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

Yao Pan*

October 15, 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 mobility restrictions, combined with selective migration policies, encouraged 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 the aggregate proportion treated is available. Findings suggest that human capital accumulation for rural residents decreased sharply when mobility restrictions were removed in 1998. This effect is bigger for males and for those getting urban identity from relatively rich areas.

JEL Classification: J24, J61, O15.

Keywords: migration restriction, human capital accumulation, regression discontinuity, China

*Department of Economics, The George Washington University (panyao@gwu.edu). I would like to thank Bryan Boulier, Paul Carrillo, Barry Chiswick, Shahe Emran, Robert Phillips, Stephen Smith, Yaohui Zhao and participants at The George Washington University Development Lunch Seminar, The 9th Midwest International Development Conference and The Chinese Economist Society 2012 Annual 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 negatively affect societies (Hamilton and Whalley, 1984). Since the household registration system in China, the Hukou system, strictly restricts rural-urban migrations, it has caught special attention from scholars. Fujita et al. (2004) and Au and 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 and 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 (Liu, 2005; Whalley and Zhang, 2007; Wu and 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 subsidy, pension, 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 find significant institutional returns to education by examining 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 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.

Nonparametric estimates show that the policy change induced a statistically significant drop of 30.5 percentage points in the high school attendance rate. This effect is bigger for males and for those able to get urban Hukou in relatively rich areas. 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. These findings 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 also 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 et al., 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 et al., 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 and Clemens, 2008) or IV estimation (Batista et al., 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 compare it with the benchmark of a closed economy without emigration at all. 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. The impact of labor mobility restrictions on educational attainment is examined in this study.

The rest of this paper is structured as follows: Section 2 provides the background of the Hukou system, related literature and 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 effect, 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 system

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 the 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 markets. The only 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.[1] However, the loosened migration restriction only guaranteed controlled and limited mobility for rural labor while the fates of migrants were still filled with discrimination and unequal treatment. Migrants without local Hukou still could not enjoy urban social benefits. 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

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

admitted students to technical high school, junior college and above are automatically granted urban Hukou.[2] 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 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 and Treiman (2004) show empirically that rural family background harms future educational attainment owing to inferior quality of previous education. They also show that higher education[3] 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)'s pioneering work 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 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

[2] Even 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 today.

[3] They define higher education as technical high school or tertiary education.

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 transfering Hukou status. Even though her article adjusts for non-wage benefits such as food coupon and subsidized housing, brought about by urban status, it does not include other benefits, such as medicare, pensions and urban residence for their children, 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 and 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 and 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 abolishing the rural/urban Hukou dichotomy 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 data scope 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.

The main data source used in this study is the 0.95‰ 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 minorities status, education level, employment status and occupation. The Hukou status of rural/urban and Hukou location at province level are reported as well.

In China, newborn children had to inherit their mother's Hukou status (rural/urban) until September 1998[4], 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.

3 Theoretical Framework

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.[5] 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

[4]The Hukou reform was initially proposed by 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.

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

subsequent effect on their optimal schooling choices.

Following De Brauw and Giles (2008), I assume each household only has one child.[6] 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^c_t. Households' human capital is accumulated by sending the child to school for e_t of his/her time and paying a cost proportional to school time, $P^e_t e_t$, where P^e_t is the unit price of schooling. Therefore, human capital evolves as follows:

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

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 $Y^h_t = \theta_t F(K_t, L^{a1}_t, L^{c1}_t)$, where L^{a1}_t and L^{c1}_t are the amount of time used in household production by adults and children, respectively, and θ_t is an exogenous productivity parameter. Income from labor market can be expressed as $Y^w_t = w(H^a)L^{a2}_t + w(H^c_t, HK)L^{c2}_t$, where L^{a2}_t and L^{c2}_t are shares of time working in labor market for adults and children. $w(H^a)$ and $w(H^c_t, HK)$ are wage functions increasing in human capital stock. The wage function for children, $w(H^c_t, HK)$ is increasing in HK, which equals one if a child gets urban Hukou and zero otherwise. The default status for rural children is $HK = 0$. Households invest all the left over income into physical capital after deducting consumption c_t and school costs from total earning.

[6] This 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.

Thus, physical capital accumulates according to

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

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. 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. $\frac{\partial V_T}{\partial H_T} > 0$.[7] 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, $w(H_t^c, HK)$, 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

[7] Liu (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 an increasing Hukou benefits associated with higher human capital stock.

$l_t^c = 1 - L_t^{c1} - L_t^{c2} - e_t$. Parents' objective is to maximize expected household lifetime utility:

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

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

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

$$U_c(t) = \lambda_t \tag{5}$$

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

$$U_{l^c}(t) = \lambda_t \Big(\theta_t F_{L^{c1}}(t) + w(H_t^c, HK) \Big) \tag{7}$$

$$U_e(t) + \mu_t \psi_t G_e(t) = \lambda_t \Big(\theta_t F_{L^{c1}}(t) + w(H_t^c, HK) - P_t^e \Big), \tag{8}$$

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:

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

Since utilities are additively separable across time, past and future decisions can only influence current decision 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:

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

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 of all, 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, $\mu_T = \frac{\partial(\Phi(K_T,H_T)+q(H_T^c)V(H_T^c))}{\partial H_T} = \frac{\partial \Phi_T}{\partial H_T} + \frac{\partial q_T}{\partial H_T}V(H_T^c) + \frac{\partial V_T}{\partial H_T}q(H_T^c)$. This marginal benefits from education investment includes three parts: the increased households' utility without Hukou transfer, the higher probability of getting urban Hukou and relating 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 *in* schooling period is ambiguous after obtaining urban Hukou and the school enrollment is uncertain.

The second effect operate 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 increased opportunity costs decreases 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 kid obtains local urban Hukou. The reduced school cost, P_t^e, reduces the school costs and is expected to result in higher school enrollment.[8] 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[9]

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 according 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 is 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.

I assume for now 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

[8] This 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.

