A look at the mechanics of compulsory primary education expansion

Alex Eble and Feng Hu*

August 2013

Abstract

Changing the length of compulsory schooling is one of the fundamental levers a government has to influence the educational attainment of its citizens. In this paper, we use exogenous variation from a policy experiment in China which extended the length of compulsory primary schooling from 5 to 6 years to understand how families respond to such policies and how these responses affect the number of years of education children attain. A rich, new nationally representative data-set from China allows us to identify county-specific policy implementation years despite having anonymized county data. We use a regression discontinuity design to estimate the impact of this policy on families' education decisions for their children. We find three major results: one, that individual take up of the policy was almost universal, two, that the policy did not induce most families to adjust on the lower margin, e.g. we do not see "displacement" of the additional year of primary school by fewer years of post-primary schooling, and three, we find suggestive but non-significant evidence that the extra primary schooling pushed some individuals who are either less wealthy or are less likely to benefit from increased education to get less post-primary education.

*Eble: Department of Economics, Brown University. 64 Waterman Street, Providence, RI 02912. Email: alexander_eble@brown.edu; Hu: Dongling School of Economics and Management, University of Science and Technology Beijing, 30 Xueyuan Road, Haidian District, Beijing, China 100083. Email: feng3hu@gmail.com. The authors would like to thank Ken Chay, Andrew Foster, Rob Jensen, Gareth Olds and participants at the Brown University Applied Micro Lunch for helpful suggestions. Eble gratefully acknowledges the support of the United States National Science Foundation. All remaining errors are our own. JEL1: O15, JEL2: I25

1 Introduction

Previous empirical studies of compulsory schooling have focused on understanding the effects of policies which change the age at which children are allowed to leave school. Angrist and Krueger (1991) showed that school leaving laws which allowed some US students to leave school at an earlier age than others did indeed result in many affected students leaving school at this earlier permitted age. They then used this exogenous variation in education levels to back out estimates of the returns to education in terms of earnings later in life. Since then, dozens of papers have used this and other similar policies to examine the effect of education on a variety of outcomes. Acemoglu and Angrist (2001), Pischke and von Wachter (2008), and Devereux and Hart (2010), among many others, have used US and European data to test Angrist and Krueger's results and more precisely estimate the relationship between education and wages. Black, Devereux, and Salvanes (2005) use a policy which extended the length of compulsory education in Norway from 7 to 9 years to understand inter-generational transmission of education. Brinch and Galloway (2012) use the same change in Norwegian compulsory education laws to estimate the impact of education on cognitive ability. Brunello and Fort (2013) use data from nine European countries with similar compulsory education extension policies to estimate the impact of additional education on a woman's body mass index.

This is only a small sampling of a much larger literature which takes a similar tack, using changes in compulsory education in the US and Europe which affect most children in their final years of schooling to understand a plethora of later life outcomes. These policy experiments lend themselves well to answering questions about the effects of education on subsequent outcomes. However, because they affect students as they are preparing to leave school, they are less well suited to use in understanding behavioral responses to compulsory education changes which influence educational attainment itself. This has left an important lacuna, as the education of a country's citizens is central to its long-term economic performance and compulsory education is a fundamental force with which to shape educational attainment. (Hanushek and Kimko, 2000)

There is a rich set of economic studies which do attempt to understand patterns of educational attainment and the decisions behind these patterns, dating back to the seminal contributions of Mincer and Becker. These authors linked decisions about educational attainment to future labor market performance, postulating a simple economic model where an internal calculus comparing costs of education (tuition and forgone earnings) and benefits (higher subsequent productivity and wages) determined attainment levels[1]. (Mincer, 1974; Becker, 1975) Their model was later enriched by economists who proposed that in addition to financial costs, there are also non-pecuniary costs (e.g. disutility of being in school) and benefits (potentially elevated social status by virtue of one's level of education) that should be considered when modeling the internal calculus used to make the decision of when to stop schooling[2]. (Akerlof and Kranton, 2002; Attanasio and Kaufmann, 2009)

This study hopes to unite these two literatures. The paper is, at its core, about understanding the decisions that families, particularly in the developing world, make about their children's education in the face of changes in compulsory education policy. Specifically, we are interested in knowing how Chinese parents and children react when they are "dosed" with a policy that extends the length of primary school from five to six years. This slows the child's progress through the educational system and makes completing any given qualification (e.g. primary school, middle school, high school) require an additional year of study, an increase which in turn may alter the internal calculus used to decide how much schooling a child will attain. In our data, which covers children graduating from primary school in China between 1980 and 2007, the median number of years of schooling is 9. In principle, affected children could adjust on either margin – either getting more or less school - within a distribution of total years of schooling attained which is similar before and after the policy.

We find that, on average, "treated" children get exactly one more year of primary school than their untreated counterparts, but have no detectable difference on a battery of measures of post-primary education, spanning graduation rates at three levels of schooling, indicators for having ever attended high school and post-secondary school, and number of years spent in post-primary education. This result suggests that the extra

[1]Spence (1973) used a similar framework but took a more cynical view on the content of education, benefits being restricted to signaling high ability and gaining higher wages, and costs also including disutility from being in school, inversely related to ability.

[2]In more recent studies, credit constraints (Lochner and Monge-Naranjo, 2012), information on returns to schooling (Jensen, 2010), and school quality (Hanushek, Lavy, and Hitomi, 2006) have also been shown to impact on these decisions. These findings have been used in other recent work (Oreopoulos, 2007) to help reconcile the empirical "paradox" of students who drop out early in the face of high estimated returns to staying on in school. (Psacharopoulos, 1985)

year of primary school and the cost of forgone wages it entailed were not enough to disincentivize students to adjust downward on the middle or high school margin. On the other hand, we might expect if the effect of the additional year of primary school was to provide students with remedial education which would help the less able prepare to pass entrance examinations they might otherwise have not, it might make more of them able to ascend to higher levels of education. We find no evidence of any movement on this margin either.

This paper also contributes to an ongoing debate about effective policy instruments for increasing education levels in the developing world. A recent literature review synthesizes a large body of research evaluating the impact of different educational interventions on school attainment and ability. (Kremer and Holla, 2009) Interventions considered include provision of school materials, after-school para-teachers, teacher monitoring systems, and deworming medicine, among several others. To the best of our knowledge, however, ours is the first study which shows the impact of a policy which extends the length of primary school on years of education attained. The benefits of such an intervention are obvious - the infrastructure currently exists, the technology for delivery is known, and implementation would require very little additional bureaucracy. On the other hand, it could have negative consequences if the extra year of primary schooling led to students being less likely to continue on to further levels of education, i.e., the policy might "displace" post-primary education pupils would have otherwise attained. We provide evidence that the displacement concern does not seem to be an issue in our representative sample of China during the period of study. Whether the policy increases skills or labor force productivity, and whether this justifies the cost of extra teachers, schooling facilities and the opportunity cost for children of this additional year of education, is treated in a companion paper. (Hu and Eble, 2013)

The rest of the paper proceeds as follows. In section 2, we discuss the history of education in China and the nature of the policy experiment. In section 3 we build a model to understand how parents might respond to the policy and derive a few testable predictions. In section 4 we describe the data and how it allows us to identify county-specific policy implementation years which we then use as our instrument. Section 5 states our identification assumption, specifies how our empirical analyses are conducted and then shows results. Section 6 concludes.

2 History of primary and secondary education in China

This section reviews the changes in the length of primary, middle, and high school in the People's Republic of China from when the country was founded in 1949 to the present. Because China's educational system was closely linked to political priorities and seriously interrupted by the Cultural Revolution (Hannum, 1999; Meng and Gregory, 2002), this review is organized around the following three periods: before, during and after the Cultural Revolution.

1. Before the Cultural Revolution (1949-1965)

At the founding of the People's Republic of China, education levels were quite low – only 20% of the population was literate, and less than 40% of school-aged children were in school. (Hannum, 1999) To address these issues, the State Council issued a series of reforms, starting with the *Decision about Reforming the Educational System*, in which the new government funneled a massive amount of resources into the primary education system, shortened the length of primary school from six years, which it had been traditionally, to five, and vastly expanded the number of schools across the country at all levels. (Liu, 1993) To ease demand on schools, this reform also delayed the age at which children could start their primary education from six to seven. The number of students in the system grew rapidly, from 24.4 million in 1949 to 93.8 million in 1960.

This was the first of many reforms in which the structure of the system fluctuated. Due to the lack of textbooks for the new system and an insufficient number of teachers, this experiment was stopped at the end of 1953. The length of primary schooling was restored to six years by the *Instructions about Rectifying and Improving Primary Education* issued by the State Council in November 1953. It remained this way until the Great Leap Forward in 1958.

