\theta)\,|\,x,W]$ works as the excluded instrument for $(I-W)Wy$ in the second round 2SLS. Theoretically, $\hat{E}[(I-W)Wy(\theta)\,|\,x,W]$ should be a good choice as the excluded instrument for $(I-W)Wy$ but it has to be supported by test results.

For the purpose of comparing, I also estimate the following two models:

1. Hedonic OLS model without spatial lags:

$$y_i = c_s + x_i\varphi + \varepsilon_i\text{, where } c_s \text{ is state fixed effect.}$$

2. Traditional Spatial-Durbin instrument variable (IV) model with spatial lags:

$y = c_s + Wy\beta + x\gamma + Wx_1\delta + \zeta_i$, where c_s is state fixed effect, and use W^2x_1 as instrument for Wy.

Because the Spatial-Durbin IV model and the G2SLS model are all based on instrumental variables, weak instruments in the first regression will lead to biased estimate of the model parameters. How well the instruments perform in the first stage regressions should be tested. Table 3.3 contains important test statistics regarding how the excluded instruments perform in all the models. For convenience, the 1st round 2SLS in the G2SLS procedure will be referred to as

[30] Please see section .33 of this paper for a detailed description of all the variables.

the 1st round 2SLS, and the 2nd round 2SLS in the G2SLS procedure will be referred to as the 2nd round 2SLS throughout this paper. The empirical results for key variables of the G2SLS models' First-stage regressions are shown in Table 3.7 to Table 3.10. Table 3.7 and Table 3.9 contain the empirical results for the 1st round 2SLS for 2000-2007 and 2000-2010, respectively; Table 3.8 and Table 3.10 contains the empirical results for the 2nd round 2SLS for 2000-2007 and 2000-2010, respectively[31]. The "first stage F statistics" in Table 3.3 are the F statistic for the excluded instruments that are used in the corresponding regressions.

From Panel A of Table 3.3, we can see that the p-value for the Hansen-J overidentification test statistics are larger than 0.1 for the first stage regression of the 1st round 2SLS for both 2000-2007 and 2000-2010. This means I do not need to worry much about the overidentification issue of the excluded instruments. The F statistics for the first stage of the 1st round 2SLS are less than 10 for both the periods[32], while the p-value for the F-statistics are both zero. The corresponding Kleibergen-Paap Wald rk F-statistics are both less than the Stock-Yogo weak ID test critical values based on 10% maximal IV relative bias while larger than the Stock-Yogo weak ID test critical values based on 20% maximal IV relative bias[33]. While those above empirical test statistics may suggest that the excluded instruments used in the 1st round 2SLS are not that strong, they still suggest that we can have confident in the overall significance of these excludes instruments used in the corresponding regressions. Whether the parameters in the model can be consistently estimated by the G2SLS procedure based heavily on how well the excluded instruments in the 2nd round 2SLS performs. Weak instruments can cause bias in the estimates of

[31] Table 3.7 to Table 3.10 contains empirical results for the 7 key variables in the first stage regressions. The full set of empirical results for the corresponding first stage regressions is available upon request.

[32] In empirical research, with a first-stage F statistics based on the excluded instruments that is equal to or greater than 10, weak identification hypothesis is considered rejected with strong confidence.

[33] The Stock-Yogo weak ID test critical values based on 5% maximal IV relative bias, 10% maximal IV relative bias and 20% maximal IV relative bias are 19.86, 11.29 and 6.73, respectively. The Stock-Yogo weak ID test critical values are calculated based on i.i.d errors. The errors in this research are clustered in commuting zone level. So the Stock-Yogo weak ID test critical values are only for reference here.

the parameters in the 1st round 2SLS, but this does not necessary mean the excluded instruments for $(I-W)Wy$ used in the 2nd round 2SLS, $\hat{E}[(I-W)Wy(\theta)\,|\,x,W]$, that based on the estimates of the parameters from the 1st round 2SLS will be a bad choice. The F statistics for the first stage of the 2nd round 2SLS for 2000-2007 and 2000-2010 are 44.41 and 51.02, respectively. The corresponding Kleibergen-Paap Wald rk F statistics is 146.77 for 2000-2007 and 130.79 for 2000-2010, and both are significantly larger than the Stock-Yogo weak ID test critical values based on 5% maximal IV relative bias. These empirical test statistics suggest that, in the 2nd round 2SLS, the weak instrument hypothesis is rejected with significant confidence.

The fact that the weak identification test statistics for the excluded instrument of the 2nd round 2SLS is significantly larger than that for the excluded instrument of the corresponding 1st round 2SLS may due to the reason that not all the variables in $(I-W)W^2x_1$ are that strong in predicting $(I-W)Wy$; while $\hat{E}[(I-W)Wy(\theta)\,|\,x,W]$ obtained following Equation (3.4.2) contains information from all the predictor variables in $(I-W)x$ and $(I-W)Wx_1$ which performs well in predicting $(I-W)Wy$. This is exactly the case implied by the empirical results shown in Table 3.7 and Table 3.9. In Panel C of Table 3.7, only $(I-W)W^2$ multiply log employment density of year 2000 is found to be significant for 2000-2007. In Panel C of Table 3.9, only $(I-W)W^2$ multiply log employment density of year 2000 and high-tech employment share of year 2000 are found to be significant for 2000-2010. $\hat{E}[(I-W)Wy(\theta)\,|\,x,W]$ contains information from all the variables that are found to be significant in Panel A and Panel B for the corresponding period. Panel A of both Table 3.7 and Table 3.9 shows that $(I-W)$ multiply property tax, safety expenditure, education expenditure, and log employment density of year 2000 are found to be significant for both 2000-2007 and 2000-2010[34]. Results in Panel B of Table 3.7 show that $(I-W)W$ multiply education expenditure and log employment density of year 2000 are found to be significant for 2000-2007, while results in Panel B of Table 3.9 show that $(I-W)W$ multiply property tax and log employment density of year 2000 are found to be significant for 2000-2010.

Panel B of Table 3.3 contains the test statistics for the first stage regressions of the Spatial-Durbin model. P value for the Hansen-J overidentification test statistics are larger than 0.1 for the first stage regression for both 2000-2007 and 2000-2010. For both periods, the F-statistics for the first stage regression are less than 10, but the p-value for the F-statistics are equal to zero. The Kleibergen-Paap Wald rk F statistic are both larger than the Stock-Yogo weak ID test critical values based on 10% maximal IV relative bias. According to the empirical results for the first stage regression of the Spatial-Durbin IV model[35], W^2 multiply safe expenditure and employment density of year 2000 are found to be significant for both 2000-2007 and 2000-2010. These results suggest that we can have confident in the overall significant of these excludes instruments used in the Spatial-Durbin IV model.

3.5. Empirical Results

The empirical results for period 2000-2007 are shown in Table 3.4. The endogenous effect is found to be 0.013 and not significant by the G2SLS model and is found to be -0.16 and not significant by the Spatial-Durbin IV model. The negative endogenous effect estimated by the Spatial-Durbin IV model is not surprising, it has been shown earlier in this paper that the spatial-Durbin IV model gives biased estimation of the endogenous effect.

---

[34] Except for those variables that are contained in the Panel A of the table 3.7 and table 3.9, some other variables that included in $(I-W)Wx$ are also found to be significant, and the information of these variables are also included in $\hat{E}[(I-W)Wy(\theta) \mid x, W]$.

[35] The empirical results for the first stage regression of the Spatial-Durbin IV model are not included in this paper, but available upon request.

For the period 2000-0007, while the empirical results based on Spatial-Durbin IV model suggests there may exist a positive contextual effect of 0.19 for the property tax of 1997, G2SLS model results show that there exists no such effect. The results for the Spatial-Durbin model suggest a negative and significant contextual effect that is equal to -0.006 for the log employment density of year 2000 variable, while the 2GSLS model results suggest such an is effect equal to -0.003 but not significant. The 2GSLS model results imply that the contextual effect of safety expenditure of year 1997 is 1.19 and significant during 2000-2007; the more safety a county's neighboring counties are at the initial of the period, the more employment growth that county can experience. This implies higher safety expenditure in the commuting zone benefits every county inside it.

