I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 11735

Returns to Higher Education in China: Evidence from the 1999 Higher Education Expansion Using Fuzzy Regression Discontinuity

Fengyan Dai
Fang Cai
Yu Zhu

AUGUST 2018

DISCUSSION PAPER SERIES

IZA DP No. 11735

Returns to Higher Education in China: Evidence from the 1999 Higher Education Expansion Using Fuzzy Regression Discontinuity

Fengyan Dai
Nanjing University of Finance and Economics

Fang Cai
Chinese Academy of Social Sciences

Yu Zhu
University of Dundee and IZA

AUGUST 2018

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Returns to Higher Education in China: Evidence from the 1999 Higher Education Expansion Using Fuzzy Regression Discontinuity*

China experienced a 47% expansion in higher education enrolment between 1998 and 1999, and a six-fold expansion in the decade to 2008. In this paper, we explore a fuzzy discontinuity in the months of births induced by the expansion to study the returns to higher education in China. We find that the mean years of education increased by roughly one full year around the cut-off point of the 1999 expansion as defined by months of births. Importantly, each additional year of university education induced by the 1999 higher education expansion increases monthly wage income by 21%, whereas the corresponding OLS estimate is only 8%. Our findings are insensitive to alternative window widths, functional forms, or the exclusion of the self-employed. Moreover, the returns to degrees also appear to vary by gender, with lowers returns to women except when they are the only child in the family.

JEL Classification: I23, I26

Keywords: returns to higher education, higher education expansion, regression discontinuity design, China

Corresponding author:
Yu Zhu
School of Business
University of Dundee
3 Perth Road
Dundee, DD1 4HN
United Kingdom
E-mail: yuzhu@dundee.ac.uk

* We thank participants of the 17th IZA/SOLE Transatlantic Meeting of Labour Economics, the 2018 China Meeting of the Econometric Society and seminar at Shanghai Lixin University of Accounting and Finance for helpful comments. We are indebted to Gerard Pfann and Lei Xu for suggestions. We are also grateful to the Chinese Academy of Social Sciences for allowing access to the China Urban Labour Force Survey data. All errors remain our own.

Highlights

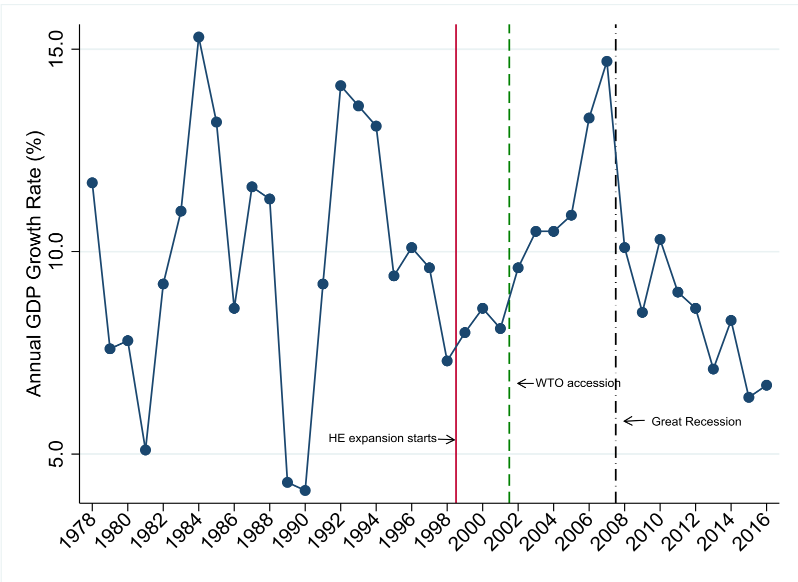
- Study an unprecedented expansion of higher education in a developing country context
- Robust estimates using fuzzy regression discontinuity design (fuzzy RDD) indicate 21% annual return for marginal graduates
- Results robust to alternative window widths, functional forms or exclusion of self-employed
- Lower returns to women except when they are the only child raises important equity concerns of the higher education expansion
- Results suggest that relaxing constraint of skilled labour conducive to development under certain conditions

Section 1: Introduction

Most economies have experienced significant expansions of the higher education (HE) sector in recent decades. While a simple textbook model would predict a decline in the college premium as a result, the empirical evidence is mixed. Evidence from the US (see e.g. Card and Lemieux (2001) and Goldin and Katz (2008)), and the UK (Blundell et al. (2000, 2016), Walker and Zhu (2008)), suggests that the returns to university degrees are largely constant over time, despite the substantial increase in college enrolment. The usual explanation is *skilled biased technological change (SBTC)* which increases the *relative* demand for more skilled workers.

The HE sector in China expanded by a factor of six over the decade from 1999, including a 40% or so annual increase in both 1999 and 2000 (Wu and Zhao (2010), Che and Zhang (2017)). The extent of the expansion is unprecedented, at least among major economies. Moreover, this period of radical HE expansion also coincides with accelerating GDP growth, as shown in Figure 1.

Figure 1: China's Annual Real GDP Growth Rate, 1978-2016



Source: China Statistical Yearbook 2017.

However, the effect of this expansion on labour market outcomes and inequality is not very well understood so far. In this paper, we take advantage of this massive expansion to study the returns to higher education in China, using the 2005 and 2010 China Urban Labour Force Surveys, conducted by the Chinese Academy of Social Sciences (CASS). Methodologically, we explore the fuzzy Regression Discontinuity Design (RDD), which relies on less restrictive assumptions for identification than the difference-in-differences (DID) or the Instrumental Variables (IV) approaches. This is important, as there was no regional pilot of the HE expansion and the general equilibrium effect arising from the scale of the expansion is too large to ignore.

We find that the mean years of education increased by more than one full year around the cut-off point of the 1999 expansion as defined by months of births. Importantly, each additional year of university education induced by the 1999 higher education expansion increases monthly wage income by 21%. Our findings are insensitive to alternative window widths, functional forms, or the exclusion of the self-employed. Moreover, the returns to degrees also appear to vary by gender, with lowers returns to women except when they are the only child in the family.

We contribute to the empirical literature on the returns to education in at least two ways. Firstly, we are the first study on the returns to higher education in China using the RDD approach exploiting the massive expansion of the sector that took place around 1999. Secondly, we provide direct evidence to the debate on whether the expansion was economically justified on the grounds of returns to education. Our finding of such a remarkable return suggests that relaxing constraint of skilled labour could be conducive to development under certain conditions, although distributional concerns must not be overlooked.

The rest of the paper is organized as follows. Section 2 provides the background of the expansion and reviews the relevant literature. Section 3 introduces the data and describes the sample selection. Section 4 discusses the identification strategy and presents summary statistics. Section 5 presents the empirical analysis and discusses the results. Finally, section 6 concludes.

Section 2: Background and Literature review

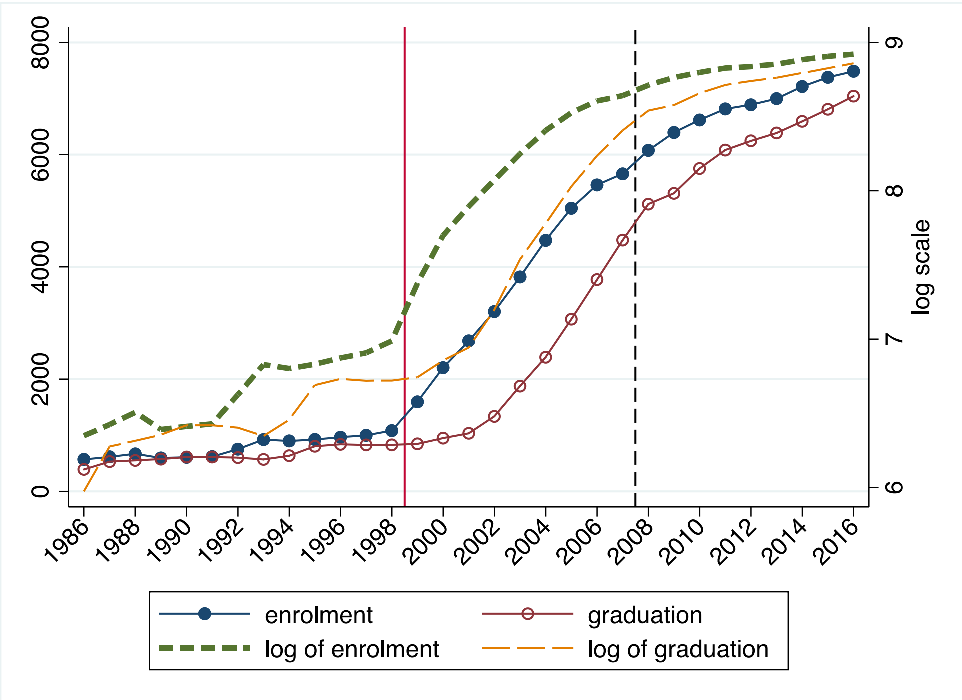
China introduced 9-year compulsory education in 1986, consisting of 6 years of primary schools and 3 years of junior high schools. Upon the completion of junior high school, students can choose to go to senior high schools, or vocational schools, both for 3 years, or enter employment directly. High school graduates can apply to universities and colleges, which normally take 4 and 3 years to complete respectively in most cases (OECD 2016).

