Human Capital and Migration: a Cautionary Tale

Salvador Navarro[1] and Jin Zhou[2]

[1]University of Western Ontario

[2]University of Chicago

Very Preliminary and Incomplete.
Please Do Not Circulate without Permission.

Abstract

We study the interaction of migration and education decisions, and their effects on labor market outcomes of individuals in sending locations. We consider the possibility that, while the level of human capital affects the migration decisions of an individual (i.e., self-selection of migrants), it is also the case that the possibility of migration itself affects the human capital accumulation decisions of agents. In particular, we first analyze how the migration option can reduce the incentives to accumulate human capital in the context of a simple Roy model with exogenous migration. As we show, even when the return to migrating is positive, if the return to education for migrants is lower in the receiving location than in the sending location, the mere possibility of migrating reduces the returns to human capital accumulation for people in the sending location. We analyze data on rural migration in China, where this pattern of returns seems to hold. We then use diff-in-diff to show that, consistent with our simple model's prediction, educational attainment in rural China slowed down compared to urban regions after an early 80's reform that relaxed the restrictions to rural migration. Finally, we build a structural model of rural-urban migration in China, where we estimate the reduction in migration costs that happened as a consequence of the reform. To quantify the effect of the policy, we simulate what would have happened had the policy not been implemented. We find that the attendance rates for high school, some college and college would have increased by 29%, 141%, and 24%, respectively.

1 Introduction

The study of individual location mobility decisions, and the effects that these decisions have on both destination and sending locations has a long history in Economics. In this paper, we study the effect that migration has on labor outcomes of individuals in the source locations. As opposed to the work of Borjas (1987), Borjas (1995), Chiquiar and Hanson (2005) or Card (2005), we don't take the human capital of the individual as given and ask whether migrants are positively or negatively selected. Instead, we allow for the possibility that, while human capital determines migration, the mere existence of the migration option can also affect the individual's human capital accumulation decision.

The point of departure of our analysis is a joint Roy model of migration and education decisions that leads to the simple observation: if the return to human capital accumulation for migrants is lower in the potential receiving location than it is at home, then the overall return to education is lower than if the possibility of migration did not exist. This pattern of returns is likely to exist in places in which low skilled migration is prevalent, and where the types of jobs available for these migrants in the receiving location are not skill intensive. For example, if a high school dropout and a high school graduate will both end up washing dishes at a restaurant if they migrate, then the skill component of the return to migration is likely to be very low. If this is the case, the mere possibility that the option of migration is there, lowers the incentive to accumulate human capital in the source location, when compared to a case in which migration is not possible (or harder).

Our analysis is in contrast to the standard assumption that the return to education (human capital) is higher at the migration destination, so the migration option makes individuals in the less developed source locations have more incentive to invest in their human capital (e.g., Dustmann and Glitz, 2011). However, as shown by Hendricks and Schoellman (forthcoming), when high skill and low skill workers do not perfectly substitute, this needs not be the case. Furthermore, the return mechanism we highlight is different than the one studied in the brain drain literature (Maria and Stryszowski, 2009; Maria and Lazarova, 2012) which argues that migration will change the composition of skills due to the emigration of skilled people. It is closer to the idea motivating the work of Stark et al. (1997), Mountford (1997), Vidal (1998), Beine et al. (2001), Stark and Wang (2002), Beine et al. (2008) where the possibility of migration raises the return to education and may lead to a higher level of human capital.

In order to examine the empirical plausibility of our hypothesis, we use the Chinese Household Income Project panel to analyze the human capital accumulation effects of one of the largest migration episodes in human history: the rural to urban migration that followed the cultural revolution. The literature on rural education in China has, for the most part, ignored the impacts of the migration option and has focused on other determinants like poverty (Knight et al., 2009), or the ineffectiveness of the college expansion (Li et. al, 2003). Brauw and Giles (2017) is the closest to our study, in that they empirically find that the reduction of migration costs has had a negative relationship with high school enrollment. They, however, do not propose a mechanism that could explain this finding.

The case of China is good fit for our analysis. As a consequence of the segregation policy of 1958, people in the rural areas were prohibited from migrating to urban areas in China until the early 1980's. At the time, the Chinese government implemented migration policies to allow and then gradually facilitate and encourage more and more rural people to migrate to urban areas. We first show that in a simple reduced form comparison, the returns to education between rural and urban areas are consistent with what is required in our simple model. We then exploit the differential educational attainment between rural and urban areas before and after the policy in a difference-in-difference approach to show that the relaxation of the migration restrictions was negatively associated with years of education for people born in rural areas.

The evidence from the difference-in-difference estimates is consistent with the story our simple Roy model suggests. However, the extent to which the policy actually reduced migration costs (i.e., facilitated migration) is unknown. Other policies that could affect educational attainment were also implemented (and have been implemented since). Therefore, in order to quantify and distinguish the migration policy effect from other policies (i.e., the college expansion policy), we develop and estimate a life-cycle dynamic discrete choice model in which we try to account for not only the migration policies implemented in the early 80's, but also other policies like the college expansion policy, etc. In the model, individuals differ in terms of two types of unobserved skills, cognitive and non-cognitive skills (Heckman et al., 2006, Navarro and Zhou, 2017 etc), and these characteristics can affect their endogenously chosen educational attainment and location.

The model estimates show that the return to education in rural areas is higher than that in urban areas, consistent with the findings in the reduced form analysis. Second, we also find that the different types of unobserved skills play different roles. For example, cognitive and non-cognitive skills are rewarded differently in rural and urban areas. In urban areas, cognitive skills have a significantly positive return, while non-cognitive skills have a significantly positive return in rural areas. Individuals with high non-cognitive skills have lower migration costs.[1]. Third, we find that the implementation of the policy in the early 80's reduced migration costs significantly, and that this lead to a significant reduction in rural individuals human capital accumulation, consistent with the reduced form evidence. When we simulate the counterfactual situation in which we eliminate the migration policies, the attendance rate for high school, some college and college level education increases by 29%, 141%, and 24%, respectively.

The rest of the paper is organized as follows. Section 2 develops a simplified Roy model of migration and education that delivers the conditions under which having an option to migrate can reduce human capital accumulation in the source locality. Section 3 provides reduced form evidence consistent with the prediction of our simple model. Section 4 presents background information on the education system and rural-urban migration in China, and describes the data in detail. In Section 5 we develop a dynamic empirical structural model or migration and educational attainment that retains the key components of our simplified model, but that tries to account for some of the salient features of the Chinese case as described in the previous section. In Section 6, the identification conditions and estimation procedure are discussed. Estimation results are presented in Section

[1]Consistent with the finding in Zhou (2017) that social networks reduce migration costs.

7, and counterfactual simulations are presented in Section 8. Section 9 concludes.

2 Educational Attainment Accumulation and Migration: A Toy Model

In this section, we use a very simple model to illustrate that, provided a particular (but plausible) pattern of returns to education holds, the option value generated by the migration decision can lead to lower human capital accumulation for people at risk of migration. To see this, consider first a world in which no migration is possible, and consider the case in which an individual faces a simple binary decision of whether to get an education or not. An individual in location r (rural), faces earnings of y_{rl} (l for low) if he does not go to school, and of y_{rh} (h for high) if he does. Trivially, the individual will invest if $y_{rh} > y_{rl}$.

We now introduce a second location u (urban). For simplicity, we assume that an individual would choose to migrate from r to u if this were a possibility. We further assume that there is some exogenous probability P that an individual can migrate. Let y_{ul} and y_{uh} denote earnings in the new location if an individual has low or high education, respectively. The individual's expected earnings when he does not invest in human capital is thus given by $Ey_l = y_{rl}(1-P)+y_{ul}P$, and the expected earnings when he gets an education are $Ey_h = y_{rh}(1 - P) + y_{uh}P$. When migration is possible, an individual will go to school provided that:

$$\begin{aligned} Ey_h > Ey_l &\iff y_{rh}(1-P) + y_{uh}P > y_{rl}(1-P) + y_{ul}P \\ &\iff \underbrace{[(y_{uh} - y_{rh}) - (y_{ul} - y_{rl})]}_{A} P + (y_{rh} - y_{rl}) > 0. \qquad (1) \end{aligned}$$

Here the term in square brackets (A), measures the difference between the return to migrating when the individual invests in human capital and the return when he does not. When this term is negative, the probability of going to a higher level of education (i.e., $Pr(Ey_h - Ey_l > 0)$) will decrease. Alternatively, we can rewrite $A = (y_{uh} - y_{ul}) - (y_{rh} - y_{rl})$, and reinterpret our condition as stating that the return to human capital investment for migrants is smaller in urban than in rural areas.

The model we present here is too simplified to be of practical use. However, it is enough to illustrate the mechanism we want to highlight in this paper: if the return of moving to the city with an higher level of education is not as high as the return of moving to the city without it, individuals will be less likely to increase their human capital. This will be the case if, for example, a rural person moving to the city will get roughly the same job (say, as a cab driver or a cook) whether they are a high school graduate or not. In both cases the person will make more in the city, but the relative gain is larger for the individual who did not get an education.

The condition that the education return rate is lower in urban (for migrants) is different from what is usually assumed in most of the migration literature. Usually, researchers assume the return is higher (or the same) in the migration destination (e.g., Dustmann and Glitz, 2011). However, as we show below, this does not seem to be the case (at least at lower levels of education) for rural-urban migration in China.[2]

[2] It is likely that a similar pattern holds for other cases, for example for Mexico-U.S. international migration (McKenzie and Rapoport, 2011).

3 Education and Rural-Urban Migration in China: Preliminary Evidence

In this section, we illustrate that the pattern described in the previous section, both in terms of education returns and migration, seems to be present in China. We employ the China Household Income Project (CHIP) panel survey for 2007-2009[3], which we describe in more detail in Section 4.

We restrict our attention to males with a rural household registration (Hukou), which essentially means that they were born in a rural area.[4] We define a migrant as someone whose work is located in an urban area outside of the county of their rural Hukou registration location. Tables 1 and 2 present some basic summary statistics for the data we employ.[5] Table 1 compares demographics between individuals with urban Hukou and those with rural Hukou. It is clear that the urban group have higher years of education than that of the rural group (i.e., 12.34 years vs. 9.22 years). Also, it is three times of urban people who take college entrance exam comparing to the rural group. However, there are no significant differences for non-cognitive skill measures between these two groups. Table 2 shows that on average, both earnings and years of education are higher for rural migrants than for non-migrants. Self-reported class performance is slightly better for non migrants, and the fraction of taking the collage entrance exam is very close between rural migrants and rural non-migrants. One thing we should notice is that rural migrants are less likely to smoke and also they smoke less than non-migrants.

[3]CHIP data (2007-2009) are part of the "Rural Urban Migration in China" data project.

[4]We restrict our attention to males in order to simplify the structural model we develop in Section 5. As we show in Appendix Tables A1-A5, the same patterns we illustrate in this section hold even more strongly for females.

[5]We deflate earnings to 1995 Yuan using regional price indices, which we construct following Brandt and Holz (2006).

We run Mincer-type regressions in Table 6. As expected, there seems to be a positive return to migration, even after controlling for education. Furthermore, the return to education in rural areas is between 5% and 6% per month, consistent with the findings in Heckman and Li (2004). For both years, we find that the difference in the return to education for rural individuals in urban areas compared to rural areas (the A term in equation 1) is negative, consistent with the pattern described in our toy model in the previous section. As we show in columns 2 to 4 of the table, the pattern holds even after we control for different measures of cognitive and non-cognitive skills.

In order to investigate whether this pattern has led to a decline in educational attainment associated with migration, we take advantage of a change in migration policy that happened in the early 1980's in China. Migration in China has been restricted by the government since 1958. In 1978, the "people's commune system" was replaced by the "household-responsibility system", which loosened the restrictions on rural residents mobility. However, the central government still kept strict restrictions limiting the opportunities of working in cities until around 1982-1983. As a consequence, we consider the policy as having taking place in 1983.

We begin by graphically analyzing in Figure 3 whether the elimination of restrictions to migration has an effect consistent with the predictions of our model in Section 2. In the figure we plot the evolution of average years of education according to the year in which individuals finish their formal education, for both individuals with an urban and a rural Hukou. In order to highlight the patterns we wish to illustrate, we plot the average educational attainment of urban individuals on the left axis and that of rural individuals on the right axis, making sure that the lines coincide in 1982. As the figure shows, the educational attainment of both groups follows closely until around 1982, at which point rural educational attainment slows down considerably relative to its urban counterpart. This pattern continues to hold over the next decades, and it is not until around 2003, four years after the massive college expansion program started in 1999, that the pattern seems to break.

We more formally investigate whether the pattern we show in Figure 3 holds by running a diff-in-diff regression for educational attainment before and after 1983. We begin by investigating whether the parallel trends assumption pre-1983 holds. In Table 7 we run a regression of years of education against an indicator for rural, year dummies up to 1982, and interactions. As we can see, the interactions (i.e., the difference in the trend for urban vs rural) is not significant, so we cannot reject the hypothesis of parallel trends. This pattern, that the interaction of the year dummy and rural is not significant, holds even after we control for measures of cognitive and non-cognitive skills

Having established that the parallel trends assumption holds, we then run a differences-in-differences specification to examine the effect that the loosening of the migration restrictions has on education choices in Table 8. The variable "Time" is an indicator that takes value 1 if the individual finishes his formal education after 1982, and "Rural" is the indicator for whether the individual has a rural Hukou. The interaction term reflects the impact of the changes in migration policy on the individual's schooling choices. As shown in the table, after 1982 rural individuals have significantly less education than the pre-1983 pattern implies, consistent with the predictions of our model. This is true even after we control for different measures of cognitive and non-cognitive skills.