For the period 2000-2007, the results for the OLS model suggest significant individual effects of safety expenditures of year 1997, education expenditures of year 1997 and log employment density of year 2000 on county employment growth, and the estimated effects for these variables are -1.92, 0.28, and -0.07, respectively. Based on results for the Spatial-Durbin IV model, the individual effects of safety expenditure of year 1997, education expenditure of year 1997, log employment density of year 2000 and high-tech employment share of year 2000 are found to be significant with the estimated effects as -2.24, 0.22, -0.04, and -0.09, respectively. But, with an estimated parameter of -0.02, only the individual effect of log employment density of year 2000 is found to be significant by the G2SLS model. The three different models give very different results regarding the individual effects of the key variables of interest and their significance. As what have been discussed earlier in this paper, the Spatial-Durbin IV model cannot provide unbiased and consistent estimates of the parameters. When there exists a contextual effect from safety expenditure of year 1997 on county employment growth, it is likely that it will be correlated with log employment density of year 2000. The OLS model does not control for this and the estimates of the individual effects based on the OLS are biased and not consistent. The negative individual effect of log employment density of year 2000 implies that, counties with higher initial level of employment density experience lower employment growth rate during 2000-2007.

The G2SLS model results for 2000-2007 suggest that spillover effects do not exist among the employment growth of counties in the same commuting zone; county safety expenditure of year 1997 has positive spillover effects on the employment growth of other counties in the same commuting zone, and county's log of employment density of year 2000 has negative effect on its own employment growth for the period 2000-2010.

Table 3.5 contains the empirical results for 2000-2010. The estimated endogenous effect based on the G2SLS model is now 0.046 which is larger than that for 2000-2007, but it is still found to be not significant. The Spatial-Durbin IV model results show an endogenous effect equal to -0.25 and it is significant at the 5% level. A negative spillover effect of the employment growth among counties in the same commuting zone for 2000-2010 suggest that, during this period, counties will have more jobs created while their neighboring counties are losing jobs, and vice versa. This is not generally what is observed in the real world.

The Spatial-Durbin IV model results imply the existence of contextual effects for sale taxes of year 1997 and log employment density of year 2000, and the estimated parameters for the two variables are 1.89 and 0.02, respectively. G2SLS model results implies that there exist no contextual effects for the two variables, but contextual effects are found for safety expenditure of year 1997, and high-tech employment share of year 2000, and the estimated effects are 3.68 and 0.21, respectively. Same as the result for 2000-2007, county safety expenditure of year 1997 is found to have a positive spillover effect on the employment growth of neighboring counties in the same commuting zone. Unlike that for 2000-2007, the county initial level of the high-tech employment share of year 2000 is now found to have positive spillover effect on the employment growth of neighboring counties in the same commuting zone.

For the period 2000-2010, results for the OLS model suggests significant individual effects of the education expenditure of year 1997 and log employment density of year 2000 on county employment growth, and the estimated effects for these variables are 0.6 and -0.04, respectively. Spatial-Durbin IV model results imply that the individual effects of property tax of year 1997, safety expenditure of year 1997, education expenditure of year 1997 and log employment density of year 2000 are significant with the estimated effects as 0.46, -0.81, 1.35, and -0.03, respectively. Only the individual effect of log employment density of year 2000 is found to be significant by the G2SLS model with an estimated parameter as -0.05. Similar to what's shown by the individual effects results for 2000-2007, the estimated individual effects of the key variables and their significance are very different based on the three models. But the Spatial-Durbin IV model results are biased and inconsistent, and when there exists contextual effects of key variables on county employment growth, the OLS results are biased and not consistent. Based on the G2SLS result, like that for 2000-2007, county's log employment density of year 2000 is found to have a negative effect on its own employment growth rate for the period 2000-2010.

The G2SLS model results for 2000-2007 and 2000-2010 suggest that there exist no spillover effects among the employment growth of counties in the same commuting zone. A county's safety expenditure of year 1997 has positive spillover effects on other counties' employment growth for both periods, while the estimated effect for 2000-2010 is larger than that for 2000-2007. A county's own initial log employment density is found to have negative effect on its employment growth for both periods.

Comparing the empirical results of the Spatial-Durbin IV model and the G2SLS model for both periods, it is clear that, in this research, the Spatial-Durbin IV model produces downwardly-biased estimates of endogenous effects and biased estimates of contextual effects and individual effects of key variables and their significance. The Spatial-Durbin IV model cannot provide credible information about the causal relationship in the economy process in this research. The consistent estimates of the parameters based on the G2SLS model are highly likely to be more reliable.

6. Monte Carlo Simulation

Lee (2007) showed that the identification of the endogenous effect and contextual effects relies on the structure of the data: the number of groups, the average group size, and the variation of the group sizes. As I have shown before, the characteristics of group structure in this research match the requirements well. But how the group structure in this research performs in estimating the model parameters in this research still needs to be investigated. Monte Carlo Simulation is applied to show how the structure of the data works in estimating the model parameters.

In the Monte Carlo Simulation, the actual weight matrix W from the data is used, and the sample size is fixed at 3039. This is to keep the interaction structure of the real data. In order to reduce computing time and without losing generality, except for the endogenous effect, I only simulate the contextual effects and individual effects of property tax, highway expenditure, safety expenditure, education expenditure, sale tax , log employment density, and High-tech employment share. I assume these variables follow a normal distribution and calibrate the moments of them according to the sample of this research: property tax with 0.0314 as mean and 0.0296 as variance , highway expenditure with 0.0085 as mean and 0.01145 as variance, safety expenditure with 0.00624 as mean and 0.0039 as variance, education expenditure with 0.0574 as mean and 0.0226 as variance, sale tax with 0.00368 as mean and 0.00515 as variance, Employment density with 60.4 as mean and 867.876[36] as variance and then log employment density is based on the natural log of it, and High-tech employment share with 0.0571 as mean and 0.067 as variance. According to the moments of the residuals that are obtained after the G2SLS for 2000-2010, the error term is generated by a normal distribution with mean equal -$7.2*10^{-11}$ and variance equal 0.1104.

The endogenous effect, contextual effects and individual effects are set equal to those estimated by the G2SLS for 2000-2010. The endogenous effect is set to be 0.046. The contextual effects of property tax, highway expenditure, safety expenditure, education expenditure, sale tax, log employment density, and High-tech employment share are set to be -0.57, -0.38, 5.68, -0.53, 0.82, -0.01 and 0.21, respectively; while the individual effects of these variables are set to be -0.06, -0.29, 0.99, 0.28, -0.27, -0.05 and 0.03, respectively. The dependent variable y is generated from the reduced-form equation in deviation form. Then the model parameters are estimated by both Spatial-Durbin with IV and G2SLS. The Monte Carlo Simulation is carried out for 1000 times, then the mean of the estimated parameters and their standard deviation from the 1000 replicates are calculated.

Table 3.6 contains the results based on the Monte Carlo Simulation. The mean value of the estimated endogenous effect from the G2SLS is 0.051, with a standard deviation of 0.09, while it's 0.009 with a standard deviation of 0.12 from the Spatial-Durbin with IV. The 0.05 mean value from the G2SLS is close to the 0.046 real value, while 0.009 is significantly smaller than 0.046. This is not a surprising result, because the G2SLS estimate of the endogenous effect is consistent while the Spatial-Durbin IV estimate of the endogenous effect is biased. For the contextual effects, the mean values of estimated parameters from G2SLS are closer to the actual values than those based on the Spatial-Durbin with IV. The dispersion between the mean value of

[36] Before it is taken natural log of, the employment density of year 2000 variable in the sample has an average of 60.4 and a variance of 867.876.

the contextual effects for highway expenditure and safety expenditure based on the simulation of Spatial-Durbin IV model and their true value are significantly larger than that for the contextual effects of other variables. The means of the individual effects of the 7 variables based on the two models are both close to their true value, and their corresponding variance based on the two models are very close. The simulation results shows that, based on group interaction structure of the sample in this research, the Spatial-Durbin IV model generally cannot provide consistent estimate of the endogenous and contextual effects of certain key variables, while the G2SLS model is highly likely to provide credible and consistent estimates of the endogenous effect and contextual effects of key variables.