China's HE sector has always been dominated by public funded universities and colleges. Until the mid-1990s, universities were highly selective but free of tuition fees. University places were highly rationed, and allocated according to the National Higher Education Entrance Examination (*gaokao*) results of applicants, with province-specific score cut-offs regardless of gender, *hukou*[1] status or family resources. The growth of the HE sector was tightly controlled by the Ministry of Education, which sets provincial, university and subject quotas annually (OECD, 2016). Between 1995 and 1998, college enrolment only increased by an average of 4.7% per annum (Che and Zhang 2017).

[1] China's *hukou (*household registration) system assign people to urban or rural status at the time of birth. Rural *hukou* holders face substantial disadvantage in terms of access to education, labour market and social welfare, compared to their urban counterparts (OECD 2016).

In early 1999, partly as a response to the economic slowdown and rising youth unemployment in the aftermath of the 1997 Asian Financial Crisis, the Ministry of Education announced a sudden 47% increase in university places (Wan (2006), Wu and Zhao (2010) and Li et al (2017)). This was followed by increases of 38% and 22% in 2000 and 2001 respectively, and subsequent more modest but still substantial double-digit growth year on year on average for the next decade (Che and Zhang 2017). It is only after the Global Financial Crisis in 2008 that the government decided to impose a tighter control on the rate of growth. Figure 2 shows the annual higher education enrolment and graduation in China, in levels and in logs, over the period 1986-2016. We focus on the period since 1986, the year the 9-year compulsory education was introduced. The decade from 1998 to 2008 witnessed the fastest expansion of the sector, with annual enrolment growing from 1.08 million in 1998 to 6.08 million in 2008.

Figure 2: Higher Education Enrolment and Graduation over Time, Unit: thousand



Note: Data resources: China Statistic Yearbook 2010 and 2017, (http://www.stats.gov.cn/tjsj/ndsj/2010/indexch.htm, http://www.stats.gov.cn/tjsj/ndsj/2016/indexch.htm.) Data from 1986-2009 is from yearbook 2010, data from 2010-2016 is from yearbook 2017.

China's education and labour market has drawn increased research interests, due to its growing economic prominence since the reform started in the late 1970s. The meta-analysis by Awaworyi and Mishra (2014) concludes that the returns to education in the post-reform era is around 10.25%. Apart from excluding all studies published in Chinese academic journals, another limitation of the meta-analysis is that it only covers studies using multivariate analyses or instrumental variables estimations typically based on family characteristics such as parent or spousal education. Therefore, one should be cautious in interpreting these estimates as causal due to the ability biases or endogeneity problems.

More recently, a few studies have attempted to uncover the returns to education in China using quasi-experimental methods, in the context of a fast-changing education and labour market.[2]

Using the 2002 Labour Market Survey by the Ministry of Labour and Social Security and the 2005 1% Population Survey, Wu and Zhao (2010) study the effect of the HE expansion in 1999 on labour market participation and unemployment using difference-in-differences (DD) and difference-in-difference-in-differences (DDD) methods. They find that the expansion adversely affects the employment of new graduates. Moreover, new graduates' hourly wages grow by over 10% less than older graduates do. However, relative to high school leavers, new graduates' wages are not significantly affected by the expansion, with an implied return to each year of university education around 14%.[3]

[2] Other studies using quasi-experimental methods have focused on the effect of HE expansion on the employment of college graduates. They find that while the expansion increased the unemployment of new college graduates in the short-run, this effect mostly vanishes after 5 year (see e.g., Li et al (2014) and Xing et al (2017)).

[3] Using the 2002 and 2007 China Household Income Project (CHIP) and 2002-2008 Chinese Urban Household

By pooling the China Health and Nutrition Survey (CHNS) from 1997, 2000, 2004 and 2006, Fang et al. (2012) estimate the returns to education by instrumenting years of schooling using the variation in the implementation of the 9-year compulsory education across provinces. Their IV estimate is as high as 20%, which is significantly above the OLS estimate. Exploiting the same policy reform starting in 1986, Liu et al. (2016) were the first study on returns to education in China using the RDD approach. Their RDD estimate based on the China Urban Household Survey is 12.8%, which is greater than the OLS estimate of 9.6%.

Section 3: Data and Sample

This paper is based on the China Urban Labour-force Survey (CULS) 2005 and 2010. CULS is a household survey of selected cities undertaken by the Institute of Population and Labour Economics, Chinese Academy of Social Sciences (IPLE-CASS). The survey uses multistage stratified sampling of communities, families and individuals respectively, with interviews undertaken by the local branches of the National Statistical Bureau.

The sampling frame of CULS was based on the 1% sample of the population in the corresponding years in the country. For each city selected, a two-stage sampling method was used to draw samples from the resident population in the main urban area: in the first stage, a predetermined number of neighbourhood committees were selected by the Probability Proportional to Size (PPS) method; in the second phase of resident extraction, a certain number of residential blocks are first drawn from each of the chosen neighborhood committee,

Survey, Knight et al (2017) present descriptive analysis of the 1999 HE expansion on graduate labour market outcomes. They conclude that relative to high school leavers, the adverse effect on wages, employment and access to managerial and professional jobs is restricted to entry-year or entry-period cohort of graduates.

then within each block a certain number of native and migrant households are selected using random equidistant method.

CULS covers cities at various levels, including municipalities, provincial capitals and mid-sized cities, making it more representative of the urban labour market in contemporary China. The 2005 CULS covers the following cities: Shanghai, Shenzhen and Zhuhai (both of Guangdong province), Fuzhou (of Fujian province), Wuhan and Yichang (both of Hubei province), Xian and Baoji (both of Shaanxi province), Shenyang and Benxi (both of Liaoning province), Daqing (of Heilongjiang province) and Wuxi (of Jiangsu province). On the other hand, the 2010 CULS only covers a subset of major cities from the 2005 survey: Shanghai, Fuzhou, Wuhan, Xian and Shenyang, plus Guangzhou (of Guangdong province). In order to maximize the sample size, we pool the two surveys together, resulting in over 40 thousand individuals in 13 cities from 8 provinces or municipalities. Our analytical sample spans the Eastern, the Central and the Western regions, and covers a mixture of municipalities, provincial capitals and mid-sized cities.

Section 4: Identification strategies: Fuzzy Regression Discontinuity (RD)

The fuzzy Regression Discontinuity Design (fuzzy RDD) is a quasi-experimental design in which the probability of receiving treatment changes discontinuously across the threshold as a function of a continuous running variable. Essentially, RDD compares the treatment group on the right-hand side of the cut-off point to the control group on the left side of the cut-off point, using the latter as the counterfactual (Hahn et al, 2001). The RDD strategy can be regarded as a weighted average treatment effect model on all individuals in the presence of heterogeneous treatment effects (Lee and Lemieux, 2010). Compared to more conventional quasi-experimental methods such as the differences-in-differences (DD) approach, RDD is much more robust as it only requires the weaker identifying assumptions of observations being randomly distributed and that the running variable is continuous. Another major merit of RDD is the graphical analysis which shows intuitively and transparently any jumps across the threshold in the running variable.