[9] I also adopt conventional Difference-in-difference (DID) strategy using OLS, Probit and Klein and Spady semi-parametric estimation. An appendix reporting the full results is available from the author upon request.

accordingly. It means birth Let the assignment index z_i represents birth month of individual i. z_i has a cutoff value of $z_0 = Sept.1980$ and has been normalized to $z_0 = 0$.[10] For individuals with an urban father and a rural mother, the eligibility of transferring Hukou status to urban is a deterministic function of individuals' birth month according to the 1998 Hukou reform,

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

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:

$$y_i = \alpha_i + x_i \times \beta_i + \epsilon_i, \tag{12}$$

where α_i is the potential education outcome without the policy change; β_i is the impact of Hukou transfer eligibility on 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 is not informative about the fact that whether an individual untilizes 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 et al., 2001; Porter, 2003). Because identification can be achieved under relatively mild continuity assumptions, RD provides more credible results compared to

[10] 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.

other conventional non-experimental strategies requiring exclusion restrictions, such as Difference-in-difference and Matching.

I adopt nonparametric RD approach for estimation based on Hahn et al. (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 be realistic in the context of this study. I now adopt an assumption that no information is revealed before the announcement of the policy. For illustration purpose, I describe the estimation method using the high school attendance outcome. 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.[11] The second criterion for individuals born just after the threshold is equivalent to finishing middle school at an actual age of 17 or older as explained in Figure 1, where the actual age is defined as the age at one's last birthday.[12] 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 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 high school choice

[11] People 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.

[12] For 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.

according to the new Hukou policy announced in 1998 and 0 otherwise and I_2 is an indicator function taking value of 1 if the actual age of finishing middle school, d_i, is greater than or equal to 17 (d_0). 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 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 literature.

This model structure is different from both conventional sharp and fuzzy design, but fits in a model described in Appendix A. 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 late 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 age enrolled in primary school, the legal length of primary school, and the probability of repeating a grade. [13] 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

[13] The legal length of middle school is three years for all counties.

of birth date.[14] Thirdly, Fertig and Kluve (2005) show that there is no effect of actual age at school entry measured in 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 $z_0 = Sept.1980$ can be simplified as follows:

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

Thus, I can obtain the treatment effect by estimating the raw jump in the high school enrollment rate, as in sharp design, and rescaling it using an estimated proportion of those finishing middle school at 17 or older for the birth cohort of Sept. 1980-Aug. 1981.[15] The only way to estimate this proportion is to exploit the school status information of the 2000 census since the age of finishing middle school is not covered in the questionnaire. Each individual reported the school status of "in school",

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

[15]This rescaling method is derived following Hahn et al. (2001). When the discontinuity 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.

"finished" or "dropout" in addition to highest education level obtained. People born between September 1983 and August 1984 finished middle school at an age of 17 or older (in 2000 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 at 17 or older can be used to infer the proportion of the Sept. 1980-Aug. 1981 cohort making a high school decision after 1998 under the following assumption:

Assumption 2. *The distribution of actual age finishing middle school 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, sharp RD estimates 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 late. 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 decreasing tuition brought by local Hukou could possibly 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, and fewer siblings on average. 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.[16] I then discuss the bandwidth and polynomial 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 rate

The group of interest is those with a father holding urban Hukou and a mother holding rural Hukou. [17]Before proceeding to formal analysis, I plot the proportion of students ever attended high school against month of birth for a subsample born

[16]High school includes technical and regular high schools. I focus on both type of high schools instead of 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 make 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 taken college entrance by the time of the census.

[17]The 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.

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.[18] 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.[19]

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 and Lemieux (2008) and Ludwig and 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 et al. (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.[20] 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

[18] Most of the individuals that 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.

[19] One concern of the threshold is possible delays in nationwide policy implementation as well as timing variations induced 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 and Miller (2007), Card et al. (2008) and Ozier (2010). I use a subsample including individuals born between 50 months before and after September 1980. For each month between September 1979 and September 1981, 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.

[20] See Lee and Lemieux (2010) for detailed discussion of this test. Using dummies for bigger bins generates similar results.

fits best compared to constant, quadratic, cubic and quartic specifications.[21] 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.[22] 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

[21] See Lee and Lemieux (2010) for a discussion of the limitations of the AIC method for polynomial choice in RD design.

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

took place, indicating an age of 17 or older when making their high school decision.[23] I use this estimated percentage to approximate the proportion of the Sept. 1980-Aug. 1981 birth cohort finishing middle school late 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 attandence 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 with is the local average treatment effect. However, this effect may be different for different genders. Table 2 report 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 find a job in China and better utilize the urban Hukou. For population aged 25 and above, the employment rate for male and female are 91% and 77%, respectively. [24] Another

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

[24] Author's calculation according to "China 2000 Population Census Data Assembly", National

possible explanation for male's 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 get. 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 Dropout Rate

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) the lowered middle school graduation rate for those ever attended middle school, and (2) the decreased likelihood to enroll in high school for middle school graduates.

First, the visual evidence in Figure 3 shows a slightly decrease in middle school

Bureau of Statistics of China.

graduation rate at the threshold of September 1980. To formally test to what extent can the change in middle school graduation rate 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 month, 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 compulsive education of primary and middle school in China.

To further test whether the impact of 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, rich region is the only one with statistically significant result 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 points more likely to complete middle school.

5.3 Dropout Rate 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 (17 or older) choosing to enroll in high school regardless of the decrease returns to education. They may be less likely to drop out compared to those bore 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 performs a few 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.[25] 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 chidren's schooling.[26] 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.[27]

Gender has also been recognized as one important determinant of educational

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

[26] Parents' 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.

[27] The 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 childrens probability of attending high school.

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 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.[28] This method excludes adopted children[29], 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 months 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 and 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

[28] This 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.

[29] This 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.

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 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 and Card (2008) study RD design with a discrete assignment variable and argue that comparing outcome 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:

$$Y_i = \alpha + \tau \times D_i + \sum \beta_{lj} \times (z_i - z_0)^j + \sum (\beta_{ri} - \beta_{lj}) \times D_i \times (z_i - z_0)^j + \epsilon_i \qquad j = 1, 2, 3, 4,$$

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 $z_0 = Sept.1980$; 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.[30] 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 could enroll in primary school in the same year. Otherwise, one has to wait for another year to start school. Therefore, children born in September

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

are older comparing 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 and Ferreira (2010) find that students who are the youngest in their school 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 an earlier age 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 et al. (2011) find no effect of school starting age on educational attainment in Norway. Furthermore, Fertig and Kluve (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.[31] 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

[31] The point estimate is 0.025 with a standard error of 0.042.

factors likely influenced individual's 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 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.[32] 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 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 (Psacharopoulos and Patrinos, 2004). This is likely to be

[32] The estimated discontinuity is -0.017 with a standard error of 0.017.

the case in China given its low education subsidy. [33] Liu (2007) is the first to empirically test for a human capital spill-over 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 index[34] of 1.04 to 1.31 in China[35](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 et al. (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.