7. Conclusion

To the best of my knowledge, this study is the first empirical application of a new G2SLS approach proposed by Lee (2007) and developed by Bramoullé et al. (2009) to investigate spillover effects of employment growth among counties in the same commuting zone and the spillover effects of certain initial fiscal policy variables and initial conditions of counties on the employment growth of their neighboring counties in the same commuting zone. No evidence of endogenous effects are found. A positive contextual effect is found for safety expenditure of year 1997 in both 2000-2007 and 2000-2010 and for the high-tech employment share of year 2000 for the period 2000-2010. Initial log employment density of a county is found to lower its own employment growth for both 2000-2007 and 2000-2010. Counties in the same commuting zone are thought to be in the same local labor market; it is reasonable to assume that there should exist spillover effects of county employment growth. But the findings of this paper suggest that, it is likely that there exist no employment spillover effects among counties in the same commuting zone. It is possible that there actually exist no such spillover effects among counties in the same commuting zone. Another possibility is that commuting zone may be not an appropriate way to define local labor market, so the interaction among the employment growth of counties in the same commuting zone is not strong enough.

In terms of public policy, the significant contextual effect of initial safety expenditure suggest that, if one county increase safety expenditure, other counties that interact with it in the same commuting zone will have higher employment growth rate. The positive contextual effect of initial high-tech employment share implies that, county employment growth is likely to benefit from the knowledge spillover that it might receive from its neighboring counties in the same commuting zone.

Theory and the Monte Carlo simulation in this paper suggest that the Spatial-Durbin IV model may not be a good choice for certain empirical spatial econometric applications. Empirical research based on the conventional Spatial-Durbin IV model may simply find the correlations between different variables instead of identifying the causality relationships in the economic processes. Instead of taking the spillover effects among spatial units for granted, empirical spatial research should focus on getting credible identification on the model parameters. The novel G2SLS procedure used in this study is based on the weighted-spatial-difference among spatial units in the same group and makes use of the variation in group size to help identify the model parameters. It does not suffer from certain serious identification issues that conventional Spatial-Durbin IV models have. Empirical spatial econometric research that is serious about the identification of causality in the economic processes should consider this G2SLS procedure over the conventional Spatial-Durbin IV model.

REFERENCES

[1] Albouy, David. 2008. “Are Big Cities Really Bad Places to Live? Improving Quality-of-Life Estimates across Cities.” NBER Working Paper No. 14472.

[2] Autor, H. David, Katz, F. Lawrence, and Kearney S. Melissa. 2008. “ Trends in U.S. Wage Inequality: Revising the Revisionists.” The Review of Economics and Statistics 90 (2), 300-323.

[3] Bai, J. and Serena, Ng. 2002. “Determining the Number of Factors in Approximate Factor Models” Econometrica 70(1), 191-221.

[4] Bai, J. and Serena, Ng. 2004. “A PANIC Attack on Unit Roots and Cointegration.” Econ-metrica 72, 1127-1177.

[5] Barro, R. and Sala-i-Martin, X. 1991. “Convergence across States and Regions.” Brookin-Papers on Economic Activity 1, 107-182.

[6] Barro, R. and Sala-i-Martin, X. 1992. “Convergence.” Journal of Politic Economics 100, 223–251.

[7] Baumol, W.J. 1986. “Productivity Growth, Convergence, and Welfare: What the Long-run Data Show.” American Economic Review 76, 1072-1085.

[8] Beeson, Patricia E. and Randall W. Eberts. 1989. “Identifying Productivity and Amenity Effects in Interurban Wage Differentials.” The Review of Economics and Statistics 71 (3), 443- 452.

[9] Bernard, AB. and Durlauf, S. 1995. “Convergence in International Output.” Journ al of Applied Econometrics 10, 97–108.

[10] Bernard, AB. and Durlauf, S. 1996. “Interpreting Tests of the Convergence Hypothesis.” Journal of Econometrics 71, 161–173.

[11] Boozer, M. A. and Cacciola, S. E. 2001. “Inside the “Black Box” of Project STAR: Estim-ation of Peer Effects Using Experimental Data”, WP no. 832, Economic Growth Center, Yale University.

[12] Bramoullé, Y., Djebbari, H. and Fortin, B. 2009. “Identification of Peer Effects Through Social networks.” Journal of Econometrics 150, 41–55.

[13] Cheng, K.M., Durmaz, N., Kim, H. and Stern, M.L. 2012. “Hysteresis vs. Natural Rate of US Unemployment.” Economic Modelling 29, 428-434.

[14] Clark, Todd E. 1995. “Rents and Prices of Housing across Areas of the United States: A Cross-Section Examination of the Present Value Model.” Regional Science and Urban Economics 25 (2), 237-247.

[15] Davezies, L., Haultfoeuille, D’ X., and Fougère, D. 2009. “Identification of Peer Effects Using Group Size Variation.” The Econometrics Journal 12, 397–413.

[16] Desmet, K. and Fafchamps, M. 2005. “Changes in the Spatial Concentration of Employment Across U.S. Counties: a Sectoral Analysis 1972-2000.” Journal of Economic Geography 5, 261-284.

[17] Dickey, DA. and Fuller, WA. 1981. “Likelihood Ratio Statistics for Autoregressive Time Series With a Unit Root.” Econometrica 49, 1057–1072.

[18] Dickie, Mark and Shelby Gerking. 1987. “Interregional Wage Differentials: A Equilibrium Perspective.” Journal of Regional Science 27 (4), 571-585.

[19] Dobnik, F. 2013. “ Long-run Money Demand in OECD Countries: What Role Do Common Factors Play?” Empirical Economics 45, 89-113.

[20] Dowrick, S. and Nguyen, D.T. 1989. “OECD Comparative Economic Growth 1950-85: Catch-up and Convergence.” American Economic Review 79, 1010-1030.

[21] Dreger, C. and Reimers, H.E. 2012. “The Long Run Relationship Between Private Consumption and Wealth: Common and Idiosyncratic Effects.” Portuguese Economic Journal 11 (1), 21-34.

[22] Dreger, C. and Zhang, Y.Q. 2014. “On the relevance of Exports for Regional Output Growth in China.” Applied Economics 46(35), 4302-4308

[23] DuMond, J. Michael, Barry T. Hirsch, and David A. Macpherson. 1999. “Wage Differentials Across Labor Markets and Workers: Does Cost of Living Matter?” Economic Inquiry 37 (4), 577-598.

[24] Evans, P. and Karras, G. 1996. “Convergence Revisited.” Journal of Monetary Economics 37, 249–265.

[25] Everaert, G. 2014. “ A Panel Analysis of the Fisher Effect with an Unobserved I(1) World Real Interest Rate.” Economic Modelling 41, 198-210.

[26] Fujita, M. and Ogawa, H. 1982. “Multiple Equilibria and Structural Transition of Nonmonocentric Urban Configurations.” Regional Science and Urban Economics 12, 161-196.

[27] Gabriel, Stuart A., Joe P. Mattey, and William L. Wascher. 2003. “Compensating Differentials and Evolution in the Quality of Life among U.S. States.” Regional Science and Urban Economics 33 (5), 619-649

[28] Gabriel, Stuart A. and Stuart S. Rosenthal. 2004. "Quality of the Business Environment Versus Quality of Life: Do Firms and Households Like the Same Cities?" Review of Economics and Statistics 86 (1), 438-444.

[29] Gibbons, S. and Overman, HG. 2012. "Most Pointless Spatial Econometrics?" Journal of regional science 52(2), 172-191.

