THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF ECONOMICS

GEOGRAPHIC HETEROGENEITY OF RETURNS TO EDUCATION IN CHINA

LIYIN JIANG

SPRING 2014

A thesis submitted in fulfillment
of the requirements
for baccalaureate degrees
in Economics and Mathematics
with honors in Economics

Reviewed and approved* by the following:
David Shapiro
Professor of Economics
Thesis Supervisor

Russell Chuderwicz
Senior Lecturer of Economics
Honors Adviser

*Signatures are on file in the Schreyer Honors College

ABSTRACT

In this thesis, I look at the geographic heterogeneity of returns to education using data from the Chinese Household Income Project (CHIP) in 1995 and 2002. I begin with an earning equation with heterogeneous returns to education, based on Mincer's work (1974). However, we confront sample selection and endogeneity where we need Heckman's selection model (1979) and an instrumental variables (IV) method to solve. For the IV method, I use regional level (ln) average income, (ln) unemployment rate and (ln) education expenditure as the instruments, test the endogeneity of education by a Durbin-Wu-Hausman (DWH) test, and then estimate returns to education. After comparing returns to education from using different models, I demonstrate the spatial autocorrelation (SAR) model and estimate the correlation coefficient of returns to education in different regions, using results from previous estimations. I find that the heterogeneity was stationary in 1995 but it displayed concavity in 2002 because dependence coefficients increase as educational level increase. This leads to the conclusion that higher level of education will reduce heterogeneity of returns to education, under the assumption that people are capable of searching for better paid jobs, which will be shown in the thesis.

ACKNOWLEDGEMENTS

I am very grateful for guidance and support from Professor David Shapiro as my thesis supervisor during my progress on this thesis. In addition, I thank Professor Sung Jae Jun and Professor Joris Pinkse for providing sophisticated advice on econometric methods. Discussions with them improved the quality of this thesis. In addition, the Business Librarian Kevin Harwell at the William and Joan Schreyer Business Library and the Social Science Librarian Stephen Woods at the Social Science Library gave me useful suggestions and made significant contributions to data collection. I would also like to thank Yuxin Chen at Stanford University, Chengwei Ge at Rice University, Chong Han at the Pennsylvania State University and Jiajie Qiu at Fudan University for their insightful comments on my thesis. This thesis would not be realized without their ongoing effort.

Contents

List of Figures

List of Tables

Symbols

$\phi(\cdot)$	Standard Normal Density Function
$\Phi(\cdot)$	Standard Normal Distribution Cumulative Distribution Function
$\ln(\cdot)$	Natural Logarithm
$(\cdot)'$	Matrix Transpose
Ω	Sample Space
Θ	Parameter Space
λ	Inverse Mill's Ratio
$\mathbb{1}[\cdot]$	Indicator Function
$\mathbb{E}(\cdot)$	Expectation Operator
$\mathbb{I}$	Identity Matrix
$\mathbb{C}$	Binary Contiguity Matrix
$\mathbb{A}$	First-Order Contiguity Matrix
$\mathbb{D}$	Contiguity Matrix with Specification of Distance Effect

Chapter 1

Introduction

1.1 Human Capital and Returns to Education

Human capital comes in various forms in our life. Generally speaking, people can invest in schooling, on-the-job training, health care and migration, and expect to receive returns from those investments. With an emphasis on education, many people pursue education as an investment in human capital because education can equip them with knowledge and skills that are generally considered as important factors in determining earnings. In addition, labor market participants are more likely to invest in education when they observe higher returns to education, measured by how much an additional amount of education can contribute to labor earnings. However, we might observe heterogeneous returns to education, which means two identically observed individuals receive different levels of earnings and returns to education, not only in China but also in other parts of the world. Intuitively speaking, people anticipate higher returns to education in regions that are more developed and seek more skilled workers.

1.2 What Is Geographic Heterogeneity?

Economists have been aware of heterogeneity as significant problem with micro-data. Heterogeneity arises mainly because people are diverse and they make diverse decisions. In addition, we might ask the following question: is there any economic theory or intepretation in depth concealed behind heterogeneity? Now, let us consider people in different geographic regions. Due to heterogeneity, it is natural to believe that people living in those regions make different decisions in their daily lives, in economic activities, and in human capital investments. Thus, we define geographic heterogeneity, or spatial heterogeneity, to explain this phenomenon. In this thesis, heterogeneity of this kind implies that the relation among returns to education for a set of regions might be different from the relation for another set of regions. In addition, geographic heterogeneity is of strong interest to researchers as the reasons behind heterogeneity might help government to make appropriate policies to raise labor earnings, maintain social welfare and sustain political stability. Although we might find other explanations for geographic heterogeneity of returns to education, the following analysis provides us with an intuitive view of such heterogeneity.

1.2.1 Local Governments

Local governments play an important role in accounting for geographic hetereogeneity. First of all, human capital of workers tends to be higher in rich provinces

Table 1.1: Educational Attainment
(*CHIP 1995 did not include Chongqing)

Province 1995/2002	Lower than middle school	Middle school	High school	College	Higher than college
Beijing	230/133 (16%/9%)	466/396 (32%/27%)	536/643 (37%/44%)	174/230 (12%/16%)	47/45 (3%/3%)
Shanxi	615/520 (33%/28%)	545/560 (29%/31%)	566/542 (30%/30%)	107/175 (6%/10%)	26/39 (3%/3%)
Liaoning	456/376 (22%/18%)	809/753 (40%/36%)	609/710 (30%/34%)	144/227 (7%/11%)	25/43 (1%/2%)
Jiangsu	586/519 (26%/24%)	775/675 (35%/31%)	699/720 (31%/34%)	149/205 (7%/10%)	39/30 (2%/1%)
Anhui	298/320 (21%/22%)	549/478 (39%/33%)	436/506 (31%/35%)	103/129 (7%/9%)	21/31 (1%/2%)
Henan	549/575 (31%/28%)	602/577 (34%/28%)	500/743 (28%/36%)	98/151 (6%/7%)	17/27 (1%/1%)
Hubei	488/471 (23%/23%)	693/572 (33%/28%)	724/749 (34%/36%)	160/220 (8%/11%)	46/44 (2%/2%)
Guangdong	564/452 (32%/24%)	555/606 (32%/32%)	520/643 (30%/35%)	87/149 (5%/8%)	31/16 (2%/1%)
Chongqing*	-/164 (-/20%)	-/282 (-/34%)	-/281 (-/34%)	-/89 (-/11%)	-/11 (-/1%)
Sichuan	581/468 (26%/28%)	767/567 (34%/33%)	689/527 (30%/31%)	180/104 (8%/6%)	48/30 (2%/2%)
Yunnan	463/561 (25%/31%)	612/518 (33%/28%)	639/517 (35%/28%)	110/193 (6%/11%)	24/47 (1%/3%)
Gansu	341/283 (30%/24%)	405/351 (35%/30%)	315/427 (27%/36%)	76/110 (7%/9%)	16/18 (1%/2%)
Total	5171/4842 (26%/24%)	6778/6263 (34%/31%)	6233/6971 (31%/34%)	1388/1982 (7%/10%)	340/381 (2%/2%)

because of the greater investment in human capital made by rich local governments. In China, education is mainly funded by local governments, which implies that rich provinces are able to invest more in education, both in quantity and quality. They can support investment in human capital by explicitly spending money in enlarging school enrollment or implicitly issuing educational subsidies and loans. As a result, people in those rich provinces have easier access to education resources than those in poor provinces and they can foster higher levels of human capital. Consider Table 1.1, which shows educational attainment in China

by province. As we can see from the table, Beijing has the greatest percentage share both in high school graduates and college graduates among all provinces in both 1995 and 2002 according to CHIP. Beijing is the capital city of China and it is comparatively richer than other regions in China. In addition, some provinces in China might also spend a lot of government income funding education and other forms of investments in human capital even if they are less developed than the most developed regions due to miscellaneous factors, such as cultural and political ones. As China grows, we can see the educational attainment was shifting from lower levels to higher levels.

1.2.2 Skill-Biased Technology

Different geographic regions in China reward education differently, which means wages and returns to education for the same person could be different depending on the region in which he works. A possible explanation comes from the following reason: returns to education are more likely to be higher for a region that demands relatively more higher-skilled workers. For instance, Shanghai has been transforming into an international metropolitan center that attracts many firms not only from China but also from the entire world. As a result, those firms in Shanghai will demand a large amount of high-skilled workers in order to complement the advanced technology brought by the large FDI inflows. If the labor market in Shanghai suffers from insufficient supply of high-skilled labors, they will be expensive because of insufficient supply and investment in human capital seems

profitable. In contrast, for workers in regions where the labor market does not require many high-skilled workers, higher education might be underpaid and returns to education tend to be lower. However, if there is no barriers to interprovincial migration, migration might greatly eliminate geographic heterogeneity as people can move to provinces where they receive higher earnings and returns to education more easily.

1.2.3 Hukou and Discrimination

Apart from technology bias, it is believed that China's household registration system, the hukou system, has a significant effect on investment decisions in human capital, especially for migrant workers in recent years (Heckman and Yi, 2012). The hukou system in China places several restrictions on labor migration, educational attainment, health care and other factors in highly populated regions, such as Beijing and Shanghai. For those who come from less developed regions in China, higher living cost and discriminating policies in urban areas can reduce their incentives to achieve education or on-the-job training, while such a discriminatory environment might also stimulate their incentives to achieve education or other forms of human capital in order to survive in such environment. Some rich people from other regions even want to buy hukou in Beijing and Shanghai because such hukou can bring them various benefits. In fact, how the hukou system affects migrants' investment in human capital might depend on their preferences towards discrimination, income, living standards, etc. Moreover, different regions set up

different hukou policies, thus migrant workers in those regions invest differently in human capital and earn wages differently.

Firms do discriminate against workers. In China, this happens because firms in local region tend to hire local workers in order to maintain economic welfare for local citizens. Especially for large cities with a great number of incoming migrants, the huge population implies fierce competition in getting a job and migrants seek local hukous since some people believe it is more likely to be employed if they have local hukous. Consider Shanghai again, given two identically observed individuals who come from different regions, firms in Shanghai prefer to hire local citizens rather than hiring migrants from other regions of China, because local government wants to protect its citizens by reducing employment competition with migrant workers. Moreover, it might be relatively more difficult for migrant workers to get job promotion than local citizens due to discrimination because of priority granted to local people. However, such protective employment also creates an externality of investment in human capital. Local citizens know they will be more likely to get jobs because their government will make an effort to secure more employment opportunities for them. This reduces the incentives for local citizens to invest in human capital. In addition, different geographic regions have different levels of discrimination against migrant workers, and these migrants may consequently make different investments in human capital in response to different levels of discrimination. If we believe hukou affects earning for some provinces, such as Beijing and Shanghai, we will have to include hukou (and possible interaction terms) as the explanatory variable(s) for those provinces in our analysis to avoid omitted

variable bias.

En route to the conclusion, I explore the data from the Chinese Household Income Project (CHIP) conducted in 1995 and 2002 for urban observations in 12 provinces, provided by the Inter-university Consortium for Political and Social Research (ICPSR). I begin with stating the earning equation with heterogeneous returns to education, based on Mincer's work (1974). However, we will confront sample selection and endogeneity where we need Heckman's selection model (1979) and instrumental variables method to solve. For the IV method, I use regional level (log) average income, (log) unemployment rate and (log) education expenditure as the instruments, test the endogeneity of education using the Durbin-Wu-Hausman (DWH) test, and compute returns to education for each region in China under IV estimators. After comparing returns to education from using different models, I demonstrate the spatial autocorrelation (SAR) model and estimate the correlation coefficient of returns to education in different regions, using results from previous estimations. Finally, I compare the results of correlation coefficients for different educational levels and arrive at a conclusion that might account for the geographic heterogeneity.

Chapter 2

Literature Review

China has been growing tremendously since its reform. Many economists believe that large inflows of FDI into China and accumulation of medium-skilled workers played an important role in China's economic growth as FDI has significantly contributed to investment in physical capital which would be complementary with medium-skilled workers (Heckman and Yi, 2012). Recently, the Chinese government has been increasing its investment in human capital due to the complementarity between human capital and physical capital by increasing the share of GDP going to education expenditure, from 2.5% in 2002 to 3.66% in 2011 according to the Ministry of Education in China. In addition, a recent news report states that China is anticipating a stock of about 195 million college graduates by 2020 (Bradsher, 2013) by making a great investment in human capital. However, earlier literature shows China suffered from a low level of return to education, at a rate of about 4% (Chow, 2001), which might account for the under-investment in human capital and the over-investment in physical capital in earlier China. Compared to

other countries, what Chow estimated is very low for a transitional economy. The return to education in many transitional economies is at least 8% and the average was about 12.8% for Asian countries (Psacharopoulos, 1981). In the United States and many Western countries, the return to education is typically estimated between 15% to 20% (Heckman, 2002). However, more recent literature indicates that workers in China were experiencing much higher returns to education in the last ten years. For instance, Fleisher and Wang (2004) estimated the return to education was as high as 30% or 40% in China. In addition, Li and Luo (2004) used the Generalized Method of Moments (GMM) estimation proposed by Hansen (1982), and they reported the return to education was 15% on average and 16.9% for females. Besides, Fang et al. (2012), use the compulsory education law in China as an instrumental variable and estimate the return to education. Their results show that the average return to education is approximately 20% per year. They also indicate that returns to education is higher in rural and costal regions of China.

While their works provide insightful information about returns to education in China, it is of stronger interest to study the heterogeneity behind the returns to education as it might help to explain economic phenomena, such as inequality and labor migration. Heckman and Li (2004) studied the heterogeneity in returns to education with distinct comparative advantages. He who enjoys comparative advantage in academics is more likely to pursue higher education; other people might pursue lower level of education because they demonstrate other forms of comparative advantages. Thus people make heterogeneous investments in human

capital and receive heterogeneous returns to education. They estimated that returns to college completion is 43% (Heckman and Li, 2004). As we see reasons in the first chapter, local provinces, technology bias, hukou system and job discrimination result in geographic heterogeneity and affect investment made in human capital. In this thesis, we will examine the geographic heterogeneity meaning the returns to education differ for people in different geographic regions and we will look at the possible reasons that account for such heterogeneity.

Although many econometricians challenge the functional form of the Mincer earnings equation (Heckman et al., 2008), I assume the standard form (education, experience, squared experience and other explanatory variables) in this thesis as we shall see in the next chapter.

Chapter 3

Model and Methodology

3.1 Model Specification

3.1.1 Heterogeneous Returns Model

Consider the Mincer earnings equation (Mincer, 1974):

$$\ln y_i = \beta s_i + \gamma X_i^{'} + \varepsilon_i \tag{3.1}$$

where for each observation i, y_i is earned income; s_i is years of education; ε_i is a stochastic error with $\mathbb{E}(\varepsilon_i) = 0$; X_i is a $1 \times k$ vector of explanatory variables such as job experience, experience squared, male dummy, province dummies, etc. Correspondingly, γ is a $1 \times k$ vector of coefficients and we are interested in estimating β in equation (3.1), which is assumed to be the same for all observations in the data. In this thesis, we consider the situation where β is heterogenous for different

geographic regions, which implies we will have a vector of $\widetilde{\beta} =< \beta_1, \beta_2, ..., \beta_j >$ for j regions. Then, the Mincer earning equation with heterogeneous returns to education is given as follows:

$$\ln y_{ij} = \beta_j s_{ij} + \gamma_j X_{ij}^{'} + \varepsilon_{ij} \tag{3.2}$$

where β_j is the return to education for observation i in region j. For each observation i in region j, y_{ij} is the earned income, X_{ij} is a $1 \times k$ vector of explanatory variables and γ_j is a $1 \times k$ vector of coefficients. In equation (3.2), $\forall j \in \{1, 2, 3, ...\}$, $\mathbb{E}(\varepsilon_{ij}) = 0$ and β_j is assumed to be the same for all observations in region j. However, for $j, j^{'} \in \{1, 2, 3, ...\}$, β_j does not equal $\beta_{j'}$ assuming the existence of heterogeneity between two regions. As a specification issue, the Mincer earnings equation suffers from an omitted variable bias by excluding an ability measure, which we will discuss in the next section.

3.1.2 Sample Selection Model

In microdata, we confront the problem of missing values, such as data censoring, data truncation and incidental truncation (Wooldridge, 2010). For example, some student observations in our data might not have earned income reported because they are not currently employed, thus not having income. However, ignoring those observations will remove randomness of data because income data is not missing randomly. In this thesis, we will deal with the incidental truncation because we only observe income when the observation is participating in the labor market.

