The Power of Demand:

A Natural Experiment in Higher Education

Mahdi Majbouri[1]

Babson College and IZA

Abstract:

A primary issue in education policy is understanding the degree to which government interventions on the supply side of the market affect educational outcomes. This study uses a natural experiment setting in one of the most under-studied countries, Iran, to contribute to this policy issue. Using a regression discontinuity design, this study measures the effect of the Iranian Cultural Revolution, during which all institutions of tertiary education were closed for 30 months, on college attainment rates of affected cohorts. The results show that this elimination of higher education supply reduced men's college attainment rate by only 2 percentage points (about 14%) while it had no impact on women's. These results provide an example for the argument that government interventions have little impact on educational outcomes when demand for education is high. It has important implications for higher education policy in both developing and developed countries. The Cultural Revolution is then used as an instrument to estimate returns to college education for men.

JEL Classification Code: I23, I28, C99

Key Words: Higher education, Natural experiment, Government Policy, Institutions

---

[1] Associate Professor of Economics at Babson College and Research Fellow at IZA. 231 Forest St., Babson Park, MA 02472. Phone: 781-239-5549. Fax: 781-239-5239. mmajbouri@babson.edu

1. Introduction

A principal problem in education policy is determining the degree to which government intervention on the supply side of the education market affects educational outcomes. There is ongoing debate on whether expansion of supply increases educational attainment, particularly learning. One viewpoint is that, because low supply means less access and higher costs, supply side interventions are necessary to expand access and increase educational attainments. There is also empirical evidence that shows access matters, including the seminal work by Duflo (2001 and 2004) demonstrating that the substantial expansion of primary schools in Indonesia in the 1970s led to cost-effective increases in educational attainment with long-term positive consequences.

The opposing viewpoint is that if the return to education is large enough, this will drive sufficient demand for education that makes markets respond by increasing the supply of education, without the need for government intervention. For example, following the invention of the printing press and the advent of the Lutheran faith, which required the faithful to read religious texts, demand for education in Northern Europe rose, which led to an endogenous increase in the supply of schools and education. Literacy rates in England, for example, rose from 5% in 1500 to about 50% in 1750 (Mitch 2004), despite the absence of free or compulsory primary education. Academics and policy makers who are pessimistic about government interventions tend to be major advocates of this view, and argue that, although interventions in the market to increase access (supply) may raise educational attainment, these interventions may not lead to actual learning, and are not necessarily cost-effective (as returns to education and, thus, demand for education, are low to begin with). There is empirical evidence demonstrating that the rise in supply of schooling in developing countries, particularly in the first decade of the 21st century, has not led to actual learning. For example, although primary education is close to universal in India, the Annual State of Education Report (ASER) has shown that 60% of second graders could not read a simple story, and 35% could not do a simple division problem (Pratham 2005).[1] Proponents of the first viewpoint, however, argue that parents may under-invest in children's education due to a variety of imperfections in the market, including systematic under-estimation of returns to education (see Nguyen (2008) for Madagascar, and Lergetporer et al. (2018) for Germany), or credit market failures. Therefore, they concluded that interventions on the supply side are critical to increase educational attainments. The ongoing debate between these two opposing views, which are significantly more complex and nuanced than what is presented here, has influenced educational policy in the last seven decades.[2] It is, however, under-researched in the context of higher education.

This paper contributes to this policy debate through a natural experiment in one of the most under-studied countries, Iran. It measures the effect of the Iranian Cultural Revolution, during which all institutions of tertiary education were closed for 30 months between 1980 and '82, on college attainment of six birth-year cohorts who were in college or were supposed to enter college during this period. As the Iran-Iraq war was escalating during this same period, the affected cohorts were also drafted into war rather than entering university. Using a Regression Discontinuity Design (RDD) to compare the last cohort for whom the doors of the universities were closed with the next cohort, who were able to enroll in university, this paper documents the causal effect of this temporary elimination of supply of higher education on the college attainment rates of these two cohorts. The results demonstrate that the Cultural Revolution and the draft that followed it reduced men's college attainment rate by only 7 to 14 percent (1 to 2 percentage points), and had no effect on women's college attainment. In other words, this elimination of supply coupled with the draft

---

[1] This is, unfortunately, not unique to India and has similarly been documented in other developing countries. In addition, randomized controlled trials have shown that increasing inputs to schools such as providing blackboards, flowcharts, textbooks, or even more teachers, had little impact on test scores.
[2] Banerjee and Duflo (2012) offer a nice summary of the debate that permeates not just education but health policy and fertility control as well (see Chapter 4).

into war had little impact on college attainment of affected cohorts, especially that they could not enter college for a few years after universities opened, because of age restrictions. This surprising result shows that demand for education was high enough that the cohorts affected by the Cultural Revolution and the draft into war were still willing to enter college years later. It provides an example in support of the second argument discussed above, particularly for higher education which is under-researched, as it shows that demand for education is crucial in educational attainment. This conclusion has important implications for higher education policy in both developing and developed countries. In particular, it suggests that increasing demand for education should be a critical part of education policy.

This study also uses the exogenous decline in college attainment rate caused by the Cultural Revolution to estimate the return to college education for men. Although OLS estimates show that individuals with college and above education have 76% higher hourly wages than those with less than a college education, the IV estimates show a 100% return on hourly wages as a result of earning college and above education. Souri (2004) has used deviations from the trend in college attainment rate for cohorts affected by the Cultural Revolution as an instrument for college attainment to estimate returns to education. This study, however, uses an RDD in the first stage of the IV estimates. It argues that the first stage satisfies the exclusion restriction better than a deviation from the trend instrument.

The rest of this paper is organized as follows: Section 2 explains the Cultural Revolution in Iran. Section 3 is about the data which is publicly available at IPUMS International. Sections 4.1, 4.2, and 4.3 present the evidence. Section 4.4 discusses the results and Section 4.5 reports the estimates on returns to education. The paper concludes in Section 5.

2. Cultural Revolution in Iran

The Islamic Revolution of 1979 in Iran has been considered a turning point in the recent history of the country. From the onset, its leaders viewed this political revolution as a cultural and ideological revolution that aimed to establish a new paradigm in Iran to replace the notions of the Western culture (Mehran 1997). Reforming cultural institutions and, particularly, the educational system was identified as a high priority early on. Universities were of particular importance, as they were one of the main centers of political activity during and after the revolution.

In the initial months after the revolution, the transitional cabinet of Prime Minister Bazargan, comprised of moderate religious nationalists, protected most political freedoms. But within days of the occupation of the US embassy in Tehran on November 4, 1979 by Islamist and leftist groups, the transitional cabinet resigned en mass. Leftist groups, whose stronghold was the universities, vied with Islamist groups to fill the resulting power vacuum. The political elite, however, viewed the political activities on university campuses as a major source of instability. In an April 18, 1980, speech, Ayatollah Khomeini voiced concerns about the universities, stating "we are not afraid of economic sanctions or military intervention. What we are afraid of is Western universities and the training of our youth in the interests of West or East." Soon after, violent fights broke out between the Islamists and leftist groups in the universities, which provided the government with the excuse to order the shutdown of all institutions of higher education on June 12, 1980, until further notice.[3] Thus began one of the most important educational projects in recent history of Iran: the overhaul of the higher educational system, a.k.a. the Iranian *Cultural Revolution*.

The Cultural Revolution Committee was established on June 13th by the government to manage the process, with the goals of redesigning the programs and curricula in universities in line with the

[3] For a detailed account of events, see Bakhash (1984) and Hiro (1985).

ideologies of the revolution, and to select and train the "right" faculty to teach the new curricula. Particular attention was paid to programs in humanities and social sciences, as curricula in these fields were considered "un-Islamic."[4] Closure of universities provided the committee with the opportunity to implement their goal. As part of this process, roughly 8000 faculty members with leftist dispositions, i.e. approximately half of the total number of professors in the country, were purged from the universities (Ashraf 1997). Most of them either left the country, or transitioned to private businesses. Leftist students were also expelled and barred from re-entering college until almost a decade later.

On September 22, 1980, less than four months after the closure of the universities, Iraq attacked Iran and sparked the longest war of the 20th century.[5] All high school students who reached age 18 in 1980 and the subsequent two years were forcibly conscripted to the military, and were thus drafted into war. During 1982, the universities gradually re-opened, first by allowing students needing 25 or fewer credits to graduate to enroll in the spring of 1982. On December 18, 1982, after thirty months of closure, the universities were re-opened to prior students requiring more than 25 credits to graduate, and the first new cohort enrolled in the fall of 1983.

The Cultural Revolution, however, did not end with the closure of universities. The Cultural Revolution Committee, whose name was changed to the Supreme Council of Cultural Revolution, still exists and, among other tasks, continuously updates textbooks and curricula. For the purposes of this paper, the term Cultural Revolution will be used to refer to the period between 1980 and 1982 during which all institutions of higher education were closed.

---

[4] This is still an important discourse in the political and social construct of the Islamic Republic.

[5] Abrahamanian (2008) compares this war, which was the longest war of the twentieth century, to World War I, in terms of military tactics, such as trench warfare, human wave attacks, and the use of chemical weapons such as mustard gas by the Iraqi government. The human wave attacks, in particular, required large body of soldiers and maximized demand for recruits.

The group of high school students fully affected by the closure of universities is comprised of three cohorts, namely those who graduated high school and could have enrolled in university in 1980, 1981, or 1982. No question in the data specifically asks whether an individual belongs to this group or not. In other words, no question asks whether the individual could potentially enter college in 1980, '81, or '82. But, with some approximation, one can argue that at least for men, this group consisted of individuals who were 18 in 1980, '81, or '82. This is because male high school graduates could try to enter college before age 18. After reaching this age, they had to go to military service. Mandatory military service was particularly well-enforced at the time,[6] as the country was at war with Iraq and soldiers were in high demand. Even after completing their two-year mandatory service, they could not immediately try to enter college due to new age limits on college entrants imposed by the government in the initial years following the Cultural Revolution.

Women were not subject to mandatory military service, and could have attempted to go to college even after they reached 18. Women who turned 18 in 1980, '81, or '82, and wanted to go to college but were unable due to the university closures had two remaining options: wait to apply to college until an unknown time in the future when universities would presumably reopen, or get married. These two options are largely mutually exclusive in Iran, as the chance to pursue higher education was negligible after marriage and starting a family. Given that the median age at first marriage for women was approximately 17 years of age at the time[7], women affected by the Cultural Revolution experienced significant pressure to marry, particularly in light of the uncertainty surrounding if and when universities would re-open. In this paper, we estimate the effect of university closure on educational attainment of both male and female cohorts who turned 18 in 1980-82.

[6] It is still very well-enforced.

[7] Median age at first marriage for women who were 10 to 29 years old in 1980 was 17, according to the Demographic and Health Survey of 2000.

The cohorts who turned 18 in 1977, ‘78, and ‘79 and were already enrolled in college when the Cultural Revolution happened experienced interruption of their education until the universities reopened. These cohorts were potentially less likely to finish their education or attain higher levels of education (e.g. graduate school). As partial and full completion are coded similarly, we cannot identify partial vs. full completion of college in this data; it is not possible to estimate the effect of the Cultural Revolution on college completion of these cohorts.

3. Data

Two sets of data are used in this study. The first set contains the combined 2% censuses of 2006 and 2011 in order to estimate the impact of Cultural Revolution on educational attainment. As the censuses do not contain hourly wages, a second set of data, the Annual Household Income and Expenditure Surveys (HIES) from 2006 through 2015, is used to estimate the return to education (using the Cultural Revolution as an instrumental variable for education.) The Statistical Center of Iran (SCI) collects these datasets, and both censuses are available on the IPUMS International website, and the HIES datasets are available on SCI website.[8]

The 2% nationally representative samples of censuses have over 2.78 million individuals and over 750,000 households combined. At the individual level, the data contains demographic information, such as gender, age, relation to the head of the household, birth year and month, education, employment and job characteristics, marital status, number of children ever born, migratory status, reasons for migration, and more.

The largest sample used in this study consists of individuals born between 1931 and 1984. Summary statistics of all variables for this sample are reported in Table 1, Panel A. As shown in Panel A, 77%

[8] IPUMS website address is https://www.ipums.org/ and the SCI website address is http://amar.org.ir/english (accessed on Jan. 5, 2018). Note that these URLs could change.

of this sample is literate and about 13% have attended college. The number of observations with recorded birth month is about 3% smaller than the sample with recorded birth year (1,372,865 vs. 1,419,591): the results are separately analyzed based on birth month and birth year.

HIES has been collected annually since 1984, and contains disaggregated expenditure and income as well as some household demographic data. Since 2006, hours worked have been collected as part of this survey (HIES is one of the few datasets in Iran that collects hours worked). This data allows us to calculate wages and estimate the return to education. Each survey since 2006 has about 31,000 to 39,000 households. Rural areas are over sampled. The surveys are generally used to estimate macro-level variables, such as inflation, and thus, have detailed disaggregated expenditure data. After Censuses, HIES is the most important and regular project of the Statistical Center of Iran, but contains limited demographic data. For example, although the Censuses have year and month of birth, HIES data includes only age, allowing only birth year (not birth month) to be constructed. Summary statistics of the variables used from the HIES survey are reported in Table 1, Panel B.

An important issue with all Iranian surveys (censuses and HIES) is that education is only identifiable by level: primary, middle school, high school, college, and graduate school. Years of education, grades, and whether the educational level was partially or fully completed are not included. Therefore, our outcome variable can best be defined as a dummy variable that shows whether the individual has partially or fully completed college or graduate school, or not. This dummy variable is termed "College and above" throughout the rest of this paper.

One technical point is that birth years and months in the non-harmonized IPUMS data are recorded according to the Persian calendar[9]: although the Persian calendar is a solar calendar, it does not match the Gregorian calendar. The Persian calendar, which has been in use in Iran, Afghanistan, and

[9] These variables are `ir2006a_birthyr`, `ir2006_birthmo`, `ir2011_birthyr`, and `ir2011_birthmo`.

Tajikistan for centuries, runs from spring equinox (the first day of spring, typically March 20th) to the day prior to the next spring equinox. Thus, each year of the Persian calendar overlaps with the last nine months of each year of the Gregorian calendar and the first three months of the following year of the Gregorian calendar.[10] For simplicity, in this paper, each year on the Persian calendar is approximated as the Gregorian year with which it has the longest overlap (i.e. nine months). For instance, the Persian year 1390 (which ran from March 20th, 2011 to March 19th 2012) is approximated as the year 2011 on the Gregorian calendar.[11] This approximation is used only to simplify presentation of the results in graphs, but is not used in the econometric analysis.