[30] Graham, BS. 2008. "Identifying Social Interactions Through Conditional Variance restrictions." Econometrica 76(3), 643–660.

[31] Harris, C. 1954. "The Market as a Factor in the Localization of Industry in the United States." Annals of the Association of American Geographers 64, 315-348.

[32] Im, K.S., Lee, J. and Tieslau, M. 2005. "Panel LM Unit Root Test with Level Shifts." Oxford Bulletin of Economics and Statistics 67, 393–419.

[33] Jian, T., Sachs, JD. and Warner, AM. 1996. "Trends in Regional Inequality in China." China Economic Review 7, 1–21.

[34] Katz, F. Lawrence and Autor, H. David. 1999. "Changes in the Wage Structure and Earnings Inequality." in O. Ashenfelter and D. Card (Eds.), Handbook of Labor Economics vol. 3, 1463-1555.

[35] Kelejian, H.H. 1998. "A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances." Journal of Real Estate Finance and Economics Vol 17(1), 99-121.

[36] Khan, R., P.F. Orazem, and D.M. Otto. 2001. "Deriving Empirical Definitions of Spatial Labor Markets: The Roles of Competing versus Complementary Growth." Journal of Regional Science 41(4), 735-756.

[37] Koo, Jahyeong, Keith R. Phillips, and Fiona D. Sigalla. 2000. "Measuring Regional Cost of Living." Journal of Business & Economic Statistics 18 (1), 127-136.

[38] Lee, LF. 2004. "Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Econometric Models." Econometrica 72, 1899-1926.

[39] Lee, LF. 2007. "Identification and Estimation of Econometric Models with Group Interactions, contextual factors and fixed effects." Journal of Econometrics 140 (2), 333–374.

[40] Lemieux, Thomas. 2008. "The Changing Nature of Wage Inequality." Journal of Population Economics 21, 21-48.

[41] Liu, PC., Lin, Ch. and Ho, IL. 2013. "Regional Convergence or Divergence in China? Evidence from unit root tests with breaks." Annals of Regional Science 50, 223-243.

[42] Malamud, Ofer and Wozniak, Abigail. 2012. "The Impact of College on Migration: Evidence from the Vietnam Generation" Journal of Human Resources 47(4), 913-950.

[43] Mankiw, N.G., Romer, D. and Weil, D.N. 1992. "A Contribution to the Empirics of Economic Growth" Quarterly Journal of Economics 107, 407-438.

[44] Manski, C. 1993. "Identification of Endogenous Social Effects : the Reflection Problem." Review of Economic Studies 60(3), 531–542.

[45] Moffitt, R. 2001. "Policy interventions, low-level equilibria, and social inter-interactions." In: Durlauf. S., Young, P. (Eds.), Social Dynamics. Brookings Institution Press and MIT Press, Cambridge, MA, 45–92.

[46] Moretti, Enrico. 2013. "Real Wage Inequality." American Economic Journal: Applied Economics,5(1), 65-103.

[47] Nagayasu, J. 2011. " The Common Component in Forward Premiums: Evidence from the Asia-Pacific Region." Review of International Economics 19(4), 750-762.

[48] NG, S. and Perron, P. 2001. "Lag Length Selection and the Construction of Unit Root tests with Good Size and Power." Econometrica 69, 1519–1554.

[49] Pantula, S. G. 1991. "Asymptotic Distributions of Unit Root Tests when the Process is Nearly Stationary." Journal of Business and Economic Statistics 9, 63–71.

[50] Pedroni, P. and Yao, JY. 2006. "Regional Income Divergence in China." Journal of Asian Econonics 17, 294–315.

[51] Perron, P. 1989. "The Great Crash, the Oil Price Shock and the Unit Root Hypothesis." Econometrica 57, 1361–1401.

[52] Pesaran, M. H. 2007. "A Simple Panel Unit Root Test in the Presence of Cross-Section Dependence." Journal of Applied Econometrics 22, 265–312.

[53] Pinkse, J. and Slade, M. 2010. "The future of spatial econometrics." Journal of regional Science. 50(1), 103-117.

[54] Roback, Jennifer. 1982. "Wages, Rents, and the Quality of Life." Journal of Political Economy 90 (6), 1257-1278.

[55] Roback, Jennifer. 1988. "Wages, Rents, and Amenities: Differences among Workers and Regions." Economic Inquiry 26 (1), 23-41.

[56] Rosen, Sherwin. 1979. "Wage-Based Indexes of Urban Quality of Life." In: Mieskowski, P., Straszheim, M. (Eds), Current Issues in Urban Economics. Johns Hopkins University Press, Baltimore, pp. 74-104.

[57] Sahling, Leonard G. and Sharon P. Smith. 1983. "Regional Wage Differentials: Has the South Risen Again?" The Review of Economics and Statistics 65 (1), 131-135.

[58] Schwert, G.W. 1989. "Tests for Unit Roots: A Monte Carlo Investigation." Journal of Business and Economic Statistics 7, 147–160.

[59] Shin, D.W. and So, B.S. 2001. “Recursive Mean Adjustment and Test for Unit Roots.” Journal of Time series Analysis 22,595–612.

[60] Shuai, X. B. 2010. “Who Benefits from job creation at county level? An Analysis of Leakage and Spillover of New Employment Opportunities in Virginia.” Business Economics 45, 38-48.

[61] So, B.S. and Shin, D.W. 1999. “Recursive Mean Adjustment in Time-Series Inferences.” Statistics & Probability Letter 43,65–73.

[62] Weeks, M. and Yao, JY. 2003. “Provincial Conditional Income Convergence in China, 1953-1997: a Panel Data Approach.” Economic Review 22, 59–77.

[63] Westerlund, J. 2015. “ The power of PANIC.” Journal of Econometrics 185, 495-509.

[64] Winters John V. 2009. “Wages and prices: Are workers fully compensated for cost of living differences?” Regional Science and Urban Economics 39, 632-643.

[65] Winters John dissertation, Essay I: “Wages and prices: Are Workers fully compensated for cost of living differences”

[66] Wolff, E.N. 199l. “Capital Formation and Productivity Convergence over the Long Tem.” American Economic Review 81, 565-579.

[67] Wozniak, Abigail. 2010. “Are College Graduates More Responsive to Distant Labor Market Opportunities?” Journal of Human Resources 45(4), 944-970.

[68] Zhang, Z., Liu, A. and Yao, S. 2001. “Convergence of China’s regional incomes 1952-1997.”China Economic Review 12, 243–258.

APPENDICES

A.1 Why CMLE is not applicable

Multiply both sides of Equation (3.3.1) with (I-W), where I is a $N \times N$ identical matrix:

$$(I-W)y = (I-W)Wy\beta + (I-W)x\gamma + (I-W)Wx_1\delta + \varepsilon^* \qquad \text{(A.1.1)}$$

Where $\varepsilon^* = (I-W)\varepsilon$. Since β is a parameter, we have:

$$(I-W)y = (I-\beta W)^{-1}[(I-W)(x\gamma + Wx_1\delta)] + v \qquad \text{(A.1.2)}$$

Where $v = (I-\beta W)^{-1}\varepsilon^*$. Because $\varepsilon^* = (I-W)\varepsilon$, we have $v = (I-\beta W)^{-1}(I-W)\varepsilon$. Denote v_{ri} as the error term for county i in commuting zone r in Equation (A.1.2). Now, for different v_{ri} (i=1, 2.. m_r) that belongs to different counties in the same commuting zone r, v_{ri} s are not i.i.d. any more. There is a joint distribution $f_r(v_r)$ for all the v_{ri} (i=1, 2.. m_r) in the same commuting zone r. For any v_{ri} and v_{ji}, as long as r ≠ j, v_{ri} and v_{ji} are still independent. As a results, the joint distributions $f_r(v_r)$ (r=1,2,..R) are independent among different commuting zones.