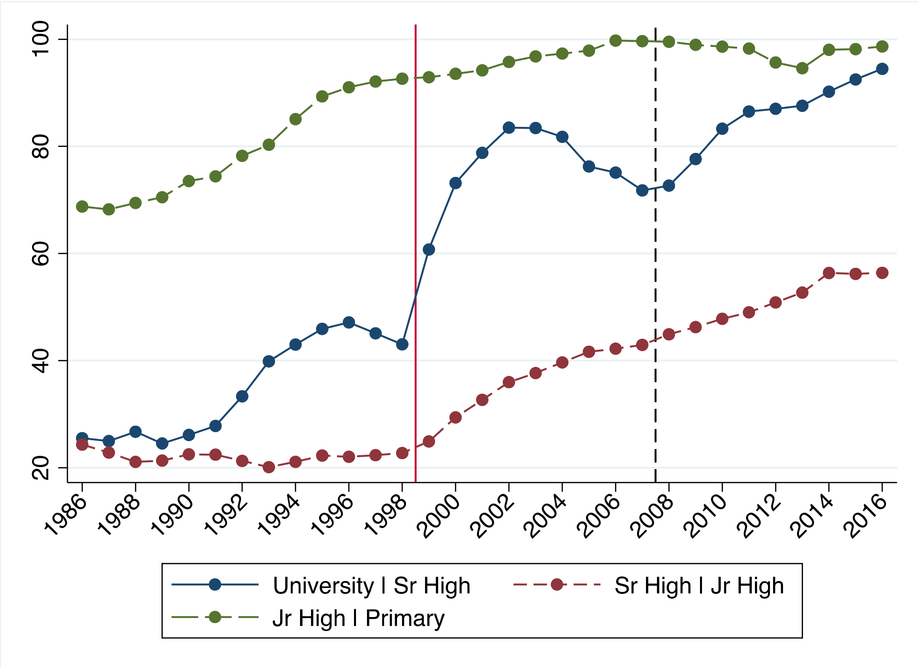
Critical to our analysis is the identification of the cut-off point in the continuous measure of month of births induced by the higher education expansion policy in 1999. This is ultimately an empirical question. According to the Chinese education system, children normally start school in the school year in which they are between 6 and 7, and are expected to complete 9 years of compulsory education in primary and junior high schools. Another 3 years of education at senior high schools or equivalent is required before one can apply to HE institutions. Therefore, taking into account the minimum 12 years of full-time education required, it works out that people who received the "treatment" of HE expansion in September 1999 or later enrolled in primary schools in 1987 or later. Given that the school year runs between September 1st and August 31st in the following year, we expect the threshold to fall between September 1st 1979 and August 31st 1981.

We choose to define the treatment group as those born on September 1st 1979 or later. This is in line with Wu and Zhao (2010) which show that among full-time students, high school students account for the vast majority of 18-year olds and a small majority of 19-year olds in the various censuses. We will undertake sensitivity checks with respect to alternative cut-off dates in the analysis.

Our preferred control group is people whose highest education is junior high school, the expected qualification at the end of the 9-year compulsory education stage. Note that the 9-year compulsory education had been enforced nationwide for more than a decade by the time of the 1999 HE expansion, and there is no reason to expect the HE expansion to affect the attainment of the Junior High School diploma, which is the school-leaving qualification.

On the other hand, one might expect a positive spill-over effect on the attainment of Senior High School qualifications, which are prerequisites for HE studies in China.

Figure 3: Conditional enrolment (progression) rate over time, by level of qualifications



Note: data resources: China Statistic Yearbook 2010 and 2017, http://www.stats.gov.cn/tjsj/ndsj/2010/indexch.htm, http://www.stats.gov.cn/tjsj/ndsj/2016/indexch.htm. Data from 1986-2009 is from yearbook 2010, data from 2010-2016 is from yearbook 2017.

Figure 3 shows the national conditional enrolment (progression) rates into HE, Senior High and Junior High Schools respectively for the time period 1986-2016, defined as the ratio of the number of enrolments at level X over the number of graduates at Level X-1 in the same year. The enrolment rate of Senior High School graduates into HE jumped from 43.1% in 1998 to 60.7% in 1999, a 40.8% increase which is very close to the 47% increase in HE enrolment.[4] While the enrollment rate of primary school graduates into junior high school already reached around 93% in 1998 and gradually converges towards 100% during the period

[4] The small discrepancy might be explained by the increased probability of enrolment of Senior High School graduates from previous cohorts. Repeating grades in order to get into (a better) college or university is a common practice in China, especially in rural areas.

of higher education expansion of 1998-2008, the enrolment rate of junior high school graduates into senior high schools appears to have experienced a significant expansion, from 22.8% in 1998 to 44.9% in 2008. This lends strong support to our choice of Junior High School graduates as the control group in our research. Including Senior High School graduates (in either the treatment or the control group) is likely to lead to biased estimates of returns to higher education.

In practice, we estimate two reduced-form equations. In the first equation, the outcome variable is years of education (or a dummy of obtaining an HE qualification).

(1) $$eduyear = \partial_0 + \delta_1 TREAT + \delta_2(a) + \partial_3 * X + \mu$$

where *TREAT* denotes the treatment status (i.e. born after September 1979), $\delta_2(a)$ is a polynomial of the running function in the continuous measure of month of birth which captures the effect of age, and X is a set of control variables. To the extent that $\delta_2(a)$ is a continuous function and the samples before and after the cut-off are as if randomly assigned, the treatment variable is the only source of discontinuity and hence δ_1 identifies the effect of discontinuity induced by the 1999 HE expansion on years of schooling.

We also estimate a second reduced-form equation, where the outcome variable is log monthly income:

(2) $$\ln(\text{income}) = ß_0 + ß_1 * TREAT + ß_2(a) + ß_3 * X + \nu$$

Using the same reasoning, one can see that β_1 identifies the effect of discontinuity induced by the 1999 HE expansion on earnings. Moreover, the returns to one year of university education is derived by the ratio $ß_1/\delta_1$.[5] Following Calonico et al. (2014, 2017), we choose

[5] We implement the two-step procedure with the help of the Stata routine *rdrobust*, which reports the correct standard errors.

the optimal fuzzy RDD bandwidth for our sample.[6] Considering the case in China, we include 36 months on either side of the cut-off. Our results are insensitive to choosing alternative bandwidths (windows) of 30 or 42 months. Following Calonico et al. (2014, 2017), we choose the *p-th* order local polynomial estimator with the *q-th* order local polynomial bias correction, and find that local linear regression with quadratic bias correction is optimal for our sample. While our baseline is based on local linear regression with triangular kernel, we report in our robustness analysis that the results are also robust to using 2nd-order polynomials of the running variable.[7]

Table 1 reports the summary statistics for the main analytical sample, with 36 months on each side of the month of birth cut-off. In other words, our main sample consists of all respondents born within 3 years of the September 1979 cut-off in the pooled CULS from 2005 and 2010, who hold either an HE degree[8] or Junior High School diploma as the highest qualification. It is worth noting that Junior High School diploma is the expected qualification for completing 9-year compulsory education whereas even the oldest cohort of our sample were subject to the 9-year compulsory education which was introduced in 1986. The final sample size is 2995, of which 2540 have non-missing monthly wage incomes. Columns 1 and 2 report the means for the employment and the wage sample respectively. The fact that they are almost identical to each other indicates that incomes are largely missing at random.

---

[6] We run bandwidth selection with the help of the Stata command *rdbwselect*, with the option *mserd* for MSE-optimal point estimation using a common bandwidth on both sides of the cutoff.

[7] We do not report results for third and fourth order polynomials. Gelman and Imbens (2017) caution against using higher-order polynomial in RDDs as they tend to inappropriately put large weights on observations further away from the threshold.

[8] We include both universities which typically last 4 years and colleges which usually last 3 years.

For the final analytical sample (N=2540), we also report summary statistics by the treatment status (i.e. before or after the cut-off), gender and *hukou* status when the respondent was completing primary education. It is worth noting that the gap in years of schooling between the treatment and the control group is 0.86 years, a remarkable jump for people who are only 3 years apart by birth on average. This is also reflected by the remarkable increase in the proportion of people holding HE degrees, from 41% to 53%. There are also significant differences in log monthly incomes (in constant 2010 prices) by gender and *hukou* status.