The Great Leap Forward was a massive reorganization of agricultural and industrial production which coincided with a series of severe adverse weather shocks. This put great strain on the resources of the entire country, causing widespread famine in rural areas. This strain was felt in the education system as well. In April 1960, at the second meeting of the second National People's Congress, Vice Premier Lu Dingyi called for a shortening of the years of primary and secondary education, a reduction in study hours, and an increase in labor time to help with the push to industrialize. This led to a new series of rapid changes mandated from on high which again reduced the number of years of primary education to five years. Due to the drastic nature of the reforms, the strain experienced by all as a result of the Great Leap Forward, and the difficulty of implementing such reforms rapidly in a country as large as China, only around 15% of primary schools nationally participated in this experiment at its zenith in 1961. (National Institute, 1984) Less than two years after its inception, the policy was reversed and, as early as 1962, less than one percent of schools were still "experimenting" with this round of mandated five year primary education. (Liu, 1993; Liao, 2004)

2. During the Cultural Revolution (1966-1976)

The ten years of Cultural Revolution constituted an even more serious interruption to China's educational system. At the beginning of the Cultural Revolution, all primary schools in China's urban areas were closed for two or three years. No teaching was carried out in these areas, and no new students were enrolled. Instead, urban students were required to participate in farming or to work in factories. Rural students remained in schools to some extent, though their experience too was frequently interrupted by frenetic policy changes. This situation continued until about 1972, when the normal school curriculum was gradually resumed. (Meng and Gregory, 2002) During this period, the educational system was standardized and shortened so that all students would study similar curricula and have more time to spend on labor. (Hannum, Behrman, Wang, and Liu, 2008) In May 1966, the Central Committee of the Communist Party of China (CC-CPC) forwarded to local governments a letter from Mao Zedong, which stated that "the years of schooling should be shortened and the educational system should be revolutionized."[3] Afterward, all provinces reformed their educational systems: some had a total of nine years of primary and secondary schooling (including five years of primary education) while others implemented systems with ten years of primary and secondary schooling, where the length of primary education could be either five or six years at the discretion of local leaders. (Liu, 1993) According to a survey by the State Council, (National Institute,

[3]This largely targeted rural areas, whose schools were still in session.

1984) in September 1973, 14 provinces had implemented a policy of nine years of primary and secondary schooling (comprising five years of primary school, two years of middle school, and two years of high school), seven provinces implemented a policy of ten years of primary and secondary schooling (comprising five years of primary school, three years of middle school, and two years of high school; or six years of primary school, and four years of a combined middle and high school) and another nine had nine years of primary and secondary schooling in rural regions and ten years in cities. Tibet allowed both five and six year systems of primary schooling to operate concurrently in different counties, and its middle school lasted three years.

3. After the Cultural Revolution (1977-present)

After the Cultural Revolution ended in 1976, China's education system gradually resumed its normal operation and educational quality rose in importance as a policy priority. (Hannum, Behrman, Wang, and Liu, 2008) In January 1978, The Ministry of Education issued the *Full-Time Ten-Year Primary and Middle Education Teaching Plan (Draft)*, which standardized the total length of primary and secondary education to be ten years, comprising five years of primary school, three years of middle school, and two years of high school. Soon afterward, responding to scholars' suggestions to institute a twelve year system of primary and secondary education, the Ministry of Education required local governments to discuss possible solutions to prolong the length of basic education. (Liu, 1993) At the end of 1980, the CCCPC and State Council issued the *Decision on Several Problems Relating to Universal Primary Education*, the policy whose changes we use for our analysis. This policy mandated that the total years of primary and secondary education should gradually be extended to twelve years. It stated that the length of primary schooling could be five or six years, but encouraged gradual adoption of six year primary school throughout the country, putting more pressure on urban schools.

In April 1981, the Ministry of Education issued another, stronger statement that the length of secondary education "should be" extended from five to six years in "most" regions by the end of 1985. (National Institute, 1984) Note that the language was neither imperative nor precise: though most regions did change in the immediate years to come, the roll-out of the change was permitted to occur according to local conditions and in many places primary education remained at five years of length until the early 2000s.

This gradual roll-out is reflected in figure 1, which shows a gradual increase of six year primary education across regions from its post-cultural revolution conception in 1980 to 2010.

It is important to note that the transition from five to six years of primary education could be done in a number of manners. Table 1 gives six examples of how the policy was enacted, taken from gazetteers in different implementing cities and counties across the country. In some cases, this was accomplished by simply forcing all fifth grade students to remain in primary school an extra year. In other cases, it was accomplished by picking a year of students (say, third graders) after which all students must take six years of primary schooling. In other instances, it was done by splitting up a fifth grade class and sending some on to middle school while retaining others at the primary school to finish a sixth year. It is also important to note that this decision was made at the county level. Though upper-level pressure certainly played a factor, as we discuss in section 4, most counties had ultimate say on the year in which the switch was made. These two facts suggest that for the few years immediately after the policy, the transition may have been a bit messy. This is borne out in our data, and in later sections we will discuss how the issues surrounding this discretionary implementation and messy transition affect our results and how we can interpret them.

3 Theory

In this section, we set up a simple conceptual framework to understand how families might respond to a policy such as this. We first write down a three-period model of an individual family's decision on whether or not to keep their child in school at two different levels of education and use this to study how the policy might change their decision criteria. We then aggregate this model to the level of a school-going population, introducing a continuum of families with heterogeneity in ability and income. This helps us to understand which type of family may change behavior as a result of the policy, and thus to understand the policy's compositional effects on the population of school-going children.

Our unit of analysis here is the family, conceived of as two parents and one child, making a series of unitary decisions[4]. The family maximizes utility over consumption across three periods, c_1, c_2, and c_3, given a utility function which is concave in its arguments and has a consumption floor in each period, $\underline{c}$, as given in equation (1):

$$u(c_1 - \underline{c},\ c_2 - \underline{c},\ c_3 - \underline{c}) \tag{1}$$

The budget constraints in each period are given in equations (2), (3), and (4):

$$y_1 = w(a)(1 - S_1) + y_f \geq c_1 + p_s S_1 \tag{2}$$

$$y_2 = w(a, S_1)(1 - S_2) + y_f \geq c_2 + p_s S_2 \tag{3}$$

$$y_3 = w(a, S_1, S_2) + y_f \geq c_3 \tag{4}$$

Income comes from the family, y_f, and the wages of the child, $w(\cdot)$, if the child works. When the first period begins, the family chooses whether or not to send the child to middle school, $S_1 \in \{0, 1\}$, where 1 represents a decision to attend school and 0 to drop out. School fees are p_s, paid only if the child attends school. If she drops out in period one, she works and earns wages, $w(a)$, which are increasing in her ability endowment, a, in this and all subsequent periods. If the child drops out of school, she cannot return. If the child goes to school in period 1, at the start of period 2 the family makes a similar schooling decision, $S_2 \in \{0, 1\}$, which corresponds to the family's decision on whether or not to send the child to high school. The benefit of going to school in a given period is that the child gains higher wages from work in subsequent periods[5]. The lifetime income profile of the family is given in figure 2. The family will choose to send their child to school if equation (5) is satisfied and at least one of equations (6) and (7) is satisfied:

$$y_f - p_s S_t > \underline{c} \tag{5}$$

$$\begin{aligned} &u(y_f - p_s - \underline{c},\ w(a, S_1 = 1) + y_f - \underline{c},\ w(a, S_1 = 1, S_2 = 0) + y_f - \underline{c}) > \\ &\quad u(w(a) + y_f - \underline{c},\ w(a, S_1 = 0) + y_f - \underline{c},\ w(a, S_1 = 0, S_2 = 0) + y_f - \underline{c}) \end{aligned} \tag{6}$$

[4] In this analysis we abstract from the case of siblings. Within our framework, expenditure on siblings can be treated as either additional consumption or as a savings device for the family. This addition does not substantially alter our conclusions.

[5] We abstract from the non-pecuniary returns to school that are covered in papers such as Oreopoulos (2007). Adding this is straightforward, but does not sufficiently enrich our model to warrant consideration here.

$$u(y_f - p_s - \underline{c},\ w(a, S_1 = 1) + y_f - \underline{c},\ w(a, S_1 = 1, S_2 = 1) + y_f - \underline{c}) >$$
$$u(w(a) + y_f - \underline{c},\ w(a, S_1 = 0) + y_f - \underline{c},\ w(a, S_1 = 0, S_2 = 0) + y_f - \underline{c}) \quad (7)$$

Equation 5 states simply that if the family doesn't have enough wealth to both satisfy their basic needs and pay for school, they will not send their child to school[6]. Equations (6) and (7) are the individual rationality (IR) conditions, which state that the family will only send their child to school if it is in their best interest. The family will choose to send the child to school in the second period if the left hand side of equation (7) is greater than the left hand side of equation (6)[7].

Within this framework, the policy we study can be seen as adding to the cost of schooling, both in terms of requiring an extra year of school fees and in terms of the forgone wages that the child would have earned. In our model, we represent this cost as an increase in p_s, which, given y_f, mechanically increases the probability that equation (5) is not satisfied[8]. The basic question we can ask using this framework is whether these forgone earnings are sufficiently large to make a family withdraw their children from schooling in a given period when they would have not done so earlier. Two testable predictions we can take from this to the data are 1), impoverished families will be more likely to have the cost increase push them to where equation (5) is binding and in doing so make schooling infeasible, and 2), that groups with lower returns to education (such as those in rural areas) are more likely to be pushed over the threshold where education ceases to satisfy the IR condition.

The second part of the model incorporates an additional margin and puts the family's decision into a system with multiple agents. Here, we assume a continuum of families which are heterogeneous in two characteristics: 1), innate ability of the child, and 2), income. Income, y_f, is distributed uniformly, $y_f \sim U[\underline{c} + \varepsilon, \overline{y}]$, where $\varepsilon > 0$ and $\underline{c} + \varepsilon < \overline{y}$. Ability is similarly distributed uniformly across a continuum, $a \sim U[0, \overline{a}]$. Here we assume

[6]We ignore the case where $y_f < \underline{c}$.

[7]Note this does not follow trivially - it could be the case that it makes economic sense for the family to send their child to middle school, but the period 3 boost in wages from going to high school does not adequately compensate the family for the forgone earnings in period 2, and so it is worthwhile to invest in school in period 1 but not in period 2.

