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

Yao Pan*
Department of Economics
The George Washington University

February 28, 2012

**Abstract**

In this paper, I study the impact of the Hukou labor mobility restrictions on human capital investment in China. Rural people have stronger incentive to pursue higher education, treating it as means to obtain urban identity and escape from underdeveloped areas. The 1998 Hukou policy reform granted urban Hukou to a specific group automaticly. Using a Regression Discontinuity strategy to fully exploit this policy change, I find the human capital decreased sharply, as measured by high school enrollment and dropout rates, when the mobility restriction is removed. This finding indicates considerable indirect returns to education stemming from the Hukou system in the presence of education-based selective migration, over and above the usual direct pecuniary returns. These results reveal the important role of the Hukou system on encouraging educational investment in China.

**JEL Classification:** J24, J61, O15.
**Keywords:** migration restriction, human capital accumulation, regression discontinuity, China

*I would like to thank Bryan Boulier, Paul Carrillo, Shahe Emran, Robert Phillips, Stephen Smith and Yaohui Zhao and participants at the George Washington University development lunch seminar for helpful comments and suggestions.

# 1 Introduction

There are clear economic reasons to anticipate that restrictions on labor mobility will reduce social welfare.

In the absence of free migration across countries, workers' potential cannot be fully exploited as a result of misallocation, generating large productivity losses. Hamilton and Whalley (1984) estimate welfare loss induced by global labor mobility restrictions and find enormous efficiency gains when immigration barriers are removed. Later studies confirm this result using different techniques (Iregui, 2003; Moses and Letnes, 2004; Walmsley and Winters, 2005).

Within-country labor mobility restrictions affect national welfare in a similar way. Fujita et al. (2004), Au and Henderson (2006a), and Au and Henderson (2006b) argue that most Chinese cities are significantly undersized and ascribe this under-urbanization to the Hukou system.[1] There are also large income losses resulting from insufficient agglomeration in rural industries (Au and Henderson, 2006b). Migration restrictions have been criticized as one major cause of rising rural-urban income inequality in China in the literature as well (Liu, 2005; Whalley and Zhang, 2007). In addition to limiting urbanization and exaggerating income disparities, the household registration system in China may generate unfair opportunities within nation and hinder the freedom of migration as pointed out by many scholars (Wu and Treiman, 2004; Chan, 2009).

Regardless of all these negative impacts, one may wonder: can there be any social benefits associated with a system that restricts migration such as Hukou system?[2]

[1] Household registration system in China, Hukou system, strictly regulates both rural-urban and intra-sector migration.

[2] This would be separate from any arguments concerning policy to compensate for excessive migration resulting from Harris-Todaro type effects.

This paper tries to answer this question by analyzing the impact of labor mobility restrictions on educational attainment. To my best knowledge, this is the first attempt to investigate possible positive impacts of household registration system 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 the case in China given its low education subsidy[3]. 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 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 low subsidization index[4] of 1.04 to 1.31 in China[5](Hossain, 1997), implies that the social return is higher than the private return to education. In addition to pecuniary returns included in the above calculations, an additional part of externality from education stems from other social impacts such as crime reduction. Moretti and Lochner (2004) find high school graduation reduces crime participation with a cost saving amounting to 14% to 26% of private returns in the United States. Groot and van den Brink (2010) find similar results for the Netherlands. Social returns will be even higher when these non-pecuniary benefits are included, resulting in a underinvestment in education. This

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

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

[5]estimates the private and social rate of return to education by levels of education in China. The subsidization indices estimated by Hossain (1997) are 1.25, 1.04 and 1.31 for primary, secondary and higher education respectively.

perception is in line with Heckman (2003)'s argument that China invested too little in human capital compared to its investment in physical capital.

The Hukou system may play a significant role of fostering education in rural area in the presence of education-based selective migration and partially correct the under-investment problem. People are categorized as rural or urban at birth according to parents' type despite where they live in China. Urban residents could enjoy a series of benefits provided by government while rural individuals make living using their own land. Even though rural people are allowed to migrate to urban areas to seek for better employment opportunities, migrants without local Hukou cannot enjoy social benefits such as medicare, unemployment insurance, housing subsidy, pension, etc.

However, people enrolled in technical high school or college are automatically granted urban Hukou. Thus, individuals with rural origin benefit more from higher education compared to their urban counterparts, causing higher incentive to invest in human capital. The private returns to education include option value of increase in the probability of being able to receive urban benefits. Individuals trapped in underdeveloped areas are expected to hold stronger incentive to invest in education, treating it as a tool to escape from poverty by becoming able to migrate. I define the extra-market gains as "institutional returns" since they stem from overcoming mobility restrictions, which are associated with certain labor mobility policies.

Empirically I demonstrate significant institutional returns to education by exploiting 1998 Hukou policy change. People had inherit their mother's Hukou type until 1998, when inheriting from their father was made possible. In addition, children under 18 years old now have the chance to transfer their Hukou to father's type, which differentially benefits the group with rural mother and urban father. They can obtain urban

Hukou and associated benefits without higher education. I apply the Regression Discontinuity approach in this study to estimate the change in high school enrollment rate for this group when institutional returns to education are removed. The nonparametric estimation results show a statistically significant drop of 35 percentage points in the high school enrollment and an increase of 4 percentage points in the dropout rate for those already in high school or above, demonstrating the existence of institutional returns to education in addition to pecuniary returns in rural China. These findings are confirmed by a series of robustness checks such as continuity assumptions required for valid Regression Discontinuity, parametric estimation and possible age induced biases.

This research contributes to a main strand in the migration literature: "brain drain" vs "brain gain". Contrary to conventional view of detrimental brain drain, a few theoretical studies in late 1990s point out the possibility of "brain gain" predicted by models with endogenous education choice 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 expected returns to schooling, resulting in a higher human capital stock at home, netting out those emigrated. Most empirical evidences are derived from cross country regressions due to data constraints (Beine et al., 2001, 2008, 2009). While their results confirm the positive effect of education based selective migration on human capital formation in less developed source countries, the identification and causal inference are always problems associated with macro studies (Docquier and Rapoport, 2011). A few micro-level studies confirming the "brain gain" effect have emerged in recent years. In order to overcome potential endogeneity problem and claim causal effect, researchers often adopt Difference-in-difference strategy (Chand and Clemens, 2008) or IV estimation (Batista et al., 2011). However, the validity of these methods requires a set of untestable exclusion restrictions. My study overcomes these limitations by exploiting an unexpected policy change that erases migration restrictions. The causal effect is identified using quasi-experimental Regression Discontinuity analysis which requires relatively mild continuity assumptions. According to my results, labor mobility restrictions combined with education-based selective migration can increase education level compared with open economy when both educated and uneducated labors are free to migrate. Previous literature analyzes human capital formation in a situation when only educated workers have positive probability of emigration, and compare it with the benchmark closed economy without emigration at all. The results could be explained as the impact of education-based selective migration on source country human capital investment. This paper, however, is the first to use free migration as the benchmark. I investigate human capital dynamics in a situation with migration restrictions and education-based selective migration, and compare it with the case when all workers are free to move regardless of education level. The impact of labor mobility restrictions on educational attainment is identified in my 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 model of human capital accumulation. 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 concludes with policy implications.

# 2 Background

## 2.1 Household registration system

China's household registration system (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, combining with the food rationing policy, effectively tied people to their registered residency place. The strictly controlled rural-urban migration played an important role of maintaining low agricultural product price and fostering rapid industrialization.

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

Regardless of the fluctuations caused by great leap forward and Cultural Revolution, the Hukou system was strictly enforced for the next few decades. Rural people rely on their land to support themselves while the government provides housing, food, pension, etc. to urban residents. The gap between these two worlds within same country widened in terms of income, political representation and employment opportunities. It was almost impossible to migrate without legally changed Hukou because of rigid food redistribution and absent of commodity markets. The only few ways of changing ones Hukou type from rural to urban are mainly through assignment after military service, enrolling in technical high school/tertiary education or employment by the states, all of which are subject to small quota and exceedingly difficult for peasants.

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 grants them legal residency for a few months and is subject to renewal. The abolishment of government subsidized food rations during the late 1980s, along with newly evolved commodity market enables abundant rural labor to seek for employment opportunity in cities, boosting the number of floating population (temporary migrants) to 147.35 million.[6] However, the loosened migration restriction only guarantees controlled and limited mobility for rural labor while the fates of peasants are still filled with discrimination and unequal treatment. Migrants without local Hukou cannot enjoy social benefits such as medicare, unemployment insurance, housing subsidies, pensions, etc. Even though primary and secondary education are possible, subject to availability by paying additional fees, their children have to go back to their legal places of residence to take college entrance examination, which usually use different textbooks and have different test questions than where they go to school.

As mentioned above, higher education is one of the effective ways of obtaining urban Hukou. After finishing nine years compulsory schooling of primary and junior middle school, students could either attend high school or work directly. There are two types of high school: regular and technical. Even though obtaining regular high school degree does not guarantee urban Hukou, it provides the opportunity for tertiary education. Newly admitted students to technical school, junior college and above

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

are automatically granted urban hukou.[7] For rural students, the returns to high school education include not only the well-studied higher future income brought by schooling, but also potential dramatic benefits associated with Hukou type change. The linkage between schooling and labor mobility creates stronger incentive to pursue high school education for rural residents than other countries, or even urban people in China.

## 2.2 Literature review

There are a few scholars trying to depict the relationship between Hukou type and educational attainment since 1990s. Using CHIP 2002 dataset, Chen et al. (2010) find the education expenses for rural migrants living in cities are lower than native urban residents, which is the biggest gap among all consumption categories. They ascribe this result to limited education chance and credit constraint facing by rural migrants.

Wu and Treiman (2004) empirically prove that higher education[8] significantly increases the odds of obtaining urban Hukou by more than four times for those from rural origin based on the survey of Life Histories and Social Change in Contemporary China. In the mean time, rural family background harm future educational attainment due to inferior quality of previous education.

