MODELLING INTERPROVINCIAL MIGRATION IN CHINA: THE CORRELATION BETWEEN IMMIGRATION AND CITIES' ECONOMY DEVELOPMENT

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Master of Science

by

Yiqi Ou

August 2016

ABSTRACT

Interprovincial migration has significant impact on Chinese population change and cities' development as previous studies find out. However, *hukou* system and a series of impeditive policies implemented by authority increase difficulty for migrants to stay in the immigrated cities. This study would like to analysis the correlation between migrants and cities' economy development. Given the absence of using simultaneous equations system with panel data in studying floating migration, this study attempts to fill up this gap by using system equations with panel data, estimated by 3sls with maximum likelihood in the last step. The results can be concluded as: in the pulling forces of drawing migration to the cities, the disparity of employment rate and average income between emigrated and immigrated cities as well as the manufacturing industry growth in immigrated cities have huge contribution. Moreover, migration do play a significant role in promote the cities' economy development.

BIOGRAPHICAL SKETCH

Born and Grown up in Canton, the fourth biggest immigrated cities in China, Yiqi Ou has been interested in studying floating migration since she was at her undergrad. During her undergrad study in Sun Yat-sen University from 2000 to 2014, she has conducted a series of researches and portrait analysis based on the migrants in Canton. After obtaining Bachelor of Arts in mass media communication and design in 2014 in China, Yiqi Ou took a further step to pursue the Master of Science degree in Regional Science at Cornell University. Given that this study yields macro aggregate level results of migration, Yiqi would like to further her study in two aspects. The first aspect is to take a farther step in the model improvement by adding dynamic model to the original one and increase the target cities number (shorten the time period) to solve the selection bias problem. The second aspect is to understand the micro level process of decision making of migrants by using nonlinear probability model and compare the result with the macro level outcomes in this study.

This document is dedicated to all Cornell graduate students.

ACKNOWLEDGMENTS

I cannot express enough thanks to my committee for their continued support and encouragement: Prof. Yuri Surtadi Mansury, my committee chair; and Prof. Kieran Donaghy. I offer my sincere appreciation for the learning opportunities provided by my committee.

My completion of this project could not have been accomplished without the support of my classmates and my friend, Yimin. Thanks to my parents as well, Mr. Ou and Mrs. Wen. The countless encouragement and support you gave to me during my hectic schedules will not be forgotten. My heartfelt thanks.

Finally, many thanks to Cornell University. It gave me unforgettable and treasurable memory that I could keep in my mind forever.

TABLE OF CONTENTS

[1] FGLS stands for Feasible Generalized Least Squares

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Migration has become an increasingly important component in population change in China. There are different kinds of migration due to implement of *hukou* system and geographical distribution. What is *hukou* system? It is a household registration system designed to control population migration and labor mobility between rural and urban areas as well as across regions in the planned economy in China. The issuing of Regulations on Household Registration of the People's Republic of China in 1958 marked the beginning of the formal establishment of the *hukou* system (Fang and Dewen, 2008). From the geographical aspect, migration can be divided into intra-provincial and inter-provincial migration. Intra-provincial migration is with emigrated and immigrated regions in the same province. While inter-provincial migration is population whose emigrated region and immigrated region are in different provinces. With detailed introduction of various types of migration in latter section, this study focuses on interprovincial floating immigration, the one without *hukou* registered in immigrated cities and staying in the immigrated cities for more than half a year. It is majority of inter-province migrants over the last 25 years (Chan, 2013). Also, it is massive population flow from rural to urban areas in post-reform China resulted of both institutional and structural changes cased by economic growth (Cai, 2008). Just as Chan (2013) points out in his study of China internal migration, since 1978, the reformation of China's economy, the success story of Chinese manufacturing in the last quarter-century is inextricably meshed with the story of migrant workers toiling for subsistence wages to produce goods for export. The total population of floating migration has raised from 6.57 million in 1978 to 221 million in 2010 (Xia *et al.,* 2015) This huge amount of migration modified the distribution of Chinese population as well as the labor market. From 2000 Censes dataset, 75% of total interprovincial migrants are moving to east region with 84.3% of total origin from central region and 68.3 of total origin from west region. Although we can see some changes in 2010 as explained in latter section, the basic blueprint has not changed through time. That is moving from undeveloped rural area to developed urban area. As De *et al*.(2002) conclude in their study, for Chinese cities to urbanized and modernized successfully, the nation must rely on labor markets to facilitate the shift from a largely rural population to an urban one. Also, numerous domestic and national scholars have viewed floating migration as the primary contributor to the cities' and national urbanization and economy development (Chan and Zhang, 1999; Fan, 2008; Bosker et al., 2012; Fu and Gabriel, 2012; Ma and Chen, 2012; Lu et al., 2013). Thus, understanding floating migration and the pulling force in immigrated cities that attracts floating migration labor are critical in studying Chinese cites' development.

However, there are many barriers for migrants to stay in the immigrated cities. These barriers are implemented mostly by authorities since they believe floating migrants will destroy local economy and social development by snatching local resources. Among these barriers, few are from the emigrated regions such as land tenure arrangements and mandatory marketing delivery quotas which continues to increase the cost of out-migration and dampen off-farm labor market participation (Mallee, 2000; Yang and Zhou, 1996). These studies are a little bit out-of-date and may not be the current situation. Other scholars focus on the obstacles that origin in the immigrated cities. They worry that several prominent urban institutions, such as the *hukou* system and the absence of social and educational services for the rural residents in cities, restrict entrance into urban labor markets (Johnson, 1995 1999; Fang and Dewen, 2008; Zhang and Shunfeng, 2003; Cindy, 2005). Some other studies even use nonlinear probability model to analyze 2000 Census data and have shown that the existence of the *hukou* system makes migrants workers much less likely to enter unban monopoly and non-competitive sectors (Wang et al. 2004; Zhang, 2004). Consequently, the key to understand whether it is necessary to carry out those impeditive policies is to finger out the correlation between floating migration and the local economy development.

As a result, this study is aiming to solve two questions. The first one is what drives floating migration to immigrate and the second one is the correlation between floating migration and the immigrated cities' economy development. In order to prove the relationship, hypothesis needs to be brought forward. In this study, we assume there's a simultaneous circulatory relation among floating migration and manufacturing industry and GDP per capital (economy development). The development of manufacturing industry in urban areas attract floating migrants moving to there to earn higher wage. Meanwhile, the floating migrants and increase in manufacturing output incite the development of local economy (GDP per capita). More labor input (floating migrants) and more capital input (GDP per capita) in turn, promote a new circle of development in manufacturing industry. Theoretical support is based on Chan's study about internal migration (2013) mentioned above. In that study, he summarizes that the total stock of rural migrant labor has been the backbone of China's manufacturing export industry since the mid-1990s. These floating migrants accounted for the great majority (70-80%) of the labor force in the early years of the 21st century (Chan, 2007), which promotes the industry reformation and also flourishes the economy. Similar conclusion is also gained from Fang (2008).

Without doubt, there are plenty of studies analyzing floating migration in China. However, this study is innovative and filling the gap of previous literatures studying floating migration in two aspects.

As for the methodology in previous literatures, many scholars have pay efforts in concluding previous findings and doing the descriptive statistic analysis (Fang and Dewen, 2008; Cindy, 2005; Cushing and Poot, 2003; Chan, 2013; Cai, 2002; Wang *et al*,. 2004, De *et al*,. 2002). Those who use model construction to analyze migration mostly apply single equation estimation (Liu *et al*,. 2015; Zhang and Shunfeng, 2003; Shen, 2016). Moreover, rare studies about China migration estimate the model using maximum likelihood estimation in simultaneous-equations system (three stage least squares, 3sls). Comparably, there are many studies about American migration using 3sls estimation (Greenwood, 1975). However, no previous study applies 3sls estimation methodology with panel data to analyze China floating migration. Thus, this study will fill up this gap by using 3sls estimation with panel data in simultaneous-equations system.

Another aspect is about dataset. Previous studies have pointed out, researches on internal immigration has proceeded along two strands (Cushing and Poot, 2003). One strand emphasizes micro-level behavior. It focuses on individual's or household's preferred choice of residential location and the influences that might trigger a shift to another location across a boundary of some sort, defined for example, by a maximum commuting range. Another strand emphasized "place" more than "individual" and tried to demonstrate the observed flow and the reasons of net inflow and outflow immigration of specific locations. The former strand uses increasingly micro-level data while the latter relies mostly on aggregate data. As Brian and Jacques (2004) mentioned in their study, although the usage of micro data has been a methodological change in migration research due to the rapid advances in computer technology and opened a door to improve understanding of the migration process in detail regarding to the role of personal characteristics and situations, aggregate analysis still has a role. Additionally, Greenwood (1997), Plane and Heins (2003) have analyzed the merits of utilizing aggregated data. They concluded that just as microeconomics which studies the behavior of individual economic agents and macroeconomics which studies the aggregated patterns, trends and cycles continue to coexist, the same should hold for micro and macro migration analyses. Besides, despite the increasing accessibility to micro-data in some countries, aggregate data remain the only data source in most developing and transition countries like China. Also, with the logit and probit analysis which is adapted from micro-data methods, aggregate data can yield very rich analytical results. So, based on previous findings and the feasibility of China's dataset, this paper will use prefectural city level cross-sectional time series aggregate data to conduct research. The time period of the dataset in this study is 14 years from 1999-2013, which remedies defect about outdated dataset usage in previous studies (Cindy, 2005).

The rest of the article is organized as follows. Section 2 displays the estimation methodology and analytical framework of this study. An identification of migration and current spatial pattern of floating migration will be introduced before getting to the data description in Section 3. The next Section shows the empirical investigation. Finally, this paper end up with concluding remarks in Section 5.

CHAPTER 2

ESTIMATION MECHANISM

Few recent studies (Shen 2016; Liu *et al*., 2015) that have attempted to estimate the magnitudes in which various factors have influenced interprovincial or intra-provincial floating migration in China employ either a single-equation, multiple regressions model or simultaneous-equation without panel data to estimate the coefficients of those variables deemed important in explaining the migratory movements that have occurred. The shortcoming of the former approach is that the estimated coefficients possess a simultaneous-equations bias (has consistency problem), since migration that has occurred has itself influenced the independent variables of the models (Greenwood 1975). While the latter approach cannot understand the effect of time on migration.

This paper attempts to fill this gap of studying floating migration in China by using simultaneous-equations system model with panel data.

In this paper, the estimation of the circulatory effect between immigrated city and migration is based on the basic structure hypothesis that the floating immigration is attracted by the development of manufacturing output in the immigrated region. Floating immigration and manufacturing output in immigrated region can help to increase its economic growth (GDP per capita). GDP per capita and Floating migration in immigrated region in turn, can contribute to the growth of manufacturing output (see following Fig.1).

