Journal of Labor Research (2019) 40:463–486
https://doi.org/10.1007/s12122-019-09293-0

Does College Education Promote Entrepreneurship in China?

Tianshu Chu[1] · Qiang Wen[2]

Published online: 14 September 2019

Abstract
There is no consensus on the impact of education on entrepreneurial choice in both theory and empirics. China's *Higher Education Expansion* (HEE) policy initiated in 1999 provides us a unique opportunity to identity the causal relationship between college education and entrepreneurship by exploiting the Fuzzy Regression Discontinuity Design (FRDD) approach. In this paper, we use the *China Household Income Project* (CHIP) 2013 database, finding that China's HEE policy significantly increases the probability of obtaining college education by 12%. There is suggestive evidence that college education decreases overall and *self-employed-type* of entrepreneurial choices, but increases *boss-type* activities; none of the coefficients are precisely estimated, though. Our results are robust to different inference approaches.

Keywords College education · Entrepreneurship · Higher education expansion policy · Fuzzy regression discontinuity design

JEL Classification I25 · J24 · L26

Introduction

Entrepreneurs play important roles in job creation, productivity growth, and innovations (van Praag and Versloot 2007), it is not surprising that personal traits determining entrepreneurship are widely explored in both theoretical (de Wit 1993; Evans and Jovanovic 1989; Lazear 2005) and empirical (Blanchflower and Oswald 1998; de Wit 1993; Evans and Leighton 1989; Le 1999; Simoes

✉ Qiang Wen
durianwen@xmu.edu.cn

[1] Research Institute of Economics and Management, Southwestern University of Finance and Economics, Chengdu, China

[2] The Wang Yanan Institute for Studies in Economics, Xiamen University, Xiamen, China

et al. 2016) analyses. Among the examined determinants of entrepreneurial choice, the impact of education is highly policy-relevant yet remains theoretically ambiguous. This study explores this impact by estimating the causal effect of education on a person's probability of becoming an entrepreneur using a unique policy setting in China.

Theoretically, the net impact of education on entrepreneurial choice is ambiguous. On the one hand, people with higher education may more likely engage in entrepreneurship: Education may enhance managerial ability (Lucas 1978), as well as the ability to identify entrepreneurial opportunities (van der Sluis et al. 2008), increasing the propensity to becoming an entrepreneur. On the other hand, higher levels of education may generate better employment options, such as higher compensation and better working condition, reducing the desirability of entrepreneurship (van der Sluis et al. 2008). These two effects offset each other and the sign and the magnitude of the net effect is an empirical question.

Empirically, the impact of education on entrepreneurial choice can also run either direction. As a result, it is not surprising that the impact of education on entrepreneurship selection is inconclusive (Simoes et al. 2016). van der Sluis et al. (2005) conduct a meta-analytical review on the impact of schooling on entrepreneurship selection in developing countries and find that ambiguous. Another meta-analysis by van der Sluis et al. (2008) is on the effect of formal education on entrepreneurship selection in industrialized countries, finding that the impact of education is insignificant. Essentially, existing studies find both positive and negative impacts of education on entrepreneurial choice in both developing and industrialized countries. One of the reasons that the literature is inconclusive may be due to the fact that the endogeneity issue of education in the studies of entrepreneurial choice is insufficiently addressed (van der Sluis et al. 2008). Endogeneity bias occurs when individuals choosing different levels of education differ systematically in unobserved characteristics that also affect their occupational choice. The causal effect of education is well explored in the literatures on earnings (Card 1999), however, the causal impact in the entrepreneurial choice researches is limited. Block et al. (2013) study the causal impact of education on entrepreneurial choice in 27 European countries and the USA. Few studies are found in developing countries. Clearly, more studies on causal impacts of education on entrepreneurship are needed.

In this paper, we pay special attention to the impact of college education on entrepreneurship in China. To identify the causal effect of education on entrepreneurial choice, we take advantage of a unique quasi-experiment – China's *Higher Education Expansion* (HEE) policy initiated in 1999. In the first half year of 1999, China's central government suddenly decided to increase the number of students admitted to colleges and universities by 0.55 million, as the result, the college entrance rate increased to more than 55.5% in 1999 (33.9% in 1998). From then on the enrollment rate of college education is usually greater than 60%, almost twice as large as the rate of the period 1978–1998. According to China's regulation on school age, students born after September 1980 were directly affected by the HEE policy (S. Li et al. 2014), and their probability of obtaining college education became much larger than that of students born before September 1980. Therefore, China's HEE policy presents an ideal setting to study causal impact of education on entrepreneurial choice, specifically it is suitable for using the *fuzzy regression discontinuity design* (FRDD) approach.[1]

We use the *China Household Income Project* (CHIP) 2013 database for the empirical analyses. A person is defined as an entrepreneur if this person's employment status is either "employer" or "self-employment".[2] Our FRDD results show that: (1) China's HEE policy has a significant positive impact on college education for a person born after Sep. 1980, the probability of obtaining college education significantly increases by 12%; (2) the impact of college education on the probability of a person being an entrepreneur is not statistically significant. Our identification passes the general validity tests, and our results are robust to alternative choices of the *bandwidth* and the *orders of polynomial* in local polynomial estimations and choices of *variance* inference procedures.

Employer (Boss-type) and self-employment (Self-employed-type) are two different types of entrepreneurs, where the former contributes to the whole economy more significantly, and economic theory pertaining to entrepreneurship concerns much more about the private enterprises (Zhang and Li 2016). We examine the potential heterogeneous impacts of college education on these two types of entrepreneurial choices. Our data results agree with Levine and Rubinstein (2017)'s finding that boss-type entrepreneurs tend to be more educated. The intuition behind may be that private enterprises are of more complex organizational structures relative to self-employed type of businesses, and that formal education can enhance an entrepreneur's capacity in managing complex business structures. Empirical results show that the signs of coefficients are consistent with intuition: the effect of college education on the *Self-employed-type* entrepreneurship is negative, whilst the effect on employer is positive. However, none of the coefficients are precisely estimated to draw a definitive statistical conclusion.

The reminder of this paper is structured as follows. Section 2 briefly introduces some related researches pertaining to exploring the impact of education on China's entrepreneurship and China's *Higher Education Expansion* policy; Section 3 discusses our identification strategy; Section 4 presents the data facts; Section 5 reports the main results, robustness checks as well as heterogeneity effects; Section 6 concludes this paper.

Literature Review and China's Higher Education Expansion Policy

Literature Review

Even though there exists numerous researches exploring the impact of education on entrepreneurial choice in China, the question remains inconclusive. Some of these researches find that education would promote entrepreneurship. For instance, Mohapatra et al. (2007) examine the determinants of self-employment in rural China

[1] See Lee and Lemieux (2010) and Skovron and Titiunik (2016) for practical guides to the implementation of the regression discontinuity design approach.

[2] This definition is consistent with Li and Wu (2014)'s research on China's entrepreneurship.

using an original survey database, finding that years of schooling has a significant positive impact on the probability of being self-employed.

In contrast, some other researches illustrate that education would impede entrepreneurial choice. Yueh (2009a) and Yueh (2009b) explore the determinants of entrepreneurs in urban China using a urban household survey database conducted by China's National Bureau of Statistics in 2000, finding that years of schooling has a significant negative impact on entrepreneurial choice. Lu and Tao (2010) examine the determinants of entrepreneurial activities in China using a life-histories survey database conducted in 1994 in twenty cities, reporting that years of education has a significant negative impact on entrepreneurial choice. Using the 2006 and 2008 *China General Social Survey* (CGSS) database, Sun et al. (2016) show that years of education has a significant negative impact on entrepreneurial choice for ethnic Han in urban China. Using a random sample of the 2005 1% *Population Sample Survey of China* database, Chu and Wen (2017) find that years of schooling has a significant negative impact on entrepreneurial choice.

Besides, there are yet researches showing that the impact of education on entrepreneurship is either insignificant or not robust. For example, Démurger and Xu (2011) explore the determinants of return migrants' self-employment decision in rural China using an original rural household survey database conducted in Wuwei County (Anhui province, China) in 2008, finding that years of schooling has no significant impact on entrepreneurial choice. Beck et al. (2015) examine the impact of access to external finance on entrepreneurship in rural China using the *Rural Finance Survey* database conducted by Peking University in 2009, finding that the impact of senior education (high school or above) has no robust impact on entrepreneurial choice.