[33] As 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%.

[34] The index of public subsidization on education shows the ratio of social costs to private costs.

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

As shown in this study, removing the institutional returns to education by directly granting urban Hukou decreases the high school attendance rate substantially by 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 an 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 does 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.[36] 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 individual with a urban father and a rural mother come from an advantaged background as 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

[36] The legal marriage age in China is 22 for male and 20 for female.

to education. The Hukou system creates stronger incentive 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 of fostering education is crucial for China during this transitional period. Because the elimination of the 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. and Henderson, J. (2006a). Are Chinese Cities Too Small? *The Review of Economic Studies*, 73(3):549.

Au, C. and Henderson, J. (2006b). How Migration Restrictions Limit Agglomeration and Productivity in China. *Journal of Development Economics*, 80(2):350–388.

Batista, C., Lacuesta, A., and Vicente, P. (2011). Testing the 'Brain Gain' Hypothesis: Micro Evidence from Cape Verde. *Journal of Development Economics*.

Beine, M., Docquier, F., and Rapoport, H. (2001). Brain Drain and Economic Growth: Theory and Evidence. *Journal of Development Economics*, 64(1):275–289.

Beine, M., Docquier, F., and Rapoport, H. (2008). Brain Drain and Human Capital Formation in Developing Countries: Winners and Losers. *Economic Journal*, 118(528):631–652.

Beine, M., Docquier, F., and Rapoport, H. (2010). On the robustness of brain gain estimates. *Annales d'Economie et de Statistique*, (97-98):143–165.

Black, S. E., Devereux, P. J., and Salvanes, K. G. (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., Mas, A., and Rothstein, J. (2008). Tipping and the Dynamics of Segregation. *The Quarterly Journal of Economics*, 123(1):177–218.

Chan, K. (2009). The Chinese Hukou System at 50. *Eurasian Geography and Economics*, 50(2):197–221.

Chand, S. and Clemens, M. (2008). Skilled Emigration and Skill Creation: A quasi-experiment. Working Papers 152, Center for Global Development.

De Brauw, A. and Giles, J. (2008). Migrant Opportunity and the Educational Attainment of Youth in Rural China. Policy Research Working Paper Series 4526, The World Bank.

Dobkin, C. and Ferreira, F. (2010). Do School Entry Laws Affect Educational Attainment and Labor Market Outcomes? *Economics of Education Review*, 29(1):40–54.

Fertig, M. and Kluve, J. (2005). The Effect of Age at School Entry on Educational Attainment in Germany. IZA Discussion Papers 1507, Institute for the Study of Labor (IZA).

Fleisher, B., Li, H., and Zhao, M. Q. (2010). Human Capital, Economic growth, and Regional Inequality in China. *Journal of Development Economics*, 92(2):215–231.

Fujita, M., Mori, T., Henderson, J., and Kanemoto, Y. (2004). Spatial Distribution of Economic Activities in Japan and China. *Handbook of Regional and Urban Economics*, 4:2911–2977.

Glewwe, P. and Jacoby, H. (2004). Economic Growth and the Demand for Education: Is There a Wealth Effect? *Journal of Development Economics*, 74(1):33–51.

Hahn, J., Todd, P., and Van der Klaauw, W. (2001). Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design. *Econometrica*, 69(1):201–209.

Hamilton, B. and Whalley, J. (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. (2000). Sample Splitting and Threshold Estimation. *Econometrica*, 68(3):575–603.

Heckman, J. J. (2003). China's Investment in Human Capital. *Economic Development and Cultural Change*, 51(4):795–804.

Hossain, S. I. (1997). Making Education in China Equitable and Efficient. Policy Research Working Paper Series 1814, The World Bank.

Imbens, G. and Lemieux, T. (2008). Regression Discontinuity Designs: A Guide to Practice. *Journal of Econometrics*, 142(2):615–635.

Lee, D. (2008). Randomized Experiments from Non-Random Selection in U.S. House Elections. *Journal of Econometrics*, 142(2):675–697.

Lee, D. and Card, D. (2008). Regression Discontinuity Inference with Specification Error. *Journal of Econometrics*, 142(2):655–674.

Lee, D. and Lemieux, T. (2010). Regression Discontinuity Designs in Economics. *Journal of Economic Literature*, 48(2):281–355.

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. (1998). On the Mechanics of Economic Development. *Frontiers of research in economic theory: the Nancy L. Schwartz Memorial Lectures, 1983-1997*, page 61.

Ludwig, J. and Miller, D. L. (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*, pages 5–19.

Psacharopoulos, G. and Patrinos, H. (2004). Returns to Investment in Education: A Further Update. *Education economics*, 12(2):111–134.

Stark, O., Helmenstein, C., and Prskawetz, A. (1997). A Brain Bain with a Brain Drain. *Economics Letters*, 55(2):227–234.

Stark, O., Helmenstein, C., and Prskawetz, A. (1998). Human Capital Depletion, Human Capital Formation, and Migration: A Blessing or a "Curse". *Economics Letters*, 60(3):363–367.

Vidal, J. (1998). The Effect of Emigration on Human Capital Formation. *Journal of Population Economics*, 11(4):589–600.

Weisbrod, B. (1962). Education and Investment in Human Capital. *Journal of Political Economy*, 70(5):106–123.

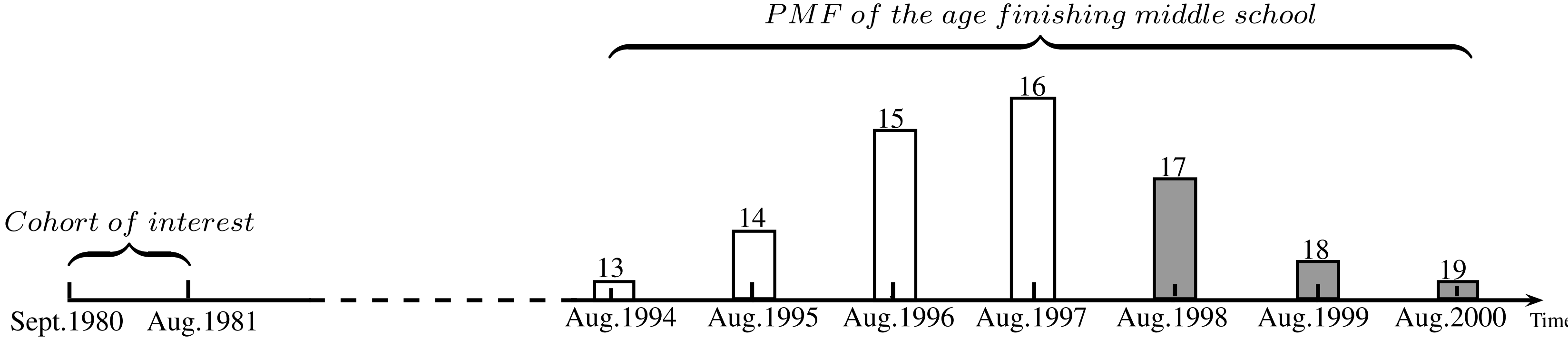
Whalley, J. and Zhang, S. (2007). A Numerical Simulation Analysis of (Hukou) Labour Mobility Restrictions in China. *Journal of Development Economics*, 83(2):392–410.