Table 1: Summary Statistics of the Main Sample

	All	Wage sample	By treatment status		By gender		By *hukou* status (primary school)		
			Before	After	Female	Male	Village	Town	City
Log monthly income	7.387	7.387	7.389	7.385	7.260	7.503	7.188	7.494	7.580
Education years	11.921	12.037	11.601	12.460	12.003	12.068	10.179	11.995	14.599
University degree	0.454	0.470	0.405	0.534	0.466	0.474	0.195	0.459	0.854
Employee	0.661	0.715	0.670	0.760	0.727	0.705	0.559	0.731	0.917
Experience	5.597	5.575	6.946	4.245	5.587	5.564	7.406	5.631	3.043
Migrant	0.649	0.648	0.664	0.633	0.620	0.674	0.882	0.798	0.221
Male	0.480	0.523	0.552	0.494	0.000	1.000	0.546	0.524	0.490
Father' years of educ	6.405	6.435	6.292	6.575	6.636	6.252	5.981	7.406	6.347
Mother's years of educ	5.082	5.113	4.909	5.310	5.364	4.883	4.514	5.728	5.479
Number of siblings	0.837	0.869	0.970	0.771	0.918	0.825	1.019	1.057	0.529
***Hukou* status (location of primary school)**									
Village	0.442	0.441	0.483	0.399	0.419	0.460	1.000	0.000	0.000
Town	0.233	0.236	0.225	0.247	0.235	0.236	0.000	1.000	0.000
City	0.325	0.324	0.293	0.354	0.346	0.303	0.000	0.000	1.000
Region of birth:									
Eastern	0.475	0.479	0.485	0.321	0.471	0.486	0.404	0.491	0.573
Central	0.363	0.352	0.382	0.161	0.353	0.352	0.411	0.351	0.274
Western	0.162	0.169	0.163	0.078	0.176	0.162	0.185	0.204	0.153
Current province of residence:									
Shanghai	0.155	0.163	0.143	0.183	0.160	0.166	0.122	0.204	0.190
Hubei	0.162	0.157	0.177	0.139	0.174	0.142	0.154	0.134	0.180
Liaoning	0.163	0.159	0.154	0.164	0.137	0.180	0.166	0.117	0.181
Fujian	0.166	0.164	0.171	0.157	0.177	0.152	0.185	0.177	0.125
Shaanxi	0.148	0.152	0.140	0.163	0.148	0.155	0.177	0.125	0.136
Heilongjiang	0.032	0.028	0.034	0.022	0.028	0.029	0.032	0.042	0.013
Guangdong	0.145	0.148	0.148	0.148	0.150	0.146	0.122	0.174	0.163
Observations	2,995	2540	1,251	1,289	1,212	1,328	1,119	599	822

Section 5: Empirical Analysis

We start off by presenting the main results using the full sample. We then present various robustness checks with respect to alternative window widths, functional forms and subsamples. Finally, we discuss our results in the context of the existing studies.

Throughout the analysis, we have implicitly assumed that the quality of HE has remained constant during the expansion. To the extent that education resources had been diluted in the process of this dramatic expansion, our estimates of the returns to education could be regarded as a lower bound.[9]

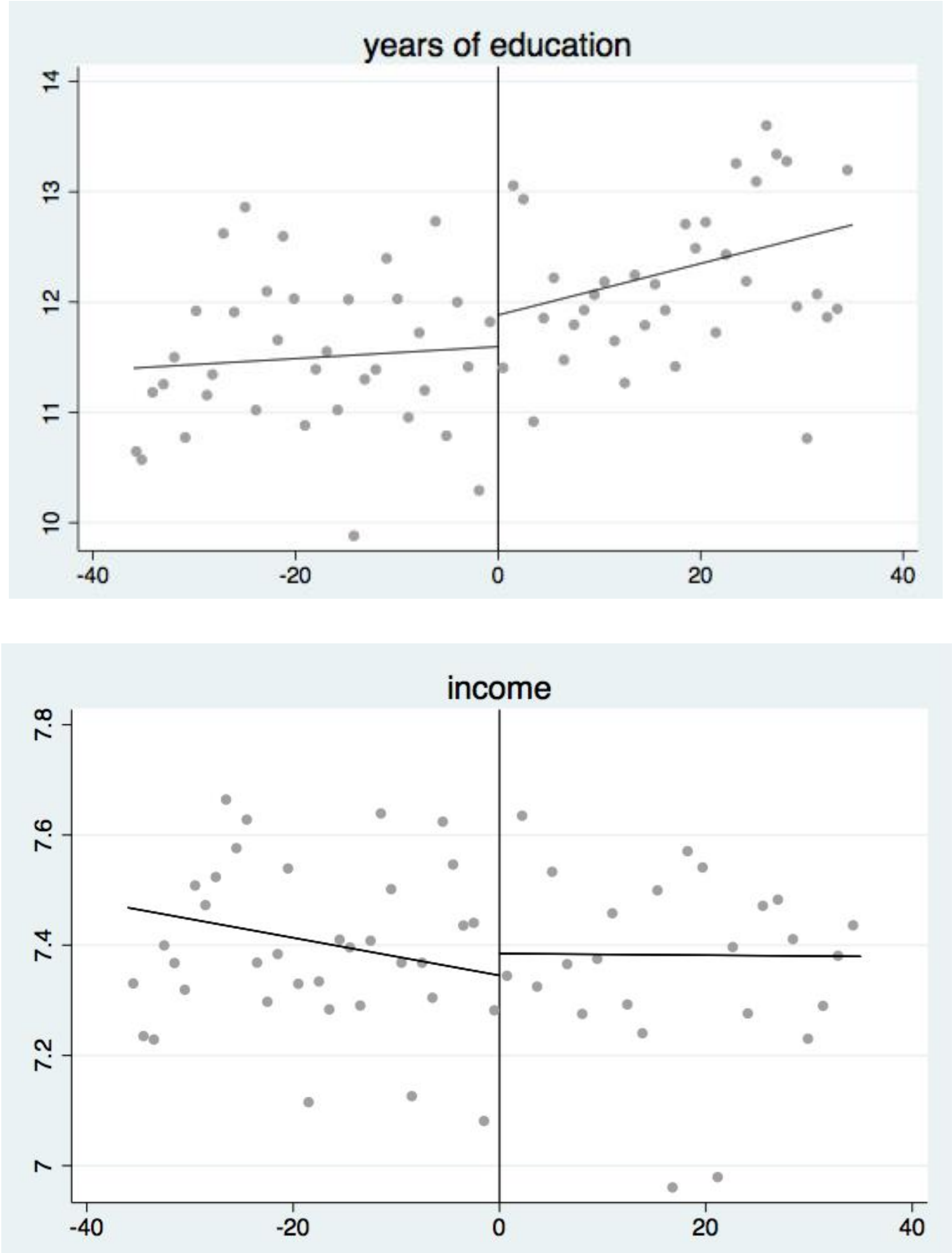
5.1 Main results

Figure 4 highlights the discontinuity in years of education and log monthly income, by months of birth. The top panel clearly indicates a jump of about 0.35 years of education around the cut-off point that corresponds to birth in September 1979. Post the cut-off, the slope also appears to have increased, consistent with the continuous expansion over a decade. In the bottom panel, the linear fit suggests a small jump in monthly income around the cut-off, but a much larger increase in the slope. A declining slope is indeed consistent with a positive effect of age (which proxies for work experience) on earnings, all else being equal. If we impose the same slope after the cut-off, the jump in monthly income would have been much higher.[10]

[9] Indeed, there is some evidence of declining resources per student during this period. E.g. Table 1 of Wan (2006) shows that the student-faculty ratio in Chinese HE jumped from 11.6 in 1998, to 13.4 in 1999, and peaked at 19.0 in 2002.

[10] Figure A1 in the Appendix shows the frequency counts by month of births (with the cut-off point defined as 0). Figure A2 in the Appendix visually checks the continuity of all covariates. All but gender appears to be continuous around the cut-off point. We will further explore the heterogeneous effect with respect to gender, however it is important to bear in mind that gender is predetermined with respect to the HE expansion.

Figure 4: Discontinuity in years of education and log monthly income.