One possibility for these mixed findings is that the endogeneity issue of education in entrepreneurial choice is not addressed. As discussed in the introduction section, if some unobserved characteristics affect education and occupational choice simultaneously, conclusions of above studies may suffer from endogeneity bias. In this paper, we examine the causal impact to address this potential endogeneity issue. As discussed in later subsection, China's government initiates one education policy in late 1990s, which significantly increases the probability of some birth cohorts obtaining college education, so we can use such policy as exogenous shock to identify the causal impact of education on entrepreneurship. Joining the effort with these insightful researches, focusing on this education policy, we implement the *Fuzzy Regression Discontinuity Design* approach to explore the causal impact of college education on entrepreneurial choice.

China's Higher Education Expansion Policy

In China, almost all high school seniors take the *National Higher Education Entrance Examination* at the end of spring semester (it is usually hold in June since 2003[3]). The exam is given once per year at the same time across the nation, commonly known as the *Gaokao* in Chinese, for admission into colleges in the coming fall semester. China's

[3] Prior to 2003, the examination was held in July, but has since been moved to the month of June in consideration of the adverse effects of hot weather on students living in southern China and possible flooding during the rainy season in July.

higher education system has been mostly centralized and the Ministry of Education sets the annual quota. The total new enrollments of college students increases stably at an annual growth rate of 8.5% for the period of 1978–1998. However, to alleviate the employment pressure resulted from the reform of state-owned enterprises started at 1998 as well as the Asian financial crisis started at 1997 and to stimulate the household consumption (S. Li et al. 2014; Yeung 2013),[4] China's central government initiated the *Higher Education Expansion* policy in the first half year of 1999, which leads to sharp increases in the total new enrollments and the enrollment rate of college education in 1999. As shown in the Fig. 1, the total new enrollments and the enrollment rate of college education are 1.1 million and 33.9%, respectively, in the fall semester of 1998, and these figures have increased to 1.6 million and 55.5% in 1999. Right after the implementation of the HEE policy, the enrollment rate of college education becomes around 60%, doubling the rate of the period of 1978–1998. Obviously, it forms a great "jump" in the probability of a typical student obtaining the college education in China before and after 1999.

One feature of the HEE policy is the unexpectedness (S. Li et al. 2014). In the early 1999, the central government suddenly decided to increase the total enrollments of college student by 0.22 million, and the central government and the Ministry of Education together announced an additional 0.33 million new admissions in June 1999. This expansion of college enrollments were announced in the eve of *Gaokao*, totally unexpected for most high school seniors and their families. Therefore, this policy constitute a "natural" experiment. Another feature of the HEE policy is that not all students benefit from the policy. In China, students take 12 years of primary and middle school education before the *Gaokao*, and the entrance age of primary school is six years old (S. Li et al. 2014; Xing et al. 2018).[5] Therefore, students born after September 1980 were the first birth cohort affected by the HEE policy, and their probability of obtaining college education became significantly higher than that of students born before September 1980.

Some researches point out that the HEE policy leads to more workers with college degrees after four years of its implementation, generating a significant "shock" to the labor supply across the country. For instance, S. Li et al. (2014) use 2000 and 2005 Population Census data to evaluate the short-run effect of the HEE policy on unemployment, using the difference-in-difference (DID) approach. They find that the HEE policy has sharply increased the unemployment rate by 6%–9% among young college graduates. Xing et al. (2018) focus on the medium-run effect and use 2000, 2005 and 2010 Population Census data and the same DID approach, they find that the HEE policy increases the unemployment rate of new college graduates in the short run, however, the unemployment effect mostly disappears after five years. Recently, Knight et al. (2017) investigate the impact of the HEE policy on wage, unemployment, and access to "good jobs" simultaneously using CHIPs database and the same DID approach. Obviously, these studies share one common feature that they treat the HEE policy as an exogenous shock and utilize the DID approach to identify treatment effect.

[4] Another two important purposes are enhancing international competitiveness with a more skilled labor force and meeting public demands for higher education (Yeung 2013; Wu and Zhang 2010).

[5] Only few provinces (mainly the minority Autonomous Regions) set the entrance age of primary school to 7 years old.

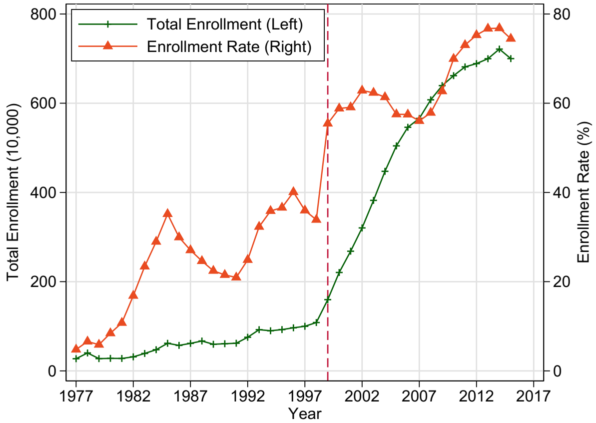


Fig. 1 Total Enrollments and Enrollment Rate of College Education in China

The studies using DID approach based on the HEE setting have only short-panel dataset, causing difficulties in testing the "parallel trend" assumption, the critical assumption validating the DID strategy. Differing from the DID strategy, we make the full use of the special feature that the HEE policy creates a "discontinuity" in probability of obtaining college education for students from different birth cohorts. This discontinuity feature permits the use of *Regression Discontinuity Design* (RDD) approach to examine the impact of college education on occupational choice. Our identification strategy is similar to Qin et al. (2017)'s work exploring the impact of being a one-child on educational attainments by using China's One Child policy. In their setting, the one-child policy which requires the majority of parents to have only one child is strictly implemented in urban Han population, and people born after the implementation of this policy have a higher probability of being the only child. The policy was initiated during the late 1979 to the early 1980 period, considering the 10-month gestation period, it can be calculated that people born in October 1980 is the first birth cohort affected by this policy. We elaborate in greater details our RDD identification strategy using the HEE policy setting in the next section.

Empirical Strategy

As discussed above, the implementation of the higher education expansion policy exhibits two features: first, there is a sharp "jump" in the enrollment rate; secondly, the HEE policy does not sharply change the probability of obtaining college education from 0 to 1 for different birth cohorts around the policy cutoff. These features allow identification of the casual effect of college education on entrepreneurial choice by utilizing the fuzzy regression discontinuity design approach. On the one hand, sharp "jump" in the enrollment rate implies that the probability of obtaining college education is quite different for students born before and after September 1980, indicating that a crucial assumption for the RDD approach is satisfied. On the other hand, the HEE policy doesn't sharply change the probability of obtaining college education from 0 to 1 for different birth cohorts around the policy cutoff, which implies that the *fuzzy* RDD should be used to identify the impact of college education on entrepreneurial choice.

Formally, the impact of college education on entrepreneurial choice can be identified as follows: there exists N persons, indexed by $i = 1, \ldots, N$. For each person, we can observe her/ his realized entrepreneurial choice status y_i ($y_i = 1$ if this person is an entrepreneur, otherwise, $y_i = 0$), college education status T_i ($T_i = 1$ if this person obtains college education, otherwise, $T_i = 0$) and birth year and month X_i. For person i, if she/he obtains college education, then this person is assigned to the "Treatment group", and the potential entrepreneurial choice is $y_i(1)$, otherwise, this person is assigned to the "Control group", and the potential entrepreneurial choice is $y_i(0)$. Therefore, the observed entrepreneurial choice can be expressed as a function of education status: $y_i = (1 - T_i) \cdot y_i(0) + T_i \cdot y_i(1)$, i.e., $y_i = y_i(0)$ when $T_i = 1$, $y_i = y_i(1)$ when $T_i = 0$. We focus on the average effects of college education on entrepreneurial choice, that is, averages of $y_i(1) - y_i(0)$, over (sub)populations, rather than on person-level effects (Lee and Lemieux 2010).

A person's treatment status (with college education degree or not) is affected by the HEE policy. As described above, persons born after Sep. 1980 have a greater probability of obtaining college education, therefore, the *running variable* affecting person's treatment status is their birth year and months in this research.[6] Assignment indicator, denoted by D_i, is defined as $D_i = \mathbf{1}(X_i \geq X_0)$, where the cutoff $X_0 =$ Sep. 1980 and $\mathbf{1}(\cdot)$ is the indicator function to specify whether this person is born after Sep.1980. In this paper, we assume that persons do not comply perfectly with their assignments, so that $T_i \neq D_i$. In this environment, under certain weak assumptions (Hahn et al. 2001), the impact of college education on entrepreneurial choice can be expressed as

$$\tau^{FRDD} = \frac{\mathrm{E}[y_i(1) - y_i(0) | X_i = X_0]}{\mathrm{E}[T_i(1) - T_i(0) | X_i = X_0]} = \frac{\lim_{x \downarrow x_0} \mathrm{E}(y_i | X_i = x) - \lim_{x \uparrow x_0} \mathrm{E}(y_i | X_i = x)}{\lim_{x \downarrow x_0} \mathrm{E}(T_i | X_i = x) - \lim_{x \uparrow x_0} \mathrm{E}(T_i | X_i = x)} \quad (1)$$

where τ^{FRDD} captures the local average treatment effect, which is the ratio of two reduced-form, sharp local-linear RD estimators.