Hence, we will use the sample selection model (Heckman, 1974) to solve this issue. The sample selection model consists of two equations, a selection equation, given as follows:

$$\omega_{ij}^{*} = \delta_j Z_{ij}^{'} + u_{ij} \qquad \Pr(\omega_{ij} = 1|Z_{ij}) = \Phi(\delta_j Z_{ij}^{'}) = \int_{-\infty}^{\delta_j Z_{ij}^{'}} \phi(x)dx \tag{3.3}$$

where ω_{ij}^{*} is a latent variable and ω_{ij} is a binary response variable. $\Pr(\omega_{ij} = 1|Z_{ij}) = \Phi(\delta_j Z_{ij})$ is a probit model and the inverse Mill's ratio λ is given by the equation $\lambda(\cdot) = \frac{\phi(\cdot)}{\Phi(\cdot)}$ that is the density over the cumulative distribution function for a standard normal random variable. In equation (3.3), Z_{ij} includes explanatory variables such as unity, marriage status, age, educational level, etc. In the probit model, we require that $u_{ij} \sim N(0, 1)$ The second equation is an outcome equation, given as follows:

$$\ln y_{ij} = \beta_j s_{ij} + \gamma_j X_{ij}^{'} + \varepsilon_{ij} \tag{3.4}$$

In this case, our outcome equation is exactly equation (3.2) for the heterogeous returns model. We will observe y_{ij} if $\omega_{ij} = \mathbb{1}[\omega_{ij}^{*} > 0] = \mathbb{1}[\delta_j Z_{ij}^{'} + u_{ij} > 0]$ whereas y_{ij} is not observed otherwise. In this case, we might regard w_{ij} as the market wage minus the reservation wage for observation i in region j. If the market wage is greater than the reservation wage, the observation i will be working in the labor market and we can observe her labor income; if the market wage is less than the reservation wage, the observation i will not be participating in the labor market and her potential labor income is unobservable to us. Applying regression to

those who have observable income ($\omega_{ij} = 1$) will ignore other observations who are potentially capable of earning income, such as homemakers, which removes randomness from the data. In order to estimate the model using equations (3.3) and (3.4), Wooldridge (2010) states the following assumptions: (1) Standard normal error term: given j, $u_{ij} \sim N(0,1)$; (2) $Cov(Z_{ij}, \varepsilon_{ij}) = 0$; (3) $Cov(Z_{ij}, u_{ij}) = 0$; (4) We always observe the explanatory vector Z_{ij}; (5) We always observe the binary response ω_{ij}; (6) $\mathbb{E}(\varepsilon_{ij}|u_{ij}) = \pi_j u_{ij}$. Then as Heckman suggested in his Nobel Prize-winning paper (1979), the second equation in (3.4) suffers from a specification problem as it omits an important explanatory variable λ_{ij}. For the sake of consistency, we need to include the inverse Mill's ratio in our analysis using the following model (see Appendix A.1):

$$\mathbb{E}(\ln y_{ij}|Z_{ij}, \omega_{ij} = 1) = \beta_j s_{ij} + \gamma_j X_{ij}^{'} + \pi_j \lambda(\delta_j Z_{ij}^{'}) \tag{3.5}$$

If $\pi_j = 0$, then there will be no selection bias and equation (3.5) is equivalent to equation (3.2). Known as the Heckit procedure, we use the probit model to estimate δ_j and λ, then use OLS to estimate coefficients of our main interest, β_j and γ_j (see Appendix A.2 for proof). This is called the two-step method where we can also alternatively use a maximum likelihood method to estimate the results.

3.1.3 Endogeneity of Education

A consistent OLS estimator demands the exogeneity condition on all explanatory variables. Thus, we require $Cov(s_{ij}, \varepsilon_{ij}) = 0$. However, it is widely believed that

the exogeneity condition fails for s_{ij}, i.e., $Cov(s_{ij}, \varepsilon_{ij}) \neq 0$. Without specifying ability in the Mincer earning equation, ability will be contained in ε_{ij} that we require $Cov(s_{ij}, \varepsilon_{ij}) = 0$. Although ability is not directly observable in real life, economic theory tells us ability is strongly correlated with educational attainment. Generally speaking, people with higher ability tend to pursue higher levels of education, invest more in human capital and expect greater economic returns. When equation (3.2) excludes a measure for ability, we violate the condition $Cov(s_{ij}, \varepsilon_{ij}) = 0$ and an OLS estimator will produce an inconsistent estimate for returns to education. In this case, we say that education is considered as an endogeneous variable such that $Cov(s_{ij}, \varepsilon_{ij}) \neq 0$ and we seek other estimation methods other than OLS. Many studies employ instrumental variables (IV) to remove the ability bias, such as tuition, quarter of birth and distance to schools. Given region j, consider the following simultaneous equations model (SEM):

$$\ln y_{ij} = \beta_j s_{ij} + \gamma_j X_{ij}^{'} + \varepsilon_{ij} \qquad s_{ij} = \theta_j V_{ij}^{'} + \alpha_j Z_{ij}^{'} + \delta_{ij} \tag{3.6}$$

where the first equation in (3.4) is called the structural form and the second one is called the reduced form. In equation (3.6), s_{ij} is the suspected endogenous variable; V_{ij} is a $1 \times k$ vector of instruments such that $\text{Cov}(V_{ij}, s_{ij}) \neq 0$, $\text{Cov}(V_{ij}, \delta_{ij}) = 0$ and Z is a $1 \times k$ vector of explanatory variables such that $\text{Cov}(Z_{ij}, \delta_{ij}) = 0$. Equation (3.6) implies that the educational level of a particular observation i in region j is a function of explanatory variables (unity, gender, province, etc.) in Z_{ij}, instruments (and their interaction terms, if possible) in V_{ij} and a stochastic error term ε_{ij}. In order to examine the endogeneity of education, we want to test

H_0: $\text{Cov}(s_{ij}, \varepsilon_{ij}) = 0$ in the structural form. Here, I use the Durbin-Wu-Hausman (DWH) test to examine the endogeneity of education. To complete a DWH test, we need to examine whether we will trust our instruments, i.e., whether we have valid instruments. First of all, we need to check whether the selected instruments are correlated with error terms in the structural equation, which in our case will be the Mincer equation. Secondly, we need to check whether the instruments are correlated with the endogenous variable(s). Equivalently, if we want to use a vector V as our instruments, two conditions have to be satisfied: (1) $Cov(V_{ij}, \varepsilon_{ij}) = 0$, where ε_{ij} is the error term in the Mincer equation; (2) $Cov(V_{ij}, s_{ij}) \neq 0$, where s_{ij} is the endogenous variable in the Mincer equation. Such V_{ij} will be considered valid instruments. To find such valid instruments, we need to start with using some common sense. In this thesis, I use the (ln) average income and the (ln) unemployment rate in a given region as my instruments. On the one hand, given region j, average income is uncorrelated with ability and people in region j are more likely to invest more in human capital if the average income in that region is also higher. On the other hand, given region j, the unemployment rate is also uncorrelated with ability and a high unemployment rate signals there is difficulty to find a job. Thus people in region j will be more likely to invest in human capital in order to get a job. Technically, we need to have more instruments than endogenous variables so that the structural parameters will be identified. In addition, the estimation might benefit from over-identification that reduces standard errors. Assuming valid instruments, we will compute the residuals $\hat{\varepsilon}_{ij}$ from the reduced form of SEM in (3.6), include the residuals $\hat{\varepsilon}_{ij}$ in the structural

form, and perform another regression for the following equation:

$$\ln y_{ij} = \beta_j s_{ij} + \gamma_j X_{ij}^{'} + \rho\hat{\varepsilon}_{ij} + \varepsilon_{ij} \tag{3.7}$$

If results show that ρ is significantly different from zero, then we will reject H_0: $\rho = 0$ and education is claimed to be endogenous. This completes a DWH test and two-stage least squares (2SLS) will produce estimation results. In general, sample selection and endogeneity apply more frequently to female and male observation respectively. On the one hand, almost every male will work, especially in a transitional economy like early China, while females will self select themselves into the labor market depending on their marriage status, number of children and other factors. Thus, the sample selection model works better for female observations. On the other hand, females with comparatively stronger ability will not always pursue a higher level of education as marriage obliges them to spend time at home. Hence, IV estimation gives more information on male observations.

At this point, we will compare estimated returns to education for different geographic regions from OLS, IV and the sample selection model. In the next section, we will look at spatial econometric techniques that provide us with a possible way to study geographic heterogeneity and use the estimated returns to education in the spatial econometric model.

3.2 Spatial Econometrics

3.2.1 Why Spatial Econometrics?

Spatial econometrics introduces econometric models that study regional effects. Many estimates ignore the regional effect that can actually exist. In this thesis, we are interested in how returns to education differ for geographic regions and why they differ. We interpret the heterogeneity by considering two essential concepts in spatial econometrics: spatial dependence and spatial heterogeneity. The first term, spatial dependence, refers to the situation in which a randomly selected observation from a geographic region might depend on observations sampled from other regions (Lesage, 2009). As for returns to education, geographic dependence can provide information about heterogeneity. For instance, when we observe a high return to education in a region, the neighboring regions are more likely to also enjoy high returns to education because large difference in returns to education might motivate people to migrate into the region with higher returns to education. The second term, spatial heterogeneity, or geographic heterogeneity, as we discussed in the previous chapter, refers to the situation where the relation among a variable in a set of regions might be different from the relation for another set of regions. In general, stronger spatial dependence implies less spatial heterogeneity. To measure the spatial dependence among a set of regions, you might find knowledge of spatial econometrics useful, as we shall see in the following section.

3.2.2 Spatial Autoregressive Regression Model

In this section, we look at the spatial autoregressive regression (SAR) model that estimates correlation coefficients among a set of regions. First, we need to model the relation among different locations, which is achieved by using the contiguity matrix $\mathbb{C}$. The general SAR model has the following functional form:

$$\widetilde{\beta} = \xi + \rho\mathbb{C}\widetilde{\beta} + \varepsilon \tag{3.8}$$

where $\widetilde{\beta}$ is a $k \times 1$ vector of a particular variable and $\widetilde{\beta}$ will be returns to education in this thesis; ξ is a $k \times 1$ vector of intercepts; ρ is a scalar which measures the spatial dependence (or heterogeneity) of $\widetilde{\beta}$. In this thesis, ρ is of interest; $\varepsilon \sim N(\mathbf{0}, \sigma^2\mathbb{I})$ and ξ is a $k \times 1$ vector of intercepts; $\mathbb{C}$ is a contiguity matrix which we can manipulate in order to make our estimate of ρ more reasonable and we will see how $\mathbb{C}$ and manipulated matrices is constructed in the following subsection.

3.2.3 Contiguity Matrix of China

3.2.3.1 Binary Contiguity Matrix

In order to model the relations for a set of geographic regions, we use the contiguity matrix that will be discussed in this section. In fact, we can input different contiguity matrices and estimate the spatial dependence coefficients using those different matrices. Start from considering a particular entry c_{ij} in the contiguity matrix $\mathbb{C}$ where c_{ij} is a binary response indicating whether region i and region j

geographically share a boundary. $c_{ij} = 1$ if they have a common boundary; $c_{ij} = 0$, otherwise. For CHIP 2002, we have twelve provinces and the binary contiguity matrix is given as follows:

$$\mathbb{C} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

3.2.3.2 First Order Contiguity Matrix

Apart from the simple case of a binary matrix where the relation between two regions is only represented by a binary response, we might want to consider using

the first order contiguity matrix, $\mathbb{A}$, given as follows:

$$\mathbb{A} = \begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1/2 & 1/2 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1/3 & 0 & 1/3 & 1/3 & 0 & 0 & 0 & 0 & 0 \\
0 & 1/4 & 0 & 1/4 & 1/4 & 0 & 1/4 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1/3 & 1/3 & 0 & 0 & 1/3 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1/2 & 0 & 0 & 1/2 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 0 & 1/3 & 1/3 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0
\end{pmatrix}$$

Consider the following reason that the first order contiguity matrix, $\mathbb{A}$, is in fact an appropriate alternative to the original binary matrix. In our case, take $i = 10$ (Sichuan) and look at the $10^{\mathbf{th}}$ row of $\mathbb{A}$. Sichuan (province 51) shares a boundary with other provinces. They are Chongqing (province 50), Yunnan (province 53) and Gansu (province 62). If for a particular educational attainment group, the return to attaining the corresponding educational level is β_{51}, β_{50}, β_{53} and β_{62} for Sichuan, Chongqing, Yunnan and Gansu, respectively. Under the first order contiguity matrix, we can represent β_{51} as a linear combination with returns to education in provinces that share a common boundary with Sichuan. In general, if we have a set of n regions and a particular region, say region i, shares a boundary

with k regions and we want to look at the dependence of a variable β_j where $j \in \{1, 2, 3, ..., n\}$ in this set of n regions, then the linear combination for β_i can be written as:

$$\beta_i = \sum_{j \in \Omega_j} \frac{c_{ij}\beta_{ij}}{k} \tag{3.9}$$

where c_{ij} is a binary response such that $c_{ij} = 1$ if they have a common boundary; $c_{ij} = 0$, otherwise. In the case of Sichuan, $\beta_{51} = \frac{1}{3}(\beta_{50} + \beta_{53} + \beta_{62})$. So using the first order contiguity matrix indeed allows us to represent a particular return to education as the average of returns in contiguous regions, such as representing β_{51} as the average of β_{50}, β_{53} and β_{62}.

Some people point out regions that do not share a boundary might also have influence on each other and a simple binary or first order matrix ignores such remote interaction. For the binary matrix, although regions that do not share a common boundary is represented by $c_{ij} = 0$, they might still be correlated in the following sense. Consider region x, y and z such that we directly observe region x and region y share a common boundary, region y and region z do not share a boundary and region x and region z do not share a boundary. Thus, $c_{xy} = 1$, $c_{yz} = 1$ and $c_{xz} = 0$. However, region x and region z are still connected in the way that their connection is built via region y since region y shares a boundary with both region x and region z. If we think this way, region x and region z can be regarded as connected, but their connection is not direct. In addition to the above reason, I provide an alternative matrix with specification of a distance effect in the following subsection.

3.2.3.3 Contiguity Matrix with Specification of Distance

Some might argue that the first order contiguity matrix fails to specify the distance effect for spatial dependence. In general, dependence tends to be stronger for close objects than for remote objects. Thus, the specification of distance produces the following matrix:

$$\mathbb{D} = \begin{pmatrix}
0 & \frac{c}{d_{1,2}} & \frac{c}{d_{1,3}} & \frac{c}{d_{1,4}} & \frac{c}{d_{1,5}} & \frac{c}{d_{1,6}} & \frac{c}{d_{1,7}} & \frac{c}{d_{1,8}} & \frac{c}{d_{1,9}} & \frac{c}{d_{1,10}} & \frac{c}{d_{1,11}} & \frac{c}{d_{1,12}} \\
\frac{c}{d_{2,1}} & 0 & \frac{c}{d_{2,3}} & \frac{c}{d_{2,4}} & \frac{c}{d_{2,5}} & \frac{c}{d_{2,6}} & \frac{c}{d_{2,7}} & \frac{c}{d_{2,8}} & \frac{c}{d_{2,9}} & \frac{c}{d_{2,10}} & \frac{c}{d_{2,11}} & \frac{c}{d_{2,12}} \\
\frac{c}{d_{3,1}} & \frac{c}{d_{3,2}} & 0 & \frac{c}{d_{3,4}} & \frac{c}{d_{3,5}} & \frac{c}{d_{3,6}} & \frac{c}{d_{3,7}} & \frac{c}{d_{3,8}} & \frac{c}{d_{3,9}} & \frac{c}{d_{3,10}} & \frac{c}{d_{3,11}} & \frac{c}{d_{3,12}} \\
\frac{c}{d_{4,1}} & \frac{c}{d_{4,2}} & \frac{c}{d_{4,3}} & 0 & \frac{c}{d_{4,5}} & \frac{c}{d_{4,6}} & \frac{c}{d_{4,7}} & \frac{c}{d_{4,8}} & \frac{c}{d_{4,9}} & \frac{c}{d_{4,10}} & \frac{c}{d_{4,11}} & \frac{c}{d_{4,12}} \\
\frac{c}{d_{5,1}} & \frac{c}{d_{1,2}} & \frac{c}{d_{5,3}} & \frac{c}{d_{5,4}} & 0 & \frac{c}{d_{5,6}} & \frac{c}{d_{5,7}} & \frac{c}{d_{5,8}} & \frac{c}{d_{5,9}} & \frac{c}{d_{5,10}} & \frac{c}{d_{5,11}} & \frac{c}{d_{5,12}} \\
\frac{c}{d_{6,1}} & \frac{c}{d_{6,2}} & \frac{c}{d_{6,3}} & \frac{c}{d_{6,4}} & \frac{c}{d_{6,5}} & 0 & \frac{c}{d_{6,7}} & \frac{c}{d_{6,8}} & \frac{c}{d_{6,9}} & \frac{c}{d_{6,10}} & \frac{c}{d_{6,11}} & \frac{c}{d_{6,12}} \\
\frac{c}{d_{7,1}} & \frac{c}{d_{7,2}} & \frac{c}{d_{7,3}} & \frac{c}{d_{7,4}} & \frac{c}{d_{7,5}} & \frac{c}{d_{7,6}} & 0 & \frac{c}{d_{7,8}} & \frac{c}{d_{7,9}} & \frac{c}{d_{7,10}} & \frac{c}{d_{7,11}} & \frac{c}{d_{7,12}} \\
\frac{c}{d_{8,1}} & \frac{c}{d_{8,2}} & \frac{c}{d_{8,3}} & \frac{c}{d_{8,4}} & \frac{c}{d_{8,5}} & \frac{c}{d_{8,6}} & \frac{c}{d_{8,7}} & 0 & \frac{c}{d_{8,9}} & \frac{c}{d_{8,10}} & \frac{c}{d_{8,11}} & \frac{c}{d_{8,12}} \\
\frac{c}{d_{9,1}} & \frac{c}{d_{9,2}} & \frac{c}{d_{9,3}} & \frac{c}{d_{9,4}} & \frac{c}{d_{9,5}} & \frac{c}{d_{9,6}} & \frac{c}{d_{9,7}} & \frac{c}{d_{9,8}} & 0 & \frac{c}{d_{9,10}} & \frac{c}{d_{9,11}} & \frac{c}{d_{9,12}} \\
\frac{c}{d_{10,1}} & \frac{c}{d_{10,2}} & \frac{c}{d_{10,3}} & \frac{c}{d_{10,4}} & \frac{c}{d_{10,5}} & \frac{c}{d_{10,6}} & \frac{c}{d_{10,7}} & \frac{c}{d_{10,8}} & \frac{c}{d_{10,9}} & 0 & \frac{c}{d_{10,11}} & \frac{c}{d_{10,12}} \\
\frac{c}{d_{11,1}} & \frac{c}{d_{11,2}} & \frac{c}{d_{11,3}} & \frac{c}{d_{11,4}} & \frac{c}{d_{11,5}} & \frac{c}{d_{11,6}} & \frac{c}{d_{11,7}} & \frac{c}{d_{11,8}} & \frac{c}{d_{11,9}} & \frac{c}{d_{11,10}} & 0 & \frac{c}{d_{11,12}} \\
\frac{c}{d_{12,1}} & \frac{c}{d_{12,2}} & \frac{c}{d_{12,3}} & \frac{c}{d_{12,4}} & \frac{c}{d_{12,5}} & \frac{c}{d_{12,6}} & \frac{c}{d_{12,7}} & \frac{c}{d_{12,8}} & \frac{c}{d_{12,9}} & \frac{c}{d_{12,10}} & \frac{c}{d_{12,11}} & 0
\end{pmatrix}$$

where $c \in \mathbb{R}$ is a scalar that I will discuss later. This is a symmetric matrix as $\forall(i,j) \in \Omega_i \times \Omega_j$ where Ω is the sample space, symmetry of the distance function (or metric) states that $d_{i,j} = d_{j,i}$. In $\mathbb{R}^2$, the distance function between two points is the Euclidean distance defined by the metric $d(x,y) = [(x_1 - y_1)^2 + (x_2 - y_2)^2]^{\frac{1}{2}}$ where $x = (x_1, x_2)$ and $y = (y_1, y_2)$. For region i and region j, if $d_{i,j}$ is small, then

their dependence will be large, as reflected in the term $\frac{c}{d_{i,j}}$. The reason to include a scalar c is when estimating the spatial dependence coefficient ρ, ρ is always in the set $[-1, 1]$. If $\rho = 1$, the β's are perfectly correlated with other regions whereas $\rho = -1$ implies that β's are perfectly negatively correlated with each other. If $\rho = 0$, then we conclude that there is no spatial dependence in the set. Distance can sometimes be very large, and including a large distance in the denominator makes this entry numerically very small, which can cause ρ to be very large or small in estimation. Thus, a rescaling of matrix $\mathbb{D}$ by a constant $c \in \mathbb{R}$ assures reasonable estimates of ρ.