[8]It could also be conceived of as increasing value of the wages in periods 1 and 2, as affected children leaving any given level of school will enter the workforce a year older than before and would likely earn more as a result of this increase in maturity (and stature, in the case of students leaving after primary or middle school), and a decrease in the value of wages in period 3, as the child will enter the period one year later than usual and so will have one year less of adult income. This reframing does not significantly change our results.

additionally that the cost of schooling is inversely related to ability[9]. This variation in cost captures a feature of the Chinese education system and of many others in East and Southeast Asia, namely that a student must pass a series of exams to make it to the next level of schooling. The student's performance on these exams depends on their test taking skill, which is a combination of their innate ability and the amount of resources a parent spends on the child's education. (Lee, 2011) Here we assume the school fee includes not only the cost of attendance, but also the cost of resources[10] (henceforth "tutoring") the child requires to pass the entrance exam. For ease of exposition, we assume that the amount of tutoring needed is inversely and linearly related to ability, with the most able student needing to purchase no tutoring. We assume that a child must be above a certain minimum level of ability, $\underline{a} < \overline{a}$, to even potentially benefit from such tutoring, e.g. if $a < \underline{a}$, $S_1 = S_2 = 0$.

To understand compositional effects, we restrict attention to the case where it is always preferable to go to school if it is affordable and the student is of sufficient ability. These assumptions, and the consumption floor in the utility function, generate a school choice threshold in ability-income space, shown in figure 3, below which families choose not to send their child to school. The threshold itself is determined by three factors - the level of income below which the parents will not invest in schooling regardless of child ability, the level of child ability below which the child cannot attend school no matter what is spent on tutoring, and the slope of the line which determines the minimum combinations of ability and income necessary for the child to pass the entrance exam.

We propose that there are two direct effects of the compulsory primary education expansion policy, as shown in the second panel of figure 3. The first, akin to the first part of the model, is to increase the income threshold for families. This is labeled as effect 1 in the figure. Effect 2 is that, by making all children go through an extra year of primary education, the schooling reform condenses the ability distribution from the left

[9]There is a large literature, starting with Becker and Tomes (1976), on whether expenditure on children's education is positively or negatively correlated with children's ability. Our assumption here is only about the likelihood of a child passing an entrance exam, and an extension of two simple assumptions: 1) a child's likelihood of passing an entrance exam is positively related to both her ability and the amount of resources spent on exam preparation, and 2) ability and exam preparation are substitutes in the production function for passing an entrance exam.

[10]These resources include both money spent on exam preparation courses and tutoring and time, both parental time helping the child with her studies and time the family allows the child to spend doing homework and not housework. For many families, particularly those too poor to afford exam prep classes, the latter is the relevant margin. (Zhang, Hannum, and Wang, 2008)

hand side toward the right[11]. If there exists an absolute ability threshold that students must exceed to proceed to the next level, the policy would thus reduce the threshold $\underline{a}$ [12]. We will test for this effect by looking for changes in schooling attainment of less and more needy students; in this part of the model, as in the first part, needier students are more likely to be near the binding income constraint. Unfortunately, the CFPS lacks data on past ability or any reliable, independent proxy for resources spent on the individual in childhood, so we are unable to test the part of the model which deals with ability at the moment. We are currently exploring the use of other data setswith ability data which will allow us to test this prediction.

4 Data and instrument

This section describes the data we use and how we obtain the county-specific policy implementation years we use for our instrument. Our data come from the China Family Panel Study (CFPS), a large nationally representative data-set containing information from over 30,000 individuals in rural and urban China across 25 provinces, representative of 94.5% of China's population[13]. In figure 4, we show a map of China with the counties sampled in CFPS highlighted in red. The CFPS is conceived of as a panel, with six waves planned, taking place in 2010, 2012, 2014, 2016, 2018, and 2020. For this analysis, we use only the 2010 wave. This project is organized by a team of economists and sociologists at Peking University, and it collects a rich set of data on family structure, income, expectations, and several other social and economic indicators. Detailed information about the sampling structure and overall plan for CFPS is available in Lv and Xie (2012). As we describe below, it is particularly well-suited to help us identify when the policy of interest happened in a given county.

Though the *Decision on Several Problems Relating to Universal Primary Education* and its impacts are well known, it happened at different times across provinces and within provinces across counties. In principle, there are two ways to identify when the policy

[11] Meghir, Palme, and Simeonova (2013) find that a compulsory education reform in Sweden had a similar differential impact on the cognitive skills of those with initially lower ability endowments.

[12] This result could also be achieved with a relative threshold that allows a fixed proportion of students to advance by including noise in the entrance exam, which introduces a nonzero probability of failing the test inversely related to ability and money spent on tutoring. The results obtained here would be identical.

[13] The data include all provinces but Tibet, Xinjiang, Inner Mongolia, Hainan, and Ningxia.

took effect in a given county. The first method is to link data with county-specific identifiers to historical records in county gazetteers, as in other recent work on China. (Zhang, 2012) The second method is to use the data itself to infer a county's year of implementation. In many data-sets, county information is suppressed to protect the anonymity of study participants, and until very recently, gazetteer records were kept only in the national archives in Beijing and were prohibitively difficult to access in large numbers. In this paper[14], we use the second method, for which the CFPS is particularly well-placed. Unlike previous large-scale data-sets from China, it asks adult survey respondents not only about attendance and graduation from primary school, but also for how many years the respondent attended primary school. We use this data to infer when the policy took effect in a given county.

One limitation of this paper is that we only have observations of individuals who were present at the time of the CFPS survey. This means that those migrants who are not observable by the surveyors will not be counted and our results will thus give us a biased estimate of the population treatment effect. The CFPS was able to reach migrants who migrated for work within their county, but is not able to reach those who have left their county for work and are not present at the time of the survey. There is not much we can do about this, but the relatively high response rate CFPS achieved (97% for households, 72% for identified adults within households) suggests that, at the very least, any migration-induced selection bias will be minimal. (Lv and Xie, 2012) Below, we explain how we use CFPS data and the second identification method to identify county-specific policy implementation years for use as instruments.

4.1 Method for identifying the policy implementation year

Our ability to identify the timing of policy implementation at the county level is best illustrated by a simple histogram. Consider figure 5, which shows, for a given county, the median number of years of primary education for all individuals graduating in each of 30

[14]We recently gained access to a large dataset of county gazetteers which give county-specific policy implementation years for our policy. In future versions of this paper, we will link these reports to household survey data with identifiable counties from the China Household Income Project (CHIP) and the China Urban Labor Survey (CULS). This will allow us to compare the two methods of identification for agreement and will greatly increase our sample size.

years, from 1976 to 2005 [15]. We see that the median in this county is consistently at or around five years until 1998, when there appears to be a structural break, after which it is at or above six. For each of the 162 counties in our sample, we look at a series of such histograms and classify each county according to the clarity of the break on a scale of 1 to 4, 1 being a clear break with only one candidate year, 2 being a clear break with at least two possible candidate years, 3 being an unclear break, and 4 being no visible break. For all counties save those in the city of Shanghai[16], we then look through a series of further histograms which

1. Dampen extreme values of reported years of primary education so that responses greater than six years are replaced as six and less than five are replaced as five;
2. Exclude those who do not graduate to address issues of confounding from students who are retained;
3. Exclude migrants who attended primary school outside the county; and
4. Exclude adult learners returning to school after an absence to finish their degree, as their duration of primary school may be influenced by factors other than this policy.

From this exercise, we are able to assign two treatment years to 112 of the 144 remaining counties, or about 78% of the post-Shanghai sample. The first treatment year is the year after which the median number of years of primary education is consistently at or above 5.5 years. The second is the year after which the median is consistently at or above six years. This was chosen in light of evidence from some county gazetteers[17] that the implementation of the primary education expansion policy was in some counties expanded gradually, with these counties splitting the first "affected" class, sending some

[15]The lower bound of this range is chosen to coincide with the end of the Cultural Revolution and the chaos it brought with it to the educational system of China, as described in section 2. The upper bound is chosen because our survey was administered in 2010. Functionally, however, we will not be able to do much with counties who implement the policy in or after 2005, as we have to exclude individuals in our analysis who graduated in or after 2005 as many of them will not have completed their education and so would bias the estimates of the impact of the policy on total years of education attained downward. We hope to resolve this shortcoming using subsequent waves of CFPS which will allow us to include those graduating later.

[16]Shanghai implemented the policy by extending the length of middle school from three to four years instead of extending the length of primary school. As this is different from the initial reaction of all other provinces in our sample, we exclude Shanghai from our analyses, 11% of our original sample.

[17]See table 1 for examples.

on to middle school and retaining the rest for a sixth year of primary education. There are often multiple primary schools in a given county and anecdotal evidence suggests that roll-out is likely to have been staggered between primary schools[18] in some regions by a year or two to smooth the flow of children to middle schools in transition years.

To further ensure that our visual inference is accurate, we use a mean shift model (Fukunaga and Hostetler, 1975) to examine the data and choose the most likely policy implementation year for each county. A mean shift model is a simple statistical tool, often used in fields such as machine learning and digital recognition, which looks through data to identify places where structural breaks are most likely. The implementation of a mean shift model in our context is quite simple - we run OLS regressions on primary education data collapsed to the county-by-graduation year median[19] at the county level as in figure 5. For each county, we run 30 OLS regressions corresponding to every possible treatment year in our data, $t^* \in [1976, 2005]$, of the following form: $y_t = \beta_0 + \beta_1 * \mathbf{1}\{t \geq t^*\}$. In this equation, y_t is the median number of years of primary education among students who graduated from primary school in year t, the indicator function is equal to one for medians from students graduating in year t^* and after. We save the sum of squared residuals (ssr) for each of these regressions, and the year (t^*) with smallest ssr is the predicted treatment year.