Hahn et al. (2001) formalize the estimation and inference in the fuzzy regression discontinuity design. Treatment effects are usually estimated using the local polynomial estimators. The regression functions above and below the cutoff are approximated by means of weighted polynomial regressions, typically of order 1 or 2, with weights computed by applying a kernel function on the distance of each observation's score (running variable) to the cutoff. The recommended choice of kernel function is *triangular* (Skovron and Titiunik 2016). In the literature, the bandwidth choice of the kernel function is somewhat controversial, as bias would occur in "large" bandwidth choice in conventional fuzzy regression discontinuity estimator (obtained by balancing squared-bias and variance of the estimator). To address this issue, Calonico et al. (2014) develop a novel data-driven bias-corrected inference procedure, where the estimator is robust to "large" bandwidth choices.[7] Recently, Calonico et al. (2018) extend Calonico

[6] In literature, the observed variable affecting treatment status is also called as *score variable*.

[7] The program operates as follows: first, recentering the usual t-statistic with an estimate of the leading bias, which can correct the bias of RD estimator to account for the effect of a "large" bandwidth choice; second, rescaling the bias-corrected t-statistic with a novel standard error formula that accounts for the additional variability introduced by the estimated bias.

et al. (2014)'s seminal work to the case where additional covariates are included in the estimation, and they offer more tools for a systematic and objective analysis. In this paper, we mainly follow Calonico et al. (2018)'s new procedures.

In FRDD, the identification typically relies on one critical assumption that observations around the cutoff are similar (Skovron and Titiunik 2016), in other words, there exists no "jump" in these pre-determined variables around the cutoff. Therefore, we construct some pre-determined variables for validity tests. As reviewed by Simoes et al. (2016), determinant factors of the entrepreneurship choice include basic individual characteristics, family background, personality characteristics, education and experience, health condition, nationality and ethnicity and access to financial resources. Unlike other variables, education may affect a number of other determinant factors, such as health condition (Cutler and Lleras-Muney 2006), thus education may affect entrepreneurial choice in both direct and indirect ways. Therefore, we construct the following four pre-determined variables: *Gender*, *Ethnicity*, *One-Child* and *Hukou*.[8] The impacts of gender and ethnicity on entrepreneurial choice are widely documented in the literature (Simoes et al. 2016), and they are determined before the education.

The impacts of *One-Child* and *Hukou* are not straightforward. For "*One-Child*" variable, Chen (2016) and Sun et al. (2016) propose that the probability of one-child being entrepreneurship is significantly lower than those with siblings. Chen (2016) emphasizes the role of personality traits (one-child is more risk averse, less competitive and less trusting), while Sun et al. (2016) stress the lack of "social insurance" (no support from siblings). In both cases, being the *One-Child* affects this person's entrepreneurial choice.

For the variable "*Hukou*", preferred indicators would be family background, such as parents' education and occupation (Djankov et al. 2006; Yueh 2009a, 2009b), however, missing values on family background are massive in our database, we thus adopt *Hukou* as a proxy for family background. In China, *Hukou* is an important social identity: a person with *Rural Hukou*, relative to a person with *Urban Hukou*, is discriminated in both resource allocation and labor market (Afridi et al. 2015; Jiang et al. 2012). Understandably, China's parents have incentives to acquire *Urban Hukou* for their children, and *Rural Hukou* can serve a good proxy for weak family background, at least for those born in 1980s. Qu and Guo (2017) explore the impact of rural Hukou on entrepreneurial choice, finding that rural Hukou decreases the probability of becoming an entrepreneur.

Data

Data Source

Our main data source is the *China Household Income Project* (CHIP) 2013 survey. CHIP2013 is conducted in July and August in 2014. Variables it contains include household income, household expenditure, individual information, work time in 2013, job information, household assets, demolition land information, agriculture business and so forth. The survey is conducted by the Annual Household Survey Office of Integration of Urban and Rural in National Bureau of Statistics (NBS). The CHIP2013

[8] We will define these variables in data section.

is a subsample of NBS's Annual Integration Household Survey (AIHS) in 2013, containing 160 thousands households in 31 provinces. The CHIP sample is selected by systematic sampling method in three layers of East, Middle and West, containing 15 provinces, 126 cities, 234 counties, 18,948 households and 64,777 individuals. Altogether, there are 7175 urban households, 11,013 rural households, and 760 migrant households. Because most people with college education are from urban areas, following L. Li and Wu (2014), Yueh (2009a) and Yueh (2009b), we focus on entrepreneurial choice in urban China, so we use only the urban sample.

Variables

In this paper, a person is defined as an entrepreneur if this person's employment status is either "employer" or "self-employment", which is consistent with Chu and Wen (2017), L. Li and Wu (2014) and Zhang and Li (2016). Following the literature (Chu and Wen 2017; Zhang and Li 2016), we distinguish *Boss-type* and *Self-employed-type* entrepreneurs. Specifically, a person is defined as *Boss-type* entrepreneur if this person's employment status is employer, and as *Self-employed-type* entrepreneur if the employment status is self-employment. This would enable study of possible heterogeneous impacts of these two types.

College Education is defined as the highest level of education completed, either *polytechnic college*, *undergraduate* (Bachelor's degree) or *graduate* (Master's degree or above). Pre-determined variables include *Female*, *Minority*, *One-Child* and *Rural Hukou*. *Female* is a dummy variable set to one if the respondent's gender is female and zero otherwise. *Minority* is set to be zero when the respondent's ethnicity is Han and one otherwise. *One-Child* equals one if the respondent has no siblings, otherwise equals to zero, and *Rural Hukou* equals one when the respondent's residence registration at birth is the rural China and zero otherwise.

We construct the *running variable* as the difference between the birth year and month and the policy cutoff (September 1980). For instance, the value of the running variable of a person born in August 1980 is set to be -1, while the value of the running variable of a person born in October 1980 is *1*. For a typical person, if she obtains college education and then enters into the labor market, she is supposed to be 23-years old. Our database provides employment information for all adults in 2013, which implies these persons we care about should be born before 1990. Therefore, we keep observations if the value of running variable is less than *120*. The logic of RDD approach is to use similar enough units around the cutoff to conduct inferences, we follow Liu et al. (2016) and keep observations who are born 10 years before and after the policy. We finally obtain 5503 observations.

Data Facts

Table 1 reports summary statistics. In our final sample, 3167 observations are born before the policy cutoff, which counts 57.6% of the sample. Figure 2 illustrates the distribution of birth year and month, which shows that there is a downtrend in the number of people born before 1981, while the number of people born after 1981 shows stationarity feature. This asymmetric distribution in the number of sample is reasonable, since China's government initiates the *One Child* policy in late 1979 (see Qin et al.

Table 1 Summary Statistics

	Total	Before Sep. 1980	After Sep. 1980
Entrepreneur	0.114	0.133	0.088
	(0.318)	(0.340)	(0.283)
Boss	0.030	0.036	0.022
	(0.171)	(0.187)	(0.146)
Self-employed	0.084	0.097	0.066
	(0.277)	(0.295)	(0.248)
College Educ.	0.482	0.410	0.580
	(0.500)	(0.492)	(0.494)
Female	0.487	0.487	0.487
	(0.500)	(0.500)	(0.500)
Minority	0.052	0.048	0.058
	(0.223)	(0.214)	(0.233)
One-Child	0.293	0.168	0.463
	(0.455)	(0.374)	(0.499)
Rural Hukou	0.348	0.361	0.330
	(0.476)	(0.480)	(0.470)
Observations	5503	3167	2336

Standard errors in parentheses

(2017) for policy review), which significantly decreases the population scale born before 1981, while stabilizes the population scale born after 1981.