3.2.4 Interpreting Heterogeneity

We will see how estimation of ρ will provide us with information on geographic heterogeneity in this section. When we estimate returns to education, we estimate them with respect to 5 levels of educational attainment: $s_1 < 9$ (less than middle school), $s_2 \in \{9, 10, 11\}$ (middle school), $s_3 \in \{12, 13, 14, 15\}$ (high school), $s_4 \in \{16, 17\}$ (college) and $s_5 > 18$ (higher than college), assuming normal length to complete the corresponding degree. For each region j in China, we will have 4 returns to education with $s_1 < 9$ being the reference group; And for each level of educational attainment, we will have $(j - 1)$ returns to education with some particular region. In most cases, we use the region with the smallest or most negative estimate of return to education as the reference group so that the rest of the estimates will all become non-negative. Then, we use the SAR model to compute

ρ for each level of educational attainment. Hence, each level of educational attainment will be associated with a particular ρ, i.e., we have ρ_{MS}, ρ_{HS}, ρ_{CL} and ρ_{GR} for return to middle school attainment, return to high school attainment, return to college attainment and return to higher degrees, respectively. Then, we can see the relation between educational levels and ρ's, i.e., if greater educational level implies greater spatial dependence (less heterogeneity). If ρ's increase in educational levels, we will conclude that greater educational levels increase spatial dependence and decrease heterogeneity of returns to education; if ρ's decrease in educational levels, we will conclude that greater educational levels decrease spatial dependence and increase heterogeneity of returns to education. We will look at our data and examine the estimation results in the upcoming chapter.

Chapter 4

Data and Results

4.1 Data Description

In this thesis, I explore the data from the Chinese Household Income Project (CHIP) conducted in 1995 and 2002 for urban observations in 12 Chinese provinces, provided by the Inter-university Consortium for Political and Social Research (ICPSR). ICPSR conducted a series of CHIPs 1988, 1995, 2002 and 2006. However, CHIP 2006 is not currently available to the public. I hope comparing results from CHIP 1995 and CHIP 2002 can make up for the unavailability of CHIP 2006. The original CHIP 1995 and 2002 datasets contain 21698 valid urban observations, 102 variables from 11 provinces and 20548 valid urban observations, 151 variables from 12 provinces (with an addition of Chongqing). As for the provinces, we find coastal provinces such as Jiangsu and Guangdong and we can also observe provinces in the western area of China, such as Gansu, Sichuan and Yunan (see Tables C.28 and C.29 in Appendix C). Notice that western China is generally

considered poorer than coastal provinces in China, as you might agree with this argument once you observe the evidence from the different intercepts in my estimation results (see Tables C.5-C.15 and C.16-C.27 in Appendix C). Working observations are also selected from a wide range of industries (see Tables C.30 and C.31 in Appendix C).

There are a lot of variables in CHIP 1995 and 2002 and we only look at some important variables in our analysis. The key variables in CHIP 1995 and 2002 are (log) labor income (annual), schooling years, educational attainment, gender, age, work experience, province, hukou, marriage status and working industry. Two CHIPs are very similar but CHIP 2002 contains more information than CHIP 1995. For example, CHIP 2002 includes Chongqing as an additional province in the dataset and we can find hukou variables from CHIP 2002. In addition, I obtained my instruments from the National Bureau of Statistics (NBS) in China. Notice that the instrumental variable method is only applied to CHIP 2002 since NBS did not include 1995 regional level data. Those are different between CHIP 1995 and 2002. We will begin the regression analysis starting from examining the endogeneity of education in the next section to determine whether we need to use an IV method.

4.2 Testing Endogeneity

Given a region j, the reduced form of the simultaneous equations model is given by the following equation:

$$s_{ij} = \beta_j \theta_j V_{ij}^{'} + \beta_j \alpha_j Z_{ij}^{'} + \gamma_j X_{ij}^{'} + \beta_j \delta_{ij} + \varepsilon_{ij} \tag{4.1}$$

Then, we compute $\hat{\varepsilon}_{ij}$ from equation (4.1), include this residual in the structural equation and perform another regression based on the following:

$$\ln y_{ij} = \beta_j s_{ij} + \gamma_j X_{ij}^{'} + \rho_j \hat{\varepsilon}_{ij} + \varepsilon_{ij} \tag{4.2}$$

where we estimate ρ_j and test H_0: $\rho_j = 0$ where education is endogenous if null hypothesis is rejected. Here, we will test for endogeneity of the whole sample in lieu of testing for endogeneity of each region, meaning we estimate ρ and test H_0: $\rho = 0$. If the whole sample suffers from endogeneity of education, then its subsample also suffers from this endogeneity. In Table C.32 (see Appendix C), we observe the $t-$statistic is -3.49 thus we reject H_0: $\rho = 0$ and education is believed to be an endogenous variable.

4.3 Returns to Education

Recall that the Mincer earnings equation with heterogenous returns has the following functional form:

$$\ln y_{ij} = \beta_j s_{ij} + \gamma_j X_{ij}^{'} + \varepsilon_{ij} \tag{4.3}$$

To be precise, we can further expand equation (4.3) by specifying the explanatory variables in $X_{ij}^{'}$ and rewrite this equation into the following form:

$$\begin{aligned} \ln y_{ij} = {} & \alpha_0 + \beta_j s_{ij} + \alpha_1 \cdot t + \alpha_2 \cdot t^2 + \alpha_3 \cdot male_{ij} + \alpha_4 \cdot male_{ij} \cdot s_{ij} \\ & + \sum_{j \in \Omega_j} \gamma_j \cdot region_j + \sum_{j \in \Omega_j} \psi_j \cdot region_j \cdot s_{ij} + \sum_{j \in \Omega_j} \zeta_j \cdot sector_j + \varepsilon_{ij} \end{aligned} \tag{4.4}$$

where s_{ij} is a vector of dummy variables indicating whether observation i at region j attains a corresponding education level (less than middle school, middle school, high school, college, higher). Notice that we have interaction terms between years of schooling and male, and years of schooling and region. Those interaction terms imply that the coefficients are heterogenous for males and for different regions. Simply differentiating $\ln y_{ij}$ with respect to years of schooling yields the following mathematical expression of returns to education:

$$\frac{\partial \ln y_{ij}}{\partial s_{ij}} = \beta_j + \alpha_4 \cdot male_{ij} + \sum_{j \in \Omega_j} \psi_j \cdot region_j \tag{4.5}$$

where different region index j indicates returns to education are different for $i \neq j$. My estimation results are shown in Tables C.5-C.27 given in Appendix C and

demonstrate the consistency with economic theory, such as diminishing returns to education, negative coefficient for female and negative coefficient for squared experience. OLS estimates are not considered consistent in the presence of either endogeneity or selection bias and they are just given in the table for the purpose of comparison. For all regions, the sample selection model produces smaller estimates than OLS for both 1995 and 2002. Given the same level of education, if OLS estimate is not very different from what the sample selection model estimates, then the selection bias is not very signficant for this given level of education.

We believe that IV estimation provides more reliable information for males and that the sample selection model produces more accurate results for females because almost every male who reaches working age works. Here selection bias will not be an issue for males, but men with stronger ability will be more likely to pursue higher levels of education than females. As we can observe from Tables C.16-C.27, there is an obvious distinction between results from the IV method and the sample selection model as the coefficients for male interacting with educational attainment are positive for IV and negative for sample selection (and OLS). Positive coefficients imply males with stronger ability have a greater return to education. In addition, my results are similar to what Heckman and Li had estimated (2004).

Notice that in 2002, some estimated returns are negative. For example, the estimated return to middle school attainment in Henan under the OLS estimator is negative at a level of 4.5% in 2002. To intepret this fact, consider the following reason. People with a higher level of education always anticipate higher labor

income and they are less preferred to work in the industry where less educated people work, such as the agricultural industry. However, as we can see from the table, working in the agricultural industry yields higher labor income than working in some other industries. In 2002, the agricultural industry has an estimated coefficient about 16% while some industries suffer from a negative coefficient, such as manufacturing.

Furthermore, hukou was not very statistically significant because most (nearly 98%) observations have urban local hukou and the variation of income for those who own a different type of hukou is very large. Another reason might be the fact that labor migration becomes a serious problem for the post-2002 period as increasingly more people move to big cities in China, such as Beijing and Shanghai, seeking higher income. I will provide more details on labor migration in the next section.

4.4 Heterogeneity

After estimating the returns to education, I use three different contiguity matrices, the binary contiguity matrix ($\mathbb{C}$), the first order contiguity matrix ($\mathbb{A}$) and the contiguity matrix with distance specification ($\mathbb{D}$), to estimate the spatial dependence coefficient of returns to education among given provinces in China using the spatial autoregressive model that I introduced at the end of the previous chapter. My results are demonstrated within Figure D.1-D.3. Specifications with $\mathbb{A}$ and $\mathbb{D}$ produce larger coefficients than a simple binary matrix $\mathbb{C}$. Both $\mathbb{A}$ and $\mathbb{D}$ yield

estimates between $(0, 0.75)$ while $\mathbb{C}$ only permits a lower bound as $\rho \in (0, 0.3)$. From the three matrix specifications, we can see that dependence in 1995 is stationary. The spatial dependence coefficients do not quite fluctuate for all levels of education. In contrast, the coefficients in 2002 fluctuate more with a weakly increasing trend. To be precise, under all three matrix specifications, we observe the lowest ρ at the middle school level and a significant increase at the high school level. Then, the increasing trend diminishes, which is analogous to the concave utility function with a single variable.

In 1995, labor migration was not very common during China's economic transition. Moving to a new location is costly in both economic and accounting senses. In addition, limited technology resources prevented people from accessing information about the labor markets and searching for better paid jobs in other regions. As a result, people would prefer to stay with their safety and not decide to migrate to another place. However, when it came to 2002, technology was more familiar to people in China and they started to search for better career possibilities. For instance, they could use a computer and the Internet to find out higher paid positions in a different geographic region and some of them moved to the new place because of greater economic return that they anticipated from doing so. Generally speaking, for those who only complete a lower level of education, the probability for them to move to a new region is smaller than those who have higher degrees. This statement is true regardless of the selected time period. As a result, less educated people are more likely to remain in their original region and more educated people are more likely to move to a new working place, assuming

they are capable of and are indeed searching for better paid jobs. Apart from limited technology resources, inadequate construction of transportation also kept the level of labor migration in 1995. Thus, heterogeneity of returns to education were flat as we observe in Figures D.1-D.3 because people decided not to move to other geographic locations. Notice that very few people in earlier China (such as in 1995) had completed a college degree or higher. Such a shortage of labor supply created a privilege so that those highly educated people might benefit from a greater level of income. This might account for the reason some estimates in 1995 are higher than those in 2002. To sum up, from my result I believe the level of labor migration will be a significant factor in determining geographic heterogeneity of returns to education.

Chapter 5

Concluding Remarks

In this thesis, we look at the geographic heterogeneity of returns to education in China and I explore the data from the Chinese Household Income Project (CHIP) conducted in 1995 and 2002. After introducing the heterogenous returns model, I examined the endogeneity of education and selection bias, estimated the returns to education with respect to different educational attainments and use the spatial autocorrelation (SAR) model to estimate the spatial dependence coefficients of returns to different educational attainments. After that, I found that the heterogeneity was stationary in 1995 but concave in 2002 and conclude that a higher level of education will reduce heterogeneity of returns to education, assuming people are capable of searching for better paid jobs. I believe the level of labor migration will be a significant factor in determining geographic heterogeneity of returns to education.

A potential update to this thesis will be to use CHIP 2006 and reexamine

the data to check whether my finding will also be consistent with results in CHIP 2006 that contain more recent information about China's economy. In addition, including more regions, such as Shanghai and Tianjin, in the data might contribute to explaining and intepreting the results.

Appendix A

Mathematical Proofs

A.1 Sample Selection Model

This section draws from Wooldridge (2010). Recall the following assumptions: (1) Standard normal error term: given j, $u_{ij} \sim N(0,1)$; (2) $Cov(Z_{ij}, \varepsilon_{ij}) = 0$; (3) $Cov(Z_{ij}, u_{ij}) = 0$; (4) We always observe the explanatory vector Z_{ij}; (5) We always observe the binary response ω_{ij}; (6) $\mathbb{E}(\varepsilon_{ij}|u_{ij}) = \pi_j u_{ij}$. Taking expectation on equation (3.2) conditioning on Z_{ij} and u_{ij} produces the following equation:

$$\mathbb{E}(\ln y_{ij}|Z_{ij}, u_{ij}) = \mathbb{E}(\beta_j s_{ij} + \gamma_j X_{ij}^{'} + \varepsilon_{ij}|Z_{ij}, u_{ij}) \tag{A.1}$$

Then linearity of $\mathbb{E}(\cdot)$ simplifies the result to the following equation:

$$\mathbb{E}(\ln y_{ij}|Z_{ij}, u_{ij}) = \beta_j s_{ij} + \gamma_j X_{ij}^{'} + \mathbb{E}(\varepsilon_{ij}|Z_{ij}, u_{ij}) \tag{A.2}$$

Under assumptions (2), (3) and (6), $\mathbb{E}(\varepsilon_{ij}|Z_{ij}, u_{ij}) = \mathbb{E}(\varepsilon_{ij}|u_{ij}) = \pi_j u_{ij}$ thus we have the following result:

$$\mathbb{E}(\ln y_{ij}|Z_{ij}, u_{ij}) = \beta_j s_{ij} + \gamma_j X_{ij}^{'} + \pi_j u_{ij} \tag{A.3}$$

Furthermore, using the law of iteration expectation with nested conditioning and taking the expectation on equation (A.3) conditioning on ω_{ij} gives us the following equation:

$$\mathbb{E}[\mathbb{E}(\ln y_{ij}|Z_{ij}, u_{ij})|\omega_{ij}] = \mathbb{E}(\ln y_{ij}|Z_{ij}, \omega_{ij}) = \beta_j s_{ij} + \gamma_j X_{ij}^{'} + \pi_j \mathbb{E}(u_{ij}|\omega_{ij}) \tag{A.4}$$

We want to estimate $\mathbb{E}(\ln y_{ij}|Z_{ij}, \omega_{ij} = 1) = \beta_j s_{ij} + \gamma_j X_{ij}^{'} + \pi_j \mathbb{E}(u_{ij}|\omega_{ij} = 1)$ and $\omega_{ij} = 1$ implies that $\omega_{ij}^{*} = \delta_j Z_{ij}^{'} + u_{ij} > 0$. Thus, we have the following:

$$\mathbb{E}(u_{ij}|\omega_{ij} = 1) = \mathbb{E}(u_{ij}|\omega_{ij}^{*} > 0) = \mathbb{E}(u_{ij}|u_{ij} > -\delta_j Z_{ij}^{'}) \tag{A.5}$$

Now we need to consider the following lemma about the inverse Mill's ratio:

Lemma 1. Given $u_{ij} \sim N(0, 1)$, we have $\lambda(-\delta_j Z_{ij}^{'}) = \mathbb{E}(u_{ij}|u_{ij} > -\delta_j Z_{ij}^{'})$.