The variance-covariance matrix Σ_r for v_{ri} (i=1, 2.. m_r) in each commuting zone r is $v_r * v_r^{'}$, so $\Sigma_r = \sigma^2 (I_r - \beta W_r)^{-1}(I_r - W_r)((I_r - \beta W_r)^{-1}(I_r - W_r))'$. Where I_r is an $m_r \times m_r$

identity matrix, W_r is the weight matrix for commuting zone r. The v_{ri} (i=1, 2.. m_r) in each commuting zone r are jointly Gaussian, and $f_r(v_r)$ is the multivariate Gaussian distribution for them:

$$f_r(v_r) = \frac{1}{\sqrt{(2\pi)^r}\left|\Sigma_r\right|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(y_r^* - x_r^*)'\Sigma_r^{-1}(y_r^* - x_r^*)\right\} \qquad \text{(A.1.3)}$$

Where $\left|\Sigma_r\right|$ is the determinant of Σ_r , $y_r^* = (I_r - W_r)y_r$, and $y_r = (y_{r1,}y_{r2,}...y_{rm_r})'$.

$x_r^* = (I_r - \beta W_r)^{-1}[(I_r - W_r)(x_r\gamma + W_r x_{r,1}\delta)]$,where

$x_r = (\mathrm{x}_{r1,}\ x_{r2,}...\mathrm{x}_{rm_r})'$, $x_{r,1} = (\mathrm{x}_{r1,1,}\ x_{r2,1,}...\mathrm{x}_{rm_{r,1}})'$. Because, as long as r$\neq$ j, $f_r(v_r)$ and $f_j(v_j)$ are independent, the likelihood function for Equation (A.1.3) can be expressed as:

$$L_{n,w}(\alpha) = \prod_{r=1}^{R}\frac{1}{\sqrt{(2\pi)^r}\left|\Sigma\right|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(y_r^* - x_r^*)'\Sigma_r^{-1}(y_r^* - x_r^*)\right\} \qquad \text{(A.1.4)}$$

where $\alpha = (\beta, \gamma', \delta', \sigma^2)'$. In Lee (2007), the weighting matrix W_r is an $m_r \times m_r$ interaction matrix with $W_{ij} = \dfrac{1}{m_r - 1}$ if $i \neq j$, and W_{ij} =0 if i=j. If all the W_r , r=1,2,…R, are in such a structure, Lee (2007) showed that there exists a very formalized expression of Equation (A.1.2), and results in an unique formalized expression of Σ_r^{-1} , and $\left|\Sigma\right|^{\frac{1}{2}}$. Eventually, there exists a generalized solution for the log-likelihood function.In this paper, W_r is defined based on the relative size of work flow, which means the structure of W_r in this paper is quite different from that defined in Lee (2007). It is not possible to solve for Σ_r^{-1}, and $\left|\Sigma\right|^{\frac{1}{2}}$ in this paper, so CML estimator is not applicable in this research.

A.2 Tables

Table1.1: Summary statistics for price indexes for years 2000, 2006, and 2011

	Std. Dev.	Min	Max
2000			
Rent-based Price index	9.3	85.7	141.2
Quality-adjusted Gross Rents	19.2	70.8	185.9
Non-housing price	6.2	88.9	134.4
2006			
Rent-based Price index	11.6	82.1	149.3
Quality-adjusted Gross Rents	24.9	64.1	196.2
Non-housing price	6.9	87.3	130.1
2011			
Rent-based Price index	10.9	81.4	137.9
Quality-adjusted Gross Rents	24.1	65.3	192.4
Non-housing price	6.9	83.6	124.3

Note: Un-weighted mean is normalized to 100

Table 1.2: Empirical results based on OLS for all workers in years 2000, 2006, and 2011

	2000	2006	2011
Log Rent-based Price index	0.636***	0.782***	0.756***
	(0.049)	(0.044)	(0.062)
Significantly different from 1? (Y/N)	Y^a	Y^a	Y^a
N	1,036,922	196,449	213,904
R^2	0.449	0.457	0.457

Note: Cluster standard error in parentheses. *** denotes significant at 1% level. a denotes significant at 1% level.

Table 1.3: Empirical Results based on OLS for education groups in years 2000, 2006, and 2011

	Less than High School	High School	Some College	College and above
Panel A: 2000				
Log Rent-based Price index	0.497***	0.671***	0.683***	0.598***
	(0.071)	(0.053)	(0.062)	(0.049)
Significantly different from 1? (Y/N)	Y^a	Y^a	Y^a	Y^a
N	197,244	260,515	324,522	254,641
R^2	0.424	0.447	0.454	0.46
Panel B: 2006				
Log Rent-based Price index	0.597***	0.755***	0.874***	0.85***
	(0.059)	(0.044)	(0.053)	(0.061)
Significantly different from 1? (Y/N)	Y^a	Y^a	Y^b	Y^b
N	30,557	53,015	59,465	53,412
R^2	0.441	0.454	0.457	0.466
Panel C: 2011				
Log Rent-based Price index	0.553***	0.619***	0.813***	0.869***
	(0.058)	(0.059)	(0.068)	(0.067)
Significantly different from 1? (Y/N)	Y^a	Y^a	Y^a	Y^c
N	29,515	52,193	68,848	63,348
R^2	0.443	0.453	0.455	0.466

Note: Cluster standard error by CBSA/CSA in parentheses. *** denotes significant at 1% level. a denotes significant at 1% level, b denotes significant at 5% level, and c denotes significant at 10% level.

Table 1.4: Empirical results based on 2SLS for all workers in years 2000, 2006, 2011

	2000	2006	2011
Second Stage results			
Log Rent-based Price index	0.796***	0.864***	0.826***
	(0.062)	(0.05)	(0.059)
Significantly different from 1?	Y^a	Y^a	Y^a
N	1,017,736	196,449	213,904
R^2	0.449	0.457	0.457
First Stage results			
Corresponding Lagged log gross rent	0.42***	0.45***	0.425***
	(0.033)	(0.018)	(0.019)
Partial R^2 of Excluded Instrument	0.574	0.87	0.869
P value for Anderson-Rubin Wald	0	0	0
Kleibergen-Paap Wald rk F stat.	182.2	21336	15740.4

Note: Cluster standard error in parentheses. *** denotes significant at 1% level. a denotes significant at 1% level.

Table 1.5: Empirical results based on 2SLS for education groups in years 2000, 2006, and 2011

	Less than high school	**High school**	**Some college**	**College and above**
Panel A:2000				
Second Stage results				
Log Rent-based Price index	0.634*** (0.099)	0.857*** (0.071)	0.917*** (0.075)	0.697*** (0.057)
Significantly different from 1? (Y/N)	Y^a	Y^b	N	Y^a
N	194,153	254,915	317,533	251,135
R^2	0.424	0.447	0.454	0.46
First Stage results				
Log gross rent, 1990	0.386*** (0.034)	0.418*** (0.03)	0.418*** (0.032)	0.44*** (0.038)
Partial R^2 of Excluded Instrument	0.528	0.588	0.573	0.593
P value for Anderson-Rubin Wald test	0	0	0	0
Kleibergen-Paap Wald rk F stat.	143.1	233.5	194.5	148.4
Panel B:2006				
Second Stage results				
Log Rent-based Price index	0.636*** (0.066)	0.815*** (0.051)	0.969*** (0.06)	0.9*** (0.071)
Significantly different from 1? (Y/N)	Y^a	Y^a	N	N
N	30,557	53,015	59,465	53,412
R^2	0.441	0.454	0.457	0.466
First Stage results				
Log gross rent,2005	0.439*** (0.018)	0.449*** (0.017)	0.451*** (0.018)	0.459*** (0.018)
Partial R^2 of Excluded Instrument	0.854	0.863	0.86	0.89
P value for Anderson-Rubin Wald test	0	0	0	0
Kleibergen-Paap Wald rk F stat.	12927.2	20206.2	18885.7	19673.6
Panel C:2011				
Second Stage results				
Log Rent-based Price index	0.626*** (0.068)	0.67*** (0.066)	0.91*** (0.077)	0.96*** (0.079)
Significantly different from 1? (Y/N)	Y^a	Y^a	N	N
N	29,515	52,193	68,848	63,348
R^2	0.443	0.453	0.455	0.466
First Stage results				
Log gross rent,2010	0.425*** (0.02)	0.43*** (0.019)	0.427*** (0.02)	0.424*** (0.018)
Partial R^2 of Excluded Instrument	0.861	0.862	0.86	0.886
P value for Anderson-Rubin Wald test	0	0	0	0
Kleibergen-Paap Wald rk F stat.	12212.2	15891.9	18022.9	12426.2

Note: Cluster standard error in parentheses. *** denotes significant at 1% level. a denotes significant at 1% level, and b denotes significant at 5% level.