There exist substantial differences in obtaining college education and pre-determined variables between persons born before Set. 1980 (the "Before Group") and after Set. 1980 (the "After Group"). As show in Table 1, in our sample, 41.0% of persons in the Before Group obtain college education, which is 17 percentage point lower than that of persons in the After Group; 4.8% of the Before Group members are not ethic Han, which is 1 percentage point lower than that of the After Group; 16.8% of

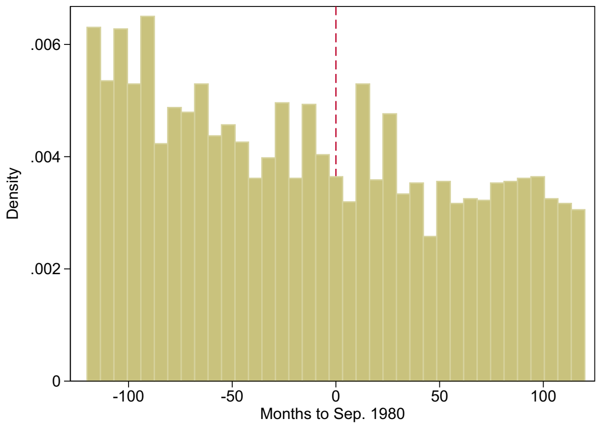


Fig. 2 Distribution of the Running Variable

the Before Group members are one-child, which is 29.5 percentage point lower than that of the After Group; and 36.1% of the Before Group members have Rural Hukou, which is 3.1 percentage point higher than that of the After Group. These differences agree with intuition because of the following factors: the *Higher Education Expansion* policy increases the population of college students, the *One Child* policy is not strictly implemented among minority, and that the *Urbanization* increases the scale of population with Urban Hukou. These differences will not constrain our *Fuzzy Regression Discontinuity Design* analyses, since the necessary identification condition is the smoothness of pre-determined variables around the cutoff, not the similarity of pre-determined variables among the Before group and the After group.

There also exist differences in entrepreneurship between the Before Group and the After group. As shown in Table 1, among our sample, 11.4% of them are entrepreneurs, where 13.3% of the Before Group members are entrepreneurs, which is 4.5 percentage point higher than that of the After Group. This is understandable. On the one hand, college education may or may not promote the probability of a person's becoming an entrepreneur. As discussed in the introduction section, higher levels of education may generate better employment options and thus decrease the likelihood of entrepreneurship as the preferred choice (van der Sluis et al. 2008). This may hedge out the positive impact of college education on entrepreneurship. On the other hand, the probability of becoming an entrepreneur is related to age. It is widely documented that age has an inverse "U" shape impact on entrepreneurial choice (Simoes et al. 2016), and the inflection point of age in China is estimated to be 48 years old (Chu and Wen 2017). The average age of the After Group members is smaller than that of the Before Group, therefore, the share of entrepreneurs in the After Group should be smaller than that in the Before group.

Differences in *Boss-type* entrepreneurs and *Self-employed-type* entrepreneurs among the Before Group and the After Group show the same pattern as the overall entrepreneurs. As shown in Table 1, in our sample, 3.0% of observations are *Boss-type* entrepreneurs, which accounts for about 26.3% of the entrepreneurs. Among our sample, 3.6% of the Before Group members are *Boss-type* entrepreneurs, which is 1.4 percentage point higher than that of the After Group, 9.7% of the Before Group members are *Self-employed-type* entrepreneurs, which is 3.1 percentage point higher than that of the After Group.

Figure 3 provides additional evidence that college education may not promote entrepreneurship. For people born before 1981, about 18% of persons without college education become entrepreneurs, while 5% of persons with college education become entrepreneurs. This fact implies that persons with college education are more likely to engage in formal employment. For persons born after 1981, the proportion of entrepreneurs among these people shows the same pattern as that of people born before 1981, although the share of entrepreneurs among persons with and without college education exhibits a downward trend.

We discussed earlier that impacts of college education on entrepreneurial choice may be heterogeneous. For college graduates, as shown in Fig. 4a, the share of *Boss-type* entrepreneurs is greater than the share of *Self-employed-type* entrepreneurs among those born before 1976 (i.e. older) and smaller among those born after 1976 (younger). This is likely due to that experiences and resources are higher for entrepreneurs with older ages. Moreover, the proportion of these two types of entrepreneurs among college

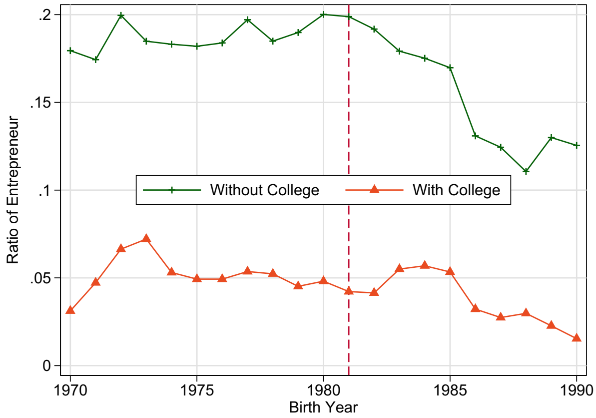


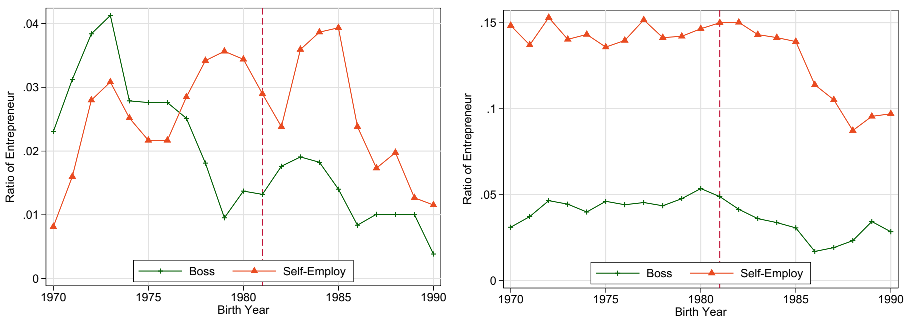
Fig. 3 Share of Entrepreneurs among Persons with or without College Education

graduates are not too much different from each other. For persons without college education, the pattern is obviously different from above. As shown in Fig. 4b, the proportion of *Self-employed-type* entrepreneurs is almost 10 percentage point higher than the proportion of *Boss-type* entrepreneurs in the entire age distribution. This observed heterogeneous relationship between college education and entrepreneurial choice is consistent with Levine and Rubinstein (2017)'s finding that *Boss-type* entrepreneur tends to be more educated.

Empirical Results

Validity Tests

Validity tests are conducted on our fuzzy regression discontinuity design as follows: First, the continuity of running variable is tested. One concern in standard FRDD is the self-selection or non-random sorting of units into control and treatment status (Lee and Lemieux 2010; Skovron and Titiunik 2016), that units in a (quasi-)experiment have



(a) With College Education (b) Without College Education

Fig. 4 Share of Different Types of Entrepreneur among Persons with or without College Education power to manipulate their score of running variable to affect their outcomes in experiment. This "manipulation" generates a discontinuity in the density of units at the known cutoff, and one way to detect the "manipulation" is to test the continuity of the running variable. The existence of "manipulation" is highly unlikely in our analysis. The "unexpectedness" of the higher education expansion policy indicates that people have no priori information to "change" their birth year and month 18 years before the implementation of the policy. As shown in Fig. 2, visual inspection shows that there is no obvious "jump" in the running variable around the cutoff. Nevertheless, Cattaneo et al. (2016)'s approach is implemented to test the continuity of running variable.[9] The resulted p value is 0.1411, indicating that there is no evidence that the density is discontinuous at the cutoff.

Second, the smoothness of all pre-determined covariates is examined around the cutoff. For the identification of FRDD, observations around the cutoff are required to be similar (Lee and Lemieux 2010; Skovron and Titiunik 2016), meaning that there exists no discontinuity in pre-determined covariates around the cutoff. We first present graphical illustrations of all pre-determined variables around the cutoff. We follow Calonico et al. (2015)'s data-driven regression discontinuity plots procedure,[10] where the default options are that the kernel function is *triangular* and bandwidth is selected by mimicking variance evenly-spaced method using *polynomial* estimators.[11] As shown in Fig. 5, there exists no obvious "jump" in these pre-determined variables, with the exception of *Minority*. We then examine treatment effects of the *Higher Education Expansion* policy on these pre-determined variables. We apply Calonico et al. (2018)'s robust bias-corrected confidence intervals and inference procedures, where the default options are that the local polynomial is *order 1*, kernel function is *triangular*, and bandwidth is selected by exploiting the one common *Mean Square Error optimal (MSE-optimal) bandwidth selector*. As shown in Table 2, the policy has no significant impact on these pre-determined variables. Therefore, above evidence indicates that all pre-determined covariates around the cutoff are smooth in our FRDD.