Proof. We will prove this lemma here. By definition of conditional expectation:

$$\mathbb{E}(u_{ij}|u_{ij} > -\delta_j Z_{ij}^{'}) = \int_{\sigma=-\delta_j Z_{ij}^{'}}^{\infty} \sigma \cdot f(\sigma|u_{ij} > -\delta_j Z_{ij}^{'}) d\sigma \tag{A.6}$$

where $f(\sigma|U_{ij} > -\delta_j Z'_{ij})$ is the density conditioning on $u_{ij} > -\delta_i Z'_{ij}$. Then we can express equation (A.6) as follows:

$$\int_{-\delta_j Z'_{ij}}^{\infty} \sigma \cdot f(\sigma|u_{ij} > -\delta_j Z'_{ij}) d\sigma = \int_{-\delta_j Z'_{ij}}^{\infty} \sigma \cdot \frac{\partial}{\partial \sigma}[F(\sigma|u_{ij} > -\delta_j Z'_{ij})] d\sigma \quad \text{(A.7)}$$

where $F(\sigma|u_{ij} > -\delta_j Z'_{ij})$ is the c.d.f. conditioning on $u_{ij} > -\delta_i Z'_{ij}$. By definition of c.d.f., $F(\sigma|u_{ij} > -\delta_j Z'_{ij}) = \Pr(z < \sigma|u_{ij} > -\delta_j Z'_{ij})$ where $z \sim N(0,1)$ and $F(\sigma|u_{ij} > -\delta_j Z'_{ij}) = \Pr(z < \sigma \cap u_{ij} > -\delta_j Z'_{ij})/\Pr(u_{ij} > -\delta_j Z'_{ij})$ by Baye's Rule. Since u_{ij} and z have the same distribution, we simplify equation (A.7) as follows:

$$\mathbb{E}(u_{ij}|u_{ij} > -\delta_j Z'_{ij}) = \int_{-\delta_j Z'_{ij}}^{\infty} \sigma \frac{\partial}{\partial \sigma} \left(\frac{\Pr(-\delta_j Z'_{ij} < u_{ij} < \sigma)}{1 - \Pr(u_{ij} < -\delta j Z'_{ij})} \right) d\sigma \quad \text{(A.8)}$$

$$= \int_{-\delta_j Z'_{ij}}^{\infty} \sigma \frac{\partial}{\partial \sigma} \left(\frac{\int_{-\delta_j Z'_{ij}}^{\sigma} \phi(z) dz}{1 - \Phi(-\delta j Z'_{ij})} \right) d\sigma = \int_{-\delta_j Z'_{ij}}^{\infty} \sigma \frac{\partial}{\partial \sigma} \left(\frac{\Phi(\sigma) - \Phi(-\delta_j Z'_{ij})}{1 - \Phi(-\delta j Z'_{ij})} \right) d\sigma \quad \text{(A.9)}$$

Since $\partial \Phi(\sigma)/\partial \sigma = \phi(\sigma)$ and $\Phi(-\delta_j Z'_{ij})$ does not depend on σ, equation (A.9) is equivalent to the following result:

$$\mathbb{E}(u_{ij}|u_{ij} > -\delta_j Z'_{ij}) = \int_{-\delta_j Z'_{ij}}^{\infty} \frac{\phi(\sigma)\sigma}{1 - \Phi(-\delta j Z'_{ij})} d\sigma = \frac{1}{1 - \Phi(-\delta j Z'_{ij})} \int_{-\delta_j Z'_{ij}}^{\infty} \phi(\sigma)\sigma d\sigma \quad \text{(A.10)}$$

Now, we can directly evaluate the integral in equation (A.10) using the method of direct substitution:

$$\int_{-\delta_j Z'_{ij}}^{\infty} \phi(\sigma)\sigma d\sigma = \int_{-\delta_j Z'_{ij}}^{\infty} \sigma \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\sigma^2} d\sigma = \frac{1}{\sqrt{2\pi}} \int_{-\delta_j Z'_{ij}}^{\infty} \sigma e^{-\frac{1}{2}\sigma^2} d\sigma \tag{A.11}$$

Let $v = -\frac{1}{2}\sigma^2$, then $dv = -\sigma d\sigma$ and equation (A.11) simplifies to the following:

$$-\frac{1}{\sqrt{2\pi}} \int_{\sigma=-\delta_j Z'_{ij}}^{\infty} e^v dv = -\left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\sigma^2}\right)_{\sigma=-\delta_j Z'_{ij}}^{\infty} = -\left(\phi(\sigma)\right)_{\delta_j Z'_{ij}}^{\infty} \tag{A.12}$$

This yields the result that $-\left(\phi(\sigma)\right)_{\delta_j Z'_{ij}}^{\infty} = -\lim_{\epsilon\to\infty} \phi(\epsilon) + \phi(-\delta_j Z'_{ij}) = \phi(-\delta_j Z'_{ij}) = \phi(\delta_j Z'_{ij})$ since $\phi(\cdot)$ is symmetric about $z = 0$. In addition, $\forall x \in \mathbb{R}$, we have $\Phi(x) + \Phi(-x) = 1$. Thus equation (A.12) is equivalent to the following equation:

$$\mathbb{E}(u_{ij} | u_{ij} > -\delta_j Z'_{ij}) = \frac{\phi(\delta_j Z'_{ij})}{\Phi(\delta_j Z'_{ij})} = \lambda(\delta_j Z'_{ij}) \tag{A.13}$$

This completes the proof. □

Hence, the inverse Mill's ratio is included in our regression analysis and we will estimate the model:

$$\mathbb{E}(\ln y_{ij} | Z_{ij}, \omega_{ij} = 1) = \beta_j s_{ij} + \gamma_j X'_{ij} + \pi_j \lambda(\delta_j Z'_{ij}) \tag{A.14}$$

A.2 Maximum Likelihood Estimation

To estimate δ_j, we use maximum likelihood estimation. First of all, we seek the likelihood function for $(X_{ij}, Z_{ij}, \omega_{ij})$. As a binary response, the ω_{ij} is defined for the entire sample as $\omega_{ij} = 1$ if we observe labor income and $\omega_{ij} = 0$ if we do not observe it. Then we consider the likelihood function. For $\omega_{ij} = 1$ $(\omega_{ij}^* > 0)$, its probability is given by the following function:

$$\Pr(\omega_{ij} = 1|Z_{ij}) = \Pr(\omega_{ij}^* > 0|Z_{ij}) = \Phi(\delta_j Z_{ij}^{'}) \tag{A.15}$$

And for $\omega_{ij} = 0$ $(\omega_{ij}* \leq 0)$, the probability is given by the following function:

$$\Pr(\omega_{ij} = 0|Z_{ij}) = 1 - \Pr(\omega_{ij} = 1|Z_{ij}) = 1 - \Phi(\delta_j Z_{ij}^{'}) \tag{A.16}$$

For the sample with a total number of n observations, $\exists N \in \mathbb{N}$ such that given j, $\forall i \geq N+1$, $\omega_{ij} = 0$ and $\forall i \in \{1, 2, 3, ..., N\}$, $\omega_{ij} = 1$. Consequently, the likelihood function has the following functional form:

$$\mathcal{L}_n(\delta_j) = \mathcal{L}_n(\delta_j; X_{ij}, Z_{ij}) = \prod_{i=1}^{N} \Pr(\omega_{ij} = 1|Z_{ij}) \prod_{i=N+1} \Pr(\omega_{ij} = 0|Z_{ij}) \tag{A.17}$$

Thus, the log likelihood function is given by:

$$\ln \mathcal{L}_n(\delta_j) = \sum_{i=1}^{N} \ln[\Phi(\delta_j Z_{ij}^{'})] + \sum_{i=N+1} \ln[1 - \Phi(\delta_j Z_{ij}^{'})] \tag{A.18}$$

The first order condition with respect to equation (A.18) produces the following result:

$$\frac{\partial \ln \mathcal{L}(\delta_j)}{\partial \delta_j} = \sum_{i=1}^{N} \frac{\partial \ln[\Phi(\delta_j Z_{ij}^{'})]}{\partial \delta_j} + \sum_{i=N+1} \frac{\partial \ln[1-\Phi(\delta_j Z_{ij}^{'})]}{\partial \delta_{ij}} = 0 \tag{A.19}$$

Using chain rule, we have the following result:

$$\sum_{i=1}^{N} \frac{\phi(\delta_j Z_{ij}^{'})}{\Phi(\delta_j Z_{ij}^{'})} Z_{ij}^{'} - \sum_{i=N+1} \frac{\phi(\delta_j Z_{ij}^{'})}{1-\Phi(\delta_j Z_{ij}^{'})} Z_{ij}^{'} = 0 \tag{A.20}$$

Hence, we need to find δ_{ij}^{*} such that equation (A.20) is satisfied. In addition, $\lambda(\delta_{ij}^{*} Z_{ij}^{'})$ produces an estimate of the inverse Mill's ratio that we need to include in our regression analysis. Thus an estimate of δ_j is given by the equation:

$$\hat{\delta_j} = \underset{\delta_j}{\operatorname{argmax}} \left(\sum_{i=1}^{N} \ln[\Phi(\delta_j Z_{ij}^{'})] + \sum_{i=N+1} \ln[1-\Phi(\delta_j Z_{ij}^{'})] \right) \tag{A.21}$$

where Bierens (2005) suggested a small Θ_{δ_j} for δ_j might help to guarantee the uniqueness of the solution to equation (A.21).

Appendix B

Map of China

FIGURE B.1: Map of China

Appendix C

Summary of Tables

*Regions highlighted in blue are shown in data.

Table C.1: Descriptive Statistics (General)

Region	Code	Capital city	Population in 2000	Population in 2010
Anhui	34	Hefei	58,999,948	59,500,510
Beijing	11	N/A	13,569,194	19,612,368
Chongqing	50	N/A	30,512,763	28,846,170
Gansu	62	Lanzhou	25,124,282	25,575,254
Guangdong	44	Guangzhou	85,225,007	104,303,132
Guaangxi	45	Nanning	43,854,,538	46,026,629
Guizhou	52	Guiyang	35,247,695	34,746,468
Hainan	46	Haikou	7,559,035	8,671,518
Heibei	13	Shijiazhuang	66,684,419	71,854,202
Henan	41	Zhengzhou	91,234,854	94,023,567
Heilongjiang	23	Harbin	36,237,576	38,312,224
Hubei	42	Wuhan	59,508,870	57,237,740
Hunan	43	Changsha	63,274,173	65,683,722
Jiangsu	32	Nanjing	73,043,577	78,659,903
Jiangxi	36	Nanchang	40,397,598	44,567,475
Jilin	22	Changchun	26,802,191	27,462,297
Liaoning	21	Shenyang	41,824,412	43,746,323
Nei Mengol	15	Huhehaote	23,323,347	24,706,321
Qinghai	63	Xining	4,822,963	5,626,722
Ningxia	64	Yinchuan	5,486,393	6,301,350
Shaanxi	61	Xi'an	35,365,072	37,327,378
Shandong	37	Jinan	89,971,789	95,793,065
Shanghai	31	N/A	16,407,734	23,019,148
Shanxi	14	Taiyuan	32,471,242	35,712,111
Sichuan	51	Chengdu	82,348,296	80,418,200
Tianjin	12	N/A	9,848,731	12,938,224
Xinjiang	65	Wulumuqi	18,459,511	21,813,334
Xizang	54	Lhasa	2,616,329	3,002,166
Yunnan	53	Kunming	42,360,089	45,966,239
Zhejiang	33	Hangzhou	45,930,651	54,426,891
Total			1,242,612,226	1,339,724,852

Table C.2: Descriptive Statistics (1995)

Variable	Observation	Mean	Std. Dev.	Min.	Max.
ln(income)	15983	8.4462	0.9770	0.693	11.704
Schooling	19910	9.4994	3.6314	0	24
Experience	15499	21.6471	11.1845	0	49
Age	21691	35.6588	18.5361	0	100
Male	21698	0.4959	0.4999	0	1
Beijing	21698	0.0705	0.2559	0	1
Shanxi	21698	0.0972	0.2963	0	1
Liaoning	21698	0.1019	0.3026	0	1
Jiangsu	21698	0.1129	0.3165	0	1
Anhui	21698	0.0704	0.2558	0	1
Henan	21698	0.0894	0.2853	0	1
Hubei	21698	0.1065	0.3084	0	1
Guangdong	21698	0.0839	0.2773	0	1
Sichuan	21698	0.1146	0.3185	0	1
Yunnan	21698	0.0927	0.2899	0	1
Gansu	21698	0.0601	0.2377	0	1
Farm & ...	14138	0.0167	0.1281	0	1
Manufacturing	14138	0.4121	0.4922	0	1
Mining &...	14138	0.0110	0.1041	0	1
Construction	14138	0.0296	0.1694	0	1
Transportation &...	14138	0.0484	0.2146	0	1
Wholesale &...	14138	0.1412	0.3482	0	1
Real Estate &...	14138	0.0388	0.1930	0	1
Health &...	14138	0.0479	0.2135	0	1
Education	14138	0.0743	0.2623	0	1
Scientifc &...	14138	0.0219	0.1462	0	1
Finance & Insurance	14138	0.0177	0.1318	0	1
Government &...	14138	0.1129	0.3165	0	1
Other & ...	14138	0.1777	0.1035	0	1

Table C.3: Descriptive Statistics (2002)

Variable	Observation	Mean	Std. Dev.	Min.	Max.
ln(income)	14971	9.010	0.8721	2.7081	11.9829
Schooling	20439	9.5483	4.2325	0	23
Experience	10239	20.0893	9.7377	0	43
Age	20547	38.2410	18.1427	0	99
Male	20547	0.4936	0.4999	0	1
Beijing	20439	0.0708	0.2564	0	1
Shanxi	20439	0.0898	0.2859	0	1
Liaoning	20439	0.1032	0.3042	0	1
Jiangsu	20439	0.1054	0.3067	0	1
Anhui	20439	0.0716	0.2579	0	1
Henan	20439	0.1014	0.3019	0	1
Hubei	20439	0.1006	0.3008	0	1
Guangdong	20439	0.0860	0.2803	0	1
Chongqing	20439	0.0405	0.1970	0	1
Sichuan	20439	0.0829	0.2759	0	1
Yunnan	20439	0.0898	0.2859	0	1
Gansu	20439	0.0582	0.2341	0	1
Local Urban Hukou	20547	0.9774	0.1487	0	1
Local Rural Hukou	20547	0.0122	0.1096	0	1
Alien Urban Hukou	20547	0.0056	0.0746	0	1
Alien Rural Hukou	20547	0.0049	0.0696	0	1
Farm &...	10239	0.0122	0.0891	0	1
Mineral	10239	0.0154	0.1233	0	1
Manufacturing	10239	0.2492	0.4326	0	1
Electricity &...	10239	0.0324	0.1771	0	1
Construction	10239	0.0326	0.1776	0	1
Geological &...	10239	0.0081	0.0896	0	1
Transportation &...	10239	0.0782	0.2685	0	1
Wholesale &...	10239	0.1227	0.328	0	1
Finance &...	10239	0.0269	0.1617	0	1
Real Estate	10239	0.0120	0.1089	0	1
Social Services	10239	0.1025	0.3034	0	1
Health, Sports &...	10239	0.0509	0.2198	0	1
Education &...	10239	0.0897	0.2857	0	1
Scientific &...	10239	0.0174	0.1307	0	1
Government &...	10239	0.1195	0.3244	0	1
Other	10239	0.0222	0.1472	0	1

TABLE C.4: Euclidean Distances Between Capital Cities

(kilometers)	Beijing	Taiyuan	Shenyang	Nanjing	Hefei	Zhengzhou	Wuhan	Guangzhou	Chongqing	Chengdu	Kunming	Lanzhou
Beijing	0	406.24	621.17	900.19	898.27	617.50	1049.90	1888.76	1465.23	1520.88	2086.07	1187.23
Taiyuan	406.24	0	1023.59	863.79	796.03	355.78	822.60	1639.55	1081.09	1117.80	1698.91	802.92
Shenyang	621.17	1023.59	0	1162.36	1234.86	1150.89	1486.13	2282.85	2040.28	2124.25	2657.84	1806.60
Nanjing	900.19	863.79	1162.36	0	146.10	566.99	454.00	1131.81	1206.67	1408.09	1742.67	1452.10
Hefei	898.27	796.03	1234.86	146.10	0	468.41	312.19	1048.06	1060.72	1262.67	1601.98	1327.04
Zhengzhou	617.50	355.78	1150.89	566.99	468.41	0	469.11	1297.99	892.83	1010.38	1507.49	905.85
Wuhan	1049.90	822.60	1486.13	454.00	312.19	469.11	0	839.21	761.58	981.95	1289.81	1148.98
Guangzhou	1888.76	1639.55	2282.85	1131.81	1048.06	1297.99	839.21	0	976.50	1233.90	1074.47	1699.24
Chongqing	1465.23	1081.09	2040.28	1206.67	1060.72	892.83	761.58	976.50	0	266.41	620.92	767.73
Chengdu	1520.88	1117.80	2124.25	1408.09	1262.67	1010.38	981.95	1233.90	266.41	0	639.59	600.96
Kunming	2086.07	1698.91	2657.84	1742.67	1601.98	1507.49	1289.81	1074.47	620.92	639.59	0	1232.07
Lanzhou	1187.23	802.92	1806.60	1452.10	1327.04	905.85	1148.98	1699.24	767.73	600.96	1232.07	0

TABLE C.5: Estimation Results (Beijing, 1995)