4. Methodology and Results

As discussed in Section 2, it is not possible definitively to identify subjects within the datasets who were eligible to apply for college but unable to do so during the Cultural Revolution, due to closure of the universities. Because of institution of mandatory military service, however, age (or birth year) at the time of the Cultural Revolution can be used to identify the cohort of men who would have entered college but were unable to do so due to closure of the universities, i.e. the cohorts who turned 18 in 1980, '81, or '82 (born between 1962-64). These cohorts were directly affected by the Cultural Revolution, while the cohorts who were already attending university at the start of the

---

[10] Each month in the Persian Calendar matches a zodiac sign. The Persian calendar was devised by mathematicians and astrologers in the 11th century upon request from the king, Jalal-e-Din Malek Shah. As a result, it is also called the Jalali calendar. Although widely in use since then, on February 21, 1911, the Iranian parliament adopted it as the official calendar. On March 31, 1925, the parliament made small modifications in the number of days in each month to make the calendar simpler. Based on that, the first six months have 31 days and the next five months have 30 days. The last month has 29 days, or 30 days in leap years. The time a year ends and a new year begins is not at midnight (like the Gregorian calendar). Instead, it is the exact time on March 20th or a close date when the spring equinox occurs (this time is when the distance between the earth and the sun is at its average). This specific time can be precisely calculated. For instance, the Year 1395 started at 8:04:30 pm on March 20, 2016 (Tehran time). Each month in the calendar matches a zodiac sign.

[11] This is in fact what one finds in the harmonized IPUMS birth year and month variables (i.e, `birthyr` and `birthmo`). They are based on Gregorian calendar. But, unfortunately, there is a mistake in the IPUMS data: January and February of each year is incorrectly attributed to the prior year. This is because 2011 corresponds to 1390. Although the last three months of 1390 belong to 2012, in the data they belong to 2011 because 1390 is solely associated with 2011.

Cultural Revolution in 1980 were partially affected, as their education was interrupted or ended due to the Cultural Revolution: these cohorts are comprised of individuals who turned 18 in 1977, '78 or '79 (birth cohorts of 1959-1961).

This section seeks to identify the impact of the Cultural Revolution on the education and wages of these two cohorts. First, the descriptive evidence on deviations from the trend is discussed. Second, regression discontinuity design (RDD) is employed to estimate the causal effect of the Cultural Revolution on education by comparing the outcomes of the last cohort that was cohort directly affected (1964) with the next cohort for which universities were opened (1965). Third, RDD and 2SLS are employed to estimate the impact of the Cultural Revolution on wages and returns to education.

4.1 Deviations from the Trend

The first step consists of examining the college attainment rates across cohorts. Figures 1(a) and (b) describe these rates for men and women, respectively. Birth year, from 1931 to 1984, is shown on the horizontal axis, while the share of the population with college or above education (undergraduate or graduate level education) is shown on the vertical axis. Each data point represents the share of any given birth cohort who had college or above education. As discussed in the data section, this includes partial completion of each level as well. These cohorts are aged between 22 to 80 years of age in the censuses collected in 2006 and 2011. Cohorts fully or partially affected by the Cultural Revolution are marked with dark gray or light gray bars, respectively.

As depicted in Figures 1(a) and 1(b), there is an upward trend in college attainment rates as one moves from older to younger cohorts. This has been documented in the past in Salehi-Isfahani (2005) and Majbouri (2010), among others. Examination of Figure 1(a) reveals that the share of men with college or above education in cohorts who were fully affected by the Cultural Revolution (i.e.

1962, '63, and '64 cohorts) deviates downwards from this general trend. The rate is approximately between 1 to 2 percentage points lower than what it could have been if the pre-Cultural Revolution trend continued uninterrupted for these male cohorts (14% vs. 15.5%).

Interestingly, it is not (visually) clear in Figure 1(b) whether there was any decline in the share of women with college and above education among the affected cohorts. The trend appears to continue upward (without any discontinuity), even for cohorts affected by the Cultural Revolution. The rise in female college attainment has been exponential and faster than males, and by the late 2000s, two thirds of students enrolled in college were women.

Aside from visual observation, more careful consideration of the data involves estimating the deviation from the trend for the affected cohorts, and testing for statistical significance of any deviation. In brief, one assumes various functional forms for the trend, and estimates the deviations from those functions at the time of the Cultural Revolution. The result of this analysis shows that deviation from the trend for the male cohorts who graduated from high school and faced the closed university doors is about 1.2 percentage points[12]. For women, however, this deviation is sensitive to the specification and switches signs (negative to positive). Detailed explanation and report of the results is provided in Online Appendix A. Note that, although the deviation from the trend for men (the regressions in Online Appendix A) is unique to the 1961-64 cohorts and is consistent with the expected effect of the Cultural Revolution, it only shows the correlation between the Cultural Revolution and college attainment and does not represent a causal effect.

[12] This is the deviation from a quadratic functional form which given the shape of the curves seems to be the best fit to both male and female college attainment rates.

4.2 Regression Discontinuity Using Birth Years

As the Cultural Revolution results in closure of universities for the three cohorts who were born just before 1964, with subsequent re-opening of universities to cohorts born after 1964, a discontinuity in attainment rates at 1964 is expected.

Figures 2(a)-(c) fit three different local polynomials (first order, second order, and local smooth polynomial, respectively) to the cohort of men born before and after the 1964 birth cohort. Data for the same process in the sample of women born before and after the 1964 cohort is shown in Figures 2(d)-(f). The horizontal axes represent the birth year cohorts measured as the number of years after the threshold year 1964. The vertical axes indicate the share of each cohort who partially or fully completed undergraduate or graduate schools. For men, the visual examination of the graph suggests a sharp discontinuity at 1964; cohorts born immediately prior to 1964 have lower attainment rates than those just after 1964. The same result is not seen for the female cohorts. Instead, the discontinuity for women is positive and small in a linear regression (Figure 2(d)), but negative and small in the others. Thus, results for women seem inconclusive.

Figure 2 suggests that an RDD is a natural solution to estimate the causal effect of the Cultural Revolution. The running variable in this RDD is birth year, which is a discrete variable.[13] There are only 54 birth year cohorts with sufficient observations who were old enough in 2006 or 2011 to have attended college, namely the birth cohorts between 1931 and 1984. The few discrete values of the running variable means that recently developed optimal bandwidth choice methods and local linear estimations may not be employed, and standard polynomial regressions should be used. To

[13] Figure OB1 in the Online Appendix B is a comparable set of figures that uses the full data, i.e. 54 birth year cohorts between 1931 and 1984.

check for robustness, however, various bandwidths can be examined in these standard regressions. Table 3 shows the result of the following standard RDD:

$$C_i = \alpha + \beta D_i + \sum_{l=1}^{L} \gamma_l (B_i - 1964)^l + \sum_{l=1}^{L} \delta_l D_i (B_i - 1964)^l + u_i \,, \quad L = 1, 2 \qquad (1)$$

in which C_i is a dummy equal to one if individual i has partially or fully attended undergraduate or graduate school. B_i is the year individual i was born (i.e. birth cohort). $B_i - 1964$ is the running variable in the RDD.[14] L is the degree of the polynomial control for the running variable; Gelman and Imbens (2018) argue that employing higher than 2nd order polynomials is not beneficial, and so the current model is restricted to 1st and 2nd order polynomials. D_i is a dummy variable and is set equal to one if individual i was born in or before 1964, and is set to zero otherwise. The coefficient of D_i, i.e. β, shows the local average treatment effect of the Cultural Revolution. Lee and Lemieux (2010) suggest that, when the running variable is discrete, standard errors should be clustered at the discrete values of the running variable. Following this, all regressions in this model include cluster-robust heteroskedastic standard errors. If standard errors are not clustered or are clustered based on the 60 geographic regions of Iran, however, we get smaller standard errors and hence, more statistically significant estimates of the effects.

The left and right sections of the top panel in Table 3 report the local average treatment effect, β, for men and women, respectively. Fifteen birth cohorts before and after the threshold (i.e. 1964) were included in these regressions. In other words, the sample contains the birth cohorts of 1950

[14] 1964 is the cutoff point based on the Gregorian Calendar. As the birth year is reported in Persian Calendar format in the data, however, 1964 is equivalent to 1343 in the data. Therefore, B_{is} is the Gregorian birth year subtracted by 1343. Following Cattaneo et al. (2017), the birth year 1343 is changed to 1342.999 to make sure it falls on the other side of the threshold.

through 1979. All specifications for men, reported in Columns (1) and (2), show a negative and statistically significant effect of Cultural Revolution on attainment rates.

Similar to Figures 2(d)-(f), the results for women (Columns (4) and (5)), depict little evidence of discontinuity. There is only a statistically significant discontinuity for the first order polynomial setting, which switches sign and becomes insignificant in the second order specification. Interestingly, as in Figure 2(d), the coefficient in the first order setting is positive, which implies the unexpected result that the Cultural Revolution increased female educational attainment. Figures 1(b) and 2(d)-(f), however, show that the trend in college educational attainment for women is not linear, but is quadratic, with the result that the correct specification is not the first order polynomial. This implies that the results of the first order polynomial are unreliable, and that there was no effect of the Cultural Revolution on women's college attainment rates.

Since the number of mass points of the discrete running variable (i.e. the number of birth years) is small, the optimal bandwidth choice methods cannot be implemented. Therefore, one needs to check the robustness of the RDD using various bandwidths. The bottom panel shows the local average treatment effect when we change the bandwidth from 15 years to 8, 10, 12, and 20 years before and after 1964. Each row represents a bandwidth size and each column is a polynomial order (first and second order). For men (the left section of the bottom panel), the results are always statistically significant (mostly at 1% level). The results robustly show that men's college attainment rate decreased by about 1.5 to 2.2 percentage points (about 10-17%) due to the Cultural Revolution.

As the right section of the bottom panel shows, for women, there is no discontinuity in almost all specifications and bandwidth choices. In the two out of ten instances that the discontinuity is statistically significant, it takes a positive sign, implying that the Cultural Revolution strangely increased educational attainment. But, both of these positive coefficients belong to the first order polynomial specification and as mentioned above, are unreliable. Therefore, there is no evidence that the Cultural Revolution affected women. These results will be discussed further in Section 4.4.

4.3 Regression Discontinuity Using Birth Months

In Iran, as in many other countries, children can enroll in primary schools at the age of six. More specifically, those who reach age six before the new educational year begins (September 23rd of every year), can start their school in the same year. But those who reach age six after the educational year starts must wait to enroll in the following year. For example, a child born in July 1964 would have turned six in July 1970, and would thus have enrolled in and started first grade in September 1970 for the 1970-71 academic year. A child born in December 1964, however, would have turned six years of age after September 1970, and thus would not have been able to enroll in and start first grade until September 1971. In this way, not only birth year but also birth month create a discontinuity in the year an individual starts in the first grade.

The educational year in Iran starts on the first day of *Mehr*[15], the seventh month of the year, which corresponds to September 23rd in the Gregorian calendar. As parents of children born in Mehr, and occasionally in *Ābān*[16] (the eighth month), often feel it is unfair to have to wait an additional year for their child to start school, simply due to missing the cut-off by a few days, they often petition (usually successfully) to enroll their child in school in the year the child turns six. Thus, in practice, Ābān is a better cut-off point for the birth month the school year starts.

At the time of the Cultural Revolution, as now, every Iranian male is obligated to join military service in the month he turns 18 years old, with the exception of men who are enrolled in school. In

[15] Mehr is pronounced /'mehr/ (using Meriam-Webster pronunciation keys) and corresponds to Libra, the seventh zodiac sign. As mentioned in footnote #7, each month in the Persian Calendar corresponds to an astrological sign.

the case of a boy born a month before Ābān in 1964, he would be able to attend first grade in 1970, while a second boy born a month after Ābān of 1964 would not be able to attend first grade until 1971, i.e. a year later than the boy born before Ābān of 1964. Eleven years later, in December of 1982, when the Cultural Revolution was about to end and the universities were about to re-open in the following year, the second boy would have been in the last year of high school, and thus able to enroll in university after universities re-opened, while the first boy would have finished high school and would have already been drafted into military service. The first boy would have also been less likely to enroll in college during his military service and for a period of several years after the end of the Cultural Revolution, as there was an upper age limit on who was eligible to enroll in university after the Cultural Revolution. As a result of these factors, there is an expected discontinuity in college educational attainment based on whether an individual was born before or after Ābān 1964.

Figures 3(a)-(c) and 3(d)-(f) illustrate this discontinuity for men and women, respectively. The running variable on the horizontal axes is the birth month, measured as the number of months since the threshold month of Ābān 1964 (months prior to Ābān 1964 are negative, and those following Ābān 1964 are positive). The vertical axes show the proportion of total population in each cohort with some degree of university-level education. Men born just before the threshold appear to be less likely to have any tertiary educational attainment. Based on the results for birth year (Table 2), one expects to see no discontinuity at the threshold for women, which is supported by figures 3(d)-(f).

The running variable, month of birth since Ābān 1964, is a discrete variable. As it can take over 600 values, we can employ optimal bandwidth selection procedures as well as local polynomial estimation (for more information, see Calonico et al., 2017; Cattaneo et al., 2014a,b; and Cattaneo et al., 2017). Table 3 reports the results for such estimations. There are two major methods of optimal bandwidth selection in the literature: 1) Coverage Error Rate (CER) and 2) Mean Square Error

(MSE). Each of these methods can find the same optimal bandwidths above and below the threshold (termed R.D.) or different optimal bandwidths above and below the threshold (termed Two-sided from now on). Thus, four different optimal bandwidths choice methods can be implemented: CER-Two-Sided, CER-R.D., MSE-Two-Sided, and MSE-R.D. For robustness check, results are estimated for all four of these choices, as shown in Table 3. The data extends from the 1931 to the 1984 birth cohorts. 1984 was chosen to ensure the youngest age in the sample is roughly 22 years of age (the 1984 cohort is about 22 years old in the 2006 census). 1931 is chosen to ensure a substantial number of observations in the oldest cohort (the 1931 cohort is 75 years old in the 2006 census and 80 years old in the 2011 census). The threshold is Ābān 1964. For the data range selected, there are 245 birth month cohorts above (after) the threshold and over 400 below (before) it. Since the data is not symmetric around the threshold, separate optimal bandwidths should be selected for the data above and the data below the threshold. Therefore, results for two-sided approaches (CER–Two Sided and MSE–Two Sided) are more relevant.