All these literature conclude rural origin is one of the major determinants of less schooling, while schooling, especially college degree, will help transfer Hukou from rural to urban. However, none of them treat education as endogenous choice variable.

[7] Even though the Hukou transfer is voluntary, most students accept urban Hukou given enormous benefits associated with it. Deny of urban Hukou is a rare case, especially in last century when Hukou played even more significant role than today.

[8] They define higher education as specialized secondary or tertiary education.

The extra return for rural people could not be identified by simple comparison of educational attainment between rural and urban people.

Zhao (1997) is the pioneer work to first incorporate the schooling choice into the calculation of expected future income and demonstrate the incentive for pursuing senior higher school education is partly rooted from the chance of changing Hukou category. She finds a decline in the senior middle school enrollment rate and interpret this change as due to increasing opportunity cost of working in rural enterprises using data from three villages near the city of Beijing. She points out that the conventional methods underestimate the returns to education in China by ignoring the possible wage and non-wage gain associated with urban Hukou type. Even though her article is able to to adjust for non-wage benefits brought about by urban status by incorporating food coupons and subsidized housing, it does not include other benefits such as medicare, pension and gains transferrable to future generations due to data limitations. The resulting estimate still underestimates the real returns to schooling for rural people since gains from Hukou type change is not fully taken in to account.

Using broader survey covering four provinces, De Brauw and Giles (2008) show in the increase of labor mobility induced by lower migration cost due to allowing temporary migration significantly reduces high school enrollment in rural China. The change of migration cost is identified using the exogenous timing of issuing national ID card, which is necessary for rural migrants to register as temporary residents registration in urban areas.

Even though scholars have been realized the important role labor mobility restrictions played on returns to education, there is no published studies that tackle the effect of Hukou on educational attainment in China to my knowledge. Relying on 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 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, sex, ethnic minorities, education level, employment status and occupation. Hukou category of rural/urban and Hukou location at province level are reported as well.

Newborn children had to inherit their mother's Hukou type (rural/urban) until Sept. 1998[9], when inheriting father's Hukou type is permitted excluding Beijing, Shanghai, Guizhou and Xinjiang.[10] Children under 18 by Sept. 1998 who followed mother's hukou now have a chance to change it according to their fathers'. Therefore, those born after Sept. 1980 with mother holding rural Hukou and father holding urban Hukou are beneficiaries of this policy change. The opportunity of obtaining urban Hukou lowers returns to education, hence is expected to reduce high school enrollment. On the other hand, rural people may face higher costs for education in cities especially for children of migrant workers. Rural migrant family could access local education subject to additional fees, which could be waived once they get local urban Hukou. lowered high school education costs may encourage them to continue study. The net effect in this case cannot be signed a priori. Nonetheless, the first effect is likely to dominate since savings from lower tuition may be well below gains brought by urban identity.

[9]The Hukou reform was initially proposed by Ministry of Public Security in June 23, 1998. Even though approved by State Counsil in July 22, 1998, it was not put into practice until the beginning of September.

[10]Shanghai never adopts the new Hukou policy. Beijing, Guizhou and Xinjiang have regional specific threshold date of August 8, 1985 ,June 1, 1994 and Jan 1,1984 respectively.

One concern is possible delay in nationwide policy implementation as well as timing variations induced by local government, which may result in a different cutoff birth time. In order to address this issue, I adopt a technique from structural break literature to test possible discontinuities across assignment variable following Card et al. (2008) and Ozier (2010). I use a subsample including individuals born between 50 months before and after Sept. 1980. For each month between Sept. 1979 and Sept. 1981, I regress high school enrollment/dropout on a dummy indicating born on/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 best fit. This method is proved to be highly consistent in Hansen (2000). As shown in Figure 1, Sept. 1980 maximized $R^2$ and is considered to be the "true" cutoff.

# 3 Theoretical Framework

In this section, I present a simple human capital investment model to frame optimal schooling decision for rural people in China. It allows me to illustrate the effect of transferring Hukou category from rural to urban on high schooling enrollment in a dynamic setting by analyzing change in returns to education.[11] I assume each household only has one child following De Brauw and Giles (2008).[12]

In each period $t$, households make an investment decision between human capital $H_t$ and physical capital $K_t$. Time endowment for both adults and children is normalized

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

[12] This simplifying assumption will not affect critical predictions of the model. Additional control for number of siblings is included in empirical regressions.

to one. Human capital is accumulated by sending the child to school for $e_t$ fraction of his/her time and paying a cost proportional to school time, $P_t^e e_t$, where $P_t^e$ is the unit price of schooling. Therefore, human capital evolves as following:

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

where $G$ is a human capital production function increasing in $e_t$. $\psi_t$ 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 market. Home production is based on physical capital and labor work according to $Y_t^h = \theta_t F(K_t, L_t^{a1}, L_t^{c1})$, where $L_t^{a1}$ and $L_t^{c1}$ are the amount of time used in household production for adults and children respectively and $\theta_t$ is an exogenous parameter that reflects productivity level. Income from the labor market could be expressed as $Y_t^w = w(H_t^a)L_t^{a2} + w(H_t^c)L_t^{c2}$, where $L_t^{a2}$ and $L_t^{c2}$ are shares of time working in labor market for adults and children. $w(H_t^a)$ and $w(H_t^c)$ are wage functions increasing in human capital stock.[13] Households invest all the left over income into physical capital after deducting consumption $c_t$ and school cost from total earning. Thus, physical capital accumulates according to

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

Assume there is no credit market available and households are facing borrowing con-

[13] For simplicity, I assume wage return of education is the same for rural and urban area. This is likely to be the case since temporary migration in search of higher wages is generally allowed in present China.

straints

$$K_{t+1} \geq 0, \forall t \tag{3}$$

Children are eligible for school from $t = 0$ to $t = T - 1$ with T a finite fixed number of periods. They work exclusively since $t = T$ and are not allowed to go back to school. Households' utility from period $T$ and beyond can be written as $\Phi(K_t, H_t) + q(H_T)V_T$ where $\Phi(K_t, H_t)$ is the terminal value function if children do not change Hukou category to urban. It includes both pecuniary and non-pecuniary benefits from children's education for entire household. $V_T$ is positive terminal value function representing utility gained from transferring Hukou. The probability of obtaining urban identity, $q(H_T)$, is increasing in final human capital stock $H_T$ as explained in section 2.1. 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 time in school $e_t$. Since time available for adults and children is normalized to one in each period, $l_t^a = 1 - L_t^{a1} - L_t^{a2}$ and $l_t^c = 1 - L_t^{c1} - L_t^{c2} - e_t$. Households maximize expected lifetime utility:

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

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

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

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

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

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

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

where $\lambda_t$ and $\mu_t$ are shadow prices of physical capital and human capital at period $t$. The school demand function can be derived from these FOCs as:

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

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

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

Now we can trace out the impact of obtaining urban Hukou on school demand for rural children. The reduced school cost, $P_t^e$, eases borrowing constraint and is expected to result in higher school enrollment. The shadow price of human capital, $\mu_t$, changes as well. The terminal condition before transferring Hukou requires that:

$$\begin{aligned}\mu_T &= \frac{\partial\Big(\Phi(K_T, H_T) + q(H_T)V_T\Big)}{\partial H_T} \\ &= \frac{\partial \Phi_T}{\partial H_T} + \frac{\partial q_T}{\partial H_T} V_T\end{aligned}$$

and the intertemporal eular equation for human capital shows that $\mu_t = \delta^{T-t} E_t \mu_T$. Once gaining urban Hukou, the terminal value function simplifies to $\Phi(K_T, H_T) + V_T$ with a new terminal condition $\mu_T = \frac{\partial(\Phi(K_T, H_T) + V_T)}{\partial H_T} = \frac{\partial \Phi_T}{\partial H_T}$. Given that $\frac{\partial q_T}{\partial H_T} V_T > 0$ based on previous assumptions, the expected returns to education falls and children are discouraged from continuing study. The sign of the mixed effect is ambiguous and is left for empirical study in the remainder of this paper.

# 4 Empirical Strategy[14].

Regression Discontinuity (RD) design was introduced by Thistlethwaite and Campbell (1960) several decades ago. It has not been widely applied to empirical studies until recent years when Hahn et al. (2001) formally presents the asymptotic distribution of the estimator. One major advantage of RD design is that the treatment is as good as randomized around the threshold if individuals only have imprecise control over the assignment variable near the cutoff value as argued by Lee (2008). This quasi-experimental nature allows researchers to identify the treatment effect under seemingly mild continuity assumptions without imposing any exclusion restrictions. This provides more credible results compared to other conventional non-experimental strategies such as Difference-in-difference and Matching.

Valid RD design requires the observability of latent index $z$ determining treatment and a jump in probability of treatment at a cutoff value of $z = z_0$. In the context of this study, cohorts born in or after Sept. 1980 inherited mothers Hukou would be eligible to change Hukou according to father's. Children with mother holding rural Hukou while father holding urban Hukou now have a chance to obtain urban status

[14] I also adopt conventional Difference-in-difference (DID) strategy for estimation using cohorts born between Sept. 1979 and Aug. 1983. The treatment group consists of children with urban father and rural mother while the control group includes those with both parents holding urban Hukou and will not be affected by the policy change. I choose urban children as comparison group instead of rural because most of the children in treatment group are living in urban areas, receiving same quality of schooling as children in control group, even though at a higher tuition rate. The estimated drop of high school enrollment rate using DID is smaller than RD estimate obtained in this paper but is still statistically significant regardless of regression specifications of OLS, Probit or Klein and Spady semi-parametric estimation. An appendix reporting the full results is available from the author upon request

immediately, which reduces the expected return from higher education. The problem fits well in regression discontinuity design, with the month of birth as assignment variable and Sept. 1980 as cutoff value.