Note: Horizontal axis indicates continuously measured month of birth (with 0 denoting September 1979).

For the main sample, Table 2 presents the baseline RDD results with different set of controls: In column (1) no control is included; column (2) controls for wave of the CULS and province dummies; column (3) additionally controls for gender (reference: male), migrant status (ref local) and a quadratic in years of experience; column (4) further controls for region of birth (reference: Western region), *hukou* status at primary school (reference: urban), interaction of region of birth with *hukou*, education of parents and number of siblings; column (5) further controls for self-employment. Our estimates are reasonably robust across the specifications. Our preferred specification is (3), which is still quite parsimonious yet controls for post-education experience and migration, indicates that the 1999 HE expansion increased the mean years of schooling by 1.05 year. Moreover, for each additional year of higher education induced by the HE expansion, monthly income increases by 0.19 log point, which is equivalent to about 21%. This is a large effect, which is also statistically significant at the 5% level. The corresponding OLS estimate in Table A1 in the Appendix is about 8.0%.

The bottom panel of Table 2 shows that the 1999 HE expansion increased the proportion of HE Degree holders in our sample by 15.0 percentage points in our preferred specification. The corresponding RDD estimate of the returns to a university degree relative to Junior High School qualification is 1.33 log points, which is equivalent to an annual return of 0.19 log points over a 7-year period (including 3 years of Senior High School).

Our estimate has the interpretation of a Local Average Treatment Effect (LATE), i.e. the returns to one year of university education for a marginal student who would not go to university in the absence of the HE expansion.

Table 2: Baseline Results

	(1)	(2)	(3)	(4)	(5)
A): Treatment = Years of education (N=2,540)					
Years of educ	1.319**	1.385***	1.048**	0.798**	1.016**
	(0.645)	(0.513)	(0.423)	(0.405)	(0.387)
Monthly income	0.178*	0.129**	0.191**	0.256*	0.213**
	(0.095)	(0.061)	(0.091)	(0.140)	(0.098)
B): Treatment = university dummy (N=2,540)					
HE Degree	0.205**	0.201***	0.150**	0.092*	0.125**
	(0.095)	(0.074)	(0.063)	(0.052)	(0.056)
Monthly income	1.124*	0.831**	1.330**	1.639	1.504*
	(0.603)	(0.412)	(0.656)	(1.043)	(0.772)

Note: ***, **, * indicate statistical significance at the 1, 5 and 10 percent level respectively. Sample of Junior High School and University graduates. Standard errors are clustered at province level. (1) with no controls; (2) controls for wave of investigation and province; (3) further controls for gender, dummy of migrant, experience and square of experience; (4) further controls for region of birth, *hukou* status, interaction terms of region of birth and *hukou* status, education of parents, numbers of siblings; (5) further controls the status of self-employed or employed.

5.2 Robustness checks

Table 3 represents a falsification test using either September 1978 or September 1980 as the cut-off point. As expected, neither specification yields significant results for years of education or monthly income.

Table 3: Falsification using Alternative Birth Cut-offs

	(1)	(2)	(3)	(4)	(5)
A): Cut-off at Sept 1978 (N=2,514)					
Years of educ.	-0.252	0.128	0.023	-0.250	-0.319
	(0.526)	(0.472)	(0.352)	(0.307)	(0.318)
Monthly income	1.243	-0.741	-2.306	0.296	0.271
	(2.404)	(3.020)	(36.994)	(0.380)	(0.303)
B): Cut-off at Sept 1980 (N=2,547)					
Years of educ.	-0.105	-0.097	0.056	-0.113	-0.008
	(0.505)	(0.443)	(0.388)	(0.327)	(0.301)
Monthly income	0.833	0.795	-1.346	0.844	-10.960
	(3.621)	(3.407)	(9.913)	(2.379)	(406.230)

Note: ***, **, * indicate statistical significance at the 1, 5 and 10 percent level respectively. Standard errors are clustered at province level.

Table 4 checks the robustness of the results with respect to narrower (30) or wider (42) window widths, instead of 36 months, on either side of the cut-off points. The patterns are remarkably similar to the baseline estimates in Table 2. This is very reassuring, and suggests that our findings are insensitive to variations in the window widths.

Table 4: Robustness w.r.t. Alternative Windows

	(1)	(2)	(3)	(4)	(5)
A): +/-30 months (N=2,082)					
Years of educ.	1.370**	1.381***	0.998**	0.884**	1.030**
	(0.595)	(0.510)	(0.409)	(0.431)	(0.411)
Monthly income	0.148*	0.128**	0.188**	0.245*	0.212**
	(0.081)	(0.061)	(0.091)	(0.130)	(0.097)
B): +/-42 months (N=2,841)					
Years of educ.	1.368**	1.382***	1.052**	0.615*	1.021**
	(0.617)	(0.511)	(0.424)	(0.359)	(0.408)
Monthly income	0.157*	0.128**	0.191**	0.258	0.213**
	(0.085)	(0.061)	(0.090)	(0.164)	(0.098)

Note: ***, **, * indicate statistical significance at the 1, 5 and 10 percent level respectively. Standard errors are clustered at province level.

Table 5 repeats Table 2, but only using the subsample of employees. While the years of education increases somewhat to 1.3 years, there is virtually no change in the returns to education estimate that remains at 0.19.

Table 5: Robustness w.r.t. Alternative Samples, Employees only (N=1,817)

	(1)	(2)	(3)	(4)
Years of educ.	1.665**	1.765**	1.298**	0.792**
	(0.793)	(0.713)	(0.515)	(0.398)
Monthly income	0.176**	0.136**	0.194**	0.226*
	(0.084)	(0.060)	(0.081)	(0.122)

Note: ***, **, * indicate statistical significance at the 1, 5 and 10 percent level respectively. Standard errors are clustered at province level.

Table 6 tests for robustness with respect to higher order polynomials. Again, there is a small increase in the effect on years of education. However, each year of education induced by the HE expansion still increases the monthly income by approximately 0.19 log point.

Table 6: Robustness w.r.t. Functional Forms, 2nd order polynomial (N=2,540)

	(1)	(2)	(3)	(4)	(5)
Years of educ.	1.195	1.515**	1.126**	0.847	1.057**
	(0.789)	(0.604)	(0.470)	(0.521)	(0.478)
Monthly income	0.252	0.121*	0.185**	0.262	0.206*
	(0.155)	(0.065)	(0.093)	(0.172)	(0.110)

Note: ***, **, * indicate statistical significance at the 1, 5 and 10 percent level respectively. Standard errors are clustered at province level.

Using a meta-analysis approach, Zeng et al (2014) present a comprehensive review of studies on gender inequality in education in China. They conclude that gender inequality has been narrowing since the 1980s, and is now mostly confined to post-compulsory education stage in rural areas. Unfortunately, there seems to be no direct evidence available on the effect of the HE expansion on educational gender inequality.

Table 7 presents the heterogeneous effect with respect to gender and whether the respondent is the only child. While the effect appears to be weaker for women on average, the effect for the female only child is as strong as for the sample as a whole.[11] For men, the estimates are higher but rather imprecisely determined. Note that estimating the returns for men and women separately also gets around the problem of discontinuity in gender around the cut-off point shown in Appendix Figure A2, but at the expense of lower statistical precision.

[11] The family planning policy in China, often known as the one-child policy, was not strictly enforced in the most parts of the country until the 1980s (Wang and Zhao 2012). For the purpose of our analysis, what matters is the fact that there is no apparent discontinuity in the birth rates around our HE expansion cut-off point.

Table 7: Robustness w.r.t. Gender and Single-child Status

	(1)	(2)	(3)	(4)	(5)
A): Female (N=1,212)					
Years of educ.	1.281	1.208	1.425***	0.651	0.486
	(0.953)	(0.769)	(0.537)	(0.540)	(0.557)
Monthly income	0.175	0.176	0.157*	0.259	0.327
	(0.131)	(0.123)	(0.088)	(0.256)	(0.397)
B): Female Only Child (N=705)					
Years of educ.	2.853***	2.702***	1.756***	0.725	0.968**
	(0.994)	(0.749)	(0.585)	(0.484)	(0.520)
Monthly income	0.112*	0.109**	0.208**	0.378	0.307
	(0.068)	(0.055)	(0.096)	(0.280)	(0.194)
C): Male (N=1,328)					
Years of educ.	1.123*	1.204**	0.484	1.173**	1.199**
	(0.596)	(0.599)	(0.553)	(0.630)	(0.556)
Monthly income	0.130	0.127	0.275	0.243*	0.220*
	(0.104)	(0.079)	(0.307)	(0.136)	(0.115)

Note: ***, **, * indicate statistical significance at the 1, 5 and 10 percent level respectively. Standard errors are clustered at province level.