Table 3 depicts the results based on each of these optimal bandwidth methods with bias correction procedures.[17] The top panel reports the discontinuity in educational attainment conditional on a first and second order local polynomial of the running variable when the optimal bandwidth choice method is CER-Two Sided. The results support those found in Table 2 using birth year as the running variable. For men, there is a discontinuity of about 1.8 to 2.2 percentage points in college attainment rates. Both estimates of discontinuity for women in Columns (3) and (4) are statistically insignificant and close to zero.

The bottom panel reports the discontinuity using other optimal choice methods: CER-R.D., MSE-Two-Sided., and MSE-R.D. Each row reports one of these methods and each column presents an

[17] Robust and Conventional Estimates are almost identical to these results. They are reported in the Online Appendix, Tables OB3 and OB4.

order of the polynomial. Interestingly, all results for men are similar and statistically significant regardless of the order of the local polynomial used or the optimal choice method implemented. The discontinuity ranges only from 1.0 to 2.2 percentage points. Coefficients for women are similar to those found in the top panel as well, i.e. they are small and statistically insignificant.

As an alternative to using the individual level data, one can employ the average of college and above dummy for each birth month cohort (i.e. the share of individuals with college and above education in each birth month cohort) as the dependent variable in the RDD. The number of observations is reduced to the number of birth month cohorts, i.e. 648. The estimates, reported in Tables OB5-OB7 in the Online Appendix, are very similar to those reported using the individual level data in Table 3, but they only range between 1.4 and 2.2.

4.4 Discussion

During the Cultural Revolution in Iran, all universities were closed for 30 months. Three cohorts of young men faced immediate draft into the longest war of the twentieth century, which resembled World War I in terms of military tactics and, hence, casualties[18]. Yet, unexpectedly, these cohorts are only slightly less college-educated than their following cohorts.

One potential explanation is that the affected cohorts disproportionately died during the war, reducing the overall population of those cohorts. As a result, fewer of the survivors would have to attend college to achieve the same overall educational attainment rates as cohorts not affected by the Cultural Revolution and war. If this is true, we should see a discontinuity in the size of cohorts at the threshold. Figures A1-A3 in the appendix plot the number of observations in each birth year and

[18] Some scholars compare this war to World War I because of Iranian human wave attacks, bayonet charges, large-scale trenches with barbed wire along the front lines, extensive use of chemical weapons by Iraq, and deliberate attacks on civilian targets (Abrahamian 2008). Iranian use of human wave attacks and Iraq's extensive use of chemical weapons resulted in casualty rates on the Iranian side comparable to casualty rates in World War I.

birth month cohort and fit multiple polynomial functional forms to the data. The figures present little evidence of discontinuity in the size of cohorts at the threshold. If anything, the cohorts affected by the Cultural Revolution (cohorts just below the threshold) have more observations than those who came afterwards.

Table A1 reports the results of estimating Equation (1) when the dependent variable is the number of observations in the birth year cohorts. This table is in the same format as Table 2. The results for the bandwidth of ±15 years are reported in the top panel, while the bottom panel depicts statistical significance levels of the discontinuity in cohort size for various bandwidths. The results for men show that there is either no statistically significant discontinuity in the cohort size or, when there is, the cohorts affected by the Cultural Revolution are larger in size. Overall, there is no evidence of a smaller cohort size for the affected cohorts. Therefore, the reason behind the relatively higher-than-expected attendance rates for the affected cohorts is not that they were less likely to survive the war. If anything, they were more likely to survive (note that the next cohorts for whom the university doors opened were also drafted into war if they could not enter college.) For women, understandably, the discontinuity in the cohort size is almost always statistically insignificant and switches signs depending on the specification.

One can use the birth month cohorts and implement an RDD on cohort size using various optimal bandwidth choice methods. The results of this approach are shown in Table A2, which is in the same format as Table 3.[19] For both men and women, there is no discontinuity in the cohort size regardless of the optimal bandwidth choice and the polynomial order of the running variable. These results are all rejecting the above hypothesis that affected cohorts achieved a relatively higher than expected college attainment rate (despite the Cultural Revolution) because they were smaller than

[19] Conventional and Robust Estimates of this RD (which are similar to Bias-Corrected estimates) are reported in Tables OB8 and OB9 in the Online Appendix.

their following cohorts. According to the best estimates, the total number of casualties during this war (on the Iranian side) ranges from about 200,000 to 600,000 (Hiro, 1991; Abrahamian, 2008; Bercovitch and Jackson, 1997; Clodfelter, 1991; Palmowski, 1997; Mikaberidze, 2011), and these casualties were spread across many birth cohorts involved in the war. In fact, war casualties increased in later years. Hence, the size of post-Cultural Revolution cohorts, who were later drafted into war, could have been affected even more by the war. Therefore, the casualties from war should have had little impact on the size of the affected cohorts relative to other cohorts.

One may argue that the small effect of the Cultural Revolution on educational attainment is due to competition between the cohorts affected by the Cultural Revolution and the post-Cultural Revolution cohorts for a limited number of spots at university, and that this reduced the chances of those in the post-Cultural Revolution cohorts to enroll in university. This is not factually correct, as there were age restrictions on who could enter college after the Cultural Revolution. These age restrictions barred the Cultural Revolution cohorts from entering college for several years, which means they did not present competition for several cohorts who came after the Cultural Revolution for positions at university. In other words, at the threshold, there is no spill-over effect from pre- to post-Cultural Revolution cohorts.

Another explanation for the small effect of the Cultural Revolution on educational attainment is that supply of higher education increased in later years, particularly during the 1990s after the war. In fact, the Cultural Revolution reduced the supply of higher education in the years that followed it; according to statistics from the Ministry of Higher Education, almost 8000 professors, or close to half of the total university faculty members in Iran, were purged[20] (Ashraf 1997). This purge significantly reduced the capacity when the universities re-opened. To compensate, many of the

[20] Faculty members who were

remaining professors were asked to teach at more than one university, and some even flew on a regular basis to other cities to teach additional courses.[21] Although there had been attempts to rejuvenate the supply of higher education by establishing a teacher/professor training college in Tehran and founding a private university with branches (campuses) across the country in small towns, it took years for the higher education system to reach the pre-Cultural Revolution supply levels.

A similar argument for the small discontinuity in the educational attainment rate is that the affected cohorts went to war. Veterans of war were subject to special quotas. These quotas gave them a higher chance to enter college than the regular college applicants. But, without a large demand for higher education, increases in supply could not have changed the outcome by its own. In developing countries, people rarely return to school once they have dropped out of the education system, which is part of the reason maintaining children in the educational system is a top priority for education policy makers. The experiences of the Iran-Iraq war left physical and mental scars on the affected cohorts, including post-traumatic stress disorder (PTSD). Moreover, these cohorts had to wait several years upon their return due to the age restrictions on who could enter college after the Cultural Revolution. The fact that these cohorts pursued education despite these barriers suggests a strong demand for higher education.

The effect on women's educational attainment is practically zero, though the situation is slightly different for women than for men. In Iran, the common perception is that a woman can only gain status in society and within her family through education. This suggests there are stronger incentives for women than for men to pursue higher education. In fact, since the 2000s, over 60% of college

[21] This had a toll on research productivity of the faculty as well. Khosrokhavar et al. (2014) report that the number of respectable journal publications decreased from 450 in 1979 (prior to Cultural Revolution) to at most 120 in 1985 (three years after universities re-opened.)

enrollees in Iran have been women. This strong demand for education suggests the effect of closure of the universities on women's educational attainment would be smaller than on men's attainment. In addition, only 5% of women in cohorts close to the Cultural Revolution period enrolled in college. In other words, this small fraction of women who enrolled during that period likely had such a strong demand for university education that they would have enrolled as soon as the opportunity presented itself.

4.5 Returns to College Education

A natural next step is to see if this exogenous decline in men's educational attainment due to the Cultural Revolution had any impact on wages. Unfortunately, wage rates (or income levels) are not reported in the census data, but job categories are. HIES data include both income and hours worked since 2006, which can be used to estimate hourly wages for salaried workers. In this section, the impact of the Cultural Revolution and its exogenous increase in education on job status using census data is calculated, and subsequently the impact on wages is estimated using the HIES data.

The data does not include a variable to indicate job status but occupation is reported in a variable named `occisco` (Occupation, based on ISCO standard), which has ten categories: 1) Legislators, senior officials and managers, 2) Professionals (doctors, lawyers, engineers, etc.), 3) Technicians and associate professionals, 4) Clerks, 5) Service workers and shop and market salespeople, 6) Skilled agricultural and fishery worker, 7) Crafts and related trades workers, 8) Plant and machine operators and assembly workers, 9) Elementary occupations, and 10) Armed forces. Excluding the armed forces, these job categories are arranged hierarchically, with legislators, senior officials, and managers in the top category and the elementary occupations in the bottom. Armed forces are excluded from this calculation; "armed forces" is not an informative job status category, as there is significant heterogeneity of jobs in the armed forces and job status in the armed forces depends on the rank of the individual. Additionally, including “armed forces” in the calculations does not affect the results (compare Table OB13 in the Online Appendix with Table 4).

As previously mentioned, job categories seem to be arranged hierarchically. Based on this hierarchy, one can convert this categorical variable into an ordinal variable for job status. Converting job categories into a quantitative measure of job status, however, is somewhat challenging. For example, how does one quantify the difference in job status between categories, e.g. between Technicians and associate professionals (category 3), and Clerks (category 4)? One subjective solution is to aggregate job categories of perceived comparable status into one group and separate those that differ significantly into other groups. For example, categories 8 and 9 can be combined into one group (and the lowest value of job status, i.e. 1 is assigned to them); categories 5, 6, and, 7 are aggregated into another group (with value 2 assigned to their job status); categories 3 and 4 into a third group (with the value of 3 for job status), and categories 1 and 2 into a fourth group (with value 4 assigned to their job status). Although the results based on this definition of job status are reported in Table OB10 in the Online Appendix, there are at least two issues with this (or any) subjective definition of job status: first, how one decides which categories should be combined, and second, what values one assigns to each of these categories (the values assigned above assume that for example, categories 1 and 2 are four times better than categories 8 and 9 and two times better than categories 5, 6, and 7.) To mitigate these issues, the log of average wage for each job category is used as a measure of status for that job category. These average wages are both a cardinal and ordinal measure, and are a measure of economic returns. Census data, however, reports neither income nor hours worked. Therefore, HIES data from 2006 through 2015 are used to calculate average wages (adjusting for inflation). The results are shown in Table A3 (Appendix), and demonstrate a clear hierarchy among these job categories. Moreover, a closer look at the results reveals that the subjective definition of job status described above is not unreasonable.

Once the job status variable was defined as above, two methods were employed to estimate the impact of the Cultural Revolution. First, job status is defined as the dependent variable in Equation (1), which allows the discontinuity in job status (log of average wage of a job category) due to the Cultural Revolution to be estimated, i.e. the reduced form. The discontinuity is reported in Panel A of Table 4. Columns (1) and (2) are first and second degree polynomial RDD estimates using birth year data, and Columns (3) and (4) are based on birth month data. All estimates are statistically significant, and three of them show a 2.2-2.4% reduction in the average wages of the job category one holds (job status) due to the Cultural Revolution. The lowest estimate is -1.2%. In Panel B, 2SLS is employed to estimate the causal impact of college attendance on job status, using Cultural Revolution as an instrument for college attendance. The first stage is the RDD reported in Tables 2 and 3. The discontinuity is the exogenous variable (instrument). Similar to Panel A, the first two columns use the first and second degree polynomial RDD, using birth year data in the first stage, and the last two columns are based on birth month data. The birth year results show that those who attended college had a 66-88% increase in their job category average wages (job status). The estimates in the last two column, which have birth month RDD in the first stage, show a return of about 100%. The estimate in the last column, which uses a second degree polynomial RDD in the first stage on birth month data, is not reliable, as the instrument in its first stage is a weak predictor of college attendance. The Kleibergen-Paap rk Wald F-statistic, which tests the weakness of the instrument, is reported under each estimate. Stock and Yogo (2005) calculate the critical value for this statistic as 16.38. In all 2SLS regressions, except the one in the last column, we reject the null hypothesis that the instrument is weak. Overall, the results show that those who attended college gain about 66-100% in their job category average wages over those without college education. Note that the average person close to the threshold who did not go to college has middle school education.[22]

A more interesting result is one that uses individual level wages rather than job status. As previously mentioned, the census data do not contain information on wages; thus, wages are calculated using HIES annual surveys. A sample of over 100,000 salaried workers whose birth year is ±15 years around the threshold is obtained by combining the HIES data from 2006 through 2015. Using these data and the same methods as used in Table 4, the impact of Cultural Revolution on individuals' wages is estimated. As mentioned in Section 3, however, only birth year can be used to identify discontinuities, as the HIES dataset does not contain birth month. The results are shown in Table 5, which is organized in a similar fashion to Table 4. Panel A reports the discontinuity in the log of wages for men due to the Cultural Revolution, which was calculated using the same first and second degree polynomial RDDs as in Equation (1). They show that the Cultural Revolution reduced wages of individuals within the affected cohorts by about 5% (4.7-5.5%). Panel B reports both the OLS regression and the 2SLS estimates of the return to college and above education. As in Table 4, the 2SLS estimates use the discontinuity due to the Cultural Revolution as an instrument for college attainment. Age and age squared are controlled for by following a simple Mincer equation. The OLS estimate shows an almost 76% return to college and above education, but the 2SLS estimate is larger, at about 100% (Column (2) in Panel B). This 2SLS estimate is close to the numbers found for job status in Table 4. These results show that those with college and above education have wages twice as high as those with less than college education. The average person close to the threshold who did not go to college in these data has middle school education.[23] Note that the result in

[22] The distribution of education around the threshold in the Census data is reported in Table OB11 in the online Appendix.
[23] The distribution of education around the threshold in the HIES data is reported in Table OB12 in the online Appendix.

Column (3) of Panel B is unreliable, as the first stage regression is weak. Overall, these high returns to college education could explain the strong demand for college in the Cultural Revolution cohorts.