There are two types of RD design in the literature: sharp and fuzzy. Sharp design requires that the probability of treatment jumps from zero to one at the threshold while fuzzy design allows the jump of probability to be less than one. The analysis of dropout decision fits in Sharp design with birth month as the only forcing variable.

However, the 1998 policy change only affects high school enrollment decision of individuals (1) born in/after Sept. 1980 AND (2) finish middle school in/after 1998.[15] Given that most middle school graduation ceremonies take place in July, the second criterion is equivalent to finishing middle school at an academic age of 17 or older. Thus fuzzy design fits for this situation. The treatment effect could be identified by taking the ratio of jump in high school enrollment rate to jump in probability of treatment, which is the product of (1)"birth month" criterion and (2)"late high school" criterion. However, the timing of finishing middle school at the individual level is not reported in my data. Therefore, I cannot implement 2SLS for fuzzy design as generally suggested in literature.

In this scenario, the treatment is defined as being able to adjust high school choice according to new Hukou policy announced in 1998. The treatment determination criterion is depicted as the following model:

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

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

where $x_i$ is unobserved treatment status. $x_i = 1$ if treated, $x_i = 0$ otherwise. $x_i$ is assignment index of birth month with the cutoff value of $z_0 = Sept.1980$. Month of birth has been normalized to $z_0 = 0$ and $z_i$ 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 Nov. 1980 and $z_i = -3$ for individuals born in Jun. 1980. $I_1$ is an indicator function taking value of 1 if $z_i \geq 0$. $I_2$ is an indicator function taking value of 1 if age finishing middle school, $d_i$, is greater than or equal to 17 ($d_0$).

This model structure is different from both conventional Sharp and Fuzzy design, but fits in semi-sharp RD design as proposed in Appendix A. One crucial assumption to identify the treatment effect is:

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

It means the probability of making one's high school choice late is uncorrelated with birth month for individual born just after Sept. 1980. There are three factors that determine $d_i$: the school starting age, length of primary school and the probability of skipping/repeating a grade. First of all, those reaching six years old by Aug. 31 each year are allowed to enroll in primary school on Sept. 1. With this primary school enrollment cutoff date, it is plausible to assume there is no variation in the distribution of age starting school. Secondly, the length of primary school is dependent on county policy, but is independent of birth date. The last fact of zero correlation between the probability of repeating/skipping grade and the time of birth is confirmed by Fertig and Kluve (2005). Therefore, it is reasonable to assume $di$ and $zi$ are independent just to the right of cutoff.

Thus, I could obtain the treatment effect by estimating the jump in high school enrollment rate, as in sharp design, and rescaling it using an estimated proportion of finishing middle school at/after 17 for birth cohort of Sept. 1980-Aug. The only way to estimate this proportion is to exploit school status information collected in 2000 census since age of finishing middle school is not covered in questionnaire. Each individual reported the school status of “in school”, “finished” or “dropout” in addition to highest education level obtained. People born between Sept. 1983 and Aug. 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 Nov. 2000. Under the following assumption

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

the proportion of Sept. 1983- Aug. 1984 birth cohort finishing middle school at 17 or older could be used to infer the proportion of individuals making higher school decision after 1998 for those born right after threshold of Sept. 1980.

# 5 Empirical Results

The group of interest are those with father holding urban Hukou while mother holding rural Hukou.[16] 2000 census lacks direct parents-child information for each surveyed individual, but only contains information of relationship to household head. The criteria to identify parenthood are as described in Table 2.

Before stepping into formal analysis, I plot the proportion of students enrolled in high school, as well as proportion of dropouts, against month of birth for a subsample born between Sept. 1971 and Aug. 1986 with a father holding urban Hukou and a mother

[16]There is little possibility for the mother to change Hukou type to urban between 1998 policy effective date and census survey date. However, it is harder for adult to transfer Hukou in general. Even if this might be problematic, the drop in educational attainment I estimate will be biased downward as explained in next section.

holding rural Hukou in Figure 2a and 3a respectively, using a bandwidth of 6 months. The curves show quadratic fit to the left and right of the threshold. The visual evidence shows a clear decrease in the high school enrollment rate and an increase in dropout rate at the cutoff value 0, which represents Sept. 1980.

## 5.1 Bandwidth and polynomial choice

Local polynomial regression is a common choice in the literature to overcome boundary problem of kernel regression. The bandwidth choice is relatively straightforward for cases with discrete assignment variable. I choose 18 months as bandwidth and use observations within this same bandwidth distance to discontinuity on each side in the following regressions. As shown in section 5.2 and section5.3, the discontinuity estimates fluctuate slightly with the choice of bandwidth, but remain similar in magnitude.

Based on Akaike Information Criterion (AIC), first order polynomial fits best compared to quadratic, cubit and quartic specifications for both of the outcome variables. To further support this specification, I include dummies for each value of the birth month along with piecewise linear controls and test the joint significance of those dummies.[17] If they are jointly significant, then the piecewise linear regression is misspecified. The test statistic fails to reject the first order polynomial specification in both cases.[18] Thus local linear regression will be used in the following analysis.

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

[18]The p-values for high school enrollment and dropout are 0.9768 and 0.6585 respectively.

### 5.2 High school enrollment rate

The sample used here includes people born between Sept. 1971 and Aug. 1986 with a father holding urban Hukou and a mother holding rural Hukou. I exclude individuals with education level of primary school or lower and those still in middle school.[19] The sample size is restricted to 4772 with both parents identified in the household and their Hukou satisfying the filtering criterion. In addition, there are two more possible situations where I could identify the treatment status with Hukou information of only one parent: (1) a child holding rural Hukou with a father holding urban, or (2) a child holding urban Hukou, which is not changed through education, with a mother holding rural.[20] When including these observations, the sample size increases to 6597.

I use observations from all provinces for estimation. Provincial governments were allowed to set up their own threshold later in the same year based on the Hukou reform announced by the Ministry of Public Security. However, the threshold of Sept. 1980 and nationwide coverage were perceived at the time of high school decision, which is earlier than realization of local specific variations. As a result of information constraint, It is reasonable to assume middle school graduates in all provinces choose "study versus work" according to the national cutoff declared by the Ministry of Public Security, and include all provinces in the following analysis. [21]

In the RD design, the underlying assumption to guarantee locally random treatment

[19]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 Sept. 1980 is unchanged with these additional observations.

[20]Another common way to obtain urban Hukou is through military service. Individuals reaching 18 years old before year's end are eligible to enroll in military. The length for selective service is two years subject to additional years of voluntary service. Those born around the threshold are still in service if they choose to join army. The possibility of getting urban Hukou by enlisting could be ruled out for these people since the household sub-sample of 2000 census used in this study excludes those in military. The local linear regression is still valid with these additional samples.

[21]Exclusion of observations in those provinces with a different cutoff later on generates equivalent results.

is the agents' imprecise control over the assignment variable near the cutoff value. Despite the possibility of giving birth at an exact anticipated day, it was almost impossible for manipulation of date of birth near the threshold since the policy change happened almost 18 years later.

Even though manipulation is hard to detect, it could be inferred from a discontinuity of the density of assignment variable. For the samples used to estimate high school enrollment rate and dropout rate, I plot the number of observations of each birth month in Figure 4a with a quadratic fit. There is no noticeable density jump. The density smoothness test proposed by McCrary (2008) fails to reject at Sept. 1980, providing additional support for continuous density of birth month.

For simplicity, I use rectangular kernel in regressions following Imbens and Lemieux (2008).[22] Regression results are presented in Column 1 of Table 3.[23] The probability of high school enrollment decreases about 10 percentage points for cohorts born just to the right of the threshold of Sept. 1980 compared to those on the left side. This result is robust to variations of bandwidth choice. As shown in Figure 2b, the magnitude of estimated jumps using the same local linear regression procedure are similar across difference lengths of bandwidth.

I use the proportion of people finishing middle school at an age of 17 or older to estimate the probability of treatment for the cohort born in 1983-1984 academic year. There were 29% of them still in junior school when 2000 census took place as displayed in Table 3, indicating an age of 17 or older when making high school decision. I use the estimated percentage to approximate the proportion of the Sept. 1980-Aug. 1981 birth cohort finishing middle school late, and having the chance to

[22]The type of kernel has little impact in practice. Even though triangular kernel performs well in boundary estimation, changing kernel to triangle type generates similar results.

[23]I modify Stata code provided by Imbens & Kalyanaram and apply it to all the following non-parametric regressions.

adjust high school choice according to new policy.

The local average treatment effect could be estimated by the ratio of the probability jump of high school enrollment to the probability jump of treatment. Being eligible to change Hukou type to urban decreases the probability of enrollment in high school by 35 percentage points, as shown in Table 3 Column 1. As the 1998 policy change happened after the annual high school exam, there was no room to adjust effort level in school. Therefore, this result can be viewed as a short run effect. In the long run, 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.

One major limitation of the estimation method adopted here is the inability to infer average treatment effect for the whole population. Those finishing middle school late 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 average treatment effect. Nonetheless, this generalization limitation is no worse than LATE obtained in fuzzy design.

### 5.3 Dropout rate

If obtaining urban Hukou decreases the returns to high school education, the 1998 Hukou policy reform would also affect the dropout decision of students in high school and above.[24]

I restrict the sample to those with at least high school education. As before, I use the same bandwidth of 18 month and rectangular kernel in the following local linear regressions. According to the regression results shown in Table4 column 1, the dropout rate of those born just after the threshold of Sept. 1980 increases by 4 percentage points compared to those born just before. I plot the magnitude of estimated discontinuity in Figure 3b as a function of the length of bandwidth. The estimate of discontinuity fluctuates slightly between 3 to 5 percentage points, but remains similar in magnitude regardless of the bandwidth choice.

However, this result may be contaminated by the non-smoothness of assignment variable. Since individuals born just after Sept. 1980 are less likely to enrollment in high school as documented in section 5.2, there may be less observations just to the right of the cutoff. This concern is confirmed when plotting the number of observations of each month in figure 4b with a quadratic fit. The discontinuity in density is not an indicator for manipulation of birth month, but is the result of another regression discontinuity.