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 Treiman, D. (2004). The Household Registration system and Social Stratification in China: 1955–1996. *Demography*, 41(2):363–384.

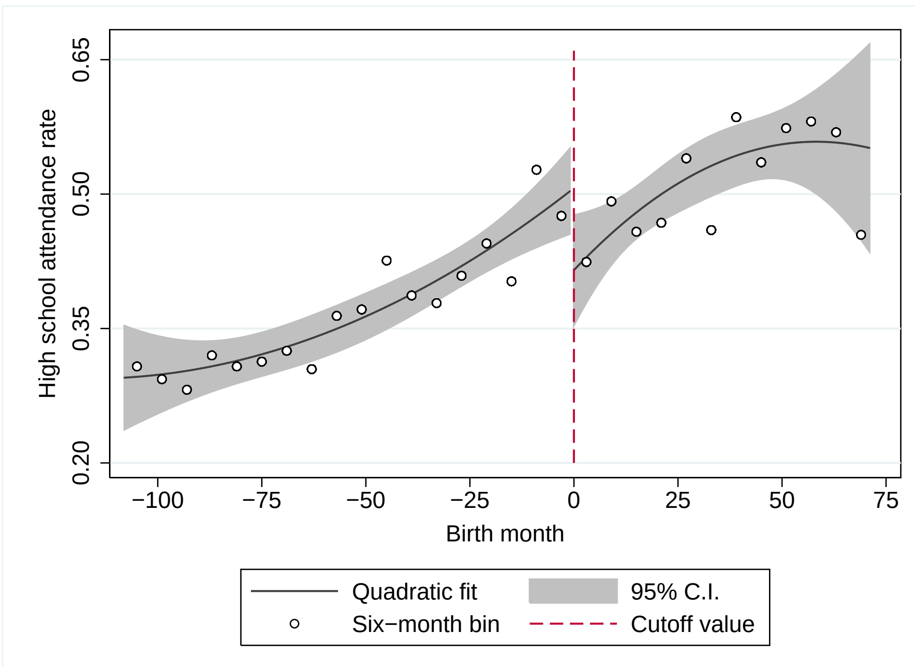
Zhao, Y. (1997). Labor Migration and Returns to Rural Education in China. *American Journal of Agricultural Economics*, 79(4):1278–1287.

Figure 1: Treatment Determination Criteria



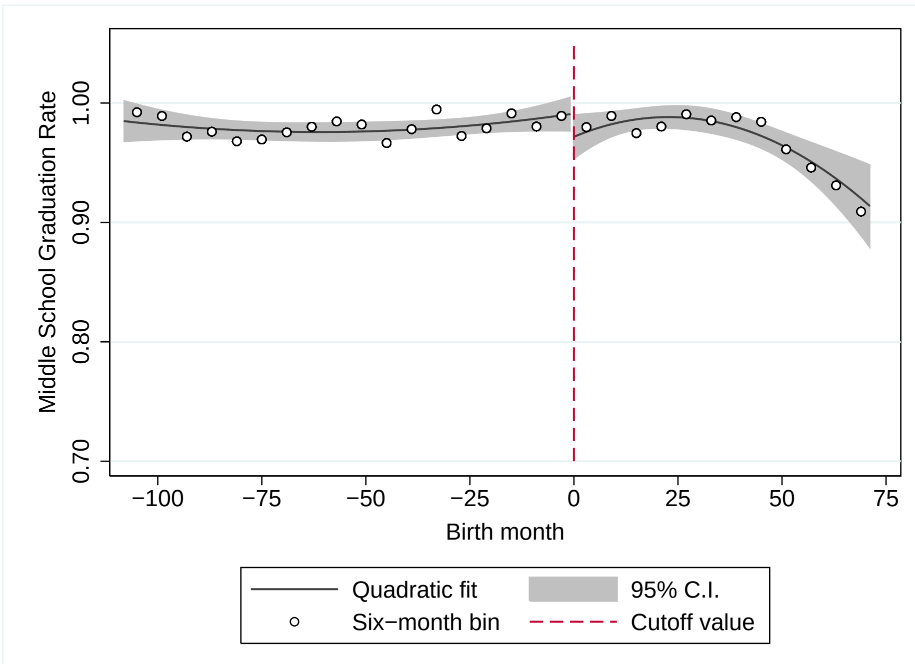
Notes: According to the 1998 Hukou reform, individuals born in or after September 1980 inherited their mother's Hukou would be eligible to change Hukou status according to their father's. Thus, children born in or after September 1980 with a mother holding rural Hukou and a father holding urban Hukou 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). For cohort born just after the threshold of September 1980, criterion 2 means that the actual age finishing middle school is equal to or greater than 17 as indicated in the shaded areas in this graph.