	Estimation Results in 1995			
	OLS		Selection	
Variables	Coef.	Std. Err.	Coef.	Std. Err.
Intercept	7.5578	0.0746	7.6425	0.0823
Middle School	0.3053	0.0848	0.2914	0.0961
High School	0.5083	0.0849	0.4926	0.0961
College	0.6636	0.0937	0.6374	0.1196
Higher	0.6672	0.1416	0.6453	0.2019
Male	0.3112	0.0303	0.3029	0.0226
Male*Middle	-0.1295	0.0341	-0.1232	0.0296
Male*High	-0.1840	0.0338	-0.1779	0.0297
Male*College	-0.2316	0.0411	-0.2226	0.0422
Male*Higher	-0.2621	0.0651	-0.2523	0.0789
Experience	0.0576	0.0020	0.0528	0.0018
Experience^2	-0.0008	0.00004	-0.0008	0.00004
Farm & ...	0.1079	0.0782	0.1095	0.0632
Manufacturing	0.0060	0.0701	0.0097	0.0519
Mining &...	0.1551	0.0777	0.1564	0.0694
Construction	0.0664	0.0743	0.0673	0.0586
Transportation &...	0.1396	0.0734	0.1432	0.0558
Wholesale &...	0.0180	0.0711	0.0206	0.0530
Real Estate &...	0.0287	0.0752	0.0290	0.0569
Health &...	0.1578	0.0720	0.1584	0.0560
Education	0.1619	0.0710	0.1644	0.0546
Scientifc &...	0.2306	0.0726	0.2342	0.0610
Finance & Insurance	0.4146	0.0764	0.4183	0.6276
Government &...	0.1804	0.0709	0.1793	0.0535

TABLE C.6: Estimation Results (Shanxi, 1995)

	Estimation Results in 1995			
	OLS		Selection	
Variables	Coef.	Std. Err.	Coef.	Std. Err.
Intercept	7.2015	0.0664	7.2893	0.0602
Middle School	0.2984	0.0778	0.2765	0.0739
High School	0.4445	0.0783	0.4153	0.0746
College	0.5218	0.0843	0.4820	0.1056
Higher	0.6093	0.1292	0.5754	0.1983
Male	0.3112	0.0303	0.3029	0.0226
Male*Middle	-0.1295	0.0341	-0.1232	0.0296
Male*High	-0.1840	0.0338	-0.1779	0.0297
Male*College	-0.2316	0.0411	-0.2226	0.0422
Male*Higher	-0.2621	0.0651	-0.2523	0.0789
Experience	0.0576	0.0020	0.0528	0.0018
Experience^2	-0.0008	0.00004	-0.0008	0.00004
Farm & ...	0.1079	0.0782	0.1095	0.0632
Manufacturing	0.0060	0.0701	0.0097	0.0519
Mining &...	0.1551	0.0777	0.1564	0.0694
Construction	0.0664	0.0743	0.0673	0.0586
Transportation &...	0.1396	0.0734	0.1432	0.0558
Wholesale &...	0.0180	0.0711	0.0206	0.0530
Real Estate &...	0.0287	0.0752	0.0290	0.0569
Health &...	0.1578	0.0720	0.1584	0.0560
Education	0.1619	0.0710	0.1644	0.0546
Scientifc &...	0.2306	0.0726	0.2342	0.0610
Finance & Insurance	0.4146	0.0764	0.4183	0.6276
Government &...	0.1804	0.0709	0.1793	0.0535

Table C.7: Estimation Results (Liaoning, 1995)

	Estimation Results in 1995			
	OLS		Selection	
Variables	Coef.	Std. Err.	Coef.	Std. Err.
Intercept	7.2475	0.0834	7.3392	0.0694
Middle School	0.2931	0.0916	0.2716	0.0802
High School	0.5045	0.925	0.4776	0.0821
College	0.6109	0.0993	0.5791	0.1079
Higher	0.8271	0.1464	0.7923	0.2097
Male	0.3112	0.0303	0.3029	0.0226
Male*Middle	-0.1295	0.0341	-0.1232	0.0296
Male*High	-0.1840	0.0338	-0.1779	0.0297
Male*College	-0.2316	0.0411	-0.2226	0.0422
Male*Higher	-0.2621	0.0651	-0.2523	0.0789
Experience	0.0576	0.0020	0.0528	0.0018
Experience^2	-0.0008	0.00004	-0.0008	0.00004
Farm & ...	0.1079	0.0782	0.1095	0.0632
Manufacturing	0.0060	0.0701	0.0097	0.0519
Mining &...	0.1551	0.0777	0.1564	0.0694
Construction	0.0664	0.0743	0.0673	0.0586
Transportation &...	0.1396	0.0734	0.1432	0.0558
Wholesale &...	0.0180	0.0711	0.0206	0.0530
Real Estate &...	0.0287	0.0752	0.0290	0.0569
Health &...	0.1578	0.0720	0.1584	0.0560
Education	0.1619	0.0710	0.1644	0.0546
Scientifc &...	0.2306	0.0726	0.2342	0.0610
Finance & Insurance	0.4146	0.0764	0.4183	0.6276
Government &...	0.1804	0.0709	0.1793	0.0535

Table C.8: Estimation Results (Jiangsu, 1995)

	Estimation Results in 1995			
	OLS		Selection	
Variables	Coef.	Std. Err.	Coef.	Std. Err.
Intercept	7.5372	0.0627	7.6279	0.0612
Middle School	0.2501	0.0721	0.2291	0.0734
High School	0.3958	0.0738	0.3715	0.0747
College	0.5738	0.0785	0.5425	0.1027
Higher	0.6435	0.1335	0.6149	0.1888
Male	0.3112	0.0303	0.3029	0.0226
Male*Middle	-0.1295	0.0341	-0.1232	0.0296
Male*High	-0.1840	0.0338	-0.1779	0.0297
Male*College	-0.2316	0.0411	-0.2226	0.0422
Male*Higher	-0.2621	0.0651	-0.2523	0.0789
Experience	0.0576	0.0020	0.0528	0.0018
Experience^2	-0.0008	0.00004	-0.0008	0.00004
Farm & ...	0.1079	0.0782	0.1095	0.0632
Manufacturing	0.0060	0.0701	0.0097	0.0519
Mining &...	0.1551	0.0777	0.1564	0.0694
Construction	0.0664	0.0743	0.0673	0.0586
Transportation &...	0.1396	0.0734	0.1432	0.0558
Wholesale &...	0.0180	0.0711	0.0206	0.0530
Real Estate &...	0.0287	0.0752	0.0290	0.0569
Health &...	0.1578	0.0720	0.1584	0.0560
Education	0.1619	0.0710	0.1644	0.0546
Scientifc &...	0.2306	0.0726	0.2342	0.0610
Finance & Insurance	0.4146	0.0764	0.4183	0.6276
Government &...	0.1804	0.0709	0.1793	0.0535

Table C.9: Estimation Results (Anhui, 1995)

	Estimation Results in 1995			
	OLS		Selection	
Variables	Coef.	Std. Err.	Coef.	Std. Err.
Intercept	7.174	0.0678	7.2631	0.0732
Middle School	0.3352	0.0743	0.3161	0.0848
High School	0.4846	0.0798	0.4605	0.0865
College	0.5946	0.0957	0.5637	0.1156
Higher	0.8813	0.1750	0.8511	0.2112
Male	0.3112	0.0303	0.3029	0.0226
Male*Middle	-0.1295	0.0341	-0.1232	0.0296
Male*High	-0.1840	0.0338	-0.1779	0.0297
Male*College	-0.2316	0.0411	-0.2226	0.0422
Male*Higher	-0.2621	0.0651	-0.2523	0.0789
Experience	0.0576	0.0020	0.0528	0.0018
Experience^2	-0.0008	0.00004	-0.0008	0.00004
Farm & ...	0.1079	0.0782	0.1095	0.0632
Manufacturing	0.0060	0.0701	0.0097	0.0519
Mining &...	0.1551	0.0777	0.1564	0.0694
Construction	0.0664	0.0743	0.0673	0.0586
Transportation &...	0.1396	0.0734	0.1432	0.0558
Wholesale &...	0.0180	0.0711	0.0206	0.0530
Real Estate &...	0.0287	0.0752	0.0290	0.0569
Health &...	0.1578	0.0720	0.1584	0.0560
Education	0.1619	0.0710	0.1644	0.0546
Scientifc &...	0.2306	0.0726	0.2342	0.0610
Finance & Insurance	0.4146	0.0764	0.4183	0.6276
Government &...	0.1804	0.0709	0.1793	0.0535

Table C.10: Estimation Results (Henan, 1995)

	Estimation Results in 1995			
	OLS		Selection	
Variables	Coef.	Std. Err.	Coef.	Std. Err.
Intercept	7.2355	0.0653	7.3192	0.0610
Middle School	0.2053	0.0743	0.1911	0.0742
High School	0.3637	0.0759	0.3447	0.0761
College	0.5308	0.0775	0.5056	0.1075
Higher	0.5314	0.0859	0.4912	0.2251
Male	0.3112	0.0303	0.3029	0.0226
Male*Middle	-0.1295	0.0341	-0.1232	0.0296
Male*High	-0.1840	0.0338	-0.1779	0.0297
Male*College	-0.2316	0.0411	-0.2226	0.0422
Male*Higher	-0.2621	0.0651	-0.2523	0.0789
Experience	0.0576	0.0020	0.0528	0.0018
Experience^2	-0.0008	0.00004	-0.0008	0.00004
Farm & ...	0.1079	0.0782	0.1095	0.0632
Manufacturing	0.0060	0.0701	0.0097	0.0519
Mining &...	0.1551	0.0777	0.1564	0.0694
Construction	0.0664	0.0743	0.0673	0.0586
Transportation &...	0.1396	0.0734	0.1432	0.0558
Wholesale &...	0.0180	0.0711	0.0206	0.0530
Real Estate &...	0.0287	0.0752	0.0290	0.0569
Health &...	0.1578	0.0720	0.1584	0.0560
Education	0.1619	0.0710	0.1644	0.0546
Scientifc &...	0.2306	0.0726	0.2342	0.0610
Finance & Insurance	0.4146	0.0764	0.4183	0.6276
Government &...	0.1804	0.0709	0.1793	0.0535

Table C.11: Estimation Results (Hubei, 1995)

	Estimation Results in 1995			
	OLS		Selection	
Variables	Coef.	Std. Err.	Coef.	Std. Err.
Intercept	7.4760	0.0647	7.5641	0.0645
Middle School	0.1668	0.0745	0.1497	0.0767
High School	0.3114	0.0752	0.2894	0.0778
College	0.4816	0.0823	0.4541	0.1054
Higher	0.5467	0.1225	0.5186	0.1852
Male	0.3112	0.0303	0.3029	0.0226
Male*Middle	-0.1295	0.0341	-0.1232	0.0296
Male*High	-0.1840	0.0338	-0.1779	0.0297
Male*College	-0.2316	0.0411	-0.2226	0.0422
Male*Higher	-0.2621	0.0651	-0.2523	0.0789
Experience	0.0576	0.0020	0.0528	0.0018
Experience^2	-0.0008	0.00004	-0.0008	0.00004
Farm & ...	0.1079	0.0782	0.1095	0.0632
Manufacturing	0.0060	0.0701	0.0097	0.0519
Mining &...	0.1551	0.0777	0.1564	0.0694
Construction	0.0664	0.0743	0.0673	0.0586
Transportation &...	0.1396	0.0734	0.1432	0.0558
Wholesale &...	0.0180	0.0711	0.0206	0.0530
Real Estate &...	0.0287	0.0752	0.0290	0.0569
Health &...	0.1578	0.0720	0.1584	0.0560
Education	0.1619	0.0710	0.1644	0.0546
Scientifc &...	0.2306	0.0726	0.2342	0.0610
Finance & Insurance	0.4146	0.0764	0.4183	0.6276
Government &...	0.1804	0.0709	0.1793	0.0535

Table C.12: Estimation Results (Guangdong, 1995)

	Estimation Results in 1995			
	OLS		Selection	
Variables	Coef.	Std. Err.	Coef.	Std. Err.
Intercept	7.8700	0.0766	7.9669	0.0595
Middle School	0.3733	0.0937	0.3498	0.0728
High School	0.5093	0.0968	0.4794	0.0741
College	0.8506	0.1241	0.8116	0.1086
Higher	1.0676	0.1685	1.0291	0.1904
Male	0.3112	0.0303	0.3029	0.0226
Male*Middle	-0.1295	0.0341	-0.1232	0.0296
Male*High	-0.1840	0.0338	-0.1779	0.0297
Male*College	-0.2316	0.0411	-0.2226	0.0422
Male*Higher	-0.2621	0.0651	-0.2523	0.0789
Experience	0.0576	0.0020	0.0528	0.0018
Experience^2	-0.0008	0.00004	-0.0008	0.00004
Farm & ...	0.1079	0.0782	0.1095	0.0632
Manufacturing	0.0060	0.0701	0.0097	0.0519
Mining &...	0.1551	0.0777	0.1564	0.0694
Construction	0.0664	0.0743	0.0673	0.0586
Transportation &...	0.1396	0.0734	0.1432	0.0558
Wholesale &...	0.0180	0.0711	0.0206	0.0530
Real Estate &...	0.0287	0.0752	0.0290	0.0569
Health &...	0.1578	0.0720	0.1584	0.0560
Education	0.1619	0.0710	0.1644	0.0546
Scientifc &...	0.2306	0.0726	0.2342	0.0610
Finance & Insurance	0.4146	0.0764	0.4183	0.6276
Government &...	0.1804	0.0709	0.1793	0.0535

Table C.13: Estimation Results (Sichuan, 1995)

	Estimation Results in 1995			
	OLS		Selection	
Variables	Coef.	Std. Err.	Coef.	Std. Err.
Intercept	7.3237	0.0651	7.4159	0.0612
Middle School	0.3232	0.0749	0.3014	0.0732
High School	0.4272	0.0766	0.3996	0.0747
College	0.5904	0.0854	0.5538	0.1016
Higher	0.6930	0.1227	0.6633	0.1841
Male	0.3112	0.0303	0.3029	0.0226
Male*Middle	-0.1295	0.0341	-0.1232	0.0296
Male*High	-0.1840	0.0338	-0.1779	0.0297
Male*College	-0.2316	0.0411	-0.2226	0.0422
Male*Higher	-0.2621	0.0651	-0.2523	0.0789
Experience	0.0576	0.0020	0.0528	0.0018
Experience^2	-0.0008	0.00004	-0.0008	0.00004
Farm & ...	0.1079	0.0782	0.1095	0.0632
Manufacturing	0.0060	0.0701	0.0097	0.0519
Mining &...	0.1551	0.0777	0.1564	0.0694
Construction	0.0664	0.0743	0.0673	0.0586
Transportation &...	0.1396	0.0734	0.1432	0.0558
Wholesale &...	0.0180	0.0711	0.0206	0.0530
Real Estate &...	0.0287	0.0752	0.0290	0.0569
Health &...	0.1578	0.0720	0.1584	0.0560
Education	0.1619	0.0710	0.1644	0.0546
Scientifc &...	0.2306	0.0726	0.2342	0.0610
Finance & Insurance	0.4146	0.0764	0.4183	0.6276
Government &...	0.1804	0.0709	0.1793	0.0535

Table C.14: Estimation Results (Yunnan, 1995)

	Estimation Results in 1995			
	OLS		Selection	
Variables	Coef.	Std. Err.	Coef.	Std. Err.
Intercept	7.4639	0.0620	7.5547	0.0639
Middle School	0.1942	0.0728	0.1739	0.0768
High School	0.3300	0.0742	0.3039	0.0774
College	0.3717	0.0876	0.3381	0.1088
Higher	0.4840	0.1154	0.4594	0.2009
Male	0.3112	0.0303	0.3029	0.0226
Male*Middle	-0.1295	0.0341	-0.1232	0.0296
Male*High	-0.1840	0.0338	-0.1779	0.0297
Male*College	-0.2316	0.0411	-0.2226	0.0422
Male*Higher	-0.2621	0.0651	-0.2523	0.0789
Experience	0.0576	0.0020	0.0528	0.0018
Experience^2	-0.0008	0.00004	-0.0008	0.00004
Farm & ...	0.1079	0.0782	0.1095	0.0632
Manufacturing	0.0060	0.0701	0.0097	0.0519
Mining &...	0.1551	0.0777	0.1564	0.0694
Construction	0.0664	0.0743	0.0673	0.0586
Transportation &...	0.1396	0.0734	0.1432	0.0558
Wholesale &...	0.0180	0.0711	0.0206	0.0530
Real Estate &...	0.0287	0.0752	0.0290	0.0569
Health &...	0.1578	0.0720	0.1584	0.0560
Education	0.1619	0.0710	0.1644	0.0546
Scientifc &...	0.2306	0.0726	0.2342	0.0610
Finance & Insurance	0.4146	0.0764	0.4183	0.6276
Government &...	0.1804	0.0709	0.1793	0.0535

Table C.15: Estimation Results (Gansu, 1995)