4 Migration and Education in China: Background and Data

4.1 Education System in China

In China, there are four stages of processing education development in the lifetime. The first stage is primary education level (Grade 1 to Grade 6). In most provinces, students who graduate need to finish Grade 6, but in some provinces students can graduate when they finish Grade 5. For example, in Jiangxi Province, before 2000, primary school stage is five years instead of six years. The second stage (junior secondary education) is named as middle school level education (Grade 7 to Grade 9), which consists the last three years of nine year compulsory education in China. Both the central and the local government provide subsidies to schools to cover student tuition fees and other payments (e.g. textbooks, sports equipment).

The third stage is the high school period (senior secondary). In terms of the target of future career, there are two types of high schools: academic high schools and vocational high schools. In China, a senior high school graduate will be considered as an educated person, especially in rural areas. Since students from both types of high schools are allowed to take college entrance exam, to simplify our study, we do not distinguish these two types of high schools. High school level education is not mandatory in China, individuals can choose to attend high school or not. Although there is an exam to sort students into different qualities of high schools, almost all individuals could attend some types of high school. We will not consider there exists a capacity constraint for high school admission in our studies.

The fourth stage is college and above level studies. After 1978, the government restarts college entrance exam to select high ability students in college level studies. Due to the capacity constraint, the admission process is highly selective. For example, before 1980, only less than 10% of individuals who take the exam can get an admission. Before the year 1999, the total admission rate is still below 35%. Before 1996 the college tuition is very low and the government also provides the subsidies to college students.

The big change for college studies is the college expansion policy which starts from 1999. Since 1999, there are additional more 60,000 admission each year. From Figure 2, we can see that the college admission offers increases from 1.6 million to more than 6 million in ten years (1999-2009).

How much this expansion policy would affect rural individuals' education choices? Potentially we may expect this policy would increase the individuals' incentive of having more years of education. In our work, we also examine the impact of this policy on rural individuals' education decisions.

4.2 Rural-Urban Migration in China

Since 1958, the Chinese central government has restricted the mobility of the population. From 1958 to 1978, the few rural people who had job offers in urban areas or recruitment letters from universities could migrate from rural to urban areas. Since 1978, the people's commune system was replaced by the household-responsibility system, which loosened the restriction of rural resident's mobility. Although the central government had strict restrictions to limit the opportunities of working in cities, from 1979 to 1983, some of rural residents began to migrate to work outside of their counties. Between 1984 and 1988, the central government did not restrict rural-urban migration. At that time, there was no market for exchanging food. People needed to use food stamps to get food. However, if rural individuals migrated, they had to provide food stamps for themselves.[6] It was still hard for rural individuals to migrate since it was not easy to have enough food stamps to support themselves. This migration policy was suspended between 1988 and 1990. After 1991, the government began to encourage rural-urban migration and starting in 2000, the government started to reform the household registration system to encourage more rural individuals to migrate.[7] For example, in 2007, 12 provinces in China canceled the rural household registration, which meant that rural individuals had the same household registration as urban households in these provinces.[8] In these provinces, the local government does not distinguish between rural and urban residents any longer.

The easing of government restrictions on migration appears to have had a significant effect on people's migration decisions. There were 9.2 million people who migrated inter-province between 1990 and 1995 and this number increased to 32 million between 1995 and 2000 and to 38 million between 2000 and 2005. The number of rural migrants increases from 78 million to 145 million within 10 years.[9]

After 2000, the central and local governments in China also proposed some policies to improve working and living conditions of rural migrants. For example, in early 2000, several provinces and cities such as Guangdong, Beijing, Shanghai and Xiamen started to set up social security schemes to cover rural labour migrants. A document issued by the State Council in May 2001 stated that local governments should provide nine years of compulsory education to migrant children through the public school system. Until the end of 2006, only a few lo-

[6] At that time, China was a planned economy. The amount of food for each individual was planned by the government. People needed to use food stamps to exchange food.

[7] A household registration record officially identifies a person as a resident of an area and includes identifying information such as name, parents, spouse, and date of birth.

[8] These 12 provinces are Chongqing, Fujian, Guangxi, Hebei, Hubei, Hunan, Jiangsu, Liaoning, Shandong, Shanxi, Sichuan and Zhejiang.

[9] All numbers referring to the measure of the migrants' number is stock value in this paragraph.

cal governments have actually implemented this policy (Liang, 2007). Although the central and local governments in China tried to change the rural household system and the associated discrimination, Chan (2012) states that the effects of those policies have not been large.

4.3 Data

We use two main sources of data: the first one is the China Household Income Project (CHIP) survey, three waves panel data from 2007 to 2009, the second is China Family Panel Studies (CFPS). The CHIP dataset is designed to study issues such as the effect of rural-urban migration on income mobility and poverty alleviation, the state of education, and the health of children in migrating families. The survey consists of three representative samples of households, including a sample of rural households, a sample of rural migrant households, and a sample of urban households. For longitudinal 2007-2009 data, we focus on the rural sample, which consists of individuals with a rural Hukou, and follow them over time, whether they migrate or not. In the analysis, we use the CHIP rural household survey sample in panel data.[10] Individuals in the rural sample are all born in rural areas and have rural household registration. It is a good representative sample for individuals who were born in rural areas. Also, the attrition rate is very low for the rural sample, i.e., less than 1 % of sample are missing over three round panel data.

Table 3 gives the summary statistics about the data we use for constructing auxiliary regression and moments in the structural model. In the sample, the average years of education is about 9 years. More than 70% of them just graduate

[10]There are two main reasons why we do not use the migration sample. First, the response rate in the migration data is quite low. The attrition rate is above 70% for the three years of panel data. Second, we cannot follow the history of migrants' work experience using the migration samples. For example, migrants who return to their hometowns are not surveyed in the migration sample.

from middle school level education. About 10 % of rural individuals go to some college or college level education. The mean of real log monthly earnings is quite stable over the three years. Also, the migration fraction is very persistent.

The average year of the first time migration in 2007 is 1999. When focusing on this measure across different cohorts, we can find that for the cohort older than 1960, the mean of first time migration year is about 1993 with large standard deviation. For the cohorts younger than 1980, the mean of first time migration is 2002, and most of them migrated just after mandatory education. When looking at the fraction of rural individuals who ever migrate across different cohorts, we can find a pattern that more and more younger cohorts migrate. For example, only 25 % of older cohorts migration but around 84 % of younger cohorts have migration experience.

We also include class performance, whether take college entrance exam as cognitive skill measures; treat the smoke behavior measures as the measures including non-cognitive skills. Only less than 10% have had the experience of taking college entrance exam. More than 50% of rural male smoke. These evidences are all consistent with skill measure datasets.

In order to have better measures for individual cognitive and non-cognitive skills, we supplement the CHIP data with the first three (biannual) waves (2010-2014) of the China Family Panel Studies (CFPS). We use both CFPS and CHIP data to estimate the distributions of individual unobserved skill endowments as in Heckman and Navarro (2007), Navarro and Zhou (2017), etc.[11] As a measurement system for cognitive skills, we include the word, math, and word recall tests from the CFPS data, and from CHIP we include whether the individual takes the college entrance exam, as well as a self- reported measure of class performance. For non-cognitive skills, we use the following categorical measures from CHIPS:

[11]See Section 6 for a detailed description of our procedure.

how easy it is for the individual to concentrate to do things, how capable is the individual of making decisions, how capable are they to face problems by themselves, and whether they lack of sef-confidence. From CFPS we use measures of how well do they get along with others, how often do they feel upset and cannot remain calm, how often they feel everything is difficult, how often can they make decisions by themselves, how often feel not overcome difficulties, how often have to face some problems, and how often not feel confident.[12]

Table 4 gives the summary statistics about the variables in CHIP data for whom with non-cognitive skill measures, which are using for skill estimation. In the sample, the average years of education is close to 8 years. The self reported class performance is concentrated at the normal level. Only 7% of rural individuals have taken the college entrance exam. For non-cognitive skill measures, most people have positive attitudes about their ability of concentrating to do things, making decisions, and overcoming difficulties. There are 70 % of people who ever smoke, and they smoke around 10 cigarettes per day.

Table 5 shows the information about CFPS data. This data provide some measures about individuals' cognitive skills. In this paper, we use six measures: two math, two language, and two word recall test measures as the cognitive skill measures. About non-cognitive skills, we have four measures including a social skill measure about whether people can get well with others. We can find that there are similar findings about the non-cognitive skill measures as what we find in CHIP data: rural individuals have very positive evaluation about their social ability, not worried about their lives, and can concentrate to work. In terms of smoking and drinking behaviors, there are high fraction of rural male who smoke in China. More than 50% of rural male drink more than 500 g alcohol per week, and even 25% of them drink more than 1000 g per week.

[12]Identification of skill density is in Appendix A.

The auxiliary parameters and moments are calculated based on the sequence of CHIP data. 2002 round is cross sectional data. In three years panel data (2007-2009), three representative samples of households were surveyed, including a sample of 8,000 rural households, a sample of 8,000 rural migrant households, and a sample of 5,000 urban households in 9 provinces. In the analysis, we use the CHIP rural household survey sample in panel data.[13] Individuals in the rural sample are all born in rural areas and have rural household registration.

In the analysis, we only focus on males in order to avoid further expanding the model by taking into account joint labor supply and fertility decisions. The sample contains information on work experience, work locations, earnings, and education choices. Using this data, we can construct the education decisions, the location choices, and work statuses for the individuals who are between 16 and 60 years old for the three-year periods. The definition of migration that we have in this paper is whether the urban residence location is out of his rural *hukou* (household registration) county.

5 Structural Model

The evidence in Section 3 shows that, consistent with the mechanism we suggest in Section 2, the existence of a rural-urban migration option can reduce the incentive to invest in human capital for rural individuals. The evidence, however, rests on us arguing that migration became easier after 1982. Furthermore, several other changes, like the massive expansion of the college system in 1999 were put in place during the period we study. To further analyze whether our hypothesized

[13]There are two main reasons why we do not use the migration sample. First, the response rate in the migration data is quite low. The attrition rate is above 70% for the three years of panel data. Second, we cannot follow the history of migrants' work experience using the migration samples. For example, migrants who return to their hometowns are not surveyed in the migration sample.

mechanism is responsible for the pattern we observe, to quantify the magnitude of the migration option on rural education choices, to parse the effects of the college expansion policy on rural individuals decisions, and to quantify the effects that the change in migration policy may have had on income inequality, we now develop a formal structural model of education and migration decisions in China which we take to the data in Section 6.

Let $x_{i,a}$ be the state vector for an individual i, who is of age a. We assume that $x_{i,a}$ contains, among other things a vector of individual endowments $\theta_i = \left(\theta_i^C, \theta_i^N\right)'$ (cognitive and non-cognitive skills) that is unobserved to the econometrician. It also contains the calendar year the individual is at, t. Let $d_{i,a} = k$ if an individual makes choice k, where $k = s$ if the individual chooses to attend school, $k = r$ if he instead works in a rural area (i.e., stays in rural or return migrates to rural), and $k = u$ if he chooses to work in an urban area (i.e., stay in urban or migrate to urban). We assume that once an individual drops out of school he does not return . Consequently, the choice set consists of $\{s, r, u\}$ if $d_{i,a-1} = s$, and of $\{r, u\}$ if $d_{i,a-1} \neq s$.

The utility flow for individual i associated with choice $d_{i,a} = k$ is specified as $u(x_{i,a}, k) + \xi_{i,a}^k$, where $\xi_{i,a}^k$ is a random variable unobserved to the econometrician. For computational simplicity, we assume that $\xi_{i,a}^k$ is distributed extreme value type-I, i.i.d. across locations and across periods, and independent of $x_{i,a}$. We let $\xi_{i,a} = \left(\xi_{i,a}^s, \xi_{i,a}^r, \xi_{i,a}^u\right)'$ if $d_{i,a-1} = s$, and $\xi_{i,a} = \left(\xi_{i,a}^r, \xi_{i,a}^u\right)'$ if $d_{i,a-1} \neq s$.