Nonetheless, the high school enrollment decision could only be affected for those finishing middle school late (17 or order). If they chose to enroll high school at first place, the dropout decision later on can not be attribute to the 1998 Hukou reform. If the dropout rate is smooth over the cutoff for the group finishing middle school late, the increased dropout rate is caused by those already in high school at the time of policy change, indicating the age finishing middle school of less than or equal to

[24] I check the dropout rate for middle school education as well. 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. The fact that dropout rate increases after the birth month threshold of Sept. 1980 confirms this statement, however, there is no statistically significant jump around the cutoff. The lack of discontinuity may result from the compulsive education of primary and middle school in China.

16. The local average treatment effect can be estimated by rescaling the raw jump for the entire sample, and rescale it with the proportion finishing middle school no late than 16 for those born in or after Sept. 1980 and with at least high school education. However, this proportion can not be estimated due to data limitations. The estimated discontinuity without rescaling is less than the true effect. However, the dropout rate is likely to be different over the cutoff for the group finishing middle school late. If people still pursue high school education after getting urban Hukou, they may have high returns to education due to other unobservables, which result in lower dropout rate compared to those born earlier. The downward bias will be even bigger in this case.

Even though the true effect can not be consistently estimated, the first stage result of 4 percentage drop can be viewed as the lower bound of the true effect. The increased dropout rate is in line with the decreased high school enrollment rate estimated in Section 5.2, reinforcing the existence of "institutional returns" to education and its important role in human capital investment decision in China.

# 6 Robustness Check

## 6.1 Continuity of covariates around cutoff

Valid RD design requires smooth covariates over cutoff. For the sample used in analysis of high school enrollment rate, I check for possible jumps over the threshold for parents' education, number of siblings and gender[25] in Figure 5 using a quadratic fit with each of these variables defined as the following.

[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.

Parents' education is used to capture inter-generational education linkage that affects chidren's schooling.[26] I construct two dummy variables for father/mother holding middle school degree and above respectively, and check for possible jumps of these middle school indicators.[27]

Siblings are expected to decrease educational attainment for a given individual since they would compete for educational resources. I calculate the number of siblings as the number of children a mother gave birth to, and still alive at the time of survey, minus one for the self.[28] The concern with this method excludes adopted children[29], who compete for household resources as well. Nevertheless, according to Chinese law, adoption is approved if a couple does not have there own children. Therefore, the adoption could be inferred if the calculated number of siblings equals to -1(the number of children mother gave birth to is zero). There are only less than one percent adoption case in the sample, thus I do not distinguish genetic child and adopted child/stepchild in this study.

Gender has been recognized as one important determinant of educational attainment. There may b cognitive difference 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 more able to concentrate. On the other hand, males may receive better childcare and educational opportunities since the Chinese culture values boys.

[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.

[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.

One interesting pattern is the gender gap of education across time. In general, boys are more likely to finish middle school due to limited resources within households. The better chance for girls to go to school over time inferred by downward sloping line might result from the tightened one child policy.

Regardless of the dramatic change in gender decomposition across different birth cohorts, there is no evidence of significant discontinuity to the left and right of Sept. 1980, as well as for all other variables. Moreover, I test the joint significance of all the discontinuities at the threshold in a Seemingly Unrelated Regression (SUR), where each equation regress one covariate on a threshold dummy, a constant and a fourth order polynomial. The coefficients of polynomials are allowed to be different on each side of the threshold and errors are allowed to be correlated across equations. This test fails to reject the hypothesis that covariates are smooth across cutoff.

I check the smoothness of the same covariate variables for the sample used in the analysis of dropout rate in Figure 6. The discontinuities appear to be bigger as compared with sample used in high school enrollment estimation. This may result from the limited sample size when I impose the restriction of at least high school education. As stated in Lee and Lemieux (2010), even if the functional form is correctly specified, the discontinuity may become significant by random chance in the case of multiple covariates. When I test the joint significance of discontinuities using SUR as suggested by Lee and Lemieux (2010), I f the hypothesis of "no discontinuity" around cutoff.

## 6.2 Incorporating covariates in estimation

Valid RD design estimates change little with additional covariates. Adding covariates that have good explanatory power may help reduce variance. As a robustness check, I add parents' education indicator (1 for at least middle school), number of siblings, and gender in the original local linear regression. As reported in Table 3 column 2, the high school enrollment estimates are robust to the additional explanatory variables. Table 4 column 2 shows the result for dropout rate. Even though they magnitude of discontinuity drops from 0.04 to 0.03, the estimate is still statistically significant.

The calculation of the number of siblings is through number of children a mother gave birth to and is based on identification of parents-children 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 sample size for both outcome variables. In order to avoid potential problems induced by fewer observations, I also report the regression results in 3 column 3 and 4 column 3, excluding the number of siblings. The estimate of discontinuity remains significant at a 5% level with similar magnitude.

## 6.3 Parametric estimation with clustered standard errors

Lee and Card (2008) study RD design with 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 case. 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 high school enrollment/dropout dependent on birth month using the fourth order polynomial[30] and allow the errors clustered at birth month level. The coefficients are allowed to be different on both sides of the cutoff:

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

where $Y_i$ is a dummy variable for high school enrollment/dropout. $Z_i$ is the birth month with a cutoff value of Sept. 1980, denoted by c. $D_i$ is an indicator that takes the value of one if birth month is no earlier than Sept.1980. $\tau$ is the estimated probability jump of high school enrollment. The coefficients of the fourth order polynomial specification are allowed to be different on either side of the discontinuity. The OLS results for high school enrollment rate and dropout rate are shown in Table 5 Column 1 and Column 3 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 2 and Column 4. The estimated discontinuities are all similar in magnitude to those obtained from local linear regression.

## 6.4 Age induced variation

The school enrollment cutoff date in China is Sept. 1. Children 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 Sept. are older comparing to students in same grade while those born in Aug. are relatively younger, which may lead to different performance in school.

[30]Fourth order polynomial passes the goodness-of-fit test proposed in Lee and Card (2008) with $G_{hs} = 0.84$ and $G_{drop} = 1.03$ respectively, both well below the 95th percentile critical value of 1.30. The results are largely unchanged using specifications of other polynomial order.

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 due to age-based mandatory school attendance laws in U.S.. Children starting primary education at and earlier age have to stay enrolled in longer before reaching legal school leaving age of 16.

Unlike 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. (2008) 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 18 months, using the same sample but a different cutoff of Sept. 1978, two 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 at threshold is not likely to be caused by school entry cutoff date.

## 6.5 Other issues

The estimate of 35 percentage points drop in high school enrollment is based on the assumption that no information is revealed before announcement of the policy. However, one may argue for the possibility of predicting the policy change one year or two year earlier. Under the assumption of perfect information, the policy would not only affect children finishing middle school in 2000 or later, but all individuals born in/after Sept. 1980, including those making high school decision earlier. As a result,

[31]The point estimate is 0.018 with the standard error of 0.048.

the first stage estimated 10 percentage points decrease without rescaling provides the treatment effect under the other extreme case of perfect foresight of the policy change. Moreover, the impact of obtaining urban Hukou on high school enrollment for all other cases with different degrees of information limitation are bounded by the two estimates, one with rescaling and the other without.

# 7 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 legal identity in more developed areas.

I adopt newly developed Regression Discontinuity approach to estimate the institutional returns to education in China under the Hukou system. People attending college(after regular high school) or technical high school are granted urban Hukou automatically. Therefore, high school education in China is considered as a bridge to urban identity.

As shown in the previous sections, removing the institutional returns to education by directly granting urban Hukou decreases the high school enrollment rate substantially by 35 percentage points for those holding rural Hukou. As a result, the Hukou system has played an important role in encouraging high school education after the nine years of mandatory schooling. The Hukou system may corrects potential under-investment problems for education induced by private returns that are lower than social optimal and contributes to a high rate of economic growth in China.

In order to mitigate the unbalanced development between rural and urban areas, the Hukou system has been relaxed since the mid 1980s when rural residents were allowed to temporarily work in urban areas. Starting from 2003, a few provinces have used uniform identity Hukou to replace the original rural/urban dichotomy, indicating a fading out of a rural-urban divide.

Understanding Hukou's role of fostering education is crucial during this transition. As projected in this study, the elimination of the Hukou type erases the institutional returns to high school education. The high school enrollment rate is expected to drop substantially for middle school graduates with rural origin. Even though they are able to earn an appealing wage by working directly after middle school, their potential for career development and long run 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, the within-urban inequality will emerge based on different education level.

The possible slow down of human capital accumulation may influence China's long run economic growth as well. According to Fleisher et al. (2010), if the portion of workers with at least some high school education increases by one percentage point, the total factor productivity (TFP) increases by about 0.5 percentage point a year. As calculated in upper panel of table 6, replacing the rural/urban Hukou type with uniform identity will cause a 0.229 percentage point reduction in proportion holding high school degree for the whole population. As a result, the first year's GDP after the reform is going to decrease by 0.1%. This seems to be a trivial amount compared to the efficiency gain of 24% GDP when rural/urban labor mobility restrictions are removed, simulated by Whalley and Zhang (2007) under the assumption that 75% of capital is mobile. However, the negative impact of lowered education attainment on GDP is going to accumulate over time since the GDP growing pattern will be changed by the slow down of TFP growth. As shown in table 6b, the efficiency gain by allowing people move freely between rural and urban area will be eaten up by the slow down of long run growth in 20 years. The growth effect will be bigger when taking in to account the increased high school dropout rate. Even though the calculation is based on simplified assumptions which may not capture all aspects of reality, it highlights the importance of high school and college education in China.

Despite the college expansion, the burden for education costs has gradually shifted from government to students. The considerable high school and college tuition combined with an under-developed college loan system discourage the poor from pursuing high education after middle school. Policies that mitigate the financial constraints and encourage continuing school after compulsory education, especially for the poor, would correct for knowledge spill over effect and erase within-urban inequality.