Third, we examine the "jump" in college education around the cutoff. In our FRDD, the *Higher Education Expansion* policy should significantly increase the probability of a person's obtaining college education. Figure 6 shows an obvious discontinuity around the cutoff. As a formal test, we compute treatment effect of the *Higher Education Expansion* policy on obtaining college education. As shown in column (1) of Table 3, this policy significantly increases the probability of a person's obtaining college education by 12.4%, which is consistent with the graphical illustration. We also conduct a placebo test. Assuming that the *Higher*

[9] Compared with the well-known McCrary's manipulation, which requires pre-binning of the data and hence introduces additional tuning parameters, this new test requires choosing only one tuning parameter, avoids pre-binning the data and permits the use of simple well-known weighting schemes, and thus it removes the need of choosing the length and positions of bins, or employing complicated boundary kernels directly. In our test, we use the corresponding Stata package with default options.

[10] In traditional analyses, RD plots are typically presented employing ad hoc choices of tuning parameters, which makes these procedures less automatic and more subjective. This new data-driven plot doesn't require additional choices of tuning parameters, which is objective and automatic.

[11] In the first version, we use the default *spacings* estimator. In fact, *spacings* is more suitable for continuous variable. We appreciate the referee points out our mistake, and we now use *polynomial* estimator.

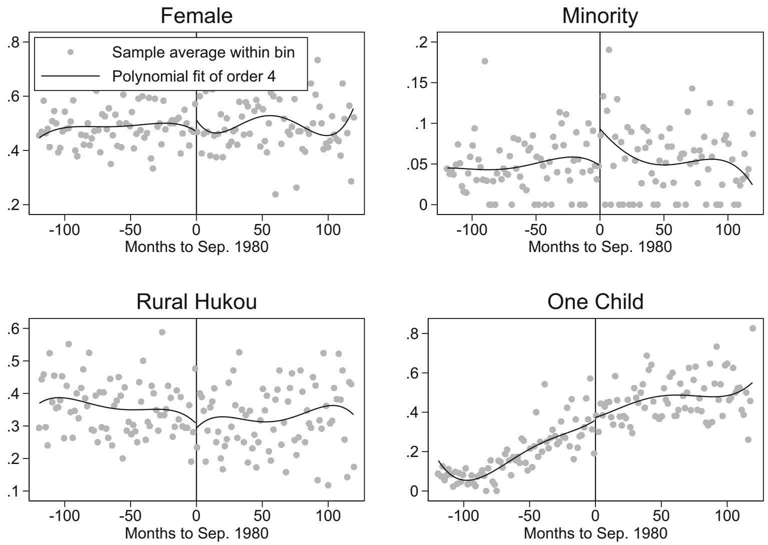


Fig. 5 Data-Driven Regression Discontinuity Plots of Pre-determined Variables

Education Expansion policy was implemented six months or one year before or after September 1998, as shown in column (2)–(5) of Table 3, our results indicate that this artificial policy has no significant impact on the probability of obtaining college education. Therefore, above evidence indicates that college education around the cutoff is discontinuous.

Table 2 The Impact of Higher Education Expansion Policy on Pre-determined Variables (Regression Discontinuity Design)

VARIABLES	(1) Female	(2) Minority	(3) Rural Hukou	(4) One-Child
Policy	0.059	0.043	−0.005	0.022
	(0.072)	(0.031)	(0.064)	(0.065)
Observations	5503	5503	5503	5503
Cutoff	Sep. 1980	Sep. 1980	Sep. 1980	Sep. 1980
Effective Obs. (L)	679	950	821	821
Effective Obs. (R)	678	923	797	797
Order loc. Poly. (p)	1	1	1	1
Order bias (q)	2	2	2	2
Bandwidth loc. Poly. (h)	29.55	42.99	35.66	35.30
Bandwidth bias (b)	52.15	67.38	55.22	53.97
Kernel Type	Triangular	Triangular	Triangular	Triangular
Bandwidth Type	mserd	mserd	mserd	mserd
VCE Type	NN	NN	NN	NN

Robust Bias-Corrected Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. "mserd" of bandwidth type means *one common mean squared error (MSE)-optimal bandwidth selector*. "NN" of VCE means *heteroskedasticity-robust nearest neighbor variance estimator.*

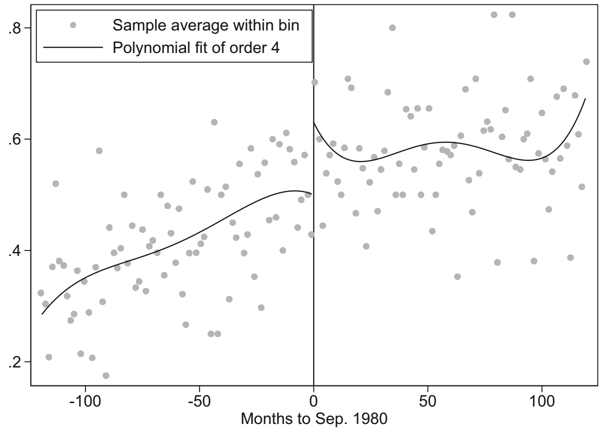


Fig. 6 Data-Driven Regression Discontinuity Plot of College Education

Main Results

The impact of college education on entrepreneurial choice is discussed in this section. Theoretically, FRDD is analogous to two stage least square regression (Lee and Lemieux 2010), where the "first" stage examines the impact of the HEE policy on college education in our analyses, and the "second" stage explores the impact of college education on the

Table 3 The Impact of Higher Education Expansion Policy on College Education (Regression Discontinuity Design)

VARIABLES		Placebo Tests			
	(1) College	(2) College	(3) College	(4) College	(5) College
Policy	0.124*	0.013	0.082	−0.027	−0.021
	(0.068)	(0.056)	(0.061)	(0.064)	(0.061)
Observations	5503	5503	5503	5503	5503
Cutoff	Sep. 1980	March 1980	Sep. 1979	March 1981	Sep. 1981
Effective Obs. (L)	793	1118	893	925	932
Effective Obs. (R)	778	1052	921	899	891
Order loc. Poly. (p)	1	1	1	1	1
Order bias (q)	2	2	2	2	2
Bandwidth loc. Poly. (h)	34.13	48.95	39.50	41.91	42.74
Bandwidth bias (b)	57.34	72.89	61.74	66.44	61.81
Kernel Type	Triangular	Triangular	Triangular	Triangular	Triangular
Bandwidth Type	mserd	mserd	mserd	mserd	mserd
VCE Type	NN	NN	NN	NN	NN

Robust Bias-Corrected Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. "mserd" of bandwidth type means *one common mean squared error (MSE)-optimal bandwidth selector.* "NN" of VCE means *heteroskedasticity-robust nearest neighbor variance estimator*

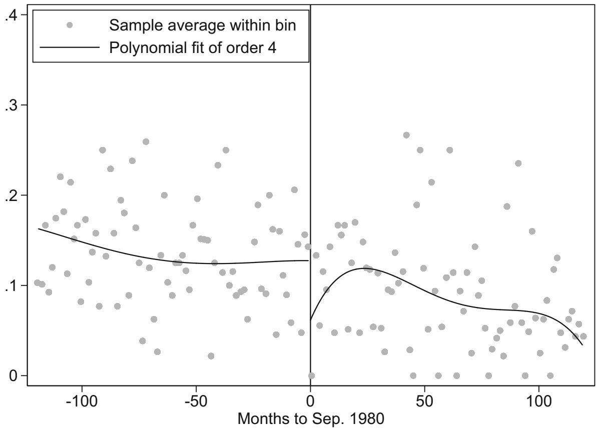


Fig. 7 Data-Driven Regression Discontinuity Plot of Entrepreneurship. **a** Boss-type **b** Self-employed

outcome variable. We first examine the graphical evidence. As shown in Fig. 7, entrepreneurship does not show obvious discontinuity around the cutoff, indicating that the HEE policy does not promote the entrepreneurship. We then conduct FRDD analyses without

Table 4 The Impact of Higher Education Expansion Policy on Entrepreneurship (Fuzzy Regression Discontinuity Design)

VARIABLES	Second Stage		First Stage	
	(1)	(2)	(3)	(4)
	Entrepreneur	Entrepreneur	College	College
College Education	−0.526	−0.554		
	(0.437)	(0.423)		
Policy			0.117*	0.126*
			(0.066)	(0.066)
Observations	5503	5503	5503	5503
Cutoff	Sep. 1980	Sep. 1980	Sep. 1980	Sep. 1980
Control Var.	No	Yes	No	Yes
Effective Obs. (L)	821	740	821	740
Effective Obs. (R)	797	743	797	743
Order loc. Poly. (p)	1	1	1	1
Order bias (q)	2	2	2	2
Bandwidth loc. Poly. (h)	35.54	32.73	35.54	32.73
Bandwidth bias (b)	61.43	57.44	61.43	57.44
Kernel Type	Triangular	Triangular	Triangular	Triangular
Bandwidth Type	mserd	mserd	mserd	mserd
VCE Type	NN	NN	NN	NN