5.1 Flow Payoffs: Urban

Let $e_{i,a}$ denotes the years of education an individual has acquired up to a, and $\exp_{i,a}^r, \exp_{i,a}^u$ denote accumulated work experience in rural and urban areas respectively. Individuals working in a city receive earnings according to

$$\ln y_{i,a}^{u} = \gamma_0^u + e_{i,a}\gamma_1^u + \exp_{i,a}^{r}\gamma_2^u + \left(\exp_{i,a}^{r}\right)^2\gamma_3^u + \exp_{i,a}^{u}\gamma_4^u + \left(\exp_{i,a}^{u}\right)^2\gamma_5^u + \mathbb{1}_{col}\gamma_6^u + \theta_i'\gamma_7^u + \varepsilon_{i,a}^{u}, \tag{2}$$

where $\varepsilon_{i,a}^{u} \sim N\left(0, \sigma_{\varepsilon^u}^2\right)$ is an i.i.d. shock to earnings. Here, we also allow extra college premium in the earning equation. $\mathbb{1}_{col}$ means college graduates. Since in China not everyone who want to go to college could attend a college, they need to attend college entrance exam. Before 1999, in total less than 35% of individuals who took the exam could get a college offer. Therefore, we expect there exists extra college wage premium γ_6^u. An individual moving from a rural area last period to an urban area this period pays a migration cost given by

$$mc_{i,a,t} = \gamma_0^m + t\gamma_1^m + a\gamma_2^m + a^2\gamma_3^m + \theta_i'\gamma_4^m + \mathbb{1}_{t>1982}\gamma_5^m + \mathbb{1}_{t>1982}t\gamma_6^m + \mathbb{1}_{t>1989}\gamma_7^m + \mathbb{1}_{t>1989}t\gamma_8^m, \tag{3}$$

where we allow for a trend in migration costs (γ_1^m). If $\gamma_1^m < 0$, it would reflect the fact that migration has become easier over time in China (i.e., the improvement of transportation condition). We further include an indicator for the post-1982 period (γ_5^m, and γ_6^m), to account for the policy change described in Section 4. During the year 1988-89, the central government temporary prohibited migration. After 1990, the government began to promote migration again. To capture these policies changes, we want to separately estimate γ_7^m, and γ_8^m. The flow utility of an individual in an urban area is

$$u\left(x_{i,a}, u\right) + \xi_{i,a}^{u} = \ln y_{i,a}^{u} - mc_{i,a,t}\mathbb{1}_{d_{i,a-1}\neq u} + \xi_{i,a}^{u}. \tag{4}$$

5.2 Flow Payoffs: Rural

Individuals working in a rural area receive earnings according to

$$\ln y^r_{i,a} = \gamma^r_0 + e_{i,a}\gamma^r_1 + \exp^r_{i,a}\gamma^r_2 + \left(\exp^r_{i,a}\right)^2\gamma^r_3 + \exp^u_{i,a}\gamma^r_4 + \left(\exp^u_{i,a}\right)^2\gamma^r_5 + \mathbb{1}_{col}\gamma^r_6 + \theta'_i\gamma^r_7 + \varepsilon^r_{i,a}, \tag{5}$$

with $\varepsilon^r_{i,a} \sim N\left(0, \sigma^2_{\varepsilon^r}\right)$. Besides deriving utility from earnings, individuals in rural areas also derive extra utility from being in a rural location (a "home premium", e.g., Kennan and Walker (2011)) given by

$$h_{i,a} = \gamma^h_0 + a\gamma^h_1 + a^2\gamma^h_2 + \theta'_i\gamma^h_3. \tag{6}$$

Finally, an individual moving back to a rural area, pays a return migration cost. Based on the identification argument, we restrict return migration cost function has the same coefficients for the variables which are not affected by the migration policies.

$$\varphi_{i,a,t} = \gamma^m_0 + t\gamma^m_1 + a\gamma^m_2 + a^2\gamma^m_3 + \theta'_i\gamma^m_4. \tag{7}$$

The flow utility of an individual in a rural area is thus given by

$$u\left(x_{i,a}, r\right) + \xi^r_{i,a} = \ln y^r_{i,a} + h_{i,a} - \varphi_{i,a,t}\mathbb{1}_{d_{i,a-1}=u} + \xi^r_{i,a}. \tag{8}$$

5.3 Flow Payoffs: Schooling

An individual derives utility from attending school that depends, in part, on the schooling level that he is attending:

$$\begin{aligned} u\left(x_{i,a}, s\right) + \xi^s_{i,a} = &\left(\gamma^s_{1,hs} + t\gamma^s_{2,hs} + \mathbb{1}_{1965<t<1977}\gamma^s_{3,hs} + \theta'_i\gamma^s_{4,hs}\right)\mathbb{1}_{hs} \\ &+ \left(\gamma^s_{1,voc} + t\gamma^s_{2,voc} + \mathbb{1}_{1965<t<1977}\gamma^s_{3,voc} + \theta'_i\gamma^s_{4,voc}\right)\mathbb{1}_{voc} \\ &+ \left(\gamma^s_{1,col} + t\gamma^s_{2,col} + \mathbb{1}_{1965<t<1977}\gamma^s_{3,col} + \theta'_i\gamma^s_{4,col}\right)\mathbb{1}_{col} + \xi^s_{i,a}, \end{aligned} \tag{9}$$

where $\mathbb{1}_{hs}$ is an indicator that takes value 1 if the individual is attending high school, and similarly for *voc* which stands for vocational college (similar to the US community college) and *col* for college. Since it is very uncommon for students to work in China, we do not allow for part-time work while in school. $\mathbb{1}_{1965<t<1977}$ indicates whether the education choices were taken during the Cultural Revolution periods. During that periods, the government expanded secondary education rapidly but destroyed both college and vocational education system. In the model, we allow individuals to evaluate schools differently between Cultural Revolution periods and not Cultural Revolution periods.

5.4 Individual Choices

The value that an individual gets from making a particular choice can be defined recursively as follows. Let

$$V_a^k\left(x_{i,a}, \xi_{i,a}^k\right) = u(x_{i,a}, k) + \xi_{i,a}^k + \beta E\left[V_{a+1}(x_{i,a+1}, \xi_{i,a+1}) | x_{i,a}, d_{i,a} = k\right]$$

the value that an individual gets if he chooses to attend school ($k = s$), or work on either $k = r$ or $k = u$, and then continues to maximize its utility every period, discounted using β.

First, consider an individual who is done with schooling. He will choose such that

$$d_{i,a} = \arg\max_{k\in\{r,u\}}\left\{V_a^r\left(x_{i,a}, \xi_{i,a}^r\right), V_a^u\left(x_{i,a}, \xi_{i,a}^u\right)\right\},$$

and

$$V_a\left(x_{i,a}, \xi_{i,a}\right) = \max\left\{V_a^r\left(x_{i,a}, \xi_{i,a}^r\right), V_a^u\left(x_{i,a}, \xi_{i,a}^u\right)\right\}.$$

Since this is a lifecycle model, individuals will face a final period A, which we set to age 60. Since we do not model what happens after retirement, etc, we simply model this terminal value as a function of the states at that point. In particular, we normalize

$$V_A\left(x_{i,A}, r\right) = 0 + \xi^r_{i,A},$$

and write

$$V_A\left(x_{i,A}, u\right) = \alpha_0 + e_{i,a}\alpha_1 + \exp^r_{i,a}\alpha_2 + \left(\exp^r_{i,a}\right)^2\alpha_3 + \exp^u_{i,a}\alpha_4 + \left(\exp^u_{i,a}\right)^2\alpha_5 + \theta_i'\alpha_6 + t\alpha_7 + \xi^u_{i,A}.$$

Next consider an individual who is already enrolled in either College or Vocational school. While we assume that the $x_{i,a}$ contains an indicator of the schooling level that the individual is attending, we abuse notation and write $x_{ia} = \left(\tilde{x}_{i,a}, \ell\right)$ for $\ell = \{hs, voc, col\}$, to make it explicit. We thus say that the value of going to school in college is given by $V^s_a\left(\tilde{x}_{i,a}, col, \xi^s_{i,a}\right)$ to recognize that his flow utility will be based on $u\left(\tilde{x}_{i,a}, col, s\right) = \gamma^s_{1,col} + \theta_i'\gamma^s_{2,col}$ in equation 9. Similarly for a vocational student we use $V^s_a\left(\tilde{x}_{i,a}, voc, \xi^s_{i,a}\right)$ to denote the value of going to school in this case. For an individual in college, his decision every period will be given by

$$d_{i,a} = \arg\max_{k\in\{s,r,u\}}\left\{V^s_a\left(\tilde{x}_{i,a}, col, \xi^s_{i,a}\right), V^r_a\left(x_{i,a}, \xi^r_{i,a}\right), V^u_a\left(x_{i,a}, \xi^u_{i,a}\right)\right\}, \tag{10}$$

his value by

$$V_a\left(x_{i,a}, \xi_{i,a}\right) = \max\left\{V^s_a\left(\tilde{x}_{i,a}, col, \xi^s_{i,a}\right), V^r_a\left(x_{i,a}, \xi^r_{i,a}\right), V^u_a\left(x_{i,a}, \xi^u_{i,a}\right)\right\}, \tag{11}$$

and similarly for vocational school.

Now, consider the decision to attend college. An individual gets to decide whether to attend college only if he is admitted into one. We assume that the probability of getting a college offer is given by:

$$\lambda_{i,t}^{col} = \frac{\exp(\gamma_0^{col} + \theta_i' \gamma_1^{col} + t\gamma_2^{col} + t\mathbb{1}_{t \geq 1999}\gamma_3^{col} + \mathbb{1}_{t \geq 1999}\gamma_4^{col})}{1 + \exp(\gamma_0^{col} + \theta_i' \gamma_1^{col} + t\gamma_2^{col} + t\mathbb{1}_{t \geq 1999}\gamma_3^{col} + \mathbb{1}_{t \geq 1999}\gamma_4^{col})}, \tag{12}$$

where γ_3, and γ_4 are included to account for the massive sustained increase in the number of college offers that happened since 1999, as shown in Figure @@@@. Individuals who do not get a college offer, may still attend a vocational college. Hence, the value for an individual who has a college offer, will be

$$V_a\left(x_{i,a}, \xi_{i,a}\right) = \max\left\{V_a^s\left(\tilde{x}_{i,a}, col, \xi_{i,a}^s\right), V_a^s\left(\tilde{x}_{i,a}, voc, \xi_{i,a}^s\right), V_a^r\left(x_{i,a}, \xi_{i,a}^r\right), V_a^u\left(x_{i,a}, \xi_{i,a}^u\right)\right\}, \tag{13}$$

while an individual who does not have a college offer (either because he did not get one, or because he did not take the college entrance exam) will decide based on

$$V_a\left(x_{i,a}, \xi_{i,a}\right) = \max\left\{V_a^s\left(\tilde{x}_{i,a}, voc, \xi_{i,a}^s\right), V_a^r\left(x_{i,a}, \xi_{i,a}^r\right), V_a^u\left(x_{i,a}, \xi_{i,a}^u\right)\right\}. \tag{14}$$

Next, consider someone who is deciding whether to enroll in their last year of high school. The timing is as follows. An individual first decides whether to enroll in the last year of high school before he gets to observe the cost of the entrance exam. Later in the same year the cost of the exam is realized and the individual decides whether to take the exam. If he takes the college exam, then at the beginning of the next year he draws according to 12 and decides based on 13. If he does not have a college offer then he chooses based on 14.

If the individual decides to take the college entrance exam, he has to pay the (psychic) cost of the college entrance exam, given by

$$ce_i = \gamma_0^{ce} + \theta_i^{'} \gamma_1^{ce} + \varepsilon_i^{ce}.$$

The decision of whether to take the college entrance exam is based on his expected value of attending college or vocational college, where the expectation is taken with respect to $\xi_{i,a}, \varepsilon^r_{i,a}, \varepsilon^u_{i,a}$. If we let

$$\begin{aligned} E\mathcal{V}^{col}\left(x_{i,a}\right) &= E_a \max\{V^s_{a+1}\left(\tilde{x}_{i,a+1}, col, \xi^s_{i,a+1}\right), V^s_{a+1}\left(\tilde{x}_{i,a+1}, voc, \xi^s_{i,a+1}\right), \\ & V^r_{a+1}\left(x_{i,a+1}, \xi^r_{i,a+1}\right), V^u_{a+1}\left(x_{i,a+1}, \xi^u_{i,a+1}\right)\}, \end{aligned} \tag{15}$$

and

$$\begin{aligned} E\mathcal{V}^{voc}\left(x_{i,a}\right) &= E_a \max\{V^s_{a+1}\left(\tilde{x}_{i,a+1}, voc, \xi^s_{i,a+1}\right), V^r_{a+1}\left(x_{i,a+1}, \xi^r_{i,a+1}\right), \\ & V^u_{a+1}\left(x_{i,a+1}, \xi^u_{i,a+1}\right)\}, \end{aligned} \tag{16}$$

then, he will take the exam if

$$\beta\lambda^{col}_{i,t} E\mathcal{V}^{col}\left(x_{i,a}\right) - ce_i > \beta\left(1 - \lambda^{col}_{i,t}\right) E\mathcal{V}^{voc}\left(x_{i,a}\right).$$

Let

$$\mathcal{V}^{hs}\left(x_{i,a}\right) = E\max\left\{\beta\lambda^{col}_{i,t} E\mathcal{V}^{col}\left(x_{i,a}\right) - ce_i, \beta\left(1 - \lambda^{col}_{i,t}\right) E\mathcal{V}^{voc}\left(x_{i,a}\right)\right\},$$

that is the value of attending high school before the individual gets to observe the cost of taking the exam, i.e., before ε^{ce}_i is realized. Hence, the value of enrolling in the last year of high school for this individual is

$$V^s_a\left(\tilde{x}_{i,a}, hs, \xi^s_{i,a}\right) = u(\tilde{x}_{i,a}, hs, s) + \xi^s_{i,a} + \mathcal{V}^{hs}\left(x_{i,a}\right).$$

Hence, the individual will enroll in the last year of high school if

$$V^s_a\left(\tilde{x}_{i,a}, hs, \xi^s_{i,a}\right) > \max\left\{V^r_a\left(x_{i,a}, \xi^r_{i,a}\right), V^u_a\left(x_{i,a}, \xi^u_{i,a}\right)\right\},$$

and his value will be

$$V_a\left(x_{i,a},\xi_{i,a}\right)=\max\left\{V_a^s\left(\tilde{x}_{i,a},hs,\xi_{i,a}^s\right),V_a^r\left(x_{i,a},\xi_{i,a}^r\right),V_a^u\left(x_{i,a},\xi_{i,a}^u\right)\right\}.$$

Finally, in any previous year, the choice for an individual deciding whether to attend (or remain) in high school is

$$d_{i,a}=\arg\max_{k\in\{s,r,u\}}\left\{V_a^s\left(\tilde{x}_{i,a},hs,\xi_{i,a}^s\right),V_a^r\left(x_{i,a},\xi_{i,a}^r\right),V_a^u\left(x_{i,a},\xi_{i,a}^u\right)\right\},$$

and his value is given by

$$V_a\left(x_{i,a},\xi_{i,a}\right)=\max\left\{V_a^s\left(\tilde{x}_{i,a},hs,\xi_{i,a}^s\right),V_a^r\left(x_{i,a},\xi_{i,a}^r\right),V_a^u\left(x_{i,a},\xi_{i,a}^u\right)\right\}.$$

6 Identification and Estimation

6.1 Identification

In this section, we briefly sketch how identification of all the parameters in the structural model is obtained.