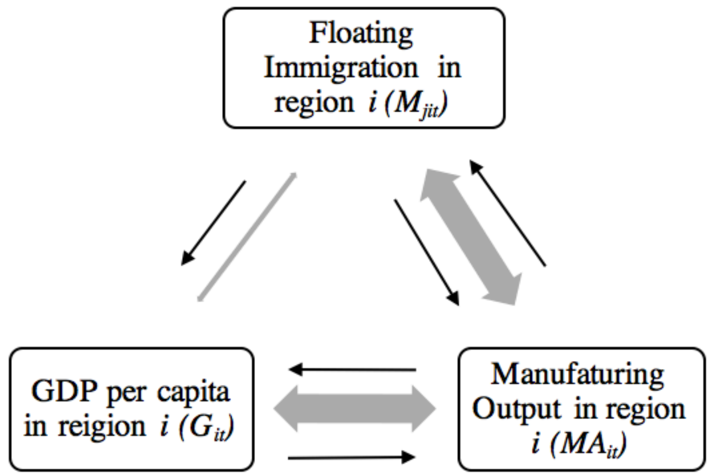

Figure 1: Basic Structure and Correlation among Dependent Variables

Depending on the structure hypothesis above, the basic model estimated for the study consists of 3 equations with 3 jointly endogenous variables. The 3 endogenous variables include 3 dependent variables which are the floating migration in immigrated region *i* from emigrated region *j* (M_{jit}), GDP per capita in immigrated region *i* (G_{it}) and the manufacturing output in immigrated region *i* (MA_{it}). Others are deemed as predetermined (lagged term of endogenous variables) or exogenous variables. *t* stands for time here. The systematic (simultaneous) equations are listed below, where *P* stands for population, *SM* stands for the ratio of service output and manufacturing output, *E* represents employment rate, *Y* indicates average income and *MA* means manufacturing output in function (1). *H* and *FG* represents average high school education rate and ratio of Fixed Asset over total GDP in function (2) separately. In function (3), *L* stands for labor, which is the product of employment rate and sum of population and floating immigration. The explanation of all the determines are listed in Table 1 and also Table 5. Three functions have all their explanatory variables on the right hand side to be lagged terms, which is to avoid reverse causal relationships between dependent and independent variables in exogenous variables list.

$$M_{jit} = f_1(\frac{P_{it-1}}{P_{jt-1}}, SM_{it-1}, \frac{E_{it-1}}{E_{jt-1}}, \frac{Y_{it-1}}{Y_{jt-1}}, \frac{MA_{it}}{MA_{it-1}}) \quad (1)$$

$$G_{it} = f_2\,(M_{it-1}, MA_{it-1}, P_{it-1}, H_{it-1}, FG_{it-1}) \quad (2)$$

$$MA_{it} = f_3\,(G_{it-1}, L_{it-1}, MA_{it-1}) \quad (3)$$

There are different ways to estimate the simultaneous-equation system above like two stage least squares (2sls) or three stage least squares (3sls). The estimation methodology in this paper is using the three stage least square (3sls) with maximum likelihood estimation in the last stage. Reasons for choosing 3sls over 2sls is because 3sls takes into account inter-temporal and not simultaneous correlations between error terms, which will be valuable in elevating the efficiency of estimated coefficients. While reasons for applying the maximum likelihood estimation in the last step is to improve the efficiency of the estimated parameters by iterating over the estimated disturbance covariance matrix and parameter estimates until the parameter estimates converge.

The basic deductive process of applying 3sls follows three steps (stata manuals13):

Consider the entire m-equation system model:

$$\begin{gathered} y_1 = \widetilde{Z_1} * \widetilde{F_1} + \varepsilon_1 \\ \vdots \\ y_m = \widetilde{Z_m} * \widetilde{F_m} + \varepsilon_m \end{gathered}$$

Where $\widetilde{Z_J} = [\widetilde{Y_J}\ \ \widetilde{X_J}]$ is vector of explanatory variables (endogenous and exogenous) in *j-th* equation and $\widehat{F_J}^{2sls}$ is vector of *j-th* equation's parameter estimates via 2sls.

The model can be denoted as matrix equation, where $\varepsilon_1,\ \varepsilon_2, ...$ are vertical vectors sized $T \times 1$:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} \widetilde{Z_1} & 0 & \cdots & 0 \\ \vdots & \widetilde{Z_2} & \ddots & \vdots \\ 0 & 0 & \cdots & \widetilde{Z_m} \end{bmatrix} \begin{bmatrix} \widetilde{F_1} \\ \widetilde{\mathrm{F}_2} \\ \vdots \\ \widetilde{F_m} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_3 \end{bmatrix}$$

The first two steps of calculating 3sls is the same as 2sls, where we estimate the variance-covariance matrix of the random disturbances $\mathrm{Var}(\varepsilon_t) = E(\varepsilon_t \varepsilon_t')$ element by element in a standard way: $\hat{\Sigma} = [\hat{\sigma}_{ij}]$

So, the first step is to compute the reduced-form estimation and theoretical values for *j-th* equation:

$$\widehat{\widetilde{Z_j}} = X\Pi_j = X(X'X)^{-1}X'\widetilde{Z_j}$$

As for the whole matrix system:

$$\hat{Z} = \begin{bmatrix} X(X'X)^{-1}X'\widetilde{Z_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & X(X'X)^{-1}X'\widetilde{Z_m} \end{bmatrix} = \{I_m \otimes [X(X'X)^{-1}X']\}Z$$

The second step is to estimate the parameters for individual equations in the structural form by 2sls estimation. Replace the empirical endogenous explanatory variables with theoretical values from first step:

$$\widehat{F^{2sls}} = (\hat{Z}'\hat{Z})^{-1}\hat{Z}'y$$

The third step is to compute the 3sls parameters by taking into account the simultaneous correlations of error terms in the model. If j-th equation's error term is spherical, its variance-covariance matrix is $\begin{bmatrix} {\sigma_{jj}}^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & {\sigma_{jj}}^2 \end{bmatrix}$

The variance-covariance of the entire error term vector ε will then be:

$$\varepsilon = \begin{bmatrix} {\sigma_{11}}^2 & \dots & {\sigma_{1m}}^2 \\ \vdots & \ddots & \vdots \\ {\sigma_{1m}}^2 & \dots & {\sigma_{mm}}^2 \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} = \Sigma \otimes I_T$$

Knowing the entire variance-covariance matrix $\hat{\Sigma}$, we can apply to the model y = ZF + ε, $\varepsilon \sim (0, \Omega)$. This is done jointly with using 2sls theoretical values for endogenous regressors:

$$\widehat{F^{3sls}} = (\hat{Z}'\Omega^{-1}\hat{Z})^{-1}\hat{Z}'\Omega^{-1}y$$

From previous steps, we have $\widehat{\Omega} = \hat{\Sigma} \otimes I_T$ and $\hat{Z} = \{I_m \otimes [X(X'X)^{-1}X']\}Z$. Thus,

$$\widehat{F^{3sls}} = [Z'\{\widehat{\Sigma^{-1}} \otimes [X(X'X)^{-1}X']\}Z]^{-1}Z'\{\widehat{\Sigma^{-1}} \otimes [X(X'X)^{-1}X']\}y$$

Next part, this paper is going to demonstrate the economic mechanism behind each simultaneous equation (function (1)(2)(3)) in detail.

2.1 describing the interprovincial floating immigration with modified gravity model

The adaption of modified gravity model in studying immigration is reanalyzed by Greenwood (2005). In that paper, he concludes that since 1960s, the modified gravity models have become commonly used in migration literature with several additional variables to the basic gravity model. Based on the property of gravity model, immigration in that model is hypothesized to be directly related to the size of relevant origin and destination populations and to be inversely related to distance, which are two major determinants. The place-to-place modified gravity migration model is then commonly taking the following form:

$$\begin{aligned}\ln M_{ji} = {} & ln\beta_0 + \beta_1 lnD_{ji} + \beta_2 lnP_i + \beta_3 lnP_j \\ & + \beta_4 lnY_i + \beta_5 lnY_j + \sum_{n=1}^{m} \beta_{in}\, lnX_{in} + \sum_{n=1}^{m} \beta_{jn}\, lnX_{jn} + \delta_{ji}\end{aligned}$$

where the *Y* terms refer to income, *P* refer to population and *D* refer to distance. Other variables that are commonly included (as reflected in terms containing *X*) are unemployment rates, degree of urbanization and median number of years of schooling and so on. Like the notation above, *i* represents the emigrated region and *j* represents the immigrated region. While X_{jn} contains factors that push migrants to move from *j* to *i*, X_{in} contains factors that draw migrants to move to *i*. This model is suitable to analyze the phenomenon of migration, however, is not appropriate to study the migration decision process since the correlation between modified gravity models and the migration decision process has not always been tight (Greenwood, 2005).

Referring to previous literature, this paper adapted modified gravity model with some adjustments to study the interprovincial immigration (notations remain the same as function (1)). The first adjustment (see function (4)) is to simplify determinants into ratio form. With this adjustment, it is much easier for us to interpret the result with classical push-pull migration theory. If the values of $In\frac{P_{it-1}}{P_{jt-1}}$, $In\frac{E_{it-1}}{E_{jt-1}}$, $In\frac{Y_{it-1}}{Y_{jt-1}}$ are bigger than zero, which means $\frac{P_{it-1}}{P_{jt-1}}$, $\frac{E_{it-1}}{E_{jt-1}}$, $\frac{Y_{it-1}}{Y_{jt-1}}$ are bigger than one, we can suggest that determinants *P, E, Y* play pull roles in drawing immigrants from *j* to *i* with coefficients $\beta_{In} > 0$ and vice versa. Here, as notated before, *P* refer to population, *E* represents employment rate and *Y* indicates average income. Also, instead of unemployment rate, here we use employment rate for the interpretation purpose. The second adjustment is to add manufacturing output increasing rate in immigrated city i as determinant. Manufacturing industry was and will still remain the pillar industry in the near future in China. City with high manufacturing output will provide high demand for low-skilled labor, which will be a pull force to attract floating immigration moving in (Fang and Dewen, 2008). However, with the reformation of Chinese industry structure, we can observe the increasing proportion of service industry output (the third industry) in total gross domestic product in recent decades. The third adjustment is to add the ratio of service output and manufacturing output in immigrated city *i*. With this variable, we can indicate whether the floating migration has transformed during the reformation of Chinese economy and whether service industry becomes a new pull force to draw floating immigration in immigrated cities. The forth adjustment is to get ride of the distance between emigrated place to immigrated place. The emigrated region j in this study is an aggregated provincial-level region with no access to measure its distance to the immigrated city i (detail explanation in chapter 3). Accordingly, the first equation of the simultaneous-equations system is shown as below, where *SM* stands for the ratio of service output and manufacturing output and *MA* is the manufacturing output in function. The expected signs of coefficients are summarized in Table 2 and explanation of each determinant is in Table 1:

$$InM_{jit} = \beta_{10} + \beta_{11}\, In\frac{P_{it-1}}{P_{jt-1}} + \beta_{12}\, InSM_{it-1} + \beta_{13}\, In\frac{E_{it-1}}{E_{jt-1}} + \beta_{14}\, In\frac{Y_{it-1}}{Y_{jt-1}} + \beta_{15}\, In\frac{MA_{it}}{MA_{it-1}} + \varepsilon_{it} \quad (4)$$