4.2 Results

We show the distribution of treatment years across the support of potential treatment years in the upper panel of figure 6. The lower panel shows a histogram which gives, for each of the counties in our sample, the difference between the mean shift rule and the six year visual inspection rule. In 56 of the 112 counties, the mean shift rule and the visual rule give the same year. In only 24 counties do the two differ by more than 2 years (21.4%). This drops to eight for counties graded as at least "somewhat clear" and to only four for "clear" counties.

In our analyses, we use the policy implementation year identified by visual inspection

[18] See table 1.

[19] We performed the same exercise using the mean, which yielded similar results. We chose the median because it made it easier for both our visual inspection and the mean-shift model to identify the structural break. Estimation results which use the policy year instruments identified using the mean are also quite similar.

using the six year rule as the main instrument, as we believe the information we obtain from comparing the different samples to be very informative in many cases and our mean shift model program is not sufficiently sophisticated to pick up these nuances. The point estimates of our results using the mean shift model year and the 5.5 year rule are in all cases quite similar to those from the six year visual inspection year which we present in section 5.

To corroborate the accuracy of our instrument, we look at province-level statistics from China's National Ministry of Education which show, from 1985 onward, the percentage of children in a given province enrolled in schools providing six years of primary education. In a series of graphs in figures 7 and 8, we plot this percentage alongside province-specific cumulative distribution functions which track, by year, the proportion of counties from that province in our sample which have implemented the policy in that year according to our identified county-specific implementation years. Visual inspection of the graphs shows the national level data and our identified policy implementation years track remarkably well. The within-province, across-time correlation between the two variables is .7957.

4.3 Threats to internal and external validity

It is important to be aware of the fact that the implementation of this policy was not random. It was decided upon by bureaucrats at the province and county level and, as described in section 2, this decision was based on a mixture of local circumstances, including ability to implement the policy, and pressure from higher levels of bureaucracy. We explain later how this does not hurt our identification strategy, as we are identifying off of differences within a given county. Still, a set of potential concerns exists which could cloud interpretation of our results. These revolve around potential correlation between certain local social and economic conditions and both the nature and timing of policy implementation.

The main concern is that such a correlation could suggest a third omitted factor influencing both timing of implementation and families' decisions on total number of years of education. A related concern is that of selection bias, e.g. whether the counties we are forced to exclude because we are unable to identify the implementation year differ systematically from those in which we can identify the policy implementation year. The issue of a third omitted factor is important, as any evidence of this will have to be taken into consideration when we attempt to understand how people are behaving in response to the policy. The issue of selection bias is of less concern to the internal validity of the study, as we are primarily interested in the "treatment on the treated" effect, which we focus on to get at the behavioral response to the policy, as opposed to the "intent to treat effect" which includes data from non-compliers. There is thus no direct threat to the internal validity of our estimates, though generalizability may be a concern if there are important differences between compliers and non-compliers (identifiable and non-identifiable counties) or between early and late adopters.

To address these concerns, we first explain why we exclude certain counties, and show evidence that the excluded counties are quite similar on a menu of characteristics to the counties which we retain in the sample, and are particularly similar to the large mass of counties which implement the policy before the mid 1990's. Table 2 lists the main reasons for exclusion alongside the number of counties excluded for this reason. We exclude 50 counties in all, 18 in Shanghai (as described in footnote 16) and 32 elsewhere. This table shows that, aside from Shanghai's peculiar implementation of the policy, there were three main reasons for exclusion. The first is that in nine counties there was no increase in the amount of primary schooling over the window in which we focus. Counties excluded for this reason came from only three provinces, two of which (Henan and Shandong) were provinces in which the majority of students, according to national statistics, only began to benefit from the policy in the mid-2000's. Given that we have very few observations in any given county after 2005, we would not expect, ex-ante, to be able to identify policy implementation in these counties.

The second reason for exclusion, responsible for 12 counties, is that over our period of analysis the median number of years of primary education in these counties was consistently at or above six. As discussed in section 2, six year primary school was the norm in China prior to Mao's frenetic reorganization of Chinese education in the 1950's and 60's. Though these educational reforms were widespread, there is evidence that implementation of some policies was less than universal (Liao, 2004) and we suspect that this is the reason for our inability to identity a policy change in these counties. The third reason, accounting for the exclusion of nine counties, is that there was no clear implementation year.

This could be the result of haphazard implementation, frequent policy experimentation[20], or schools "gaming" the system, perhaps, for example, inducing weaker students to drop out before primary school graduation and encouraging more able students to carry on. This would be a larger concern were we worried about measuring the effect of the policy on all students in China. As our primary outcome of interest is the behavior of families after being "dosed" with an extra year of primary school, this type of idiosyncratic policy behavior at the school level, while interesting, does not immediately affect our ability to estimate the relationships of interest in this paper[21]. We are further comforted by the fact that inability to identify the policy year for this particular reason is the case in less than six percent of our total sample of counties.

The next step is to compare characteristics of non-compliers, early takers and late takers at the county level. To classify counties as early and late takers, we generated a histogram of policy implementation years. Based on the visible break in the bimodal distribution (see figure 6) and our knowledge of the policy, we assigned 1995 as the cutoff year. Counties implementing the policy before this date were considered early implementers, and those implementing afterward, late. County level data is given in table 3 and individual level data is given in table 4. For county-level characteristics, relative to villages which are early compliers, late compliers have fewer residents by about half, and slightly larger households. These characteristics often go together - in rural parts of China, there are allowed exceptions to and lenient enforcement of the one child policy, population is more sparsely distributed across space, and policy changes take longer to trickle down. The above village level differences between late and early implementers are different at the 5% level. Importantly, late compliers have only a slightly higher proportion of families on welfare, a proxy for the poverty level of the village, and seem to get public works such as electricity and connection to a major road about the same time, though on average they get running water a few years later. These differences are not significant at conventional levels. This similarity is important, because these are major milestones which might be correlated with government officials' predisposition to

[20] In some counties, the form of implementation changed over time, e.g. going from five years of primary and three of middle school (5+3) to 6+3, then to 5+4 as in Shanghai, and so on, and such changes could also explain some of the patterns we see. This is corroborated by the anecdotes in table 1.

[21] This is of greater importance in the companion paper, Hu and Eble (2013), and is dealt with more thoroughly there

implement progressive education policies.

The individual level data shows a few noteworthy differences between the major demographic and schooling variables of interest. Late implementers are 20 percent less likely than early implementers to be in urban areas and get, on average, and get about one less year of schooling. They are also much less likely than early implementers to graduate from university. Wald tests, performed but not shown, reject equality of means on these characteristics at the five percent level. In light of the differences we find, we investigate the possibility of heterogeneous treatment effects of the policy between early and late implementers in section 5. Non-compliers are strikingly similar to early implementers on all of the individual variables we examine.

We estimate a Cox proportional hazards model using these same village/community characteristics to understand which are most predictive of early policy implementation. We use 1980 as our base year, as it is the first year in which the policy could be implemented. The results for this exercise are given in table 5. The results here largely follow what is shown in tables 3 and 4 above. The major predictor of early implementation is whether the county is located in a western province (Chongqing, Gansu, Guangxi, Guizhou, Shaanxi or Sichuan). Western provinces in China have typically been the least developed, more sparsely populated, and have fewer large metropolitan areas to which temporary migration is easy. There are 23 counties in western provinces and 89 in non-western provinces. In the empirics section, we also look for heterogeneous treatment effects between western and non-western counties. As in the tables, counties with larger populations and smaller household sizes also implement the policy earlier; counties with more primary schools implement the policy later.

Finally, underlying every regression discontinuity is the assumption that observations immediately on either side of the discontinuity are comparable on important characteristics. Our ability to provide data justifying this comparison is hindered by the fact that our data is collected long after the policy has taken place, and so characteristics such as parents' income and occupation, which we expect may have been influenced by the policy and will have changed over time, cannot be easily compared for this purpose. We can, however, compare characteristics which are unlikely to change over time. In table 6, we compare gender, household registry status, ethnic minority status, and residence between treated and untreated groups. Columns 1 and 2 compare values for these variables for all treated and untreated observations in our main sample. Here there are some notable differences - the treated individuals are 10 percentage points more likely to be from urban areas, 8.7 percentage points more likely to have their household registry be from an urban area, and 4.4 percentage points more likely to be from a non-Han (i.e. minority) ethnic group. There are also many more untreated individuals than there are treated. Restricting the sample to a five year bandwidth around the treatment year, however, these differences all attenuate substantially, particularly the imbalance in number of observations. Wald tests fail to reject equality of means for each of the four characteristics using the restricted bandwidth sample. Still, in the regressions that follow, we control for these variables, and will investigate heterogeneous treatment effects based on urban/rural residence and gender in the next section[22].