# References


Acemoglu, D. (1996). A Microfoundation for Social Increasing Returns in Human Capital Accumulation. *Quarterly Journal of Economics*, 111(3):779.

Acemoglu, D. (1998). Why Do New Technologies Complement Skills? Directed Technical Change and Wage Inequality. *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. (2009). *On the Robustness of Brain Gain Estimates*. Forschungsinst. zur Zukunft der Arbeit.

Black, S., Devereux, P., and Salvanes, K. (2008). Too Young to Leave the Nest: The Effects of School Starting Age.

Card, D., Mas, A., and Rothstein, J. (2008). Tipping and the Dynamics of Segregation. *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*.

Chen, B., Lu, M., and Zhong, N. (2010). Household Consumption Constrained by Hukou System. *Economic Research Journal*, Special Issue:62–71.

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.

Docquier, F. and Rapoport, H. (2011). Globalization, Brain Drain and Development. IZA Discussion Papers 5590, Institute for the Study of Labor (IZA).

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.

Groot, W. and van den Brink, H. (2010). The Effects of Education on Crime. *Applied Economics*, 42(3):279–289.

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.

Iregui, A. (2003). Efficiency Gains from the Elimination of Global Restrictions on Labour Mobility: An Analysis Using a Multiregional CGE Model. Discussion Paper No. 2003/27, World Institute for Development Economic Research (UNU-WIDER).

Lee, D. (2008). Randomized Experiments from Non-Random Selection in Us 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.

McCrary, J. (2008). Manipulation of the Running Variable in the Regression Discontinuity Design: A Density test. *Journal of Econometrics*, 142(2):698–714.

Moretti, E. and Lochner, L. (2004). The Effect of Education on Criminal Activity: Evidence from Prison Inmates, Arrests and Self-Reports. *American Economic Review*, 94(1):2004.

Moses, J. and Letnes, B. (2004). The Economic Costs to International Labor Restrictions: Revisiting the Empirical Discussion. *World Development*, 32(10):1609–1626.

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.*

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.

Thistlethwaite, D. and Campbell, D. (1960). Regression-Discontinuity Aanalysis: An Alternative to the Ex Post Facto Experiment. *Journal of Educational Psychology*, 51(6):309.

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

Walmsley, T. and Winters, L. (2005). Relaxing the Restrictions on the Temporary Movement of Natural Persons: A Simulation Analysis. *Journal of Economic Integration*, 20(4):688–726.

Weisbrod, B. (1962). Education and Investment in Human Capital. *The 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. 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: Discontinuity Search Result

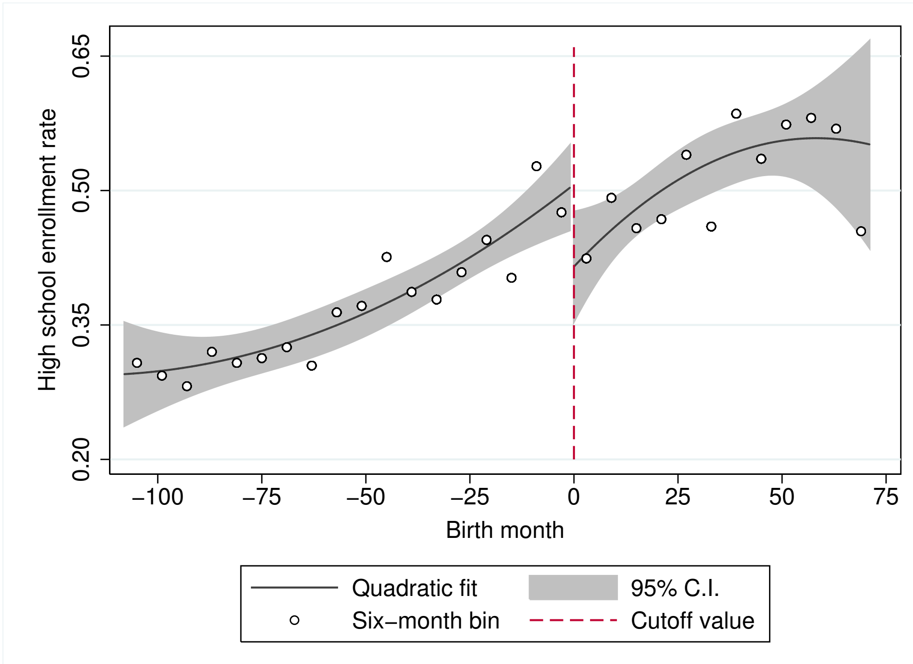
(a) High School Enrollment Rate


(b) Dropout Rate of High School or Above

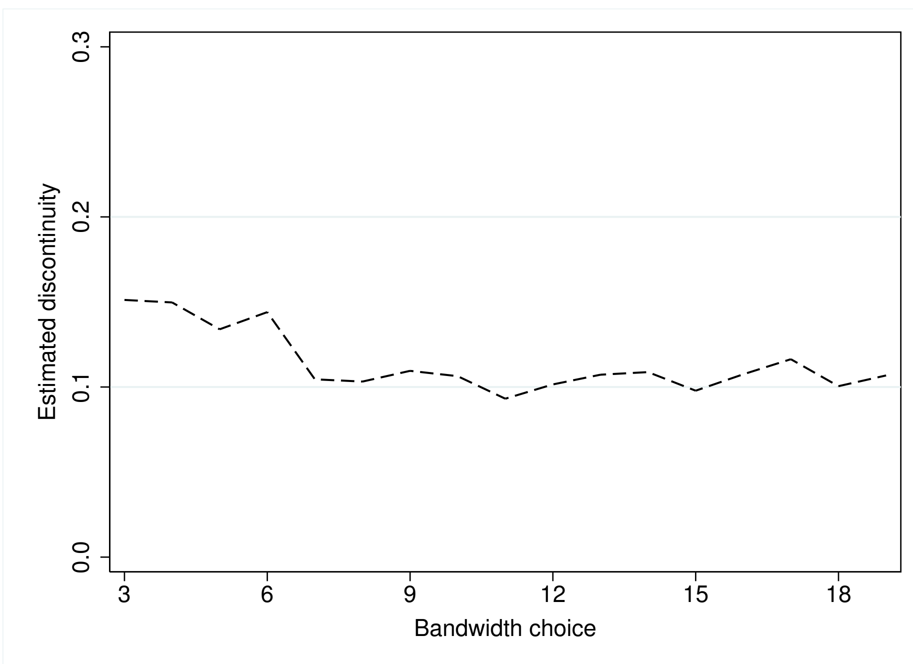

Note: Search for possible discontinuities between Mar. 1979 and Mar. 1981. Sample used here includes individuals born between 50 months before and after Sept. 1980. Birth month is normalized with Sept. 1980=0.

Figure 2: Discontinuity for High School Enrollment Rate

(a) High School Enrollment Rate by Birth Month with Quadratic Fit



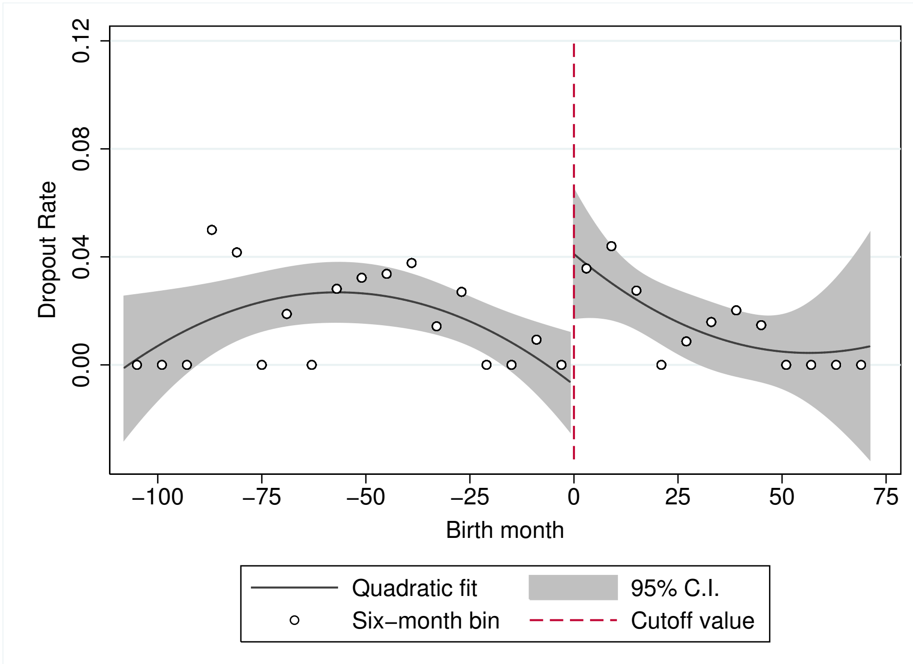

(b) Estimated Discontinuity for Different Bandwidths



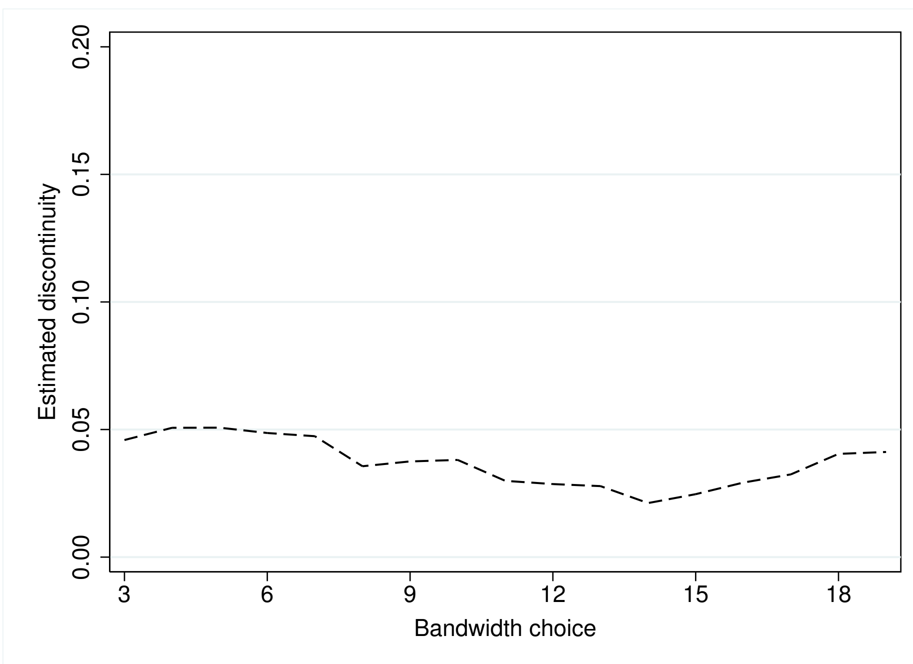

Notes: High school enrollment rate for children with father holding urban Hukou and mother holding rural Hukou. Sample used here includes individuals born between Sept. 1971 and Aug. 1986 with at least middle school education. Those still in middle school are excluded. Birth month is normalized with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