	Estimation Results in 1995			
	OLS		Selection	
Variables	Coef.	Std. Err.	Coef.	Std. Err.
Intercept	7.1746	0.0988	7.2632	0.0746
Middle School	0.2788	0.0618	0.2569	0.0609
High School	0.3882	0.0648	0.3637	0.0627
College	0.6037	0.0687	0.5677	0.0873
Higher	0.7084	0.1108	0.6748	0.1693
Male	0.3112	0.0303	0.3029	0.0226
Male*Middle	-0.1295	0.0341	-0.1232	0.0296
Male*High	-0.1840	0.0338	-0.1779	0.0297
Male*College	-0.2316	0.0411	-0.2226	0.0422
Male*Higher	-0.2621	0.0651	-0.2523	0.0789
Experience	0.0576	0.0020	0.0528	0.0018
Experience^2	-0.0008	0.00004	-0.0008	0.00004
Farm & ...	0.1079	0.0782	0.1095	0.0632
Manufacturing	0.0060	0.0701	0.0097	0.0519
Mining &...	0.1551	0.0777	0.1564	0.0694
Construction	0.0664	0.0743	0.0673	0.0586
Transportation &...	0.1396	0.0734	0.1432	0.0558
Wholesale &...	0.0180	0.0711	0.0206	0.0530
Real Estate &...	0.0287	0.0752	0.0290	0.0569
Health &...	0.1578	0.0720	0.1584	0.0560
Education	0.1619	0.0710	0.1644	0.0546
Scientifc &...	0.2306	0.0726	0.2342	0.0610
Finance & Insurance	0.4146	0.0764	0.4183	0.6276
Government &...	0.1804	0.0709	0.1793	0.0535

Table C.16: Estimation Results (Beijing, 2002)

	Estimation Results in 2002					
	OLS		IV		Selection	
Variables	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Intercept	8.2598	0.3409	9.0272	0.5627	8.8945	0.2182
Middle School	0.1332	0.1412	0.0614	0.2527	0.0879	0.2253
High School	0.2889	0.1753	0.1409	0.2492	0.2381	0.2222
College	0.6105	0.1933	0.5548	0.2536	0.5511	0.2301
Higher	0.7615	0.2488	0.5397	0.2941	0.6933	0.2785
Male	0.2458	0.0453	0.0761	0.3524	0.2182	0.0446
Male*Middle	–0.0396	0.0493	0.0654	0.3578	-0.0258	0.0482
Male*High	-0.1238	0.0478	0.1413	0.3553	-0.1062	0.0471
Male*College	-0.0739	0.0539	0.0287	0.3593	-0.0551	0.0527
Male*Higher	-0.1222	0.0738	0.2228	0.4012	-0.0972	0.0801
Experience	0.0352	0.0026	0.0166	0.0064	0.0289	0.0024
Experience^2	-0.0005	0.00006	-0.0001	0.0002	-0.0004	0.00006
Local Rural Hukou	0.0177	0.3031	-0.1065	0.4930	0.0037	0.5502
Alien Urban Hukou	-0.4593	0.2958	-0.5864	0.2863	-0.4811	0.3258
Alien Rural Hukou	-0.3891	0.1373	0.1443	0.2552	0.1258	0.0842
Farm &...	0.1567	0.0546	0.1874	0.1843	0.1659	0.0586
Mineral	-0.0121	0.0547	0.1069	0.1251	-0.0028	0.5443
Manufacturing	-0.0326	0.0348	0.0916	0.1143	-0.0262	0.0336
Electricity &...	0.2679	0.0437	0.2434	0.1300	0.2738	0.0439
Construction	0.0183	0.0475	0.2189	0.1548	0.0251	0.0437
Geological &...	0.2538	0.0638	0.4127	0.2532	0.2583	0.0682
Transportation &...	0.1515	0.0391	0.3259	0.1306	0.1604	0.0373
Wholesale &...	-0.1341	0.0380	-0.0621	0.1222	-0.1252	0.0354
Finance &...	0.2718	0.0457	0.4287	0.1461	0.2774	0.0458
Real Estate	0.1953	0.0544	0.2627	0.2115	0.2049	0.0585
Social Services	-0.1432	0.0384	0.0661	0.1333	-0.1354	0.0360
Health, Sports &...	0.2700	0.0408	0.3184	0.1292	0.2773	0.0399
Education &...	0.2900	0.0372	0.3792	0.1183	0.2973	0.0368
Scientific &...	0.3281	0.0541	0.6268	0.1585	0.3322	0.0521
Government &...	0.2307	0.0358	0.3177	0.1184	0.2357	0.0356

Table C.17: Estimation Results (Shanxi, 2002)

	Estimation Results in 2002					
	OLS		IV		Selection	
Variables	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Intercept	8.3507	0.1226	8.2521	0.2493	8.4642	0.1236
Middle School	-0.0378	0.1723	0.0198	0.1209	-0.0941	0.1348
High School	0.1868	0.1706	0.2728	0.1185	0.1127	0.1309
College	0.3710	0.1764	0.4207	0.1305	0.2852	0.1443
Higher	0.5471	0.1877	0.6628	0.1962	0.4522	0.2129
Male	0.2458	0.0453	0.0761	0.3524	0.2182	0.0446
Male*Middle	–0.0396	0.0493	0.0654	0.3578	-0.0258	0.0482
Male*High	-0.1238	0.0478	0.1413	0.3553	-0.1062	0.0471
Male*College	-0.0739	0.0539	0.0287	0.3593	-0.0551	0.0527
Male*Higher	-0.1222	0.0738	0.2228	0.4012	-0.0972	0.0801
Experience	0.0352	0.0026	0.0166	0.0064	0.0289	0.0024
Experience^2	-0.0005	0.00006	-0.0001	0.0002	-0.0004	0.00006
Local Rural Hukou	-0.0923	0.1349	0.0958	0.1964	-0.1061	0.0537
Alien Urban Hukou	0.0726	0.1194	-0.2225	0.3092	0.0606	0.0804
Alien Rural Hukou	0.1428	0.1373	0.1443	0.2552	0.1256	0.0842
Farm &...	0.1567	0.0546	0.1874	0.1843	0.1659	0.0586
Mineral	-0.0121	0.0547	0.1069	0.1251	-0.0028	0.5443
Manufacturing	-0.0326	0.0348	0.0916	0.1143	-0.0262	0.0336
Electricity &...	0.2679	0.0437	0.2434	0.1300	0.2738	0.0439
Construction	0.0183	0.0475	0.2189	0.1548	0.0251	0.0437
Geological &...	0.2538	0.0638	0.4127	0.2532	0.2583	0.0682
Transportation &...	0.1515	0.0391	0.3259	0.1306	0.1604	0.0373
Wholesale &...	-0.1341	0.0380	-0.0621	0.1222	-0.1252	0.0354
Finance &...	0.2718	0.0457	0.4287	0.1461	0.2774	0.0458
Real Estate	0.1953	0.0544	0.2627	0.2115	0.2049	0.0585
Social Services	-0.1432	0.0384	0.0661	0.1333	-0.1354	0.0360
Health, Sports &...	0.2700	0.0408	0.3184	0.1292	0.2773	0.0399
Education &...	0.2900	0.0372	0.3792	0.1183	0.2973	0.0368
Scientific &...	0.3281	0.0541	0.6268	0.1585	0.3322	0.0521
Government &...	0.2307	0.0358	0.3177	0.1184	0.2357	0.0356

Table C.18: Estimation Results (Liaoning, 2002)

	Estimation Results in 2002					
	OLS		IV		Selection	
Variables	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Intercept	8.3574	0.1415	8.7476	0.3917	8.4852	0.1401
Middle School	-0.0017	0.1849	0.0085	0.1876	-0.0714	0.1487
High School	0.3399	0.1851	0.4140	0.1850	0.2597	0.1461
College	0.4334	0.1389	0.4805	0.1961	0.3424	0.1581
Higher	0.6866	0.1915	0.7699	0.2511	0.5785	0.2261
Male	0.2458	0.0453	0.0761	0.3524	0.2182	0.0446
Male*Middle	–0.0396	0.0493	0.0654	0.3578	-0.0258	0.0482
Male*High	-0.1238	0.0478	0.1413	0.3553	-0.1062	0.0471
Male*College	-0.0739	0.0539	0.0287	0.3593	-0.0551	0.0527
Male*Higher	-0.1222	0.0738	0.2228	0.4012	-0.0972	0.0801
Experience	0.0352	0.0026	0.0166	0.0064	0.0289	0.0024
Experience^2	-0.0005	0.00006	-0.0001	0.0002	-0.0004	0.00006
Local Rural Hukou	-0.0923	0.1349	0.0958	0.1964	-0.1061	0.0537
Alien Urban Hukou	0.0726	0.1194	-0.2225	0.3092	0.0606	0.0804
Alien Rural Hukou	0.1428	0.1373	0.1443	0.2552	0.1256	0.0842
Farm &...	0.1567	0.0546	0.1874	0.1843	0.1659	0.0586
Mineral	-0.0121	0.0547	0.1069	0.1251	-0.0028	0.5443
Manufacturing	-0.0326	0.0348	0.0916	0.1143	-0.0262	0.0336
Electricity &...	0.2679	0.0437	0.2434	0.1300	0.2738	0.0439
Construction	0.0183	0.0475	0.2189	0.1548	0.0251	0.0437
Geological &...	0.2538	0.0638	0.4127	0.2532	0.2583	0.0682
Transportation &...	0.1515	0.0391	0.3259	0.1306	0.1604	0.0373
Wholesale &...	-0.1341	0.0380	-0.0621	0.1222	-0.1252	0.0354
Finance &...	0.2718	0.0457	0.4287	0.1461	0.2774	0.0458
Real Estate	0.1953	0.0544	0.2627	0.2115	0.2049	0.0585
Social Services	-0.1432	0.0384	0.0661	0.1333	-0.1354	0.0360
Health, Sports &...	0.2700	0.0408	0.3184	0.1292	0.2773	0.0399
Education &...	0.2900	0.0372	0.3792	0.1183	0.2973	0.0368
Scientific &...	0.3281	0.0541	0.6268	0.1585	0.3322	0.0521
Government &...	0.2307	0.0358	0.3177	0.1184	0.2357	0.0356

Table C.19: Estimation Results (Jiangsu, 2002)

	Estimation Results in 2002					
	OLS		IV		Selection	
Variables	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Intercept	8.3681	0.1292	9.3832	0.3267	8.4841	0.1250
Middle School	0.1403	0.1766	0.2728	0.2675	0.0790	0.1354
High School	0.3964	0.1753	0.5420	0.2424	0.3235	0.1317
College	0.5494	0.1819	0.6728	0.2499	0.5132	0.1455
Higher	0.9269	0.2057	1.0431	0.265	0.8369	0.2206
Male	0.2458	0.0453	0.0761	0.3524	0.2182	0.0446
Male*Middle	–0.0396	0.0493	0.0654	0.3578	-0.0258	0.0482
Male*High	-0.1238	0.0478	0.1413	0.3553	-0.1062	0.0471
Male*College	-0.0739	0.0539	0.0287	0.3593	-0.0551	0.0527
Male*Higher	-0.1222	0.0738	0.2228	0.4012	-0.0972	0.0801
Experience	0.0352	0.0026	0.0166	0.0064	0.0289	0.0024
Experience^2	-0.0005	0.00006	-0.0001	0.0002	-0.0004	0.00006
Local Rural Hukou	-0.0923	0.1349	0.0958	0.1964	-0.1061	0.0537
Alien Urban Hukou	0.0726	0.1194	-0.2225	0.3092	0.0606	0.0804
Alien Rural Hukou	0.1428	0.1373	0.1443	0.2552	0.1256	0.0842
Farm &...	0.1567	0.0546	0.1874	0.1843	0.1659	0.0586
Mineral	-0.0121	0.0547	0.1069	0.1251	-0.0028	0.5443
Manufacturing	-0.0326	0.0348	0.0916	0.1143	-0.0262	0.0336
Electricity &...	0.2679	0.0437	0.2434	0.1300	0.2738	0.0439
Construction	0.0183	0.0475	0.2189	0.1548	0.0251	0.0437
Geological &...	0.2538	0.0638	0.4127	0.2532	0.2583	0.0682
Transportation &...	0.1515	0.0391	0.3259	0.1306	0.1604	0.0373
Wholesale &...	-0.1341	0.0380	-0.0621	0.1222	-0.1252	0.0354
Finance &...	0.2718	0.0457	0.4287	0.1461	0.2774	0.0458
Real Estate	0.1953	0.0544	0.2627	0.2115	0.2049	0.0585
Social Services	-0.1432	0.0384	0.0661	0.1333	-0.1354	0.0360
Health, Sports &...	0.2700	0.0408	0.3184	0.1292	0.2773	0.0399
Education &...	0.2900	0.0372	0.3792	0.1183	0.2973	0.0368
Scientific &...	0.3281	0.0541	0.6268	0.1585	0.3322	0.0521
Government &...	0.2307	0.0358	0.3177	0.1184	0.2357	0.0356

Table C.20: Estimation Results (Anhui, 2002)

	Estimation Results in 2002					
	OLS		IV		Selection	
Variables	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Intercept	8.1779	0.1307	8.1769	0.3099	8.2869	0.1474
Middle School	0.1902	0.1801	0.1556	0.1702	0.1372	0.1579
High School	0.3856	0.1763	0.4024	0.1652	0.3215	0.1541
College	0.6002	0.1860	0.5899	0.1811	0.5244	0.1690
Higher	0.6698	0.2236	0.6546	0.2399	0.5831	0.2307
Male	0.2458	0.0453	0.0761	0.3524	0.2182	0.0446
Male*Middle	–0.0396	0.0493	0.0654	0.3578	-0.0258	0.0482
Male*High	-0.1238	0.0478	0.1413	0.3553	-0.1062	0.0471
Male*College	-0.0739	0.0539	0.0287	0.3593	-0.0551	0.0527
Male*Higher	-0.1222	0.0738	0.2228	0.4012	-0.0972	0.0801
Experience	0.0352	0.0026	0.0166	0.0064	0.0289	0.0024
Experience^2	-0.0005	0.00006	-0.0001	0.0002	-0.0004	0.00006
Local Rural Hukou	-0.0923	0.1349	0.0958	0.1964	-0.1061	0.0537
Alien Urban Hukou	0.0726	0.1194	-0.2225	0.3092	0.0606	0.0804
Alien Rural Hukou	0.1428	0.1373	0.1443	0.2552	0.1256	0.0842
Farm &...	0.1567	0.0546	0.1874	0.1843	0.1659	0.0586
Mineral	-0.0121	0.0547	0.1069	0.1251	-0.0028	0.5443
Manufacturing	-0.0326	0.0348	0.0916	0.1143	-0.0262	0.0336
Electricity &...	0.2679	0.0437	0.2434	0.1300	0.2738	0.0439
Construction	0.0183	0.0475	0.2189	0.1548	0.0251	0.0437
Geological &...	0.2538	0.0638	0.4127	0.2532	0.2583	0.0682
Transportation &...	0.1515	0.0391	0.3259	0.1306	0.1604	0.0373
Wholesale &...	-0.1341	0.0380	-0.0621	0.1222	-0.1252	0.0354
Finance &...	0.2718	0.0457	0.4287	0.1461	0.2774	0.0458
Real Estate	0.1953	0.0544	0.2627	0.2115	0.2049	0.0585
Social Services	-0.1432	0.0384	0.0661	0.1333	-0.1354	0.0360
Health, Sports &...	0.2700	0.0408	0.3184	0.1292	0.2773	0.0399
Education &...	0.2900	0.0372	0.3792	0.1183	0.2973	0.0368
Scientific &...	0.3281	0.0541	0.6268	0.1585	0.3322	0.0521
Government &...	0.2307	0.0358	0.3177	0.1184	0.2357	0.0356

Table C.21: Estimation Results (Henan, 2002)

	Estimation Results in 2002					
	OLS		IV		Selection	
Variables	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Intercept	8.2490	0.1181	8.3653	0.4103	8.3620	0.1244
Middle School	-0.0448	0.1693	0.0019	0.1140	-0.1043	0.1353
High School	0.2099	0.1675	0.2278	0.1129	0.1365	0.1308
College	0.3797	0.1748	0.3979	0.1317	0.2977	0.1471
Higher	0.5885	0.2019	0.8356	0.2302	0.4878	0.2245
Male	0.2458	0.0453	0.0761	0.3524	0.2182	0.0446
Male*Middle	–0.0396	0.0493	0.0654	0.3578	-0.0258	0.0482
Male*High	-0.1238	0.0478	0.1413	0.3553	-0.1062	0.0471
Male*College	-0.0739	0.0539	0.0287	0.3593	-0.0551	0.0527
Male*Higher	-0.1222	0.0738	0.2228	0.4012	-0.0972	0.0801
Experience	0.0352	0.0026	0.0166	0.0064	0.0289	0.0024
Experience^2	-0.0005	0.00006	-0.0001	0.0002	-0.0004	0.00006
Local Rural Hukou	-0.0923	0.1349	0.0958	0.1964	-0.1061	0.0537
Alien Urban Hukou	0.0726	0.1194	-0.2225	0.3092	0.0606	0.0804
Alien Rural Hukou	0.1428	0.1373	0.1443	0.2552	0.1256	0.0842
Farm &...	0.1567	0.0546	0.1874	0.1843	0.1659	0.0586
Mineral	-0.0121	0.0547	0.1069	0.1251	-0.0028	0.5443
Manufacturing	-0.0326	0.0348	0.0916	0.1143	-0.0262	0.0336
Electricity &...	0.2679	0.0437	0.2434	0.1300	0.2738	0.0439
Construction	0.0183	0.0475	0.2189	0.1548	0.0251	0.0437
Geological &...	0.2538	0.0638	0.4127	0.2532	0.2583	0.0682
Transportation &...	0.1515	0.0391	0.3259	0.1306	0.1604	0.0373
Wholesale &...	-0.1341	0.0380	-0.0621	0.1222	-0.1252	0.0354
Finance &...	0.2718	0.0457	0.4287	0.1461	0.2774	0.0458
Real Estate	0.1953	0.0544	0.2627	0.2115	0.2049	0.0585
Social Services	-0.1432	0.0384	0.0661	0.1333	-0.1354	0.0360
Health, Sports &...	0.2700	0.0408	0.3184	0.1292	0.2773	0.0399
Education &...	0.2900	0.0372	0.3792	0.1183	0.2973	0.0368
Scientific &...	0.3281	0.0541	0.6268	0.1585	0.3322	0.0521
Government &...	0.2307	0.0358	0.3177	0.1184	0.2357	0.0356