Table 8 presents the heterogeneous effect by *hukou* status at the time of primary education. We did not use the current *hukou* status which might be contaminated by endogenous migration after completing formal education (see Xing 2013). Unfortunately, the estimates are all poorly determined due to the small cell sizes except for urban residents, for whom the effect on years of schooling are not only large at 1.5 years, but also statistically significant at the 5% level. This pattern is consistent with Xing (2013) who finds much lower OLS returns to education in rural areas than in urban areas using the China Household Income Project (CHIP) data from 1995, 2002 and 2007.

Table 8: Robustness w.r.t. *Hukou* Status

	(1)	(2)	(3)	(4)	(5)
A): Village (N=1,119)					
Years of educ.	-0.071	0.207	0.567	0.507	0.619
	(0.547)	(0.512)	(0.452)	(0.424)	(0.439)
Monthly income	0.386	0.776	0.385	0.413	0.372
	(2.765)	(1.801)	(0.300)	(0.323)	(0.267)
B): Town (N=599)					
Years of educ.	0.826	-0.196	-0.093	-0.433	-0.102
	(1.052)	(0.915)	(0.854)	(0.685)	(0.702)
Monthly income	0.192	0.949	1.703	0.496	1.403
	(0.276)	(4.216)	(15.315)	(0.631)	(9.384)
C): City (N=822)					
Years of educ.	2.260**	2.179***	1.453**	1.766**	1.434**
	(0.988)	(0.795)	(0.721)	(0.742)	(0.649)
Monthly income	0.109	0.120*	0.134	0.122	0.151
	(0.094)	(0.069)	(0.103)	(0.092)	(0.110)

Note: ***, **, * indicate statistical significance at the 1, 5 and 10 percent level respectively. Standard errors are clustered at province level.

Finally, Table A2 in the Appendix explores the effect of the 1999 HE expansion on the relative returns of Senior High School to Junior High School, and the relative returns of university to Senior High School respectively. While the estimates are both positive, none of them are statistically significant at any conventional level. On the other hand, the HE expansion appears to have increased the years of schooling among Junior and Senior High School leavers by almost 0.4 year, with a 10% level of significance. We interpret this as evidence of a significant spill-over effect of the HE expansion on Senior High School completion, which is a prerequisite for HE enrolment. This finding lends further support to our choice of the control group. On the other hand, while the magnitude of this spill-over effect is of interest in itself, our modest sample size precludes precise estimates unfortunately.

5.3 Comparison with existing studies

Compared to the existing quasi-experimental estimates of returns to education in China, our returns to education estimate of 21% is on the high side, and significantly higher than e.g. the 12.8% RDD estimate by Liu et al. (2016), although it is not far off the 20% IV estimate by Fang et al. (2012). Note that both of these studies are based on the variation in time across provinces in the implementation of 9-year compulsory education.

There are good reasons to expect a higher estimate in our case:

a) There is a general consensus that the returns to education in China is higher at higher qualification levels, presumably reflecting the relative scarcity of more skilled labour (Awaworyi and Mishra (2014)). Therefore, our LATE estimates on the returns based on higher education expansion should be higher than LATE estimates based on compulsory schooling reforms.

b) There is also consensus that returns to education in China are increasing over time (see also Awaworyi and Mishra (2014)). The meta-analysis by Liu and Zhang (2013) concludes that return to education in China since the economic reform started has increased by approximately 0.2 percentage points a year till the onset of the Global Financial Crisis. Since our sample members are interviewed in 2005 and 2010, we would expect higher returns than studies based on older data, including Wu and Zhao (2010).

c) Using potential expansion based on pre-expansion provincial enrolment share as instruments for the supply of young skilled worker, Li et al (2017) find that the 1999 expansion increased the college premium of workers aged 25 or above whereas it decreased the college premium of those aged 20-24. For instance, the IV estimates of the college premium for those aged 20-24 and 25-29 are 0.164 and 0.370 respectively. Since our sample members interviewed in 2005 and 2010 have a mean age of 26 and 31 respectively, we would also expect higher returns than studies such as Wu and Zhao (2010) which is based on samples of younger graduates collected in 2002 and 2005.

d) Compared to Wu and Zhao (2010) who also exploit the HE expansion but uses DD and DDD, it is conceivable that LATE are higher than ATE, e.g. if marginal students are better motivated, have higher unobservable skills or more favourable family background. Our results are also consistent with Che and Zhang (2017), who suggest that more human-capital intensive industries in China experienced more rapid growth in total factor productivity after 2003 because the HE expansion substantially relaxed the constraint of skilled labour.

Section 6: Concluding Remarks

We take advantage of the substantial HE expansion, which started in 1999 to study the returns to higher education in China. Using pooled China Urban Labour-force Survey (CULS) 2005 and 2010 which contains year and month of birth, we are able to apply fuzzy RDD which relies on weaker assumptions for identification compared to other quasi-experimental methods such as Difference-in-Differences or Instrumental Variables.

For our sample of people with either HE or Junior High School qualifications and born within 3 years of the relevant month-of-birth cut-off, our RDD results indicate that the mean years of education increased by more than one full year around the cut-off point. Importantly, each additional year of university education induced by the 1999 higher education expansion increases monthly wage income by about 21%, which is much higher than the corresponding OLS estimate of 8%. Our findings are insensitive to alternative window widths, functional forms, or the exclusion of self-employed. Moreover, the returns to degrees also appear to vary by gender, with lowers returns to women except when they are the only child in the family.

Nearly 20 years on, the overall assessment of the massive expansion starting 1999 still remains controversial. Our RDD estimate of such a high return to higher education seems to provide strong *ex post* justification of the HE expansion, at least from a pure cost-benefit perspective. However, our Local Average Treatment Effect (LATE) estimate is based on the comparison of 3 school-year cohorts on each side of the month-of-birth cut-off, and only applies to marginal students who would not go to university in the absence of the HE expansion. It would be heroic to try to generalize our findings to more recent cohorts of graduates, who have not only been exposed to sustained HE expansion, but also are faced with a much weaker labour market following the Global Financial Crisis. With sustained expansion of the HE sector, one might speculate that returns to higher education will eventually decrease. However, a full investigation on the extent, if any, of a decline in the college premium over time is beyond the scope of this paper.

One limitation of our study is the inability to explore in greater detail the spill-over effect of the HE expansion on upper-secondary education including the vocational track, or the full scale of the heterogeneous effect of the expansion on returns to education along dimensions such as *hukou* status and gender, due to the modest sample size. This is particularly important for a good understanding of the equity implications of the HE expansion.

Nevertheless, this study represents an important first attempt to estimate the returns to HE, in a developing country context, by exploring an unprecedented HE expansion reform in China. A key lesson to be learnt is that relaxing constraint of skilled labour could be conducive to development under some favourable conditions. On the other hand, further studies using larger surveys will be needed to fully understand the distributional effects of such HE expansion reforms.

References

Awaworyi, S. and Mishra, V. (2014) Returns to education in China: A meta-analysis. *Monash University Department of Economics Discussion Paper* 41/14.

Blundell, R.W., Dearden, L., Goodman, A. and H. Reed. (2000) The returns to higher education in Britain. *Economic Journal*, 110, F82-F99.

Blundell, R.W., Green, D.A. and Jin, W. (2016) The UK wage premium puzzle: how did a large increase in university graduates leave the education premium unchanged? *IFS Working Paper* W16/01.

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, *82*(6), 2295–2326.

Calonico, S., M.D. Cattaneo, Farrell, M. H., & Titiunik, R. (2017). rdrobust: software for regression-discontinuity designs. *Stata Journal*, *17*, 372-404.

Card, D. and Lemieux, T. (2001) Can Falling Supply Explain the Rising Return to College for Young Men? *Quarterly Journal of Economics 116*, 705–746.