One can split the sample into rural and urban areas and estimate the same RDD and 2SLS procedures on these subsamples. The results show that the discontinuity in individuals' wages is statistically significant in both areas, but seems to be larger in urban areas. The 2SLS estimates, however, are only reliable in urban areas, as the first stage is weak in rural areas. The return to college education and above in urban areas is estimated at 88%. These results are reported in Tables OB14 and OB15 in the Online Appendix. One should note that these results are subject to selection, as there has been substantial migration from rural to urban areas over time.[24]

5. Conclusion

The contemporary history of Iran is fascinating and puzzling. In the span of a few months, the monarchy collapsed, and new institutions emerged to dominate the political landscape. Almost overnight, the ruling class were replaced with a new elite class that did not have such political power over the previous centuries. For the first two years following the collapse of the monarchy, Iran was in turmoil and had to deal with civil war, emigration and diaspora of the old elite and educated classes, as well as the Iraqi invasion and ensuring Iran-Iraq war, which became the longest war of the twentieth century.

In the midst of these crises, the Iranian government was eager to establish domestic stability. To this end, the new Iranian government closed all the universities for thirty months. The cohorts who would have normally matriculated to college but were unable to do so due to the closure of the universities, were instead drafted into the Iranian armed forces and fought in the Iran-Iraq war. In

[24] Therefore, I refrain from including them as part of the paper.

this paper, I show that cohorts who were affected by the closure of universities and the draft, were almost as likely to have been college educated as the cohorts after the universities were re-opened. This is evidence for a strong demand for education (among these cohorts) in Iran. This strong demand is also evident in the rapid rise in average years of education during and after the Iran-Iraq war. Human capital investment has been a corner stone of the Iranian household after World War II, and it has remained one of the resilient features of the economics of the household in Iran.

Further anthropological and sociological research is necessary to understand how these affected cohorts gained the education they wanted years after they left high school.

References

Abrahamian, Ervand (2008). *A History of Modern Iran* (3rd print ed.). Cambridge, U.K.: Cambridge University Press.

Ashraf, Ahmad (1997). "Education: General Survey of Modern Education." *Encyclopædia Iranica*, VIII/Fasc. 2, pp. 189-196. Last updated on December 9, 2011, available online at http://www.iranicaonline.org/articles/education-vii-general-survey-of-modern-education (accessed online at June 30, 2018).

Bakhash, Shaul (1984). *The Reign of the Ayatollahs: Iran and the Islamic Revolution*. New York: Basic Books, p. 122.

Banerjee, Abhijit V., and Esther Duflo (2012). *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty.* New York, NY: Public Affairs.

Bercovitch, J., and Jackson, R. (1997). *International Conflict: A Chronological Encyclopedia of Conflicts and Their Management 1945–1995*. Congressional Quarterly Press, p. 195.

Calonico, S., Cattaneo, M. D., Farrell, M. H., and Titiunik, R. (2017). "`rdrobust`: Software for regression discontinuity designs," *Stata Journal*, forthcoming.

Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014a). "Robust data-driven Inference in the Regression-Discontinuity Design," *Stata Journal*, 14, 909-946.

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

Cattaneo, M. D., Idrobo, N., and Titiunik, R. (2017). "A practical introduction to regression discontinuity," Monograph prepared for *Cambridge Elements: Quantitative and Computational Methods for Social Science*. Cambridge, UK: Cambridge University Press.

Clodfelter, Michael (2002). *Warfare and Armed Conflict: A Statistical Reference to Casualty and Other Figures*. Jefferson, NC: McFarland & Company, p. 1618–1991.

Duflo, Esther (2001). "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment," *American Economic Review*, 91(4):795-813.

Duflo, Esther (2004). "The medium run effects of educational expansion: evidence from a large school construction program in Indonesia," *Journal of Development Economics*, 74(1):163-197,

Khosrokhavar, Farhad, Etemad, Shapour and Masoud Mehrabi (2004). "Report on Science in Post-Revolutionary Iran—Part I: Emergence of a Scientific Community?" *Critique: Critical Middle Eastern Studies*, 13(2): 209–224.

Gelman, Andrew and Guido Imbens (2018). "Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs." *Journal of Business and Economic Statistics*, published online at https://doi.org/10.1080/07350015.2017.1366909.

Hiro, Dilip (1985). *Iran under the Ayatollahs*. London: Routledge and K. Paul.

Imbens, G. and K. Kalyanaraman (2012). "Optimal bandwidth choice for the regression discontinuity estimator." *Review of Economic Studies,* 79(3): 933-959.

Lee, D. S. and T. Lemieux (2010). "Regression discontinuity designs in economics." *Journal of* Economic Literature, 48(2): 281-355.

Lergetporer, Philipp, Werner, Katharina, and Ludger Woessmann (2018). "Does Ignorance of Economic Returns and Costs Explain the Educational Aspiration Gap? Evidence from Representative Survey Experiments," IZA Discussion Papers No. 11453.

Majbouri, Mahdi (2010). *Against the Wind: Labor Force Participation of Women in Iran*. Ph.D. thesis. University of Southern California.

Mehran, Golnar (1997). "Education in post-revolutionary Persia, 1979-95." In Yarshater, Ehsan (Ed). *Encyclopedia Iranica*, Vol. VIII, Fasc. 3, pp. 230-233. Also available online at http://www.iranicaonline.org/articles/education-xxiv-education-in-postrevolutionary-persia-1979-95 (accessed on Dec. 23, 2015)

Mitch, David (2004). "Education and Skill of the British Labour Force." In Roderick Floud and Paul Johnson (Eds.), *The Cambridge Economic History of Modern Britain, Vol. I: Industrialisation, 1700-1860*, Cambridge: Cambridge University Press, p. 344.

Mikaberidze, Alexander (2011). *Conflict and Conquest in the Islamic World: A Historical Encyclopedia*. Santa Barbara, California: ABC-CLIO, p. 418.

Nguyen, Trang (2008). "Information, Role Models and Perceived Returns to Education: Experimental Evidence from Madagascar," Working Paper

Palmowski, Jan (1997). *A Dictionary of Twentieth Century World History*. Oxford, UK: Oxford University Press.

Pratham (2005). *Annual Status of Education Report*. Final Edition, available at http://www.asercentre.org/p/85.html (accessed on June 4, 2018.)

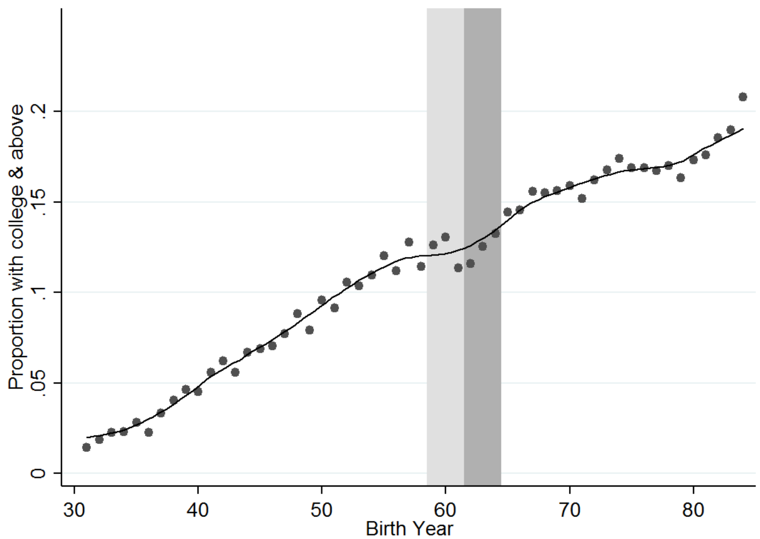
Salehi-Isfahani, Djavad (2005). "Human resources in Iran: potentials and challenges." *Iranian Studies* 38 (1), 117-147.

Souri, Davoud (2004). "Using the IV Method to Estimate Omitted Variable Models: An Application of the IV Method to Estimate the Rate of Return to Education in Iran," Ph.D. Dissertation, Virgina Tech.

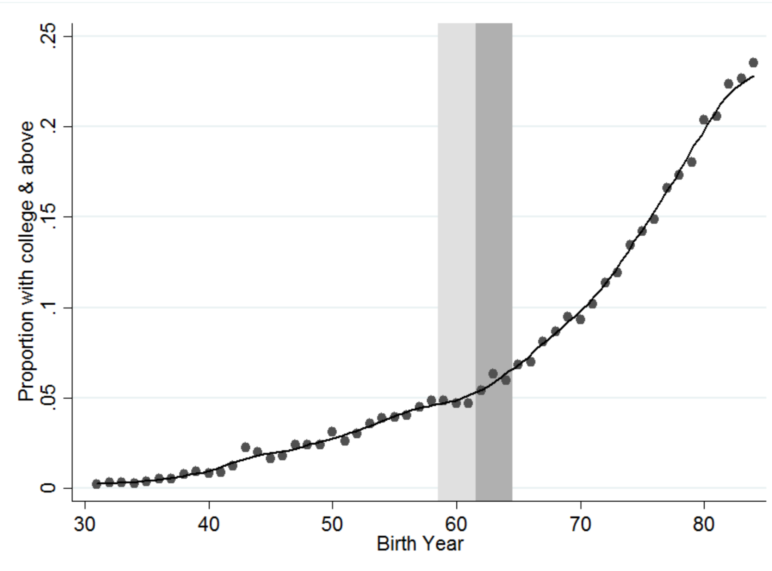
Stock J. and M. Yogo (2005). "Testing for Weak Instruments in Linear IV Regression." In: DWK Andrews (Ed.) *Identification and Inference for Econometric Models*. New York: Cambridge University Press, pp. 80-108.

Figures

Figure 1 – Proportion of men and women with college and above education



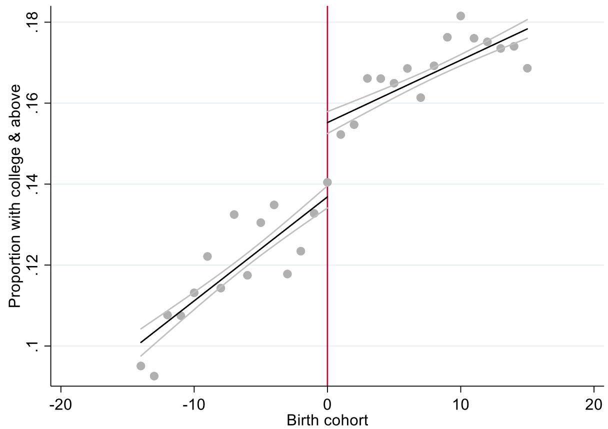
(a) Men



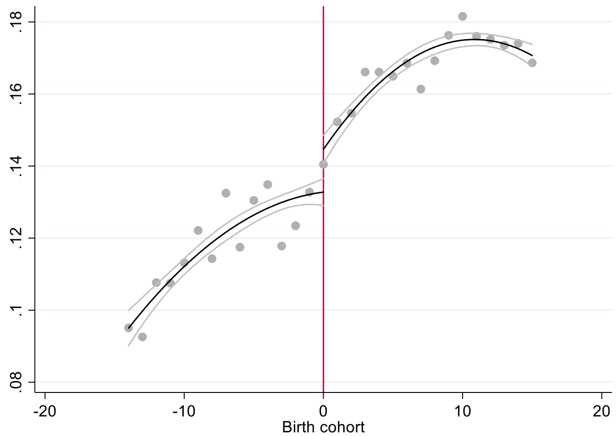
(b)

Note: College and above represents attainment of partial or complete undergraduate or graduate level education. Birth cohorts of 1962, 63, and 64 are shaded with dark gray as they reached age 18 when universities were close. Birth cohorts of 1959, 60, and 61 were in school just before the universities were closed. So their education could have remained incomplete. They are shaded with light gray.

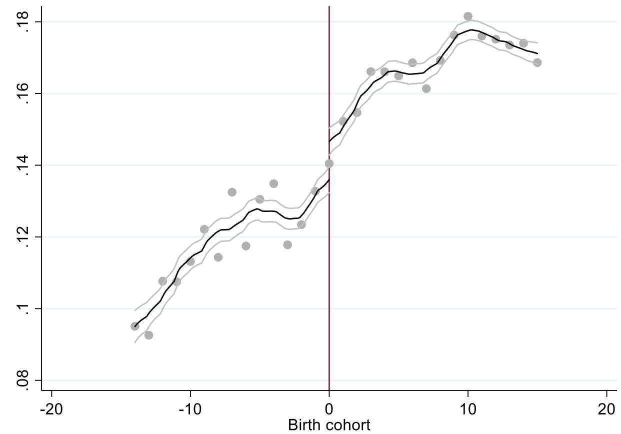
Figure 2 – Discontinuity in Proportion of men and women with college and above education across birth cohorts
(1950 through 1979 birth cohorts)