2.2 estimating the impact of interprovincial floating immigration on immigrated city's economy

Depending on the previous study of economy development (Jones, 1998), production growth basically relies on three factors, capital input, labor input and the productivity (which is measured by Solow residual). In the study of understanding the impact of floating immigration on immigrated city's economy, immigration itself should definitely be an explanatory variable as labor input. The hypothetic impact should be positive. In addition to that, the local registered population in city *i* should also be considered as explanatory variable in understanding the production growth in city *i*. It reflects the hypothesis that greater urban population and urban concentrations in a city tend to grow faster in a cumulative manner due to agglomeration economies (Alonso, 1971; Zhang and Shunfeng, 2003). To some extent, part of the local registered population can be taken as labor input. For the other side of input, capital input, this study chooses manufacturing output in city *i*, high school education rate in city *i* and proportion of fixed asset investment over total GDP as determinants. Basically, increase in fixed asset investment will improve the labor productivity and thus accelerate the production growth. High school education rate, seen as human capital investment, will also improve the labor productivity and trigger the reformation of economy changing from a low efficiency and high pollution one to a high efficiency high-tech and light pollution one in the meanwhile. Thus, it will assist in the growth of production in city *i*. Manufacturing output in city i, as a traditional and mainstream of capital input, will still play supportive effect on the growth of production in city i under the hypothesis of this paper. The Solow residual is represented by the constant term β_{20} in the function. As result, the second equation of the simultaneous-equations system is developed as function (5), where *G* stands for GDP per capita, *MA* stands for manufacturing output, *H* stands for high school education rate and *FG* means ratio of Fixed Asset over total GDP, *M* represents the total floating migration from j to i as the same as previous notation (the expected signs of coefficients in the equation are in Table 2 and each notation's explanation is listed in Table 1) :

$$InG_{it} = \beta_{20} + \beta_{21}\, InM_{jit-1} + \beta_{22}\, InMA_{it-1} + \beta_{23}\, InP_{it-1} + \beta_{24}\, InH_{it-1} + \beta_{25} InFG_{it-1} + \mu_{it} \quad (5)$$

2.3 describing the production function of manufacture industry

The production function of manufacture industry is adapted from Cobb-Douglas production function with capital input and labor input. Capital input is using GDP per capital(notated as *G*) from one lagged period (the production output in the second equation above) and labor input is applying the product of employment rate (notated as *E*) in city *i* times the summation of population(notated as *P*) and floating immigration(notated as *M*) in city *i* as a proxy. Here, assumption is making that both registered population and floating migration are employed at the same rate. Given imperfection of accessible data, this is the most accurate proxy we could have.

Under the hypothesis, the increase in capital input as well as the labor input will have a positive effect on the growth of manufacturing output. Besides, rather than estimating the static equation, this study estimates the dynamic production process by letting one time lagged dependent variable to be explanatory variable. With dynamic model, it is possible to obtain an estimate of the "short run" and "long run" behavior patterns of an economic unit over time (Grunfeld, 1961).

Under the hypothesis of this study, (t-1) period manufacturing output should have positive effect on t period manufacturing output. All parameters' expected effects are listed in Table 2 and explanation of each variable is listed in Table 1. The third equation of the simultaneous-equation system is illustrated as function (6):

$$InMA_{it} = \beta_{30} + \beta_{31} InG_{it-1} + \beta_{32} In\,[E * (P+M)]_{it-1} + \beta_{33} InMA_{it-1} + \xi_{it} \quad (6)$$

Table 1: Description of all the determinants

Variables	Definition
InM_{jit}	Inter-provincial floating migration in *i from j* (logarithm)
$In\frac{P_{it-1}}{P_{jt-1}}$	One time lagged of total *hukou* registered population in i divided by total *hokou* registered population in *j* (logarithm)
$InSM_{it-1}$	One time lagged of service industry output devided by manufacture industry output in *i* (logarithm)
$In\frac{E_{it-1}}{E_{jt-1}}$	One time lagged of employment rate in *i* divided by employment rate in *j* (logarithm)
$In\frac{MA_{it}}{MA_{it-1}}$	Ratio of manufacture industry ouput at time t divided by at time (t-1) in *i* (logarithm)
$In\frac{Y_{it-1}}{Y_{jt-1}}$	One time lagged of average income per person in *i* divided by average income per person in *j* (logarithm)

InG_{it}	GDP per capita in *i* (logarithm)
$InMA_{it-1}$	One time lagged manufacture industry output in *i* (logarithm)
InP_{it-1}	One time lagged total *hukou* registered population in i (logarithm)
InH_{it-1}	One time lagged high school education rate in *i* (logarithm)
$InFG_{it-1}$	One time lagged of total fixed asset divided by GDP in *i* (logarithm)
InG_{it-1}	One time lagged GDP per capita in *i* (logarithm)
$In(P+M)_{it-1}$	One time lagged the sum of population plus migration in i
InE_{it-1}	One time lagged employment rate in *i* (logarithm)
$InMA_{it}$	Manufacture industry output in *i* (logarithm)
InM_{jit-1}	One time lagged total floating migration in *i* from *j* (logarithm)

Table 2: Expected effect of each parameter
*Note: the four parameters with determinants ranging between negative values (refer to chapter 3)

Parameter	Description	Expected effect
β_{11}	Effect of population in *i* divided by population in *j at t-1*	-*
β_{12}	Effect of service output divided by manufacturing output in *i* at t-1	-*
β_{13}	Effect of employment rate in *i* divided by that in *j* at t-1	+
β_{14}	Effect of average income in *i* divided by that in *j* at t-1	+
β_{15}	Effect of manufacturing output at t divided by that at t-1 in *i*	+
β_{21}	Effect of floating immigration in *i* at t-1	+
β_{22}	Effect of manufacturing output in *i* at t-1	+
β_{23}	Effect of total registered population in *i* at t-1	+
β_{24}	Effect of high school education rate in *i* at t-1	+
β_{25}	Effect of fixed asset over total GDP in *i* at t-1	-*
β_{31}	Effect of GDP per capital in *i* at t-1	+
β_{32}	Effect of total labor in *i* at t-1	-*
β_{33}	Effect of manufacturing output in *i* at t-1	+

CHAPTER 3

DATA

For better understanding the characters of floating migration movement, this section displays the content by doing descriptive statistics analysis of the spatial pattern change of migration in China from 2000 to 2010 in the first part. By comparing two indicators, GDP per capital change and hukou registered population change in each city with immigrated city ranking in these two years, the characters of migration's spatial pattern can be concluded into two points. In the second part, this paper will give detailed data description of data processing and parameters.

3.1 migration's definitions and spatial patterns in China

Given the complexity of China's population components, there is a variety of ways to define migration. In this study, I primarily use National Bureau of Statistics (NBS)'s definition of migration, with adaptations of data availability. Basically, estimates of the number of migrants in China can be related to three elements: the length of stay, the geographic boundary crossed (township or county) and official status (with or without *hukou*). Based on this standard, the three categories are: (1) planned *hukou* migrants; (2) permanent migration with or without *hukou* change; (3) the "floating" rural labor force. Here, migration with a change of *hukou* is planned migration approved annually by the Ministry of Public Security and it reflects officially recognized population reallocation. While "permanent" stands for population that have resided in this administrative region for more than half a year. And "floating" rural labor equals permanent migration with or without *hukou* change minus *hukou* registered population plus *hukou* registered population that have been outside of the city more than half a year minus planned *hukou* migrants.[2] As mentioned in the previous chapter, this study focus on the "floating" rural labor migrant from other province. However, data set about them is only available in National Population Census every decade and in very few cites' yearly statistics book. An alternative method would be to use permanent population of the immigrated administrative region, which contains both population with and without *hukou* registration in this region, minus local *hukou*-registered population as a proxy. Both these two variables can be obtained in City Statistical Yearbook of each city. According to the statistical calculation of permanent population defined by National Bureau of Statistics (NBS), this method cannot rule out 'noise' from *hukou* backlog population[3]. To some extent, although this may cause overestimation of the true number of "floating" rural labor migrant from other province, it is the most accurate proxy can be got from existing reachable database. Again, *hukou* backlog population partially contains rural labor from other provinces with an undefined *hukou* status.

The complexity of studying China's migration not only lies in the multiple definitions of migration as mentioned above and statistical inaccuracy caused by the huge number of "floating" rural labor migration, but also the high level of mobility. From previous study, China's migration characteristics differ through time. Based on data from 2000 and 2010 Population Census of the People's Republic of China, inter-province migration gross[4] has expanded three times during one decade, from 28.24 million to 83.28 million. Not only the volume, but also the spatial pattern has

[2] Based on the definition of NBS

[3] Population with undecided *hukou* status like people with relocation certificate.

[4] Population with *hukou* registered in other provinces that resides in this region more than half a year.

changed during this period. Floating labor migration flows are closely linked to significant disparities in income between the urban and rural sectors and between regions in China (Chan 1994, 2013; Cai 2000; Fan 2005a). This indicates the basic blueprint of floating rural labor migration flow will not change. That is moving from relatively undeveloped landlocked region to developed east coast. Figure 2 shows the GDP per capita of each prefecture level city in China based on four classes quartile in 2000 and 2010.