Figure 2: Discontinuity in High School Attendance Rate



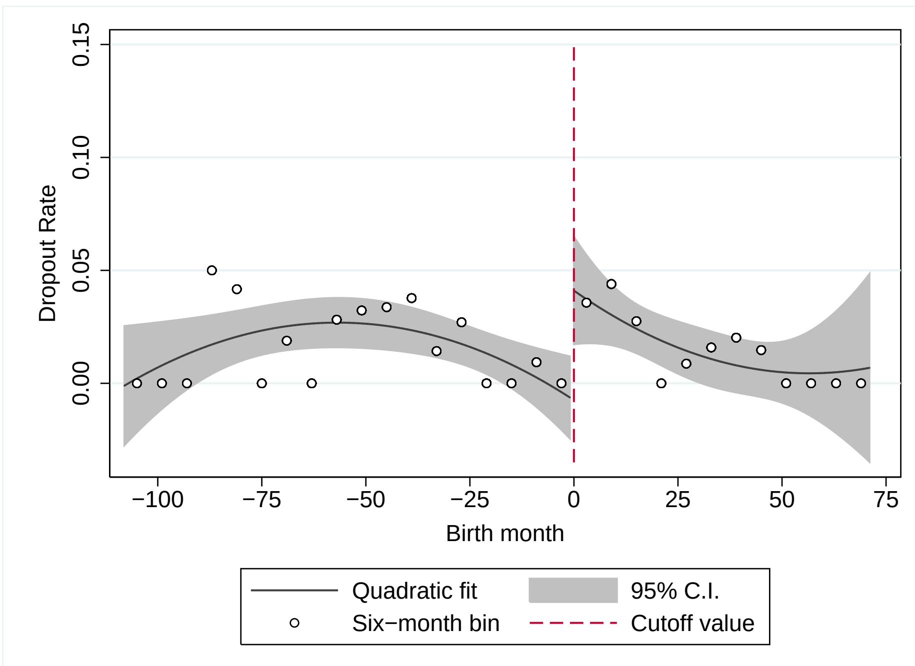
Notes: High school enrollment rate for children with a father holding urban Hukou and a mother holding rural Hukou. Sample used here includes individuals born between September 1971 and August 1986 with at least some middle school education. Those still in middle school are excluded. Birth month is normalized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

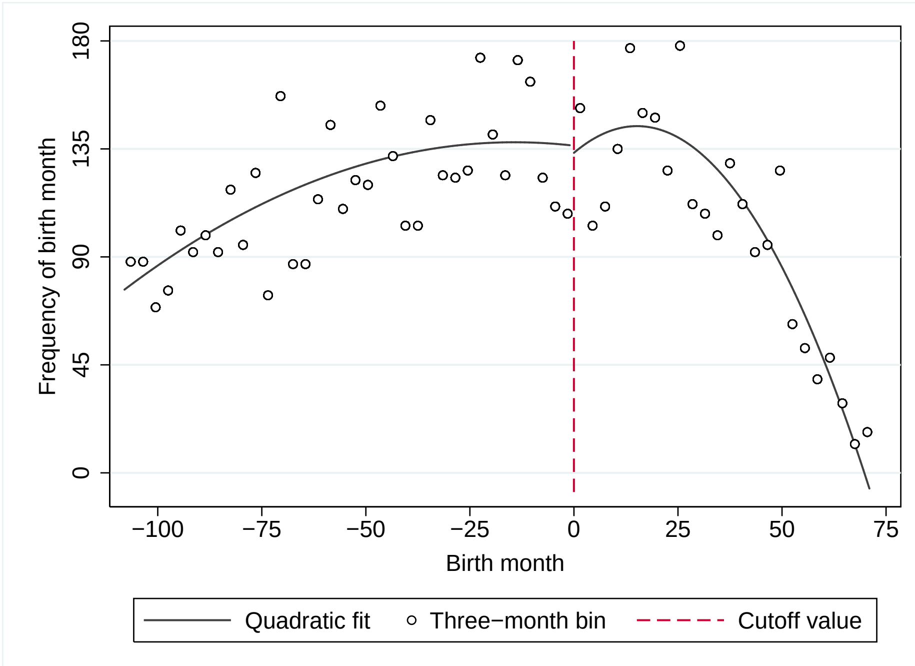
Figure 4: Discontinuity in Dropout Rate of High School or Above



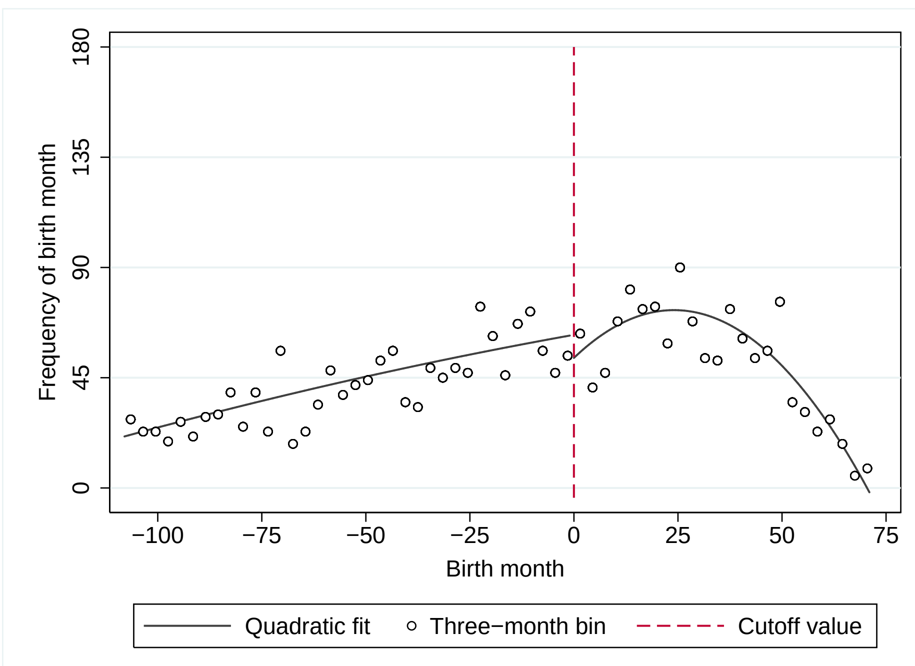
Notes: Dropout rate for children with a father holding urban Hukou and a mother holding rural Hukou. Sample used here includes individuals born between September 1971 and August 1986 with at least some high school education. Birth month is normalized with $Sept.1980 = 0$. The estimated discontinuity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