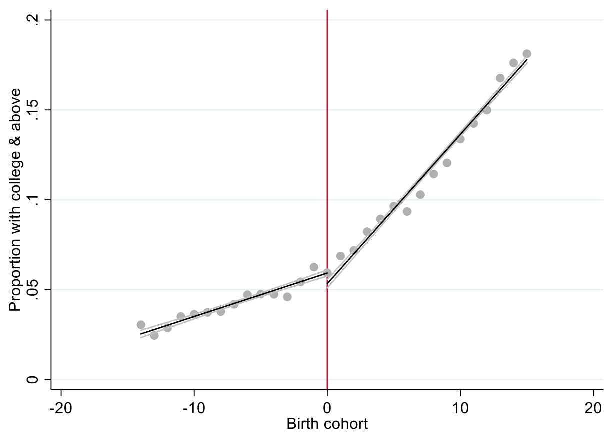
(a) linear – men



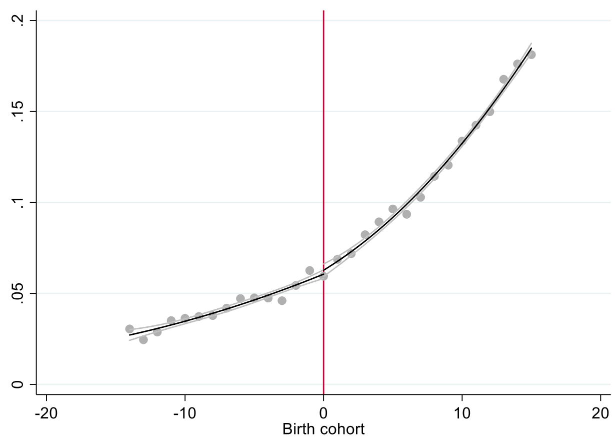
(b) quadratic - men



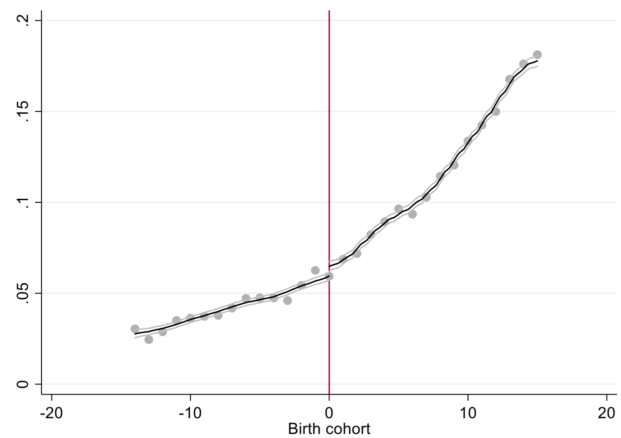
(c) local polynomial smooth line - men



(d) linear - women



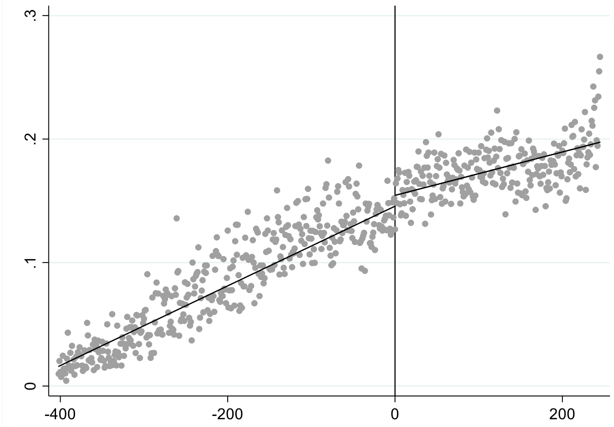
(e) quadratic - women



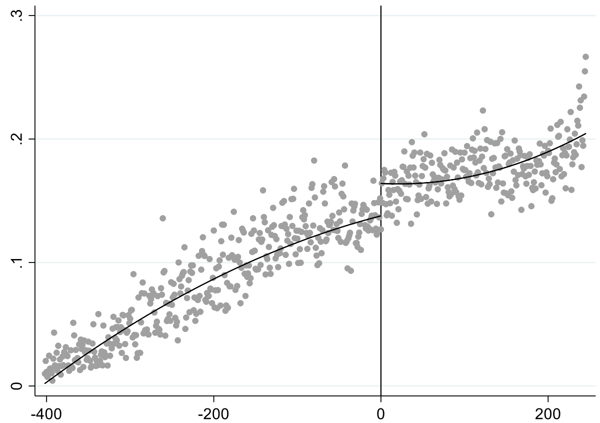
(f) local polynomial smooth line - women

Note: The dots represent the proportion in each birth year cohort who partially or fully completed an undergraduate or graduate degree. The curves depict linear, quadratic and local polynomial predictions respectively.

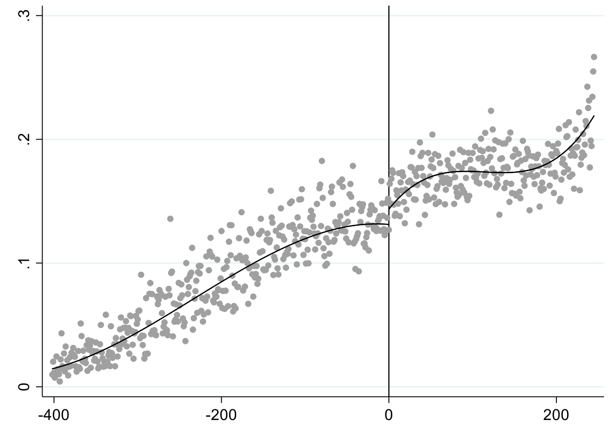
Figure 3 – Discontinuity in the proportion of men and women with college and above education (vertical axis) across monthly birth cohorts (horizontal axis) (1931 through 1984 birth cohorts)



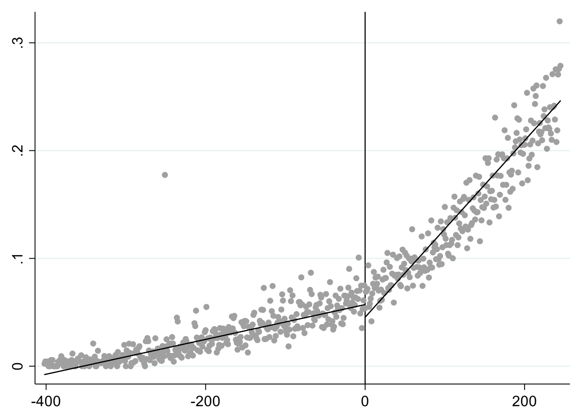
(a) local linear - men



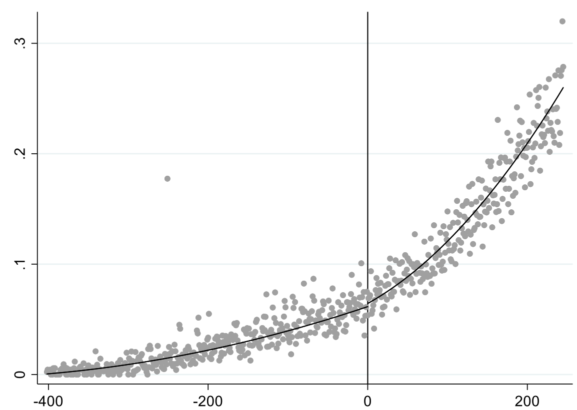
(b) local quadratic - men



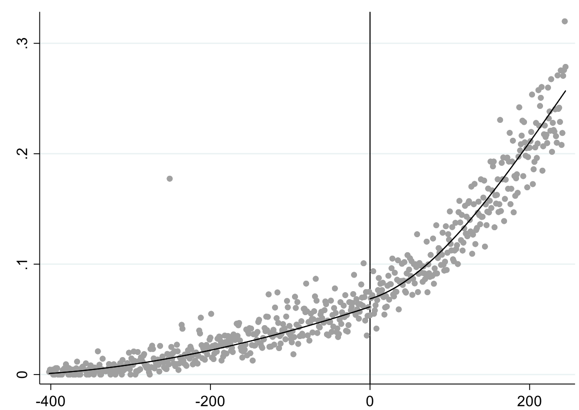
(c) local polynomial of the 3rd order - men



(d) local linear - women



(e) local quadratic - women



(f) local polynomial of the 3rd order - women

Note: The vertical axis represents the proportion of the birth cohort who partially or fully completed an undergraduate or graduate level education. The figures are drawn with `rdplot` command in Stata to find the optimal bin size. The dots represent the proportion in the bin who partially or fully completed an undergraduate or graduate degree. The curves represent first, second, and third order local polynomial predictions.

Tables

Table 1 – Summary Statistics

Variable	Observations	Mean	St. dev.	Min	Max
Panel A – 2006 and 2011 Censuses					
Female	1,419,591	0.50	0.50	0	1
Age	1,419,591	41.31	13.77	22	80
Literate	1,419,591	0.77	0.42	0	1
Primary school	1,419,591	0.25	0.43	0	1
Middle school	1,419,591	0.15	0.36	0	1
High school	1,419,591	0.19	0.39	0	1
College & above	1,419,591	0.13	0.34	0	1
Birth year	1,419,591	1967.03	13.59	1931	1984
Birth month	1,372,865	6.20	3.09	1	12
Panel B – Household Income and Expenditure Surveys, 2006-2015 (Male Wage Earners in 1950 to 1979 Birth Cohorts)					
ln(wage)	107,814	11.47	0.62	7.20	16.45
Age	107,814	41.79	7.75	27	65
Urban	107,814	0.55	0.50	0	1
College & above	107,814	0.12	0.33	0	1

Note: Female is a dummy equal to one if the individual is a female and zero otherwise. College & Above is a dummy variable equal to one if the individual has attended college or graduate school. Note that when Birth years are reported in Gregorian calendar but they are in Solar Hijri (Persian) calendar in the dataset. The sample for censuses includes individuals whose birth year is between 1931 and 1984. The sample for Household Income and Expenditure Surveys include only male wage earners in 1950-1979 cohorts. ln(wage) is the natural log of gross wages for wage earners. Gross annual salary is divided by the hours worked per week.

Table 2 – College Attendance and Cultural Revolution,
Bandwidth in the Top Panel: ±15 years
(i.e. the birth cohorts are between 1950 and 1979)

	Men		Women	
	(1)	(2)	(3)	(4)
Discontinuity	-0.022***	-0.015**	0.007*	-0.004
	(0.004)	(0.005)	(0.004)	(0.003)
Polynomial degree	1st	2nd	1st	2nd
Observations	470,850	470,850	467,912	467,912
Coefficient of D when the bandwidth is:				
±8 [i.e. 1957-1972]	-0.022***	-0.008**	-0.002	-0.001
	(0.006)	(0.003)	(0.003)	(0.004)
±10 [i.e. 1955-1974]	-0.018***	-0.018***	0.000	-0.005
	(0.005)	(0.006)	(0.003)	(0.004)
±12 [i.e. 1953-1976]	-0.019***	-0.017***	0.002	-0.005
	(0.004)	(0.006)	(0.003)	(0.003)
±20 [i.e. 1946-1984]	-0.012**	-0.031***	0.016***	-0.003
	(0.005)	(0.008)	(0.005)	(0.003)

Note: Dependent variable is a dummy equal to one if the individual partially or fully completed college and above levels, and zero otherwise. Following suggestion by Lee and Lemieux (2010), robust-heteroskedastic standard errors are corrected for correlation within each birth year cohort, as this running variable (birth year) is a discrete variable. No covariates other than those in Equation (1) is controlled for. See Tables OB1 and OB2 in the Online Appendix B for results with no clustering of standard errors and geographic clustering of standard errors respectively.
*** p<0.01, ** p<0.05, * p<0.10

Table 3 – College Attendance and Cultural Revolution - Bias-Corrected Optimal Bandwidth Choice, Birth Month Cohort

	Men		Women	
	(1)	(2)	(3)	(4)
Discontinuity†	-0.018***	-0.022***	-0.003	-0.006
	(0.006)	(0.008)	(0.004)	(0.005)
Polynomial degree	1st	2nd	1st	2nd
Observations	686,188	686,188	686,677	686,677
Discontinuity when the optimal bandwidth choice method is:				
Coverage Error Rate – RD	-0.019***	-0.022***	-0.007	-0.010*
	(0.007)	(0.009)	(0.005)	(0.006)
Mean Square Error – Two Sided	-0.016***	-0.010*	-0.001	0.002
	(0.004)	(0.005)	(0.003)	(0.003)
Mean Square Error – R.D.	-0.010**	-0.009*	0.001	0.003
	(0.005)	(0.005)	(0.003)	(0.004)

Note: This table reports RDD estimates of discontinuity in a dummy which is equal to one if the individual went to college or above and zero otherwise. Birth month is the running variable. The top panel uses Coverage Error Rate – Two Sided (two different optimal bandwidths above and below the threshold) as the optimal bandwidth method. The bottom panel reports the same coefficient for other optimal bandwidth choice methods, for robustness check. All estimates reported in this table are Bias-Corrected RD estimates. For Conventional and Robust RD estimates, see Tables OB3 and OB4 in the Online Appendix B.
*** $p<0.01$, ** $p<0.05$, * $p<0.10$
† Optimal bandwidth choice method is Coverage Error Rate – Two Sided (two different optimal bandwidth choices above and below the threshold).

Table 4 – Men's Job Status, College Attendance, and Cultural Revolution
(Census data; using birth year and birth month)

	Birth Year†		Birth Months‡	
	(1)	(2)	(3)	(4)
Panel A – RDD Estimate of the Impact of Cultural Revolution on Job Status				
Discontinuity in Job Status	-0.023***	-0.014**	-0.022***	-0.024***
	(0.006)	(0.005)	(0.006)	(0.007)
Polynomial degree	1st	2nd	1st	2nd
Observations	370,805	370,805	495,928	495,928
Panel B – Second Stage of the 2SLS Estimate of the Impact of College Attendance on Job Status (using Discontinuity in College Attendance, due to Cultural Rev., as IV)				
College Attendance	0.660***	0.876***	0.942***	1.221***
	(0.049)	(0.173)	(0.180)	(0.327)
Kleibergen-Paap rk Wald F-Statistic$^{\Pi}$	223.7	18.50	18.05	7.602
Polynomial order of RD in the 1st stage	1st	2nd	1st	2nd
Observations	370,805	370,805	495,928	495,928

Note: The table consists of two panels and four columns. The first two columns in both panels use the birth year analysis and the last two columns are based on birth month analysis. For birth year data, the bandwidth is chosen as ±15 years, i.e. those born between 1950 and 1979. For birth month data optimal bandwidth and bin size is selected using Coverage Error Rate-Two sided procedure. The top panel reports the discontinuity in men's job status (measured as the average wage in a job category). The bottom panel reports the second stage of a 2SLS that uses discontinuity in Cultural Revolution in an RD as instrument for college & above education to estimate the impact of college education on job status. In the analysis using the birth year data (Columns (1) and (2)) and following a suggestion by Lee and Lemieux (2010), robust-heteroskedastic standard errors corrected for correlation within each birth year cohort (the running variable) are estimated and reported in parentheses, because the running variable is discrete. For birth month data (Columns (3) and (4)), Bias-Corrected RD estimates are reported in this table.

*** p<0.01, ** p<0.05, * p<0.10

† Bandwidth is ±15 years, i.e. those born between 1950 and 1979.

‡ Optimal bandwidth choice method is Coverage Error Rate – Two Sided (two different optimal bandwidth choices above and below the threshold). Bias-Corrected coefficients and standard errors are reported.

$^{\Pi}$ This is the Kleibergen-Paap rk Wald F statistic for weak instruments. The 10% critical value of this statistic, calculated by Stock and Yogo (2005), is 16.38.