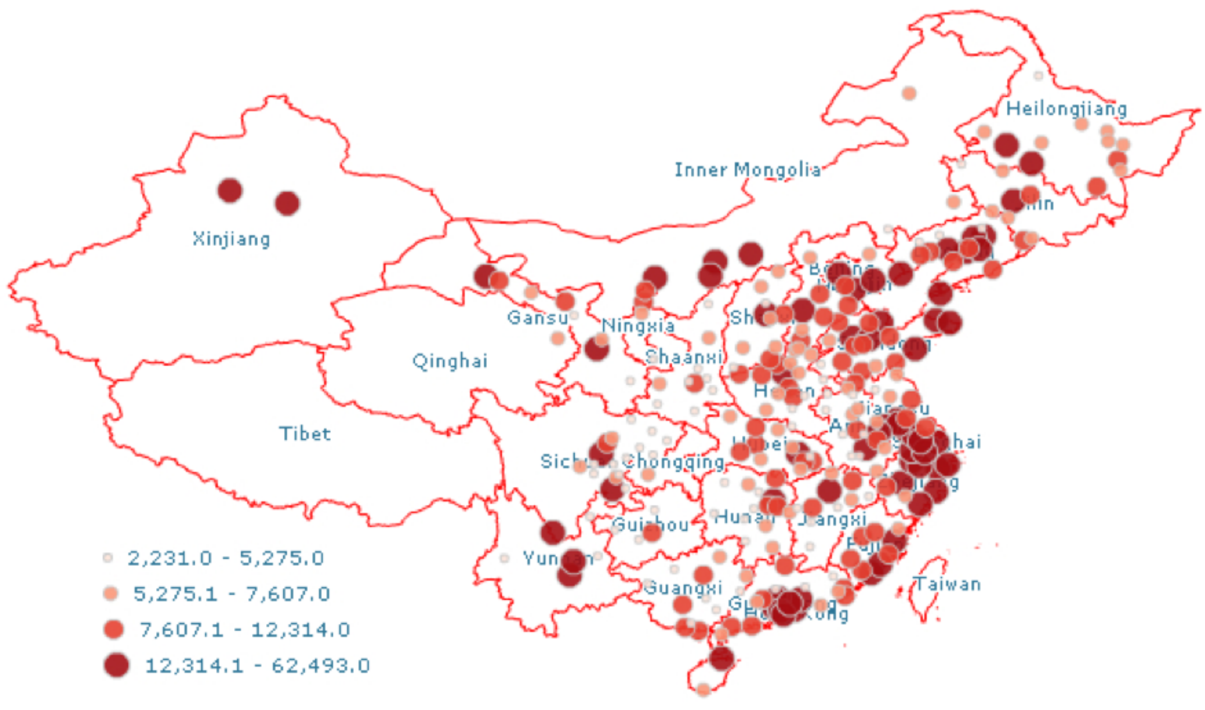

Figure 2(a): *2000*

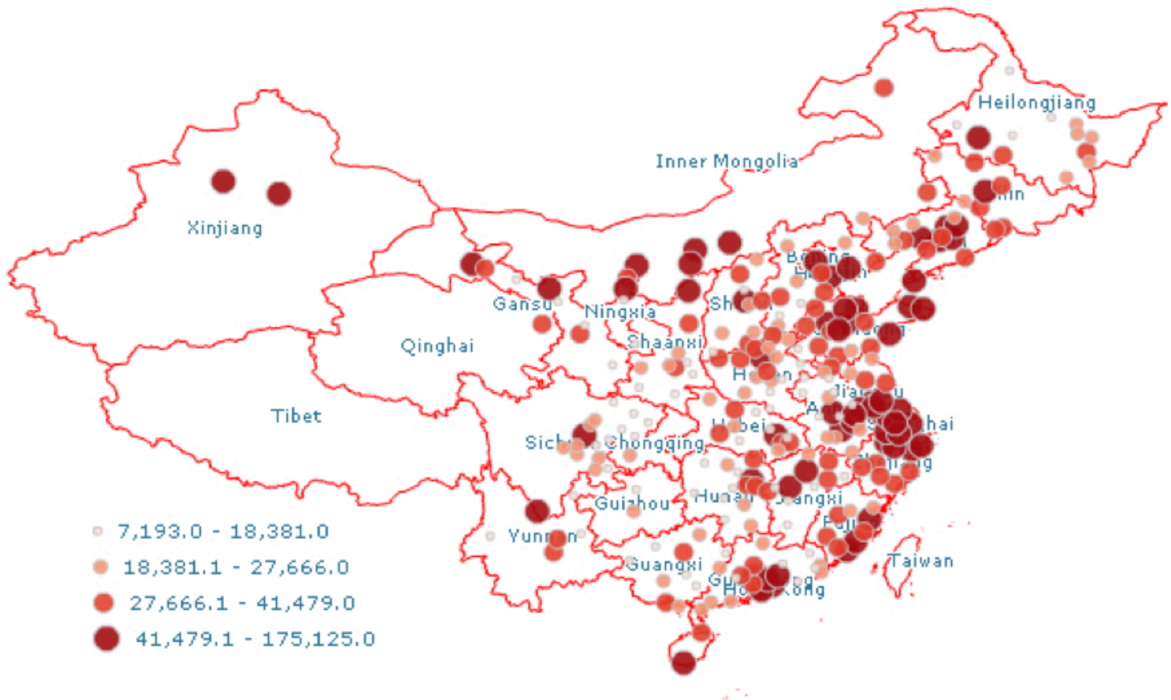

Figure 2(b): *2010*

Figure 2: Prefecture Level Cities' GDP Per Capita in China 2000 and 2010 (yuan/person)

Sources: China Statistical Yearbook (2000: $1=￥8.28 2010: $1 =￥6.83)

Comparing these two years, we can conclude that the overall arrangement did not alter with most of the biggest and darkest circles aggregated along east coast. While taking a closer look, we can find out that more big and dark circles aggregated in Beijing-Tianjin-Hebei Region and Yangtze River Delta Region in 2010 than 2000. Yet the number of the biggest circle shank in Pearl River Delta in 2010. In addition, the number of circles in the mid-western area built up gradually.

Another aspect we should look at, which respects the urbanization level of each city (Alonso, 1971), is population. The circles layout (see Figure 3) has not modified a lot during 2000-2010.

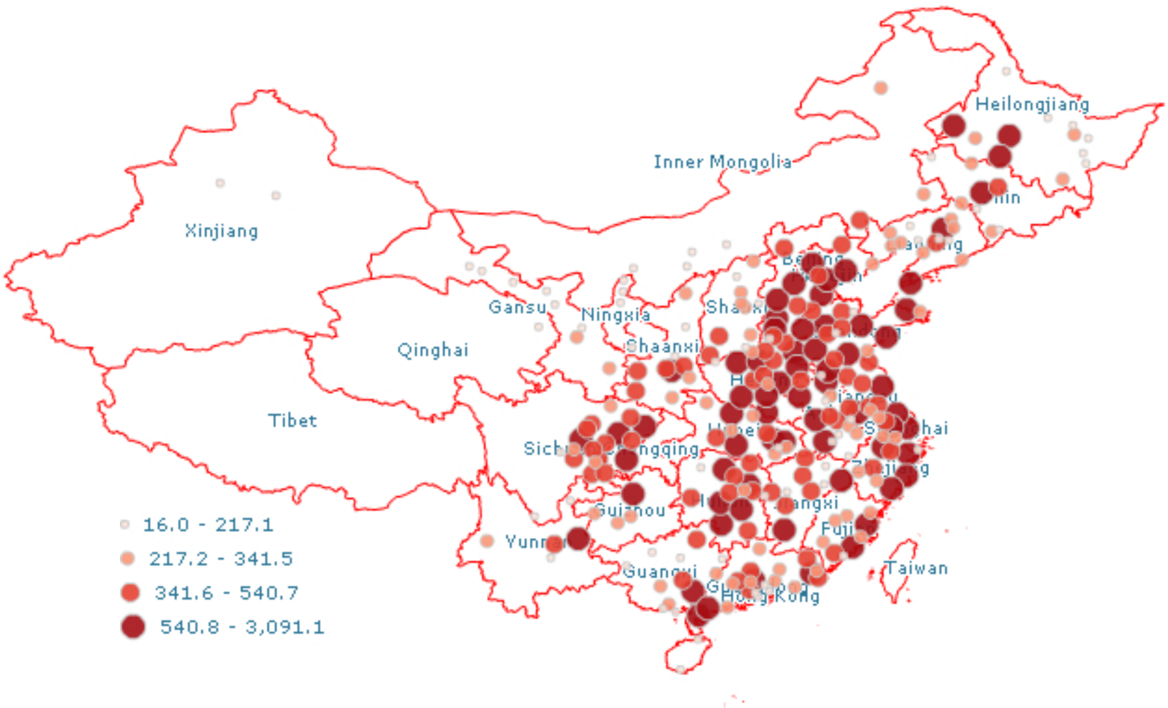

Figure 3(a):*2000*

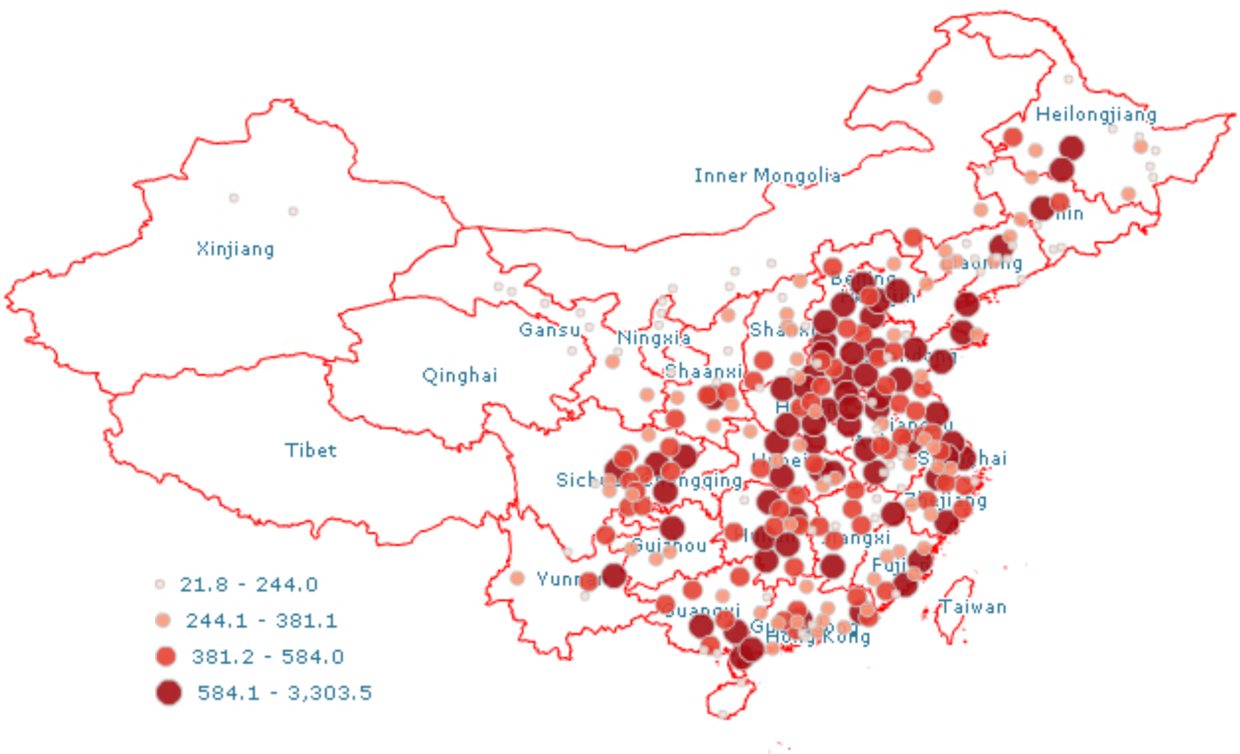

Figure 3(b): 2010

Figure 3: Prefecture Level Cities' Population in China 2000 and 2010 (10000 persons)

Source: China Statistical Yearbook

Darkest circles number augmented in mid-western area in 2010, especially along the Yangtze River area like Anhui and Henan. South East area like Guangdong also had more circles in 2010 than 2000.