Figure 3: Discontinuity for Dropout Rate of High School or Above

(a) Dropout Rate by Birth Month with Quadratic Fit



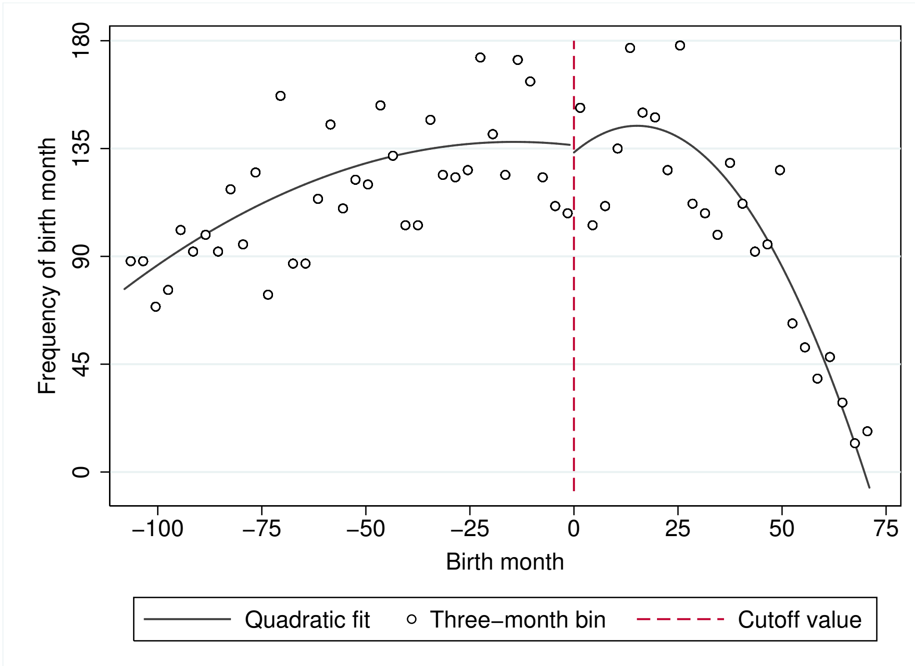

(b) Estimated Discontinuity for Different Bandwidths




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

Figure 4: Density Continuity for Birth Month with Quadratic Fit

(a) Sample to Estimate High School Enrollment




(b) Sample to Estimate Dropout Rate

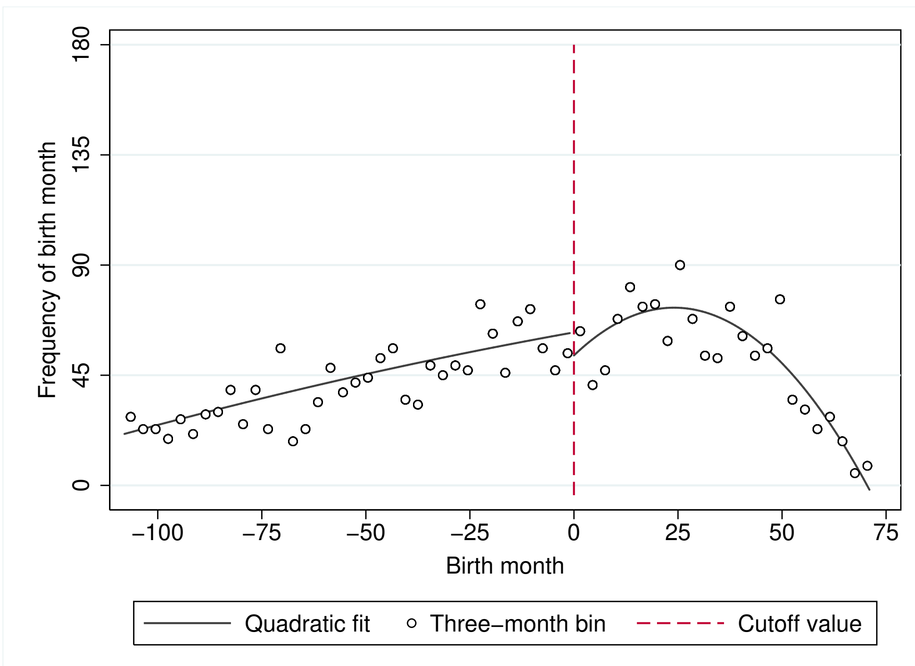
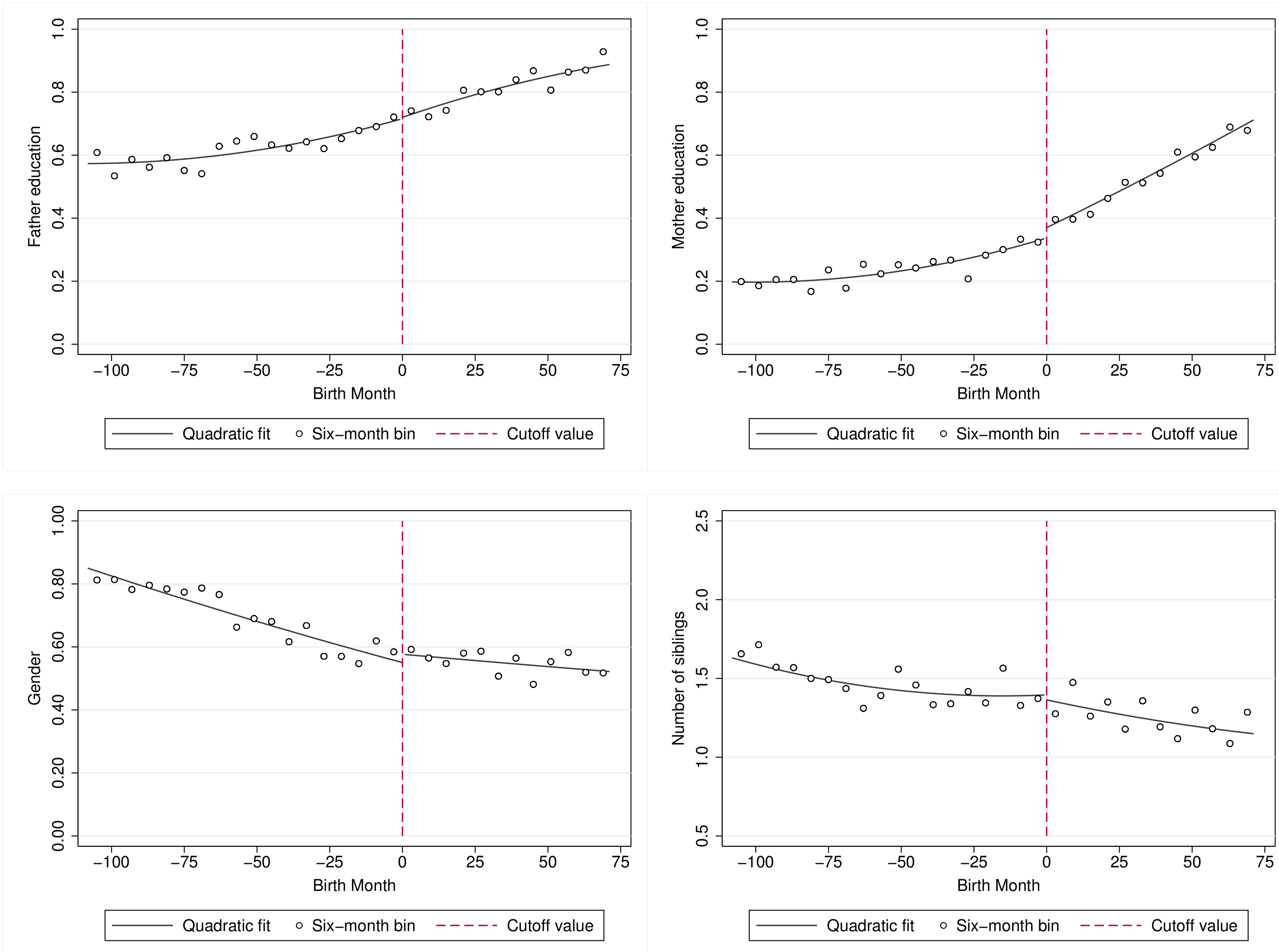
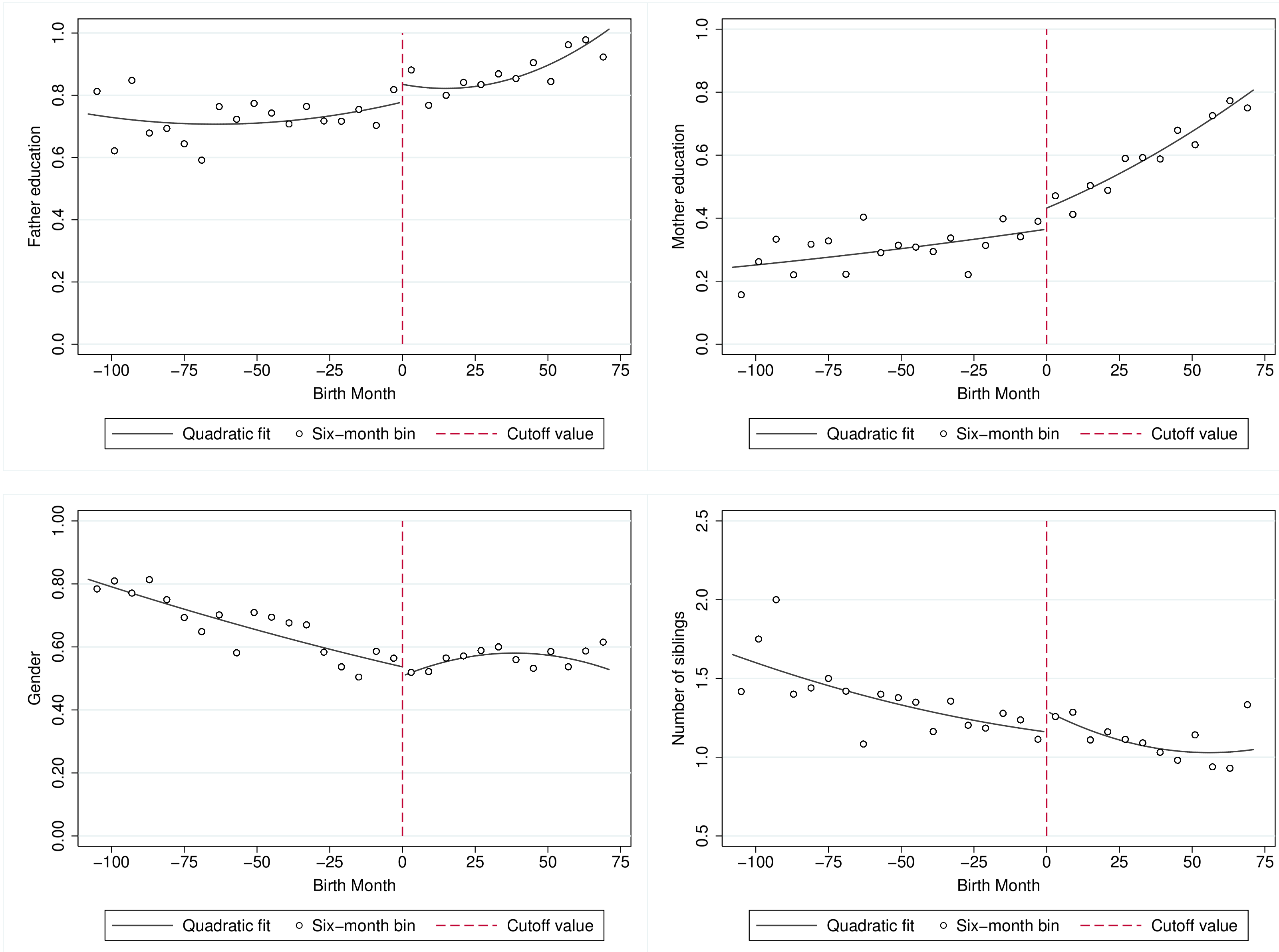