Robust Bias-Corrected Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. “mserd” of bandwidth type means *one common mean squared error (MSE)-optimal bandwidth selector*. “NN” of VCE type means *heteroskedasticity-robust nearest neighbor variance estimator.*

controlling any pre-determined variables. As shown in column (3) of Table 4, this policy significantly increases the probability of obtaining college education by 11.7%. As shown in column (1) of Table 4, college education decreases the probability of becoming an entrepreneur by 52.6%, however, this treatment effect is statistically insignificant. Further FRDD analyses controlling all pre-determined variables show similar results. Column (2) of Table 4 shows that college education still has no significant impact on entrepreneurial choice. As discussed before, education may increase the propensity becoming an entrepreneur though enhancing person's managerial ability as well as the ability to identify entrepreneurial opportunities, while decrease the likelihood of becoming an entrepreneur though generating better outside options, therefore, the above results may be either because college education has no impact on entrepreneurship or the two opposing effects of college education on entrepreneurship offset each other. Regardless the exact mechanisms, above empirical evidence suggests that college education does not permanently promote China's entrepreneurship.

Robustness Checks

We conduct robustness checks and find that our main results are robust. First, we adopt alternative choices of bandwidth. In the standard FRDD, estimation and inference are sensitive to bandwidth choice of kernel function in local polynomial estimation (Lee and Lemieux 2010; Skovron and Titiunik 2016). Calonico et al. (2018) develop five types of mean square error optimal (*MSE-optimal*) bandwidth choices and five types of coverage error rate optimal (*CER-optimal*) bandwidth choices, and we take these 10 alternative choices to conduct our first robustness analyses. As shown in Table 5, the *Higher Education Expansion* policy has a significant positive impact on college education, and college education has no significant impact on entrepreneurial choice, which implies that our main results are robust to bandwidth choices.[12]

Second, we use alternative choice of order of local polynomial. In FRDD, the typical choice of order of local polynomial estimation is 1 or 2 (Skovron and Titiunik 2016). Now we conduct our second robustness analyses by taking order 2 of local polynomial and 10 alternative bandwidth choices. As shown in Table 6, the policy has a significant positive impact on college education, and college education still has no significant impact on entrepreneurial choice, which implies that our main results are robust to order choices.

Third, we use alternative choices of variance. The computation method of variance-covariance matrix will directly affect inferences. Calonico et al. (2018) develop a set of different heteroskedasticity-robust (HC_k class) estimation methods, and we take these alternative choices to conduct our third robustness analyses by using the one common *MSE-optimal* bandwidth selector. When we use order 1 of local polynomial, as show in column (1)–(5) of Table 7, the policy has a significant positive impact on college education, and college education still has no statistically significant impact on entrepreneurial choice. When we use order 2 of local polynomial, as show in column (6)–(10) of Table 7, these findings are robust. Therefore, our main results are robust to variance estimation methods.[13]

[12] In unreported results, we control all pre-determined variables in all robustness checks, and find that all results are still robust.

[13] In our main analyses, we restrict our sample to a 10 years window. Other windows, such as 6 years, 7 years, and so on, have also been used but not reported. Results are still robust.

Table 5 Alternative Bandwidth Choices (Robustness I)

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Second Stage Regression: Dependent Variable is Entrepreneurship										
College Education	−0.526	−0.529	−0.525	−0.526	−0.529	−0.678	−0.638	−0.591	−0.678	−0.638
	(0.437)	(0.471)	(0.499)	(0.437)	(0.472)	(0.454)	(0.447)	(0.425)	(0.454)	(0.448)
First Stage Regression: Dependent Variable is College Education										
Policy	0.117*	0.104*	0.101	0.117*	0.105*	0.137*	0.125*	0.126*	0.137*	0.125*
	(0.066)	(0.062)	(0.062)	(0.066)	(0.063)	(0.075)	(0.071)	(0.069)	(0.075)	(0.071)
Effective Obs. (L)	821	934	934	821	934	539	610	638	539	610
Effective Obs. (R)	797	866	908	797	866	535	602	639	535	602
Order loc. Poly. (p)	1	1	1	1	1	1	1	1	1	1
Order bias (q)	2	2	2	2	2	2	2	2	2	2
BW loc. Poly. (h) (L)	35.54	41.04	41.79	35.54	41.04	23.10	26.68	27.17	23.10	26.68
BW loc. Poly. (h) (R)	35.54	38.49	41.79	35.54	38.49	23.10	25.02	27.17	23.10	25.02
BW bias (b) (L)	61.43	65.77	65.81	61.43	65.77	61.43	65.77	65.81	61.43	65.77
BW bias (b) (R)	61.43	67.24	65.81	61.43	65.81	61.43	67.24	65.81	61.43	65.81
Kernel Type	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular
Bandwidth (BW) Type	mserd	msetwo	msesum	msecomb1	msecomb2	cerrd	certwo	cersum	cercomb1	cercomb2
VCE Type	NN	NN	NN	NN	NN	NN	NN	NN	NN	NN

Note: Robust Bias-Corrected Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. In all analyses, the total observation is 5503, the cutoff of running variable is *Sep. 1980*, and we don't control pre-determinants. "mserd" of bandwidth (BW) Type means *one common mean squared error (MSE)-optimal bandwidth selector*; "msetwo" means *two different MSE-optimal bandwidth selectors*; "msesum" means *one common MSE-optimal bandwidth selector*; "msecomb1" specifies min(mserd, msesum); "msecomb2" specifies median(msetwo, mserd, msesum); "cerrd" means *one common coverage error-rate (CER)-optimal bandwidth selector*; "certwo" means *two different CER-optimal bandwidth selectors*; "cersum" means *one common CER-optimal bandwidth selector*; "cercomb1" specifies min(cerrd, cersum); "ercomb2" specifies median(certwo, cerrd, cersum). VCE type means *heteroskedasticity-robust nearest neighbor variance estimator.*

Table 6 Alternative Order of polynomial (Robustness II)

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Second Stage Regression: Dependent Variable is Entrepreneurship										
College Education	−0.602	−0.786	−0.609	−0.607	−0.613	−0.732	−0.608*	−0.741	−0.732	−0.738
	(0.426)	(0.514)	(0.434)	(0.429)	(0.432)	(0.449)	(0.318)	(0.461)	(0.450)	(0.457)
First Stage Regression: Dependent Variable is College Education										
Policy	0.159*	0.147*	0.158*	0.159*	0.158*	0.185*	0.233**	0.180*	0.185*	0.181*
	(0.083)	(0.086)	(0.083)	(0.084)	(0.083)	(0.098)	(0.101)	(0.097)	(0.098)	(0.097)
Effective Obs. (L)	1016	1069	1048	1016	1048	638	658	658	638	658
Effective Obs. (R)	966	719	987	966	966	639	442	656	639	639
Order loc. Poly. (p)	2	2	2	2	2	2	2	2	2	2
Order bias (q)	3	3	3	3	3	3	3	3	3	3
BW loc. Poly. (h) (L)	45.57	47.42	46.76	45.57	46.76	27.86	28.99	28.58	27.86	28.58
BW loc. Poly. (h) (R)	45.57	31.22	46.76	45.57	45.57	27.86	19.08	28.58	27.86	27.86
BW bias (b) (L)	64.84	64.55	63.65	63.65	64.55	64.84	64.55	63.65	63.65	64.55
BW bias (b) (R)	64.84	51.05	63.65	63.65	63.65	64.84	51.05	63.65	63.65	63.65
Kernel Type	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular
Bandwidth (BW) Type	mserd	msetwo	msesum	msecomb1	msecomb2	cerrd	certwo	cersum	cercomb1	cercomb2
VCE Type	NN	NN	NN	NN	NN	NN	NN	NN	NN	NN