5 Empirics

This section gives the empirical identification strategy we use in the paper and then presents the main empirical results of our analysis. Using descriptive statistics, graphical representation of the data and then a regression discontinuity design, we first show evidence that the policy seems to have increased total years of education by exactly one year, which suggests that, aside from the extra year of primary school, affected individuals are attaining around the same number of years of education as they were before the policy. We look at differences between treated and untreated groups in post-primary educational attainment only and confirm this result. Drilling a bit deeper, we look at graduation rates and indicators for ever having attended a given level of education to try to disaggregate our effect across different levels of schooling. We then look for differential effects between certain subgroups - late vs. early implementers, rural vs. urban areas, western vs. central and eastern provinces, and men vs. women. Finally, we address potential concerns about implementation and interpret the results we find.

[22]Though it is of potential interest, our sample size does not allow us to investigate heterogeneous treatment effects between minorities and Han Chinese.

5.1 Empirical strategy

The empirical identification strategy used in this paper is a regression discontinuity design. We use instrumental variables regression on a sample restricted to a few years' bandwidth around the implementation year to estimate the causal effect of increasing the length of primary school by one year on the educational decisions made by families later on in their children's schooling. We use the policy experiment described in section 2 to provide exogenous variation in the number of years of primary school children are required to attend. The main equations we use for estimation are as follows.

$$x_{ci} = \beta_0 + \beta_1 * Treated_{ci} + \beta_2 * [t_{ci} - t_c^*] + \lambda_t + \mu_c + \epsilon_{ci} \tag{8}$$

$$y_{ci} = \gamma_0 + \gamma_1 * Treated_{ci} + \gamma_2 * [t_{ci} - t_c^*] + \lambda_t + \mu_c + \epsilon_{ci} \tag{9}$$

$$y_{ci} = \delta_0 + \delta_1 * \overline{x_{ci}} + \delta_2 * [t_{ci} - t_c^*] + \lambda_t + \mu_c + \epsilon_{ci} \tag{10}$$

Equation (8) gives us the first stage. x_{ci} is the total number of years of primary school attained by individual i in county c. $Treated_{ci}$ is an indicator variable equal to 1 if the individual graduated from primary school in or after the identified treatment year in her county. The term $t_{ci} - t_c^*$ captures the time elapsed between the county-level year of treatment, t_c^*, and the year in which individual i in county c graduated from primary school, t_{ci}, included to capture the possibility that effects may have been different as the counties adjusted to the policy. County (μ_c) and year (λ_t) fixed effects are also included in all specifications save two of the initial "naive" regressions. Equation (9) is the reduced form estimation strategy, with identical controls. Our dependent variables are total years of schooling, total years of schooling after primary school (henceforth "post-primary schooling"), middle school graduation, high school graduation, university or technical school (henceforth "post-secondary") graduation, and indicators for having ever attended high school or post-secondary school. Equation (10) gives the instrumental variables specification, where $\overline{x_{ci}}$ is the number of extra primary school years predicted by the first stage or the predicted value of the probability, given treatment, that an individual would have received at least six years of primary education. These estimates can also be obtained from the inverse ratio of coefficients estimated in equations (8) and (9). The identifying assumptions required for causal inference in these specifications are 1), after controlling flexibly for time and county effects, there is no other systematic difference between treated and untreated individuals within a sufficiently narrow bandwidth around the treatment year within a given county, and 2) we have continuity in the conditional expectation of the outcome variable on either side of the treatment threshold[23]. Under these assumptions, we attribute any difference that arises between treated and untreated individuals in such an analysis to be caused by the treatment.

Our survey begins with 33,610 observations. We exclude the 6,859 observations which report graduating from primary school before 1976 because, as discussed earlier, they are likely to have their trajectory influenced by the various and tumultuous changes of the Cultural Revolution, the most severe parts of which ended in 1971, before the first cohort of students included under this rule would have entered primary school. We similarly exclude those 10,153 observations who do are missing a year of graduation and are too old to have been affected by the policy[24]. We have to drop those 2,250 observations who do not have a primary school graduation year or the number of years they attended primary school, as they give us neither information on our dependent variable nor on our treatment variable. For 1,637 students who do give the number of years of primary school attended, we do not have a year in which they graduate. We impute this by assuming they all entered primary school at age 7 and graduated five years later if they were old enough to be in the "treated" portion of their county, and six years otherwise. This imputation is imprecise, as children in our sample begin attending primary school between the ages of 5 and 10, and so we drop all 226 observations (from the original pool of 1,637) with imputed values for which the imputed graduation year is within two years of the relevant county treatment year. Excluding those in counties with no treatment year, those who graduated from primary school after 2004, and migrants brings our main sample to 8,146. All of our results are robust to excluding the 2,289 individuals which are in counties with an "unclear" quality of policy year identification[25].

The regressions we present all use robust standard errors which are clustered at the

[23]Which we examined in section 4.3.

[24]This worked out to be all individuals who were 42 years or older at the time of our survey. These individuals, if they had entered primary school at age 6, would have graduated from primary school before 1976 and would have been similarly affected by the cultural revolution.

[25]Graded "3", as discussed in section 4.

county level. Our total years of schooling variable refers to the sum total of years spent in primary school, middle school, high school and university/technical school. Post-primary education refers to this sum excluding the contribution from primary school. In our main regressions on educational attainment, as above, we dampen values so that the number of years of education contributed from each level of schooling cannot exceed the nationally mandated number of years for that level by more than one year. This means that high school, which normally is three years in duration, is allowed to be no more than four years. In primary school, this means that before policy implementation in a given county, we trim primary school attainment observations which are greater than six to be six, and after implementation, we trim values greater than seven to be seven. For robustness, we also conduct the same analyses without dampening values and by fully dampening post-primary values, e.g. high school can only contribute up to 3 years. In neither case do our point estimates change substantially, but using the raw data expands the confidence intervals, due to numerous observations reporting seven, eight, nine and even 10 years of primary school.

5.2 First stage

The "first stage" of our analysis is the proportion of students in our sample graduating from primary school in a given "distance to treatment" year (e.g. one year before the policy is implemented in the student's county, or five years after) who are getting six years of primary education. This is of limited interest because we are mechanically choosing the year in each county which maximizes the jump from five years to six years of education, but we feel it is still instructive to view the graph as we can learn from it what proportion of the population is affected by this policy. Figure 9 shows this data, including students within a 20 year bandwidth of being treated. Prior to implementation of the policy, the proportion of students getting at least six years of primary school is consistently around 25% of the population. At the chosen policy implementation year it then jumps to approximately 75%, increasing to over 95% of the sample in the 10 years following implementation. In other words, the first stage of our subsequent IV regressions will be an approximately 50% increase in the probability of completing an extra year of primary school. Regression output showing similar results is presented in table 8. These regressions show that, as we add controls, county and year fixed effects, we get a precise estimate of about .5, mirroring the pattern in figure 9.

5.3 Descriptive statistics and graphs

Our main outcomes of interest are the total years of education an individual attains and the years of post-primary education an individual attains, as described in section 5.1. We are also interested in the graduation rates of individuals from middle school, high school, and post-secondary school, and the probability they have ever attended high school or university. In table 7, we show the unconditional means for these data for the same four groups (treated and non, full sample and restricted to observations which graduated from primary school within five years of the treatment year, henceforth "within a five year bandwidth"). This table shows a few patterns: 1), treated individuals have higher educational attainment levels in all outcomes we consider, 2), the differences between treatment and control individuals attenuate substantially for the limited bandwidth sample relative to the entire sample, and 3), even in the unconditional means, the difference in post-primary years of education between untreated and treated is quite small in the restricted bandwidth.

To portray these patterns a bit more clearly, in figure 10 we present a simple histogram which shows, separately for untreated and treated observations, the distribution of the number of years of post-primary education. The sample is again restricted to students who graduate from primary school within five years of the policy implementation year, though the 10 year bandwidth results look quite similar. The figure shows that though the treated group seems to have slightly fewer individuals getting no post-primary education and slightly more individuals getting middle school, the differences are a few percent of the sample at most. The regression output we discuss below corroborates what inspection of these simple graphs suggests, namely that any differences we observe in post-primary schooling are not statistically significant.

We next present a series of graphs which plots the trend in total and post-primary years of education over time. In the top left panel of figure 11, which shows the raw data for total years of education, there is a clear upward trend. At the policy implementation year, there is a jump of about one year, but visually this appears to follow an overall trend displayed in the adjacent data. In the top right panel, we remove year-specific fixed effects and plot the residual data. Here we see a smaller trend and a more apparent, though still noisy, jump of about one year at the treatment year. We next look at years of post-primary education. In the bottom two panels, we see a similar pattern as in the total years of education graphs - the raw data show a clear upward trend in number of years of post-primary education as distance to the treatment year increases, and the residual data shows a smaller but still visible trend. Unlike for total years of education, in neither the raw nor residual data is there any visible jump in the number of years of post-primary education at the treatment year.

This graphical representation of the data over such a wide time horizon obscures the fact that we do not have a balanced panel - only some counties are contributing to the observations in the -20 bandwidth bin, and they are in all cases not the same as those contributing to the +20 bandwidth bin. This also prevents us from removing county-specific fixed effects without biasing the results, as point estimates in the various bins are comprised of different combinations of counties. To address this, we present similar graphs, first of raw data, then removing year and county fixed effects, for a "balanced panel", wherein we have observations in each distance-to-treatment bin from all counties in the sub-sample. To obtain a balanced panel with an eight year bandwidth, we have to drop counties implementing before 1984 and after 1997. This leaves us with data from 60 counties, a little more than half of the sample in which we can identify treatment years. These plots are given in figure 12. The residual data here is very noisy, but the trends apparent in the previous analysis describe what we see here with some degree of accuracy: we observe, on average, a higher level of total years of education after the treatment year, but there is no clear pattern for post-primary years of education. The dip in the residuals a few years after the treatment year is intriguing, but given we have only a few observations per county in this bin with the current data, this suggests to us that we require a larger sample to understand anything beyond the first moment of the estimated treatment effect. We are currently working on adding data sets to address this shortcoming.