First, to identify the joint distribution of unobserved skills, we use two different data sets which include both cognitive and non-cognitive measures of individual's skills. Following the method of Cunha, Heckman, and Schennach (2010), we can identify the joint distribution of two different unobserved skills.

Second, we assume the location specific offered log earning distributions are normal distribution. Therefore, the observed log earnings identify the location specific earning function and variances of earning shocks $\sigma^2_{\varepsilon^u}$, and $\sigma^2_{\varepsilon^r}$.

Third, to identify location amenity value, migration and return migration costs, since we only can identify relative location amenity value, we normalize the location amenity value of living in urban areas equals to zero. Then we follow Kennan and Walker (2011) identification strategy to separate identify location amenity value and migration costs.

Fourth, since we observe individuals drop out from different levels of schooling, these choices help us to identify the level of flow utility at different schooling levels. Given the flow utility level at different schooling levels, the probability of getting college offer, $\lambda_{i,t}^{col}$ can be identified when comparing the fraction of individuals who gets college offers and the fraction of individuals who taking the college entrance exam.

Identification of the other components of our model, follows directly from the analysis of Heckman and Navarro (2007), and French and Taber (2011).

6.2 Estimation

We use the indirect inference method to estimate our model. There are two sets of parameters we need to estimate. One set of parameters are the density of unobserved endowments, Θ_1. The second set of parameters are the rest parameters in the structural model Θ_2.

To estimate the joint distribution of two types of unobserved skills, we use both cognitive and non-cognitive skill measures in both CHIP and CFPS data sets. For non-continuous measures, we assume they follow ordered probit process. The auxiliary regressions and skill distribution parameters are estimated by indirect inference method.

The estimation procedure is as follows:

1. First, we estimate Θ_1 using the two types of skill measures in the data
2. Then, given Θ_1, with the initial guess of Θ_2, we solve the dynamic discrete choice model recursively

3. Given the solution of the model from Step 2, we simulate individuals' decisions and get auxiliary parameters and moments. Using GMM estimation methods, we update parameters Θ_2

4. After repeating Step 2-3 R_1 times, we update Θ_1

5. Get Θ_1, Θ_2 until they converge

6.2.1 Skill Distribution Estimation

We group data measures into three different categories: cognitive skill measures, non-cognitive skill measures, and measures for both skills. In CFPS data, we use 6 cognitive skill measures (math and Chinese test in 2010 and 2014), and immediate memory record and delayed memory record in 2012; three discrete non-cognitive skill measures and three measures for both cognitive and non-cognitive skills. In CHIP Data, we use two cognitive skill measures (self-reported class performance, and whether take college entrance exam); thirteen non-cognitive skill measures. Also, two measures which are evaluated in both dataset (i.e., one non-cognitive skill measure, and one measure for both skills).

For continuous skill measures, we use OLS regression as potential model, and also use the same regression as auxiliary regression. In the OLS model, we control for the age and years of education when individuals who take the tests.

For discrete skill measures, we model the choices are formed as the following ordered probit process:

$$y_i = \begin{cases} 1, & \text{if} \quad x_i\beta_1 + u_i \geq M_1 \\ k, & \text{if} \quad M_{k-1} \leq x_i\beta_1 + u_i \leq M_k \\ K, & \text{if} \quad x_i\beta_1 + u_i \geq M_{K-1} \end{cases} \tag{17}$$

Since we use the indirect inference method to estimate the skill distributions, we use the following OLS regression as the auxiliary regression:

$$1_{y_i=1} = \sum_{k=2}^{K} \gamma_k 1_{y_i=k} + x_i' \beta_2 + \varepsilon_i \tag{18}$$

For the measures which are in both datasets, we also add data dummy variables.

6.2.2 Structural Model Estimation

For the structural model, we choose three wage regressions: the first wage regression is comparable with the reduced form studies. We pool migrants and non-migrants together and control for age, years of education, migration status, the interaction term of migration times years of education, college graduate indicator, and cognitive and non-cognitive skills. The other two wage regressions are the wage regression for whom never migrate and for the individuals we can observe the full migration history.

We have two migration and two return migration regressions to pin down migration and return migration decisions. There are four regressions for different education level choices. Also, we have the moments for different cohorts: the first year of migration and fraction of ever migration by 2007 across four cohorts.

7 Estimation Results

7.1 Model Fit

As we described in Section 6, we use the indirect inference method to estimate the skill distributions first, and then estimate the structural model. Tables 17-23 provide auxiliary models for the estimates for skill distributions. The joint moments of model Goodness fit test $\chi^2_{328} = 165.53$ and the probability (right-tail) is almost equal to 1, which shows that the model estimates fits are really well. For example, Table 16 shows the auxiliary regression for the cognitive skill measures using CFPS data. All model estimates are in the one standard deviation of the estimates from data regression.

Tables 13-15 give the auxiliary model regression parameters and moments. The model joint moments goodness fit test $\chi^2_{110} = 22.48$, and the probability (right-tail) is also close to 1. We cannot reject the model estimates from data moments or regression parameters.

7.2 Estimation Results

We can find structural model estimation parameters from Tables 9 to 12. In Table 9, we show the estimates for rural and urban earning equations. First, as we expected, the education return rate is higher in rural areas than that in urban areas for individuals with rural Hukou. In rural areas, the monthly education return rate is around 0.08, and the return in urban areas is around 0.04. The rural education return rate is almost double than that in urban areas. But one thing should be noticed is that the extra college graduate premium is higher in urban areas. For example, the extra college premium is closed to 11 percent in urban areas, however, in rural areas, there is almost no significant extra college premium.

Furthermore, when we focus on the constant term for both earning equations, the constant term (i.e., 5.76) in urban equations is much higher than that (i.e., 4.77) in rural earning equation. This big difference would come from the productivity differences between rural and urban areas, and the general equilibrium effect of the relative high skilled and low skilled workers substitution in rural and urban areas.

Both the findings about the differences of education return and the constant term in rural and urban areas support our reduced form analysis, which is the term A in Equation 1 is negative. Later we examine how the migration policy affect rural individuals education choices.

The role of cognitive and non-cognitive skills are rewarded differently in rural and urban areas. We can find that cognitive skill has no significant return, but non-cognitive skills has positive return in rural areas. However, cognitive skill gives a significantly positive return in urban areas and non-cognitive skill does not reward in urban earnings.

Table 14 gives the estimates for migration costs. We can find that the migration policy introduced in 1983 and 1990 significantly reduce the migration costs. Regardless the effect from policies, the model also capture the decreasing migration cost trend over time, which can be explained as the development of transportation system in China. One interesting finding is that non-cognitive skill could reduce migration costs with a large magnitude. It would reflect social skills would help rural individuals to migrate to urban areas, which is consistent with Zhou (2017).

In Table 11, we can see the utility values for different levels of education. First, individuals evaluate most about college, then some college, and least for high school level education. In general, individuals with high cognitive skills have higher utility values for schooling. From the parameters of the probability of getting college offer, we can see the college entrance exam will have positive effect on college admission. One interesting thing in psychic cost of taking college entrance exam is that individuals with high non-cognitive skills have lower psychic costs of taking exams.

In Figures 5, we compare the years of education distributions between rural migrants and rural non-migrants. We can find that migrants have higher education level. Most individuals with some college or college education choose to work in urban areas. Figure 6 shows the log earning distribution comparison between migrants and non-migrants. The earnings have a larger dispersion in rural areas than those in urban.

8 Policy Simulations

In this section, we examine how the migration policy and college expansion policy affect individuals' education and migration decisions.

Table 12 gives us the counterfactual simulations under different policy environment. In the column of "No Migration Policy", we set policy parameters in migration costs equal to zero, which we shut down all migration promotion policies, but in this case, we still allow the college expansion policy in 1999 in this case. In the column of " No College Expansion", we set policy related parameters in admission offer probability equals zero to eliminate the impact of education policy. In the column of " No policy", we eliminate both policies in the model.

To evaluate the impact of migration policy, first, we want to examine the impact on individuals' education choices. To analyze this, we compare the column of "No College Expansion" with the column of "No Policy". When we eliminating the effect from college expansion policy, the migration policy would reduce average years of education by 0.4 years. Furthermore, the fraction of rural individuals who attending high school would increase to 36 percentage points if there were no migration policies. That means the high school attendance rate would increase round 29%. The number of individuals who go to some college would be double if there were no migration policies and college attendance rate would increase around 24%. From here, we can see the migration policy of reducing migration costs significantly reduce the rural individuals' incentive to invest in their human capital.

When comparing the column of “No Migration Policy” with “No policy”, we can evaluate the effect of college expansion policy. In general, the impact is very small and almost no impact on individuals’ years of education or attendance rates across different education levels. Figure 7 shows the distribution changes across different policy environment. It is clearly that the migration policy reduce the incentive for rural individuals to attending high schools.

Now let us switch our attention to migration behaviors and earning outcomes. We give the comparison of migration fraction across four different cohorts. When erasing the promoting migration policy, in general less individuals migrate, especially the younger cohorts are less likely to migrate. Education policy does not change people’s migration behaviors. Since when eliminating the migration policy, rural individuals would like to invest in their education more, which increase their log earnings in general.

9 Conclusion

In this paper, we reexamine the question who migrate and how migration affect the labor market outcomes in sending locations with the considering of individuals can endogenously choose their education. To examine why the migration option may reduce the education incentive, first, we give a simplified model and show a condition under which potential migrants would have less incentive to invest in their education. Then, we reduced form analyze whether the condition holds by using the rural urban migration in China. To quantify the impact of migration policy, which significantly reducing the migration costs, we also structurally estimate a dynamic discrete choices model. We find that eliminating the policies, promoting rural urban migration, the attendance rate for high school, some college and college levels would increase 29%, 141%, and 24% respectively.

References

BEINE, M., F. DOCQUIER, AND H. RAPOPORT (2001): "Brain Drain and Economic Growth: Theory and Evidence," *Journal of Development Economics*, 64, 275 – 289. 1

——— (2008): "Brain Drain and Human Capital Formation in Developing Countries: Winners and Losers," *The Economic Journal*, 118, 631–652. 1

BORJAS, G. J. (1987): "Self-Selection and the Earnings of Immigrants," *The American Economic Review*, 77, 531–553. 1

——— (1995): "The Economic Benefits from Immigration," *Journal of Economic Perspectives*, 9, 3–22. 1

BRAUW, A. AND J. GILES (2017): "Migrant Opportunity and the Educational Attainment of Youth in Rural China," *The Journal of Human Resources*, 52, 272–311. 1

CARD, D. (2005): "Is the New Immigration Really so Bad?" *The Economic Journal*, 115, 300–323. 1

CHAN, K. W. (2012): "Migration and Development in China: Trends, Geography and Current Issues," *Migration and Development*, 1, 187–205. 4.2

CHIQUIAR, D. AND G. HANSON (2005): "International Migration, Self-Selection, and the Distribution of Wages: Evidence from Mexico and the United States," *Journal of Political Economy*, 113, 239–281. 1

CUNHA, F., J. J. HECKMAN, AND S. M. SCHENNACH (2010): "Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Econometrica*, 78, 883–931. 6.1

Dustmann, C. and A. Glitz (2011): "Chapter 4 - Migration and Education," in *Handbook of The Economics of Education*, ed. by E. A. Hanushek, S. Machin, and L. Woessmann, Elsevier, vol. 4 of *Handbook of the Economics of Education*, 327 – 439. 1, 2

French, E. and C. Taber (2011): "Chapter 6 - Identification of Models of the Labor Market," in *Handbook of Labor Economics*, ed. by O. Ashenfelter and D. Card, Elsevier, vol. 4, 537 – 617. 6.1

Harris, J. R. and M. P. Todaro (1970): "Migration, Unemployment and Development: A Two-Sector Analysis," *American Economic Review*, 60, 126–142.

Heckman, J. ., J. . Stixrud, and S. . Urzua (2006): "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior," *Journal of Labor Economics*, 24, 411–482. 1

Heckman, J. and B. Singer (1984): "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," *Econometrica*, 52, 271–320.

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

Heckman, J. J. and S. Navarro (2007): "Dynamic discrete choice and dynamic treatment effects," *Journal of Econometrics*, 136, 341 – 396. 4.3, 6.1

Heckman, J. J. and G. Sedlacek (1985): "Heterogeneity, Aggregation, and Market Wage Functions: An Empirical Model of Self-Selection in the Labor Market," *Journal of Political Economy*, 93, 1077–1125.

Hendricks, L. and T. Schoellman (forthcoming): "Human Capital and Development Accounting: New Evidence from Wage Gains at Migration," *The Quarterly Journal of Economics.* 1

Keane, M. P. and K. I. Wolpin (1997): "The Career Decisions of Young Men," *Journal of Political Economy*, 105, 473–522.