Table 5 – Men's Log of Wages, College Attendance, and the Cultural Revolution (Bandwidth: ±15 Birth Years; 1950 through 1979)

Panel A – Discontinuity in Men's Log of Wages Due to Cultural Revolution

	(1)	(2)
Discontinuity in log of Wages	-0.055***	-0.047**
	(0.011)	(0.019)
Polynomial degree	1st	2nd
Observations	107,814	107,814

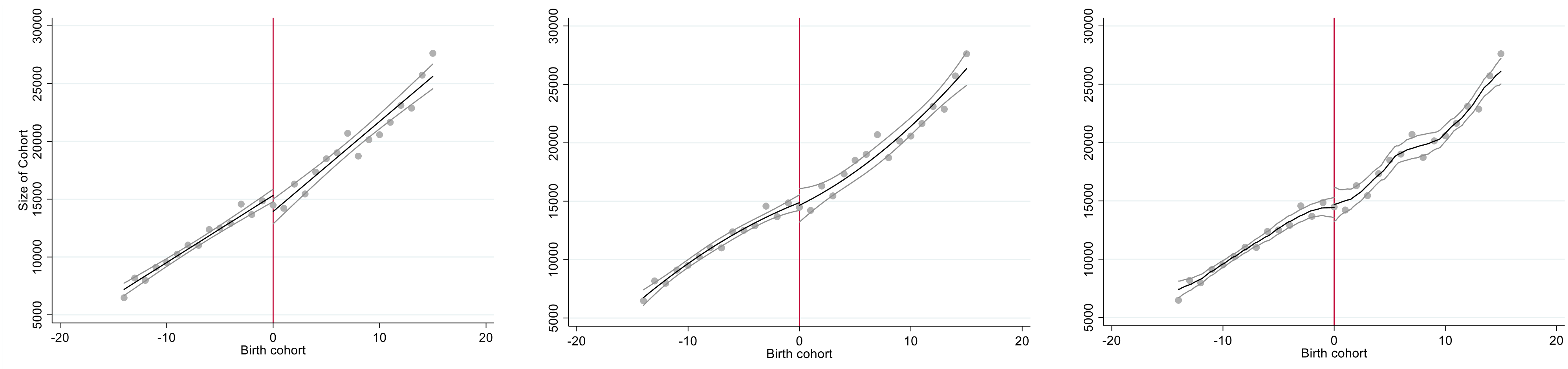
Panel B – Returns to College Education for Men Using the Discontinuity from the Cultural Revolution as an Instrument

	OLS	IV	
	(1)	(2)	(3)
College & above	0.755***	1.007***	2.580***
	(0.020)	(0.212)	(0.812)
Age	0.070***	0.034***	-0.002
	(0.007)	(0.007)	(0.022)
$Age^2 \times 10^{-3}$	-0.763***	-0.568***	-0.187
	(0.083)	(0.078)	(0.242)
Constant	9.805***	11.029***	11.623***
	(0.146)	(0.128)	(0.382)
Polynomial order of the RD in the 1st stage		1st	2nd
Kleibergen-Paap rk Wald F-Statistic†		54.88	8.165
Observations	107,814	107,814	107,814

Note: Dependent variable is the log of wages. The sample for all these regressions consists of 1950 and 1979 birth cohorts (i.e. ±15 years) in the Household Income and Expenditure Surveys of 2006-2015. The top panels contains the discontinuity in wages in an RDD based on Equation (1). The bottom panel reports the OLS and the 2nd stage of a 2SLS regression that uses discontinuity in Cultural Revolution in an RD as instrument for college & above education to estimate the return of college education on wages. See notes in Table 4 for more information.
*** p<0.01, ** p<0.05, * p<0.10
† This is the Kleibergen-Paap rk Wald F statistic for weak instruments. The 10% critical value of this statistic, calculated by Stock and Yogo (2005), is 16.38.

Appendix A

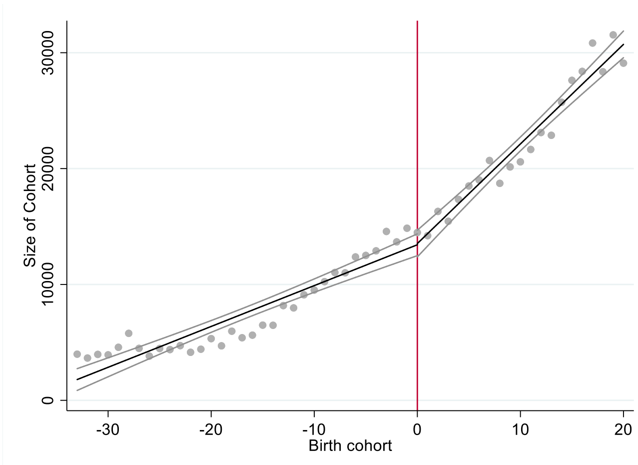
Figure A1 – Size of male cohorts and the Cultural Revolution – Birth year cohorts are between 1950 to 1979



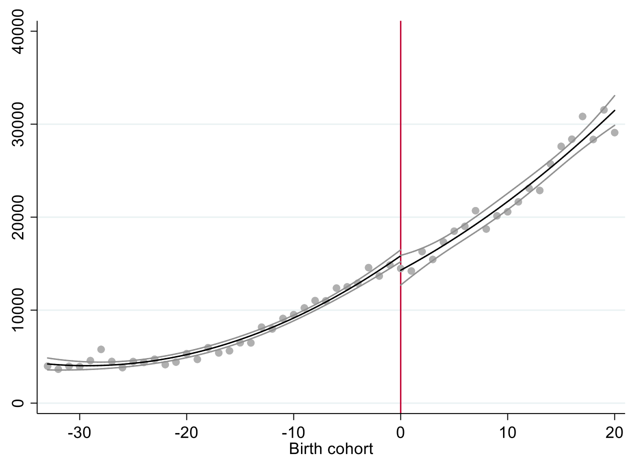
(a) linear – birth year is 1950-79 (b) quadratic – birth year is 1950-79 (c) local polynomial – birth year is 1950-79

Note: The dots represent the proportion in each birth year cohort who partially or fully completed an undergraduate or graduate degree. The curves depict linear, quadratic and local polynomial predictions respectively.

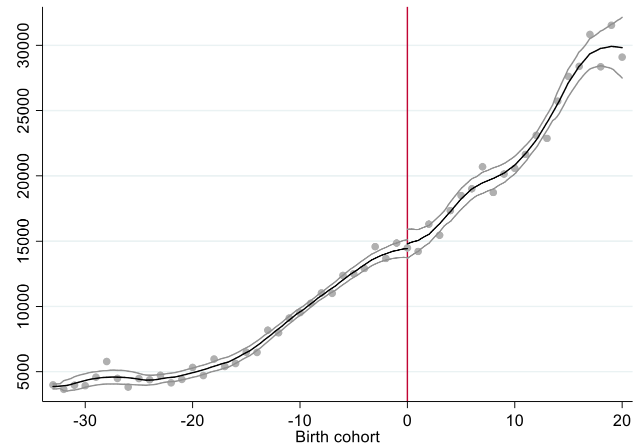
Figure A2 – Size of male cohorts and Cultural Revolution – Birth year cohorts are between 1931 and 1984



(d) linear – birth year is 1931-84



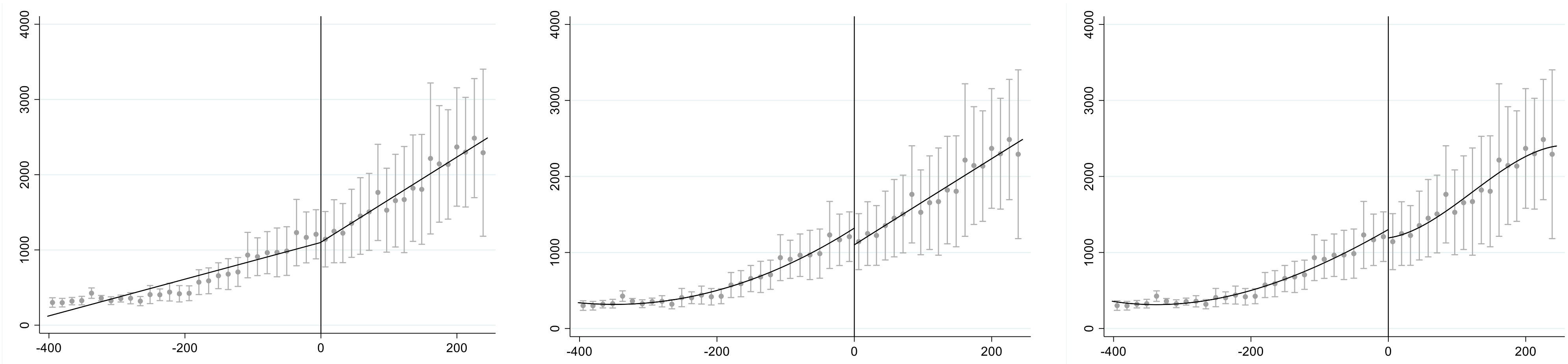
(e) quadratic – birth year is 1931-84



(f) local polynomial – birth year is 1931-84

Note: The dots represent the proportion in each birth year cohort who partially or fully completed an undergraduate or graduate degree. The curves depict linear, quadratic and local polynomial predictions respectively.

Figure A3 – Size of male cohorts (vertical axis) across birth month cohorts (horizontal axis)



(a) linear (b) quadratic (c) polynomial fir of order 3

Note: The dots represent the proportion in each birth year cohort who partially or fully completed an undergraduate or graduate degree. The curves depict linear, quadratic and local polynomial predictions respectively.

Table A1 – Cohort Size and Cultural Revolution
Bandwidth: ±15 years, i.e. Birth cohort is between 1950 and 1979

	Men		Women	
	(1)	(2)	(3)	(4)
Discontinuity in Cohort Size	1,553**	94	851	-975
	(589)	(950)	(651)	(1,121)
Polynomial degree	1st	2nd	1st	2nd
Observations	30	30	30	30
Coefficient of Discontinuity when the bandwidth is:				
±8 [i.e. 1957-1972]	1,016	1,927**	-9	868
	(839)	(715)	(896)	(801)
±10 [i.e. 1955-1974]	698	1,767**	-211	583
	(657)	(765)	(750)	(889)
±12 [i.e. 1953-1976]	861	1,067	100	-154
	(576)	(885)	(648)	(1,046)
±20 [i.e. 1946-1984]	1,908***	1,049	1,773**	-241
	(540)	(805)	(660)	(932)

Note: Dependent variable is the size of each birth year cohort. Robust standard errors are in parentheses. All statistically significant coefficients in the bottom panel (for various bandwidth sizes) are positive (for both men and women).
*** p<0.01, ** p<0.05, * p<0.10

Table A2 – Cohort Size and Cultural Revolution – Bias-Corrected RD
Optimal Bandwidth Choice, Birth Month Cohorts (1931 through 1984)

	Men		Women	
	(1)	(2)	(3)	(4)
Discontinuity in Cohort Size†	257	297	188	224
	(202)	(248)	(183)	(227)
Polynomial degree	1st	2nd	1st	2nd
Observations	648	648	648	648
Coefficient of Discontinuity when the optimal bandwidth choice method is:				
Coverage Error Rate – R.D	264	312	197	242
	(235)	(273)	(217)	(249)
Mean Square Error – Two Sided	237	264	180	198
	(179)	(217)	(162)	(198)
Mean Square Error – R.D.	250	287	165	203
	(207)	(240)	(192)	(220)

Note: Dependent variable is the number of observations for each birth month cohort. The top panel reports the coefficient of D when the optimal bandwidth method used is Mean Square Error – Two Sided (two different optimal bandwidth choices above and below the threshold). The bottom panel reports the same coefficient for other optimal bandwidth choices, for robustness check. Robust-heteroskedastic standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.10
† Optimal bandwidth choice method is Coverage Error Rate – Two Sided (two different optimal bandwidth choices above and below the threshold).

Table A3 – Log of Average Wages for each Job Category

Occupation (Job) Category	Log of Average Wages
1) Legislators, senior officials and managers	12.24590
2) Professionals (doctors, lawyers, engineers, etc.)	12.37025
3) Technicians and associate professionals	11.86571
4) Clerks	11.83416
5) Service workers and shop and market salespeople	11.47818
6) Skilled agricultural and fishery worker	11.25746
7) Crafts and related trades workers	11.51462
8) Plant and machine operators and assembly workers	11.53646
9) Elementary occupation	11.24500

Data Source: Household Income and Expenditure Surveys, 2006-2015; Author's Calculations.

Online Appendix A

We need a more careful consideration of data than a visual observation to see if there has been a statistically significant difference in the educational attainment of the affected and partially affected cohorts relative to others. One way to do this is to look at the deviation from the trend for these cohorts. In other words, one may run the following regression:

$$C_{is} = \alpha_s + \beta_1 T1_{is} + \beta_2 T2_{is} + f(B_{is}) + u_{is} \quad \text{(A1)}$$

in which C_{is} is a dummy equal to one if individual i in region s has partially or fully attended undergraduate or graduate schools. α_s are region fixed effects. There are two regions per province: one representing all the urban areas in a province and another for all the rural areas in that province. Because there were 30 provinces in 2006, the total number of clusters is 60.[1] All results have robust-heteroskedastic standard errors corrected for within region correlations. $T1_{is}$ is a dummy variable equal to 1 if individual i in region s belongs to birth cohorts that were fully affected (i.e. 1962 through 64) and zero otherwise. $T2_{is}$ is a dummy variable equal to 1 if individual i in region s belongs to birth cohorts that were partially affected (i.e. 1959 through 61) and zero otherwise. B_{is} is the year individual i in region s was born (i.e. it is the birth cohort). $f(B_{is})$ is a function that picks up a linear or non-linear trend. Based on the shape of the graphs in Figures 1(a) and 1(b), we use 1) a linear trend (only B_{is}), 2) quadratic trends (B_{is} and B_{is} squared), and 3) differential linear trends before and after the Cultural Revolution. β_1 and β_2 show deviations from the trend for cohorts who were fully or partially affected by the Cultural Revolution. Table B1 reports the results.

Column (1) in Table B1 depicts the regression in Equation (A1) using the sample of males in birth cohorts of 1931 through 1984 and a linear trend. The coefficient of T1 shows that cohorts of men who were fully

[1] There were 31 provinces in 2011. So, the provinces in 2011 were matched with those in 2006 to create a consistent regional data over time.

affected by the Cultural Revolution had about 1.2 percentage points lower rate of college (and above) attainment than the trend. Coefficient of T2 depicts that for those who were partially affected, the attainment rate was the same as the linear trend.

Considering a linear trend for college attainment rate may be strong (see Figures 1(a) and 1(b)). One can consider a quadratic trend. But as Column (2) of Table B1 shows, adding the trend squared, i.e. B_{is}^2, has little impact on the results for the sample of men.

One may argue that after the Cultural Revolution, the transformation of institutions of higher education might have affected the trend altogether. This is easy to control by adding AR_{is} and its interaction with the trend variable, B_{is}, to Equation (A1). AR_{is} is a dummy equal to one if individual i in region s was born after 1961 and zero otherwise. The interaction of this variable and the trend picks up the change in trend after the Cultural Revolution. The results for men are reported in Column (3). Coefficient of T1 shows a -1.5 percentage points deviation from the trend before the Cultural Revolution. But the deviation for the partially affected cohorts remains insignificant and small.