5.4 Naive regressions

Here we conduct a set of "naive" reduced form and instrumental variables regressions to show how our treatment effect estimates change as we refine our specification. The reduced form estimates are shown in table 9, in which the dependent variable is total years of education, and in table 10, in which the dependent variable is years of post-primary education. In the first column, we show a simple regression of the dependent variable on the treatment variable. In these results, we see a large and significant effect of the treatment in both sets of regressions. In column two of both tables, we add the control for distance to treatment year, and the estimated coefficient for the treatment dummy is attenuated slightly. In column three we exclude the counties where we are least sure about our identification of the treatment year. This actually raises coefficients in both tables, suggesting that the measurement error in our identification of county policy implementation year may attenuate our results. In column four, we add year-specific fixed effects to control flexibly for time. This reduces our coefficient estimates drastically, which is to be expected, given that over the duration of our study period, 1976-2010, educational attainment increased substantially. In column five we add county fixed effects, which further attenuates our estimated coefficients and, in the post-primary educational attainment regression, renders the estimated coefficient insignificantly different from zero. If we were to take these results literally, we would interpret these as saying that the treatment raised total educational attainment by about .7 years and did not significantly change educational attainment levels in the post-primary period. In the next two columns, we add quadratic and then cubic measures of the distance to treatment variable to control more flexibly for effects which may vary relative to the distance to treatment. In no instance do these substantially alter the results we find.

The instrumental variables regression here inflates our reduced form with the fraction affected from the first stage and is shown in tables 11 and 12. The patterns in these results are largely similar - the final specification suggests that the treatment raised total educational attainment levels by 1.35 years, but there is no significant difference in the educational attainment of treated and untreated individuals in the post-primary period. While these are suggestive, given the nature of our regression discontinuity specification and our identification assumption, we are primarily interested in these analyses restricted to a narrow bandwidth around the treatment year, analyses which we provide in the following subsection.

5.5 Reduced form and instrumental variables results

In the analyses in this section, we present our results in two forms. One is graphical - we show the estimated regression coefficient along with a 95 percent confidence interval for the treatment indicator in the reduced form specification, and for the fitted value of the probability of getting at least six years of education in the IV specification, over fifteen different bandwidths corresponding to the number of years around the policy implementation year we include in our sample. We also provide regression tables providing more complete output for these regressions for 8 bandwidth values: 1, 3, 5, 7, 9, 11, 13, and 15 years around the policy treatment[26].

Total years of education

We first examine results of analyses where the total number of years of education is the dependent variable. In figure 13, we present coefficients on the treatment dummy variable and confidence intervals from reduced form and instrumental variables estimates of regressions as specified above. These results show that, in the reduced form, treatment corresponds with an approximately .5 year increase in total years of schooling across most bandwidths, and in instrumental variables estimates, "treated" individuals received approximately one full additional year relative to untreated individuals. The accompanying regression results are given in tables 13 and 15.

Post-primary education

This result of visual inspection is borne out by the reduced form and instrumental variables coefficients we estimate here, given in figure 14 and in tables 14 and 16. The estimated reduced form and instrumental variables are never significantly different from zero and are slightly negative after including more than seven years in the bandwidth. From these graphs and the related tables, we conclude that, at least on the whole, that the policy did not lead to any measurable increase or decrease in post-primary educational attainment.

[26]We do not include the other 7 regression outputs as they do not fit nicely on a one page table, but point estimates and confidence intervals are given in the graphical results and full tabular results are available from us on request.

Graduation rates

Here we probe deeper into the "no effect" result by looking at the effect of the policy on graduation rates from middle school, high school, and university/technical school as well as indicators for ever having attended these three levels of school. These graduation estimates are given in figures 15, 16, and 17 and tables 17, 18, and 19. Estimated coefficients on the treatment effect for restricted bandwidth reduced form and instrumental variables regressions are presented as above. These estimates show a fairly consistent "zero effect" for middle school, high school and post-secondary graduation with point estimates close to zero and confidence intervals which are symmetric around the x-axis. Estimates of the treatment's effect on ever having attended high school and university, as opposed to graduating, are given in figure 18, and mirror the graduation results.

Subgroup estimates

We next investigate the potential for heterogeneous treatment effects within four groups: 1), men and women, 2), urban and rural residents, 3), counties in western provinces and those which are not, and 4), early and late implementers. These are given in figures 19, 20, 21, and 22, respectively[27]. Our analyses here are motivated by our theoretical predictions that less wealthy groups and groups with lower returns to education are more likely to adjust the amount of schooling they get downward. These results should be taken with a grain of salt, as in most of our counties, particularly for narrow bandwidths, there are only a few observations on either side of the cutoff. Subgroup analyses such as these reduce the number of observations, often by more than a half, and so we might expect (and indeed see) large, insignificant, and unstable treatment effect estimates for narrow bandwidths.

We find some evidence that men may get slightly more additional post-primary schooling as a result, but this tapers off with time as in previous analyses, further suggesting the need for a larger sample size to give us the power necessary to investigate whether there is something special about the first few years of implementation. For women, on the other hand, we estimate consistently negative treatment effect estimates. Though these are never significantly different from zero, they are quite sizable. Taken literally, the five year bandwidth point estimate suggests that treated women attain between .5

[27]Tables for these results are available upon request.

and 1 year less post-primary education than non-treated women, though the result is not significant at traditional confidence levels. Women have higher returns to education in China as estimated by Mincerian regressions, but receive lower education, and previous work has suggested that families either perceive returns to education for women as lower than for men, which would match our theoretical predictions, or choose to allocate fewer resources to them for some other reason. (Li, 2003)

The difference between rural and urban residents follows a similar pattern. Estimates of the effect of the treatment on post-primary education for urban residents are consistently positive, large and in a few cases, significantly different from zero. Estimates for rural residents are consistently negative as the bandwidth grows, with the point estimate for a five year bandwidth and beyond at about half a year of education less than non-treated residents. As rural residents are generally of lower income than urban residents and have fewer work opportunities which reward high levels of education than urban residents, our findings here match well with our model's predictions. We find very little evidence of heterogeneous treatment effects on the the early/late and western/non margin.

5.6 Concerns and competing hypotheses

Here we address a series of concerns and potential competing explanations for the results that we find. The first and foremost concern is that China's 1986 compulsory education law, which stipulated that all Chinese students had to complete primary and middle school, may have prevented many students in our study from adjusting downward the number of years of middle school they attained. Recent empirical work has shown the law was gradually implemented and remains porously enforced. (Fang, Eggleston, Rizzo, Rozelle, and Zeckhauser, 2012) Even as recently as 2010, after the policy had been implemented across the country, dropout from middle school remained an issue, estimated at over 20% in some rural areas. (Mo, Zhang, Yi, Luo, Rozelle, and Brinton, 2013) To address these concerns, in figure 23 we plot a graph using CFPS data which shows the mean years of education for students graduating from primary school in a given year over time, and superimpose a graph of the proportion of students in this same time period who get less than eight total years of primary education. Here we see that though there is a dip after 1986, the proportion of students not completing middle school is greater than 20% in the majority of the years we study and this number varies surprisingly little. From this, we draw the conclusion that a large proportion of students in this period were in fact able to adjust on the lower margin were they inclined to do so.

Another concern with the ability to adjust is how flexible the upper margin is. If there were binding supply constraints on individuals which prevented them from advancing to the next level of schooling, we might observe a pattern similar to what we find in our data, that, on average, affected students did not increase their amount of post-primary education, but this result would be the result of having a binding supply constraint as opposed to the result of a concerted decision by the family. Though this concern is difficult to untangle without county and province-specific supply data (which we do not currently have access to), national level data provides suggestive evidence that this is not the case. (Ministry of Education, 2013) In figure 24, we show data from the Chinese Ministry of Educational Statistics which plots the proportion of Chinese students advancing from a given level of education to the next over our period of study. This shows that the proportion is steadily increasing at each of the major school transitions. While not conclusive, this suggests that any supply constraints that did exist were consistently being relaxed over the duration of our study.

5.7 Discussion

Based on our regression results, the evidence of clumping provided in figure 6, and the evidence provided above that both margins of adjustment were available to affected students, we attribute our finding that there was no change in the number of years of post-primary education to the importance of credentials in China, either for status, wages, or some other purpose unknown to us. A story in line with such an explanation is that wages matter, and attaining a given credential, e.g., a middle school diploma, carries with it a wage premium. In figure 25, we provide a bar chart which gives the mean non-agricultural annual income for individuals with a given number of years of post-primary education. We place vertical lines at the levels of degree attainment, i.e. at the attainment of a middle school degree, a high school degree, a technical degree and university degree. Though this is only correlation, it gives us an idea of what the wage structure in the economy is and what information a family might use to make their educational decisions.