Figure 5: Density Continuity for Birth Month with Quadratic Fit

(a) Sample to Estimate High School Attendance

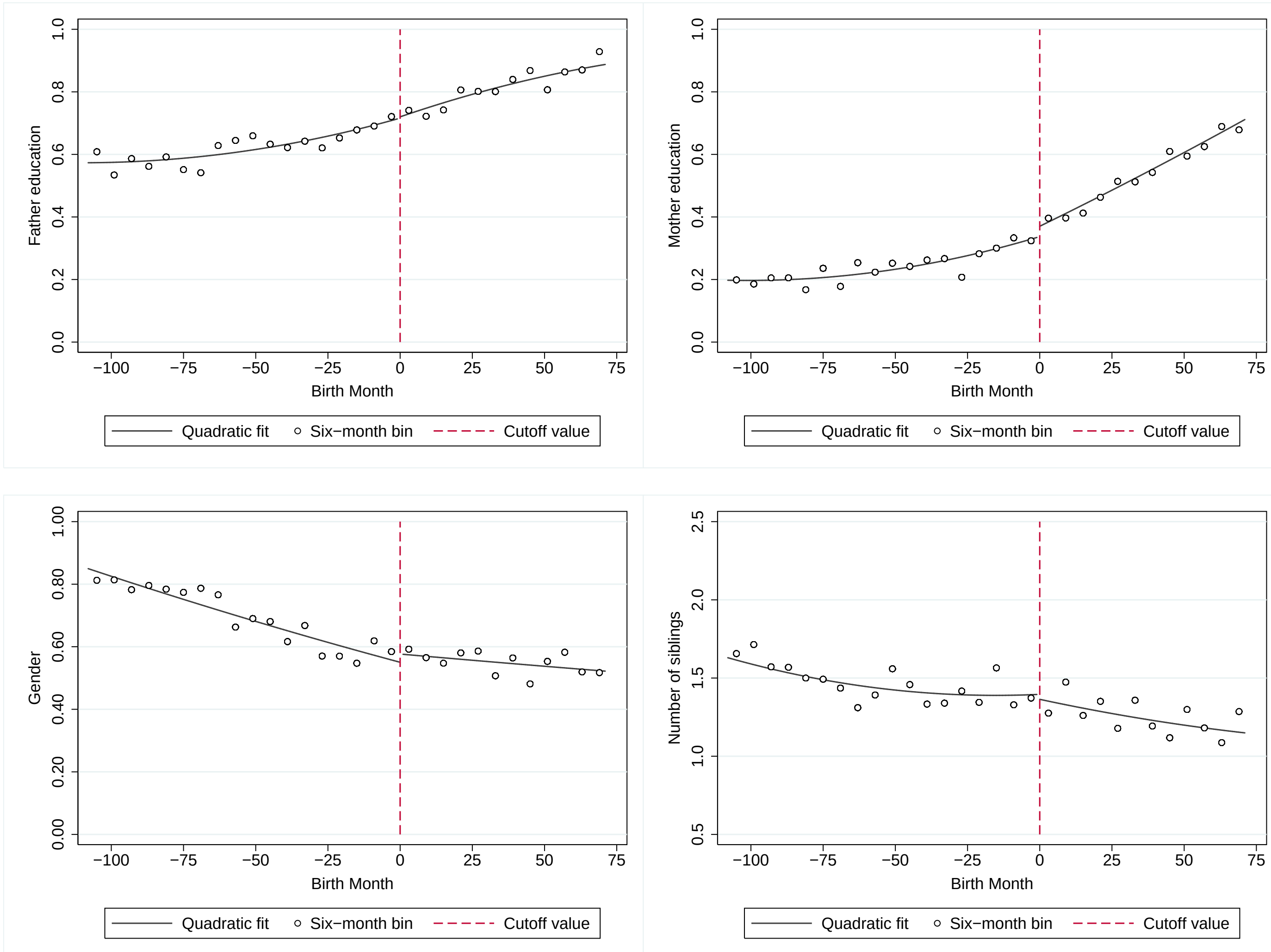


(b) Sample to Estimate Dropout Rate of High School or Above



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
Panel A: Characteristics of the full sample			
Father education (middle school and above)	0.69	(0.46)	6324
Mother education (middle school and above)	0.34	(0.47)	6476
Gender (male=1)	0.64	(0.48)	6597
Number of siblings	1.35	(0.96)	4364
High school attendance rate	0.41	0.49	6597
Dropout of High school or above	0.02	0.13	2709
Panel B: High school attendance and covariates in subsample used for estimation			
Father education (middle school and above)	0.72	(0.45)	1585
Mother education(middle school and above)	0.36	(0.48)	1607
Gender (male=1)	0.57	(0.49)	1629
Number of siblings	1.37	(0.95)	1309
High school attendance rate	0.44	(0.50)	1629
Panel C: Dropout rate and covariates in subsample used for estimation			
Father education (middle school and above)	0.78	(0.41)	700
Mother education(middle school and above)	0.42	(0.49)	710
Gender (male=1)	0.55	(0.50)	715
Number of siblings	1.21	(0.87)	604
Dropout of high school or above	0.02	(0.13)	715

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 also include: (1) a child holding rural Hukou with a father holding urban, or (2) a child holding urban Hukou that is not changed through education with a mother holding rural. I exclude respondents with education level lower than middle school and those still in middle school for the first five variables. Panel B and C restrict the samples for enrollment and dropout estimations to individuals born between March 1979 and February 1982 (e.g. 18 months before and after the threshold of September 1980), respectively. For the dropout variable, I further restrict 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**	-0.0.072*	-0.086**
	(0.042)	(0.045)	(0.041)
By Gender			
Male	-0.113**	-0.113*	-0.101*
	(0.054)	(0.060)	(0.053)
Female	-0.058	-0.022	-0.069
	(0.065)	(0.068)	(0.064)
By Region			
Rich	-0.124**	-0.151**	-0.140**
	(0.058)	(0.061)	(0.056)
Poor	-0.049	-0.004	-0.034
	(0.059)	(0.064)	(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	-0.087**	-0.113**	-0.058	-0.124**	-0.049
High School Enrollment	(0.042)	(0.054)	(0.065)	(0.058)	(0.059)
Proportion Treated	0.285	0.298	0.269	0.272	0.298
	-0.305**	-0.379**	-0.216	-0.456**	0.164
LATE	(0.147)	(0.181)	(0.242)	(0.213)	(0.198)
	[0.143]	[0.186]	[0.243]	[0.226]	[0.206]
N	2762	1576	1186	1421	1341

Note: 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	0.009	0.008
	(0.010)	(0.011)	(0.011)
By Gender			
Male	0.010	0.011	0.010
	(0.015)	(0.017)	(0.015)
Female	0.007	0.009	0.009
	(0.009)	(0.010)	(0.009)
By Region			
Rich	0.025**	0.021*	0.025**
	(0.012)	(0.012)	(0.012)
Poor	-0.010	-0.010	
	(0.016)	(0.017)	(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	0.008	0.010	0.007	0.025**	-0.010
High School Enrollment	(0.010)	(0.015)	(0.009)	(0.012)	(0.016)
Proportion Treated	0.285	0.298	0.269	0.272	0.298
	0.028	0.034	0.026	0.092**	-0.034
LATE	(0.035)	(0.050)	(0.033)	(0.044)	(0.054)
	[0.035]	[0.052]	[0.044]	[0.045]	[0.052]
N	2762	1576	1186	1421	1341