Robust Bias-Corrected Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. In all analyses, the total observation is 5503, the cutoff of running variable is *Sep. 1980*, and we don't control pre-determinants. "mserd" of bandwidth (BW) Type means *one common mean squared error (MSE)-optimal bandwidth selector*; "msetwo" means *two different MSE-optimal bandwidth selectors*; "msesum" means *one common MSE-optimal bandwidth selector*; "msecomb1" specifies min(mserd, msesum); "msecomb2" specifies median(msetwo, mserd, msesum); "cerrd" means *one common coverage error-rate (CER)-optimal bandwidth selector*; "certwo" means *two different CER-optimal bandwidth selectors*; "cersum" means *one common CER-optimal bandwidth selector*; "cercomb1" specifies min(cerrd, cersum); "ercomb2" specifies median(certwo, cerrd, cersum). VCE type means *heteroskedasticity-robust nearest neighbor variance estimator.*

Table 7 Alternative Variance (Robustness III)

Variables	Order of polynomial (1)					Order of polynomial (2)				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Second Stage Regression: Dependent Variable is Entrepreneurship										
College Education	−0.526	−0.526	−0.526	−0.526	−0.526	−0.602	−0.604	−0.603	−0.603	−0.602
	(0.437)	(0.432)	(0.434)	(0.435)	(0.438)	(0.426)	(0.424)	(0.425)	(0.426)	(0.427)
First Stage Regression: Dependent Variable is College Education										
Policy	0.117*	0.118*	0.118*	0.117*	0.117*	0.159*	0.159*	0.159*	0.159*	0.159*
	(0.066)	(0.065)	(0.065)	(0.065)	(0.065)	(0.083)	(0.082)	(0.083)	(0.083)	(0.083)
Effective Obs. (L)	821	821	821	821	821	1016	1016	1016	1016	1016
Effective Obs. (R)	797	797	797	797	797	966	966	966	966	966
Order loc. Poly. (p)	1	1	1	1	1	2	2	2	2	2
Order bias (q)	2	2	2	2	2	3	3	3	3	3
BW loc. Poly. (h) (L)	35.54	35.36	35.41	35.47	35.58	45.57	45.32	45.38	45.45	45.58
BW loc. Poly. (h) (R)	35.54	35.36	35.41	35.47	35.58	45.57	45.32	45.38	45.45	45.58
BW bias (b) (L)	61.43	61.24	61.30	61.36	61.48	64.84	64.64	64.71	64.78	64.92
BW bias (b) (R)	61.43	61.24	61.30	61.36	61.48	64.84	64.64	64.71	64.78	64.92
Kernel Type	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular
Bandwidth (BW) Type	mserd	mserd	mserd	mserd	mserd	mserd	mserd	mserd	mserd	mserd
VCE Type	NN	HC0	HC1	HC2	HC3	NN	HC0	HC1	HC2	HC3

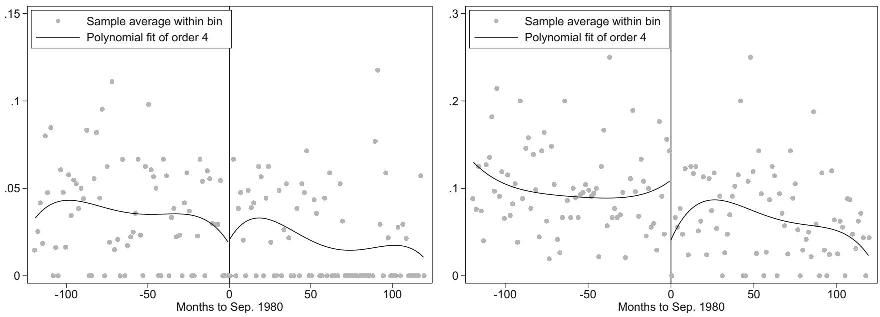
Robust Bias-Corrected Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. In all analyses, the total observation is 5503, the cutoff of running variable is *Sep. 1980*, and we don't control pre-determinants. "mserd" of bandwidth (BW) Type means *one common mean squared error (MSE)-optimal bandwidth selector*. "NN" of VCE type means *heteroskedasticity-robust nearest neighbor variance estimator*; "HC0" means *a heteroskedasticity-robust plug-in residuals variance estimator without weights*; "HC1" means *a heteroskedasticity-robust plug-in residuals variance estimator with hc1 weights*; "HC2" means *a heteroskedasticity-robust plug-in residuals variance estimator with hc2 weights*; "HC3" means *a heteroskedasticity-robust plug-in residuals variance estimator with hc3 weights.*

Heterogeneity Effects

The contribution of private enterprise (managed by *Boss-type* entrepreneur) to the whole economy is much more profound than that of individual business (operated by *Self-employed-type* entrepreneur) (Zhang and Li 2016), suggesting the importance of examining the potential heterogeneous impact of college education. Moreover, the education level of *Boss-type* entrepreneur is significantly higher than that of *Self-employed-type* entrepreneur (Levine and Rubinstein 2017), indicating that the relationship between college education and these two entrepreneurial choices may be heterogeneous. Therefore, we explore the heterogeneous impacts of college education on *Boss-type* and *Self-employed-type* entrepreneurial choice. Besides reasons stated above, this exploration can also check the robustness due to that the entrepreneurial choice is measured differently. The result is shown in Fig. 8. There exist almost no sharp "jumps" in *Boss-type* and *Self-employed-type* entrepreneurial choice around the cutoff. Column (2) and column (4) of Table 8 show that, conditional on all pre-determined variables, the *Higher Education Expansion* policy has a significant positive impact on college education. However, the impacts of college education on *Boss-type* and *Self-employed-type* entrepreneurial choice are both statistically insignificant. Even though effects are insignificant, they are obviously heterogeneous, the college education increase the probability of being *Boss-type* entrepreneur by 0.2%, and decreases the possibility of being *Self-employed-type* entrepreneur by 53.5%. This heterogeneity suggests a possibility of temporary dominance of substitution effects of education on *Self-employed-type* entrepreneurial choice, which agree with the intuition. Therefore, our main results are robust regardless the possible likelihood of heterogeneity.

Conclusion

It is universally agreed that entrepreneurship is critical for economic development and so is education. As governments tend to promote both, it is important to understand interactions of the two, in particular, identification of causal relationship of education on entrepreneurship can be useful in theory and practice. This paper estimates the causal impact of college education on entrepreneurial choice in urban China applying the *Fuzzy Regression Discontinuity Design* approach. We find that China's *Higher*



(a) Boss-type (b) Self-employed

Fig. 8 Data-Driven Regression Discontinuity Plot of Boss-type as well as Self-employed Entrepreneurship

Table 8 The Heterogeneous Impact of Higher Education Expansion Policy on Entrepreneurship (Fuzzy Regression Discontinuity Design)

Variables	Boss		Self-employed	
	(1) Entrepreneur	(2) Entrepreneur	(3) Entrepreneur	(4) Entrepreneur
Second Stage Regression: Dependent Variable is Boss-type/Self-employed-type Entrepreneur				
College Education	0.041	0.002	−0.523	−0.535
	(0.245)	(0.222)	(0.466)	(0.424)
First Stage Regression: Dependent Variable is College Education				
Policy	0.083	0.099*	0.103*	0.121*
	(0.056)	(0.060)	(0.062)	(0.064)
Observations	5503	5503	5503	5503
Cutoff	Sep. 1980	Sep. 1980	Sep. 1980	Sep. 1980
Control Var.	No	Yes	No	Yes
Effective Obs. (L)	1153	908	904	793
Effective Obs. (R)	1057	934	882	778
Order loc. Poly. (p)	1	1	1	1
Order bias (q)	2	2	2	2
Bandwidth loc. Poly. (h)	50.26	41.26	39.66	34.69
Bandwidth bias (b)	79.65	65.76	66.89	59.93
Kernel Type	Triangular	Triangular	Triangular	Triangular
Bandwidth Type	mserd	mserd	mserd	mserd
VCE Type	NN	NN	NN	NN

Robust Bias-Corrected Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. "mserd" of bandwidth type means *one common mean squared error (MSE)-optimal bandwidth selector*. "NN" of VCE type means *heteroskedasticity-robust nearest neighbor variance estimator.*

Education Expansion policy initiated in 1999 significantly increases the probability of a person born after September 1980 obtaining college education by more than 12%; however, college education has no significant impact on overall entrepreneurial choice.

Heterogeneous impacts are also examined due to varying importance of two different types of entrepreneurship. There is suggestive evidence of heterogeneity in the impacts of college education on entrepreneurship, but the coefficients are not precisely estimated. No definitive conclusion can be drawn to support the claim that an increase in the level of education may necessarily led to more entrepreneurial activities.