This change explains the phenomenon found in the absorbing capacity ranking change of immigrated cities, on the basis of 2000 and 2010 National Population Census (see Appendix Table 3).

First, the ability of East coastal cities like Shenzhen, Dongguan and Foshan for absorbing floating migration receded in 2010, comparing to 2000. The proportion of Shenzhen to country's floating migration decreased to 5.02% in 2010 from 7.85% in 2000. Ranking slipped from No.1 in 2000 to No.3 in 2010. While the proportion of Dongguan to China's floating migration dropped from 6.61% to 3.87% in these ten years. Other top 10 ranking east coastal cities like Foshan, Wenzhou and Quanzhou fell from No.6, 7, 8 to No. 9, 10, 15. However, rankings of cities in the Yangtze River Delta rose in 2010. Suzhou, with 1.55% proportion of the whole country's floating migration in 2000, increased sharply to 2.74% in 2010. Its ranking rose from No.10 to No.6. Shanghai ranked No.3 in 2000 and went up to Top 1 in 2010. Hefei[5] rose its ranking from No.44 in 2000 No.29 in 2010, with proportion increasing from 0.46% to 0.9%. Consistent with previous finding, Yangtze

[5] The capital city of Anhui Province.

River Delta has replaced Pearl River Delta, becoming the main region that absorbs interprovincial floating immigration. (Wang and Pan, 2012).

Second, the capability of the inland region for attracting migrants was enhanced during these 10 years. Especially those capital cities like Chengdu, Xi'an and Chongqing, increased rapidly in rankings as well as proportion scale. From Table 3, Chengdu, Xi'an and Chongqing's floating migration scale increased 2.7, 2.1 and 2.2 times separately, with ranking changing from No.9 to No.7, No.34 to No.20 and No.17 to No.12.

In the next part, I would use the top 10 immigrated cities of 2000 and 2010 in the Appendix Table 3 (total will be 11 cities) as my study cities and describe the characters of determinants used in the model.

3.2 data description

For better understanding the effect of inter-provincial floating immigration, this study is using fourteen years' panel data from 1999-2013. The determinants of the simultaneous-equations model discussed in Section 2 are collected separately from 11 immigrated cities' City Statistical Yearbook, China data online and National Population Census. The reason for me to use the top 10 immigrated cities of 2000 and 2010 from Table 1 as my immigrated regions *i* in the dataset is because that first, the proportion of these 10 cities has already occupied nearly 40% of the total floating migration in China. Second and most importantly, regarding to the imperfection of China's database, many immigrated cities don't have enough data to cover fourteen years' estimation. This is a trade off between accuracy of dataset and size of dataset. Former one is more significant in this study. In order to decide the emigrated region *j* for each immigrated city *i,* I use the data of migration with *hukou* registered in other provinces (inter-provincial floating migration) in 2000 and 2010 National Population Census. After calculating the proportion of each province's migration to total immigration of the immigrated city *i*, top 5 provinces ranked by their proportion weights as the emigrated region *j* for *i* will be aggregated. Thus, the explanatory variables of emigrated region *j* are also aggregated variables with the proportion weights calculated before. This method is feasible as each top 5 emigrated provinces for each immigrated city *i* occupied over 50% and even up to 70% of total inter-provincial immigration (see Appendix Table 4).

The definition and key information of each determinant is displayed in Table 5. Especially, for some determinants, there are no direct data that can be drawn from the database. This paper uses some indirect methods to calculate them. For instance, the total labor in *i* at time *t* is the product of employment rate in *i* at time *t* times the summation of registered population and inter-provincial floating immigration at time *t*. While the employment rate in *i* at time *t* is computed by 1 minus the value of total registered unemployed population divided by registered employed population in *i* at time *t*. The proxy of high school education rate in *i* at *t* is the result of total number of unretired labor with high school education divided by total registered employed population in *i* at time *t*. In China, the legal retiring age and the legal working age is 60 and 18. The total number of unretired labor with high school education is calculating by using the recorded yearly high school registered student population data from City Statistical Yearbook. Let's take year 2000 as an example, the earliest year for counting the total number of unretired labor with high school education should be 1959 (= 2000 – 60 + 18). So, the time period for counting the total number of unretired labor with high school education is 1959-2000. However, this method can not get ride of the immigrated student with *hukou* registered in other places and other students who may not work in this city after graduation. Besides that, it is the most accurate method based on reachable data sources. Moreover, according to the three industries division rule (GB/T 4754—2011) from China National Bureau, service industry can be divided into 18 categories. But not all of them are feasible in explaining the increase in floating immigration in city i. Some categories like finance service and information transportation service will not provide jobs for low-skilled workers like floating immigration. Hence, in this paper, just two categories' output (wholesale and retail; transportation and postal service) in service industry are collected as the total service industry output in city i.

Apart from that, due to the usage of logarithm, some determinants' ranges are between negative values like $In\frac{P_{it-1}}{P_{jt-1}}$, $InSM_{it-1}$, $InFG_{it-1}$ and InE_{it-1} . In the next section, when explaining the effects of each determinant, the signs of them need to be taken into consideration as well.

Table 5: Key information of determinants
*Notes: i---Immigrated region (prefecture-level city), j--- Emigrated region (aggregated), n --- Groups of variables, T --- time period, N --- number of observation

Variables	Definition	Observation	Mean (overall)	Std. Dev (overall)	Min (overall)	Max (overall)
InM_{jit}	Inter-provincial floating migration in i *from j* (logarithm)	N = 165 T = 15 n = 11	14.5475	.9708912	11.7736	16.10806
$In\frac{P_{it-1}}{P_{jt-1}}$	One time lagged of total *hukou* registered population in i divided by total	N = 154 T = 14 n = 11	-2.129272	.8789759	-3.639708	-.7216428

	hokou registered population in *j* (logarithm)					
$InSM_{it-1}$	One time lagged of service industry output devided by manufacture industry output in *i* (logarithm)	N = 154 T = 14 n = 11	-1.056991	.3618722	-1.782519	-.1293737
$In\frac{E_{it-1}}{E_{jt-1}}$	One time lagged of employment rate in *i* divided by employment rate in *j* (logarithm)	N =154 T = 14 n = 11	.0263442	.0234841	-.0241618	.0685399
$In\frac{MA_{it}}{MA_{it-1}}$	Ratio of manufacture industry ouput at time t divided by at time (t-1) in *i* (logarithm)	N = 151 T = 15 n = 11	.142628	.0837762	-.0694382	.6574305
$In\frac{Y_{it-1}}{Y_{jt-1}}$	One time lagged of average income per person in *i* divided by average income per person in *j* (logarithm)	N =154 T = 14 n = 11	.8417008	.2572709	.3829241	1.407992
InG_{it}	GDP per capita in *i* (logarithm)	N = 165 T = 15 n = 11	10.66669	.6631404	9.22877	12.20601
$InMA_{it-1}$	One time lagged manufacture industry output in *i* (logarithm)	N = 154 T = 14 n = 11	7.382418	.8699081	4.965744	8.968876

InP_{it-1}	One time lagged total *hukou* registered population in i (logarithm)	N = 154 T = 14 n = 11	15.52841	.7792929	13.99658	16.564
InH_{it-1}	One time lagged high school education rate in *i* (logarithm)	N = 154 T = 14 n = 11	.9347209	.8033839	-1.471762	2.311984
$InFG_{it-1}$	One time lagged of total fixed asset divided by GDP in *i* (logarithm)	N = 154 T = 14 n = 11	-.9511028	.3271199	-1.775177	-.2694397
InG_{it-1}	One time lagged GDP per capita in *i* (logarithm)	N = 154 T = 14 n = 11	10.61113	.6443185	9.228769	12.13461
$In(P+M)_{it-1}$	One time lagged the sum of population plus migration in i	N = 154 T = 14 n = 11	15.98622	.5342654	14.57228	16.98538
InE_{it-1}	One time lagged employment rate in *i* (logarithm)	N = 154 T = 14 n = 11	-.0384348	.021146	-.0920971	-.0006511
$InMA_{it}$	Manufacture industry output in *i* (logarithm)	N = 165 T = 15 n = 11	7.446816	.8849912	4.965745	8.990662
InM_{jit-1}	One time lagged total floating migration in *i* from *j* (logarithm)	N = 154 T = 14 n = 11	14.50321	.9743477	11.7736	16.07048

CHAPTER 4

RESULT ANALYSIS

This paper presents result by starting with examining in depth statistical aspects of observed regression result. Follow up with comparing the result with the one in the single equation robust OLS model with fixed effect. Final is to do the regression diagnostics.

4.1 3sls model

According to the regression result (see Table 6), most attributes have the expected effect to the dependent variables as the hypothesis demonstrated above (see Table 2). For the whole simultaneous equations system, the chi-square goodness of fit value of the three equations are at significant level with p = 0.0000, which means all parameters in this system equations are significantly different from 0. Moreover, the R^2 = -1.4452[6], 0.9882, 0.9950 separately for function (4)(5)(6). As a summary measure of the overall in-sample predictive power of

[6] Especially, R^2 in function (4) is negative. This is common for 2sls or 3sls estimation given some of the regressors enter the model as instruments when parameters are estimated. The model residuals are computed over a different set of regressors from those used to fit the model. Thus, 2sls or 3sls no longer nested within a constant-only model of the dependent variable, and the residual sum of squares is no longer constrained to be smaller than the total sum of squares, which means $R^2 \notin [0,1]$. But a negative R^2 doesn't indicate the model doesn't fit well when we compare the associated standard errors and the overall model significance with the OLS estimation result below (Stata Manuals13 rreg3).

estimators, normally higher the R^2 is, better the model fits. It means the system equations model is doing a good job here in fitting the real data.