Figure 5: Covariates Continuity of High School Enrollment Rate




Notes: The joint significance test using SUR fails to reject “no discontinuity” in covariates with a p-value of 0.9781.

Figure 6: Covariates Continuity of Dropout Rate




Notes: The joint significance test using SUR fails to reject "no discontinuity" in covariates with a p-value of 0.4619.

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 | 0.64 | (0.48) | 6597 |
| Number of Siblings | 1.35 | (0.96) | 4364 |
| High School Enrollment | 0.41 | 0.49 | 6597 |
| Dropout of High school or above | 0.02 | 0.13 | 2709 |
| *Panel B: High school enrollment 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 | 0.57 | (0.49) | 1629 |
| Number of Siblings | 1.37 | (0.95) | 1309 |
| High school enrollment 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 | 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: This is a subsample of 0.95‰ of 2000 census. I exclude respondents with education level lower than middle school and those still in middle school.

Table 2: Criterion for Parent-Child Identification

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

Note: The indentification criterion is category specific based on the relationship to household head listed in first column. "Relationship" refers to relationship to household head.

Table 3: Estimation Results for High School Enrollment Rate

| Variables | ESTIMATION (1) | (2) | (3) |
|---|---|---|---|
| Father Education | No | Yes | Yes |
| Mother Education | No | Yes | Yes |
| Gender | No | Yes | Yes |
| Number of Siblings | No | Yes | No |
| Bandwidth | 18 month | 18 month | 18 month |
| Prob. Discontinuity of High School Enrollment | -0.100** | -0.116** | -0.118** |
| | (0.051) | (0.056) | (0.051) |
| Prob. Discontinuity of Treatment | | 0.287 | |
| | | (0.453) | |
| LATE | -0.351** | -0.403** | -0.411** |
| | (0.179) | (0.196) | (0.178) |
| | [0.181] | [0.198] | [0.182] |
| $N$ | 1629 | 1272 | 1563 |

Note: Sample restricted to those with at least 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 depicted in 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 probability jump of high school enrollment to probability jump of getting treated. The standard error of LATE is rescaled by treating the denominator as a constant as a result of slower convergence rate of numerator. I also report the bootstrapped standard errors with 2000 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

| Variables | ESTIMATION (1) | (2) | (3) |
|---|---|---|---|
| Father Education | No | Yes | Yes |
| Mother Education | No | Yes | Yes |
| Gender | No | Yes | Yes |
| Number of Siblings | No | Yes | No |
| Bandwidth | 18 month | 18 month | 18 month |
| Prob. Discontinuity of Dropout Rate | 0.040** | 0.030* | 0.045** |
| | (0.018) | (0.017) | (0.021) |
| $N$ | 715 | 514 | 591 |

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 depicted 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: Parametric Regression Discontinuity Estimation Results

| | OUTCOME | | | |
|---|---|---|---|---|
| | HIGH SCH. ENR. | | DROPOUT | |
| Variables | (1) | (2) | (3) | (4) |
| Threshold | -0.1121*** | -0.1398*** | 0.0445*** | 0.0384** |
| (Birth Month $\geq$ 0) | (0.0419) | (0.0463) | (0.0154) | (0.0187) |
| Birth Month | 0.0075** | 0.0054 | -0.0021 | -0.0017 |
| | (0.0038) | (0.0049) | (0.0016) | (0.0014) |
| (Birth Month)$^2$ | 1.84e-04 | 9.34e-05 | -5.93e-05 | -5.40e-05 |
| | (1.48e-04) | (1.99e-04) | (7.04e-05) | (6.47e-05) |
| (Birth Month)$^3$ | 2.42e-06 | 9.04e-07 | -7.57e-07 | -6.68e-07 |
| | (2.10e-06) | (3.01e-06) | (1.10e-06) | (1.04e-06) |
| (Birth Month)$^4$ | 1.13e-08 | 3.56e-09 | -3.52e-09 | -2.88e-09 |
| | (9.75e-09) | (1.50e-08) | (5.39e-09) | (5.14e-09) |
| (Birth Month) $\times$ Threshold | 0.0062 | 0.0164** | 0.0014 | 0.0010 |
| | (0.0078) | (0.0084) | (0.0037) | (0.0040) |
| (Birth Month)$^2\times$ Threshold | -8.42e-4* | -0.0014*** | 2.55e-05 | 1.85e-05 |
| | (4.50e-4) | (4.65e-04) | (2.31e-4) | (2.43e-4) |
| (Birth Month)$^3\times$ Threshold | 1.37e-5 | 2.84e-05*** | 2.21e-06 | 2.21e-06 |
| | (9.97e-06) | (1.00e-05) | (5.02e-06) | (5.19e-06) |
| (Birth Month)$^4\times$Threshold | -1.44e-07* | -2.29e-07 *** | -9.65e-09 | -1.12e-08 |
| | (7.42e-08) | (7.30e-08) | (3.55e-08) | (3.65e-08) |
| Father Education | - | 0.1256*** | - | -0.0024 |
| | | (0.0171) | | (0.0083) |
| Mother Education | - | 0.541*** | - | 0.0041 |
| | | (0.0165) | | (0.0055) |
| Gender | - | -0.0211 | - | 0.0061 |
| | | (0.0158) | | (0.0054) |
| Number of Siblings | - | -0.0670*** | - | 0.0031 |
| | | (0.0071) | | (0.0021) |
| $N$ | 6597 | 4211 | 2709 | 1936 |
| $R^2$ | 0.038 | 0.065 | 0.004 | 0.004 |

Notes: Birth month is rescaled as $Sept.1980 = 0$. Binary variable threshold takes value of 1 if individual is born in or after Sept. 1980. The parametric model used here is fourth order polynomials with constant term. Coefficients are allowed to be different to the left and right of threshold. Standard errors are allowed to be clustered at birth month level. Asterisks *, ** and *** denote significant levels of 10%, 5% and 1% respectively.

Table 6: Welfare Analysis of Removing Rural/Urban Hukou Type

(a) Decrease in TFP Growth Rate

| Year | 2001 |
|---|---|
| (1). Rural Middle School Graduates (Million) | 8.33 |
| (2). Decrease in HS Enrollment Rate (Percentage Point) | 35.1 |
| (3). Total Population (Million) | 1276.27 |
| (4). Decrease in Proportion of Pop. with High School Edu.(1) × (2)/3 | 0.229 |
| (5). Impact of 1 ppt increase in h.s. worker's share on TFP Growth | 0.5 |
| (6). Decrease in Growth Rate (4) × (5) | 0.1145 |

Notes: (1) (3) are measured in million. (2) (4) (5) (6) are measured in percentage point.