Note: 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 or Above

	ESTIMATION		
Sample	(1)	(2)	(3)
All	0.044**	0.037**	0.046***
	(0.018)	(0.018)	(0.018)
By Gender			
Male	0.046*	0.046*	0.047*
	(0.024)	(0.027)	(0.024)
Female	0.042*	0.027	0.043*
	(0.026)	(0.021)	(0.026)
By Region			
Rich	0.050**	0.033	0.051**
	(0.025)	(0.021)	(0.026)
Poor	0.037	0.033	0.034*
	(0.024)	(0.021)	(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: Sample 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

					Parametric			
					Quadratic		Quartic	
Sample	Nonparametric				No Cont.	Controls	No Cont.	Controls
Bandwidth (months)	20	30	40	50	-	-	-	-
All	-0.114*	-0.097**	-0.089**	-0.077**	0.092**	0.067**	-0.127**	-0.133**
	(0.061)	(0.049)	(0.043)	(0.038)	(0.037)	(0.033)	(0.054)	(0.046)
By Gender								
Male	-0.162**	-0.140**	-0.120**	-0.092*	-0.116**	-0.076	-0.163**	-0.217***
	(0.079)	(0.064)	(0.056)	(0.050)	(0.049)	(0.051)	(0.074)	(0.080)
Female	-0.052	-0.046	-0.053	-0.062	-0.060	-0.054	-0.077	-0.026
	(0.094)	(0.077)	(0.067)	(0.060)	(0.066)	(0.060)	(0.101)	(0.090)
By Region								
Rich	-0.164*	-0.144**	-0.128**	-0.112**	-0.115**	-0.141***	-0.129*	-0.206***
	(0.085)	(0.069)	(0.060)	(0.053)	(0.054)	(0.053)	(0.076)	(0.073)
Poor	-0.049	-0.049	-0.035	-0.040	-0.061	0.013	-0.127*	-0.083
	(0.086)	(0.070)	(0.061)	(0.055)	(0.053)	(0.057)	(0.076)	(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

Sample	Nonparametric				Parametric			
					Quadratic		Quartic	
					No Cont.	Controls	No Cont.	Controls
Bandwidth (months)	30	40	50	60	-	-	-	-
All	0.007	0.008	0.008	0.006	0.019*	0.020*	0.006	0.004
	(0.013)	(0.011)	(0.010)	(0.009)	(0.011)	(0.012)	(0.017)	(0.016)
By Gender								
Male	0.011	0.010	0.010	0.009	0.027*	0.027*	0.012	0.002
	(0.022)	(0.018)	(0.016)	(0.014)	(0.015)	(0.015)	(0.025)	(0.025)
Female	0.004	0.007	0.004	0.011	0.011	0.011	0.001	0.008
	(0.009)	(0.009)	(0.009)	(0.009)	(0.012)	(0.013)	(0.012)	(0.014)
By Region								
Rich	0.022	0.023*	0.025**	0.023**	0.034*	0.023	0.022	0.018
	(0.016)	(0.014)	(0.012)	(0.012)	(0.018)	(0.016)	(0.032)	(0.024)
Poor	-0.009	-0.008	-0.009	-0.011	0.003	0.014	-0.011	-0.012
	(0.021)	(0.018)	(0.016)	(0.014)	(0.017)	(0.020)	(0.025)	(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 Design 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 $E[x_i \mid z_i = z] = Pr[x_i = 1 \mid z_i = z]$ is discoutinuous 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 et al. (2001), RD design is valid under the following assumptions:

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

Assumption A1
$E[\alpha_i \mid z_i = z]$ 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 exits $\epsilon > 0$ 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

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

which identifies the local average treatment effect for whose treatment status changes discontinuously 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:

$$n^{\frac{2}{5}}(\hat{\beta}_{fuzzy} - \beta_{fuzzy}) = n^{\frac{2}{5}}\left(\frac{\hat{y}^+ - \hat{y}^-}{\hat{x}^+ - \hat{x}^-} - \frac{y^+ - y^-}{x^+ - x^-}\right) \to N(\mu_f,\ \Omega_f)^{37} \tag{A.1}$$

Theorem 1':

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

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:

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

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

[37] Please refer to Hahn et al. (2001) for detailed expression.

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 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

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

The third equality is derived since $E[x_i \mid z_i = z, I_2 = 1] = 1$ and $E[x_i \mid z_i = z, I_2 = 0] = 0$. 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:

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

$$x_i = I_{1i}(z_i \geq z_0) \times I_{2i}(d_i \geq d_0)^{38} \tag{A.4}$$

Assumption 1:
z_i and d_i are independent for $z_i \in [z_0, \infty)$.[39]

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

[38] In 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 finish middle school at an age no less than 17.

[39] The 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.

By Central Limit Theorem,

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

Therefore,

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

with a convergence rate of $n^{\frac{1}{2}}$. From Hahn et al. (2001) Theorem 1',

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

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

$$n^{\frac{2}{5}}(\frac{\hat{y}^+ - \hat{y}^-}{\hat{x}^+} - \frac{(y^+ - y^-)}{x^+}) \to N(\frac{\mu_s}{\mu_d},\ \frac{\Omega_s}{\mu_d^2})\ . \tag{A.5}$$

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:

$$y_i = \alpha_i + x_i \times \beta_i + \epsilon_i \tag{A.6}$$
$$x_i = f(z_i) + \nu_i, \tag{A.7}$$

where ν_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:

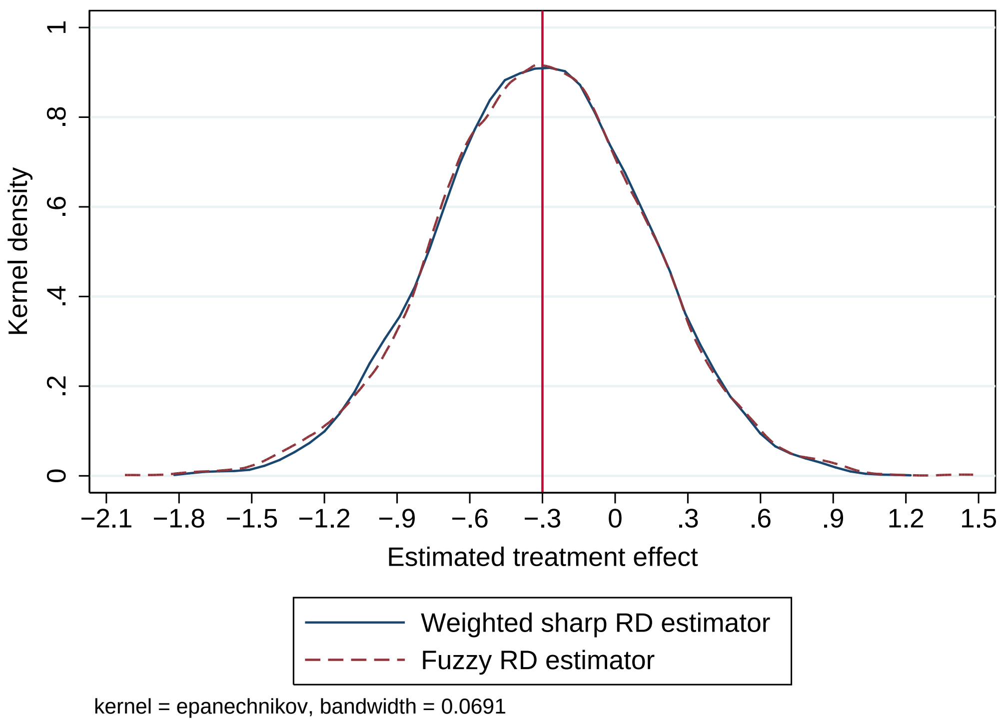
$$x_i = I_{1i}(z_i \geq z_0) \times I_{2i}(f(d_{1i}, d_{2i} \cdots d_{ki}, d_0) \geq 0)$$

where d_{ji} for $j = 1, \cdots, k$ 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 prevailed problem.

B Additional Figures and Tables

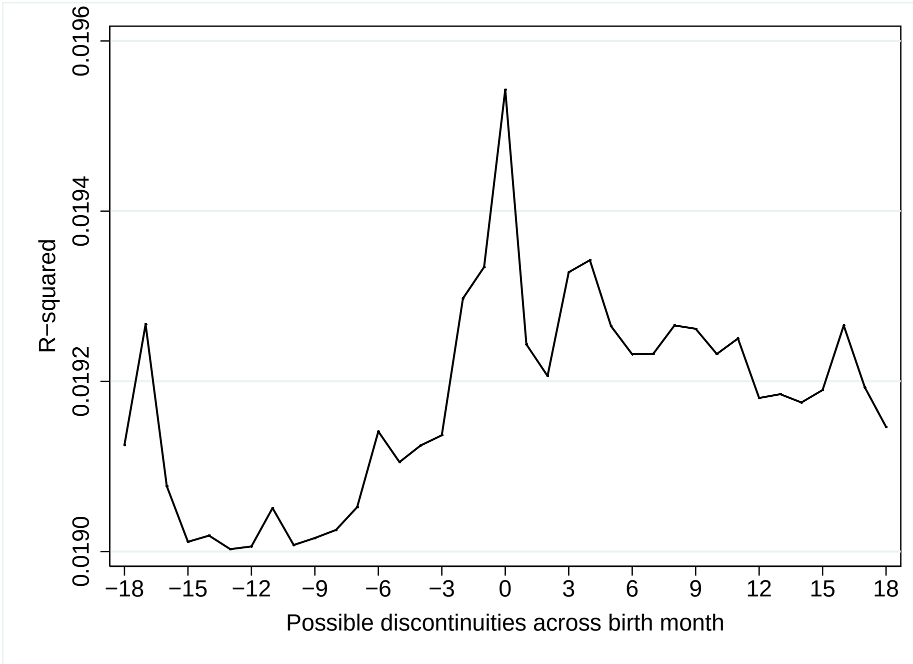
Figure B.1: Comparison Between the Weighted sharp and Fuzzy estimators



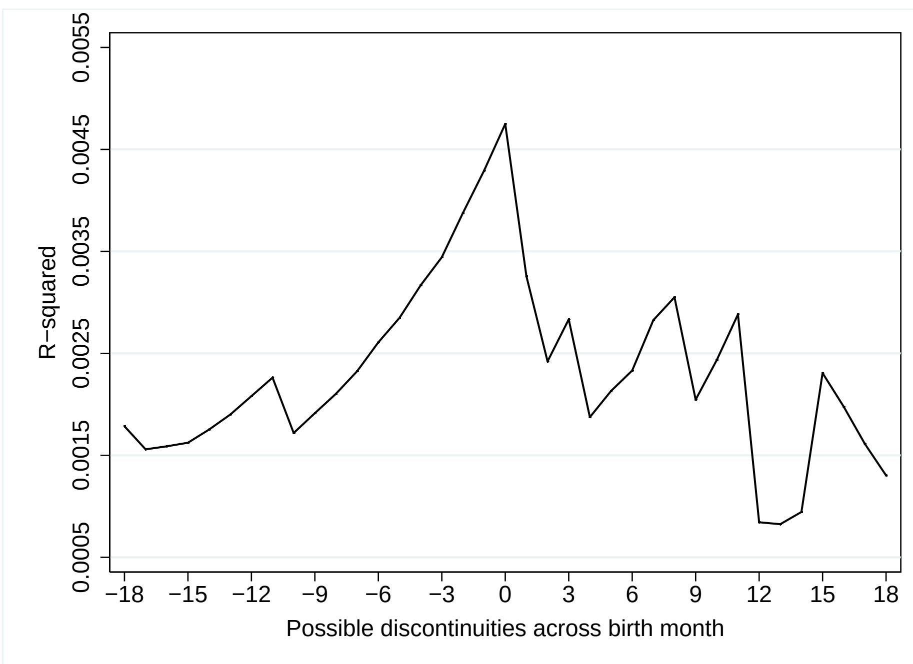
Notes: number of obs=1000; repetition=5000; real effect=-0.3

Figure B.2: Discontinuity Search Result

(a) High School Enrollment Rate



(b) Dropout Rate of High School or Above



Note: Search for possible discontinuities between March 1979 and March 1981. Sample used here includes individuals born between 50 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 indentification 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".