As for the first modified gravity model equation (function (4)) in the simultaneous equation system, expect $In\frac{P_{it-1}}{P_{jt-1}}$, $InSM_{it-1}$, $In\frac{\boldsymbol{Y_{it-1}}}{\boldsymbol{Y_{jt-1}}}$, all other parameters are significant at 99% with $p < 0.01$. Different from previous prediction in Table 2, $In\frac{P_{it-1}}{P_{jt-1}}$ has a negative range between -3.639 to -0.7216 while has an insignificant estimated coefficient .1343896 with $p = 0.201 > 0.1$ This means $\frac{P_{it-1}}{P_{jt-1}}$ actually plays a negative role in attracting floating immigration from emigrated region *j* to immigrated city *i*. However, with larger P_{it-1} (*hukou* registered population in immigrated city *i*), the negative effect will be smaller. This observation contradicts Greenwood's modified gravity migration model (2005) in which total population in the immigrated city *i* plays a positive role in pulling floating migration to move in. Reasons for this phenomenon can be partially explained by the data source in chapter 3 as well as the special characters of Chinese interprovincial migration. Due to the incompleteness of Chinese data, there's no record of prefectural city level origins of floating immigration in city i. The aggregated provincial level data of emigrated city j will cause P_{jt-1} larger than P_{it-1}. In order to test these guests, another simultaneous-equation system model (using 3sls estimation as well) was run with InP_{it-1} and InP_{jt-1} as two separate parameters[7]. The estimated coefficient for lnP_{jt-1} is -0.0855177 with $p = 0.728 > 0.1$ and the estimated coefficient for InP_{it-1} is -

[7] This model just modifies the first equation of immigration as $InM_{jit} = \beta_{10} + \beta_{11}\, InP_{it-1} + \beta_{12} \ln P_{jt-1} + \beta_{13}\, InSM_{it-1} + \beta_{14}\, In\frac{E_{it-1}}{E_{jt-1}} + \beta_{15}\, In\frac{Y_{it-1}}{Y_{jt-1}} + \beta_{16}\, In\frac{MA_{it}}{MA_{it-1}} + \varepsilon_{it-1}$, others remain the same.

0.0131159 with $p = 0.880 > 0.1$. This means P_{it-1} itself also has a negative effect on floating immigration in city *i* while P_{jt-1} has also have a negative effect on floating immigration in city *i* although they are very insignificant. Another possible reason causing this is based on the characters of Chinese interprovincial migration. According to the statistics from the sixth China Population Census, Guangdong Province is the No.1 province with the most population (104.303 million) among all provinces in China. Guangdong is also the one absorbs the most immigration from other provinces in China. While Henan Province, which ranked No.3 (94.024 million) in the provincial total population ranking in China, is the one with the most emigration among all provinces in China. People from Henan will immigrant to Zhejiang Province, Tianjing and Sichuan Province whose rank behind it (see Appendix Table 4). Thus, to some extent, city population can not perfectly explain the reason for people to immigrate in China. $In\frac{E_{it-1}}{E_{jt-1}}$ and $In\frac{MA_{it}}{MA_{it-1}}$ have greater effect on the floating immigration with coefficient equaling 7.262058 and 25.50028 in turn. The *p-value* for them both equal 0.000. This is highly coherent with previous literature. For floating migrant, seeking nonagricultural employment opportunities that offer high wages is the primary reason for leaving their hometowns (Zhu et al., 2001; Lu et al., 2005; Cao et al., 2012; Shen, 2012). This is also, in this study, a significant drive of pulling floating immigration from *j* to *i*. Besides, as mentioned in the previous chapter, manufacturing industry as the pillar industry in China, still plays pivotal role in creating employment opportunities in city economy. Thus, the growth of manufacturing output comparing to last year will draw floating immigration to move in. Still, the disparity of average income between emigrated and immigrated cities $In\frac{\boldsymbol{Y_{it-1}}}{\boldsymbol{Y_{jt-1}}}$ implements a pulling power to draw migration moving from a relatively low average income region *j* to high average income city *i,* just as the hypothesis supposes.

For the second equation (function (5)), InM_{jit-1} , $InMA_{it-1}$, $InFG_{it-1}$ are significant in explaining the economy development in immigrated city i. Their coefficients are 0.1736701, 0.3930946, 0.0627316 accordingly. Consistent with the hypothesis in this study, the increase in floating immigration will stimulate development in immigrated city's economy since it guarantees rich and cheap labor supply. Manufacturing industry as the pillar industry in most Chinese cities, continuously promotes local economy development. Fixed asset investment, which theoretically assists in improving the labor productivity and increase enterprise's demand for labor, is proven to be supportive in promoting city's economy development in this study. Yet, contradict to previous assumption, InH_{it-1} is highly insignificant in demonstrating the cause of city i's economy development with $p = 0.240 > 0.1$ and InP_{it-1}, with estimated coefficient equaling -0.0517915, becomes an impeditive factor in the economy development of city i. Just as previous chapter illustrates, the collection of high school education rate has some imperfection that cannot be avoided. This might account for the insignificance of InH_{it-1}. For InP_{it-1}, in order to better test its effect on GDP per capital, this paper constructs a single equation estimation for equation 2 (function(5)) by 2sls with fixed effect. The regression result is listed in Appendix Table 7. For this analysis, equation 3 (function (6)) is used as instrumental variable with all the explanatory variables are lagged two times[8]. From this result, the coefficient of InP_{it-1} is highly insignificant with $p\text{-}value = 0.485 > 0.1$. While the sign of the coefficient is positive and equals 0.0202515. the parameter for InM_{jit-1} is as prediction and equals 0.6719983 with $p\text{-}value = 0.0000$. One possible explanatory theory is that GDP per capita equals GDP divided by total population, the higher volume of total population, given GDP per capita constant, indicates lower value of GDP

[8] The instrumental variable is: $InMA_{it-1} = \beta_{30} + \beta_{31} InG_{it-2} + \beta_{32} In\ [E * (P+M)]_{it-2} + \beta_{33} InMA_{it-2} + \xi_{it}$

per capita. Also, as we can interpret from the result, manufacturing industry is still the pillar industry in these cities. In the supply labor market, migrants provide sufficient and cheaper labor supply than local *hukou* registered population labor to the manufacturing industry and thus comparatively better in promoting local economy growth (Cai, 2008). Additionally, the constant equals 5.816036 with *p-value* = 0.0000. In section 2 model construction, we know that the constant term indicates the total factor productivity, which means the efficiency of production such as technology input. This suggests that the technology improvement has a significantly positive effect on the city's economy development as plenty of papers have pointed out.

Regarding to the third equation (function (6)), InG_{it-1} , $InMA_{it-1}$, $In(P+M)_{it-1}$ are all showing positive effect in triggering manufacturing industry output increase in highly significance level. Expect that $In(P+M)_{it-1}$ is at 95% significance level, InG_{it-1} and $InMA_{it-1}$ are at 99% significance level. [9] While coefficient of InE_{it-1} is insignificant with negative sign. [10] Among all the explanatory variables, manufacturing output at city *i* in *t-1* plays the most important role in explaining the manufacturing output improvement with coefficient 0.9362445. Cause of this phenomenon can be demonstrated by the structure model. The manufacturing output in city *i* is the main drive of attracting floating immigration to city *i*. Then these immigrations enter labor market of city *i* and become important labor supply to promote growth of GDP per capita in city *i*. The GDP per capita in city *i* from last period will then work as accumulated capital input, together with manufacturing output in city *i* from last period *t-1*, motivate the yield of manufacturing industry. Reversely, without a high demand of labor in the manufacturing

[9] To remind, as previous notation, *G* here represents GDP per capita and labor is represented by $E*(P+M)$, where *E* stands for employment rate and (*P+M*) equals the summation of population and floating immigration
[10] To remind, the range of *lemi1* is [-0.0920971, -0.0006511].

industry due to the increasingly growing output of manufacturing, immigration will not be attracted to city *i* and then will not provide labor force to promote city *i*'s economy development, which in turn impedes the accumulation of capital in manufacturing production.

Table 6: 3sls estimation result with endogenous variables list (Standard errors in parentheses)
*** Significance at 1% level
** Significance at 5% level
* Significance at 10% level

	3SLS Estimation		
	Function (4)	Function (5)	Function (6)
$In\frac{P_{it-1}}{P_{jt-1}}$	.1343896 (.1051859)		
$InSM_{it-1}$	-.1049024 (.1109548)		
$In\frac{E_{it-1}}{E_{jt-1}}$	7.262058*** (2.617267)		
$In\frac{Y_{it-1}}{Y_{jt-1}}$	.3779822* (.2307419)		
$In\frac{MA_{it}}{MA_{it-1}}$	25.50028*** (.3140125)		
_cons	11.34481*** (.6852645)		
InM_{jit-1}		.1736701*** (.0136405)	
$InMA_{it-1}$		.3930946*** (.037041)	
InP_{it-1}		-.0517915** (.0212817)	
InH_{it-1}		.0209049 (.0178074)	
$InFG_{it-1}$		.0627316** (.0252724)	
_cons		5.816036*** (.4908795)	
InG_{it-1}			.1259668*** (.0115937)

InE_{it-1}			-.149906 (.1131136)
$In(P+M)_{it-1}$			.0103025** (.004891)
$InMA_{it-1}$			.9362445*** (.0088418)
_cons			-.8938938*** (.1299621)
R-square	-1.4452	0.9863	0.9956
N	154	154	154
chi^2	6824.16	11328.88	53116.86
P-value for chi^2	0.0000	0.0000	0.0000

Endogenous variables list: InM_{jit} , $InMA_{it}$, InG_{it}

4.2 Robust OLS model with fixed effect

For comparison, this paper also uses robust OLS with fixed effect estimation to estimate the three equations by single equation method (see Appendix Table 8). General speaking, OLS estimation is barely satisfactory under this context. Just as presented in previous content, using OLS model to estimate system equations will cause inconsistency and econometrics bias in error terms. But for the purpose of enriching the aspects of estimation, this paper decides to include the result as well. Before the estimation part, we need to test the over identification restriction for each single equation in the system. There're two conditions to decide whether we can determine the unknown parameters in a reduced form. A necessary but not sufficient condition for the testing the parameter to be identifiable is that the number of unknown parameters (endogenous variables) is not greater $K+1$ (where K is the number of exogenous variables). This is also called order condition. If $K+1$ = number of endogenous variables, then the equation is just identified. Otherwise, $K+1$ > number of endogenous variables means the equation is over identified. In this case, the system equations are over identified since the number of endogenous variables is 3 and $K = 13$. Another rank condition will not be necessary to test in this case. When equations are over identified, the estimation problem becomes complicated. Three kinds of procedure have been proposed: (1) full system methods (maximum likelihood estimator), (2) single equation methods and (3) subsystem methods. Among all these asymptotically equivalent classes of estimators corresponding to different information structures it has been established that the maximum likelihood estimators have asymptotically higher-order efficiency than other estimators, and Monte Carlo and numerical studies show that they are in most cases better than others if properly adjusted for the biases (Itô, 1993). From the result in Table 8, regardless of other estimators, InM_{jit-1} is still playing a highly significant role in promoting the city *i*'s economy development with coefficient equals 0.137879 at 99% significance level.