Table 1.6: Empirical results based on group-specific weights and 2SLS for all workers in years 2000, 2006 and 2011

	2000	2006	2011
Second Stage results			
Log Rent-based Price index	0.823***	0.982***	0.946***
	(0.074)	(0.14)	(0.147)
Significantly different from 1?	Y^b	N	N
N	1,017,736	196,449	213,904
R^2	0.449	0.456	0.456
First Stage results			
Corresponding Lagged log gross rent	0.405***	0.398***	0.371***
	(0.036)	(0.056)	(0.051)
Partial R^2 of Excluded Instrument	0.374	0.443	0.372
P value for Anderson-Rubin Wald	0	0	0
Kleibergen-Paap Wald rk F stat.	144.1	2607.6	2320

Note: Cluster standard error in parentheses. *** denotes significant at 1% level. b denotes significant at 5% level.

Table 1.7: Empirical results based on group-specific weights and 2SLS for education groups in years 2000, 2006, and 2011

	Less than high school	**High school**	**Some college**	**College and above**
Panel A:2000				
Second Stage results				
Log Rent-based Price index	0.647*** (0.108)	0.889*** (0.089)	0.958*** (0.09)	0.717*** (0.061)
Significantly different from 1? (Y/N)	Y^a	N	N	Y^a
N	194,153	254,915	317,533	251,135
R^2	0.424	0.446	0.454	0.46
First Stage results				
Log gross rent, 1990	0.378*** (0.038)	0.403*** (0.03)	0.4*** (0.034)	0.43*** (0.038e)
Partial R^2 of Excluded Instrument	0.342	0.362	0.368	0.41
P value for Anderson-Rubin Wald test	0	0	0	0
Kleibergen-Paap Wald rk F stat.	108.6	156.7	151.3	136.8
Panel B:2006				
Second Stage results				
Log Rent-based Price index	0.685*** (0.112)	0.938*** (0.155)	1.11*** (0.154)	1.03*** (0.184)
Significantly different from 1? (Y/N)	Y^a	N	N	N
N	30,557	53,015	59,465	53,412
R^2	0.441	0.454	0.456	0.465
First Stage results				
Log gross rent,2005	0.407*** (0.056)	0.342*** (0.067)	0.393*** (0.05)	0.4*** (0.06)
Partial R^2 of Excluded Instrument	0.488	0.306	0.449	0.417
P value for Anderson-Rubin Wald test	0	0	0	0
Kleibergen-Paap Wald rk F stat.	980.7	1629.3	1668.2	2103.9
Panel C:2011				
Second Stage results				
Log Rent-based Price index	0.682*** (0.122)	0.769*** (0.125)	1.01*** (0.141)	1.13*** (0.197)
Significantly different from 1? (Y/N)	Y^b	Y^c	N	N
N	29,515	52,193	68,848	63,348
R^2	0.443	0.452	0.455	0.465
First Stage results				
Log gross rent,2010	0.39*** (0.058)	0.37*** (0.05)	0.374*** (0.046)	0.361*** (0.052)
Partial R^2 of Excluded Instrument	0.418	0.368	0.403	0.326
P value for Anderson-Rubin Wald test	0	0	0	0
Kleibergen-Paap Wald rk F stat.	1105.4	1362.6	1701.5	1870.8

Note: Cluster standard error in parentheses. *** denotes significant at 1% level. a denotes significant at 1% level, b denotes significant at 5% level, and c denotes significant at 10% level.

Table 2.1: Cross-section dependence (CD) test results

Area	CD test	P_value
China	-2.39	0.017
East region	-2.97	0.03
Central region	-3.55	0
West region	-4.23	0

Note: CD test denotes Pesaran's (2007) test statistic with the null hypothesis of cross-section independence

Table 2.2: Information criteria results

No. of common factors	China		East region		Central region		West region	
	IC1	IC2	IC1	IC2	IC1	IC2	IC1	IC2
1	-6.592	-6.553	-7.212	-7.017	-7.717	-7.519	-8.291	-8.095
2	-6.581	-6.502	-7.025	-6.784	-6.942	-6.81	-7.858	-7.695
3	-6.575	-6.458	-6.821	-6.756	-7.234	-7.069	-7.602	-7.471
4	-6.58	-6.424	-6.791	-6.753	-6.745	-6.619	-7.311	-7.213
5	-6.588	-6.392	-6.918	-6.857	-6.685	-6.646	-7.115	-7.049
6	-6.565	-6.33	-6.857	-6.757	-6.605	-6.572	-6.939	-6.906
K^*	1	1	1	1	1	1	1	1

Table 2.3: Unit root test results for common factor

Areas	1978-2007		1978-2012	
	ADF t-statistic	RMA-ADF t-statistic	ADF t-statistic	RMA-ADF t-statistic
China	-1.279	-0.364	-1.413	-0.483
East region	-1.776*	-0.589	-1.596*	-0.718
Central region	-0.985	-0.096	-1.453	-0.629
West region	-2.442**	-1.01	-1.846*	-1.219

Note: * and ** denotes significant at 10% and 5% level, respectively.

Table 2.4: Unit root test results for idiosyncratic components for China

Provinces	1978-2007	1978-2012
	ADF t-statistic	ADF t-statistic
Beijing	-0.15	2.79
Tianjing	-0.57	-0.34
Hebei	-0.51	-1.02
Liaoling	0.93	-0.03
Shanghai	0.77	1.28
Jiangsu	1.93	1.69
Zhejiang	1.92	0.87
Shandong	4.14	1.05
Fujian	1.24	1.43
Guangdong	0.80	-0.26
Guangxi	-4.77***	-3.27**
Shanxi	-3.42**	-1.96
Inner Mongolia	2.20	2.19
Jilin	-2.11	-1.16
Heilongjiang	1.50	1.47
Anhui	-2.06	-2.18
Jiangxi	-0.62	-0.91
Henan	2.12	1.28
Hubei	-0.76	-1.67
Hunan	0.10	-0.49
Sichuan	1.01	-0.80
Guizhou	-1.81	-1.79
Yunnan	-1.11	-1.29
Shaaxi	0.47	1.67
Gansu	1.35	1.80
Qinghai	1.64	1.38
Ningxia	0.61	-0.43
XinJiang	-0.61	-0.72
$P_{\hat{e}}$	0.67	0.25

Note: * *and *** denotes significant at 5% and 1% level, respectively.

Table 2.5: Unit root test results for idiosyncratic component for east region

Provinces	1978-2007	1978-2012
	ADF t-statistic	ADF t-statistic
Beijing	2.47	0.04
Tianjing	-0.76	-0.52
Hebei	-2.10	-1.94
Liaoling	-0.74	0.47
Shanghai	0.89	0.78
Jiangsu	1.22	0.74
Zhejiang	0.67	1.33
Shandong	2.61	2.91
Fujian	1.01	0.59
Guangdong	-0.34	0.42
Guangxi	-2.84*	-2.37
$P_{\hat{e}}$	-0.57	0.12

Note: * denotes significant at 10% level.