In May 13, 2015, China's State Council issued the *Opinion to Promote Education in Innovation and Entrepreneurship at Universities and Colleges*, aiming at establishing a new innovation-and-entrepreneurship-oriented education system in 2020. In accordance with this policy, universities and colleges were to establish curriculums promoting innovation and entrepreneurship, and local government and universities were to provide students with an environment that fosters creativity and create multi-channels to support innovation and startup programs. According to the State Council, this reform started in 2015 would witness major progress by 2017 after the concept has spread widely and leading to bourgeoning of startups by college students. Our results indicate that the old-fashioned college education has no significant impact on entrepreneurial choice in the past, suggesting that entrepreneurship may not necessarily be a preferred choice for people with college education degree in the absence targeted policy intervention. Therefore, by implementing reform on curriculum to enhance learning on entrepreneurship, it may alter the impact of education on entrepreneurship choice, thus the new policy of 2015 has a potential to be effective in achieving intended objectives. When new data become available, more research is warranted to examine effectiveness of the new policy enhancing college learning of entrepreneurship.

Acknowledgements We appreciate support from the Income Distribution and Development Policy Research Project at the Research Institute of Economics and Management of SWUFE. We are very grateful for the editor and reviewer's insightful feedback. All errors are mine.

Compliance with Ethical Standards

Conflict of Interest We have no conflict of interest.

References

Afridi F, Li SX, Ren Y (2015) Social identity and inequality: the impact of China's hukou system. J Public Econ 123:17–29. https://doi.org/10.1016/j.jpubeco.2014.12.011

Beck T, Lu L, Yang R (2015) Finance and growth for microenterprises: evidence from rural China. World Dev 67:38–56. https://doi.org/10.1016/j.worlddev.2014.10.008

Blanchflower DG, Oswald AJ (1998) What makes an entrepreneur? J Labor Econ 16(1):26–60. https://doi.org/10.1086/209881

Block JH, Hoogerheide L, Thurik R (2013) Education and entrepreneurial choice: an instrumental variables analysis. Int Small Bus J 31(1):23–33. https://doi.org/10.1177/0266242611400470

Calonico S, Cattaneo MD, Farrell MH, Titiunik R (2018) Regression discontinuity designs using covariates. Rev Econ Stat *forthcoming*

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

Calonico S, Cattaneo MD, Titiunik R (2015) Optimal data-driven regression discontinuity plots. J Am Stat Assoc 110(512):1753–1769. https://doi.org/10.1080/01621459.2015.1017578

Card, D. (1999). Chapter 30 - The Causal Effect of Education on Earnings. In C. A. Orley, & C. David (Eds.), *Handbook of Labor Economics* (Vol. Volume 3, Part A, pp. 1801–1863): Elsevier

Cattaneo MD, Jansson M, Ma X (2016) Simple local regression distribution estimators with an application to manipulation testing. Working Paper

Chen G (2016) China's one-child policy and the missing of entrepreneurship. Economic Research Journal Working Paper

Chu T, Wen Q (2017) State-owned economy and entrepreneurship in China. Working Paper

Cutler DM, Lleras-Muney A (2006) Education and health: evaluating theories and evidence. National Bureau of Economic Research Working Paper Series, No. 12352, doi: 10.3386/w12352

de Wit G (1993) Models of self-employment in a competitive market. J Econ Surv 7(4):367–397. https://doi.org/10.1111/j.1467-6419.1993.tb00171.x

Démurger S, Xu H (2011) Return migrants: the rise of new entrepreneurs in rural China. World Dev 39(10): 1847–1861. https://doi.org/10.1016/j.worlddev.2011.04.027

Djankov S, Qian Y, Roland G, Zhuravskaya E (2006) Who are China's entrepreneurs? Am Econ Rev 96(2): 348–352

Evans DS, Jovanovic B (1989) An estimated model of entrepreneurial choice under liquidity constraints. J Polit Econ 97(4):808–827

Evans DS, Leighton LS (1989) Some empirical aspects of entrepreneurship. Am Econ Rev 79(3):519–535

Hahn J, Todd PE, van der Klaauw W (2001) Identification and estimation of treatment effects with a regression-discontinuity design. Econometrica 69(1):201–209

Jiang S, Lu M, Sato H (2012) Identity, inequality, and happiness: evidence from urban China. World Dev 40(6):1190–1200. https://doi.org/10.1016/j.worlddev.2011.11.002

Knight J, Deng Q, Li S (2017) China’s expansion of higher education: the labour market consequences of a supply shock. China Econ Rev 43:127–141. https://doi.org/10.1016/j.chieco.2017.01.008

Lazear EP (2005) Entrepreneurship. J Labor Econ 23(4):649–680. https://doi.org/10.1086/491605

Le AT (1999) Empirical studies of self-employment. J Econ Surv 13(4):381–416. https://doi.org/10.1111/1467-6419.00088

Lee DS, Lemieux T (2010) Regression discontinuity designs in economics. J Econ Lit 48(2):281–355. https://doi.org/10.1257/jel.48.2.281

Levine R, Rubinstein Y (2017) Smart and illicit: who becomes an entrepreneur and do they earn more? Q J Econ 132(2):963–1018

Li L, Wu X (2014) Housing price and entrepreneurship in China. J Comp Econ 42(2):436–449. https://doi.org/10.1016/j.jce.2013.09.001

Li S, Whalley J, Xing C (2014) China's higher education expansion and unemployment of college graduates. China Econ Rev 30:567–582. https://doi.org/10.1016/j.chieco.2013.08.002

Liu S, Zhou S, Hu A (2016) Compulsory education law and economic return to education in urban China: based on regression discontinuity design. Econ Res J 51(2):154–167

Lu J, Tao Z (2010) Determinants of entrepreneurial activities in China. J Bus Ventur 25(3):261–273. https://doi.org/10.1016/j.jbusvent.2008.10.005

Lucas RE (1978) On the size distribution of business firms. Bell J Econ 9(2):508–523. https://doi.org/10.2307/3003596

Mohapatra S, Rozelle S, Goodhue R (2007) The rise of self-employment in rural China: development or distress? World Dev 35(1):163–181. https://doi.org/10.1016/j.worlddev.2006.09.007

Qin X, Zhuang CC, Yang R (2017) Does the one-child policy improve children's human capital in urban China? A regression discontinuity design. J Comp Econ 45(2):287–303. https://doi.org/10.1016/j.jce.2016.09.001

Qu Z, Guo S (2017) Employed v.s self-employed, the role of Hukou: evidence from urban China. *China Economic*. Studies 305(6):72–86

Simoes N, Crespo N, Moreira SB (2016) Individual determinants of self-employment entry: what do we really know? J Econ Surv 30(4):783–806. https://doi.org/10.1111/joes.12111

Skovron C, Titiunik R (2016) A practical guide to regression discontinuity designs in political science. Working Paper

Sun W, Sun A, Wang Y (2016) Does one-child policy corrode Chinese entrepreneurship? China Journal of Economics 3(2):138–160

van der Sluis J, van Praag M, Vijverberg W (2005) Entrepreneurship selection and performance: a meta-analysis of the impact of education in developing economies. World Bank Econ Rev 19(2):225–261

van der Sluis J, van Praag M, Vijverberg W (2008) Education and entrepreneurship selection and performance: a review of the empirical literature. J Econ Surv 22(5):795–841. https://doi.org/10.1111/j.1467-6419.2008.00550.x

van Praag CM, Versloot PH (2007) What is the value of entrepreneurship? A review of recent research. Small Bus Econ 29(4):351–382. https://doi.org/10.1007/s11187-007-9074-x

Wu X, Zhang Z (2010) Changes in educational inequality in China, 1990–2005: evidence from the population census data. Hannum E, Park H, Goto Butler Y (ed) Globalization, Changing Demographics, and Educational Challenges in East Asia (Research in the Sociology of Education, Vol. 17), Emerald Group Publishing Limited, Bingley, pp, 123–152. https://doi.org/10.1108/S1479-3539(2010)0000017007

Xing C, Yang P, Li Z (2018) The medium-run effect of China's higher education expansion on the unemployment of college graduates. China Econ Rev 51:181–193. https://doi.org/10.1016/j.chieco.2017.05.005

Yeung W-JJ (2013) Higher education expansion and social stratification in China. Chinese Sociological Review 45(4):54–80. https://doi.org/10.2753/csa2162-0555450403

Yueh L (2009a) China’s Entrepreneurs. World Dev 37(4):778–786. https://doi.org/10.1016/j.worlddev.2008.07.010

Yueh L (2009b) Self-employment in urban China: networking in a transition economy. China Econ Rev 20(3):471–484. https://doi.org/10.1016/j.chieco.2009.01.002

Zhang C, Li T (2016) Clan and entrepreneurship from the perspectives of both network and cultural norms: effects and mechanism. *Economic Research Journal Working Paper*