Column (4) represents the same regression as Column (1) but for women. The coefficients of T1 and T2 imply that attainment rates for cohorts fully or partially affected by the Cultural Revolution were 4 percentage points below the trend. But, as shown in Figure 1(b), the trend for women is more like a quadratic one. Hence, squared of birth year should be added to the specification. Column (5) reports the results for such a specification. The deviations found in Column (4) become quite smaller (from about four to less than one percentage points).

Another specification is to control for separate trends before and after the Cultural Revolution. Column (6) depicts the results. The coefficients imply that attainment rate for female cohorts who were fully affected by the Cultural Revolution was 2.2 percentage points *larger* than the trend. This is an odd result and clearly shows that the deviations from the trend for women are sensitive to the specification.

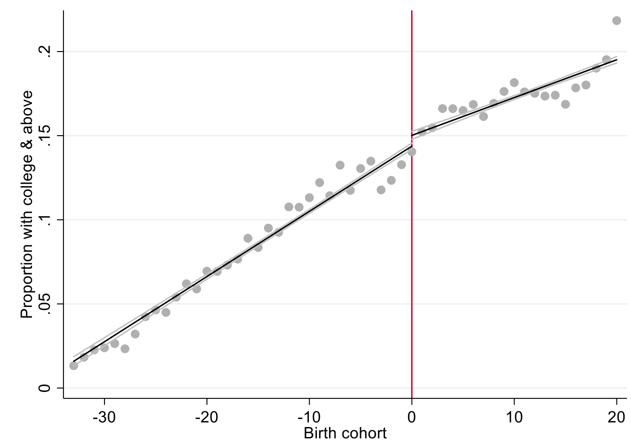
Table OA1 – Deviations from the Trend for College Attendance Because of Cultural Revolution

	Men			Women		
	(1)	(2)	(3)	(4)	(5)	(6)
T1	-0.010***	-0.012***	-0.018***	-0.044***	-0.008***	0.024***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
T2	-0.004**	-0.007***	-0.004*	-0.039***	-0.003*	0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Birth year	0.003***	0.047***	0.003***	0.005***	-0.590***	0.001***
	(0.000)	(0.009)	(0.000)	(0.000)	(0.008)	(0.000)
Birth year^2 / 1000		-0.011***			0.152***	
		(0.002)			(0.002)	
AR			1.471***			-17.465***
			(0.267)			(0.234)
Birth year × AR			-0.001***			0.009***
			(0.000)			(0.000)
Region FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	708,919	708,919	708,919	710,672	710,672	710,672
Number of Regions	60	60	60	60	60	60

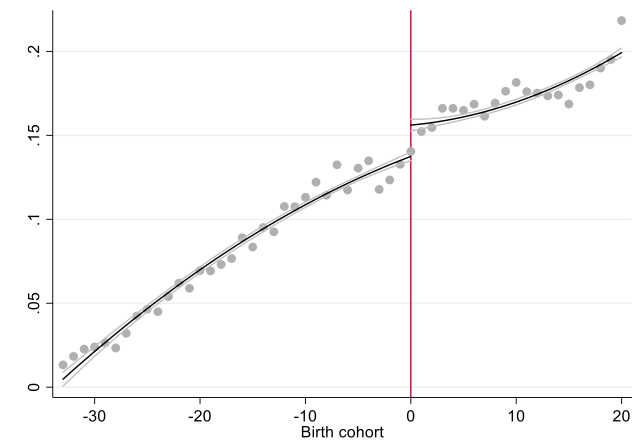
Note: Dependent variable is a dummy equal to one if the individual partially or fully completed college and above levels, and zero otherwise. Robust-heteroskedastic standard errors corrected for correlation within regions in parentheses. A Region is all urban or rural areas within a province. The sample consists of birth cohorts between 1931 and 1981.
*** p<0.01, ** p<0.05, * p<0.10

Online Appendix B
Figures

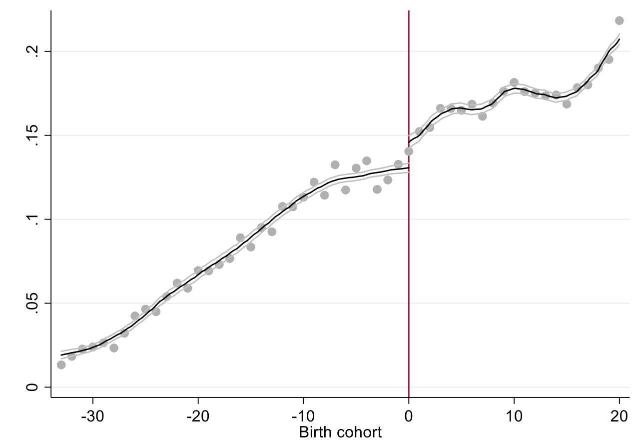
Figure OB1 – Discontinuity in the proportion of men and women with college and above education across birth year cohorts (1931 through 1984 birth cohorts)



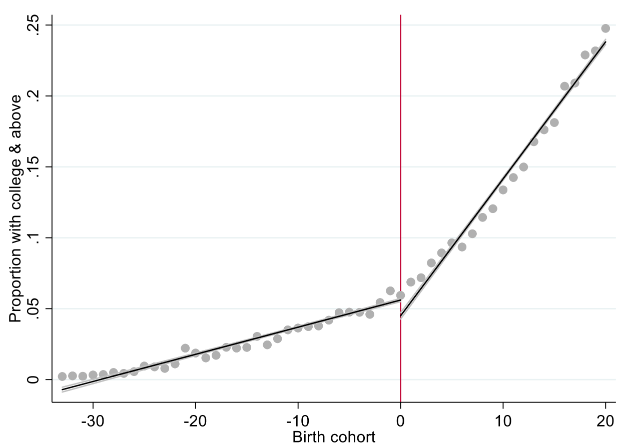
(a) linear - men



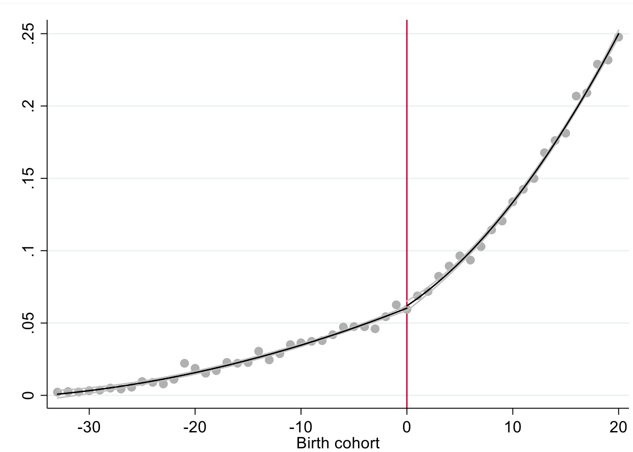
(b) quadratic - men



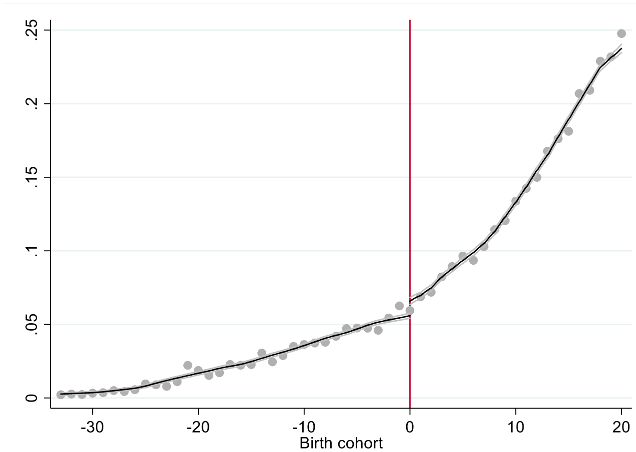
(c) local polynomial smooth line - men



(d) linear - women



(e) quadratic - women



(f) local polynomial smooth line - women

Note: The dots represent the proportion in each birth year cohort who partially or fully completed an undergraduate or graduate degree. The curves depict linear, quadratic and local polynomial predictions respectively.

Tables

Table OB1 – College Attendance and Cultural Revolution – No Clustering of Standard Errors,
Bandwidth in the Top Panel: ±15 years
(i.e. the birth cohorts are between 1950 and 1979)

	Men		Women	
	(1)	(2)	(3)	(4)
Discontinuity	-0.022***	-0.015***	0.007***	-0.004*
	(0.002)	(0.003)	(0.002)	(0.003)
Polynomial degree	1st	2nd	1st	2nd
Observations	470,850	470,850	467,912	467,912
Coefficient of D when the bandwidth is:				
±8 [i.e. 1957-1972]	-0.022***	-0.008*	-0.002	-0.001
	(0.003)	(0.005)	(0.002)	(0.003)
±10 [i.e. 1955-1974]	-0.018***	-0.018***	0.000	-0.005
	(0.003)	(0.004)	(0.002)	(0.003)
±12 [i.e. 1953-1976]	-0.019***	-0.017***	0.002	-0.005
	(0.002)	(0.004)	(0.002)	(0.003)
±20 [i.e. 1946-1984]	-0.012***	-0.031***	0.016***	-0.003
	(0.002)	(0.003)	(0.002)	(0.002)

Note: Dependent variable is a dummy equal to one if the individual partially or fully completed college and above levels, and zero otherwise. Regular standard errors without any clustering are reported. No covariates other than those in Equation (1) is controlled for.
*** p<0.01, ** p<0.05, * p<0.10

Table OB2 – College Attendance and Cultural Revolution – Clustering of Standard Errors by 60 Geographic Regions,
Bandwidth in the Top Panel: ±15 years
(i.e. the birth cohorts are between 1950 and 1979)

	Men		Women	
	(1)	(2)	(3)	(4)
Discontinuity	-0.022***	-0.015***	0.007***	-0.004*
	(0.003)	(0.004)	(0.002)	(0.002)
Polynomial degree	1st	2nd	1st	2nd
Observations	470,850	470,850	467,912	467,912
Coefficient of D when the bandwidth is:				
±8 [i.e. 1957-1972]	-0.022***	-0.008**	-0.002	-0.001
	(0.003)	(0.004)	(0.002)	(0.003)
±10 [i.e. 1955-1974]	-0.018***	-0.018***	0.000	-0.005
	(0.003)	(0.004)	(0.002)	(0.003)
±12 [i.e. 1953-1976]	-0.019***	-0.017***	0.002	-0.005
	(0.003)	(0.004)	(0.002)	(0.003)
±20 [i.e. 1946-1984]	-0.012***	-0.031***	0.016***	-0.003
	(0.003)	(0.004)	(0.003)	(0.003)

Note: Dependent variable is a dummy equal to one if the individual partially or fully completed college and above levels, and zero otherwise. Standard errors are corrected for correlation within 60 geographic clusters. No covariates other than those in Equation (1) is controlled for. *** p<0.01, ** p<0.05, * p<0.10

Table OB3 – College Attendance and Cultural Revolution – Conventional RD Estimates
Optimal Bandwidth Choice, Birth Month Cohorts

	Men		Women	
	(1)	(2)	(3)	(4)
Discontinuity†	-0.018***	-0.022***	-0.003	-0.006
	(0.006)	(0.008)	(0.004)	(0.005)
Polynomial degree	1st	2nd	1st	2nd
Observations	686,188	686,188	686,677	686,677
Discontinuity when the optimal bandwidth choice method is:				
Coverage Error Rate – RD	-0.019***	-0.023***	-0.007	-0.010*
	(0.007)	(0.009)	(0.005)	(0.006)
Mean Square Error – Two Sided	-0.017***	-0.011**	-0.001	0.001
	(0.004)	(0.005)	(0.003)	(0.003)
Mean Square Error – R.D.	-0.012**	-0.010*	0.001	0.002
	(0.005)	(0.005)	(0.003)	(0.004)

Note: This table reports RDD estimates of discontinuity in a dummy which is equal to one if the individual went to college or above and zero otherwise. Birth month is the running variable. The top panel uses Coverage Error Rate – Two Sided (two different optimal bandwidths above and below the threshold) as the optimal bandwidth method. The bottom panel reports the same coefficient for other optimal bandwidth choice methods, for robustness check. All estimates reported in this table are Conventional RD estimates. For Bias-Corrected and Robust RD estimates, see Tables 3 and OB2.
*** $p<0.01$, ** $p<0.05$, * $p<0.10$
† Optimal bandwidth choice method is Coverage Error Rate – Two Sided (two different optimal bandwidth choices above and below the threshold).

Table OB4 – College Attendance and Cultural Revolution – Bias-Corrected RD Estimates with Robust Standard Errors
Optimal Bandwidth Choice, Birth Month Cohort

	Men		Women	
	(1)	(2)	(3)	(4)
Discontinuity†	-0.018***	-0.022***	-0.003	-0.006
	(0.006)	(0.008)	(0.004)	(0.005)
Polynomial degree	1st	2nd	1st	2nd
Observations	686,188	686,188	686,677	686,677
Discontinuity when the optimal bandwidth choice method is:				
Coverage Error Rate – RD	-0.019***	-0.022***	-0.007	-0.010*
	(0.007)	(0.009)	(0.005)	(0.006)
Mean Square Error – Two Sided	-0.016***	-0.010*	-0.001	0.002
	(0.004)	(0.005)	(0.003)	(0.003)
Mean Square Error – R.D.	-0.010**	-0.009	0.001	0.003
	(0.005)	(0.006)	(0.003)	(0.004)

Note: This table reports RDD estimates of discontinuity in a dummy which is equal to one if the individual went to college or above and zero otherwise. Birth month is the running variable. The top panel uses Coverage Error Rate – Two Sided (two different optimal bandwidths above and below the threshold) as the optimal bandwidth method. The bottom panel reports the same coefficient for other optimal bandwidth choice methods, for robustness check. All estimates reported in this table are Robust RD estimates. For Bias-Corrected and Conventional RD estimates, see Tables 3 and OB1.
*** p<0.01, ** p<0.05, * p<0.10
† Optimal bandwidth choice method is Coverage Error Rate – Two Sided (two different optimal bandwidth choices above and below the threshold).