Che, Y. and Zhang, L. (2017) Human capital, technology adoption and firm performance: Impacts of China's Higher Education expansion in the late 1990s. *Economic Journal* (*early view*) https://onlinelibrary.wiley.com/doi/abs/10.1111/ecoj.12524.

National Bureau of Statistics (NBS) of China, China Statistic Yearbook, various years.

Fang, H. , Rizzo, J.A. , Rozelle , S. and Zeckhauser , R.J. (2012) The Returns to Education in China: Evidence from the 1986 Compulsory Education Law, *NBER Working Paper* No. 18189.

Gelman, A., & Imbens, G. (2017). Why high-order polynomials should not be used in regression discontinuity designs. *Journal of Business & Economic Statistics (early view)*. DOI: 10.1080/07350015.2017.1366909.

Goldin, C. and Katz, L. (2008) *The Race between Education and Technology*. Harvard University Press, Harvard.

Hahn, J., Todd, P, Van der Klaauw, W. (2001) Identification and estimation of treatment effects a Regression-Discontinuity Design. *Econometrica* 69 (1), 201-209.

Knight, J., Deng, Q. and Li, S. (2017) China's expansion of higher education: The labour market consequences of a supply shock. *China Economic Review 43*: 127-141.

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

Li, H., Ma, Y., Meng, L., Qiao, X and Shi, X. (2017) Skill complementarities and returns to higher education: Evidence from college enrollment expansion in China. *China Economic Review* 46, 10-26.

Li, S., Whalley, J. and Xing, C. (2014) China's higher education expansion and unemployment of college graduates. *China Economic Review* 30, 567-582.

Liu, E. and Zhang, S. (2013) A meta-analysis of the estimates of returns to schooling in China. Mimeo. University of Houston.

Liu, S., Zhou, S. and Hu, A. (2016) Compulsory Education Law and economic return to education in urban China: Based on Regression Discontinuity Design. *Economic Research* 2016 (2), 154-167 (*in Chinese*).

OECD (Organisation for Economic Co-Operation and Development) (2016) Education in China – A Snap Shot. OECD.

Walker, I. and Zhu, Y. (2008) The College Wage Premium and the Expansion of Higher Education in the UK. *Scandinavian Journal of Economics* 110 (4), 695–709.

Wan, Y. (2006) Expansion of Chinese Higher Education since 1998: Its causes and outcomes. *Asia Pacific Education Review* 7 (1), 19-31.

Wang, J. and Zhao, B. (2012) The recognition of the family planning policy from the public policy perspective. *Population Journal* 194(4), 81-89 (*in Chinese*).

Wu, Y. and Zhao, Q. (2010) Higher Education expansion and employment of university graduates. *Economic Research* 2010 (9), 93-108 (*in Chinese*).

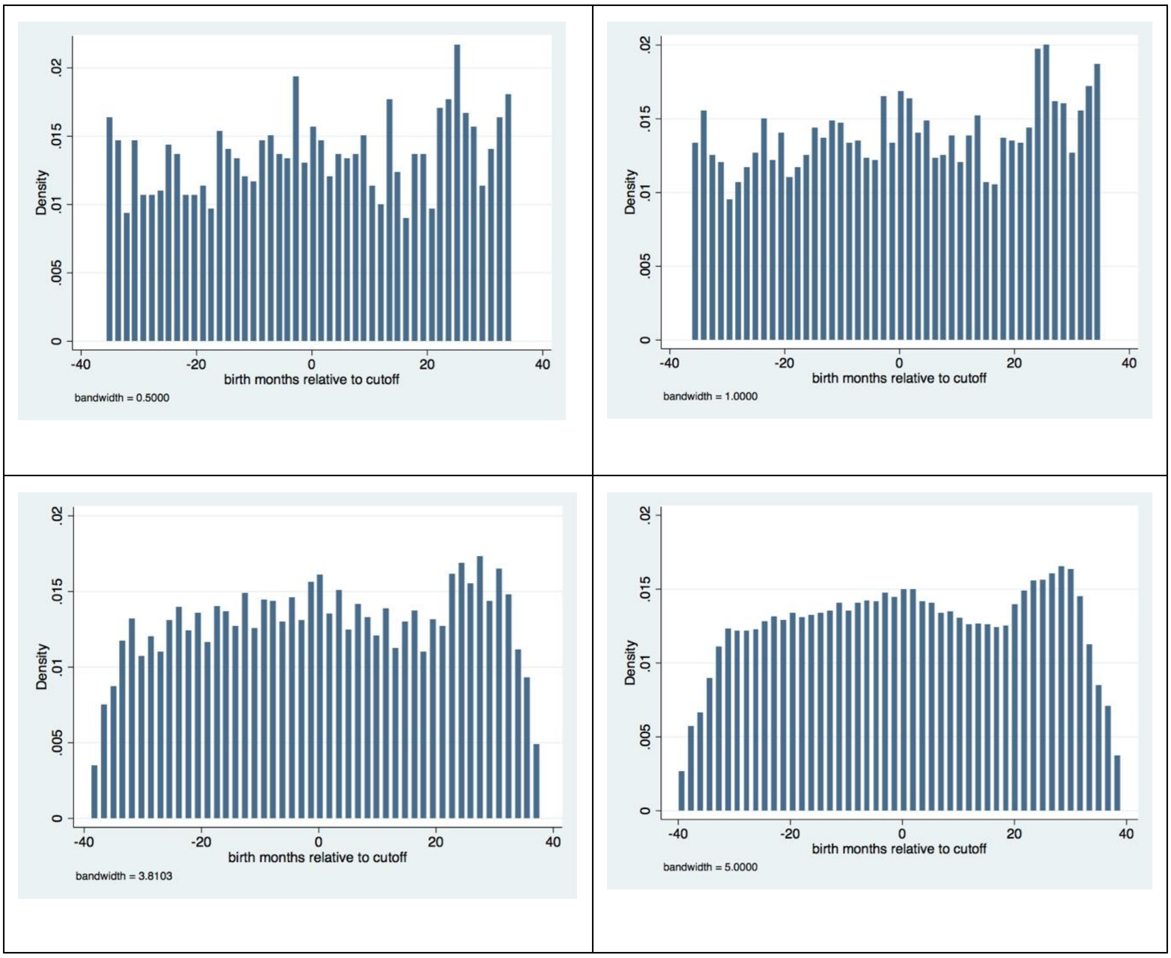
Xing, C. (2013) Education expansion, migration and rural-urban education gap: A case study on the effect of university expansion. *China Economic Quarterly* 13(1), 207-232 (*in Chinese*).

Xing, C., Yang, P. and Li, Z. (2017) The medium-run effect of China's higher education expansion on the unemployment of college graduates. *China Economic Review,* forthcoming.

Zeng, J., Pang, X., Zhang, L., Medina, A. and Rozelle, S. (2014) Gender inequality in education in China: A meta-regression analysis. *Contemporary Economic Policy* 32(2), 474-491.

Appendix

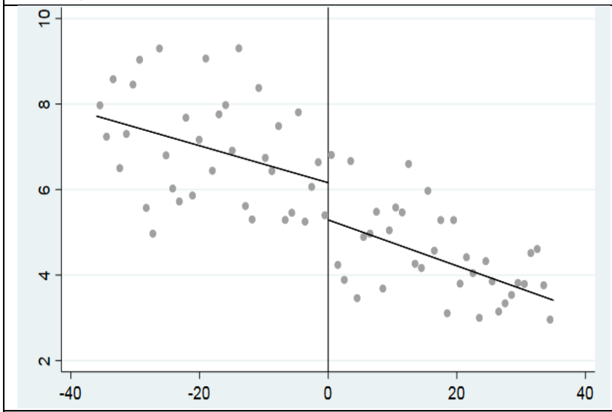
Figure A1: Validation test



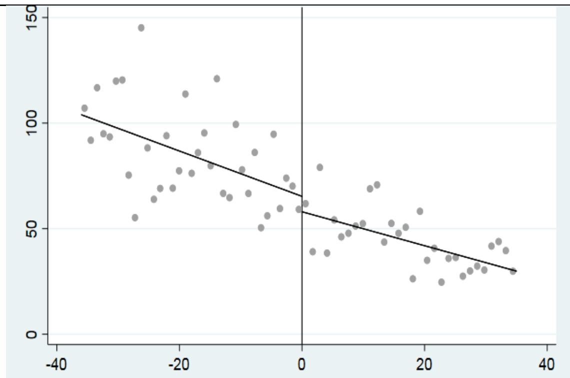
Note: we estimate the kernel density of the running variable of months of birth. The graphs indicate a lack of discontinuity around the cutoff in the running variable, thus satisfying the critical RDD assumption of no non-random sorting into treatment status. Our results are insensitive to different bandwidths at 0.5, 1, 3.8 and 5 months.