Table C.22: Estimation Results (Hubei, 2002)

	Estimation Results in 2002					
	OLS		IV		Selection	
Variables	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Intercept	8.2839	0.1221	8.4995	0.3490	8.3843	0.1273
Middle School	0.1990	0.1716	0.3424	0.1576	0.1544	0.1381
High School	0.2887	0.1704	0.4149	0.1531	0.2291	0.1336
College	0.4728	0.1764	0.5889	0.1615	0.4033	0.1470
Higher	0.6239	0.1912	0.8872	0.1783	0.5446	0.2111
Male	0.2458	0.0453	0.0761	0.3524	0.2182	0.0446
Male*Middle	–0.0396	0.0493	0.0654	0.3578	-0.0258	0.0482
Male*High	-0.1238	0.0478	0.1413	0.3553	-0.1062	0.0471
Male*College	-0.0739	0.0539	0.0287	0.3593	-0.0551	0.0527
Male*Higher	-0.1222	0.0738	0.2228	0.4012	-0.0972	0.0801
Experience	0.0352	0.0026	0.0166	0.0064	0.0289	0.0024
Experience^2	-0.0005	0.00006	-0.0001	0.0002	-0.0004	0.00006
Local Rural Hukou	-0.0923	0.1349	0.0958	0.1964	-0.1061	0.0537
Alien Urban Hukou	0.0726	0.1194	-0.2225	0.3092	0.0606	0.0804
Alien Rural Hukou	0.1428	0.1373	0.1443	0.2552	0.1256	0.0842
Farm &...	0.1567	0.0546	0.1874	0.1843	0.1659	0.0586
Mineral	-0.0121	0.0547	0.1069	0.1251	-0.0028	0.5443
Manufacturing	-0.0326	0.0348	0.0916	0.1143	-0.0262	0.0336
Electricity &...	0.2679	0.0437	0.2434	0.1300	0.2738	0.0439
Construction	0.0183	0.0475	0.2189	0.1548	0.0251	0.0437
Geological &...	0.2538	0.0638	0.4127	0.2532	0.2583	0.0682
Transportation &...	0.1515	0.0391	0.3259	0.1306	0.1604	0.0373
Wholesale &...	-0.1341	0.0380	-0.0621	0.1222	-0.1252	0.0354
Finance &...	0.2718	0.0457	0.4287	0.1461	0.2774	0.0458
Real Estate	0.1953	0.0544	0.2627	0.2115	0.2049	0.0585
Social Services	-0.1432	0.0384	0.0661	0.1333	-0.1354	0.0360
Health, Sports &...	0.2700	0.0408	0.3184	0.1292	0.2773	0.0399
Education &...	0.2900	0.0372	0.3792	0.1183	0.2973	0.0368
Scientific &...	0.3281	0.0541	0.6268	0.1585	0.3322	0.0521
Government &...	0.2307	0.0358	0.3177	0.1184	0.2357	0.0356

Table C.23: Estimation Results (Guangdong, 2002)

	Estimation Results in 2002					
	OLS		IV		Selection	
Variables	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Intercept	8.7280	0.1229	9.3775	0.4075	8.8399	0.1228
Middle School	0.2063	0.1728	0.2861	0.3113	0.1526	0.1334
High School	0.4769	0.1712	0.5668	0.3097	0.4065	0.1297
College	0.6454	0.1821	0.7110	0.3182	0.5645	0.1454
Higher	0.7803	0.2616	0.9271	0.3292	0.7059	0.2495
Male	0.2458	0.0453	0.0761	0.3524	0.2182	0.0446
Male*Middle	–0.0396	0.0493	0.0654	0.3578	-0.0258	0.0482
Male*High	-0.1238	0.0478	0.1413	0.3553	-0.1062	0.0471
Male*College	-0.0739	0.0539	0.0287	0.3593	-0.0551	0.0527
Male*Higher	-0.1222	0.0738	0.2228	0.4012	-0.0972	0.0801
Experience	0.0352	0.0026	0.0166	0.0064	0.0289	0.0024
Experience^2	-0.0005	0.00006	-0.0001	0.0002	-0.0004	0.00006
Local Rural Hukou	-0.0923	0.1349	0.0958	0.1964	-0.1061	0.0537
Alien Urban Hukou	0.0726	0.1194	-0.2225	0.3092	0.0606	0.0804
Alien Rural Hukou	0.1428	0.1373	0.1443	0.2552	0.1256	0.0842
Farm &...	0.1567	0.0546	0.1874	0.1843	0.1659	0.0586
Mineral	-0.0121	0.0547	0.1069	0.1251	-0.0028	0.5443
Manufacturing	-0.0326	0.0348	0.0916	0.1143	-0.0262	0.0336
Electricity &...	0.2679	0.0437	0.2434	0.1300	0.2738	0.0439
Construction	0.0183	0.0475	0.2189	0.1548	0.0251	0.0437
Geological &...	0.2538	0.0638	0.4127	0.2532	0.2583	0.0682
Transportation &...	0.1515	0.0391	0.3259	0.1306	0.1604	0.0373
Wholesale &...	-0.1341	0.0380	-0.0621	0.1222	-0.1252	0.0354
Finance &...	0.2718	0.0457	0.4287	0.1461	0.2774	0.0458
Real Estate	0.1953	0.0544	0.2627	0.2115	0.2049	0.0585
Social Services	-0.1432	0.0384	0.0661	0.1333	-0.1354	0.0360
Health, Sports &...	0.2700	0.0408	0.3184	0.1292	0.2773	0.0399
Education &...	0.2900	0.0372	0.3792	0.1183	0.2973	0.0368
Scientific &...	0.3281	0.0541	0.6268	0.1585	0.3322	0.0521
Government &...	0.2307	0.0358	0.3177	0.1184	0.2357	0.0356

Table C.24: Estimation Results (Chongqing, 2002)

	Estimation Results in 2002					
	OLS		IV		Selection	
Variables	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Intercept	8.3604	0.1674	8.3402	0.5727	8.4599	0.1522
Middle School	0.1063	0.1914	0.3661	0.2745	0.0649	0.1644
High School	0.4214	0.1907	0.3546	0.2707	0.3684	0.1612
College	0.5494	0.2079	0.6810	0.2862	0.4829	0.1766
Higher	0.6816	0.2491	0.8584	0.4189	0.6053	0.2653
Male	0.2458	0.0453	0.0761	0.3524	0.2182	0.0446
Male*Middle	–0.0396	0.0493	0.0654	0.3578	-0.0258	0.0482
Male*High	-0.1238	0.0478	0.1413	0.3553	-0.1062	0.0471
Male*College	-0.0739	0.0539	0.0287	0.3593	-0.0551	0.0527
Male*Higher	-0.1222	0.0738	0.2228	0.4012	-0.0972	0.0801
Experience	0.0352	0.0026	0.0166	0.0064	0.0289	0.0024
Experience^2	-0.0005	0.00006	-0.0001	0.0002	-0.0004	0.00006
Local Rural Hukou	-0.0923	0.1349	0.0958	0.1964	-0.1061	0.0537
Alien Urban Hukou	0.0726	0.1194	-0.2225	0.3092	0.0606	0.0804
Alien Rural Hukou	0.1428	0.1373	0.1443	0.2552	0.1256	0.0842
Farm &...	0.1567	0.0546	0.1874	0.1843	0.1659	0.0586
Mineral	-0.0121	0.0547	0.1069	0.1251	-0.0028	0.5443
Manufacturing	-0.0326	0.0348	0.0916	0.1143	-0.0262	0.0336
Electricity &...	0.2679	0.0437	0.2434	0.1300	0.2738	0.0439
Construction	0.0183	0.0475	0.2189	0.1548	0.0251	0.0437
Geological &...	0.2538	0.0638	0.4127	0.2532	0.2583	0.0682
Transportation &...	0.1515	0.0391	0.3259	0.1306	0.1604	0.0373
Wholesale &...	-0.1341	0.0380	-0.0621	0.1222	-0.1252	0.0354
Finance &...	0.2718	0.0457	0.4287	0.1461	0.2774	0.0458
Real Estate	0.1953	0.0544	0.2627	0.2115	0.2049	0.0585
Social Services	-0.1432	0.0384	0.0661	0.1333	-0.1354	0.0360
Health, Sports &...	0.2700	0.0408	0.3184	0.1292	0.2773	0.0399
Education &...	0.2900	0.0372	0.3792	0.1183	0.2973	0.0368
Scientific &...	0.3281	0.0541	0.6268	0.1585	0.3322	0.0521
Government &...	0.2307	0.0358	0.3177	0.1184	0.2357	0.0356

Table C.25: Estimation Results (Sichuan, 2002)

	Estimation Results in 2002					
	OLS		IV		Selection	
Variables	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Intercept	8.1747	0.1185	8.9364	0.2926	8.2832	0.1169
Middle School	0.1704	0.1696	0.0941	0.1873	0.1177	0.1281
High School	0.4436	0.1689	0.3053	0.1855	0.3785	0.1249
College	0.6681	0.1754	0.5298	0.2000	0.5899	0.1454
Higher	0.7320	0.2158	0.6945	0.1989	0.6522	0.2135
Male	0.2458	0.0453	0.0761	0.3524	0.2182	0.0446
Male*Middle	–0.0396	0.0493	0.0654	0.3578	-0.0258	0.0482
Male*High	-0.1238	0.0478	0.1413	0.3553	-0.1062	0.0471
Male*College	-0.0739	0.0539	0.0287	0.3593	-0.0551	0.0527
Male*Higher	-0.1222	0.0738	0.2228	0.4012	-0.0972	0.0801
Experience	0.0352	0.0026	0.0166	0.0064	0.0289	0.0024
Experience^2	-0.0005	0.00006	-0.0001	0.0002	-0.0004	0.00006
Local Rural Hukou	-0.0923	0.1349	0.0958	0.1964	-0.1061	0.0537
Alien Urban Hukou	0.0726	0.1194	-0.2225	0.3092	0.0606	0.0804
Alien Rural Hukou	0.1428	0.1373	0.1443	0.2552	0.1256	0.0842
Farm &...	0.1567	0.0546	0.1874	0.1843	0.1659	0.0586
Mineral	-0.0121	0.0547	0.1069	0.1251	-0.0028	0.5443
Manufacturing	-0.0326	0.0348	0.0916	0.1143	-0.0262	0.0336
Electricity &...	0.2679	0.0437	0.2434	0.1300	0.2738	0.0439
Construction	0.0183	0.0475	0.2189	0.1548	0.0251	0.0437
Geological &...	0.2538	0.0638	0.4127	0.2532	0.2583	0.0682
Transportation &...	0.1515	0.0391	0.3259	0.1306	0.1604	0.0373
Wholesale &...	-0.1341	0.0380	-0.0621	0.1222	-0.1252	0.0354
Finance &...	0.2718	0.0457	0.4287	0.1461	0.2774	0.0458
Real Estate	0.1953	0.0544	0.2627	0.2115	0.2049	0.0585
Social Services	-0.1432	0.0384	0.0661	0.1333	-0.1354	0.0360
Health, Sports &...	0.2700	0.0408	0.3184	0.1292	0.2773	0.0399
Education &...	0.2900	0.0372	0.3792	0.1183	0.2973	0.0368
Scientific &...	0.3281	0.0541	0.6268	0.1585	0.3322	0.0521
Government &...	0.2307	0.0358	0.3177	0.1184	0.2357	0.0356

Table C.26: Estimation Results (Yunnan, 2002)

	Estimation Results in 2002					
	OLS		IV		Selection	
Variables	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Intercept	8.3259	0.1171	8.1491	0.5287	8.4370	0.1192
Middle School	0.1766	0.1681	0.1270	0.1420	0.1202	0.1308
High School	0.3738	0.1672	0.2431	0.1392	0.3058	0.1267
College	0.4952	0.1740	0.3933	0.1457	0.4098	0.1405
Higher	0.4671	0.1796	0.3753	0.1569	0.3787	0.2038
Male	0.2458	0.0453	0.0761	0.3524	0.2182	0.0446
Male*Middle	–0.0396	0.0493	0.0654	0.3578	-0.0258	0.0482
Male*High	-0.1238	0.0478	0.1413	0.3553	-0.1062	0.0471
Male*College	-0.0739	0.0539	0.0287	0.3593	-0.0551	0.0527
Male*Higher	-0.1222	0.0738	0.2228	0.4012	-0.0972	0.0801
Experience	0.0352	0.0026	0.0166	0.0064	0.0289	0.0024
Experience^2	-0.0005	0.00006	-0.0001	0.0002	-0.0004	0.00006
Local Rural Hukou	-0.0923	0.1349	0.0958	0.1964	-0.1061	0.0537
Alien Urban Hukou	0.0726	0.1194	-0.2225	0.3092	0.0606	0.0804
Alien Rural Hukou	0.1428	0.1373	0.1443	0.2552	0.1256	0.0842
Farm &...	0.1567	0.0546	0.1874	0.1843	0.1659	0.0586
Mineral	-0.0121	0.0547	0.1069	0.1251	-0.0028	0.5443
Manufacturing	-0.0326	0.0348	0.0916	0.1143	-0.0262	0.0336
Electricity &...	0.2679	0.0437	0.2434	0.1300	0.2738	0.0439
Construction	0.0183	0.0475	0.2189	0.1548	0.0251	0.0437
Geological &...	0.2538	0.0638	0.4127	0.2532	0.2583	0.0682
Transportation &...	0.1515	0.0391	0.3259	0.1306	0.1604	0.0373
Wholesale &...	-0.1341	0.0380	-0.0621	0.1222	-0.1252	0.0354
Finance &...	0.2718	0.0457	0.4287	0.1461	0.2774	0.0458
Real Estate	0.1953	0.0544	0.2627	0.2115	0.2049	0.0585
Social Services	-0.1432	0.0384	0.0661	0.1333	-0.1354	0.0360
Health, Sports &...	0.2700	0.0408	0.3184	0.1292	0.2773	0.0399
Education &...	0.2900	0.0372	0.3792	0.1183	0.2973	0.0368
Scientific &...	0.3281	0.0541	0.6268	0.1585	0.3322	0.0521
Government &...	0.2307	0.0358	0.3177	0.1184	0.2357	0.0356

Table C.27: Estimation Results (Gansu, 2002)

	Estimation Results in 2002					
	OLS		IV		Selection	
Variables	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Intercept	8.3940	0.1837	7.9309	0.4616	8.5112	0.1179
Middle School	-0.1031	0.1613	0.0167	0.2265	-0.1642	0.1174
High School	0.1662	0.1607	0.1935	0.2206	0.0944	0.1144
College	0.2712	0.1653	0.3737	0.2346	0.1827	0.1268
Higher	0.5737	0.1748	0.5445	0.2745	0.4748	0.1839
Male	0.2458	0.0453	0.0761	0.3524	0.2182	0.0446
Male*Middle	–0.0396	0.0493	0.0654	0.3578	-0.0258	0.0482
Male*High	-0.1238	0.0478	0.1413	0.3553	-0.1062	0.0471
Male*College	-0.0739	0.0539	0.0287	0.3593	-0.0551	0.0527
Male*Higher	-0.1222	0.0738	0.2228	0.4012	-0.0972	0.0801
Experience	0.0352	0.0026	0.0166	0.0064	0.0289	0.0024
Experience^2	-0.0005	0.00006	-0.0001	0.0002	-0.0004	0.00006
Local Rural Hukou	-0.0923	0.1349	0.0958	0.1964	-0.1061	0.0537
Alien Urban Hukou	0.0726	0.1194	-0.2225	0.3092	0.0606	0.0804
Alien Rural Hukou	0.1428	0.1373	0.1443	0.2552	0.1256	0.0842
Farm &...	0.1567	0.0546	0.1874	0.1843	0.1659	0.0586
Mineral	-0.0121	0.0547	0.1069	0.1251	-0.0028	0.5443
Manufacturing	-0.0326	0.0348	0.0916	0.1143	-0.0262	0.0336
Electricity &...	0.2679	0.0437	0.2434	0.1300	0.2738	0.0439
Construction	0.0183	0.0475	0.2189	0.1548	0.0251	0.0437
Geological &...	0.2538	0.0638	0.4127	0.2532	0.2583	0.0682
Transportation &...	0.1515	0.0391	0.3259	0.1306	0.1604	0.0373
Wholesale &...	-0.1341	0.0380	-0.0621	0.1222	-0.1252	0.0354
Finance &...	0.2718	0.0457	0.4287	0.1461	0.2774	0.0458
Real Estate	0.1953	0.0544	0.2627	0.2115	0.2049	0.0585
Social Services	-0.1432	0.0384	0.0661	0.1333	-0.1354	0.0360
Health, Sports &...	0.2700	0.0408	0.3184	0.1292	0.2773	0.0399
Education &...	0.2900	0.0372	0.3792	0.1183	0.2973	0.0368
Scientific &...	0.3281	0.0541	0.6268	0.1585	0.3322	0.0521
Government &...	0.2307	0.0358	0.3177	0.1184	0.2357	0.0356

Table C.28: Distribution of Provinces (1995)

Province	Freq.	Percent	Cum.
11	1,529	7.05	7.05
14	2,110	9.72	16.77
21	2,212	10.19	26.97
32	2,450	11.29	38.26
34	1,527	7.04	45.29
41	1,939	8.94	54.23
42	2,310	10.65	64.88
44	1,821	8.39	73.27
51	2,486	11.46	84.73
53	2,010	9.26	93.99
62	1,304	6.01	100.00
Total	21,698	100.00	