Kennan and Walker (2011): "The Effect of Expected Income on Individual Migration Decisions," *Econometrica*, 79, 211–251. 5.2, 6.1

Knight, J., S. Li, and Q. Deng (2009): "Education and the Poverty Trap in Rural China: Setting the Trap," *Oxford Development Studies*, 37, 311–332. 1

Liang, Z. (2007): *Chapter 12 Internal Migration: Policy Changes, Recent Trends and New Challenges.*, New York: Oxford Unveristy Press. 4.2

Maria, C. D. and E. A. Lazarova (2012): "Migration, Human Capital Formation, and Growth: An Empirical Investigation," *World Development*, 40, 938 – 955. 1

Maria, C. D. and P. Stryszowski (2009): "Migration, Human Capital Accumulation and Economic Development," *Journal of Development Economics*, 90, 306 – 313. 1

Meng, X. (2000): *Labout Market Reform in China*, Cambridge University Press.

Mountford, A. (1997): "Can a Brain Drain be Good for Growth in the Source Economy?" *Journal of Development Economics*, 53, 287 – 303. 1

Navarro, S. and J. Zhou (2017): "Identifying Agents Information Sets: an Application to a Lifecycle Model of Schooling, Consumption and Labor Supply," *Review of Economic Dynamics*, 25, 58–92. 1, 4.3

Stark, O., C. Helmenstein, and A. Prskawetz (1997): "A Brain Gain with a Brain Drain," *Economics Letters*, 55, 227 – 234. 1

——— (1998): "Human Capital Depletion, Human Capital Formation, and Migration: a Blessing or a Curse?" *Economics Letters*, 60, 363 – 367.

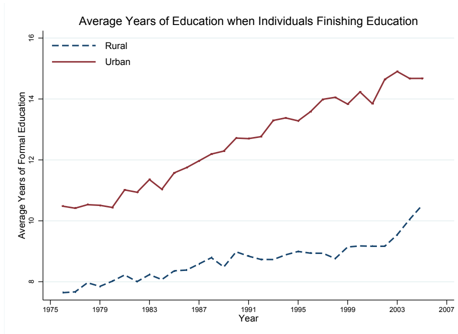
Stark, O. and Y. Wang (2002): "Inducing Human Capital Formation: Migration as a Substitute for Subsidies," *Journal of Public Economics*, 86, 29 – 46. 1

Vidal, J.-P. (1998): "The Effect of Emigration on Human Capital Formation," *Journal of Population Economics*, 11, 589–600. 1

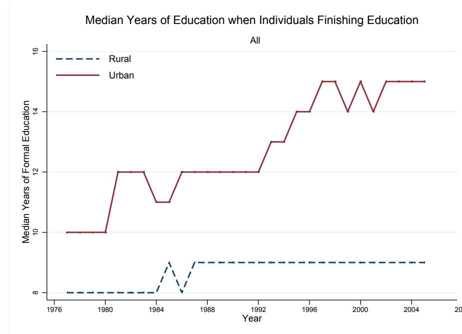
Zhao, Y. (1997): "Labor Migration and Returns to Rural Education in China," *American Journal of Agricultural Economics*, 79, 1278–87.

Zhou, J. (2017): "Internal Migration with Social Networks in China," *Working paper*. 1, 7.2

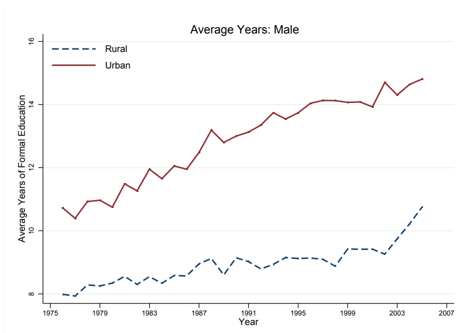
Figures



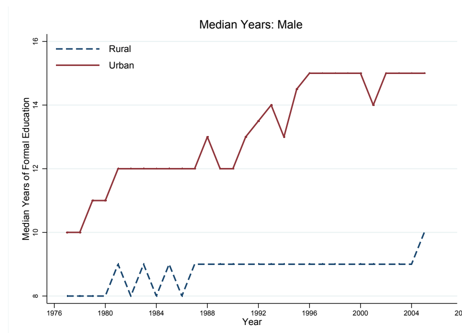
(a) All individuals (Average)



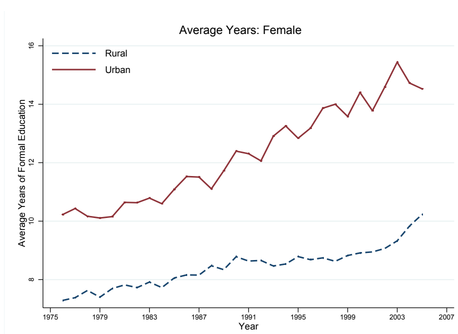
(b) All individuals (Median)



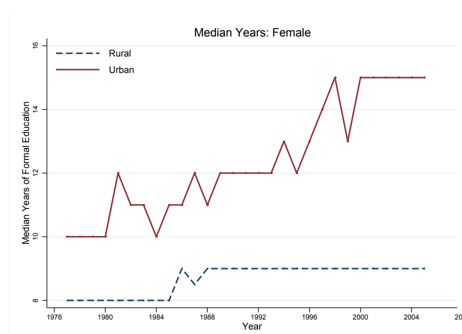
(c) Male (Average)



(d) Male (Median)



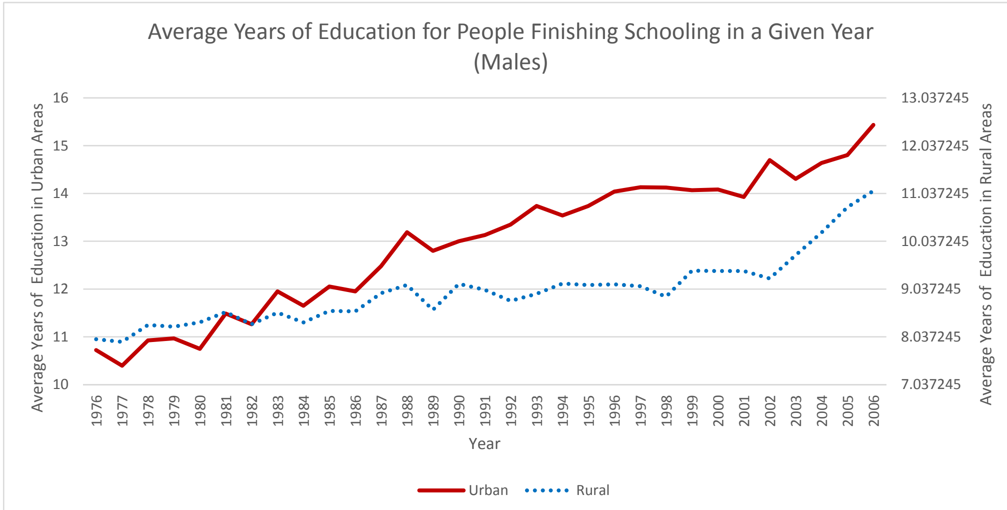
(e) Female (Average)



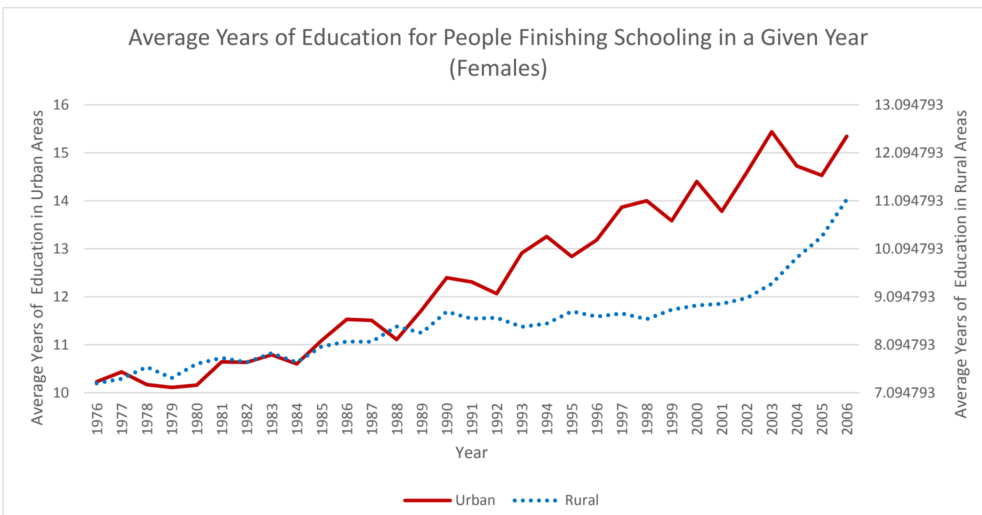
(f) Female (Median)

Figure 1: The Comparison of Years of Education in China

Figure 2: Rural Residents College Attendance Rate across Different Cohorts



(a) Male (Average)



(b) Female (Average)

Figure 3: **The Comparison of Years of Education in China**

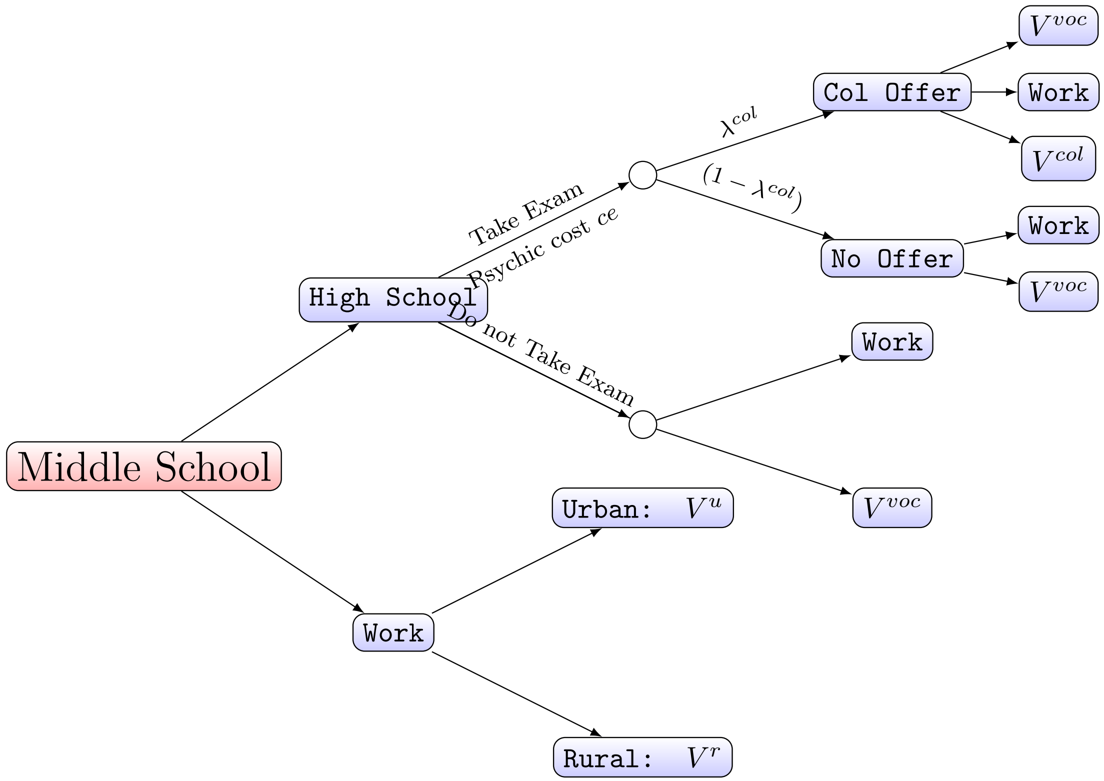


Figure 4: Model Timing: Education Decisions

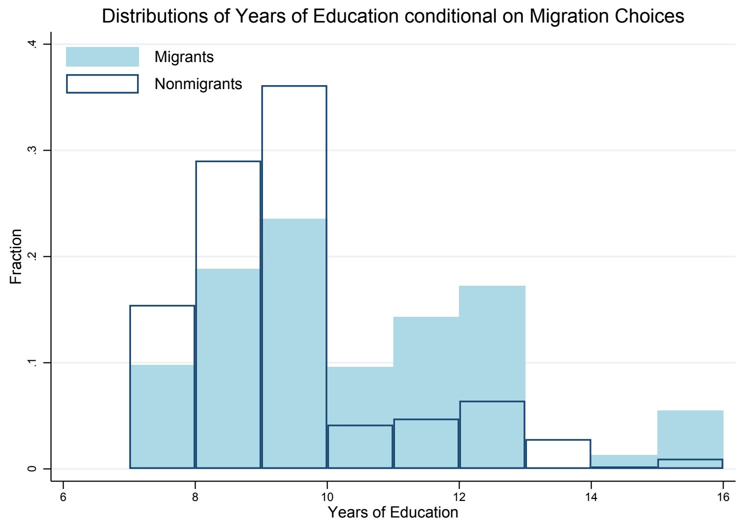


Figure 5: Distribution of Years of Education conditional on Migration Choices

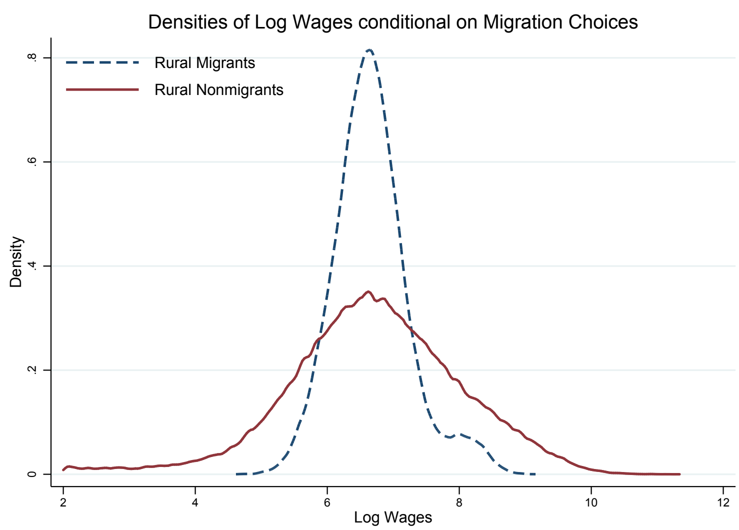
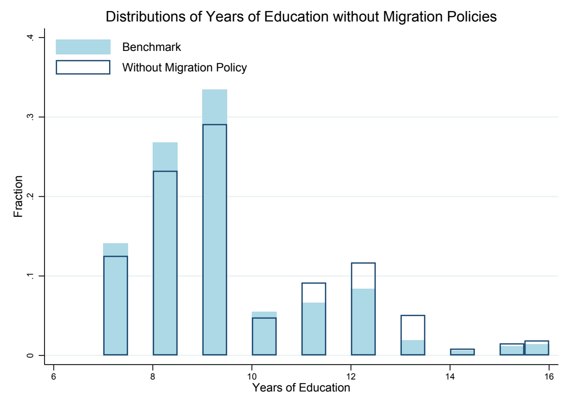


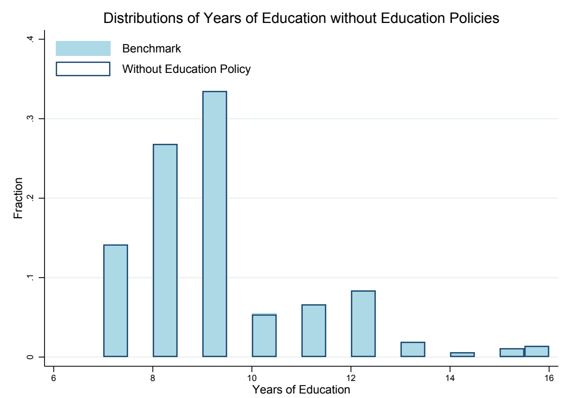
Figure 6: Log Wage Distributions conditional on Migration Choices

Figure 7: The Change of Distributions of Years of Education
(Under Different Policies)

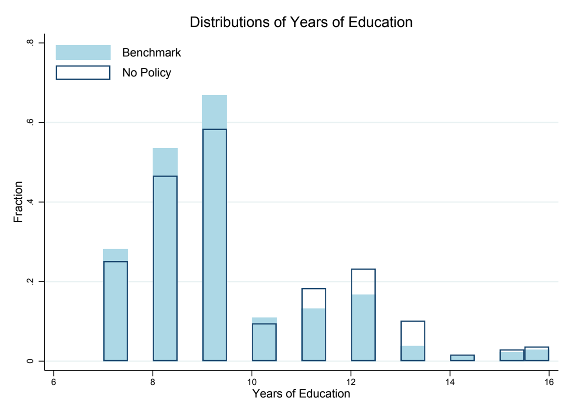
(a) No Migration Policy



(b) No Education Policy)



(c) No Policy



Data

Table 1

Data Summary Statistics

Variables	Observations	Mean	Standard Deviation	Min	Max
		Urban			
Birth	4,251	1969.04	11.23	1951	1991
Years of Education	4,251	12.34	2.88	7	22
Class Performance	3,829	2.32	0.74	1	5
Whether Take College Entrance Exam	4,251	0.34	0.47	0	1
Can you concentrate to do something	3,037	1.61	0.66	1(Can)	4(Cannot)
Are you capable of making decisions	3,039	1.56	0.52	1(Always)	4(Do not)
Do you feel you could not overcome difficulties?	3,039	1.59	0.50	1(Never)	4(Very Often)
Are you able to face problem?	3,039	1.53	0.61	1(Never)	4(Always)
Do you always lack of confidence	3,039	1.41	0.47	1(Not at All)	4(Very Seriously)
		Rural			
Birth	8,292	1971.84	11.51	1951	1991
Years of Education	8,292	9.22	1.91	7	22
Class Performance	8,138	2.61	0.68	1	5
Whether Take College Entrance Exam	8,292	0.09	0.29	0	1
Can you concentrate to do something	4,873	1.46	0.63	1(Can)	4(Cannot)
Are you capable of making decisions	4,873	1.57	0.54	1(Always)	4(Do not)
Do you feel you could not overcome difficulties?	4,872	1.56	0.50	1(Never)	4(Very Often)
Are you able to face problem?	4,874	1.51	0.56	1(Never)	4(Always)
Do you always lack of confidence	4,872	1.42	0.48	1(Not at All)	4(Very Seriously)

Table 2

Data Statistics Between Rural Migrants and Rural Non-migrants

Variables	Observations	Mean	Standard Deviation	Min	Max
	Rural Non-migrants				
Log Monthly Earnings	4,543	6.51	1.11	-0.21	11.26
Birth	4,543	1966.82	10.83	1950	1991
Years of Education	4,543	9.01	1.70	7	19
Class Performance	4,543	2.59	0.68	1	5
Whether Take College Entrance Exam	4,543	0.09	0.28	0	1
Smoke	4,543	0.60	0.49	0	1
How many cigarettes usually smoke per day	4,543	8.58	10.46	0	91
	Rural Migrants				
Log Monthly Earnings	2,148	6.57	0.57	-0.08	10.19
Birth	2,148	1977.81	8.75	1950	1991
Years of Education	2,148	9.26	1.84	7	19
Class Performance	2,148	2.71	0.65	1	5
Whether Take College Entrance Exam	2,148	0.10	0.29	0	1
Smoke	2,148	0.50	0.50	0	1
How many cigarettes usually smoke per day	2,148	6.04	8.99	0	60

Table 3

The CHIPs Data Summary Statistics (For Auxiliary Model)

Variables	Observations	Mean	Standard Deviation	Min	Max
		2007-2009 Panel Data			
Birth	6,691	1970.35	11.42	1950	1991
Years of education	6,691	9.09	1.75	7	19
Education levels	6,691	1.43	0.80	1	5
Log monthly earnings					
in 2007	6,691	6.53	0.97	-0.21	11.26
in 2008	6,691	6.52	0.96	-0.26	10.43
in 2009	6,691	6.59	1.00	-0.26	10.67
Migration status					
in 2007	6,691	0.32	0.47	0	1
in 2008	6,691	0.31	0.46	0	1
in 2009	6,691	0.34	0.47	0	1
The year of first time migration	3,275	1999.33	6.80	1967	2008
Ever migrate	6,691	0.58	0.49	0	1
Class performance	6,691	2.63	0.68	1	5
Whether take college entrance exam	6,691	0.09	0.28	0	1
Smoke	6,691	0.57	0.50	0	1
How many cigarettes usually smoke per day	6,691	7.76	10.08	0	91

Table 4

CHIPS Data Cognitive and Non-Cognitive Skills Summary Statistics

Variables	Mean	Standard Deviation	Min	Max
Birth	1960.80	7.78	1950	1989
Years of education	7.87	2.32	1	19
Age at first round Test	46.20	7.78	18	57
Class performance	2.61	0.70	1	5
Taking college entrance exam	0.07	0.25	0	1
Non-Cognitive skill measures				
Can you concentrate to do something?				
2007	1.45	0.69	1 (Can)	4 (Cannot)
2008	1.50	0.72	1 (Can)	4 (Cannot)
2009	1.58	0.74	1 (Can)	4 (Cannot)
Are you capable of making decisions?				
2007	1.59	0.58	1 (Always)	4 (Do not)
2008	1.64	0.61	1 (Always)	4 (Do not)
2009	1.57	0.56	1 (Always)	4 (Do not)
Do you feel you could not overcome difficulties?				
2007	1.50	0.61	1 (Never)	4 (Very Often)
2008	1.52	0.63	1 (Never)	4 (Very Often)
2009	1.59	0.64	1 (Never)	4 (Very Often)
Are you able to face problem?				
2007	1.50	0.61	1 (Never)	4 (Always)
2008	1.52	0.63	1 (Never)	4 (Always)
2009	1.59	0.64	1 (Never)	4 (Always)
Do you always lack of confidence?				
2007	1.46	0.54	1 (Not at All)	4 (Very Seriously)
2008	1.45	0.54	1 (Not at All)	4 (Very Seriously)
Whether Smoke	0.69	0.46	0	1
How many cigarettes usually smoke per day	9.91	10.83	0	91
Observations	3,742			

Table 5

CFPS Data Cognitive and Non-Cognitive Skills Summary Statistics

Variables	Mean	Standard Deviation	Min	Max
Birth	1966.99	10.70	1950	1991
Years of education	8.32	4.01	0	19
Age at first round test	43.01	10.70	19	60
Standardized cognitive skills				
Math in 2010	0.29	0.87	-1.47	2.15
Language in 2010	0.28	0.84	-1.56	1.60
Immediately word recall in 2012	0.11	0.91	-2.26	2.82
Delayed word recall in 2012	0.10	0.93	-1.53	3.09
Math in 2014	0.27	0.89	-1.42	2.24
Language in 2014	0.27	0.87	-1.44	1.63
Get on well with others	4.07	0.84	1 (Very hard)	5 (Very easy)
Feel upset and cannot remain calm	4.56	0.79	1 (Almost everyday)	5 (Never)
Feel everything is difficult	4.46	0.86	1 (Almost everyday)	5 (Never)
Can you concentrate to do something	1.55	0.74	1 (Almost never)	4 (Most of the time)
Whether Smoke	0.59	0.49	0	1
How many cigarettes usually smoke per day	11.03	17.45	0	400
How much alcohol do you drink last month in 2010	21.33	52.03	0	770
How much alcohol do you drink last week in 2012	11.42	11.69	0	80
How much alcohol do you drink last week in 2014	11.32	11.46	0	100
Observations	6,033			

Reduced Form Analysis Tables

Table 6

Return to Education
(Rural Males)

	2007 Data			
Years of Education×Migration	-0.036***	-0.037**	-0.037***	-0.037**
	(0.014)	(0.014)	(0.014)	(0.014)
Age	0.124***	0.123***	0.123***	0.122***
	(0.007)	(0.007)	(0.007)	(0.007)
Age2	-0.002***	-0.002***	-0.002***	-0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
Years of Education	0.056***	0.056***	0.050***	0.050***
	(0.008)	(0.008)	(0.009)	(0.009)
Migration	0.368***	0.366**	0.380***	0.379***
	(0.130)	(0.130)	(0.131)	(0.131)
Constant	3.982***	3.986***	4.080***	4.089***
	(0.149)	(0.149)	(0.167)	(0.167)
Non-Cognitive Skill Measures	No	Yes	No	Yes
Cognitive Skill Measures	No	No	Yes	Yes
Observations	6,691	6,691	6,691	6,691

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Rural migrants are the individuals working in the urban area, which is out of the county of their rural hukou location.

Table 7

Migration and Education Decisions: Pretrend Analysis

(Males)

	Explained Variable: Years of Education			
Year 1977× Rural	0.646*	0.609	0.280	0.268
	(0.319)	(0.318)	(0.284)	(0.284)
Year 1978× Rural	0.303	0.307	0.411	0.426
	(0.321)	(0.321)	(0.286)	(0.286)
Year 1979× Rural	0.317	0.330	0.142	0.147
	(0.327)	(0.326)	(0.290)	(0.290)
Year 1980× Rural	0.532	0.507	0.159	0.155
	(0.321)	(0.320)	(0.287)	(0.288)
Year 1981× Rural	0.160	0.159	0.207	0.210
	(0.330)	(0.329)	(0.293)	(0.294)
Year 1977	-1.015***	-0.985***	-0.691**	-0.680**
	(0.264)	(0.263)	(0.236)	(0.236)
Year 1978	-0.319	-0.324	-0.521*	-0.532*
	(0.266)	(0.265)	(0.237)	(0.238)
Year 1979	-0.367	-0.382	-0.315	-0.315
	(0.273)	(0.272)	(0.242)	(0.242)
Year 1980	-0.492	-0.474	-0.206	-0.203
	(0.266)	(0.265)	(0.239)	(0.239)
Year 1981	0.094	0.114	-0.013	-0.008
	(0.277)	(0.277)	(0.247)	(0.247)
Rural	-3.058***	-3.060***	-2.361***	-2.375***
	(0.238)	(0.238)	(0.214)	(0.214)
Constant	11.358***	12.243***	16.500***	16.906
	(0.196)	(0.269)	(0.271)	(0.308)
Cognitive Skill Measures	No	No	Yes	Yes
Non-Cognitive Skill Measures	No	Yes	No	Yes
Observations	2,367	2,367	2,348	2,348

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Year is the calender year when individuals finish their education.

Rural is an indicator variable for individuals with rural Hukou.

Table 8

The Effect of Relaxing Migration Restrictions on Education
(Males)

Explained Variable: Years of Education				
$> 1982 \times$Rural	-1.939***	-1.919***	-1.007***	-1.013***
	(0.081)	(0.082)	(0.069)	(0.070)
> 1982	2.910***	2.894***	1.814***	1.803***
	(0.065)	(0.066)	(0.057)	(0.058)
Rural	-1.965***	-1.975***	-1.584***	-1.592***
	(0.062)	(0.062)	(0.051)	(0.051)
Constant	10.605***	11.396***	11.309***	11.679***
	(0.051)	(0.107)	(0.072)	(0.102)
Cognitive Skill Measures	No	No	Yes	Yes
Non-Cognitive Skill Meaures	No	Yes	No	Yes
Observations	12,543	12,543	11,967	11,967

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

> 1982 is an indicator for whether the individual finishes his schooling after the migration reform 1982.

10 Structural Model Estimates

Table 9

Structural Model Estimation Parameters (Part I)

Earning Equations		
	Rural	Urban
Years of Education	0.0777	0.0398
	(0.0366)	(0.0052)
Rural Work Experience	0.1658	0.0206
	(0.0246)	(0.0006)
Urban Work Experience	1.5480	0.1606
	(0.1266)	(0.0018)
Rural Work Experience^2	-0.0044	-0.0003
	(0.0063)	(0.0027)
Urban Work Experience^2	-0.3347	-0.0045
	(0.0226)	(0.0026)
Cognitive Skill	-0.0049	0.5000
	(0.0347)	(0.0005)
Non-cognitive Skill	0.0743	-0.0019
	(0.0104)	(0.0013)
Whether College Graduate	0.0053	0.1099
	(0.0149)	(0.0018)
Constant	4.7693	5.7564
	(0.1652)	(0.0044)
Variance of shocks	0.4113	0.0568
	(0.0041)	(0.0267)

Table 10

Structural Model Estimation Parameters (Part II)

Migration costs		Psychic Value of Living in Rural	
Age	0.0839	Age	-11.5918
	(0.0022)		(0.0003)
Age^2	0.0020	Age^2	0.2152
	(0.0006)		(0.0000)
Year Trend	-0.3482	Cognitive Skill	29.8322
	(0.0028)		(0.0029)
Cognitive Skill	0.1633	Non-Cognitive Skill	0.3243
	(0.0013)		(0.0002)
Non-Cognitive Skill	-4.1509	Constant	2.3947
	(0.0139)		(0.0001)
Whether After 1982		Terminal Values	
Year Trend	-0.0567	Rural Work Experience	0.0639
	(0.0007)		(0.0000)
Constant	-0.2479	Urban Work Experience	5.4146
	(0.0017)		(0.0001)
Whether After 1990		Rural $Experience^2$	11.4459
Year Trend	-0.1141		(0.0000)
	(0.0015)	Urban Work $Experience^2$	0.3718
Constant	-0.0029		(0.0348)
	(0.0010)	Years of Education	0.4572
Constant	29.8249		(0.0001)
	(0.0663)	Cognitive Skill	-0.6711
			(0.0527)
		Non-Cognitive Skill	-0.0369
			(0.0451)
		Constant	-0.0016
			(0.0049)

Table 11

Structural Model Estimation Parameters (Part III)

Education Utility		Probability of Getting College Offer	
High School Level		Cognitive Skill	-7.1281
Cognitive Skill	0.1949		(0.0005)
	(0.0010)	Non cognitive Skill	2.4212
Noncognitive Skill	-0.1167		(0.0000)
	(0.0016)	Year Trend	-0.0094
Year Trend	-0.0079		(0.0000)
	(0.0008)	After 1999	
$1_{1965<year<1977}$	0.0820	Year Trend	0.0077
	(0.0004)		(0.0000)
Constant	-273.8736	Constant	-0.0053
	(0.1268)		(0.0000)
Some College Level		Constant	-4.4308
Cognitive Skill	0.5720		(0.0006)
	(0.0003)		
Noncognitive Skill	-0.3171	Psychic Cost of Taking College Entrance Exam	
	(0.0008)	Cognitive Skill	-0.0339
Year Trend	-2.1580		(0.0029)
	(0.0001)	Non cognitive Skill	20.9558
$1_{1965<year<1977}$	0.0023		(0.0534)
	(0.0001)	Constant	-426.2201
Constant	-86.3824		(9.6534)
	(0.0391)		
College Level			
Cognitive Skill	6.3682		
	(0.0016)		
Noncognitive Skill	-0.0018		
	(0.0001)		
Year Trend	-0.0080		
	(0.0002)		
$1_{1965<year<1977}$	-0.0005		
	(0.0002)		
Constant	283.5237		
	(0.0651)		

Table 12

Policy Effects Analysis

	Data	Benchmark	No Migration Policy	No College Expansion	No Policy
Average Years of Education	9.094	9.166	9.572	9.161	9.560
Fraction of Attending High School or Above	0.295	0.285	0.364	0.281	0.362
Fraction of Attending Some College	0.061	0.054	0.128	0.053	0.128
Fraction of Attending College	0.054	0.039	0.048	0.038	0.047
Fraction of Taking College Entrance Exam	0.103	0.095	0.123	0.094	0.120
Fraction of Ever Migration					
Cohort:1950-1960	0.254	0.243	0.41	0.243	0.41
Cohort:1961-1970	0.507	0.457	0.25	0.457	0.25
Cohort:1971-1980	0.710	0.703	0.41	0.703	0.41
Cohort: ¿1980	0.844	0.830	0.51	0.831	0.51
Mean of Log Earnings in 2007	6.53	6.60	6.76	6.60	6.76
Mean of Log Earnings for Never Migrants in 2007	6.42	6.39	6.56	6.39	6.56

Table 13

Log Earnings Regression

2007 Log Earnings Regression	Data	Model	Log Earning Regression for Never Migrants	Data	Model	Migrant's Log Earnings (with Full History)	Data	Model
Years of Education	0.0406	0.0283	Years of Education	0.0381	-0.0175	Years of Education	0.0191	-0.0335
	(0.0100)			(0.0159)			(0.0130)	
Age	0.1220	0.1604	Rural Experience	0.0730	0.1787	Rural Experience	0.0081	0.0663
	(0.0073)			(0.0089)			(0.0063)	
Age^2	-0.0017	-0.0024	Rural Experience^2	-0.0016	-0.0044	Rural Experience^2	-0.0001	-0.0020
	(0.0001)			(0.0002)			(0.0002)	
Migration	0.4081	0.3841	Cog 1	-0.0467	-0.0016	Urban Experience	0.0532	0.0869
	(0.1309)			(0.0320)			(0.0088)	
Years of Education×Migration	-0.0400	-0.0370	Cognitive 2	0.1397	0.0105	Urban Experience^2	-0.0013	-0.0014
	(0.0139)			(0.0831)			(0.0004)	
Cognitive 1	-0.0138	-0.0221	Non cognitive 1	0.0074	0.0032	Cognitive 1	-0.0127	0.0018
	(0.0178)			(0.0488)			(0.0236)	
Cognitive 2	0.0447	-0.0043	Non cognitive 2	-0.0002	-0.0001	Cognitive 2	0.0343	-0.0100
	(0.0479)			(0.0022)			(0.0673)	
Non cognitive 1	0.0122	0.0129	College	0.3217	0.4090	Non cognitive 1	0.0425	-0.0202
	(0.0273)			(0.1363)			(0.0339)	
Non cognitive 2	0.0007	-0.0010	Constant	5.5918	5.3902	Non cognitive 2	-0.0026	-0.0004
	(0.0013)			(0.2168)			(0.0020)	
College	0.1843	0.4410				College	0.2273	0.1491
	(0.0754)						(0.0995)	
Constant	4.1715	4.1087				Constant	6.0915	6.1087
	(0.1703)						(0.1539)	

Table 14

Migration and Education Moments

	Data	Model
Year of First Time Migration		
Cohort:1950-1960	1993.48	1993.75
Cohort:1961-1970	1996.05	1994.13
Cohort:1971-1980	1999.30	1994.07
Cohort: >1980	2004.29	2002.58
Fraction of Ever Migration		
Cohort:1950-1960	0.25	0.24
Cohort:1961-1970	0.51	0.46
Cohort:1971-1980	0.71	0.70
Cohort: >1980	0.84	0.83
Average Years of Education	9.09	9.17
Fraction of Attending High School or Above	0.29	0.28
Fraction of Attending Some College	0.06	0.05
Fraction of Attending College	0.05	0.04
Fraction of Taking College Entrance Exam	0.10	0.10

Table 15

Migration Regression

Return Migration Regression I	Data	Model
Years of Education	0.0027	0.0053
	(0.0059)	
Age	-0.0272	-0.0300
	(0.0064)	
Age^2	0.0005	0.0004
	(0.0001)	
Cognitive Skill 1	0.0136	0.0028
	(0.0135)	
Cognitive Skill 2	-0.0078	-0.0095
	(0.0357)	
Noncognitive Skill 1	0.0234	-0.0063
	(0.0195)	
Noncognitive Skill 2	-0.0017	0.0000
	(0.0011)	
Constant	0.4381	0.5387
	(0.1222)	

Return Migration Regression II	Data	Model
Years of Education	-0.0078	-0.0066
	(0.0031)	
Age	0.0055	-0.0145
	(0.0029)	
Age^2	-0.0001	0.0002
	(0.0000)	
Cognitive Skill 1	-0.0051	0.0008
	(0.0071)	
Cognitive Skill 2	0.0095	-0.0020
	(0.0188)	
Noncognitive Skill 1	0.0130	0.0007
	(0.0108)	
Noncognitive Skill 2	0.0004	-0.0001
	(0.0005)	
Location in 2007	-0.9887	-0.9456
	(0.0019)	
Constant	0.9494	1.2841
	(0.0635)	

Migration Regression I	Data	Model
Age	-0.0020	-0.0480
	(0.0061)	
Age^2	-0.0002	0.0005
	(0.0001)	
Years of Education	-0.0058	0.0520
	(0.0071)	
Cohort(1961-1970)	0.0567	-0.5727
	(0.0872)	
Cohort(1971-1980)	-0.0656	0.0562
	(0.0925)	
Cohort(>1980)	0.1353	0.5965
	(0.0990)	
Years of Education×Cohort(1961-1970)	-0.0106	0.0644
	(0.0093)	
Years of Education×Cohort(1971-1980)	0.0161	-0.0046
	(0.0091)	
Years of Education×Cohort(¿1980)	-0.0013	-0.0723
	(0.0088)	
Cognitive Skill 1	0.0159	-0.0010
	(0.0078)	
Cognitive Skill 2	0.0271	-0.0043
	(0.0209)	
Noncognitive Skill 1	0.0303	-0.0017
	(0.0120)	
Noncognitive Skill 2	-0.0009	0.0002
	(0.0006)	
Constant	0.5908	0.8358
	(0.1447)	

Migration Regression II	Data	Model
Years of Education	0.0007	-0.0069
	(0.0081)	
Cognitive Skill 1	-0.0027	0.0055
	(0.0155)	
Cognitive Skill 2	0.0721	0.0090
	(0.0674)	
Noncognitive Skill 1	-0.0066	-0.0296
	(0.0252)	
Noncognitive Skill 2	0.0009	0.0012
	(0.0010)	
Constant	0.0308	0.2161
	(0.0826)	

11 Cognitive and Non-cognitive Skills Estimates

Table 16

Unobserved Skill Distribution

Distribution 1		Distribution 2	
Mean (Cognitive)	-0.6950	Probability	0.4120
	(0.0460)		(0.0132)
Mean (Non-Cognitive)	-2.1589	Variance (Cognitive)	0.0620
	(0.0600)		(0.0332)
Variance (Cognitive)	0.0340	Variance (Non-cognitive)	0.1449
	(0.0009)		(0.1556)
Variance (Non-cognitive)	0.7726		
	(0.2046)		

Table 17

Cognitive Skill Measures (CFPS)

	Math 2010		Word 2010	
	Data	Model	Data	Model
Years of Education at the Test Time	0.1775	0.1710	0.1368	0.1354
	(0.0016)		(0.0020)	
Age at the Test Time	-0.0029	-0.0026	-0.0076	-0.0085
	(0.0006)		(0.0008)	
Constant	-1.0653	-1.0751	-0.5360	-0.5343
	(0.0325)		(0.0408)	
Residual^2	0.2346	0.3963	0.3694	0.3044
	Delayed Record		**Immediate Record**	
	Data	Model	Data	Model
Years of Education at the Test Time	0.0621	0.0583	0.0638	0.0657
	(0.0028)		(0.0027)	
Age at the Test Time	-0.0254	-0.0285	-0.0222	-0.0210
	(0.0010)		(0.0010)	
Constant	0.7275	0.7100	0.5783	0.5827
	(0.0578)		(0.0569)	
Residual^2	0.6964	0.6681	0.6734	0.6034
	Math 2014		**Word 2014**	
	Data	Model	Data	Model
Years of Education at the Test Time	0.1884	0.1769	0.1384	0.1286
	(0.0016)		(0.0021)	
Age at the Test Time	-0.0010	-0.0016	-0.0113	-0.0114
	(0.0006)		(0.0008)	
Constant	-1.2535	-1.2275	-0.3528	-0.3513
	(0.0316)		(0.0450)	
Residual^2	0.2220	0.1068	0.3960	0.3926

Table 18

Non-cognitive Skill Measures (CFPS 2010)

	Get on Well		Upset		Feel Difficult	
	Data	Model	Data	Model	Data	Model
Years of Education at the Test Time	0.0043	0.0309	0.0067	0.0101	0.0067	0.0012
	(0.0030)		(0.0031)		(0.0003)	
Age at the Test Time	0.0028	0.0164	0.0043	0.0001	0.0044	0.0039
	(0.0010)		(0.0001)		(0.0001)	
D=2	-0.1520	-0.1633	-0.2380	-0.2077	-0.2500	-0.7538
	(0.0080)		(0.0089)		(0.0082)	
D=3	-0.1543	-0.4645	-0.2367	-0.2230	-0.2399	-0.2413
	(0.0052)		(0.0090)		(0.0087)	
D=4	-0.1557	-0.3188	-0.2378	-0.2234	-0.2403	-0.2977
	(0.0050)		(0.0060)		(0.0060)	
D=5	-0.1578	-0.1849	-0.2427	-0.2539	-0.2487	-0.5963
	(0.0051)		(0.0057)		(0.0059)	
$Residual^2$	0.0062	0.0341	0.0091	0.0116	0.0094	0.0236

Table 19

Joint Measures: Amount of Cigarette (CFPS)

	Smoke 2010		Smoke 2012		Smoke 2014	
	Data	Model	Data	Model	Data	Model
Years of Education at the Test Time	0.4546	0.4382	-0.1206	-0.1288	-0.1083	-0.1137
	(0.1722)		(0.0283)		(0.0278)	
Age at the Test Time	-0.2115	-0.2168	0.0712	0.0734	0.0593	0.0485
	(0.0643)		(0.0106)		(0.0104)	
Whether Smoke	7.3199	3.3107	15.9805	2.0623	15.6562	4.3621
	(1.3610)		(0.2237)		(0.2200)	
Constant	22.3564	22.2506	-0.1406	-0.0491	0.2685	0.4974
	(3.5872)		(0.6081)		(0.6161)	
Residual^2	2685.3307	2685.5114	72.5551	73.1819	70.1309	70.5611

Table 20

Cognitive Skill Measures (CHIP)

	Whether Take College Entrance Exam		Class Performance	
	Data	Model	Data	Model
Years of Education at the Test Time	0.0386	0.0387	-0.0915	-0.0918
	(0.0017)		(0.0049)	
Age at the Test Time	-0.0006	-0.0006	-0.0105	-0.0105
	(0.0005)		(0.0015)	
Constant	-0.2074	-0.2061	3.8114	3.8396
	(0.0306)		(0.0863)	
$Residual^2$	0.0562	0.0564	0.4464	0.4470

Table 21

Non-cognitive Skill Measures CHIP

	Concentration 2007		Concentration 2008		Whether Confidence in yourself (2007-08)			
	Data	Model	Data	Model	Data	Model	Data	Model
Years of Education at the Test Time	0.0366	0.03337	0.0366	0.03337	0.0367	0.0393	0.0369	0.0371
	(0.0007)		(0.0007)		(0.0007)		(0.0007)	
Age at the Test Time	0.0150	0.0141	0.0150	0.0141	0.0147	0.0145	0.0147	0.0150
	(0.0001)		(0.0001)		(0.0001)		(0.0001)	
D=2	-0.9890	-0.9902	-0.9843	-0.9535	-0.9649	-0.9690	-0.9670	-0.9431
	(0.0048)		(0.0047)		(0.0042)		(0.0042)	
D=3	-0.9921	-0.9050	-0.9939	-0.9797	-0.9362	-0.9391	-0.9451	-0.9532
	(0.0085)		(0.0079)		(0.0165)		(0.0158)	
D=4	-0.9948	-0.9213	-1.0068	-0.9877	-0.9765	-0.8197	-1.0594	-0.9617
	(0.0164)		(0.0151)		(0.0397)		(0.0445)	
$Residual^2$	0.0158	0.0155	0.0157	0.0160	0.0155	0.0160	0.0155	0.0159

Table 22

Non-cognitive Skill Measures CHIP

	Whether Deal with Things Decisively (2007-2009)					
	Data	Model	Data	Model	Data	Model
Years of Education at the Test Time	0.0363	0.0363	0.0365	0.0325	0.0362	0.0367
	(0.0007)		(0.0007)		(0.0007)	
Age at the Test Time	0.0147	0.0149	0.0147	0.0089	0.0147	0.0149
	(0.0001)		(0.0001)		(0.0001)	
D=2	-0.9676	-0.6099	-0.9675	-0.9644	-0.9628	-0.9703
	(0.0042)		(0.0042)		(0.0042)	
D=3	-0.9208	-0.9196	-0.9472	-0.9364	-0.9457	-0.9556
	(0.0110)		(0.0105)		(0.0091)	
D=4	-0.9529	-0.6662	-0.9513	-1.0154	-0.9776	-0.9458
	(0.0282)		(0.0475)		(0.0289)	
$Residual^2$	0.0154	0.0123	0.0155	0.0158	0.0154	0.0158
	Whether Impossible to Overcome Difficulties (2007-2009)					
	Data	Model	Data	Model	Data	Model
Years of Education at the Test Time	0.0364	0.0420	0.0363	0.0364	0.0363	0.0364
	(0.0007)		(0.0007)		(0.0007)	
Age at the Test Time	0.0147	0.0134	0.0147	0.0151	0.0147	0.0142
	(0.0001)		(0.0001)		(0.0001)	
D=2	-0.9631	-0.9339	-0.9641	-0.9235	-0.9620	-0.9690
	(0.0041)		(0.0041)		(0.0042)	
D=3	-0.9495	-0.9226	-0.9381	-0.9315	-0.9527	-0.8497
	(0.0136)		(0.0135)		(0.0122)	
D=4	-0.9494	-0.8903	-0.9903	-0.9468	-1.0070	-0.8197
	(0.0379)		(0.0419)		(0.0349)	
$Residual^2$	0.0154	0.0160	0.0154	0.0157	0.0154	0.0160
	Whether Escape from the Difficulties (2007-2009)					
	Data	Model	Data	Model	Data	Model
Years of Education at the Test Time	0.0364	0.0371	0.0365	0.0288	0.0364	0.0393
	(0.0007)		(0.0007)		(0.0007)	
Age at the Test Time	0.0147	0.0150	0.0147	0.0153	0.0146	0.0120
	(0.0001)		(0.0001)		(0.0001)	
D=2	-0.9636	-0.9426	-0.9643	-0.9220	-0.9556	-0.9353
	(0.0042)		(0.0043)		(0.0042)	
D=3	-0.9469	-0.9628	-0.9564	-0.9333	-0.9411	-0.9225
	(0.0096)		(0.0084)		(0.0081)	
D=4	-0.8954	-0.9296	-0.8926	-0.8656	-0.9343	-0.8931
	(0.0324)		(0.0512)		(0.0334)	
$Residual^2$	0.0154	0.0156	0.0154	0.0150	0.0152	0.0156

Table 23

Non-cognitive Skill Measures CHIP and CFPS

	Concentration			Smoke	
	Data	Model		Data	Model
Years of Education at the Test Time	0.0276	0.0284	Years of Education at the Test Time	-0.1075	-0.1078
	(0.0004)			(0.0396)	
Age at the Test Time	0.0150	0.0154	Age at the Test Time	0.0641	0.0644
	(0.0063)			(0.0141)	
D=2	-0.9502	-0.9392	Whether Smoke	16.0555	16.0557
	(0.0035)			(0.2770)	
D=3	-0.9569	-0.9135	CFPS	2.9963	2.9842
	(0.0063)			(0.2798)	
D=4	-0.9677	-0.9065	Constant	-3.3554	-3.3541
	(0.0104)			(0.8510)	
CFPS	0.1036	0.1029	Residual^2	171.5688	171.5581
	(0.0032)				
Residual^2	0.0251	0.0257			

Appendix

Table A1: Return to Education

(Rural Females)

	2002 Data	2007 Data			
Age	0.115***	0.077***	0.077***	0.077***	0.077***
	(0.008)	(0.004)	(0.005)	(0.004)	(0.005)
Age2	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Years of Education	0.037***	0.052***	0.048***	0.046***	0.043***
	(0.005)	(0.003)	(0.003)	(0.003)	(0.003)
Migration	0.180***	0.433***	0.414***	0.455***	0.436***
	(0.072)	(0.070)	(0.070)	(0.070)	(0.070)
Years of Education Migration	0.008	-0.020***	-0.018**	-0.023**	-0.020***
	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)
Constant	3.127***	4.364***	4.831***	4.514***	4.929***
	(0.140)	(0.090)	(0.100)	(0.097)	(0.105)
Non-Cognitive Skill Measures	No	No	Yes	No	Yes
Cognitive Skill Measures	No	No	No	Yes	Yes
Observations	6451	7,576	7,576	7,576	7,576

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Rural migrants are the individuals working in the urban area, which is out of the county of their rural hukou location.

Table A2: The Effect of Relaxing Migration Restrictions on Education
(Females)

Explained Variable: Years of Education				
> 1982	3.267***	3.194***	2.161***	2.134***
	(0.064)	(0.065)	(0.060)	(0.061)
Rural	-2.849***	-2.851***	-2.358***	-2.373***
	(0.058)	(0.059)	(0.051)	(0.052)
> 1982× Rural	-1.226***	-1.270***	-0.409***	-0.448***
	(0.079)	(0.079)	(0.072)	(0.073)
Constant	9.764***	10.788***	16.798***	17.242***
	(0.048)	(0.104)	(0.120)	(0.139)
Cognitive Skill Measures	No	No	Yes	Yes
Non-Cognitive Skill Measures	No	Yes	No	Yes
Observations	13,699	13,699	13,013	13,013

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

> 1982 is an indicator for whether the individual finishes his schooling after the migration reform 1982.

Table A3: The Effect of Relaxing Migration Restrictions on Education
(All)

Explained Variable: Years of Education				
> 1982	3.215***	3.163***	2.062***	2.037***
	(0.047)	(0.047)	(0.044)	(0.044)
Rural	-2.684***	-2.678***	-2.277***	-2.285***
	(0.042)	(0.042)	(0.036)	(0.036)
> 1982×Rural	-1.363***	-1.379***	-0.450***	-0.478***
	(0.057)	(0.057)	(0.051)	(0.052)
Constant	10.025***	11.190***	17.048***	17.622***
	(0.035)	(0.072)	(0.079)	(0.093)
Cognitive Skill Measures	No	No	Yes	Yes
Non-Cognitive Skill Measures	No	Yes	No	Yes
Observations	29,628	29,628	28,189	28,189

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

> 1982 is an indicator for whether the individual finishes his schooling after the migration reform 1982.

Table A4: Migration and Education Decisions: Pretrend Analysis
(Females)

Explained Variable: Years of Education				
Year 1977× Rural	-0.196	-0.254	-0.341	-0.380
	(0.301)	(0.300)	(0.282)	(0.282)
Year 1978× Rural	0.311	0.306	0.210	0.194
	(0.299)	(0.298)	(0.280)	(0.280)
Year 1979× Rural	0.110	0.114	-0.010	-0.002
	(0.306)	(0.305)	(0.287)	(0.286)
Year 1980× Rural	0.309	0.281	0.412	0.392
	(0.297)	(0.295)	(0.278)	(0.278)
Year 1981× Rural	-0.095	-0.101	-0.109	-0.112
	(0.304)	(0.303)	(0.285)	(0.284)
Year 1977	-0.148	-0.094	-0.035	0.003
	(0.248)	(0.247)	(0.231)	(0.231)
Year 1978	-0.408	-0.406	-0.260	-0.248
	(0.245)	(0.244)	(0.229)	(0.228)
Year 1979	-0.437	-0.428	-0.347	-0.344
	(0.252)	(0.251)	(0.235)	(0.235)
Year 1980	-0.340	-0.307	-0.407	-0.383
	(0.243)	(0.242)	(0.228)	(0.228)
Year 1981	0.188	0.183	0.165	0.162
	(0.248)	(0.246)	(0.231)	(0.231)
Rural	-2.833***	-2.800***	-2.262***	-2.266***
	(0.219)	(0.219)	(0.209)	(0.209)
Constant	10.560***	11.517***	15.224***	15.698***
	(0.180)	(0.256)	(0.300)	(0.334)
Cognitive Skill Measures	No	No	Yes	Yes
Non-Cognitive Skill Measures	No	Yes	No	Yes
Observations	2,337	2,337	2,319	2,319

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Year is the calender year when individuals finish their education.

Rural is an indicator variable for individuals with rural Hukou.

Table A5: Migration and Education Decisions: Pre-trend Analysis
(All)

Explained Variable: Years of Education				
Year 1977× Rural	0.227	0.174	-0.039	-0.065
	(0.222)	(0.220)	(0.202)	(0.202)
Year 1978× Rural	0.327	0.332	0.304	0.307
	(0.222)	(0.221)	(0.202)	(0.202)
Year 1979× Rural	0.242	0.252	0.077	0.087
	(0.227)	(0.225)	(0.206)	(0.205)
Year 1980× Rural	0.429	0.398	0.306	0.290
	(0.221)	(0.220)	(0.202)	(0.201)
Year 1981× Rural	0.097	0.088	0.089	0.083
	(0.227)	(0.225)	(0.206)	(0.205)
Year 1977	-0.564**	-0.517**	-0.342*	-0.313
	(0.183)	(0.182)	(0.167)	(0.166)
Year 1978	-0.364*	-0.370*	-0.369*	-0.368*
	(0.183)	(0.182)	(0.166)	(0.166)
Year 1979	-0.399*	-0.405*	-0.321	-0.321
	(0.188)	(0.186)	(0.171)	(0.170)
Year 1980	-0.415*	-0.380*	-0.322	-0.299
	(0.182)	(0.181)	(0.166)	(0.166)
Year 1981	0.122	0.132	0.069	0.076
	(0.188)	(0.186)	(0.170)	(0.170)
Rural	-2.937***	-2.930***	-2.277***	-2.294***
	(0.164)	(0.163)	(0.151)	(0.150)
Constant	10.941***	12.069***	16.015***	16.582***
	(0.135)	(0.186)	(0.201)	(0.225)
Cognitive Skill Measures	No	No	Yes	Yes
Non-Cognitive Skill Measures	No	Yes	No	Yes
Observations	4,704	4,704	4,667	4,667

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Year is the calender year when individuals finish their education.

Rural is an indicator variable for individuals with rural Hukou.