The credentials story suggests we might see humps at these lines, but in most cases we do not. Rather, we see some evidence of a wage premium to having ever attended middle school, having ever attended university, and having either a technical or university degree. The lack of evidence for degree premiums/sheepskin effects, albeit only correlational, is surprising, and is seemingly in conflict with the clumping at degree levels we see in the data. Interpretation of this pattern, however, needs to be taken with a grain of salt given the large amount of clumping around each of these degree attainment levels as shown in figure 10. It could be, for example, that those who are induced to drop out after entering a degree level but before attaining a degree are of higher ability and so receive similar wages to those who stay on to continue, but are not otherwise comparable. If this were not the case, it would be curious to see so many individuals staying on until finishing their middle school or high school degree despite the fact that expected earnings do not seem to vary much with an extra year of middle or high school. One plausible explanation is that individuals are learning about their ability as they go through the educational system, and due to either noise in the examination process, uncertainty about ability which the entrance exams resolve, or both, it is worthwhile for most students to remain until their ultimate year of a given level of schooling when these entrance examinations are administered. This, however, is just conjecture and the distribution of income across years of education we find here is a novel finding that we will give more attention to in future research.

6 Conclusion

In the analysis presented here, we find that China's primary compulsory education policy reform of 1980 was successful in inducing people to attend one more year of primary education, but had no measurable impact on the education attained after primary school by the affected population. This result is shown both in self-reports on number of years of education attained and in coarser measures of graduation and ever having attended a given level of schooling. Looking into the possibility of heterogeneous treatment effects among different subgroups, we find weakly suggestive evidence that poorer groups and those with lower returns to education may have been induced to get less education. These results, however, are not statistically significant and require further analysis and more statistical power. We are currently working on incorporating into this paper data from two other large data-sets which will more than quadruple our sample size and decrease our reliance on our current method of identifying the policy year by giving us counties which we can link directly to county gazetteer data on when the policy was introduced. We hope this will give us more precise ancillary estimates with which to pin down compositional effects.

This study contributes what is, to the best of our knowledge, the first investigation of the impact that a compulsory primary education reform has on families' education decisions. In addition, it takes place in the developing world, where there is a paucity of studies on the impacts of compulsory education reform. We believe that the central messages - 1), that the additional year of educational reform was adopted by virtually all affected individuals, 2), it did not induce people, on average, to get more or less education after primary school, and 3), it may have induced some of the poor and those with low lifetime earnings expectations to drop out earlier - are important pieces of evidence to consider as other developing countries are making decisions about how to reform their primary education systems.

Particularly in places with large rural/urban disparities, limited mobility, and a population which is in school, on average, at least until part-way through middle school, such a policy has obvious uses. By extending the length of primary school, which usually takes place in or near students' place of residence, the government gains a year in which to drive home competencies and skills which it deems important for middle school and beyond and, at least under the circumstances we study, is guaranteed essentially full compliance. On the other hand, this may be costly to an economy. It will shrink the labor force temporarily, as there will be a gap when the first batch of students comes out of school a year later than anticipated. It may also shrink the labor force permanently, as, if families react to such a policy as they did in our analysis, the majority of laborers will consistently enter the labor market a year later than they would have otherwise and the labor market will have one less young cohort in it at any given time.

The large, significant first stage we find suggests several obvious lines for future work. In the companion paper to this study, we follow along the lines of the literature mentioned in the introduction, using the exogenous variation in education levels we find to estimate the relationship between years of education and labor market outcomes. In future work, we hope to expand the analysis to issues such as health and migration. We believe this would provide some of the first rigorous evidence from the developing world to several of the literatures which link national schooling policy changes to later life outcomes. We also hope to use the educational decisions we study and the patterns we uncover as our first "foot in the door" to understand, along the lines of papers such as Atkin (2012), the relationship between changes in the labor market and people's decisions about educational attainment. Finally, and perhaps most importantly, will be to combine these lines of inquiry to give a comprehensive assessment of what the overall costs and benefits of the policy were. The usefulness of such a policy will ultimately rest on whether its beneficiaries are more productive or otherwise more valuable to society, and whether this increase in value justifies the costs of teaching children for an extra year and the forgone wages that come from these individuals spending an additional year outside of the workforce.

References

Acemoglu, D., and J. Angrist (2001): "How large are human-capital externalities? evidence from compulsory-schooling laws," in *NBER Macroeconomics Annual 2000, Volume 15*, pp. 9–74. MIT Press.

Akerlof, G., and R. Kranton (2002): "Identity and schooling: Some lessons for the economics of education," *Journal of Economic Literature*, 40(4), 1167–1201.

Angrist, J., and A. Krueger (1991): "Does Compulsory School Attendance Affect Schooling and Earnings?," *Quarterly Journal of Economics*, 106(4).

Atkin, D. G. (2012): "Endogenous Skill Acquisition and Export Manufacturing in Mexico," Discussion paper, National Bureau of Economic Research.

Attanasio, O., and K. Kaufmann (2009): "Educational choices, subjective expectations, and credit constraints," Discussion paper, National Bureau of Economic Research.

Becker, G. S. (1975): *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. National Bureau of Economic Research, Inc.

Becker, G. S., and N. Tomes (1976): "Child Endowments and the Quantity and Quality of Children," *The Journal of Political Economy*, 84(4), S143–S162.

Black, S. E., P. J. Devereux, and K. G. Salvanes (2005): "Why the Apple Doesn't Fall Far: Understanding Intergenerational Transmission of Human Capital," *The American Economic Review*, 95(1), 437–449.

Brinch, C. N., and T. A. Galloway (2012): "Schooling in adolescence raises IQ scores," *Proceedings of the National Academy of Sciences*, 109(2), 425–430.

Brunello, G., F. D., and M. Fort (2013): "The causal effect of education on body mass: Evidence from Europe," *Journal of Labor Economics*.

Devereux, P. J., and R. A. Hart (2010): "Forced to be Rich? Returns to Compulsory Schooling in Britain*," *The Economic Journal*, 120(549), 1345–1364.

Fang, H., K. N. Eggleston, J. A. Rizzo, S. Rozelle, and R. J. Zeckhauser (2012): "The Returns to Education in China: Evidence from the 1986 Compulsory Education Law," Working Paper 18189, National Bureau of Economic Research.

Fukunaga, K., and L. Hostetler (1975): "The estimation of the gradient of a density function, with applications in pattern recognition," *IEEE Transactions on Information Theory*, 21(1), 32–40.

Hannum, E. (1999): "Political change and the urban-rural gap in basic education in China, 1949-1990," *Comparative Education Review*, 43(2), 193–211.

Hannum, E. C., J. Behrman, M. Wang, and J. Liu (2008): "Education in the reform era," .

Hanushek, E. A., and D. D. Kimko (2000): "Schooling, labor-force quality, and the growth of nations," *American economic review*, pp. 1184–1208.

Hanushek, E. A., V. Lavy, and K. Hitomi (2006): "Do Students Care about School Quality? Determinants of Dropout Behavior in Developing Countries," Working Paper 12737, National Bureau of Economic Research.

Hu, F., and A. Eble (2013): "Labor market consequences of expanding compulsory education: New evidence from a Chinese policy experiment," *Mimeo, Brown University*.

Jensen, R. (2010): "The (perceived) returns to education and the demand for schooling," *The Quarterly Journal of Economics*, 125(2), 515–548.

Kremer, M., and A. Holla (2009): "Improving Education in the Developing World: What Have We Learned from Randomized Evaluations?," *Annual Review of Economics*, 1(1), 513–542.

Lee, Y. S. (2011): "School Districting and the Origins of Residential Land Price Inequality," Working paper.

Li, H. (2003): "Economic transition and returns to education in China," *Economics of education review*, 22(3), 317–328.

Liao, Q. (2004): "History and Lessons from Contemporary China¡¯ s School System Reform (in Chinese, dangdai zhongguo xuezhi gaige de fazhan licheng yu jingyan jiaoxun)," *Journal of Nanjing Xiaozhuang College*, 20(2), 9–16.

Liu, Y. (1993): *Book of Major Educational Events in China 1949-1990 (in Chinese)*. Zhejiang Education Publishing House.

Lochner, L., and A. Monge-Naranjo (2012): "Credit Constraints in Education," *Annual Review of Economics*, 4(1), 225–256.

Lv, P., and Y. Xie (2012): "Sampling Design of the Chinese Family Panel Studies," Working paper, Institute of Social Sciences Surveys, Peking University.

Meghir, C., M. Palme, and E. Simeonova (2013): "Education, Cognition and Health: Evidence from a Social Experiment," Working Paper 19002, National Bureau of Economic Research.

Meng, X., and R. G. Gregory (2002): "The Impact of Interrupted Education on Subsequent Educational Attainment: A Cost of the Chinese Cultural Revolution*," *Economic Development and Cultural Change*, 50(4), 935–959.

Mincer, J. (1974): *School, experience and earnings*. NBER, New York.

Ministry of Education, o. t. P. R. o. C. (2013): "Promotion rate of graduates of schools of all levels," .

Mo, D., L. Zhang, H. Yi, R. Luo, S. Rozelle, and C. Brinton (2013): "School Dropouts and Conditional Cash Transfers: Evidence from a Randomised Controlled Trial in Rural China's Junior High Schools," *The Journal of Development Studies*, 49(2), 190–207.

National Institute, o. E. S. (1984): *Chronicle of Education Events in China (in Chinese)*. Educational Science Publishing House.

Oreopoulos, P. (2007): "Do dropouts drop out too soon? Wealth, health and happiness from compulsory schooling," *Journal of Public Economics*, 91(11), 2213–2229.

Pischke, J.-S., and T. von Wachter (2008): "Zero Returns to Compulsory Schooling in Germany: Evidence and Interpretation," *The Review of Economics and Statistics*, 90(3), 592–598.