4.3 Heteroskedasticity test and FGLS[11] model

For increasing the credibility of analysis, test of overall system heteroskedasticity is implemented here. Results are listed in Appendix Table 9. From the results we can know that, there are heteroskedasticity in three equations since the null hypothesis with homoscedasticity is rejected in Breusch-Pagan LM Test, Likelihood Ratio LR Test and Wald Test with p-value > chi^2 at 0.0000. However, in the previous paper and study, there's no pretty satisfactory solution for this question (Biorn *et al*, 2003 2004). Either using SUR (Seemingly Unrelated Regression) to estimate the

[11] FGLS stands for Feasible Generalized Least Squares

system equations or using Eicker-Huber-White-sandwich covariance estimator in a set of equations has its defect. While the limitation for the former one is the assumption of homoscedasticity in error terms and for the latter one is the lack of efficiency by omitting the cross-equation correlation estimation, which is the same as using GLS (Generalized Least Squares) and FGLS. Also, no package in STATA could be used to apply robust effect in 3sls estimator. Thus, there's always a tradeoff between correcting heterogeneity and estimators' efficiency. For comparison, this paper also includes the FGLS estimators result for solving the heteroskedasticity problem with assumption that there is heteroskedasticity on panel dataset in Appendix Table 10. In this case, we sacrifice the efficiency of estimators for correcting the heteroskedasticity problem by applying FGLS single equation estimation. From Table 9, we can see that the effects of all parameters are mostly similar to the result we get from 3sls estimation.

Selection bias

Given the data feasibility in China, this paper picks top 11 immigrated cities in 2000 and 2010 as study object (not each city's City Statistical Yearbook has complete 14 years' data resource covering all explanatory variables). The process of picking immigrated city may generate sample selection bias since it excludes other non immigrated cities, which means there may be omitted variables due to self-selection. In this case, only immigrated cities with $InM_{jit} > 0$ will be selected. However, the Heckman fits regression model with selection for correcting selection bias is not feasible here. Just as 2sls estimation, the application of Heckman model is using either Heckman's two-step consistent estimator or full maximum likelihood which is running the equation one by one and ignoring the correlation among system equations' errors terms (Stata manual 13 heckman).

.

CHAPTER 5

CONCLUSION

This study ends by summarizing the key findings, some of which are coherent with previous literatures, and others that implement the gap in studying Chinese interprovincial migration. One observation is about the attraction in the immigrated city that pull floating immigration to come. Different from previous study about migration, population of the immigrated city doesn't play a keen role in drawing migrants. The possible reason for that is due to the special characters of China's migration as pointing out in the previous section. Instead, employment rate in immigrated city is important in providing attraction for migrants. While the disparity of average income between regions and growth of manufacturing industry output, as previous study points out (Cindy, 2005), play significant roles in drawing floating migration to move from their hometown to the cities. To be specific, one Additionally, despite the reformation of industry structure has taken place with higher and higher proportion of service industry in the whole system, no significant finding in the results that service industry has replaced manufacturing industry in attracting migration.

The other observation is about the impact of migrants to the immigrated city. It is an important fact that floating immigration do have positive effect on immigrated cities' economy development, which has also been found out by previous scholars (Fang and Dewen, 2008; Cai and *et al*., 2002). Another finding about local population playing negative role in GDP per capital growth is conflicting to previous study. One possible explanation is that the large amount and low-wage migrants fits the demand of manufacturing industry labor market. They are more competitive in the manufacturing industry labor market than the city *i*'s *hukou* registered population. Moreover, manufacturing industry is the main contributor to city's economy development, which is proved in the result of system equations in this study. While the *hukou* registered population will consume the total GDP and diminish the GDP per capital since it serves as a denominator of total GDP. However, attention should be given to the constraints in this conclusion. Given previous study (Fang and Dewen, 2008; Cai and *et al*., 2002) and data from China Statistical Yearbook, manufacturing industry still is and will remain the pillar industry in China economy structure in the near future. This conclusion is only applicable to the manufacturing-domain cities, which is the case in this study. Based on the findings in the model result, suggestions can be made to the authorities that if they want to promote the local economy development and accelerate the industry development, abolishing impeditive policies to the entry of migrants and fastening the *hukou* system reformation is necessary and essential.

Last but not the least, limitation of this study should also be pointed out. The selection bias caused by the imperfection of China's dataset and the heteroskedasticity in the residuals will impact the efficiency of the estimators. To some extent, it weakens the power of interpretation in this study. However, when we comparing the results from 3sls estimation and FGLS estimation, we can obtain similar outcomes (with a little difference in coefficients' values but no difference in the effects of signs), which in turn, inspires readers' confidence in the analysis of this study. As last section has discussed, there's no perfect solution about this in current econometrics study. Every method has its defects and merits. Consequently, It's reasonable to interpret the results under the constraints of limitation of the methodology, which is exactly the case in this study.

Based on current findings, for future study, I would like to prefect the system equations by adding lagged dependent variables on the right hand side of each equation as determines to see the dynamic result of the system. In order to avoid overweight emphasis of the one time lagged dependent variables in explaining the dependent variable in each equation, I would like to apply partial adjustment before each lagged dependent variable. What's more, I would like to include more cities in the dataset and shorten the time period to see whether there will be difference in result. This will be a supplementary method to solve the selection bias. Besides, since this paper has analyzed the macro aspect of floating migration, I would like to understand the micro aspect of migration like the decision making process for migrants and the reaction of the migrants-receiving region by using dynamic model.

APPENDIX

Table 3: Top 50 immigrated cities and its proportion over national floating migration in 2000 and 2010, based on National Population Census

	2000			2010		
Ranking of city	city	floating migration (10 thousand person)	propotion %	city	floating migration (10 thousand person)	propotio n %
1	Shenzhen	585	7.85	Shanghai	1102	6.68
2	Dongguan	492	6.61	Beijing	892	5.41
3	Shanghai	436	5.85	Shenzhen	828	5.02
4	Guangzhou	331	4.45	Dongguan	639	3.87
5	Beijing	260	3.49	Guangzhou	542	3.29
6	Foshan	221	2.96	Suzhou	452	2.74
7	Wenzhou	135	1.82	Chengdu	387	2.35
8	Quanzhou	129	1.74	Tianjing	386	2.34
9	Chengdu	126	1.69	Foshan	358	2.17
10	Suzhou	115	1.55	Wenzhou	324	1.96
11	Kunming	113	1.52	Hangzhou	285	1.73
12	Zhongshan	104	1.40	Chongqing	280	1.70
13	Wuhan	98	1.32	Wuhan	277	1.68
14	Hangzhou	96	1.29	Ningbo	259	1.57
15	Ningbo	94	1.26	Quanzhou	241	1.46
16	Huizhou	91	1.23	Nanjing	225	1.36
17	Chongqing	88	1.19	Wuxi	214	1.30
18	Wuxi	86	1.15	Zhengzhou	200	1.21
19	Fuzhou	83	1.12	Xiamen	199	1.21
20	Nanjing	80	1.07	Xi'an	180	1.09
21	Tianjin	79	1.06	Qingdao	179	1.09
22	Xiamen	75	1.00	Fuzhou	176	1.07
23	Dalian	63	0.85	Shenyang	169	1.02
24	Zhengzhou	62	0.83	Huizhou	167	1.01
25	Urumchi	60	0.81	Kunming	167	1.01
26	Changzhou	60	0.80	Zhongshan	165	1.00
27	Guiyang	58	0.78	Dalian	165	1.00
28	Zhuhai	58	0.78	Jinhua	148	0.90
29	Jinhua	57	0.77	Hefei	148	0.90
30	Taizhou	55	0.73	Changzhou	143	0.87
31	Qingdao	54	0.73	Taizhou	143	0.86
32	Shenyang	52	0.70	Changsha	142	0.86

33	Changsha	52	0.70	Urumchi	131	0.79
34	Xi'an	49	0.66	Jiaxing	130	0.79
35	Harbin	49	0.66	Nanning	128	0.78
36	Shijiazhuang	47	0.63	Jinan	123	0.74
37	Jiangmen	46	0.62	Guiyang	118	0.71
38	Nanning	46	0.62	Harbin	113	0.69
39	Liuzhou	40	0.54	Taiyuan	110	0.67
40	Taiyuan	38	0.51	Hohhot	110	0.67
41	Changchun	36	0.48	Shaoxing	110	0.67
42	Hohhot	36	0.48	Changchun	97	0.59
43	Yantai	34	0.46	Nanchang	92	0.56
44	Hefei	34	0.46	Shijiazhuang	91	0.55
45	Baotou	33	0.44	Baotou	80	0.48
46	Shaoxing	33	0.44	Lanzhou	78	0.47
47	Baoding	33	0.44	Yantai	78	0.47
48	Nanchang	33	0.44	Jiangmen	76	0.46
49	Jiaxing	32	0.43	Erdos	76	0.46
50	Lanzhou	32	0.43	Haikou	74	0.45

Table 4: Decomposing aggregated emigrated region *j* for each city *i* by top 5 ranking of proportion over total migration, based on National Population Census in 2000 and 2010

	2000		2010	
City *i*	Emigrated region *j* (province)	Proportion %	Emigrated region *j* (province)	Proportion %
Dongguan				
	Hunan	18.4752	Hunan	23.2076
	Sichuan	15.6242	Sichuan	15.6237
	Guangxi	11.7986	Guangxi	11.3126
	Hubei	9.63	Hubei	10.3675
	Jiangxi	7.7045	Jiangxi	10.0426
Foshan				
	Guangxi	25.7101	Guangxi	30.5386
	Hunan	18.6488	Hunan	18.6014947
	Sichuan	18.5754	Sichuan	15.9211
	Hubei	6.3402	Hubei	6.6746
	Jiangxi	5.7678	Jiangxi	6.4519
Guangzhou				
	Hunan	29.3212	Heilongjiang	26.5423
	Sichuan	17.5939	Hebei	13.554
	Guangxi	12.4251	Jiangsu	11.7578
	Jiangxi	10.9295	Shanxi	10.0910
	Hubei	8.9839	Hainan	8.8502
Zhongshan				
	Guangxi	24.1676	Guangxi	26.5439
	Sichuan	22.3809	Hunan	18.8292
	Hunan	18.7201	Sichuan	15.5855
	Jiangxi	8.8989	Hubei	8.1943
	Hubei	7.3965	Jiangxi	6.8276
Shenzhen				
	Hunan	19.8013	Hunan	20.3549
	Sichuan	18.7958	Hubei	14.0103
	Hubei	12.2605	Guangxi	11.4717
	Jiangxi	11.3113	Sichuan	11.1441
	Guangxi	9.6902	jiangxi	9.7631
Shanghai				
	Anhui	32.8080	Anhui	26.8395
	Jiangsu	23.9187	Jiangsu	18.3673
	Zhejiang	9.9773	Henan	8.3224
	Sichuan	7.3399	Sichuan	6.4998
	Jiangxi	6.0725	Zhejiang	5.6598
Beijing				
	Hebei	22.5321	Hebei	22.1308
	Henan	13.5840	Henan	13.9078
	Anhui	9.2721	Anhui	8.4851