Table 2.6: Unit root test results for the idiosyncratic component for the Central region

Provinces	1978-2007	1978-2012
	ADF t-statistic	ADF t-statistic
Shanxi	0.04	-0.20
Inner Mongolia	-1.12	-1.09
Jilin	0.21	0.32
Heilongjiang	-0.09	0.38
Anhui	-1.28	-1.55
Jiangxi	-1.00	-0.34
Henan	1.04	0.01
Hubei	0.53	0.88
Hunan	0.09	-0.18
$P_{\hat{e}}$	-1.29	-1.7

Table 2.7: Unit root test results for idiosyncratic component for the West region

Provinces	1978-2007	1978-2012
	ADF t-statistic	ADF t-statistic
Sichuan	-1.38	-1.22
Guizhou	0.29	-0.11
Yunnan	-0.66	-0.17
Shaaxi	0.75	0.54
Gansu	-2.69*	-1.82
Qinghai	-1.41	-1.57
Ningxia	-3.23**	-2.36
XinJiang	-0.04	-0.33
$P_{\hat{e}}$	1.53	1.01

Note: * and **denotes significant at 5%, and 10% level, respectively.

Table 3.1: Description and summary statistics of data

Variable	Description	Obs	Mean	Std.
Dependent variable				
emp_grow0007	employment growth 2000-2007	3039	0.066	0.145
emp_grow0010	employment growth 2000-2010	3039	0.045	0.17
County Fiscal Variables 1997 (Share of personal income)				
Property tax	Revenue from property tax	3039	0.031	0.03
Highway expenditure	Expend. on highway	3039	0.009	0.011
Safety expenditure	Expend. on public safety(police + fire protection)	3039	0.006	0.004
Education expenditure	Expend. on education	3039	0.057	0.023
Sales tax	Revenue from sales tax	3039	0.004	0.005
Knowledge spillover variables (2000)				
Log Emp. Density 2000	Natural log of employment density 2000 (Thousand person per squared kilometer)	3039	2.103	1.658
High_tech emp. Share 2000	Employment share of high-tech industry[37] of year 2000	3039	0.057	0.067
Demographic Variables (2000)				
African	Percent population of African American	3039	0.286	2.853
Native	Percent population of Native American	3039	0.037	0.768
Asianpacific	Percent population of Asian	3039	0.027	0.414
other	Percent population of other races	3039	0.063	0.705
Hispanic	Percent population of Hispanic	3039	0.159	1.769
Married	Percent population (15years over) that are married	3039	0.605	0.053
Female	Percent population that are female	3039	0.505	0.019
disable	Percent Civilian non-institutionalized population 16 to 64 years with a work disability	3039	0.112	0.03
lingiso	Percent household with linguistic isolation prob.	3039	0.014	0.022

[37] High technology industries are chosen according to the BEA high-tech industry definition. For certain industries, the US census data only show the number of establishments and the employment group those establishments belong to. For these industries, employment is calculated as the product of the number of establishments with the mean employment of the employment group that those establishments belong to.

Table 3.1 continued.

Variable	Description	Obs	Mean	Std.
Education Variables (2000)				
High-school	Percent population 25 years and over that are high school graduates	3039	0.348	0.065
Some college	Percent population 25 years and over that attend college but without a bachelor's degree	3039	0.204	0.044
Associate	Percent population 25 years and over that have an associate degree	3039	0.057	0.02
Bachelor	Percent population 25 years and over with education attainment equal or higher than a bachelor's degree	3039	0.164	0.076
Amenity Variables				
Jantemp	Mean temperature for January, 1941-71	3039	32.942	12.055
Jansun	Mean hours of sunshine for January, 1941-71	3039	151.479	33.142
Julytem	Mean temperature for July, 1941-71	3039	75.914	5.337
Julyhumid	Mean relative humidity for July, 1941-71	3039	56.136	14.522
topography	Topography score ranging from 1-21, where 1 represents flat plain and 21 represent most mountainous land	3039	8.84	6.576
Additional Control Variables				
RC2003	Rural-Urban Continuum Codes 2003	3039	5.111	2.673
Navigable river distance	distance to navigable river (KM)	3039	245.934	242.766

Table 3.2: Summary statistics of group sizes

	Obs	**Mean**	**Std.**	**Min**	**Max**
Group size	656	4.63	2.26	2	17

Table 3.3: F test stat., weak instrument test stat., and overidentification test stat. for first stage regressions for 2000-2007 and 2000-2010.

	2000-2007	**2000-2010**
Panel A: G2SLS Procedure		
1st round 2SLS :		
$(I-W)W^2x_1$ as excluded instrument for $(I-W)Wy$		
First stage F stat.	5.41	6.62
P value for first stage F stat.	0	0
Kleibergen-Paap Wald F stat.	7.48	8.67
Hansen-J test stats.	9.81	4.89
P value for Hansen-J test	0.13	0.56
2nd round 2SLS :		
$\hat{E}[(I-W)Wy(\theta) \mid x, W]$ as excluded instrument for $(I-W)Wy$		
First stage F stat.	44.41	51.02
P value for first stage F stat.	0	0
Kleibergen-Paap Wald rk F	146.77	130.79
Panel B: Spatial-Durbin IV		
W^2x_1 as excluded instrument for Wy		
First stage F stat.	7.7	9.18
P value for first stage F stat.	0	0
Kleibergen-Paap Wald rk F	11.48	12.75
Hansen-J test stats.	7.35	3.11
P value for Hansen-J test	0.29	0.79

Table 3.4: Empirical results for 2000-2007

	Hedonic OLS		**Spatial-Durbin IV**		**G2SLS**	
	Coeff.	t-stat.	Coeff.	t-stat.	Coeff.	t-stat.
Endogenous effect			-0.16	-1.09	0.013	0.1
Contextual effect						
Property tax 1997			0.19*	1.67	-0.24	-0.84
Highway expenditure 1997			0.08	0.36	-0.76	-1.27
Safety expenditure 1997			-0.59	-0.93	1.19*	1.74
Education expenditure 1997			-0.1	-0.43	-0.24	-0.64
Sale tax 1997			2	1.14	1.15	0.54
Log Emp. density 2000			-0.006***	-3.38	-0.003	-1.09
High-tech emp. Share 2000			0.01	0.11	0.03	0.31
Individual effect						
Property tax 1997	-0.1	-0.93	-0.08	-0.42	-0.29	-1.49
Highway expenditure 1997	-0.32	-1.18	-0.33	-1.03	-0.87	-1.54
Safety expenditure 1997	-1.92*	-1.79	-2.24**	-2.3	-1.27	-0.79
Education expenditure 1997	0.28*	1.84	0.22***	2.72	0.12	0.54
Sale tax 1997	0.91	1	1.42	1.54	0.82	0.69
Log Emp. density 2000	-0.07***	-3.69	-0.04*	-1.78	-0.02***	-3.74
High-tech emp. Share 2000	-0.05	-1.48	-0.09*	-1.86	-0.06	-1.37

Note: *,**, and *** denote significance at the 10%, 5% and 1% level, respectively. The t-statistics are based on residuals clustered over commuting zone.

Table 3.5: Empirical results for 2000-2010

	Hedonic OLS		**Spatial-Durbin IV**		**G2SLS**	
	Coeff.	t-stat.	Coeff.	t-stat.	Coeff.	t-stat.
Endogenous effect			-0.25**	-2.29	0.046	0.39
Contextual effect						
Property tax 1997			0.5	1.43	-0.57	-1.32
Highway expenditure 1997			0.19	0.63	-0.38	-0.62
Safety expenditure 1997			1.26	0.92	3.68*	1.79
Education expenditure 1997			0.38	1.35	-0.53	-1.25
Sale tax 1997			1.89*	1.75	0.82	0.32
Log Emp. density 2000			0.02***	3.71	-0.01	-1.33
High-tech emp. Share 2000			0.18	1.18	0.21*	1.75
Individual effect						
Property tax 1997	0.2	1.45	0.46**	2.4	-0.06	-0.31
Highway expenditure 1997	0.14	0.69	0.33	1.57	-0.29	-0.51
Safety expenditure 1997	-0.89	-0.83	-0.81*	-1.76	0.99	0.53
Education expenditure 1997	0.6***	3.03	1.35***	4.52	0.28	1.1
Sale tax 1997	-0.1	-0.8	0.09	0.65	-0.27	-0.17
Log Emp. density 2000	-0.04***	-8.89	-0.03*	-1.69	-0.05***	-6.39
High-tech emp. Share 2000	-0.01	-0.75	0.06	0.61	0.03	0.33

Note: *,**, and *** denote significant at 10%, 5% and 1% level, respectively. T-stat. based on residuals cluster over commuting zone.