Figure A2: Continuity test of covariates

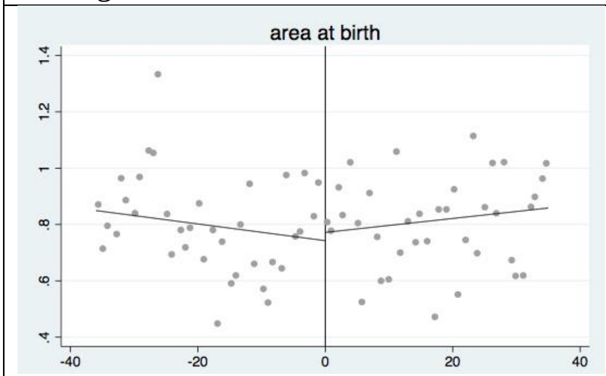
A：gender



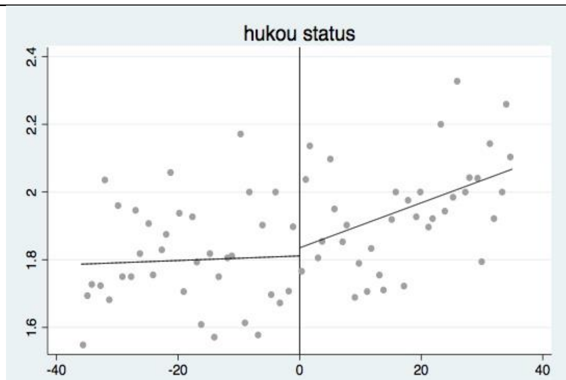
B：experience



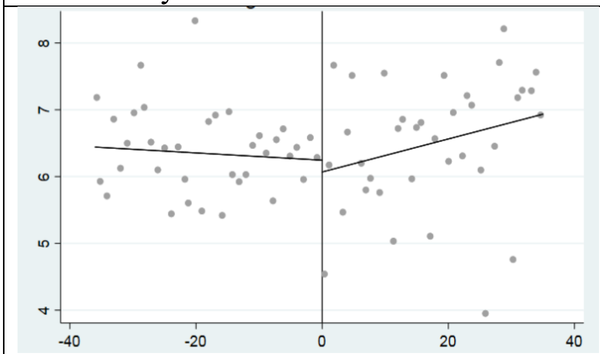
C：region of birth



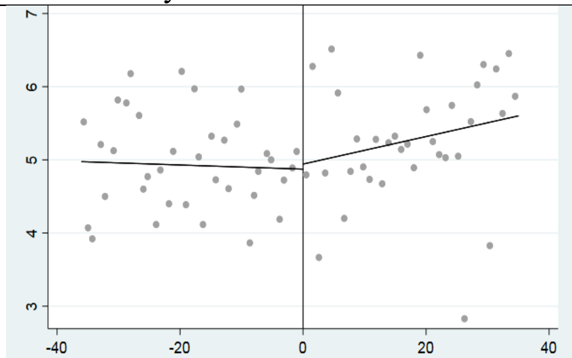
D: *hukou* status (location of primary school)



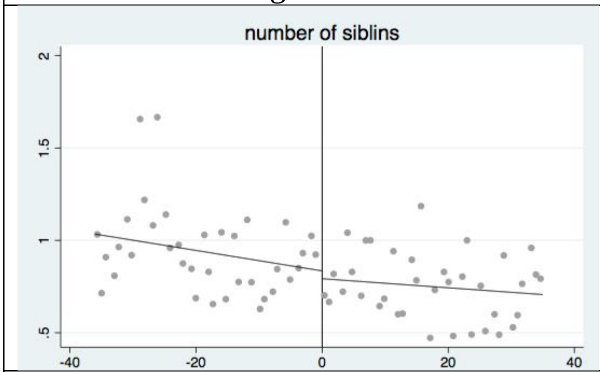
E: father’s years of education



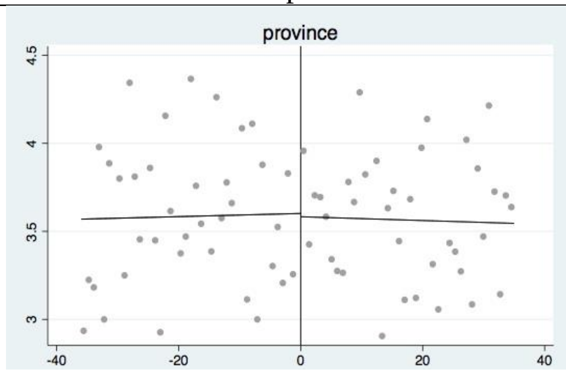
F: mother’s years of education



G: number of siblings

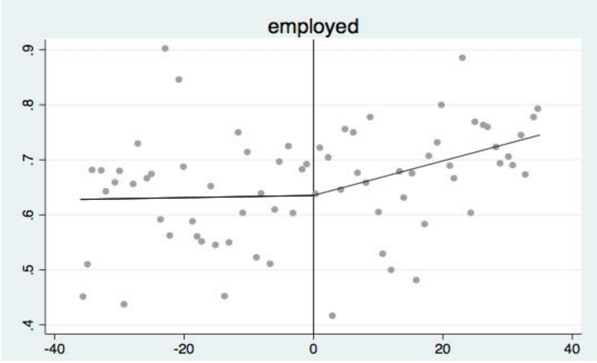


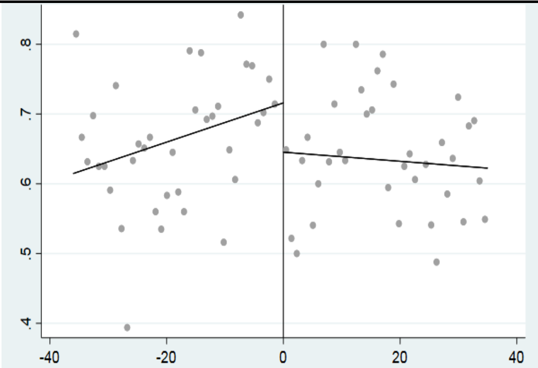
H: current residential province



I: employed

J: migrant





Note: We use the Stata programme *rdrobust* to perform the continuity test of covariates. All the control variables except for gender satisfy the requirement of continuity around the cut-off. Table 7 reports the RDD estimates for men and women separately, at the expense of lower precision.

Table A1: OLS Estimates, N=2,540

	(1)	(2)	(3)	(4)	(5)
Years of educ.	0.103***	0.066***	0.080***	0.075***	0.078***
	(0.013)	(0.011)	(0.011)	(0.011)	(0.011)

Note: ***, **, * significant at the 1, 5 and 10 percent level. Standard errors are clustered at province level. Model specifications same as in Table 2.

Table A2: Alternative Samples

	(1)	(2)	(3)	(4)	(5)
A): Junior High and Senior High School graduates (N=2,287)					
Years of educ.	0.220	0.346	0.383*	0.365	0.284
	(0.225)	(0.274)	(0.230)	(0.248)	(0.205)
Monthly income	0.173	0.009	0.143	0.104	0.132
	(0.467)	(0.266)	(0.206)	(0.229)	(0.249)
B): Senior High and University graduates (N=2,137)					
Years of educ.	0.345	0.610	0.321	0.263	0.288
	(0.364)	(0.264)	(0.248)	(0.222)	(0.220)
Monthly income	0.070	0.160	0.375	0.438	0.394
	(0.399)	(0.145)	(0.346)	(0.435)	(0.373)

Note: ***, **, * significant at the 1, 5 and 10 percent level. Standard errors are clustered at province level. Model specifications same as in Table 2.