	Shandong	7.6766	Shandong	6.1053
	Sichuan	6.8085	Sichuan	5.7248
Suzhou				
	Anhui	46.6255	Anhui	31.9453
	Sichuan	10.7960	Henan	18.4021
	Zhejiang	10.0248	Sichuan	7.3901
	Henan	8.6462	Hubei	6.7205
	Hubei	4.3448	Shandong	6.0740
Chengdu				
	Chongqing	32.2681	Chongqing	30.2995
	Zhejiang	11.2497	Hubei	6.1814
	Hubei	5.4471	Henan	5.3191
	Guangdong	4.1266	Zhejiang	5.1288
	Xinjiang	3.6686	Hunan	4.1653
Wenzhou				
	Jiangxi	27.0965	Jiangxi	19.3505
	Anhui	15.4288	Anhui	15.5007
	Sichuan	14.7180	Hubei	13.9851
	Hubei	11.3677	Guizhou	12.1523
	Guizhou	7.6815	Sichuan	8.9357
Tianjing				
	Hebei	27.5749	Hebei	25.1199
	Shandong	16.6017	Shandong	17.5977
	Henan	10.0383	Henan	10.9226
	Anhui	8.0120	Heilongjiang	5.2861
	Heilongjiang	7.1224	Anhui	4.5706

Table 7: Result of fixed effect 2sls estimation for equation 2 (function(5)) with equation 3 (function (6)) as IV (Standard errors in parentheses)
*** Significance at 1% level
** Significance at 5% level
* Significance at 10% level

	2SLS with fixed effect
	Function (5)
InM_{jit-1}	.1376771***
	(.0254993)
$InMA_{it-1}$	.6719983***
	(.021457)
InP_{it-1}	.0202515
	(.0289871)
InH_{it-1}	.0818608
	(.0547241)
$InFG_{it-1}$	.0600149
	(.0447938)
_cons	3.436486***
	(.4878797)
R-square(overall)	0.6690
N of obs	143
Wald chi^2	1.99e+06
Prob > chi^2	0.0000
F test for all u_i = 0	F(10,127) = 89.77
Prob>F	0.0000
corr(u_i, Xb)	-0.4575
sigma_u	.39439273
sigma_e	.09155887
rho	.94886173

Table 8: Result of robust OLS with fixed effect model (Robust standard errors in parentheses)
*** Significance at 1% level
** Significance at 5% level
* Significance at 10% level

	Robust OLS with Fixed Effect		
	Function (4)	Function (5)	Function (6)
$In\frac{P_{it-1}}{P_{jt-1}}$	.2401385 (.202071)		
$InSM_{it-1}$	-.1030857 (.2975583)		
$In\frac{E_{it-1}}{E_{jt-1}}$	10.23825 (7.767014)		
$In\frac{Y_{it-1}}{Y_{jt-1}}$	-1.05998* (.5346859)		
$In\frac{MA_{it}}{MA_{it-1}}$	-.2992297 (.6505993)		
_cons	15.89376*** (.8374729)		
InM_{jit-1}		.137879*** (.0370019)	
$InMA_{it-1}$		.6737367*** (.0526572)	
InP_{it-1}		.0121793 (.0423048)	
InH_{it-1}		.0134882 (.0337404)	
$InFG_{it-1}$		.0853905 (.091738)	
_cons		3.635411*** (.8302594)	
InG_{it-1}			.2601905* (.0959002)
InE_{it-1}			-2.414495*** (.6445526)
$In(P+M)_{it-1}$			.0293115 (.02668)
			.7450166***

InMA$_{it-1}$			(.0785266)
_cons			-1.297273 (.7253376)
R-square(overall)	0.0104	0.7196	0.9877
N of obs	154	154	154
F	F(5,10) = 5.33	F(5,10) = 720.86	F(4,10) = 2476.6
Prob>F	0.0000	0.0000	0.0000
corr (u_i, Xb)	-0.4603	-0.4024	0.3846
sigma_u	.95748638	.3684874	.07510659
sigma_e	.49519263	.09629078	.06965617
rho	.78897022	.93607996	.53759744

Table 9: Overall heteroskedasticity test for simultaneous system equations

Overall System Heteroskedasticity Tests H_0: No overall system heteroskedasticity	
test	*p-value* > $Chi^2(3)$
Breusch-Pagan LM Test = 201.3584	0.0000
Likelihood Ratio LR Test = 637.5827	0.0000
Wald Test = 8883.3326	0.0000

Table 10: Result of FGLS estimators under assumption that there's heteroskedasticity in residuals (Standard errors in parentheses)
*** Significance at 1% level
** Significance at 5% level
* Significance at 10% level

	FGLS estimation		
	Function (4)	Function (5)	Function (6)
$In\frac{P_{it-1}}{P_{jt-1}}$	.0727741		
	(.0439161)		
$InSM_{it-1}$	-.3824828		
	(.0726009)		
$In\frac{E_{it-1}}{E_{jt-1}}$	26.89391***		
	(1.934314)		
$In\frac{Y_{it-1}}{Y_{jt-1}}$	.031335*		
	(.1012334)		
$In\frac{MA_{it}}{MA_{it-1}}$	-.6955011*		
	(.3402426)		
_cons	14.73611***		
	(.1509404)		
InM_{jit-1}		.070646***	
		(.0173823)	
$InMA_{it-1}$		.7009976***	
		(.0185901)	
InP_{it-1}		-.2305573***	
		(.0203036)	
InH_{it-1}		-.1445443***	
		(.0140347)	
$InFG_{it-1}$		.1453028***	
		(.0351732)	
_cons		8.251809***	
		(.3871136)	
InG_{it-1}			.0673489***
			(.0216856)
InE_{it-1}			-1.341788***
			(.2516745)
$In(P+M)_{it-1}$			.0290971*
			(.0169998)
			.9176304***

$lnMA_{it-1}$			(.0206918)
			-.4901022
_cons			(.3327087)
N of obs	154	154	154
Wald $Chi^2(4)$	359.33	4008.37	32447.85
Prob>Chi^2	0.0000	0.0000	0.0000
Log likelihood	-135.5455	56.07888	205.5292

BIBLOGRAPHY

Au, C. C., & Henderson, J. V. (2006). How migration restrictions limit agglomeration and productivity in China. Journal of development economics,80(2), 350-388.

Alonso, W. (1971). THE ECONOMICS OF URBAN SIZE†. *Papers in Regional Science*, *26*(1), 67-83.

Biorn (2004) "Regression system for unbalanced panel data: a stepwise maximum likelihood procedure." Journal of Econometrics 122: 28.

Bosker M, Brakman S, Garretsen H et al., (2012). Relaxing hukou: Increased labor mobility and China's economic geography. Journal of Urban Economics, 72(2/3): 252–266

Cai, F., Wang, D., & Du, Y. (2002). Regional disparity and economic growth in China: The impact of labor market distortions. *China Economic Review*,*13*(2), 197-212.

C. Cindy Fan (2005) Interprovincial Migration, Population Redistribution and Regional Development in China: 1990 and 2000 Census Comparisons, The Professional Geographer, 57:2, 295-311.

Cushing, B., & Poot, J. (2003). Crossing boundaries and borders: Regional science advances in migration modelling. Papers in regional science, 83(1), 317-338.

De Brauw, A., Huang, J., Rozelle, S., Zhang, L., & Zhang, Y. (2002). The evolution of China's rural labor markets during the reforms. *Journal of Comparative Economics*, *30*(2), 329-353.

Fang, C., & Dewen, W. (2008). Impacts of internal migration on economic growth and urban development in China. In IOM is committed to the principle that humane and orderly migration benefits migrants and society. As an intergovernmental organization, IOM acts with its partners in the international community to: assist in meeting the operational challenges of migration; advance understanding of migration issues; encourage social and economic development through migration; and uphold the human dignity and well-being of migrants. (p. 245).

Fu Y M, Gabriel S A, (2012). Labor migration, human capital agglomeration and regional development in China. Regional Science and Urban Economics, 42(3): 473–484
Greenwood, M. J. (2005). Modeling migration. Encyclopedia of social measurement, 2, 725-734.
Chan, K. W. (2013). China: internal migration. The encyclopedia of global human migration.

Grunfeld, Y. (1961). The Interpretation of Cross Section Estimates in a Dynamic Model *Econometrica, 29*(3), 397-404.

Greenwood, M. J. (1975). A simultaneous-equations model of urban growth and migration. *Journal of the American Statistical Association*, *70*(352), 797-810.

Itô, K. (Ed.). (1993). *Encyclopedic dictionary of mathematics* (Vol. 1), 510-512.

Jones, C. (1998). Introduction to Economic Growth 2nd Edition.

Lu C, Wu Y Z, Shen Q P et al., (2013). Driving force of urban growth and regional planning: A case study of China's Guangdong Province. Habitat International, 40: 35–41.

Liu, T., Qi, Y., Cao, G., & Liu, H. (2015). Spatial patterns, driving forces, and urbanization effects of China's internal migration: County-level analysis based on the 2000 and 2010 censuses. *Journal of Geographical Sciences*,*25*(2), 236-256.

Shen, J. (2016). Modelling Interregional Migration in China in 2005–2010: the Roles of Regional Attributes and Spatial Interaction Effects in Modelling Error. *Population, Space and Place*.

Wang, D., Yaowu, W., & Fang, C. (2004). Migration, unemployment and urban labor market segregation in China's economic transition. *The World Economy*, *4*, 40-43.

Xia Yi Ran, Su Jinhong, & Michael (2015) where migrants gather to? --- City inflow characteristics and change trend. *population and economy*. (3), 13-22.

Zhang, Zhe (2004). Research on the issues of current Chinese farmers with losing land. http://www.eco.sdu.edu.cn/.

Zhang, K. H., & Shunfeng, S. (2003). Rural–urban migration and urbanization in China: Evidence from time-series and cross-section analyses. China Economic Review, 14(4), 386-400.