Table OB5 – Average College Attendance by Birth Month and Cultural Revolution – Bias-Corrected RD Estimates
Optimal Bandwidth Choice, Birth Month Cohort

	Men		Women	
	(1)	(2)	(3)	(4)
Discontinuity†	-0.019***	-0.022***	-0.001	-0.002
	(0.005)	(0.006)	(0.005)	(0.007)
Polynomial degree	1st	2nd	1st	2nd
Observations	648	648	648	648
Discontinuity when the optimal bandwidth choice method is:				
Coverage Error Rate – RD	-0.015***	-0.017**	-0.001	-0.002
	(0.005)	(0.007)	(0.007)	(0.008)
Mean Square Error – Two Sided	-0.020***	-0.020***	-0.001	-0.000
	(0.004)	(0.005)	(0.004)	(0.005)
Mean Square Error – R.D.	-0.016***	-0.014**	0.002	0.004
	(0.005)	(0.006)	(0.006)	(0.007)

Note: This table reports RDD estimates of discontinuity in the share of individuals in each birth month cohort who went to college or above. Birth month is the running variable. The top panel uses Coverage Error Rate – Two Sided (two different optimal bandwidths above and below the threshold) as the optimal bandwidth method. The bottom panel reports the same coefficient for other optimal bandwidth choice methods, for robustness check. All estimates reported in this table are Bias-Corrected RD estimates. For Conventional and Robust RD estimates, see Tables OB4 and OB5.
*** p<0.01, ** p<0.05, * p<0.10
† Optimal bandwidth choice method is Coverage Error Rate – Two Sided (two different optimal bandwidth choices above and below the threshold).

Table OB6 – Average College Attendance by Birth Month and Cultural Revolution – Conventional RD Estimates
Optimal Bandwidth Choice, Birth Month Cohort

	Men		Women	
	(1)	(2)	(3)	(4)
Discontinuity[†]	-0.019***	-0.022***	-0.002	-0.002
	(0.005)	(0.006)	(0.005)	(0.007)
Polynomial degree	1st	2nd	1st	2nd
Observations	648	648	648	648
Discontinuity when the optimal bandwidth choice method is:				
Coverage Error Rate – RD	-0.016***	-0.017***	-0.000	-0.002
	(0.005)	(0.007)	(0.007)	(0.008)
Mean Square Error – Two Sided	-0.020***	-0.020***	-0.002	-0.000
	(0.004)	(0.005)	(0.004)	(0.005)
Mean Square Error – R.D.	-0.017***	-0.016***	0.001	0.002
	(0.005)	(0.006)	(0.006)	(0.007)

Note: This table reports RDD estimates of discontinuity in the share of individuals in each birth month cohort who went to college or above. Birth month is the running variable. The top panel uses Coverage Error Rate – Two Sided (two different optimal bandwidths above and below the threshold) as the optimal bandwidth method. The bottom panel reports the same coefficient for other optimal bandwidth choice methods, for robustness check. All estimates reported in this table are Conventional RD estimates. For Bias-Corrected and Robust RD estimates, see Tables OB3 and OB5.

*** p<0.01, ** p<0.05, * p<0.10

† Optimal bandwidth choice method is Coverage Error Rate – Two Sided (two different optimal bandwidth choices above and below the threshold).

Table OB7 – Average College Attendance by Birth Month and Cultural Revolution – Robust RD Estimates
Optimal Bandwidth Choice, Birth Month Cohort

	Men		Women	
	(1)	(2)	(3)	(4)
Discontinuity†	-0.019***	-0.022***	-0.001	-0.002
	(0.005)	(0.006)	(0.005)	(0.007)
Polynomial degree	1st	2nd	1st	2nd
Observations	648	648	648	648
Discontinuity when the optimal bandwidth choice method is:				
Coverage Error Rate – RD	-0.015***	-0.017**	-0.001	-0.002
	(0.005)	(0.007)	(0.007)	(0.008)
Mean Square Error – Two Sided	-0.020***	-0.020***	-0.001	-0.000
	(0.004)	(0.005)	(0.004)	(0.005)
Mean Square Error – R.D.	-0.016***	-0.014**	0.002	0.004
	(0.005)	(0.006)	(0.007)	(0.007)

Note: This table reports RDD estimates of discontinuity in the share of individuals in each birth month cohort who went to college or above. Birth month is the running variable. The top panel uses Coverage Error Rate – Two Sided (two different optimal bandwidths above and below the threshold) as the optimal bandwidth method. The bottom panel reports the same coefficient for other optimal bandwidth choice methods, for robustness check. All estimates reported in this table are Robust RD estimates. For Bias-Corrected and Conventional RD estimates, see Tables OB3 and OB4.
*** p<0.01, ** p<0.05, * p<0.10
† Optimal bandwidth choice method is Coverage Error Rate – Two Sided (two different optimal bandwidth choices above and below the threshold).

Table OB8 – Cohort Size and Cultural Revolution – Conventional RD Estimates
Optimal Bandwidth Choice, Birth Month Cohorts (1931 through 1984)

	Men		Women	
	(1)	(2)	(3)	(4)
Discontinuity in Cohort Size†	244	292	172	217
	(202)	(248)	(183)	(227)
Polynomial degree	1st	2nd	1st	2nd
Observations	648	648	648	648
Coefficient of Discontinuity when the optimal bandwidth choice method is:				
Coverage Error Rate – R.D	245	302	178	231
	(235)	(273)	(217)	(249)
Mean Square Error – Two Sided	214	249	149	177
	(179)	(217)	(162)	(198)
Mean Square Error – R.D.	215	256	130	170
	(207)	(240)	(192)	(220)

Note: The dependent variable is the number of observations for each birth month cohort. The top panel reports the coefficient of Discontinuity in an RDD when the optimal bandwidth method used is Coverage Error Rate – Two Sided (two different optimal bandwidth choices above and below the threshold). The bottom panel reports the same coefficient for other optimal bandwidth choices, for robustness check. All estimates reported in this table are Conventional RD estimates. For Bias-Corrected and Robust RD estimates, see Tables A2 and OB7.
*** p<0.01, ** p<0.05, * p<0.10
† Optimal bandwidth choice method is Coverage Error Rate – Two Sided (two different optimal bandwidth choices above and below the threshold).

Table OB9 – Cohort Size and Cultural Revolution – Robust RD Estimates
Optimal Bandwidth Choice, Birth Month Cohorts (1931 through 1984)

	Men		Women	
	(1)	(2)	(3)	(4)
Discontinuity in Cohort Size†	257	297	188	224
	(202)	(248)	(183)	(227)
Polynomial degree	1st	2nd	1st	2nd
Observations	648	648	648	648
Coefficient of Discontinuity when the optimal bandwidth choice method is:				
Coverage Error Rate – R.D	264	312	197	242
	(235)	(273)	(217)	(249)
Mean Square Error – Two Sided	237	264	180	198
	(179)	(217)	(162)	(198)
Mean Square Error – R.D.	250	287	165	203
	(207)	(240)	(192)	(220)

Note: The dependent variable is the number of observations for each birth month cohort. The top panel reports the coefficient of Discontinuity in an RDD when the optimal bandwidth method used is Coverage Error Rate – Two Sided (two different optimal bandwidth choices above and below the threshold). The bottom panel reports the same coefficient for other optimal bandwidth choices, for robustness check. All estimates reported in this table are Robust RD estimates. For Bias-Corrected and Conventional RD estimates, see Tables A2 and OB6.
*** $p<0.01$, ** $p<0.05$, * $p<0.10$
† Optimal bandwidth choice method is Coverage Error Rate – Two Sided (two different optimal bandwidth choices above and below the threshold).

Table OB10 – Men's Job Status (Subjective Measure), College Attendance, and Cultural Revolution
(Census data; using birth year and birth month)

	Birth Year†		Birth Months‡	
	(1)	(2)	(3)	(4)
Panel A – RDD Estimate of the Impact of Cultural Revolution on Job Status				
Discontinuity in Job Status	-0.054***	-0.028**	-0.045***	-0.046***
	(0.013)	(0.011)	(0.014)	(0.017)
Polynomial degree	1st	2nd	1st	2nd
Observations	370,805	370,805	495,928	495,928
Panel B – 2SLS Estimate of the Impact of College Attendance on Job Status (using Discontinuity in College Attendance, due to Cultural Rev., as IV)				
College Attendance	1.566***	1.803***	1.958***	2.386***
	(0.125)	(0.431)	(0.441)	(0.726)
Kleibergen-Paap rk Wald F-Statistic[Π]	223.7	18.50	18.05	7.602
Polynomial degree	1st	2nd	1st	2nd
Observations	370,805	370,805	495,928	495,928

Note: See notes in Table 4. The only difference between this table and Table 4 is that the job status variable is a categorical variable defined as follows: plant and machinery workers and elementary occupations are combined into one group and the lowest value of job status, i.e. 1 is assigned to them. Service workers, salespeople, skilled agricultural workers, and crafts and trades workers are aggregated into another group and value 2 is assigned to their job status. Technicians and clerks are combined into a third group with the value of 3 for job status, and government and private sector managers, and doctors, lawyers, engineers and other professionals are aggregated into a fourth group and value 4 is assigned to their job status.

*** p<0.01, ** p<0.05, * p<0.10

† Bandwidth is ±15 years, i.e. those born between 1950 and 1979.

‡ Optimal bandwidth choice method is Coverage Error Rate – Two Sided (two different optimal bandwidth choices above and below the threshold).

Π This is the Kleibergen-Paap rk Wald F statistic for weak instruments. The 10% critical value of this statistic calculated by Stock and Yogo (2005) is 16.38.

Table OB11 – Distribution of Educational Attainment Around the Threshold (Birth Month Data – 2006 and 2011 Censuses)

Variable	Observations	In %
Illiterate	797	25.9
Primary	899	29.2
Middle School	531	17.2
High School	519	16.9
College & above	335	10.9
Average	Middle School	

Table OB12 – Distribution of Educational Attainment Around the Threshold (Birth Year Data – HIES 2006-2015)

Variable	Observations	In %
Illiterate	1,139	15.8
Primary	2,535	35.1
Middle School	1,137	15.7
High School	1,043	14.4
College & above	1,371	19.0
Average	Middle School	

Table OB13 – Men's Job Status (Average Wages), College Attendance, and Cultural Revolution – Including Armed Forces
(Census data; using birth year and birth month)

	Birth Year†		Birth Months‡	
	(1)	(2)	(3)	(4)
Panel A – RDD Estimate of the Impact of Cultural Revolution on Job Status				
Discontinuity in Job Status	-0.021***	-0.012**	-0.020***	-0.023***
	(0.006)	(0.005)	(0.006)	(0.007)
Polynomial degree	1st	2nd	1st	2nd
Observations	378,122	378,122	506,087	506,087
Panel B – Second Stage of the 2SLS Estimate of the Impact of College Attendance on Job Status (using Discontinuity in College Attendance, due to Cultural Rev., as IV)				
College Attendance	0.653***	0.864***	0.968***	1.267***
	(0.053)	(0.187)	(0.199)	(0.375)
Kleibergen-Paap rk Wald F-Statistic Π	188	15.77	15.32	6.316
Polynomial order of RD in the 1st stage	1st	2nd	1st	2nd
Observations	378,122	378,122	506,087	506,087

Note: This table is the same as Table 4. The only difference is that Armed Forces are also included among the occupations and average wages of armed forces is part of the data.
*** p<0.01, ** p<0.05, * p<0.10
† Bandwidth is ±15 years, i.e. those born between 1950 and 1979.
‡ Optimal bandwidth choice method is Coverage Error Rate – Two Sided (two different optimal bandwidth choices above and below the threshold). Bias-Corrected coefficients and standard errors are reported.
Π This is the Kleibergen-Paap rk Wald F statistic for weak instruments. The 10% critical value of this statistic, calculated by Stock and Yogo (2005), is 16.38.

Table OB14 – Rural Men's Log of Wages and the Cultural Revolution
(Bandwidth: ±15 Birth Years; 1950 through 1979)

Panel A – Discontinuity in Men's Log of Wages Due to Cultural Revolution

	(1)	(2)
Discontinuity in log of Wages	-0.050***	-0.018**
	(0.010)	(0.008)
Polynomial degree	1st	2nd
Observations	48,182	48,182

Panel B – Returns to College Education for Men Using the Discontinuity from the Cultural Revolution as an Instrument

	OLS	IV	
	(1)	(2)	(3)
College & above	0.837***	2.100***	-4.484
	(0.037)	(0.762)	(9.048)
Age	0.049***	0.024**	0.088
	(0.005)	(0.010)	(0.083)
$Age^2 \times 10^{-3}$	-0.549***	-0.369***	-1.162
	(0.053)	(0.125)	(1.034)
Constant	9.805***	10.898***	10.086***
	(0.146)	(0.170)	(1.054)
Polynomial order of the RD in the 1^{st} Stage		1st	2nd
Kleibergen-Paap rk Wald F-Statistic†		11.68	0.423
Observations	48,182	48,182	48,182

Note: This table reports the same estimates as Table 5 for the subsample of rural areas. See notes for Table 5 for more information.
*** p<0.01, ** p<0.05, * p<0.10
† This is the Kleibergen-Paap rk Wald F statistic for weak instruments. The 10% critical value of this statistic, calculated by Stock and Yogo (2005), is 16.38.

Table OB15 – Urban Men's Log of Wages and the Cultural Revolution
(Bandwidth: ±15 Birth Years; 1950 through 1979)

Panel A – Discontinuity in Men's Log of Wages Due to Cultural Revolution

	(1)	(2)
Discontinuity in log of Wages	-0.064***	-0.055**
	(0.014)	(0.026)
Polynomial degree	1st	2nd
Observations	59,632	59,632

Panel B – Returns to College Education for Men Using the Discontinuity from the Cultural Revolution as an Instrument

	OLS	IV	
	(1)	(2)	(3)
College & above	0.658***	0.883***	2.106***
	(0.014)	(0.200)	(0.703)
Age	0.085***	0.035***	-0.006
	(0.009)	(0.010)	(0.029)
$Age^2 \times 10^{-3}$	-0.924***	-0.631***	-0.204
	(0.108)	(0.107)	(0.306)
Constant	9.580***	11.225***	11.952***
	(0.190)	(0.186)	(0.520)
Polynomial order of the RD in the 1^{st} Stage		1st	2nd
Kleibergen-Paap rk Wald F-Statistic†		45.14	7.374
Observations	59,632	59,632	59,632

Note: This table reports the same estimates as Table 5 for the subsample of urban areas. See notes for Table 5 for more information.
*** p<0.01, ** p<0.05, * p<0.10
† This is the Kleibergen-Paap rk Wald F statistic for weak instruments. The 10% critical value of this statistic, calculated by Stock and Yogo (2005), is 16.38.