(b) Welare Effect

| | Growth Rate | GDP | Growth Rate | GDP | GDP Loss | GDP |
|---|---|---|---|---|---|---|
| Year | (1) | (2) | (3) | (4) | (5) | (6) |
| 0 | - | 9,731.5 | - | 9,731.5 | - | 9,731.5 |
| 1 | 10.00% | 10,704.6 | 9.89% | 10,693.5 | 0.10% | 13,259.9 |
| 2 | 10.00% | 11,775.1 | 9.77% | 11,738.3 | 0.31% | 14,555.6 |
| 3 | 10.00% | 12,952.6 | 9.66% | 12,871.9 | 0.62% | 15,961.1 |
| 4 | 10.00% | 14,247.9 | 9.54% | 14,100.1 | 1.04% | 17,484.2 |
| 5 | 10.00% | 15,672.6 | 9.43% | 15,429.4 | 1.55% | 19,132.5 |
| 6 | 10.00% | 17,239.9 | 9.31% | 16,866.3 | 2.17% | 20,914.3 |
| 7 | 10.00% | 18,963.9 | 9.20% | 18,417.8 | 2.88% | 22,838.1 |
| 8 | 10.00% | 20,860.3 | 9.08% | 20,090.8 | 3.69% | 24,912.7 |
| 9 | 10.00% | 22,946.3 | 8.97% | 21,892.9 | 4.59% | 27,147.2 |
| 10 | 10.00% | 25,241.0 | 8.86% | 23,831.5 | 5.58% | 29,551.1 |
| 11 | 10.00% | 27,765.0 | 8.74% | 25,914.5 | 6.67% | 32,134.0 |
| 12 | 10.00% | 30,541.6 | 8.63% | 28,149.9 | 7.83% | 34,905.9 |
| 13 | 10.00% | 33,595.7 | 8.51% | 30,545.8 | 9.08% | 37,876.9 |
| 14 | 10.00% | 36,955.3 | 8.40% | 33,110.8 | 10.40% | 41,057.5 |
| 15 | 10.00% | 40,650.8 | 8.28% | 35,853.2 | 11.80% | 44,458.0 |
| 16 | 10.00% | 44,715.9 | 8.17% | 38,781.7 | 13.27% | 48,089.4 |
| 17 | 10.00% | 49,187.5 | 8.05% | 41,905.0 | 14.81% | 51,962.2 |
| 18 | 10.00% | 54,106.2 | 7.94% | 45,231.8 | 16.40% | 56,087.5 |
| 19 | 10.00% | 59,516.8 | 7.82% | 48,770.9 | 18.06% | 60,476.1 |
| 20 | 10.00% | 65,468.5 | 7.71% | 52,531.2 | 19.76% | 65,138.8 |
| R/U Hukou | Yes | Yes | No | No | No | No |
| Eff. Gain | - | - | No | No | No | Yes |

Notes: I use 2001 as base year to match Whalley and Zhang (2007)'s simulation data. GDP data is from China Statistical Yearbook. For simplicity, the initial GDP growth rate is assumed to be a constant of 10%. Column shows potential GDP under this growth rate. Column (2) and (3) show the impact of decreased high school enrollment on growth rate and GDP, ignoring the efficiency gain. The potential GDP loss is calculated in column (5). The GDP calculation in the last column incorporates both the negative effect cause by decreased schooling and 23% efficiency gain simulated by Whalley and Zhang (2007). Even though the gain is greater than loss in initial periods, there relative importance reverses in the 20th year.

# Appendices

## A Regression Discontinuity Design with Unobserved Treatment Variable

### A.1 Sharp and fuzzy regression discontinuity design

There are two types of Regression Discontinuity designs in the literature: Sharp and Fuzzy. Sharp design requires the probability of treatment jumps from zero to one at 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. It equals to one if individual $i$ is treated, zero otherwise
$\epsilon_i$ is a random error term.
$z_i$ is the assignment variable with 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 $E[x_i \mid z_i = z] = Pr[x_i = 1 \mid z_i = z]$ 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 $\hat{y}^+, \hat{y}^-, \hat{x}^+, \hat{x}^-$, 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 boundary problem of kernel regression. Under certain assumptions, the asymptotic distribution can be derived as:

*Theorem 1:*

$$n^{\frac{2}{5}}(\hat{\beta}_{fuzzy} - \beta_{fuzzy}) = n^{\frac{2}{5}}(\frac{\hat{y}^+ - \hat{y}^-}{\hat{x}^+ - \hat{x}^-} - \frac{y^+ - y^-}{x^+ - x^-}) \to N(\mu_f,\ \Omega_f)^{32} \tag{A.1}$$

*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 Semi-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 those two when aggregate level information of $x_i$ is enough for estimation. I call it *semi-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 $d_i$

$$y_i = \alpha_i + x_i \times \beta_i + \epsilon_i$$
$$x_i = I_1(z_i \geq z_0) \times I_2(f(d_i, d_0) \geq 0)$$

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

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

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

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

Even though the rescaling formula above is similar to fuzzy RD, semi-sharp RD design has a different nature from fuzzy case. In fuzzy RD, with unobserved $x_i$, $x^+$ and $x^-$ could not be calculated due the unknown additional variables that determine treatment status $x_i$ together with $z$, and how they interact with $z_i$. In semi-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 obtained 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 $x^+$ given conditional probability for $I_2 = 1$. A simple case is when $z_i$ is independent to $d_i$ just to the right of $z_0$. 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

$$y_i = \alpha_i + x_i \times \beta_i + \epsilon_i \tag{A.3}$$
$$x_i = I_{1i}(z_i \geq z_0) \times I_{2i}(d_i \geq d_0) \tag{A.4}$$

*Assumption 1:*
$z_i$ and $d_i$ are independent for $z_i \in (z_0, z_0 + \epsilon)$, where $\epsilon$ is a small positive number

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

By Central Limit Theorem,

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

Therefore

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

with 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 $\mu_d$ is required for the estimation. This result can be applied to situations where individual level treatment information is missing but aggregate level data (proportion getting treatment) is available in either original data set or additional data set using same sampling rule.

## A.4 Treatment effect

The model setting of semi-sharp design is another form of departure from classic sharp design besides fuzzy design. A two-step procedure is suggested: first estimate outcome discontinuity at threshold using sharp design estimator shown in equation

(A.2); then rescale it with the proportion treated. The effect identified could be explained as weighted average treatment effect on treated(ATT) for the subgroup with $d_i = 1$, where weights are determined by ex ante probability that one's assignment variable will fall in the neighborhood of the threshold.[33] The major drawback is the estimated effect cannot be used to infer average treatment effect(ATE) for the population. Nevertheless, it is no worse than weighted local average treatment effect(LATE) obtained in 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 $\nu_i$ 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 $z_0$. This is exactly the same effect identified using semi-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 semi-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 $\beta$ is -0.3;
- $y_i = \alpha + x_i \times \beta + \epsilon_i$ with $\epsilon_i$ normally distributed with mean 0 and variance 1.

I use data on $y_i, z_i, d_i$ in semi-sharp estimation and $y_i, z_i, x_i$ in fuzzy estimation. I draw 1000 observations each time, and repeat the estimation procedure for 5000 times with bandwidth of 18. As shown in Figure A.1, the two methods generate similar results.

[33] Please refer to Lee and Lemieux (2010) for detailed information about the weighted nature of RD estimates.

## A.5 Generalization

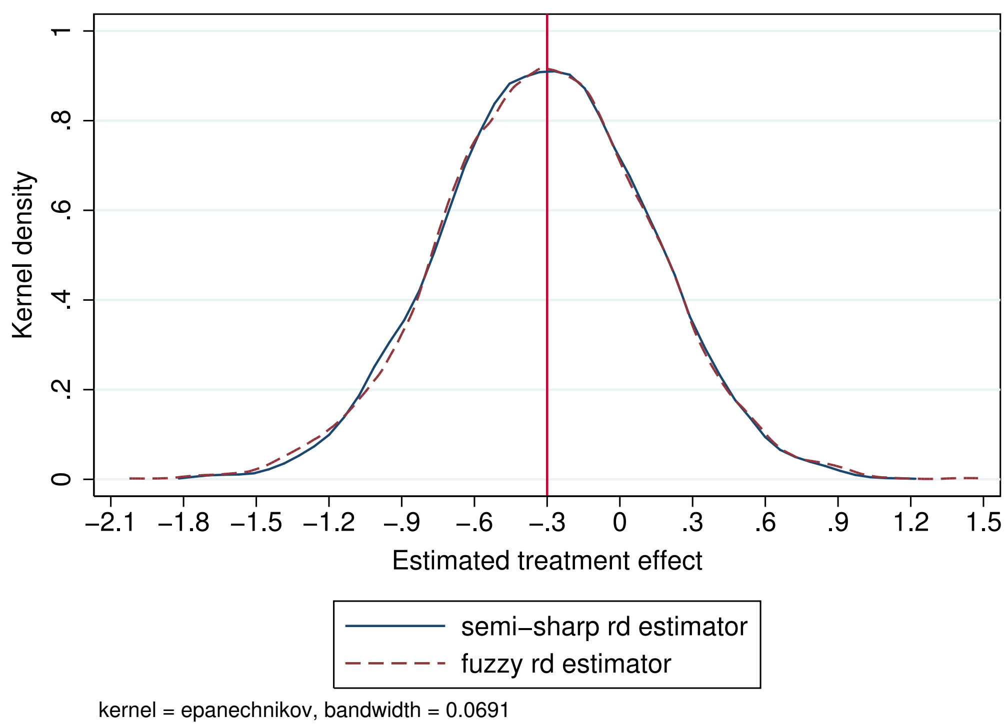
This result could be generalized to situations with more than one additional variables determining treatment status:

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

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

The "quasi-experimental" Regression Discontinuity design could be applied in more empirical cases, especially for research on developing countries where the observability of individual level data is a prevailed problem.

Figure A.1: Comparison Between Semi-sharp and Fuzzy estimator




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