Table C.29: Distribution of Provinces (2002)

Province	Freq.	Percent	Cum.
11	1447	7.08	7.08
14	1836	8.98	16.06
21	2109	10.51	26.38
32	2149	7.16	44.06
34	1464	7.16	44.06
41	2073	10.14	54.20
42	2056	10.06	64.26
44	1757	8.60	72.86
50	827	4.05	76.90
51	1696	8.30	85.20
53	1836	8.98	94.18
62	1189	5.82	100.00
Total	20439	100.00	

Table C.30: Distribution of Industries (1995)

Industry	Freq.	Percent	Cum.
Farm, Forest, Husbandry & Fishery	236	1.67	1.67
Manufacturing	5,826	41.21	42.88
Geological Prospecting, Irrigation Administration	155	1.10	43.73
construction	418	2.96	46.93
Transportation, Storage, Post Office	684	4.84	51.95
Wholesale, Retail & Food Services	1,996	14.12	65.89
Real Estate & Public Utilities	548	3.88	69.78
Health, Sports and Social Welfare	677	4.79	74.55
Education, Culture & Arts, Mass Media	1,051	7.43	81.99
Scientific Research & Professional Servicei	309	2.19	84.17
Finance & Insurance	250	1.77	85.94
Government Agents, Party Organizations	1,596	11.29	97.13
Other	392	2.77	100.00
Total	14138	100.00	

Table C.31: Distribution of Industries (2002)

Industry	Freq.	Percent	Cum.
Farm, Forest, Husbandry & Fishery	125	1.22	1.22
Mineral	158	1.54	2.76
Manufacturing	2552	24.92	27.69
Electricity, Gas & Water Supply	332	3.24	30.93
Construction	334	3.26	34.19
Geological Prospecting, Irrigation Administration	83	0.81	35.00
Transportation, Storage, Post Office	801	7.82	42.83
Wholesale, Retail & Food Services	1256	12.27	55.09
Finance & Insurance	275	2.69	57.88
Real Estate	123	1.20	58.98
Social Services	1050	10.25	69.24
Health, Sports and Social Welfare	521	5.09	74.32
Education, Culture & Arts, Mass Media	918	8.97	83.29
Scientific Research & Professional Service	178	1.74	85.03
Government Agents, Party Organizations	1224	11.95	96.98
Other	309	3.02	100.00
Total	10239	100.00	

Table C.32: Testing Endogeneity

ln(Income)	Coef.	Std. Err.	t	P$>\|t\|$
Intercept	7.7292	0.0652	118.52	0.000
Schooling	0.1760	0.0349	5.05	0.000
Residual	-0.1217	0.0349	-3.49	0.000
Male	0.1234	0.0159	7.75	0.000
Experience	0.0400	0.0030	13.15	0.000
Experience^2	-0.0004	0.0005	–7.29	0.000
Beijing	0.3058	0.0411	7.44	0.000
Shanxi	-0.1725	0.0315	-5.48	0.000
Liaoning	-0.0959	0.0324	-2.96	0.003
Jiangsu	0.0362	0.0318	1.14	0.256
Anhui	-0.1577	0.0328	-4.80	0.000
Henan	-0.2655	0.0316	-8.41	0.000
Hubei	-0.1231	0.0306	-4.02	0.000
Guangdong	0.4428	0.0327	13.54	0.000
Sichuan	-0.1207	0.0324	-3.73	0.000
Yunnan	-0.0308	0.0313	-0.98	0.326
Gansu	-0.1875	0.0341	-5.50	0.000
Farm, Forest, Husbandry & Fishery	0.1932	0.0564	3.43	0.025
Mineral	0.0643	0.0572	1.12	0.885
Manufacturing	0.0412	0.0387	1.07	0.365
Electricity, Gas & Water Supply	0.3367	0.0468	7.19	0.000
Construction	0.0879	0.0502	1.75	0.747
Geological Prospecting, &	0.2774	0.0654	4.24	0.001
Transportation, Storage, Post Office	0.2261	0.0426	5.31	0.000
Wholesale, Retail & Food Services	-0.0488	0.0416	-1.17	0.001
Finance & Insurance	0.3091	0.0489	6.32	0.000
Real Estate	0.2472	0.0571	4.33	0.001
Social Services	-0.0654	0.0418	-1.56	0.000
Health, Sports and Social Welfare	0.3067	0.0442	6.94	0.000
Education, Culture & Arts, Mass Media	0.3114	0.0411	7.58	0.000
Scientific Research & Professional Service	0.3633	0.0575	6.32	0.000
Government Agents, Party Organizations	0.2562	0.0399	6.43	0.000

Appendix D

Summary of Figures

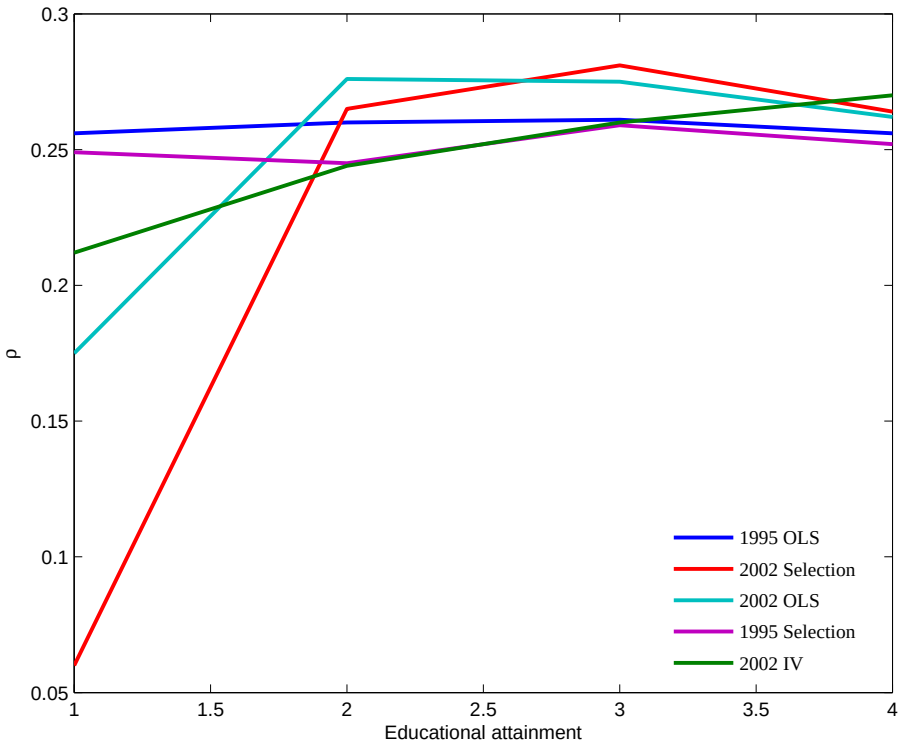

FIGURE D.1: Spatial Dependence (using matrix $\mathbb{C}$)

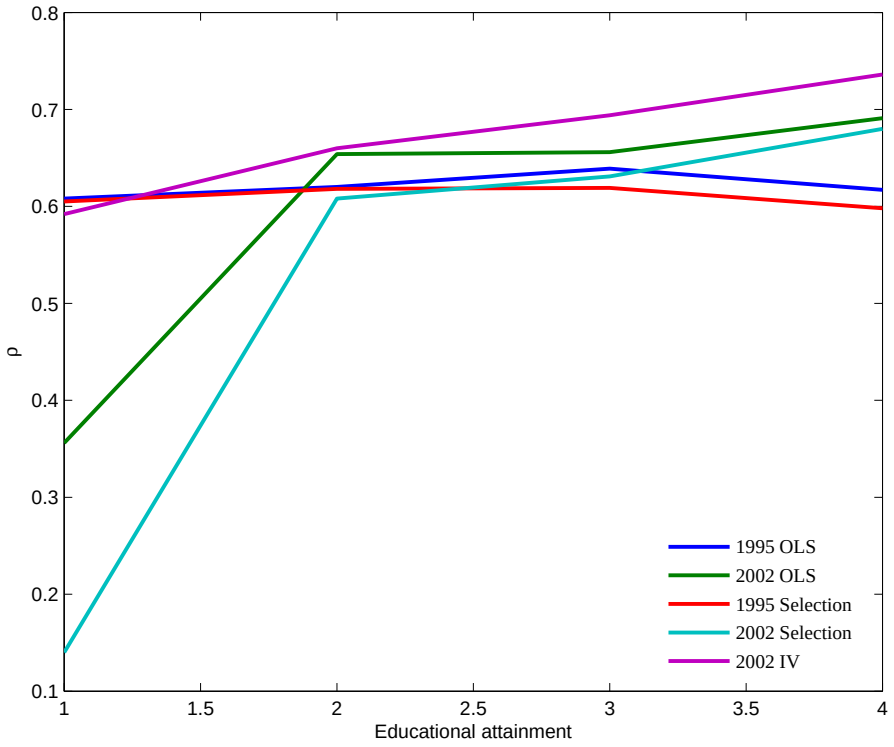

FIGURE D.2: Spatial Dependence (using matrix $\mathbb{A}$)

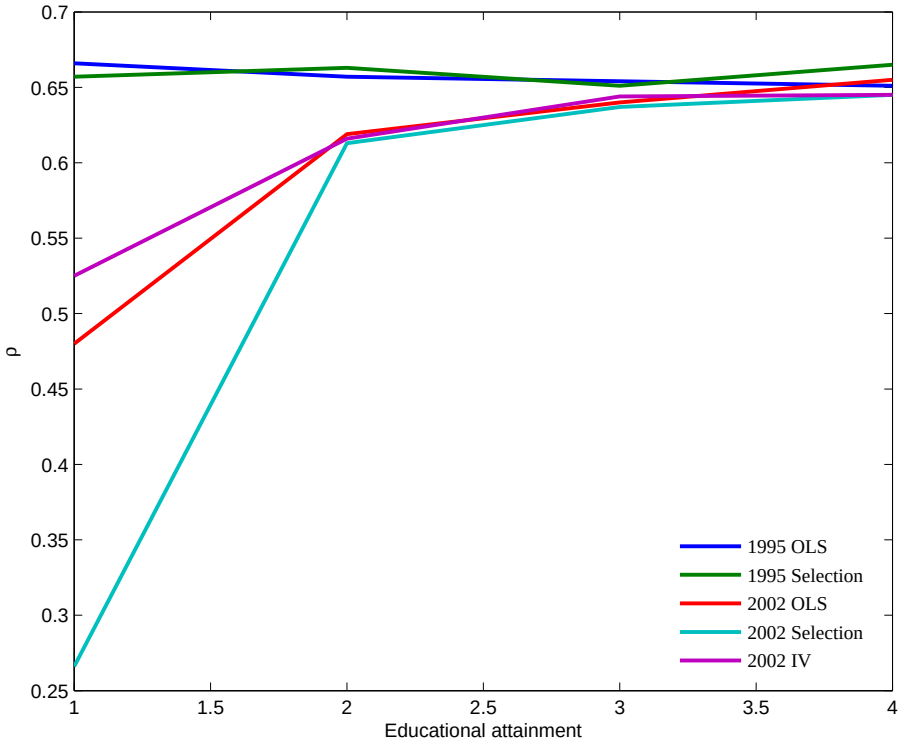

FIGURE D.3: Spatial Dependence (using matrix $\mathbb{D}$)

Bibliography

Angrist, J., 2000. "Estimation of Limited Dependent Variable Models with Dummy Endogenous Regressors: Simple Strategies for Empirical Practice", NBER working papers No. 248.

Angrist, J., Krueger, A., 1999. "Empirical Strategies in Labor Economics", *Handbook of Labor Economics*, Vol. 3, Part A, pp. 1277-1366.

Becker, G., 1964. *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*, Columbia University Press, New York.

Bierens, H., 2004. *Introduction to The Mathematical and Statistical Foundations of Econometrics*, Cambridge University Press, Cambridge.

Borjas, G., 2010. *Labor Economics*, 5th edition, McGraw-Hill Irwin, New York

Chow, G., 2001. *China's Economic Transformation*, Princeton University, Oxford, Blackwell.

Fang, H., Eggleston, K., Rizzo, J., Rozelle, S., Zeckhauser, R., 2012. *The Returns to Education in China: Evidence from the 1986 Compulsory Education Law*, NBER working papers No. 18189

Fleisher, B., Li, H., Zhao, M., 2011. Human Capital, Economic Growth, and Regional Inequality in China, *Journal of Development Economics*, Vol. 92, #2, pp. 215-231.

Fleisher, B., Wang, X., 2004. "Skill Differentials, Return to Schooling, and Market Segmentation in a Transition Economy: the Case of Mainland China", *Journal of Development Economics*, Vol. 73, #1, pp. 315-328.

Hansen, L., 1982. "Large Sample Properties of Generalized Method of Moments Estimators", *Econometrica*, Vol. 50, #4, pp. 1029-1054.

Heckman, J., 1978. "Dummy Endogenous Variable in a Simultaneous Equations System", *Econometrica*, Vol. 46, #4, pp. 931-959.

Heckman, J., 1979. "Sample Selection Bias as a Specification Error", *Econometrica*, Vol. 47, # 1, pp. 153-161.

Heckman, J., 1997. "Instrumental Variable: A Study of Implicit Behavioral Assumption in Making Program Evaluations", *Journal of Human Resources*, Vol. 32, # 3, pp. 441-462

Heckman, J., 2002. "Chinas Investment in Human Capital", NBER working paper No. 9296.

Heckman, J., Li, X., 2004. "Selection Bias, Comparative Advantage and Heterogeneous Returns to Education: Evidence from China in 2000", *Pacific Economic Review*, Vol. 9, #3, pp. 155-171.

Heckman, J., Lochner, L., Todd, P., 2003. "Fifty Years of Mincer Earnings Regressions", NBER working papers No. 9732

Heckman, J., Lochner, L., Todd, P., 2008. "Earning Functions and Rates of Return", *Journal of Human Capital*. Vol. 2, #1, pp. 1-31.

Heckman, J., Yi, J., 2012. "Human Capital, Economics Growth, and Inequality in China", NBER working papers No. 18100.

Bradsher, K., 2013. "The Education Revolution", *The New York Times*. The New York Times, Web. 22 Apr. 2013.

Lesage, J., Pace, R., 2009. *Introducation to Spatial Econometrics*, Chapman and Hall

Li, H., Luo, Y., 2004. "Reporting Errors, Ability Heterogeneity, and Returns to Schooling in China", *Pacific Economic Review*, Vol. 9, #3, pp. 191-207.

Mincer, J., 1974. *Schooling, Experience and Earning*, Columbia University Press: New York

Psacharopoulos, G., 1981. "Returns to Education: An Updated International Comparison", *Comparative Education*, Vol. 17, #3, pp. 321-.341.

Statistical Year Book, 2003. National Bureau of Statistics of China.

Statistical Year Book, 2012. Ministry of Education in China.

Rivers, D., Vuong, Q., 1988. "Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models", *Journal of Econometrics*, Vol. 39, #3, pp. 347-366.

Wooldridge, J., *Econometric Analysis of Cross Section and Panel Data*, 2nd edition, MIT Press, Cambridge.

Sandell, S., Shapiro, D., 1978. "The Theory of Human Capital and The Earnings of Women: A Reexamination of The Evidence", *Journal of Human Resources*, Vol. 13, #1, pp. 103-117.

LIYIN JIANG

Academic Vitae updated January 18, 2014

Contact Information

501 Vario Boulevard
Apartment 2524D
State College, PA, 16803

+1(814)-441-7532
lyj5044@psu.edu
http://sites.google.com/site/liyinjiang1
Citizenship: Chinese F-1 visa

Intended Fields

Labor Economics (Human Capital), Applied Econometrics, Behavioural Economics

Education

Schreyer Honors College, The Pennsylvania State University, University Park

B.S. in Mathematics, B.S. in Economics (Honors), May 2014

- Completed 16 Honors courses
- Advanced Economics Courses: Coordination Games, Finance, Labor, Mathematical Econ, Microeconometrics, PhD Micro I
- Advanced Mathematics Courses: Complex Analysis, Discrete Mathematics, Linear Algebra, Differential Equations, Real Analysis, Classical Analysis, Topology, Vector Calculus
- Advanced Statistics Courses: Computational Statistics, Mathematical Statistics, Probability, Stochastic Modeling

Thesis: *Geographic Heterogeneity of Returns to Education in China*, Spring 2013
Thesis Supervisor: David Shapiro, Professor of Economics

Honors and Awards

Department of Economics Honors Program Participant, August 2012 - May 2013

- Completed the honors program as the only junior student in class.

Dean's List (all semesters)

Theory and Quantitative Methods Modulo Award, Fall 2011 (completed in the 3$^{\mathbf{rd}}$ semester)

Plenary speaker for international student advising (FTCAP), Fall 2011

Working Papers

Strategic Complementarity in Labor Markets

Research Experience

REU Program (Bates White), June 2013 - August 2013

- Research Assistant for Professor Russell Chuderewicz

Teaching Experience

Grading Assistant, Spring 2014

- Grading undergraduate econometrics for Professor Patrik Guggenberger
- Attended regular lectures, prepared solutions to problem sets and proctored exams.

Tutor at Econese

- Hosted regular review sessions on undergraduate econ and math courses.

Professional Experience

Student Association

- Founder of Econese, Spring 2011

Internship

- Shanghai International Office, University of Southern California, Spring 2010

Computer skills

C++, EViews, LaTeX, Matlab, Microsoft Office, R, STATA

Language Skills

Chinese (native), English (fluent), Japanese (elementary)