Table 3.6: Monte Carlo Simulation results (1000 replicates).

	G2SLS		Spatial-Durbin IV	
	Mean	Std. Dev.	Mean	Std. Dev.
Endogenous effect	0.05	0.07	0.009	0.11
Contextual effect				
Property tax 1997	-0.57	0.16	-0.55	0.15
Highway expenditure 1997	-0.36	0.32	-0.47	0.32
Safety expenditure 1997	5.51	1.27	4.86	1.22
Education expenditure 1997	-0.54	0.24	-0.5	0.23
Sale tax 1997	0.81	0.75	0.73	0.71
Log Emp. density 2000	-0.11	0.01	-0.01	0.01
High-tech emp. Share 2000	0.21	0.07	0.31	0.06
Individual effect				
Property tax 1997	-0.06	0.09	-0.06	0.11
Highway expenditure 1997	-0.26	0.22	-0.29	0.23
Safety expenditure 1997	0.96	0.67	0.94	0.66
Education expenditure 1997	0.27	0.12	0.28	0.13
Sale tax 1997	-0.28	0.51	-0.27	0.5
Log Emp. density 2000	-0.05	0.02	-0.05	0.02
High-tech emp. Share 2000	0.03	0.04	0.03	0.05

Table 3.7: First-stage results for 1st round 2SLS in the G2SLS procedure for period 2000-2007

	Coefficient	t-stat.
Panel A: $(I - W)$ multiply		
Property tax 1997	0.36***	3.00
Highway expenditure 1997	-0.34	-1.62
Safety expenditure 1997	-1.75**	-2.51
Education expenditure 1997	-0.42***	-3.05
Sale tax 1997	-0.67	-0.81
Log Emp. Density 2000	0.02***	4.06
High-tech emp. Share 2000	0.05	1.34
Panel B: $(I - W)W$ multiply		
Property tax 1997	0.28	0.55
Highway expenditure 1997	-0.17	-0.15
Safety expenditure 1997	-2.40	-0.70
Education expenditure 1997	-0.84**	-2.31
Sale tax 1997	0.65	0.17
Log Emp. Density 2000	0.01*	1.69
High-tech emp. Share 2000	0.17	1.29
Panel C: $(I - W)W^2$ multiply		
Property tax 1997	-0.17	-0.33
Highway expenditure 1997	0.29	0.29
Safety expenditure 1997	5.83	1.22
Education expenditure 1997	-0.38	-0.45
Sale tax 1997	2.86	0.44
Log Emp. Density 2000	0.02***	3.59
High-tech emp. Share 2000	0.17	1.13

Note: *, **, and *** denotes significant at the 10%, 5% and 1% level, respectively. T-stat. based on residuals cluster over commuting zone.

Table 3.8: First-stage results for 2nd round 2SLS in the G2SLS procedure for period 2000-2007

	Coefficient	**t-stat.**
Panel A: $(I-W)$ multiply		
Property tax 1997	0.24***	2.80
Highway expenditure 1997	-0.29*	-1.67
Safety expenditure 1997	-1.41**	-2.53
Education expenditure 1997	-0.10	-1.03
Sale tax 1997	-0.44	-0.64
Log Emp. Density 2000	0.01***	3.09
High-tech emp. Share 2000	0.02	0.71
Panel B: $(I-W)W$ multiply:		
Property tax 1997	0.34*	1.72
Highway expenditure 1997	-0.26	-1.09
Safety expenditure 1997	-7.57***	-3.94
Education expenditure 1997	-0.50	-1.41
Sale tax 1997	-2.61	-1.02
Log Emp. Density 2000	-0.01	-1.57
High-tech emp. Share 2000	0.07	0.93
$\hat{E}[(I-W)Wy(\theta) \mid x, W]$	0.5***	6.66

Note: *, **, and *** denotes significant at the 10%, 5% and 1% level, respectively. T-stat. based on residuals cluster over commuting zone.

Table 3.9: First-stage results for 1st round 2SLS in the G2SLS procedure for period 2000-2010

	Coefficient	t-stat.
Panel A: $(I-W)$ multiply		
Property tax 1997	0.49***	3.47
Highway expenditure 1997	-0.23	-0.83
Safety expenditure 1997	-2.00**	-2.53
Education expenditure 1997	-0.51***	-3.09
Sale tax 1997	-0.68	-0.76
Log Emp. Density 2000	0.02***	5.28
High-tech emp. Share 2000	0.06	1.30
Panel B: $(I-W)W$ multiply		
Property tax 1997	1.92*	1.66
Highway expenditure 1997	0.69	0.51
Safety expenditure 1997	-4.07	-1.00
Education expenditure 1997	0.36	0.51
Sale tax 1997	-0.84	-0.20
Log Emp. Density 2000	0.02*	1.82
High-tech emp. Share 2000	0.26	1.58
Panel C: $(I-W)W^2$ multiply		
Property tax 1997	0.88	0.79
Highway expenditure 1997	0.88	0.68
Safety expenditure 1997	5.87	1.16
Education expenditure 1997	1.00	1.16
Sale tax 1997	2.76	0.39
Log Emp. Density 2000	0.03**	2.37
High-tech emp. Share 2000	0.35*	1.85

Note: *, **, and *** denotes significant at the 10%, 5% and 1% level, respectively. T-stat. based on residuals cluster over commuting zone.

Table 3.10: First-stage results for 2nd round 2SLS in the G2SLS procedure for period 2000-2010

	Coefficient	t-stat.
Panel A: $(I-W)$ multiply		
Property tax 1997	0.2**	2.30
Highway expenditure 1997	-0.27	-1.39
Safety expenditure 1997	-1.32**	-2.24
Education expenditure 1997	-0.04	-0.34
Sale tax 1997	-0.39	-0.51
Log Emp. Density 2000	0.01***	2.74
High-tech emp. Share 2000	0.01	0.40
Panel B: $(I-W)W$ multiply		
Property tax 1997	0.32	1.47
Highway expenditure 1997	-0.21	-0.70
Safety expenditure 1997	-5.87***	-2.95
Education expenditure 1997	-0.49	-1.19
Sale tax 1997	-3.34	-1.20
Log Emp. Density 2000	-0.01	-1.22
High-tech emp. Share 2000	0.04	0.46
$\hat{E}[(I-W)Wy(\theta) \mid x, W]$	0.73***	7.14

Note: *, **, and *** denotes significant at the 10%, 5% and 1% level, respectively. T-stat. based on residuals cluster over commuting zone.

VITA

Ying Tan

Candidate for the Degree of

Doctor of Philosophy

Thesis: ESSAYS IN APPLIED MICROECONOMICS

Major Field: Economics

Biographical:

Education:

Completed the requirements for the Doctor of Philosophy in Economics at Oklahoma State University, Stillwater, Oklahoma in December, 2015.

Completed the requirements for the Master of Arts in Economics at Nationalities University of Yunnan, Kunming, China in 2010.

Completed the requirements for the Bachelor of Arts in Economics at Xiangtan University, Xiangtan, China in 2005.

Experience:

Instructor of Principal of Economics, full responsibility,
Oklahoma State University, Fall 2014, Spring 2015.

Research Assistant at the CAER, Oklahoma State University, Fall 2010-Spring 2014.

Professional Memberships: