

The Positive Effect of Labor Mobility Restrictions on Human Capital Accumulation in China

Yao Pan*

Job Market Paper

November 8, 2012

Abstract: The Hukou system restricted most rural-urban migration in China for over 50 years. Under this system, rural residents could permanently migrate to urban areas by acquiring higher education. In this paper, I test the hypothesis that these mobility restrictions, combined with selective migration policies, encouraged rural investment in education. The test is based on an extension of the regression discontinuity design that allows it to work when individual treatment status is missing but sufficient aggregate data are available. Findings suggest that human capital accumulation for rural residents decreased sharply when mobility restrictions were removed in 1998. Nonparametric estimates show a statistically significant drop in high school attendance rate, ranging between 8.7 to 30.5 percentage points, depending on specifications and assumptions. There are also signs of a decrease in the middle school graduation rate and an increase in the dropout rate of high school and above. These effects are bigger for males and for those able to permanently migrate to relatively rich areas.

Keywords: Migration; Education; Brain Drain; Regression Discontinuity; China
JEL Classification: O15; J61; J24; R23; J22

*Department of Economics, The George Washington University (panyao@gwu.edu). I would like to thank Bryan Boulier, Paul Carrillo, Barry Chiswick, Shahe Emran, Vincy Fon, James Foster, Remi Jedwab, Sumit Joshi, Donald Parsons, Robert Phillips, Stephen Smith, Benjamin Williams, Yaohui Zhao and participants at The George Washington University Development Lunch Seminar, The 9th Midwest International Development Conference, The Chinese Economist Society 2012 Annual Conference and The 2012 Northeast Universities Development Consortium Conference for helpful comments and suggestions. I acknowledge financial support from The Institute for International Economic Policy and The George Washington University. All mistakes are my own.

1. INTRODUCTION

Mobility restrictions are said to affect societies negatively (Hamilton & Whalley, 1984). Since the household registration system in China (the Hukou system) tightly restricts rural-urban migrations, it has attracted special attention from scholars. Fujita et al. (2004) and Au & Henderson (2006a,b) argue that most Chinese cities are significantly undersized and ascribe this under-urbanization to the Hukou system. There are also large income losses resulting from insufficient agglomeration in rural industries (Au & Henderson, 2006b). In addition, the Hukou system has been criticized as a major cause of rising rural-urban income inequality in China and as a generator of unfair opportunities (Yang, 1999; Liu, 2005; Whalley & Zhang, 2007; Wu & Treiman, 2004; Chan, 2009).

This paper examines an additional effect of the Hukou system that has not been explored in the literature: its effect on human capital accumulation in rural China. In particular, the Hukou system may play a significant role in fostering education in rural areas given selective migration policies.

The Hukou system categorizes people as rural or urban at birth according to their parents' status. Urban residents enjoy benefits provided by the government while rural individuals do not. Even though rural people are allowed to temporarily migrate to urban areas to seek better employment opportunities, migrants without local urban Hukou cannot enjoy benefits such as medicare, unemployment insurance, housing subsidies, pensions, etc.

However, rural residents enrolled in technical high school or college are automatically granted urban Hukou. Thus, individuals of rural origin have additional gains from higher education compared to their urban counterparts, giving rural residents a greater incentive to invest in human capital. The private returns to education in rural areas include the option value of increased probability of being able to receive urban benefits. I define these extra gains as "institutional returns" since they stem from government constraints on mobility.

Empirically, I examine institutional returns to education using a 1998 Hukou policy change. Before 1998, individuals inherited their mother's Hukou status. After 1998, it became possible to inherit one's father's status. Children under 18 years old subsequently had the chance to transfer their Hukou to their father's status, which differentially benefited individuals with a rural mother and an urban father. These individuals could obtain urban Hukou and its associated benefits without higher education. I apply a regression discontinuity (RD) approach in this study to estimate the changes in the high school attendance rate for this group when institutional returns to education were removed. If individuals could perfectly predict the 1998 Hukou policy change at the time of their school decisions, all eligible individuals under the birth month rule are able to adjust their education choices accordingly. My scenario fits into a sharp design. On the other hand, if this Hukou policy change was fully unexpected, then it only affected the high school attendance for the group that had not made the high school decision by the time of the Hukou reform. Unfortunately, the data does not provide information about the date when individuals finished middle school. I provide an extension of the traditional RD estimation that allows it to work when only the aggregate proportion treated is available. The local average treatment effect can be obtained by rescaling the first stage sharp RD estimator using the discontinuity in the treated proportion.

Nonparametric estimates show a statistically significant drop of 8.7 percentage points in the

probability of attending high school for those individuals satisfying the birth month criterion of the 1998 Hukou policy reform. This result is based on the assumption that people can perfectly predict this policy change at the time of their high school decision, and it can be viewed as a lower bound of the true effect. On the other hand, if the policy change is fully unexpected, then it only affected people's high school decisions if they were made after the policy change. In this case, the drop in high school attendance rate increases to 30.5 percentage points after rescaling. In both cases, the estimated effects are bigger for males and for those able to get urban Hukou in relatively rich areas. The probability of attending high school decreases by 11.3 to 31.9 percentage points for males and 12.4 to 45.6 percentage points for those getting urban Hukou in rich regions. These effects are not significant for females and for those getting urban Hukou in relatively poor areas. In addition, there are also signs of a negative discontinuity in the middle school graduation rate and a positive discontinuity in the dropout rate of high school and above. The findings above are confirmed by a series of robustness checks such as continuity assumptions required for a valid regression discontinuity, parametric estimations and possible age induced biases. These findings demonstrate the existence of substantial institutional returns to education for people with rural Hukou in China.

This research contributes to a main strand in the migration literature: brain drain versus brain gain. Contrary to the conventional view of detrimental brain drain, some theoretical studies in the late 1990s point out the possibility of "brain gain" when there are endogenous education choices and uncertain migration prospects (Mountford, 1997; Stark, Helmenstein & Prskawetz, 1997, 1998; Vidal, 1998). Under certain conditions, the future chance of emigrating to developed countries raises the expected returns to schooling, resulting in a higher human capital stock at home (minus those who emigrated). Most empirical evidence of this issue is derived from cross country regressions (Beine, Docquier & Rapoport, 2001, 2008, 2010). A few micro-level studies confirming the "brain gain" effect have emerged in recent years. In order to claim causal effect, researchers often adopt a difference-in-difference strategy (Chand & Clemens, 2008) or IV estimation (Batista, Lacuesta & Vicente, 2011). The validity of these methods requires a set of untestable exclusion restrictions. This study overcomes these limitations by identifying the causal effect using quasi-experimental RD analysis which requires relatively mild continuity assumptions.

Previous literature analyzes human capital formation in a situation when only educated workers have positive probability of emigration, and compares it with the benchmark of a closed economy with no emigration. The results of the literature can be interpreted as the impact of selective migration on the source country's human capital investment. This paper complements previous studies by using free migration as the benchmark. I investigate human capital dynamics in a situation with migration restrictions and selective migration, and then I compare it with the case when all workers are free to move regardless of education levels.

This study is also closely related to the literature on the private returns to education and individuals' schooling choices. In particular, this paper analyzes the institutional returns to education, which are defined as educational benefits stemming from overcoming geographical¹ and/or occupational² labor mobility restrictions. These returns include both market

¹The returns to education under geographical migration restrictions are well documented in the brain gain literature (Mountford, 1997; Stark, Helmenstein & Prskawetz, 1997, 1998; Vidal, 1998).

²See Sicherman & Galor (1990) for a theoretical and empirical discussion of returns to education in the form of increased probability of occupational upgrading. Such returns will be amplified by the occupational mobility restrictions associated with the Hukou system in China.

returns, such as higher wages, widened job choices and increased access to social welfare programs, and nonmarket returns, such as improved health, enhanced political representation, better marriage market outcomes and benefits for one's offspring. Even though the institutional returns in China are hard to quantify, the paper supports the existence of such returns. The magnitude of institutional returns constitutes a significant portion of rural returns to education in China given that a removal of these returns results in a significantly lower investment in human capital.

This study has policy implications for contemporary China. In order to mitigate the unbalanced development between rural and urban areas, a few provinces have used uniform identity Hukou to replace the original rural/urban dichotomy since 2003. The elimination of rural/urban Hukou status erases the institutional returns to education for the previous rural Hukou holders, which may in turn negatively affect their investment in human capital.

The rest of this paper is structured as follows: Section 2 provides the background of the Hukou system, reviews related literature, and describes the main data source used in this study. Section 3 presents a simple human capital accumulation model. The identification strategy and local linear regression results for regression discontinuity approach are provided in Section 4 and Section 5, respectively. In Section 6, I conduct a series of robustness checks including the continuity assumption required for valid RD and other potential problems. Section 7 briefly discusses welfare effects, and section 8 concludes.

2. BACKGROUND

This section briefly describes the development of the Hukou system in China and summarizes a few attempts in the literature to relate the Hukou system to educational attainment. A description of the main data set used in this study is provided at the end of this section.

2.1 Household registration system³

China's household registration system (the Hukou system) is one of the strictest population regulation mechanisms in the world. Unlike most of the unsuccessful attempts made by other countries to regulate migration, the Hukou system, combined with China's food rationing policy, effectively tied people to their registered residency place.

At the early stage of development in the 1950s, the Hukou system was used for civil records. People were allowed to migrate freely until 1958, when the system was formally incorporated into law. The two most important pieces of information of Hukou record are Hukou status (urban/rural) and legal residence address, which were inherited from one's mother until 1998, when inheritance from one's father was permitted. This information is registered at birth for every legal Chinese citizen, follows a person for his/her lifetime, and is extremely hard to change.

Rural people relied on their land to support themselves while the government provided housing, food, pension, etc. to urban residents. It was almost impossible to migrate without legally changed Hukou because of rigid food rationing and the absence of commodity

³Please refer to Wu & Treiman (2004) and Chan (2009) for a detailed description of the Hukou system.

markets. The ways of changing one's Hukou status from rural to urban were mainly through a job assignment after military service, enrollment in technical high school or tertiary education, or employment through the states, all of which were subject to small quotas and were exceedingly difficult for peasants to achieve.

The profound economic reform launched in 1978 altered the pattern of the Hukou system in a significant way. People temporarily migrating to urban areas could apply for a temporary resident permit, which granted them legal residency for a few months and was subject to renewal. Abolishment of government subsidized food rations during the late 1980s, along with newly evolved commodity markets, enabled abundant rural labor to seek employment opportunity in cities, boosting the temporary migrants stock to 147.35 million in 2005.⁴ However, the loosened migration restriction only guaranteed controlled and limited mobility for rural labor. These temporary migrants normally took up low-pay jobs with poor working conditions, and they were still denied access to urban social welfare benefits (Wang & Zuo, 1999). This situation persists today.

As mentioned above, higher education is one of the few ways of obtaining urban Hukou. After finishing nine years of compulsory primary and middle school education, students can either attend high school or work directly. There are two types of high school: regular and technical. Even though obtaining regular high school degree does not guarantee urban Hukou, it provides the opportunity for tertiary education. Newly admitted students to technical high school, junior college and above are automatically granted urban Hukou.^{5,6} For rural students, the returns to high school education include not only the higher future income, which is well-studied in the literature, but also potential dramatic social welfare benefits associated with Hukou status change. The link between schooling and labor mobility creates stronger incentives to pursue high school education for rural residents than their urban counterparts in China.

2.2 Hukou and education

There have been a few scholars analyzing the relation between Hukou status and educational attainment since the 1990s. Wu (2010) finds that rural people are significantly less likely to enroll in high school after the nine years of compulsory education. Wu & Treiman (2004) show empirically that rural family background harms future educational attainment owing to inferior quality of previous education. They also show that higher education⁷ significantly increases the odds of obtaining urban Hukou by more than four times for those of rural origin. Even though these studies document that rural origin is one of the major determinants of less schooling, and that schooling itself, especially college degrees, will help transfer Hukou from rural to urban areas, none treats education as a choice variable.

Zhao (1997) is the first to incorporate the schooling choice into the calculation of expected future income. She demonstrates that the incentive for pursuing higher school education

⁴Communique on Major Data of 1% National Population Sample Survey in 2005, National Bureau of Statistics.

⁵Admission to these schools are based on competitive exams, which are equally available to all Chinese citizens. However, rural children are less likely to attend these schools due to the inferior quality of primary education in rural areas.

⁶Even though the Hukou transfer is voluntary, most students accept urban Hukou given the enormous benefits associated with it. Denial of urban Hukou is rare, especially in the last century when Hukou played an even more significant role than it does today.

⁷They define higher education as technical high school or tertiary education.

is partly rooted in the chance of changing Hukou status. Using data from three villages near the city of Beijing, she finds a decline in the high school enrollment rate in 1980s and interprets this change as a result of increased opportunity of non-farm employment in rural enterprises. She points out that conventional methods underestimate the returns to education in China by ignoring the possible wage and non-wage gains associated with transferring Hukou status. Even though her article adjusts for non-wage benefits such as food ration coupons and subsidized housing, brought about by urban status, it does not include other benefits, such as medicare, pensions and benefits that are inter-generational transferable, due to data limitations. Thus, her data still underestimate the real returns to schooling for rural people since gains from Hukou status change are not fully taken in to account.

Using a broader survey covering four provinces, De Brauw & Giles (2008) show that increased labor mobility from allowing temporary migration significantly reduces high school enrollment in rural China. The effect of increased temporary migration is identified using the exogenous timing of issuing the national ID cards, which are necessary for rural migrants to register as temporary residents in urban areas. However, De Brauw & Giles (2008) only analyse the impact on high school enrollment of a relaxation of the Hukou system, which allows rural residents to temporarily work in urban areas without any urban benefits. They do not investigate the effect of obtaining urban Hukou on schooling, which allows peasants to permanently migrate to urban areas and enjoy associated local benefits.

Even though scholars have realized the important role that labor mobility restrictions played on returns to education, there are no published studies that tackle the *overall* effect of Hukou on educational attainment in China. Moreover, most of the studies that take into account the endogenous education decision only use data sets covering a few villages or provinces. The restricted scope of the data limits the generalization of the results in these studies. Relying on the 1998 Hukou policy reform, this paper is the first to directly address the effect of the Hukou system on educational attainment for rural people using a nationally representative dataset.

In China, newborn children had to inherit their mother's Hukou status (rural/urban) until September 1998⁸, when inheriting father's Hukou status was permitted by a Hukou policy reform. Children under 18 by September 1998 who followed their mother's Hukou then had a chance to change it according to their fathers'. Therefore, those born in or after September 1980 with their mother holding rural Hukou and father holding urban Hukou are beneficiaries of this reform. In the next section, I show how this policy change affects the returns to investment in education in a simple model.

The main data source used in this study is the 0.95 per thousand sample of China's fifth wave population census conducted in 2000. It contains individual level demographic information as of November 1, 2000, such as month of birth, gender, ethnic minority status, education level, employment status and occupation. The Hukou status of rural/urban and Hukou location at province level are reported as well.

3. THEORETICAL FRAMEWORK

⁸The Hukou reform was initially proposed by the Ministry of Public Security on June 23, 1998. Even though approved by State Council on July 22, 1998, it was not put into practice until the beginning of September.

In this section, I present a simple human capital investment model to frame optimal schooling decisions for rural people in China. It allows me to illustrate the effect of transferring Hukou status from rural to urban on schooling in a dynamic setting by analyzing changes in returns to education.⁹ I proceed by first introducing the accumulation of human capital and the sources of household income, and then analyzing the household's utility maximization problem. Finally I discuss the changes of returns to education for rural people once the government grants them urban Hukou and the subsequent effect on their optimal schooling choices.

Following De Brauw & Giles (2008), I assume each household only has one child.¹⁰ In each period t , households make an investment decision between human capital H_t and physical capital K_t . H_t is the sum of fixed adults' human capital H^a and the child's human capital H_t^c . Households' human capital is accumulated by sending the child to school for the share e_t of his/her time and paying a cost proportional to school time, $P_t^e e_t$, where P_t^e is the unit price of schooling. Therefore, human capital evolves as follows:

$$H_{t+1} = H_t + \psi_t G(e_t), \quad (1)$$

where G is a human capital production function increasing in e_t and ψ_t is a positive time-varying productivity parameter that captures differences in ability, motivation, effort level, school quality, etc.

Households generate income from home production or labor employment in the labor market. Home production is based on physical capital and labor according to $Y_t^h = \theta_t F(K_t, L_t^{a1}, L_t^{c1})$, where L_t^{a1} and L_t^{c1} are the amount of time used in household production by adults and the child, respectively, and θ_t is an exogenous productivity parameter. Income from the labor market can be expressed as $Y_t^w = w(H^a)L_t^{a2} + w(H_t^c, S)L_t^{c2}$, where L_t^{a2} and L_t^{c2} are shares of time working in the labor market for adults and children. $w(H^a)$ is the adults' wage function increasing in human capital stock. The wage function for children, $w(H_t^c, S)$, is increasing in their human capital, H_t^c , and increasing their Hukou status, S , which equals one if a child gets urban Hukou and zero otherwise. The default status for rural children is $s = 0$. Households invest all the left over income into physical capital after deducting consumption c_t and school costs from total earnings. Thus, physical capital accumulates according to

$$K_{t+1} = K_t + \theta_t F(K_t, L_t^{a1}, L_t^{c1}) + w(H^a)L_t^{a2} + w(H_t^c, S)L_t^{c2} - c_t - P_t^e e_t. \quad (2)$$

Assume there are no credit markets and that households face borrowing constraints

$$K_{t+1} \geq 0, \forall t. \quad (3)$$

Children are eligible for school from $t = 0$ to $t = T - 1$ with T a fixed number of periods. They work exclusively once $t = T$ and are not allowed to go back to school. Households' utility from period T and beyond can be written as $\Phi(K_T, H_T) + q(H_T^c)V(H_T^c)$, where $\Phi(K_T, H_T)$ is the terminal value function if children do not change Hukou status to urban.

⁹The model presented here is based on the general human capital investment model developed by Glewwe & Jacoby (2004). I incorporate the institutional returns to education associated with Hukou status change and allow for market employment in addition to home production.

¹⁰This assumption is likely to be valid, particularly for the sample I use for estimation. Families with a rural mother and an urban father are only allowed to have one child according to the one child policy in China. This simplifying assumption will not affect critical predictions of the model. Additional control for number of siblings is included in empirical regressions.

It includes both pecuniary and non-pecuniary benefits from children's education for the entire household. $V(H_T^c)$, the institutional returns to education, is a positive terminal value function representing utility gained from transferring Hukou. It consists of the increased wage and better employment opportunities as well as a series of other urban benefits. I assume the benefits premium of an urban Hukou increases with children's terminal education levels, e.g. $\frac{\partial V_T}{\partial H_T} > 0$.¹¹ The probability of obtaining urban identity, $q(H_T^c)$, is increasing in final human capital stock of children H_T^c . I assume the transfer from rural to urban Hukou via higher education only takes place at the beginning of period T . Thus, the probability q does not enter children's wage function, $w(H_t^c, S)$, during schooling periods. Households' current utility is a function of consumption c_t , leisure of both adults and children, l_t^a and l_t^c , and children's school enrollment e_t . Assume time endowments for both adults and children is normalized to one in each period, then $l_t^a = 1 - L_t^{a1} - L_t^{a2}$ and $l_t^c = 1 - L_t^{c1} - L_t^{c2} - e_t$. Parents' objective is to maximize expected household lifetime utility:

$$E_0 \left[\sum_{t=0}^{T-1} \delta^t U(c_t, l_t^a, l_t^c, e_t) + \Phi(K_T, H_T) + q(H_T^c)V(H_T^c) \right], \quad (4)$$

subject to constraints (1) (2) and (3), where δ is the discount factor. At time 0, households are uncertain about future values of $\psi_t, \theta_t, w(H_t^a), w(H_t^c, S), P_t^e, \Phi_T, q(H_T^c)$ and $V(H_T^c)$.

The first-order conditions for an interior solution to this maximization problem are:

$$U_c(t) = \lambda_t \quad (5)$$

$$U_{l^a}(t) = \lambda_t \left(\theta_t F_{L^{a1}}(t) + w(H^a) \right) \quad (6)$$

$$U_{l^c}(t) = \lambda_t \left(\theta_t F_{L^{c1}}(t) + w(H_t^c, S) \right) \quad (7)$$

$$U_e(t) + \mu_t \psi_t G_e(t) = \lambda_t \left(\theta_t F_{L^{c1}}(t) + w(H_t^c, S) - P_t^e \right), \quad (8)$$

where λ_t and μ_t are shadow prices of physical capital and human capital at period t , respectively. The school demand function can be derived from these FOCs as:

$$e_t^* = e^* \left(\lambda_t, \mu_t, \psi_t, \theta_t F_{L^{c1}}(t), \theta_t F_{L^{a1}}(t), w(H_t^c, S), w(H^a), P_t^e \right). \quad (9)$$

Since utilities are additively separable across time, past and future decisions can only influence current decisions through shadow prices λ_t and μ_t . In addition, the borrowing constraint (3) only affects intertemporal decisions conditional on λ_t , but not intratemporal decisions since the coefficient of borrowing constraint v_t does not appear in equation (5) (6) (7) or (8) but only in the following intertemporal Euler equation for the physical capital price:

$$\lambda_t = \delta E_t(\lambda_{t+1} + v_{t+1})(1 + \theta_{t+1} F_k(t+1)), \quad (10)$$

where v_t is the multiplier on the borrowing constraint.

Now we can trace out the impact of obtaining urban Hukou on school demand for rural children, which consists of three effects. First, the shadow price of human capital, μ_t , changes. To show this, note that the terminal condition before transferring Hukou requires that the marginal value of human capital after the schooling period, $\mu_T = \frac{\partial(\Phi(K_T, H_T) + q(H_T^c)V(H_T^c))}{\partial H_T} =$

¹¹Liu (2005) finds the wage returns to education for urban Hukou holders is higher than for rural Hukou holders. I treat his results as a sign of increasing Hukou benefits associated with higher human capital stock.

$\frac{\partial \Phi_T}{\partial H_T} + \frac{\partial q_T}{\partial H_T} V(H_T^c) + \frac{\partial V_T}{\partial H_T} q(H_T^c)$. This marginal benefit from education investment includes three parts: the increased households' utility without Hukou transfer, the higher probability of getting urban Hukou and related benefits, and a higher wage premium brought by higher education conditional on Hukou transfer prospects. Once gaining urban Hukou, this marginal value function simplifies to $\mu'_T = \frac{\partial(\Phi(K_T, H_T) + V(H_T^c))}{\partial H_T} = \frac{\partial \Phi_T}{\partial H_T} + \frac{\partial V_T}{\partial H_T}$. The change in the marginal value is $\mu'_t - \mu_t = (-\frac{\partial q_T}{\partial H_T} V(H_T^c)) + \frac{\partial V_T}{\partial H_T} (1 - q(H_T^c))$. The first term is negative, showing that directly granting individuals urban Hukou eliminates their incentive to go for higher education with the aim of possible urban benefits. The second term is positive, showing an increased probability of reaping the urban-status wage premium. The marginal returns to education *after* schooling periods cannot be signed. Since the intertemporal Euler equation for human capital shows that $\mu_t = \delta^{T-t} E_t \mu_T$, the expected returns to education while *in* the schooling period is ambiguous after obtaining urban Hukou and the school enrollment is uncertain.

The second effect operates through the shadow price of children's time. Obtaining an urban Hukou increases children's wage at any education level from $w(H_T^c, 0)$ to $w(H_T^c, 1)$. The increase in opportunity costs will lower investment in human capital based on equation (8).

Thirdly, the school cost, P_t^e , may change for some rural children as well. Many children of migrant workers live with their parents in urban areas. Without local Hukou, families of these children have to pay additional fees for the enrollment in urban schools. These extra fees can be waived once the child obtains local urban Hukou. The reduced school cost, P_t^e , reduces the school costs and is expected to result in higher school enrollment.¹² The overall effect of obtaining urban Hukou on schooling is the combination of all three effects. The sign of the net impact is ambiguous and is left for empirical study.

4. EMPIRICAL STRATEGY

The 1998 Hukou policy reform granted urban Hukou to a specific group of children, allowing me to identify the effect of the Hukou system on education. According to this reform, children born in or after September 1980 who inherited their mother's Hukou status then had a chance to change it to their father's Hukou, which differentially benefited individuals with their mother holding rural Hukou and their father holding urban. Thus, the main focus of this study is to examine whether schooling decisions are different between cohorts born before and after September 1980 for individuals with a rural mother and an urban father. A natural approach to estimating this is regression discontinuity (RD) design.

Assume for the moment that individuals can perfectly predict the 1998 Hukou policy change at the time of their school decisions. Under this assumption, all eligible individuals under the birth month rule are able to adjust their education choices accordingly. Let the assignment index z_i represent birth month of individual i . z_i has a cutoff value of $z_0 = \text{Sept.1980}$ and has been normalized to $z_0 = 0$.¹³ For individuals with an urban father and a rural mother, the eligibility of transferring Hukou status to urban is a deterministic function of

¹²This is especially true for my sample. Children with a rural mother and an urban father live in urban areas because women normally join their husband's family after marriage in Chinese culture.

¹³ z_i now is the difference between original z_i and z_0 , with a negative sign indicating "before". For example, $z_i = 2$ for individuals born in November 1980 and $z_i = -3$ for individuals born in June 1980.

individuals' birth month according to the 1998 Hukou reform,

$$x_i = I_{1i}(z_i \geq z_0), \quad (11)$$

where I_1 is an indicator function taking value of 1 if $z_i \geq z_0 = 0$. The relationship between an individual's eligibility, x_i , and his/her education outcome y_i is described in the following equation:

$$y_i = \alpha_i + x_i \times \beta_i + \epsilon_i, \quad (12)$$

where α_i is the potential education outcome without the policy change; β_i is the impact of Hukou transfer eligibility on the schooling decision, and ϵ_i is a random error term. The problem fits in sharp RD design with individuals' treatment status defined as their eligibility to get urban Hukou. I focus on this intent-to-treat (ITT) effect because the data are not informative about whether an individual utilizes this opportunity to transfer Hukou status or not.

One major advantage of RD is that the identification of the treatment effect does not require the error term ϵ_i to be uncorrelated with z_i and x_i . The causal treatment effect can be nonparametrically identified by assuming smoothness in potential education outcomes, α_i , around the birth month threshold of September 1980 (Hahn, Todd & Van der Klaauw, 2001; Porter, 2003). Because identification can be achieved under relatively mild continuity assumptions, RD provides more credible results compared to other conventional non-experimental strategies that requires exclusion restrictions such as difference-in-difference and matching.

I adopt a nonparametric RD approach for estimation based on Hahn, Todd & Van der Klaauw (2001) and Porter (2003). They suggest estimating the left and right limits of the discontinuity using local polynomial regression, which overcomes the boundary problem of kernel regression, and then taking the difference to estimate the treatment effect. To achieve better performance in boundary estimation, I use a triangular kernel with kernel weights $K(\frac{z_i - z_0}{h})$ in all regressions.

The perfect prediction assumption may not apply in the context of this study. Now assume the opposite (i.e., that no information is revealed before the announcement of the policy). For illustration purposes, I describe the estimation method using high school attendance outcomes. Under this new assumption, the 1998 policy change only affects the high school attendance decision of individuals (1) born in or after September 1980 and (2) finishing middle school in or after 1998.^{14,15} The treatment determination criterion is depicted in the following equation:

$$x_i = I_{1i}(z_i \geq z_0) \times I_{2i}(d_i \geq d_0),$$

where I_2 is an indicator function taking value of 1 if the year of finishing middle school, d_i , is greater than or equal to 1998 (d_0) and x_i is the unobserved treatment status taking the value of 1 if individual i is eligible to transfer to urban Hukou status and able to adjust his/her high school choice according to the new Hukou policy announced in 1998, and it is 0 otherwise. Thus, the jump of treatment probability is less than one at the cutoff of September 1980. Fuzzy design seems to fit this situation. However, because the timing of

¹⁴The Hukou policy reform was announced in July 1998, which took place after the high school entrance exam in June and before the high school state date in September.

¹⁵People finishing middle school earlier had already made their high school decision when Hukou reform was announced. Even though they may be eligible to transfer Hukou based on the birth month rule, this eligibility does not change their high school enrollment decision.

finishing middle school at the individual level is not reported in my data, I cannot implement 2SLS for the fuzzy design as generally suggested in the literature.

This model structure is different from both conventional sharp and fuzzy models. I provide an extension of these standard models as described in the Appendix A, in which the treatment status is a *deterministic* function of the assignment index, birth month, and other known but unobserved variables. One crucial assumption for identifying the treatment effect is the following:

Assumption 1. z_i and d_i are independent for $z_i \in [z_0, z_0 + \epsilon)$, where ϵ is a small positive number.

This assumption means the probability of making one's high school choice in or after 1998 is uncorrelated with the birth month for individuals born in or *just* after September 1980. For this assumption to be valid, the birth month z_i needs to be uncorrelated with the three factors that determine d_i : the year enrolled in primary school, the legal length of primary school, and the probability of repeating a grade.¹⁶ First of all, those reaching six years old by August 31 each year are allowed to enroll in primary school on September 1. It indicates that those born in August and September of the same year t are eligible for school at different years, e.g. $t + 6$ and $t + 7$ respectively. However, this primary school enrollment cutoff date will not cause a difference in school starting year between those born in September and October of the same year t . Thus, it is plausible to assume there is no variation in the distribution of age starting school for a small range of birth date no earlier than September 1. Secondly, the legal length of primary school is dependent on county policy but is independent of birth date.¹⁷ Thirdly, Fertig & Kluve (2005) show that there is no effect of actual age at school entry measured in months on the probability of repeating a grade in Germany, providing some evidence of zero correlation between the probability of repeating a grade and the time of birth. Therefore, it is reasonable to assume that d_i and z_i are independent just after the cutoff.

Under assumption 1, the jump of treatment probability at the cutoff of $z_0 = \text{Sept. 1980}$ can be simplified as follows:

$$\begin{aligned}
& \lim_{z \rightarrow z_0^+} E[x_i | z_i = z] - \lim_{z \rightarrow z_0^-} E[x_i | z_i = z] \\
&= \lim_{z \rightarrow z_0^+} E[x_i | z_i = z] \quad (\text{since } \lim_{z \rightarrow z_0^-} E[x_i | z_i = z] = 0) \\
&= \lim_{z \rightarrow z_0^+} \{E[x_i | z_i = z, I_2 = 1] \times Pr[I_2 = 1 | z_i = z] + \\
&\quad E[x_i | z_i = z, I_2 = 0] \times Pr[I_2 = 0 | z_i = z]\} \\
&= \lim_{z \rightarrow z_0^+} Pr[I_2 = 1 | z_i = z] \\
&= \lim_{z \rightarrow z_0^+} Pr[I_2 = 1] \quad (\text{assumption 1})
\end{aligned}$$

Thus, I can obtain the treatment effect by estimating the raw jump in the high school enrollment rate, as in the sharp design, and rescaling it using an estimated proportion of those finishing middle school in or after 1998 for the birth cohort of Sept. 1980-Aug. 1981 as illustrated in Figure 1.¹⁸ The only way to estimate this proportion is to exploit the school

¹⁶The legal length of middle school is three years for all counties.

¹⁷Primary school education in a few counties takes five years instead of the usual length of six years.

¹⁸This rescaling method is derived following Hahn, Todd & Van der Klaauw (2001). When the discontinuity

status information of the 2000 census since the year of finishing middle school is not covered in the questionnaire. Each individual reported the school status of “in school”, “finished” or “dropout” in addition to highest education level obtained. People born between September 1983 and August 1984 finished middle school in 2001 or later if they are still in middle school when the census took place in November 2000. The proportion of the Sept. 1983-Aug. 1984 birth cohort finishing middle school in or after 2001 can be used to infer the proportion of the Sept. 1980-Aug. 1981 cohort making a high school decision in or after 1998 under the following assumption:

Assumption 2. *The distribution of the middle school finishing age is the same across birth year cohorts.*

Note that the two different assumptions I adopt represent two extreme cases. If individuals have perfect foresight about the policy change, the sharp RD estimate provides the local average treatment effect (LATE). In contrast, if no information is revealed before the policy change, then the treatment effect can be obtained by rescaling the first stage sharp RD estimator using the discontinuity in the treated proportion. Given that the policy change is hard to predict for the majority, I report empirical results based on this second assumption. In addition, the impact of obtaining urban Hukou on education for all other cases with different degrees of information limitation lies between these two estimates, one with rescaling and the other without.

One major limitation of the estimation method adopted here is the inability to infer the average treatment effect for the whole population. The effect identified here can only be interpreted as the local average treatment effect for individuals finishing middle school in or after 1998. This is equivalent to finishing middle school at an age of 17 or older for individuals born just after the threshold of September 1980.¹⁹ On the one hand, these late finishers may have smaller wage returns to education due to unobserved below average ability. Thus, their institutional returns contribute to a bigger share of the returns to education, and the 1998 Hukou policy change affects them more than average. On the other hand, if finishing middle school late is due to credit constraint, the possible decrease in tuition brought by local Hukou can ease this constraint, resulting in an underestimation of the average treatment effect. Moreover, individuals with a rural mother and an urban father are from a relatively well-off group as compared to children from typical rural households. They are more likely to have better educated parents, especially fathers, on average.²⁰ I return to this generalization issue in the conclusion.

5. EMPIRICAL RESULTS

In this section, I first show the visual evidence of a drop in high school attendance rate for those born in or after September 1980.²¹ I then discuss the bandwidth and polynomial

of treatment probability is less than one, they show that the local treatment effect can be identified using $\frac{y^+ - y^-}{x^+ - x^-}$, where the numerator represents the discontinuity in an outcome and the denominator is the discontinuity of treatment probability. I bootstrap to get the standard errors.

¹⁹Age here is defined as the age at one’s last birthday. For example, the age of a child born in September 1, 1980 is 17 on any date between September 1, 1997 and August 31, 1998. His/her age reaches 18 on September 1, 1998.

²⁰Define educated parents as those with at least some middle school education. For children with an urban father and a rural mother, 69% of them have an educated father and 34% have an educated mother according to my data. These proportions drop to 54% and 28%, respectively, for children with both parents holding rural Hukou.

²¹High school includes technical and regular high schools. I focus on both types of high schools instead of

choices for the nonparametric estimation and present the regression results for high school attendance rate. For the same sample, I also examine possible changes in the middle school graduation rate at the September 1980 threshold. At the end of this section, I also present the additional evidence of an increased dropout rate of high school and above for those born in or after September 1980, which also supports the existence of institutional returns to education.

5.1 High school attendance

The group of interest is those with a father holding urban Hukou and a mother holding rural Hukou.²² Before proceeding to formal analysis, I plot the proportion of students who ever attended high school against month of birth for a subsample born between September 1971 and August 1986 with a father holding urban Hukou and a mother holding rural Hukou (see Figure 2). I restrict the sample to individuals with at least some middle school education and exclude those still in middle school at the time of the census.²³ The curves show quadratic fit to the left and right of the threshold. The visual evidence shows a clear decrease in the high school attendance rate at the cutoff value 0, which represents September 1980.²⁴

The next estimation issue has to do with the bandwidth and polynomial choice in the local polynomial regression. I choose a bandwidth of 43 months for the main results based on the leave-one-out cross-validation procedure proposed by Imbens & Lemieux (2008) and Ludwig & Miller (2007). I also report the results using different bandwidths in Table 5. The estimates remain similar in magnitude.

I adopt a local linear regression according to Hahn, Todd & Van der Klaauw (2001). To support this specification, I include dummies for each value of the birth month along with a piecewise linear control and test the joint significance of those dummies.²⁵ If they are jointly significant, then the piecewise linear regression is mis-specified. The test statistic fails to reject the first order polynomial specification with a p-value of 0.9768. In addition, based on the Akaike Information Criterion (AIC), a first order polynomial fits best compared

technical high school and college for two reasons. On the one hand, the timing of making a certain education decision is important for estimations. High school decisions are made after middle school while college decisions are made after regular high school with a delay of around three years. On the other hand, due to this delay, many individuals in my sample had not made college decisions by the time of the census.

²²The 2000 census lacks direct parent-child information for each surveyed individual, but only contains information about the relationship of each member to household head. The criteria for identifying parent-child relationship are described in Table B.1.

²³Most of the individuals who are still in middle school were born in 1984 or later. Including them will result in low high school enrollment rate for later birth cohorts. Nonetheless, the estimated discontinuity at the cutoff of September 1980 is unchanged with these additional observations.

²⁴One concern of the threshold is possible delays in nationwide policy implementation as well as timing variations introduced by local government, which may result in a different cutoff birth date. In order to address this issue, I perform a goodness-of-fit exercise to test for possible discontinuities along the assignment variable following Ludwig & Miller (2007), Card, Mas & Rothstein (2008) and Ozier (2010). I use a subsample including individuals born between 43 months, the optimal bandwidth based on cross-validation method, before and after September 1980. For each month between March 1979 and March 1982, I regress high school enrollment/dropout on a dummy indicating born on or after this potential discontinuity and a piecewise linear control for birth month, allowing the slope to be different on each side of the threshold. The estimated cutoff is the one with the best fit. This method is proved to be highly consistent by Hansen (2000). As shown in Figure B.2, September 1980 maximizes R^2 and is considered to be the “true” cutoff.

²⁵See Lee & Lemieux (2010) for detailed discussion of this test. Using dummies for bigger bins generates similar results.

to constant, quadratic, cubic and quartic specifications.²⁶ Thus, local linear regression with a bandwidth of 43 months will be used in the remaining analysis.

In the RD design, the underlying assumption guaranteeing locally random treatment is the smoothness in the potential outcomes, which indicates agents' inability to precisely manipulate the assignment variable and sort around the threshold (Lee, 2008). Because the births of affected individuals occurred around 18 years prior to the policy change, the birth dates of individuals around the threshold cannot be manipulated. Consistent with this argument, the density of birth dates around the cutoff date displays no noticeable jump when I plot the number of observations of each birth month in Figure 5a with a quadratic fit. The density smoothness test proposed by McCrary (2008) fails rejection at September 1980, providing additional support for continuous density of birth month.

The sample used here for nonparametric estimation includes people born within 43 months before and after the threshold of September 1980 (i.e. February 1977 - March 1984) with a father holding urban Hukou and a mother holding rural Hukou. Table 1 presents the summary statistics of a few selected variables. Regression results are presented in Table 2a Column 1.²⁷ The probability of high school enrollment decreases by about 10 percentage points for cohorts born just after the threshold of September 1980 compared to those born just before. This result is robust to variations of bandwidth choice as shown in Table 5 Column 1-5.

I use the school status of Sept. 1983-Aug. 1984 birth cohort to infer the probability of treatment. 28.5% of this cohort was still in middle school when the 2000 census took place, indicating an year of 2001 or later when making their high school decision.²⁸ I use this estimated percentage to approximate the proportion of the Sept. 1980-Aug. 1981 birth cohort finishing middle school in or after 1998 and having the opportunity to adjust high school choice according to the new policy.

The local average treatment effect can be estimated by the ratio of the probability jump of high school attendance to the probability jump of treatment. Being eligible to change Hukou status to urban decreases the probability of enrollment in high school by 30.5 percentage points, as shown in Table 2b Column 1. As the 1998 policy change happened after the annual high school entrance exam, there was no time to adjust the effort level in middle school. Therefore, this result can be viewed as a short term effect. In the long term, the negative impact on high school enrollment will be even bigger since individuals who get urban Hukou will invest less time and money on their middle school education.

The estimated 30.5 percentage points drop in high school attendance rate is the local average treatment effect. However, this effect may differ by gender. Table 2 reports the estimation results for male and female separately using a same bandwidth of 43 months. The estimated drop in high school attendance rate is 37.9 percentage point points for males, which is statistically significant and is bigger than the average effect. In contrast, the effect on females is relatively smaller and is no longer significant. The stronger effect on males indicates a higher institutional returns to education brought by the urban Hukou. One possible explanation for this return difference between genders is that males are more likely to

²⁶See Lee & Lemieux (2010) for a discussion of the limitations of the AIC method for polynomial choice in RD design.

²⁷I modify Stata code provided by Imbens & Kalyanaram to allow for user specified bandwidth and apply it to all the following nonparametric regressions.

²⁸Among the total 414 individuals born between September 1983 and August 1984, 118 were still in middle school at the time of the census.

find a job in China and better utilize urban Hukou status. Among those aged 25 and above, the employment rates for male and female are 91% and 77%, respectively.²⁹ Another possible explanation for males' higher valuation on the urban Hukou stems from the unbalanced sex ratio in contemporary China. Males with urban Hukou are more likely to succeed in the highly competitive marriage market.

The heterogeneous effects can also arise between different urban Hukou locations that one can obtain. The value of getting an urban Hukou of a big city in coastal areas may be different from getting an urban Hukou of a small town in the relatively poorer western China. In order to test this hypothesis, I rank all provinces in China according to their per capita income and categorize the top 50% as the rich region and the rest as the poor region. I then group individuals into these two regions according to their father's urban Hukou locations. Table 2 shows the different effects for the two groups. The likelihood to enroll in high school significantly decreases by 45.6 percentage points for individuals that are able to get urban Hukou from relatively rich areas. In contrast, getting an urban Hukou from relatively poor areas does not affect one's high school decision much.

5.2 Middle school graduation

The sample in the high school attendance estimation includes those with at least some middle school education. Since part of the returns to middle school education stems from the option value of attending high school, the reduced benefits associated with high school degree is expected to weaken the incentive to graduate from middle school. Thus, the estimated drop in high school attendance rate can be divided into two parts: (1) a lowered middle school graduation rate for those ever attended middle school, and (2) a decreased likelihood for enrolling in high school for middle school graduates.

First, the visual evidence in Figure 3 shows a slightly decrease in middle school graduation rate at the threshold of September 1980. To formally test to what extent the change in middle school graduation rate can explain the decreased high school attendance, I run local linear regression for middle school dropout rate using the same sample. The estimation focuses on a bandwidth of 52 months, chosen by the leave-one-out cross-validation procedure. According to the regression results in Table 3b Column 1, the change in middle school dropout rate at the cutoff of September 1980 is small and not statistically significant. The result is similar using different bandwidths as shown in Table 6. The lack of discontinuity may result from the compulsory education for primary and middle school in China.

To further test whether the Hukou reform has different impacts on different people, I estimate the change in middle school graduation rate for each gender and each previously defined region, separately. Among all four groups, the rich region is the only one with statistically significant results in the first stage estimation. Being able to obtain an urban Hukou from relatively rich areas decreases the middle school graduation rate by 2.5 percentage points for those born after the cutoff. After rescaling, this amounts to a 9.2 percentage point decrease in the probability of completing middle school.

²⁹Author's calculation according to "China 2000 Population Census Data Assembly", National Bureau of Statistics of China.

5.3 Dropout of high school and above

If obtaining urban Hukou decreases the returns to high school education, the 1998 Hukou policy reform would also affect the dropout decisions of students already in high school and above at the time of the policy change, e.g. early finishers of middle school. I restrict the sample to those with at least some high school education and plot the proportion of dropouts against month of birth in Figure 4. There is an increase in the dropout rate at the cutoff. Table 4 presents the regression results using a bandwidth of 37 months selected by the cross-validation procedure. The dropout rate of those born just after the threshold of September 1980 increases by 4.4 percentage points compared to those born just before. In addition, this effect is stronger for males and for obtaining urban Hukou from relatively rich areas.

However, this result may be contaminated by the non-smoothness of the assignment variable, birth month, at the cutoff. Since individuals born just after September 1980 are less likely to enroll in high school as documented in section 5.1, there may be fewer observations just to the right of the cutoff. This concern is confirmed by a plot of the density of observations in Figure 5b. The discontinuity in density is not an indicator for the manipulation of birth month but is the result of another regression discontinuity. Moreover, the sample composition of individuals born after the threshold may change as well in terms of their gender and their father's Hukou location. Nonetheless, individuals born just after September 1980 includes those late middle school finishers (1998 or later) choosing to enroll in high school regardless of the decreased returns to education. They may be less likely to drop out compared to those born before September 1980. Thus, the increase in dropout rate caused by early finishers may be underestimated when comparing dropout rate before and after the birth month threshold.

Even though the true effect cannot be consistently estimated in RD without a smooth density of birth month at September 1980, the increased dropout rate of high school and above is in line with previous results for high school enrollment rate and middle school graduation rate, reinforcing the existence of "institutional returns" to education and its important role in human capital investment decision in China.

6. ROBUSTNESS CHECKS

This section presents results of robustness checks such as testing the continuity assumption of the covariates required for a valid RD design, conducting local linear regressions including these additional covariates and presenting parametric regression results taking into account the clustered nature of the errors. Moreover, I discuss the potential issue of age-induced variations in educational achievement. At the end of this section, I carry out a placebo test for a group that were not affected by the 1998 Hukou reform (i.e. with both parents holding rural Hukou) to show that the drop in high school enrollment rate for the group with a rural mother and an urban father is not likely to be a result of other nationwide policy changes.

6.1 Continuity of covariates around cutoff

Valid RD design requires smooth covariates over the cutoff. For the sample used in the analysis of high school enrollment rate, I check for possible jumps over the threshold for

parents' education, number of siblings and gender.³⁰ Following a brief description of these covariates, I present the empirical results.

Parents' education is used to capture the inter-generational education linkage that affects children's schooling.³¹ I construct two dummy variables for fathers and mothers holding middle school degrees and above, respectively, and check for possible jumps of these middle school indicators.³²

Gender has also been recognized as one important determinant of educational attainment: males may receive better childcare and educational opportunities since the Chinese culture values boys; on the other hand, there may be cognitive differences between boys and girls. For instance, the latter have an advantage in lower-level school since they mature earlier, have better control of their behavior and are better able to concentrate. Thus, I also check for the smoothness of gender composition in the sample.

The number of siblings is expected to decrease educational attainment for a given individual since they would compete for educational resources. To calculate the number of living siblings, I subtract one from the mother's number of children ever born and still alive at the time of survey.³³ This method excludes adopted children³⁴, who compete for household resources as well. The number of children who are adopted appears to be very small. According to Chinese law, adoption is approved if a couple does not have their own children. Therefore, the adoption can be inferred if the calculated number of living siblings equals -1 (i.e. the mother's number of birth equals zero). There are fewer than one percent adoption cases in the sample. Thus, I do not distinguish between genetic and adopted children in this study.

Figure 6 graphically presents the mean value of each covariate in 6 month bins separately with a quadratic fit. The visual evidence shows no significant discontinuity before and after September 1980 for all of these variables. As suggested by Lee & Lemieux (2010), I test the joint significance of all the discontinuities at the threshold in a Seemingly Unrelated Regression (SUR), where each equation regresses one covariate on a threshold dummy, a constant and a fourth order polynomial of birth month. The coefficients of polynomials are allowed to be different on each side of the threshold and errors are allowed to be correlated across equations. This test fails to reject the hypothesis that covariates are smooth across this cutoff.

6.2 Incorporating covariates in estimation

Valid RD estimates change little with additional covariates. Adding covariates that have good explanatory power may help reduce variance of the RD estimator. As a robustness

³⁰All these variables are significant at 1% level in the DID regression. An appendix reporting the full results is available from the author upon request.

³¹Parents' education ranges from no school to college in census as follows: illiteracy, primary school, middle school, regular high school, technical high school, junior college, college, graduate and above.

³²The threshold of middle school is chosen according to the DID results. They show parents with middle school degree or higher have positive influence on their children's probability of attending high school.

³³This measure has the advantage of including older children who already left the household, as compared with the number of siblings calculated based on presence at the household. Besides, the results are largely unchanged when I use the latter measure of the number of siblings.

³⁴This measure excludes stepchildren as well. The identification of stepchildren is not feasible given limited information collected by the 2000 census. However, this is unlikely to invalidate my result given the low divorce and remarriage rates in China.

check, I add parents' education indicators (1 for at least middle school), number of siblings, and gender in the original local linear regression. As reported in Table 2 column 2, the high school attendance estimates are robust to the additional explanatory variables.

The calculation of the number of siblings is through the mother's number of children ever born and is based on identification of parent-child relationship, which is not directly reported in 2000 census and cannot be fully exploited. Including the number of siblings as one of the explanatory variables reduces the sample to two thirds of the original size for both outcome variables. In order to avoid potential problems induced by fewer observations, I also report the regression results in Table 2 column 3, excluding the number of siblings. The estimate of discontinuity remains significant at a 5% level with unchanged magnitude. Regression results for other education outcomes are also robust to the inclusion of these covariates.

6.3 Parametric estimation with clustered standard errors

Lee & Card (2008) study RD design with a discrete assignment variable and argue that comparing outcomes in very narrow bins just to the right and left of the cutoff is not possible in discrete cases. Parametric estimation is more efficient if the functional form is correctly specified and the clustered nature of errors is taken into account. I estimate the equation of some education outcome to be dependent on birth month and allow the errors to be clustered at birth month level. The coefficients are allowed to be different on both sides of the cutoff:

$$Y_i = \alpha + \tau \times D_i + \sum \beta_{lj} \times (z_i - z_0)^j + \sum (\beta_{ri} - \beta_{lj}) \times D_i \times (z_i - z_0)^j + \epsilon_i,$$

where Y_i is a dummy variable for high school attendance or middle school graduation; z_i is the birth month with a cutoff value of $z_0 = \text{Sept.1980}$; D_i is an indicator that takes the value of one if birth month is in or after September 1980 and τ is the estimated probability jump of high school enrollment. I focus on quadratic and quartic models, allowing the coefficients to be different on either side of the discontinuity.³⁵ The OLS results for high school attendance rate and middle school graduation rate are shown in Column 5 and Column 7 of Table 5 and Table 6, respectively. To further check the robustness of the estimates, I also report regression results including parents' education, gender and number of siblings as additional controls in Column 6 and Column 8. The estimated discontinuities are all similar in magnitude to those obtained from the local linear regressions.

6.4 Age induced variation

The school enrollment cutoff date in China is September 1. Children who become 6 years old before this date can enroll in primary school in the same year. Otherwise, they have to wait for another year to start school. Therefore, children born in September are older compared to students in the same grade while those born in August are relatively younger, which may lead to different performance in school.

There is a broad literature analyzing the impact of school starting time on educational attainment. Dobkin & Ferreira (2010) find that students who are the youngest in their school

³⁵Both second and fourth order polynomials pass the goodness-of-fit test proposed in Lee & Card (2008).

cohort have slightly higher educational attainment using U.S. data. However, this may be due to age-based mandatory school attendance laws in U.S. Children starting primary education at earlier ages have to stay enrolled longer before reaching the legal school leaving age of 16.

Unlike the U.S., China has compulsory school laws based on years of schooling instead of age. Sharing the same feature in mandatory school policy, research results using European countries' data are more suitable to make inferences for China. Black, Devereux & Salvanes (2011) find no effect of school starting age on educational attainment in Norway. Furthermore, Fertig & Kluge (2005)'s study of Germany generates similar results. To further test possible age induced schooling variation in China, I run local linear regression with the same bandwidth of 43 months using the same sample but a different cutoff of September 1976, which is four years before the true cutoff. There is no significant change in high school enrollment rate before and after this threshold.³⁶ Hence, the probability jump of high school enrollment at September 1980 is not likely to be caused by school entry cutoff date.

6.5 A placebo test

Policies in China changed drastically in the last two decades as well as economic conditions. One may argue that the drop in high school enrollment is a possible result of other changes, such as the tuition reform and college expansion in the late 1990s or the rising unemployment problem for college graduates in 2000s. These factors likely influenced individuals' educational achievement and could have affected individuals of all birth months, not just the cohort born after a specific date.

Some support for the idea that the drop in human capital that is documented in the previous sections is not caused by other nationwide changes comes from the results of a placebo test for individuals with both parents holding urban Hukou. They should not be affected by the 1998 Hukou reform since the allowance of transferring their Hukou from mother's status to father's did not change their Hukou. I use the same 43 months bandwidth and test for possible discontinuity at the same threshold of September 1980 as I did before. If there were other factors only affecting individuals born in or after September 1980, the high school attendance of this placebo group should decrease as well. However, the change in high school enrollment rate at September 1980 is small and not statistically significant for this placebo group.³⁷ Therefore, the drop in high school enrollment for the beneficiaries of the 1998 Hukou reform is not likely to have been caused by other nationwide reforms.

7. DISCUSSION

The significant role that the Hukou system plays in fostering education in rural areas may partially correct the potential underinvestment problem in China. Economists argue that the social benefits from education may significantly exceed the private benefits resulting from substantial positive externalities of knowledge (Weisbrod, 1962; Acemoglu, 1996, 1998; Lucas Jr, 1998). When taking into account the spill-over effect, social returns may be well above private returns even if the government subsidy is included in the calculation

³⁶The point estimate is 0.025 with a standard error of 0.042.

³⁷The estimated discontinuity is -0.017 with a standard error of 0.017.

(Psacharopoulos & Patrinos, 2004). This is likely to be the case in China given its low education subsidy.³⁸ Liu (2007) is the first to empirically test for a human capital spillover effect in China using individual level data. He finds substantial external benefits of an additional year of schooling. These range from an 11% to a 13% increment in wages caused by externalities, which is at least as great as the private benefits. His finding, combined with a low subsidization index³⁹ of 1.04 to 1.31 in China⁴⁰ (Hossain, 1997), implies that the social returns exceed the private returns to education. This perception is in line with Heckman (2003)'s argument that China invested too little in human capital compared to its investment in physical capital.

Now, China is replacing the rural/urban Hukou with uniform identity Hukou. The gap between social and private returns to education may be widened, which will hinder the human capital accumulation in China. According to Fleisher, Li & Zhao (2010), if the portion of workers with at least some high school education decreases by one percentage point, the total factor productivity (TFP) decreases by about 0.5 percentage point a year in China. Therefore, the slow down of human capital accumulation may influence China's long run economic growth pattern.

8. CONCLUSION

This paper analyzes the institutional returns to education under labor mobility restrictions. Higher education serves as a tool to escape from poverty if it increases the probability of obtaining identity in more developed areas. In particular, I estimate the additional returns to education in China under the Hukou system.

As shown in this study, removing the institutional returns to education by directly granting urban Hukou decreases the high school attendance rate substantially by 8.7 to 30.5 percentage points among those holding rural Hukou and with at least some middle school education. This negative effect is stronger for males and for those getting urban Hukou in relatively rich areas. The drop in the high school attendance rate is mainly due to decreased high school enrollment rate for middle school graduates, except that the middle school completion rate also drops for those obtaining urban Hukou in rich areas. In addition, there is suggestive evidence of an increased dropout rate for high school and above as well. Therefore, the Hukou system has played an important role in encouraging high school education after the nine years of mandatory schooling.

Unfortunately, the data do not allow for analyzing long term effects of obtaining urban Hukou on other outcomes such as employment perspectives, wage earnings or marriage market performance. Many individuals in my regression sample were still in school at the time of the survey and most of them were under the legal marriage age in China.⁴¹ The effects of obtaining urban Hukou on other long-term outcomes are important aspects for future research.

Another limitation of this study is the generalizability of the results. On the one hand, an

³⁸As one indicator, according to United Nations Educational, Scientific, and Cultural Organization (UNESCO) Institute for Statistics, the Chinese government spent 1.9% of GDP on Education in 1999. This figure is well below the world average of 4.2%.

³⁹The index of public subsidization on education shows the ratio of social costs to private costs.

⁴⁰The subsidization indices estimated by Hossain (1997) are 1.25, 1.04 and 1.31 for primary, secondary and higher education respectively.

⁴¹The legal marriage age in China is 22 for male and 20 for female.

individual with a urban father and a rural mother comes from an advantaged background compared to a typical rural child with both parents holding rural Hukou. In addition, their families are more likely to locate in urban areas, which grants easy access to better education opportunities. On the other hand, late middle school finishers are likely to have lower than average ability. Their response to obtaining urban status may be different from early middle school graduates. Nonetheless, one thing that can be drawn from this study is the existence of institutional returns to education. The Hukou system creates stronger incentives for rural households to invest in human capital in China.

A few provinces have used uniform identity Hukou to replace the original rural/urban dichotomy since 2003. Understanding Hukou's role in fostering education is crucial for China during this transitional period. Because the elimination of the rural/urban Hukou status erases the additional returns to high school education, human capital investment for rural people may be negatively affected in terms of the high school attendance rate, the middle school graduation rate and the dropout rate for high school and above. Even though individuals with rural origin are able to enjoy urban benefits and higher wages instantly, their potential for career development and long term income will be restricted by limited education. In addition, this low education trap is transferable across generations as a result of positive correlation between education levels of parents and children. Even though the disparity between rural and urban areas may be removed by uniform identity, within-urban inequality may emerge based on different education levels.

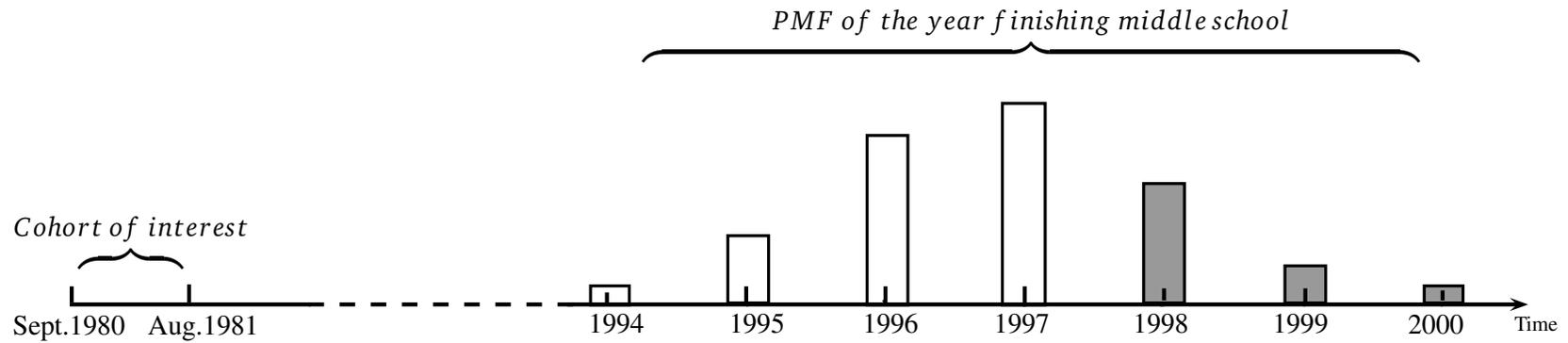
REFERENCES

- Acemoglu, D.** 1996. "A Microfoundation for Social Increasing Returns in Human Capital Accumulation." *The Quarterly Journal of Economics*, 111(3): 779.
- Acemoglu, D.** 1998. "Why Do New Technologies Complement Skills? Directed Technical Change and Wage Inequality." *The Quarterly Journal of Economics*, 113(4): 1055–1089.
- Au, C.C., and J.V. Henderson.** 2006a. "Are Chinese Cities Too Small?" *The Review of Economic Studies*, 73(3): 549.
- Au, C.C., and J.V. Henderson.** 2006b. "How Migration Restrictions Limit Agglomeration and Productivity in China." *Journal of Development Economics*, 80(2): 350–388.
- Batista, C., A. Lacuesta, and P.C. Vicente.** 2011. "Testing the 'Brain Gain' Hypothesis: Micro Evidence from Cape Verde." *Journal of Development Economics*.
- Beine, M., F. Docquier, and H. Rapoport.** 2001. "Brain Drain and Economic Growth: Theory and Evidence." *Journal of Development Economics*, 64(1): 275–289.
- Beine, M., F. Docquier, and H. Rapoport.** 2008. "Brain Drain and Human Capital Formation in Developing Countries: Winners and Losers." *Economic Journal*, 118(528): 631–652.
- Beine, Michel, Fr d ric Docquier, and Hillel Rapoport.** 2010. "On the Robustness of Brain Gain Estimates." *Annales d'Economie et de Statistique*, (97-98): 143–165.
- Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes.** 2011. "Too Young to Leave the Nest? The Effects of School Starting Age." *The Review of Economics and Statistics*, 93(2): 455–467.
- Card, D., A. Mas, and J. Rothstein.** 2008. "Tipping and the Dynamics of Segregation." *The Quarterly Journal of Economics*, 123(1): 177–218.
- Chand, Satish, and Michael Clemens.** 2008. "Skilled Emigration and Skill Creation: A quasi-experiment." Center for Global Development Working Papers 152.
- Chan, K.W.** 2009. "The Chinese Hukou System at 50." *Eurasian Geography and Economics*, 50(2): 197–221.
- De Brauw, A., and J. Giles.** 2008. "Migrant Opportunity and the Educational Attainment of Youth in Rural China." The World Bank Policy Research Working Paper Series 4526.
- Dobkin, C., and F. Ferreira.** 2010. "Do School Entry Laws Affect Educational Attainment and Labor Market Outcomes?" *Economics of Education Review*, 29(1): 40–54.
- Fertig, M., and J. Kluve.** 2005. "The Effect of Age at School Entry on Educational Attainment in Germany." Institute for the Study of Labor (IZA) IZA Discussion Papers 1507.
- Fleisher, Belton, Haizheng Li, and Min Qiang Zhao.** 2010. "Human Capital, Economic growth, and Regional Inequality in China." *Journal of Development Economics*, 92(2): 215–231.
- Fujita, M., T. Mori, J.V. Henderson, and Y. Kanemoto.** 2004. "Spatial Distribution of Economic Activities in Japan and China." *Handbook of Regional and Urban Economics*, 4: 2911–2977.

- Glewwe, P., and H.G. Jacoby.** 2004. "Economic Growth and the Demand for Education: Is There a Wealth Effect?" *Journal of Development Economics*, 74(1): 33–51.
- Hahn, J., P. Todd, and W. Van der Klaauw.** 2001. "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design." *Econometrica*, 69(1): 201–209.
- Hamilton, B., and J. Whalley.** 1984. "Efficiency and Distributional Implications of Global Restrictions on Labour Mobility: Calculations and Policy Implications." *Journal of Development Economics*, 14(1): 61–75.
- Hansen, B.E.** 2000. "Sample Splitting and Threshold Estimation." *Econometrica*, 68(3): 575–603.
- Heckman, James J.** 2003. "China's Investment in Human Capital." *Economic Development and Cultural Change*, 51(4): 795–804.
- Hossain, Shaikh I.** 1997. "Making Education in China Equitable and Efficient." The World Bank Policy Research Working Paper Series 1814.
- Imbens, G.W., and T. Lemieux.** 2008. "Regression Discontinuity Designs: A Guide to Practice." *Journal of Econometrics*, 142(2): 615–635.
- Lee, D., and T. Lemieux.** 2010. "Regression Discontinuity Designs in Economics." *Journal of Economic Literature*, 48(2): 281–355.
- Lee, D.S.** 2008. "Randomized Experiments from Non-Random Selection in U.S. House Elections." *Journal of Econometrics*, 142(2): 675–697.
- Lee, D.S., and D. Card.** 2008. "Regression Discontinuity Inference with Specification Error." *Journal of Econometrics*, 142(2): 655–674.
- Liu, Z.** 2005. "Institution and Inequality: The Hukou System in China." *Journal of Comparative Economics*, 33(1): 133–157.
- Liu, Z.** 2007. "The External Returns to Education: Evidence from Chinese Cities." *Journal of Urban Economics*, 61(3): 542–564.
- Lucas Jr, R.E.** 1998. "On the Mechanics of Economic Development." *Frontiers of research in economic theory: the Nancy L. Schwartz Memorial Lectures, 1983-1997*, 61.
- Ludwig, Jens, and Douglas L Miller.** 2007. "Does Head Start Improve Children's Life Chances? Evidence from a Regression Discontinuity Design." *The Quarterly Journal of Economics*, 122(1): 159–208.
- McCrary, J.** 2008. "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density test." *Journal of Econometrics*, 142(2): 698–714.
- Mountford, A.** 1997. "Can a Brain Drain be Good for Growth in the Source Economy?" *Journal of Development Economics*, 53(2): 287–303.
- Ozier, O.** 2010. "The Impact of Secondary Schooling in Kenya: A Regression Discontinuity Analysis." *University of California at Berkeley, mimeo*.
- Porter, J.** 2003. "Estimation in the regression discontinuity model." *Unpublished Manuscript, Department of Economics, University of Wisconsin at Madison*, 5–19.
- Psacharopoulos, G., and H.A. Patrinos.** 2004. "Returns to Investment in Education: A Further Update." *Education economics*, 12(2): 111–134.

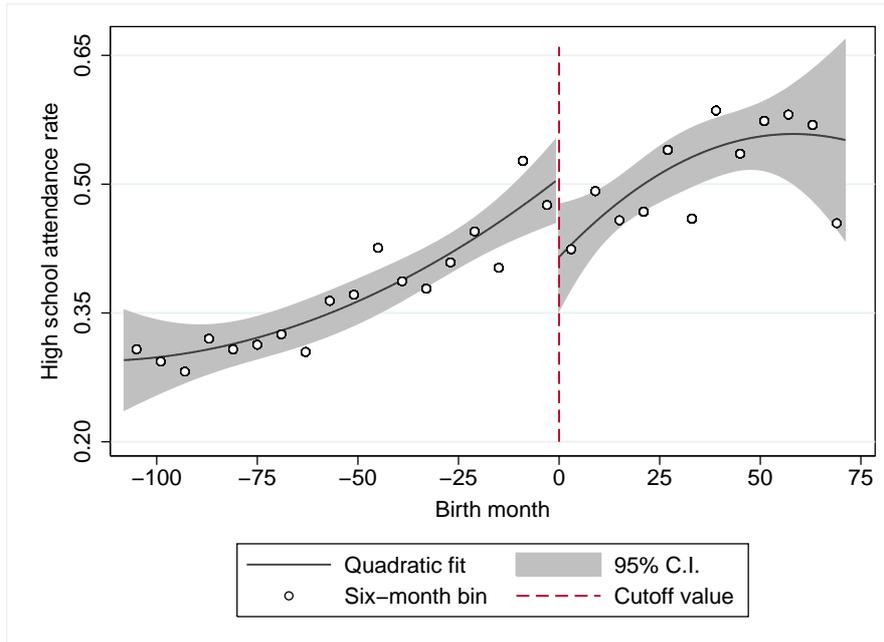
- Sicherman, Nachum, and Oded Galor.** 1990. "A Theory of Career Mobility." *Journal of Political Economy*, 98(1): pp. 169–192.
- Stark, O., C. Helmenstein, and A. Prskawetz.** 1997. "A Brain Gain with a Brain Drain." *Economics Letters*, 55(2): 227–234.
- Stark, O., C. Helmenstein, and A. Prskawetz.** 1998. "Human Capital Depletion, Human Capital Formation, and Migration: A Blessing or a "Curse"." *Economics Letters*, 60(3): 363–367.
- Vidal, J.P.** 1998. "The Effect of Emigration on Human Capital Formation." *Journal of Population Economics*, 11(4): 589–600.
- Wang, Feng, and Xuejin Zuo.** 1999. "Inside China's Cities: Institutional Barriers and Opportunities for Urban Migrants." *The American Economic Review*, 89(2): pp. 276–280.
- Weisbrod, B.A.** 1962. "Education and Investment in Human Capital." *Journal of Political Economy*, 70(5): 106–123.
- Whalley, J., and S. Zhang.** 2007. "A Numerical Simulation Analysis of (Hukou) Labour Mobility Restrictions in China." *Journal of Development Economics*, 83(2): 392–410.
- Wu, X.** 2010. "Economic transition, school expansion and educational inequality in china, 1990-2000." *Research in Social Stratification and Mobility*, 28(1): 91–108.
- Wu, X., and D.J. Treiman.** 2004. "The Household Registration system and Social Stratification in China: 1955–1996." *Demography*, 41(2): 363–384.
- Yang, Dennis Tao.** 1999. "Urban-Biased Policies and Rising Income Inequality in China." *The American Economic Review*, 89(2): pp. 306–310.
- Zhao, Y.** 1997. "Labor Migration and Returns to Rural Education in China." *American Journal of Agricultural Economics*, 79(4): 1278–1287.

Figure 1: Treatment Determination Criteria



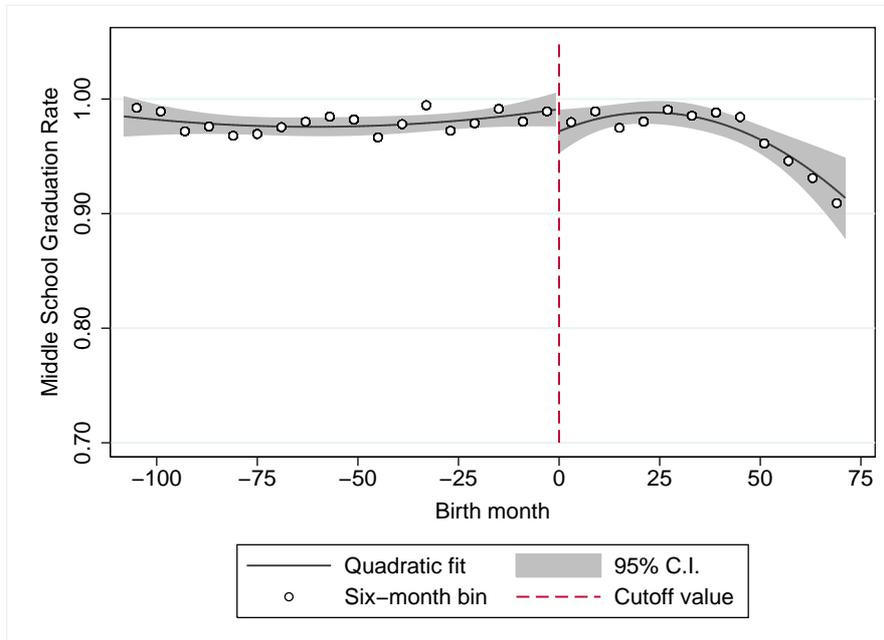
Notes: According to the 1998 Hukou reform, individuals born in or after September 1980 inherited their mother's Hukou would be eligible to change Hukou status according to their father's. Thus, children born in or after September 1980 with a mother holding rural Hukou and a father holding urban Hukou then had a chance to obtain urban status immediately (criterion 1). However, not all these children's high school enrollment decision would be affected by this policy change. Only those finishing middle school in or after 1998 had the chance to adjust their high school enrollment decision according to the new policy (criterion 2) as indicated in the shaded areas in this graph. The year distribution is drawn according to the China Health and Nutrition Survey, which may differ from the information found in the 2000 census.

Figure 2: Discontinuity in High School Attendance Rate



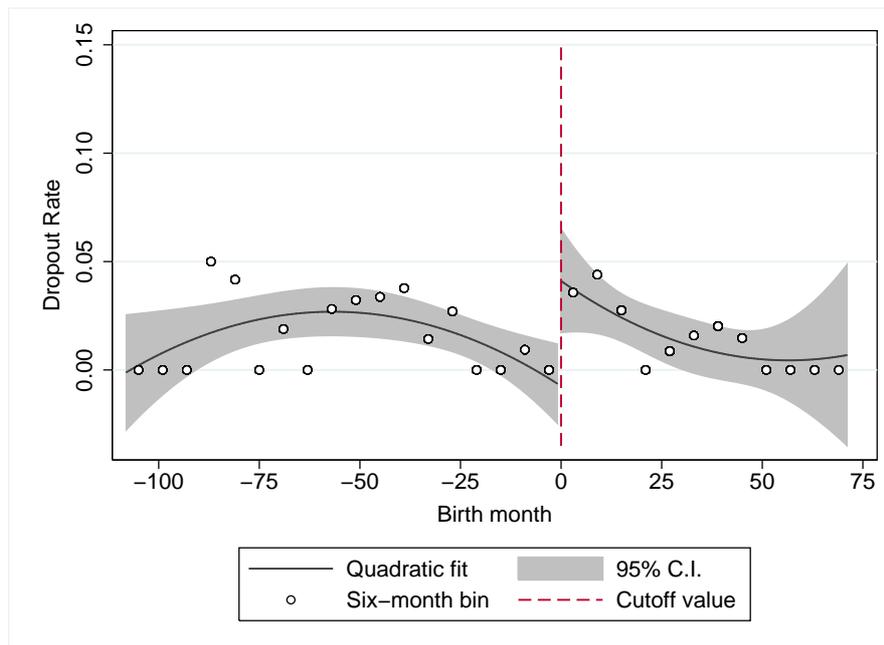
Notes: High school enrollment rate for children with a father holding urban Hukou and a mother holding rural Hukou. Sample used here includes individuals born between September 1971 and August 1986 with at least some middle school education. Those still in middle school are excluded. Birth month is normalized with Sept.1980 = 0. The estimated discontinuity is reported in absolute value.

Figure 3: Discontinuity in Middle School Graduation Rate



Notes: Middle School dropout rate for children with a father holding urban Hukou and a mother holding rural Hukou. Sample used here includes individuals born between September 1971 and August 1986 with at least some middle school education. Those still in middle school are excluded. Birth month is normalized with Sept.1980 = 0. The estimated discontinuity is reported in absolute value.

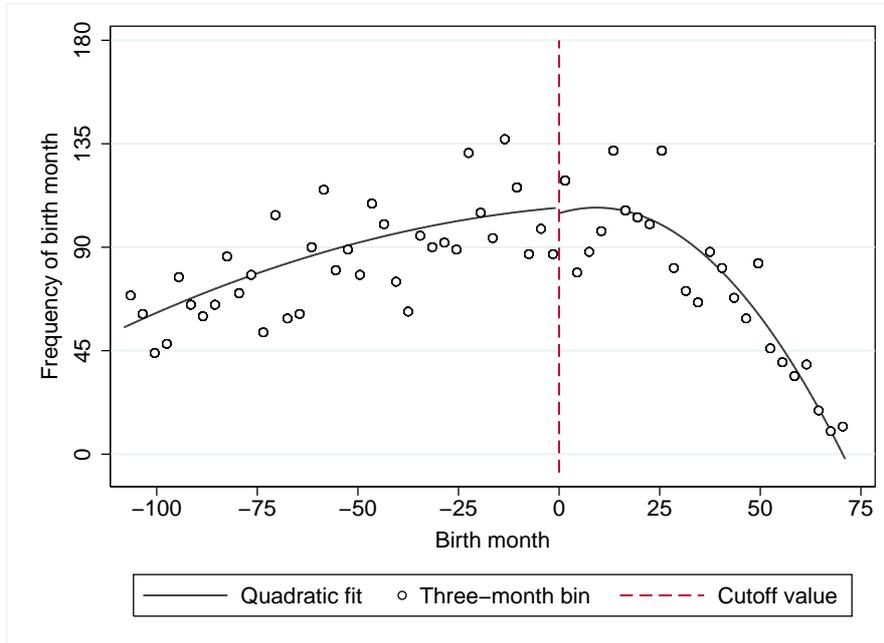
Figure 4: Discontinuity in Dropout Rate of High School and Above



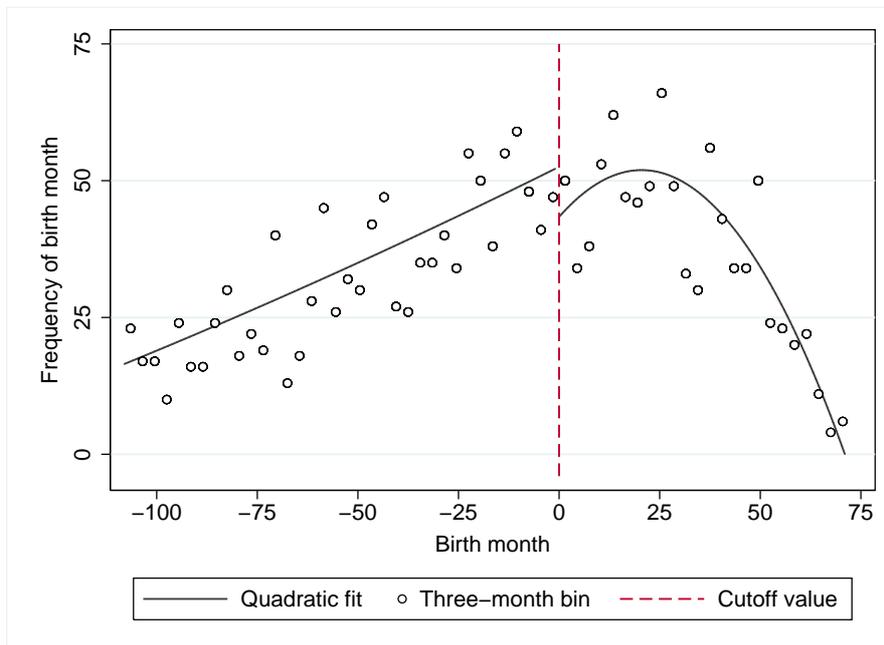
Notes: Dropout rate for children with a father holding urban Hukou and a mother holding rural Hukou. Sample used here includes individuals born between September 1971 and August 1986 with at least some high school education. Birth month is normalized with *Sept.*1980 = 0. The estimated discontinuity is reported in absolute value.

Figure 5: Density Continuity for Birth Month with Quadratic Fit

(a) Sample to Estimate High School Attendance

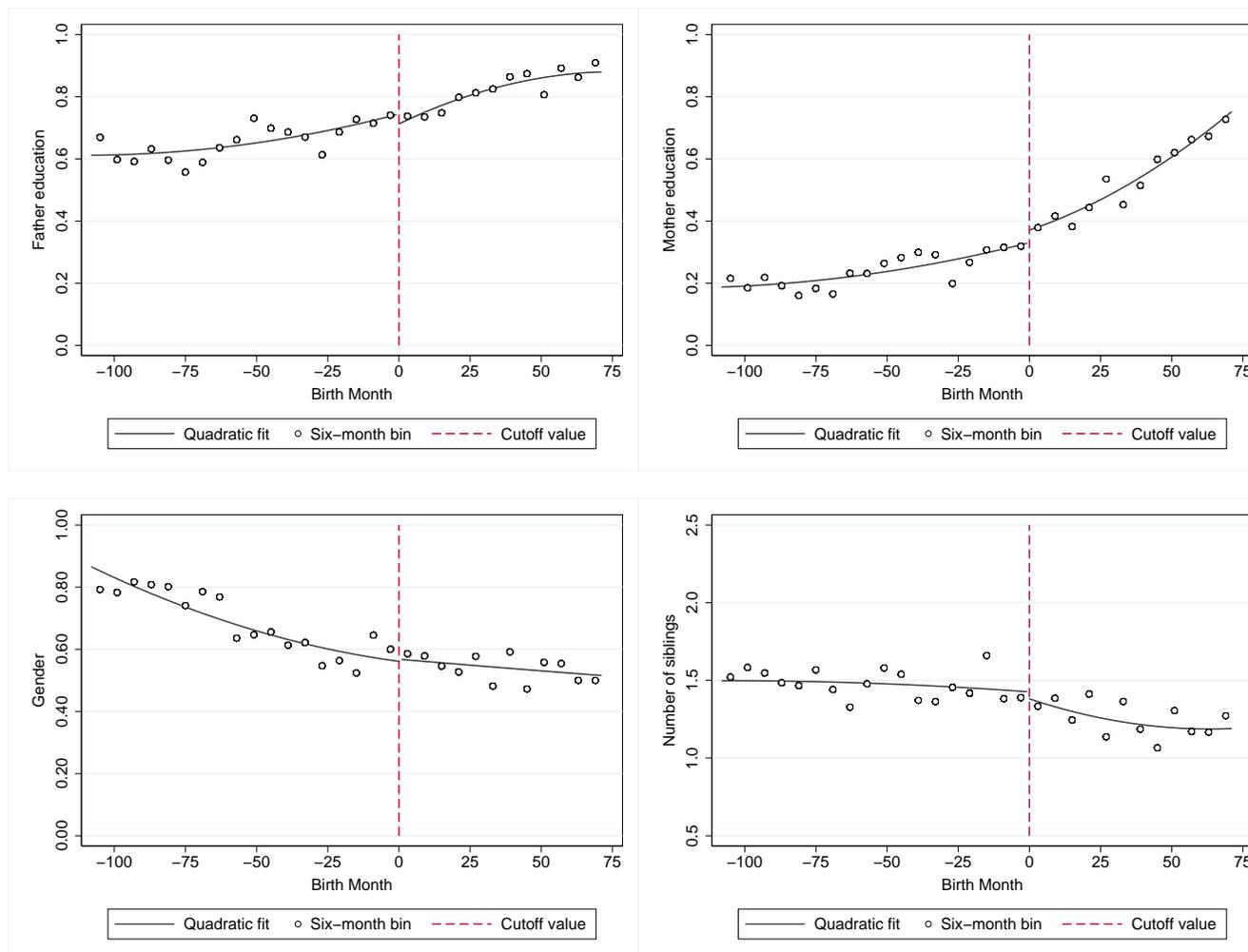


(b) Sample to Estimate Dropout Rate of High School and Above



Notes: Birth month density for children with a father holding urban Hukou and a mother holding rural Hukou. Birth month is normalized with *Sept.*1980 = 0. Sample used in the upper panel includes individuals born between September 1971 and August 1986 with at least middle school education. Those still in middle school are excluded. The lower panel further restrict the sample to individuals with at least some high school education.

Figure 6: Covariates Continuity of High School Attendance Rate



Notes: Covariates continuity of the high school enrollment analysis for children with a father holding urban Hukou and a mother holding rural Hukou. Sample used here includes individuals born between September 1971 and August 1986 with at least middle school education. Those still in middle school are excluded. Birth month is normalized with *Sept.*1980 = 0. The joint significance test using SUR fails to reject smoothness in covariates with a p-value of 0.9781.

Table 1: Summary Statistics

| SAMPLE/VARIABLE | MEAN | STANDARD DEV. | N |
|---|------|---------------|------|
| <i>Panel A: Characteristics of the full sample</i> | | | |
| Father education (middle school and above) | 0.71 | (0.45) | 4772 |
| Mother education (middle school and above) | 0.34 | (0.47) | 4772 |
| Gender (male=1) | 0.63 | (0.48) | 4772 |
| Number of siblings | 1.36 | (0.97) | 3152 |
| <i>Panel B: Subsample used for estimations of the high school attendance rate</i> | | | |
| Father education (middle school and above) | 0.74 | (0.44) | 2762 |
| Mother education(middle school and above) | 0.37 | (0.48) | 2762 |
| Gender (male=1) | 0.57 | (0.50) | 2762 |
| Number of siblings | 1.36 | (0.98) | 2191 |
| High school attendance rate | 0.46 | (0.50) | 2762 |
| <i>Panel C: Subsample used for estimations of the middle school graduation rate</i> | | | |
| Father education (middle school and above) | 0.74 | (0.44) | 3246 |
| Mother education(middle school and above) | 0.37 | (0.48) | 3246 |
| Gender (male=1) | 0.57 | (0.49) | 3246 |
| Number of siblings | 1.36 | (0.99) | 2534 |
| Middle school graduation rate | 0.98 | (0.13) | 3246 |
| <i>Panel D: Subsample used for estimations of the dropout rate of high school and above</i> | | | |
| Father education (middle school and above) | 0.81 | (0.40) | 1113 |
| Mother education(middle school and above) | 0.42 | (0.49) | 1113 |
| Gender (male=1) | 0.56 | (0.50) | 1113 |
| Number of siblings | 1.19 | (0.90) | 930 |
| Dropout of high school and above | 0.01 | (0.12) | 1113 |

Note: Data drawn from the 0.95% sample of the 2000 census. Panel A consists of individuals born between September 1971 and August 1986 with a father holding urban Hukou and a mother holding rural Hukou. I exclude respondents with education level lower than middle school and those still in middle school. Panel B, Panel C and Panel D restrict the sample to individuals born within 43 months, 52 months and 37 months before and after the threshold of September 1980, respectively. Panel D further restricts the sample to individuals with at least some high school education.

Table 2: Estimation Results for High School Attendance Rate

(a) First Stage: Sharp Design

| Sample | ESTIMATION | | |
|--------------------|---------------------|---------------------|---------------------|
| | (1) | (2) | (3) |
| All | -0.087** (0.042) | -0.072* (0.045) | -0.086** (0.041) |
| By Gender | | | |
| Male | -0.113** (0.054) | -0.113* (0.060) | -0.101* (0.053) |
| Female | -0.058 (0.065) | -0.022 (0.068) | -0.069 (0.064) |
| By Region | | | |
| Rich | -0.124** (0.058) | -0.151** (0.061) | -0.140** (0.056) |
| Poor | -0.049 (0.059) | -0.004 (0.064) | -0.034 (0.058) |
| Control Variables: | | | |
| Father Education | No | Yes | Yes |
| Mother Education | No | Yes | Yes |
| Gender | No | Yes | Yes |
| Number of Siblings | No | Yes | No |
| Bandwidth (months) | 43 | 43 | 43 |

(b) Second Stage: Rescaling

| | By Gender | | | By Region | |
|---|---------------------|---------------------|-------------------|---------------------|-------------------|
| | All | Male | Female | Rich | Poor |
| Prob. Discontinuity of High School Attendance | -0.087** (0.042) | -0.113** (0.054) | -0.058 (0.065) | -0.124** (0.058) | -0.049 (0.059) |
| Proportion Treated | 0.285 | 0.298 | 0.269 | 0.272 | 0.298 |
| Local Average Treatment Effect | -0.305** [0.143] | -0.379** [0.186] | -0.216 [0.243] | -0.456** [0.226] | -0.164 [0.206] |
| N | 2762 | 1576 | 1186 | 1421 | 1341 |

Note: The sample is restricted to people with at least some middle school education, excluding those still in middle school. Father education and mother education are measured as binary variables indicating middle school and above. Sample used for local linear regressions with covariates as presented in panel (a) column 2 and 3 only consists of individuals with non-missing value of these additional explanatory variables. Local average treatment effect is obtained as the ratio of the probability jump of high school enrollment to the probability jump of treatment. The standard error of LATE is rescaled by treating the denominator as a constant as a result of a slower convergence rate of the numerator. I also report the bootstrapped standard errors with 1000 replications in squared brackets. Asterisks *, ** and *** denote significant levels of 10%, 5% and 1% respectively.

Table 3: Estimation Results for Middle School Graduation Rate

(a) First Stage: Sharp Design

| Sample | ESTIMATION | | |
|--------------------|---------------------|--------------------|---------------------|
| | (1) | (2) | (3) |
| All | -0.008 (0.010) | -0.009 (0.011) | -0.008 (0.011) |
| By Gender | | | |
| Male | -0.010 (0.015) | -0.011 (0.017) | -0.010 (0.015) |
| Female | -0.007 (0.009) | -0.009 (0.010) | 0.009 (0.009) |
| By Region | | | |
| Rich | -0.025** (0.012) | -0.021* (0.012) | -0.025** (0.012) |
| Poor | 0.010 (0.016) | 0.004 (0.017) | 0.010 (0.016) |
| Control Variables: | | | |
| Father Education | No | Yes | Yes |
| Mother Education | No | Yes | Yes |
| Gender | No | Yes | Yes |
| Number of Siblings | No | Yes | No |
| Bandwidth (months) | 52 | 52 | 52 |

(b) Second Stage: Rescaling

| | By Gender | | | By Region | |
|---|-------------------|-------------------|-------------------|---------------------|------------------|
| | All | Male | Female | Rich | Poor |
| Prob. Discontinuity of Middle School Graduation | -0.008 (0.010) | -0.010 (0.015) | -0.007 (0.009) | -0.025** (0.012) | 0.010 (0.016) |
| Proportion Treated | 0.285 | 0.298 | 0.269 | 0.272 | 0.298 |
| Local Average Treatment Effect | -0.028 [0.035] | -0.034 [0.052] | -0.026 [0.044] | -0.092** [0.045] | 0.034 [0.052] |
| <i>N</i> | 2762 | 1576 | 1186 | 1421 | 1341 |

Note: The sample is restricted to people with at least some middle school education, excluding those still in middle school. Father education and mother education are measured as binary variables indicating middle school and above. Sample used for local linear regressions with covariates as presented in panel (a) column 2 and 3 only consists of individuals with non-missing value of these additional explanatory variables. Local average treatment effect is obtained as the ratio of the probability jump of high school enrollment to the probability jump of treatment. The standard error of LATE is rescaled by treating the denominator as a constant as a result of a slower convergence rate of the numerator. I also report the bootstrapped standard errors with 1000 replications in squared brackets. Asterisks *, ** and *** denote significant levels of 10%, 5% and 1% respectively.

Table 4: Estimation Results for Dropout Rate of High School and Above

| Sample | ESTIMATION | | |
|--------------------|--------------------|--------------------|---------------------|
| | (1) | (2) | (3) |
| All | 0.044** (0.018) | 0.037** (0.018) | 0.046*** (0.018) |
| By Gender | | | |
| Male | 0.046* (0.024) | 0.046* (0.027) | 0.047* (0.024) |
| Female | 0.042* (0.026) | 0.027 (0.021) | 0.043* (0.026) |
| By Region | | | |
| Rich | 0.050** (0.025) | 0.033 (0.021) | 0.051** (0.026) |
| Poor | 0.037 (0.024) | 0.033 (0.022) | 0.034* (0.020) |
| Control Variables: | | | |
| Father Education | No | Yes | Yes |
| Mother Education | No | Yes | Yes |
| Gender | No | Yes | Yes |
| Number of Siblings | No | Yes | No |
| Bandwidth (months) | 37 | 37 | 37 |

Note: The sample is restricted to those with at least some high school education. Father education and mother education are measured as binary variables indicating middle school and above. Sample used for local linear regressions with covariates as presented in column 2 and 3 only consists of individuals with non-missing value of these additional explanatory variables. Asterisks *, ** and *** denote significant levels of 10%, 5% and 1% respectively.

Table 5: Bandwidth Sensitivity of High School Attendance Rate (First Stage)

| Sample | Parametric | | | | | | | |
|--------------------|---------------------|---------------------|---------------------|---------------------|---------------------|----------------------|---------------------|----------------------|
| | Nonparametric | | | | Quadratic | | Quartic | |
| | 20 | 30 | 40 | 50 | No Cont. | Controls | No Cont. | Controls |
| Bandwidth (months) | 20 | 30 | 40 | 50 | - | - | - | - |
| All | -0.114* (0.061) | -0.097** (0.049) | -0.089** (0.043) | -0.077** (0.038) | -0.092** (0.037) | -0.067** (0.033) | -0.127** (0.054) | -0.133** (0.046) |
| By Gender | | | | | | | | |
| Male | -0.162** (0.079) | -0.140** (0.064) | -0.120** (0.056) | -0.092* (0.050) | -0.116** (0.049) | -0.076 (0.051) | -0.163** (0.074) | -0.217*** (0.080) |
| Female | -0.052 (0.094) | -0.046 (0.077) | -0.053 (0.067) | -0.062 (0.060) | -0.060 (0.066) | -0.054 (0.060) | -0.077 (0.101) | -0.026 (0.090) |
| By Region | | | | | | | | |
| Rich | -0.164* (0.085) | -0.144** (0.069) | -0.128** (0.060) | -0.112** (0.053) | -0.115** (0.054) | -0.141*** (0.053) | -0.129* (0.076) | -0.206*** (0.073) |
| Poor | -0.065 (0.086) | -0.049 (0.070) | -0.049 (0.061) | -0.040 (0.055) | -0.061 (0.053) | 0.013 (0.057) | -0.127* (0.076) | -0.083 (0.085) |

Note: Sample restricted to those with at least high school education. Father education and mother education are measured as binary variables indicating middle school and above. Sample used for local linear regressions with covariates as presented in column 2 and 3 only consists of individuals with non-missing value of these additional explanatory variables. Asterisks *, ** and *** denote significant levels of 10%, 5% and 1% respectively.

Table 6: Bandwidth Sensitivity of Middle School Graduation Rate (First Stage)

| Sample | Parametric | | | | | | | |
|--------------------|-------------------|--------------------|---------------------|---------------------|--------------------|--------------------|-------------------|-------------------|
| | Nonparametric | | | | Quadratic | | Quartic | |
| | 30 | 40 | 50 | 60 | No Cont. | Controls | No Cont. | Controls |
| Bandwidth (months) | 30 | 40 | 50 | 60 | - | - | - | - |
| All | -0.007 (0.013) | -0.008 (0.011) | -0.008 (0.010) | -0.006 (0.009) | -0.019* (0.011) | -0.020* (0.012) | -0.006 (0.017) | -0.004 (0.016) |
| By Gender | | | | | | | | |
| Male | -0.011 (0.022) | -0.010 (0.018) | -0.010 (0.016) | -0.009 (0.014) | -0.027* (0.015) | -0.027* (0.015) | -0.012 (0.025) | -0.002 (0.025) |
| Female | -0.004 (0.009) | -0.007 (0.009) | -0.008 (0.009) | -0.004 (0.009) | -0.011 (0.012) | -0.011 (0.013) | -0.001 (0.012) | -0.008 (0.014) |
| By Region | | | | | | | | |
| Rich | -0.022 (0.016) | -0.023* (0.014) | -0.025** (0.012) | -0.023** (0.012) | -0.034* (0.018) | -0.023 (0.016) | -0.022 (0.032) | -0.018 (0.024) |
| Poor | 0.009 (0.021) | 0.008 (0.018) | 0.009 (0.016) | 0.011 (0.014) | -0.003 (0.017) | -0.014 (0.020) | 0.011 (0.025) | 0.012 (0.028) |

Note: Sample restricted to those with at least high school education. Father education and mother education are measured as binary variables indicating middle school and above. Sample used for local linear regressions with covariates as presented in column 2 and 3 only consists of individuals with non-missing value of these additional explanatory variables. Asterisks *, ** and *** denote significant levels of 10%, 5% and 1% respectively.

Appendices

A. REGRESSION DISCONTINUITY WITH UNOBSERVED TREATMENT VARIABLE

A.1 Sharp and fuzzy regression discontinuity design

There are two types of regression discontinuity design in the literature: sharp and fuzzy. Sharp design requires the probability of treatment jumps from zero to one at the cutoff point of the assignment variable. It could be illustrated in the following model:

$$y_i = \alpha_i + x_i \times \beta_i + \epsilon_i$$
$$x_i = \begin{cases} 1 & \text{if } z_i \geq z_0 \\ 0 & \text{if } z_i < z_0, \end{cases}$$

where

y_i is the outcome variable;

x_i is the treatment indicator taking the value of one if individual i is treated and zero otherwise;

ϵ_i is a random error term;

z_i is the assignment variable with a cutoff value of z_0 .

For all $z_i \geq z_0$, the probability of treatment is one. For all $z_i < z_0$, the probability of treatment is zero.

Fuzzy design only requires that $E[x_i | z_i = z] = Pr[x_i = 1 | z_i = z]$ is discontinuous at z_0 , allowing the jump of the probability of treatment to be less than one. x_i is not a deterministic function of z_i anymore. Instead, it is determined by z_i along with other unknown variables.

Following Hahn, Todd & Van der Klaauw (2001), RD design is valid under the following assumptions:

Assumption RD:

- (i) The limits $x^+ \equiv \lim_{z \rightarrow z_0^+} E[x_i | z_i = z]$ and $x^- \equiv \lim_{z \rightarrow z_0^-} E[x_i | z_i = z]$ exist.
- (ii) $x^+ \neq x^-$.

Assumption A1

$E[\alpha_i | z_i = z]$ is continuous in z at z_0 .

Assumption A3

- (i) $(\beta_i, x_i(z))$ is jointly independent of z_i near z_0 .
- (ii) There exists $\epsilon > 0$ such that $x_i(z_0 + e) \geq x_i(z_0 - e)$ for all $0 < e < \epsilon$.

The treatment effect of fuzzy design could be identified as

$$\beta_{fuzzy} = \lim_{e \rightarrow 0^+} E[\beta_i | x_i(z_0 + e) - x_i(z_0 - e)] = \frac{y^+ - y^-}{x^+ - x^-},$$

which identifies the local average treatment effect for whose treatment status changes dis-

continuously at z_0 . A special case is sharp design:

$$\beta_{sharp} = y^+ - y^-$$

Given consistent estimators \hat{y}^+ , \hat{y}^- , \hat{x}^+ and \hat{x}^- , the treatment effect can be consistently estimated by $\frac{\hat{y}^+ - \hat{y}^-}{\hat{x}^+ - \hat{x}^-}$. Local linear regression is a common choice here to overcome the boundary problem of kernel regression. Under certain assumptions, the asymptotic distribution can be derived as:

Theorem 1:

$$n^{\frac{2}{5}}(\hat{\beta}_{fuzzy} - \beta_{fuzzy}) = n^{\frac{2}{5}}\left(\frac{\hat{y}^+ - \hat{y}^-}{\hat{x}^+ - \hat{x}^-} - \frac{y^+ - y^-}{x^+ - x^-}\right) \rightarrow N(\mu_f, \Omega_f)^{42} \quad (\text{A.1})$$

Theorem 1':

$$n^{\frac{2}{5}}(\hat{\beta}_{sharp} - \beta_{sharp}) = n^{\frac{2}{5}}(\hat{y}^+ - \hat{y}^- - (y^+ - y^-)) \rightarrow N(\mu_s, \Omega_s) \quad (\text{A.2})$$

A.2 Weighted sharp design

In sharp design, z_i perfectly predicts x_i . There is no need to collect data on x_i . In fuzzy design, however, individual level of x_i is required to perform 2SLS using z_i to instrument x_i and estimate x^+ and x^- . There are also cases between these two when aggregate level information of x_i is enough for estimation. I call it *weighted sharp regression discontinuity design*. It has the following structure with I_1 as an indicator function of z_i and I_2 as an indicator function of the other treatment determinant d_i :

$$\begin{aligned} y_i &= \alpha_i + x_i \times \beta_i + \epsilon_i \\ x_i &= I_1(z_i \geq z_0) \times I_2(f(d_i, d_0) \geq 0), \end{aligned}$$

where d_0 is a known constant. The treatment effect can be identified as:

$$\begin{aligned} \beta_{weighted} &= \frac{\lim_{z \rightarrow z_0^+} E[y_i | z_i = z] - \lim_{z \rightarrow z_0^-} E[y_i | z_i = z]}{\lim_{z \rightarrow z_0^+} E[x_i | z_i = z] - \lim_{z \rightarrow z_0^-} E[x_i | z_i = z]} \\ &= \frac{y^+ - y^-}{x^+ - x^-} \\ &= \frac{y^+ - y^-}{x^+} \quad (\text{since } x^- = 0) \end{aligned}$$

There are two potential ways to estimate the treatment effect. On the one hand, if information on d_i is available and $f(\cdot)$ is known, sharp design fits the scenario using a subsample with $I_2(\cdot) = 1$. On the other hand, if d_i is unobserved but x_i is known, we can use z_i to instrument x_i and implement 2SLS as in fuzzy design. However, neither of the conventional methods is applicable when both x_i and d_i are unobserved.

Even though the rescaling formula above is similar to fuzzy RD, weighted sharp RD design has a different nature from fuzzy case. In fuzzy RD, x_i is a random variable given z_i . In

⁴²Please refer to Hahn, Todd & Van der Klaauw (2001) for detailed expression.

weighted sharp case, however, the treatment determination process is clear(or “sharp”). The only problem is the observability of d_i . The transparent treatment determination rule allows for the estimation with aggregate data on d_i .

In order to estimate x_i^+ , I rewrite it as

$$\begin{aligned} x_i^+ &= \lim_{z \rightarrow z_0^+} E[x_i | z_i = z] \\ &= \lim_{z \rightarrow z_0^+} \{E[x_i | z_i = z, I_2 = 1] \times Pr[I_2 = 1 | z_i = z] + \\ &\quad E[x_i | z_i = z, I_2 = 0] \times Pr[I_2 = 0 | z_i = z]\} \\ &= \lim_{z \rightarrow z_0^+} Pr[I_2 = 1 | z_i = z]. \end{aligned}$$

The third equality is derived since $E[x_i | z_i = z, I_2 = 1] = 1$ and $E[x_i | z_i = z, I_2 = 0] = 0$. Without individual observations of d_i , I can calculate x^+ using conditional probability of $I_2 = 1$. A simple case is when z_i is independent to d_i for $z_i \geq z_0$. Then

$$x_i^+ = Pr[I_2 = 1].$$

A.3 Estimation of the simple case

$Pr(I_2 = 1)$ can be estimated if another data set is available from the same population with the same sampling rule. Consider the following simplified model:

$$y_i = \alpha_i + x_i \times \beta_i + \epsilon_i \tag{A.3}$$

$$x_i = I_{1i}(z_i \geq z_0) \times I_{2i}(d_i \geq d_0)^{43} \tag{A.4}$$

Assumption 1:

z_i and d_i are independent for $z_i \in [z_0, \infty)$.⁴⁴

Assumption 2:

I_{2i} is i.i.d. with $E[I_{2i}] = \mu_d$ and $Var[I_{2i}] = \sigma_d^2$.

By Central Limit Theorem,

$$\frac{\hat{\mu}_d - \mu_d}{\sigma_d / \sqrt{n}} \rightarrow N(0, 1).$$

Therefore,

$$n^{\frac{1}{2}}(\hat{x}^+ - x^+) = n^{\frac{1}{2}}(\hat{\mu}_d^+ - \mu_d^+) \rightarrow N(0, \sigma_d^2)$$

with a convergence rate of $n^{\frac{1}{2}}$. From Hahn, Todd & Van der Klaauw (2001) Theorem

⁴³In this paper, y_i is the high school enrollment decision for each individual. z_i denotes for birth month with a threshold of z_0 =Sept. 1980, and d_i is the age making high school decision with d_0 =17. x_i is the treatment status which equals to one if individual was born in/after Sept. 1980 AND finished middle school at an age no less than 17.

⁴⁴The context of this study only satisfies this independence assumption locally as discussed in Section 4. Thus, I bootstrap to get standard error in the empirical analysis.

1',

$$n^{\frac{2}{5}}(\hat{y}^+ - \hat{y}^- - (y^+ - y^-)) \rightarrow N(\mu_s, \Omega_s)$$

with convergence rate of $n^{\frac{2}{5}}$. Because $(\hat{y}^+ - \hat{y}^-)$ converges at a slower rate than \hat{x}^+ , it is easy to show that

$$n^{\frac{2}{5}}\left(\frac{\hat{y}^+ - \hat{y}^-}{\hat{x}^+} - \frac{(y^+ - y^-)}{x^+}\right) \rightarrow N\left(\frac{\mu_s}{\mu_d}, \frac{\Omega_s}{\mu_d^2}\right). \quad (\text{A.5})$$

Only sample mean μ_d is required for estimation. This result can be applied to situations where individual level treatment information is missing but aggregate level data (proportion treated) is available either in original data set or in additional data set using same sampling rule.

A.4 Treatment effect

The model setting of weighted sharp design is another form of departure from the classic sharp design in addition to the fuzzy design. A two-step procedure is suggested: first estimate outcome discontinuity at a threshold using the sharp design estimator shown in equation (A.2); then rescale it with the proportion treated (weight). The effect identified could be explained as the average treatment effect on treated (ATT) for the subgroup with $d_i = 1$. The major drawback of this method is that the estimated effect cannot be used to infer the average treatment effect (ATE) for the population. Nevertheless, it is no worse than the local average treatment effect (LATE) obtained in the fuzzy design using 2SLS. Consider the following fuzzy setup with observed x_i and z_i but unknown relationship between them:

$$y_i = \alpha_i + x_i \times \beta_i + \epsilon_i \quad (\text{A.6})$$

$$x_i = f(z_i) + v_i, \quad (\text{A.7})$$

where v_i is a pure random error and all other notations are the same as before. The local linear estimator for fuzzy design is numerically equivalent to IV estimator using z_i as instrument for x_i . The estimated LATE is only applicable to the subgroup whose treatment status changes when z_i moves from below z_0 to above. This is exactly the same effect identified using weighted sharp setup if treatment status in equation (A.7) is truly determined by z_i and d_i as shown in equation (A.4).

In order to compare the results obtained from fuzzy estimator (A.1) and weighted sharp estimator (A.5), I use each method separately to estimate the treatment effect using the same randomly generated data:

- z_i is uniformly distributed in $[-50, 50]$;
- $P[d_i = 1] = 0.5$ if $z_i \geq 0$ and $P[d_i = 1] = 0$ otherwise;
- treatment status $x_i = I_{1i}(z_i \geq 0) \times I_{2i}(d_i = 1)$
- $\alpha = 0.5$ and the underlying real treatment effect β is -0.3 ;
- $y_i = \alpha + x_i \times \beta + \epsilon_i$ with ϵ_i normally distributed with mean 0 and variance 1.

I use data on y_i, z_i , and d_i in weighted sharp estimation and y_i, z_i , and x_i in fuzzy estimation. I draw 1000 observations each time, and repeat the estimation procedure for 5000

times with a bandwidth of 18. As shown in Figure B.1, the two methods generate similar results.

A.5 Generalization

This result could be generalized to situations with more than one additional variables determining treatment status:

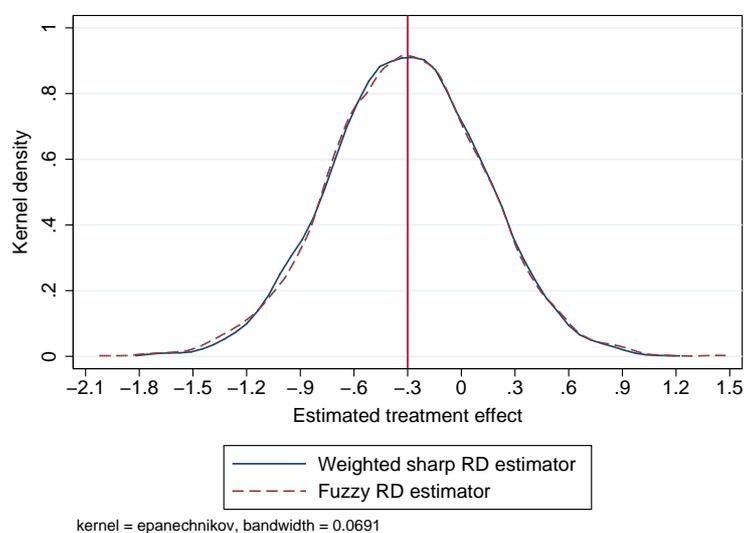
$$x_i = I_{1i}(z_i \geq z_0) \times I_{2i}(f(d_{1i}, d_{2i} \cdots d_{ki}, d_0) \geq 0)$$

where d_{ji} for $j = 1, \dots, k$ could be either discrete or continuous. If d_{ji} s are jointly independent of z_i near z_0^+ , then aggregate level statistics is enough for estimation.

The “quasi-experimental” regression discontinuity design could be applied to more empirical cases, especially for research on developing countries where the observability of individual level data is a prevalent problem.

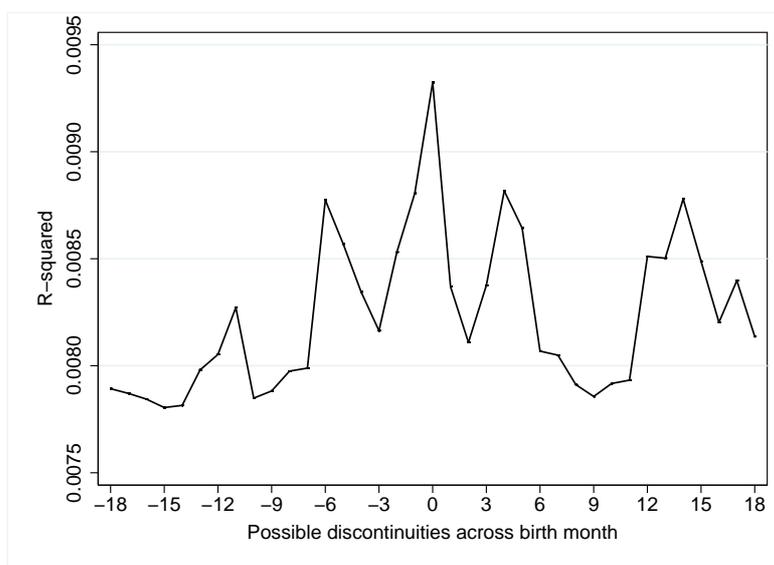
B. ADDITIONAL FIGURES AND TABLES

Figure B.1: Comparison Between the Weighted sharp and Fuzzy estimators



Notes: number of obs=1000; repetition=5000; real effect=-0.3

Figure B.2: Discontinuity Search in High School Attendance Rate



Note: Search for possible discontinuities between March 1979 and March 1981. Sample used here includes individuals born between 43 months before and after Sept. 1980. Birth month is normalized with Sept. 1980=0.

Table B.1: Criteria for Identifying Parent-Child Relationship

| RELATIONSHIP | FATHER IDENTIFICATION | MOTHER IDENTIFICATION |
|----------------|---|--|
| Household head | Relationship: parent Gender: male | Relationship: parent Gender: female |
| Sibling | Relationship: parent Gender: male | Relationship: parent Gender: female |
| Spouse | Relationship: parent in law Gender: male | Relationship: parent in law Gender: female |
| Child | Relationship: HH head/spouse Gender: male | Relationship: HH head/spouse Gender: female |
| Grandchild | Relationship: child/son in law Gender: male If relationship to HH head is child, he/she has to be the only child | Relationship: child/daughter in law Gender: female If relationship to HH head is child, he/she has to be the only child |

Note: “Relationship” refers to relationship to household head. The identification criteria are based on the individual’s relationship to household head listed in the first column. For example, the father of a household head can be identified if one’s relationship to household head is “parent” and gender is “male” in a given household. The mother of the spouse of a household head can be identified if one’s relationship to household head is “parent in law” and gender is “female”.

C. DIFFERENCE IN DIFFERENCE ESTIMATION

C.1 Identification Strategy

According to the Hukou reform in 1998, Children under 18 by September 1998 who followed their mother's Hukou then had a chance to change it according to their father's. Therefore, those born in or after September 1980 with their mother holding rural Hukou and father holding urban Hukou are beneficiaries of this reform. Nevertheless, their high school attendance decisions could only be affected if they were made after the reform if the 1998 Hukou policy change is fully unexpected.

China has a school starting date of September 1. Children reach six years old before September 1 that year are eligible to start nine years of mandatory education, with six years of primary school and three years of middle school. The majority of students finish middle school at an actual age of 15 or older, where the actual age is defined as the age at one's last birthday.⁴⁵ I use the sub-sample born between September 1979 and August 1983 with at least some middle school education in the following analysis. The 1998 Hukou reform affected the high school decisions of most of the Sept.1982-Aug.1983 cohort, part of the Sept.1980-Sept.1981 and Sept.1981-Sept.1982 cohorts, and none of those born between September 1979 and August 1980. The high school attendance rate is expected to drop for people born in later academic years. The effect of policy change could be identified by comparing cohort born in the first academic year in the sub-sample and those born in the last.

However, simple comparison of high school attendance outcomes across birth year cohorts cannot disentangle the effect of obtaining urban Hukou on one's education from the time trend. Chinese government expanded educational investment in 1990s. Private investment increased as well due to the relaxed of school regulations. The dramatic increase in supply side may lead to an increase in educational attainment, disguising the decreased demand. Figure C.3 shows the proportion enrolled in high school for cohorts born in different academic years. It drops at first then rises for the treatment group. The youngest cohort was slightly more likely to enroll in high school than the oldest cohort whose high school decisions were not affected by the new policy.

I use a difference in difference strategy in order to capture unobserved changes that affect schooling decision across time. The control group includes children born between September 1979 and August 1983 with both parents holding urban Hukou. They should not be affected by the 1998 Hukou reform. I also include year dummies and province dummies for additional control. As shown in Figure C.3, the high school enrollment gap between the treatment and control groups increased from around 26 percentage points to 35 percentage points across birth years.

⁴⁵For example, the actual age of a child born in September 1, 1980 is 17 on any date between September 1, 1997 and August 31, 1998. His/her actual age reaches 18 on September 1, 1998.

C.2 Regression Specifications

I estimate the effect of obtaining urban Hukou on education using the following regression model:

$$Y_{itj} = \alpha \times \sum t_i + \beta * \sum (t_i \times I_i(M = rural)) + \gamma \times I_i(M = rural) + \eta \times \sum PROV_i + \delta \times x_{itj} + \epsilon_{itj} \quad (C.8)$$

where Y_{itj} is a binary variable taking value of one if person i born in academic year t with family type of j ($j = 1$ if mother's hukou is rural) attend high school. t_i s are birth year dummies indicating if individual i was born in 1979-1980, 1980-1981, 1981-1982 or 1982-1983 birth year with the oldest cohort as default. All birth years are measured in academic calendar year. For example, people born between Sept.1979 and Aug.1980 are assigned birth year of 1979-1980. I_i is an indication function taking value of 1 if the mother of individual i 's holds rural Hukou. $PROV_i$ is a set of province dummies indicating where the individual i 's Hukou belongs when taking high school entrance exam. X_{itj} is a vector of individual characteristics including gender, minority status, number of siblings, parent's education and occupation of father. ϵ_{itj} is assumed to be exogenous and captures unexplained variations of the dependent variable. The interaction term $\sum (t_i \times I(M = rural))$ for $t_i=1982-1983$ is the variable of interest. Its coefficient captures the effect of the policy change on education of the treatment group.

Hukou province dummies are used to control for different competition level across province since there's admission quota for each province. The census data only report current Hukou location at the time of survey. It might be different from Hukou location at the time of taking the high school entrance exam as a result of transferring Hukou to father's side or changing through enrollment of technical school/college. Since Hukou is inherited from mother for newborn children, mother's Hukou location recorded in 2000 census is used here for approximation. It is also relatively consistent over time.

Parents' educations are used to capture inter-generational education linkage that affects Y_i . In addition, mother's education influenced the variable of interest: mother's Hukou type. Excluding mother's education may cause omitted variable bias. I use dummies for each education level⁴⁶ instead of years of school to better account for possible nonlinear relationship of education across generations.

Lacking of income information in the 2000 census, I use father's occupation variable to proxy financial capability. There are two occupation variables used: dummy variable of employment status and social rank variable. I convert the Chinese census 3-digit occupational code to the International Standard Classification of Occupations (ISCO68), and then convert the ISCO code to International Social-Economic Index (ISEI) following Ganzeboom, De Graaf and Treiman(1992) and Deng and Treiman(1997). The ISEI score scales occupations by the average level of education and average earnings of jobholders from 0 to 100. I use it as proxy for income. Mother's occupation may affect children's educational attainment as well. However, it may be determined by the variable of interest, mother's Hukou type. I exclude mother's occupation to avoid potential biases.

I first estimate equation (C.8) using ordinary least square (OLS) with robust standard error.

⁴⁶They range from no school to college in census as following: illiteracy, primary school, middle school, regular high school, technical high school, junior college, college, graduate and above. I combine the last two categories, as there are few observations in each of them.

The Linear Probability Model is simple to estimate and allows for heteroscedasticity in error terms.

However, the predicted probability in OLS models may exceed the range between zero and one. A nonlinear probability model has the advantage of restricting probability within $[0,1]$. Assuming error terms are normally distributed, I estimate the following equation using Probit regression with latent variable Y_{itj}^* . It is more efficient than OLS given correctly specified error distribution.

$$Y_{itj}^* = \alpha \times \sum t_i + \beta * \sum (t_i \times I_i(M = rural)) + \gamma \times I_i(M = rural) + \eta \times \sum PROV_i + \delta \times x_{itj} + \epsilon_{itj} \quad (C.9)$$

$$Y_{itj} = \begin{cases} 1 & \text{if } Y_{itj}^* \geq 0 \\ 0 & \text{if } Y_{itj}^* < 0 \end{cases}$$

The Probit results may be biased if the error distribution is not normal. I re-estimate the model using semi-parametric Klein and Spady estimator,⁴⁷ which allow for more flexible functional form and does not impose any restriction on the structure of error terms. It achieves root N convergence rate under the single index assumption: the probability that the dependant variable is equal to one is a function only of the index $Z'_{itj}\Gamma$, where Z_{itj} is the vector of all explanatory variables in previous specification and Γ denotes their respective coefficients.

$$Y_{itj} = f(Z'_{itj}\Gamma) + \epsilon_{itj} \quad (C.10)$$

C.3 Estimation Results

The descriptive statistics for major variables used in regression is shown in Table C.2 for treatment group, control group and total sample separately. There are significant differences in demographic characteristics between the treatment and control groups. Even though high school choice for these two groups are different, the effect of the policy change could be correctly estimated as long as the time trend of high school enrollment is the same across groups.

As show in Table C.3 Column 1-4, All three methods discussed above show significant 7 to 10 percentage points drop of high school attendance rate given the opportunity to change Hukou status. It implies that future chance of obtaining urban Hukou amounts for a considerable portion of returns to high school education.

One concern is the county level variation of the primary school length. Primary school takes six years in most counties while some lasts for five years in rural China. Those attending five years of primary school graduate from middle school one year earlier than their counterpart, thus the birth year variable fails to predict whether they are influenced by 1998 policy or not. The 2000 census does not have information regarding the length of primary school each individual attended. However, I could infer the primary school pattern for each county from another nationwide survey: the Household Income Project (CHIP) 2002. It does not directly ask for primary school length but contains information about years of schooling. For a subsample with only middle school degree, the length of primary school is expected

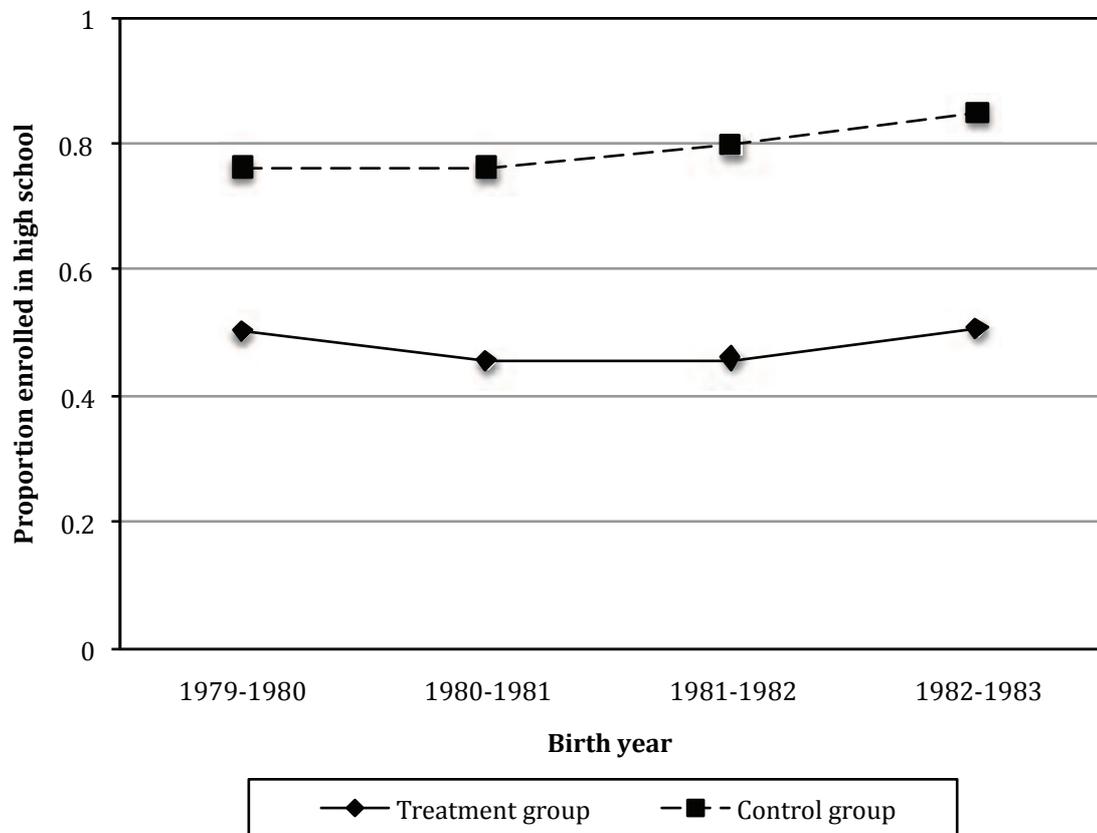
⁴⁷For detailed information, please refer to Klein, R. and R. Spady (1993)

to be five if there are more people obtaining eight years of schooling instead of nine in each county. Since both 2000 census and CHIP2002 share the same area code, I can match counties across datasets and infer the length of primary school for each county in 2000 census. I exclude those attending five years of primary school and re-estimate equation (C.9) using Probit.

Based on this data mapping method, 600 individuals are identified as attending five years of primary school, which accounts for 7% of the total sample of 8478. I re-estimate the Probit model excluding these individuals as a robustness check. The results are presented in Table C.3, Column 5 with the average marginal effects reported in Column 6. There is 9 percentage points drop in the high school attendance rate due to decreased returns to education. Even though only part of the 1981-1982 birth cohort are able to change high school decision according to the new Hukou policy, their high school attendance rate also significantly decreases by 6 percentage points.

C.4 Figures

Figure C.3: High School Enrollment Rate by Birth Years



Notes: Cohorts are grouped by academic years instead of calendar years. Eg. 1979-1980 denotes Sept. 1979 to Aug. 1980.

C.5 Tables

Table C.2: Summary Statistics for 2000 Census

| Variables | TREATMENT | | CONTROL | | FULL SAMPLE | |
|------------------------------|-----------|-----------|---------|-----------|-------------|-----------|
| | Mean | Std. dev. | Mean | Std. dev. | Mean | Std. dev. |
| <i>Father Education</i> | | | | | | |
| Literacy | 0.004 | (0.062) | 0.001 | (0.035) | 0.002 | (0.041) |
| Primary School | 0.204 | (0.403) | 0.107 | (0.309) | 0.122 | (0.327) |
| Middle school | 0.408 | (0.492) | 0.455 | (0.498) | 0.448 | (0.497) |
| Reg. High school | 0.170 | (0.376) | 0.213 | (0.410) | 0.207 | (0.405) |
| Tech. High School | 0.154 | (0.361) | 0.075 | (0.264) | 0.087 | (0.282) |
| Junior College | 0.051 | (0.220) | 0.111 | (0.314) | 0.102 | (0.303) |
| College and Above | 0.003 | (0.056) | 0.032 | (0.175) | 0.027 | (0.163) |
| <i>Mother Education</i> | | | | | | |
| Literacy | 0.027 | (0.162) | 0.008 | (0.090) | 0.011 | (0.105) |
| Primary School | 0.468 | (0.499) | 0.164 | (0.370) | 0.210 | (0.407) |
| Middle school | 0.320 | (0.467) | 0.443 | (0.497) | 0.424 | (0.494) |
| Reg. High school | 0.100 | (0.301) | 0.231 | (0.421) | 0.211 | (0.408) |
| Tech. High School | 0.005 | (0.068) | 0.071 | (0.258) | 0.061 | (0.240) |
| Junior College | 0.001 | (0.028) | 0.048 | (0.214) | 0.041 | (0.198) |
| College and Above | 0.000 | (0.000) | 0.012 | (0.108) | 0.010 | (0.100) |
| Number of Siblings | 1.328 | (0.902) | 0.523 | (0.789) | 0.646 | (0.858) |
| Gender(Male=1) | 0.561 | (0.496) | 0.521 | (0.500) | 0.527 | (0.499) |
| Minority(Yes=1) | 0.086 | (0.281) | 0.074 | (0.261) | 0.075 | (0.264) |
| Father Employment (Yes=1) | 0.845 | (0.362) | 0.826 | (0.379) | 0.829 | (0.377) |
| Father ISEI | 38.392 | (24.361) | 37.689 | (22.188) | 37.796 | (22.534) |
| <i>N</i> | 1295 | | 7183 | | 8478 | |

Note: This is a subsample drawn from 0.95% of 2000 census based on birth month and parents' Hukou type as described in text. I exclude respondents with education level lower than middle school and those still in middle school. The subsample is restricted to individuals with non-missing information of all the variables listed in first column.

Table C.3: Estimation Results of Difference in Difference Approach

| Variables | OLS | PROBIT | | KLEIN & SPADY | PROBIT (6 yrs pri. sch.) | |
|--------------------------------|----------------------|----------------------|--------|---------------|-----------------------------|--------|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| <i>Father Education</i> | | | | | | |
| Literacy | -0.011 (0.123) | -0.047 (0.427) | -0.013 | 0.072 | -0.204 (0.453) | -0.059 |
| Primary School | 0.089 (0.069) | 0.262 (0.198) | 0.064 | -0.044 | 0.222 (0.218) | 0.054 |
| Middle school | 0.155** (0.068) | 0.457** (0.196) | 0.120 | -0.008 | 0.426** (0.215) | 0.110 |
| Regular High school | 0.213*** (0.069) | 0.654*** (0.199) | 0.146 | 0.064 | 0.615*** (0.218) | 0.136 |
| Technical High School | 0.224*** (0.070) | 0.713*** (0.206) | 0.143 | 0.091 | 0.651*** (0.225) | 0.131 |
| Junior College | 0.246*** (0.069) | 1.007*** (0.212) | 0.180 | 0.184 | 0.975*** (0.231) | 0.172 |
| College and Above | 0.223** (0.070) | 1.064*** (0.273) | 0.168 | 0.131 | 0.973*** (0.290) | 0.157 |
| <i>Mother Education</i> | | | | | | |
| Literacy | -0.012 (0.060) | -0.087 (0.159) | -0.024 | -0.086 | -0.138 (0.169) | -0.039 |
| Primary School | 0.070** (0.033) | 0.157* (0.091) | 0.041 | -0.006 | 0.101 (0.094) | 0.026 |
| Middle school | 0.109*** (0.032) | 0.284*** (0.091) | 0.075 | 0.019 | 0.237** (0.095) | 0.062 |
| Regular High school | 0.170*** (0.033) | 0.549*** (0.098) | 0.127 | 0.092 | 0.511*** (0.102) | 0.117 |
| Technical High School | 0.196*** (0.034) | 0.923*** (0.138) | 0.163 | 0.217 | 0.894*** (0.144) | 0.157 |
| Junior College | 0.220*** (0.034) | 1.469*** (0.225) | 0.193 | 0.244 | 1.406*** (0.228) | 0.186 |
| College and Above | 0.194*** (0.038) | 1.120*** (0.363) | 0.167 | 0.048 | 1.043*** (0.368) | 0.159 |
| Number of Siblings | -0.087*** (0.007) | -0.293*** (0.022) | -0.079 | -0.126 | -0.289*** (0.023) | -0.076 |
| Gender(Male=1) | -0.051*** (0.008) | -0.208*** (0.034) | -0.056 | -0.093 | -0.200*** (0.035) | -0.053 |
| Minority(Yes=1) | 0.025 (0.018) | 0.072 (0.068) | 0.019 | 0.041 | 0.093 (0.070) | 0.024 |
| Father Emp.(Yes=1) | -0.058*** (0.020) | -0.203*** (0.068) | -0.051 | -0.082 | -0.195*** (0.071) | -0.048 |
| Father ISEI | 0.002*** (0.000) | 0.009*** (0.001) | 0.002 | 0.003 | 0.009*** (0.001) | 0.002 |
| Treatment(Yes=1) | -0.154*** (0.030) | -0.422*** (0.088) | -0.128 | -0.171 | -0.391*** (0.093) | -0.116 |
| Y80-81*treatment | -0.033 (0.041) | -0.102 (0.120) | -0.029 | -0.030 | -0.136 (0.127) | -0.038 |
| Y81-82*treatment | -0.043 (0.040) | -0.153 (0.115) | -0.044 | -0.057 | -0.215 (0.121) | -0.062 |
| Y82-83*treatment | -0.074* (0.041) | -0.321*** (0.119) | -0.097 | -0.100 | -0.311** (0.126) | -0.093 |
| <i>N</i> | 8478 | 8477 | | 8478 | | 7877 |
| (Pseudo) <i>R</i> ² | 0.204 | 0.200 | | - | | 0.203 |

Notes: Coefficients for birth cohort and Hukou province dummies are not reported. Column 1-4 reports regression results based on a full sample of 8478 while the last two columns excludes individuals attending five year primary school. Coefficients reported in Column 3 and column 6 are calculated marginal effect for corresponding Probit model. Asterisks *, ** and *** denote significant levels of 10%, 5% and 1% respectively.