Psacharopoulos, G. (1985): "Returns to education: a further international update and implications," *Journal of Human Resources*, pp. 583–604.

Spence, M. (1973): "Job market signaling," *The Quarterly Journal of Economics*, 87(3), 355–374.

Zhang, S. (2012): "Mother's Education and Infant Health: Evidence from High School Closures in China," *Mimeo, Cornell University.*

Zhang, Y., E. Hannum, and M. Wang (2008): "Gender-based employment and income differences in urban China: Considering the contributions of marriage and parenthood," *Social Forces*, 86(4), 1529–1560.

Figures and tables

Figure 1: National data on proportion of students in six year primary education

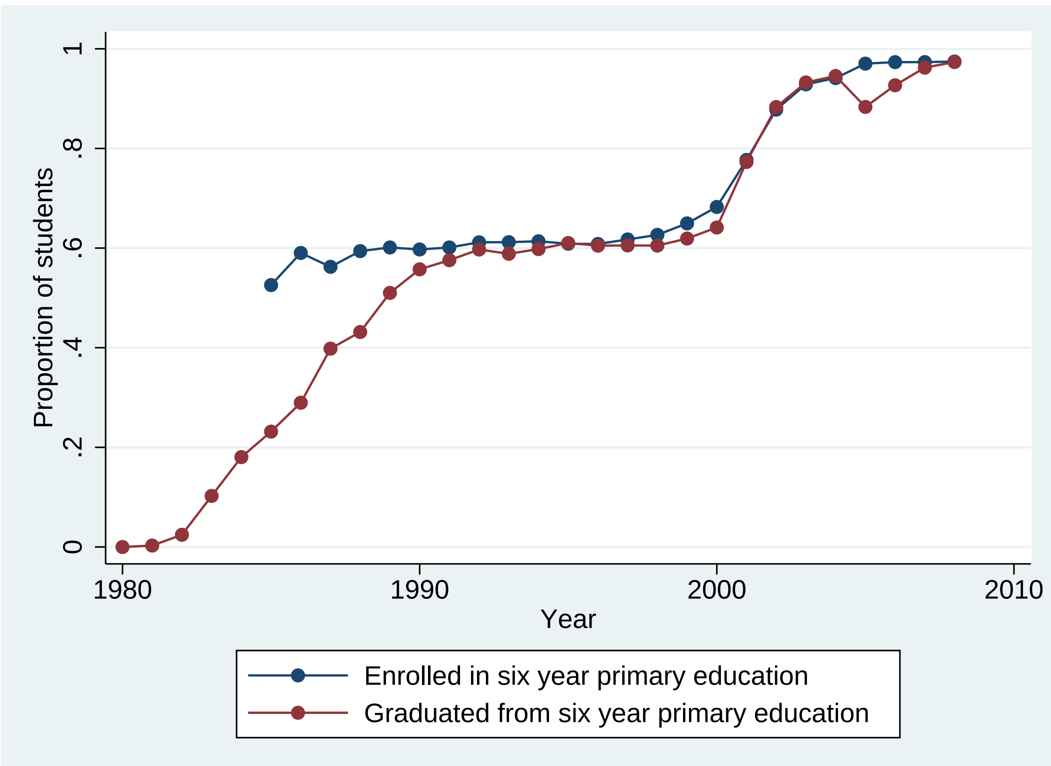


Figure 2: The family's decision between income and schooling

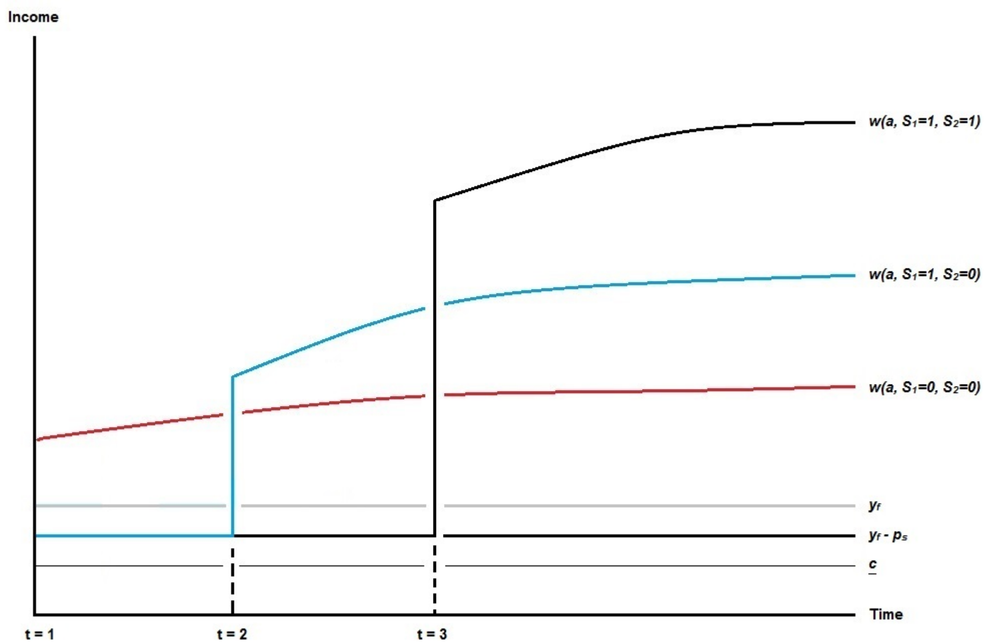
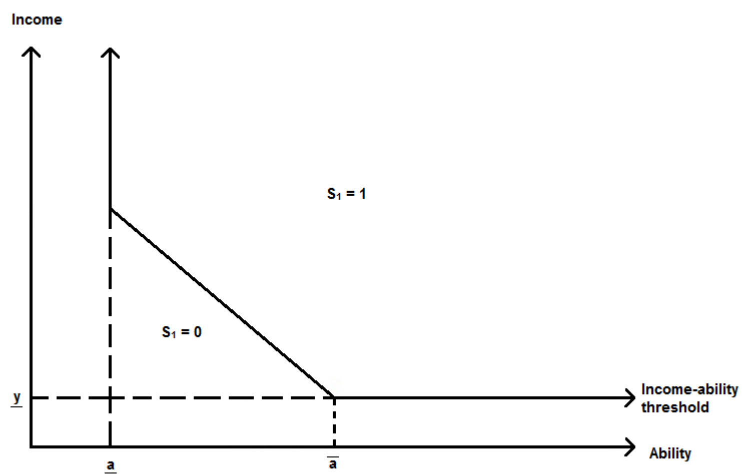


Figure 3: The schooling choice frontier in income and ability

Initial choice frontier



Choice frontier after policy is introduced

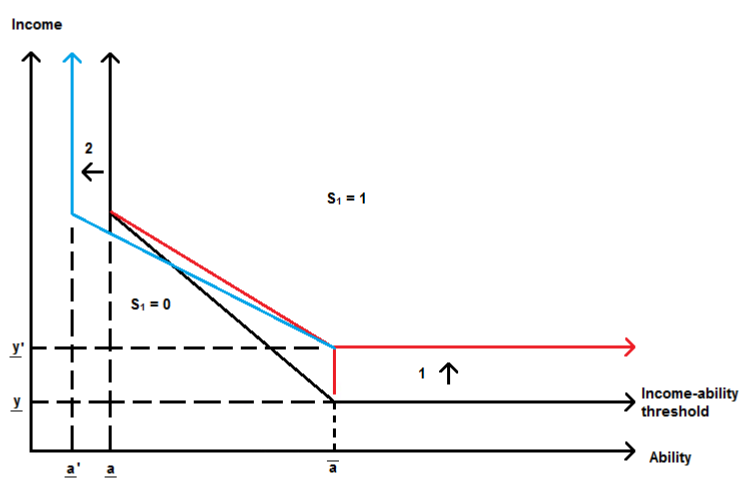


Figure 4: Sampled counties in CFPS

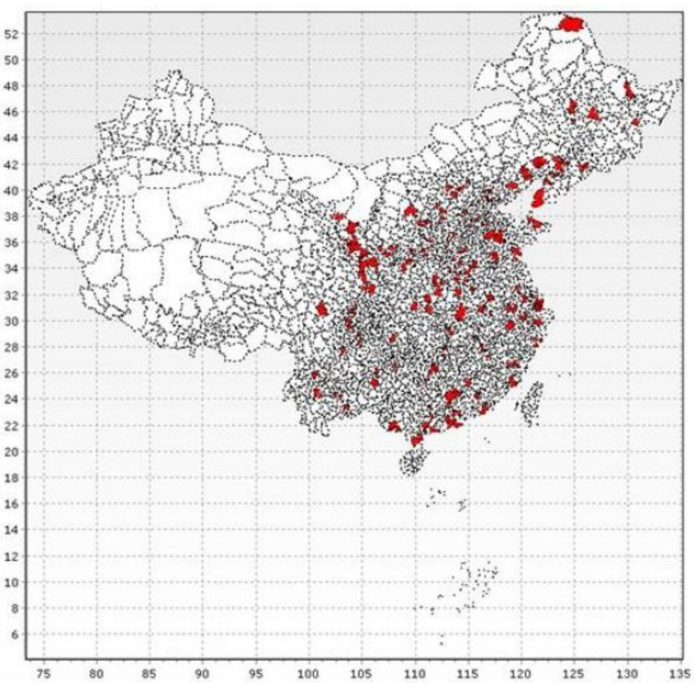


Figure 5: Example histogram for identifying policy year

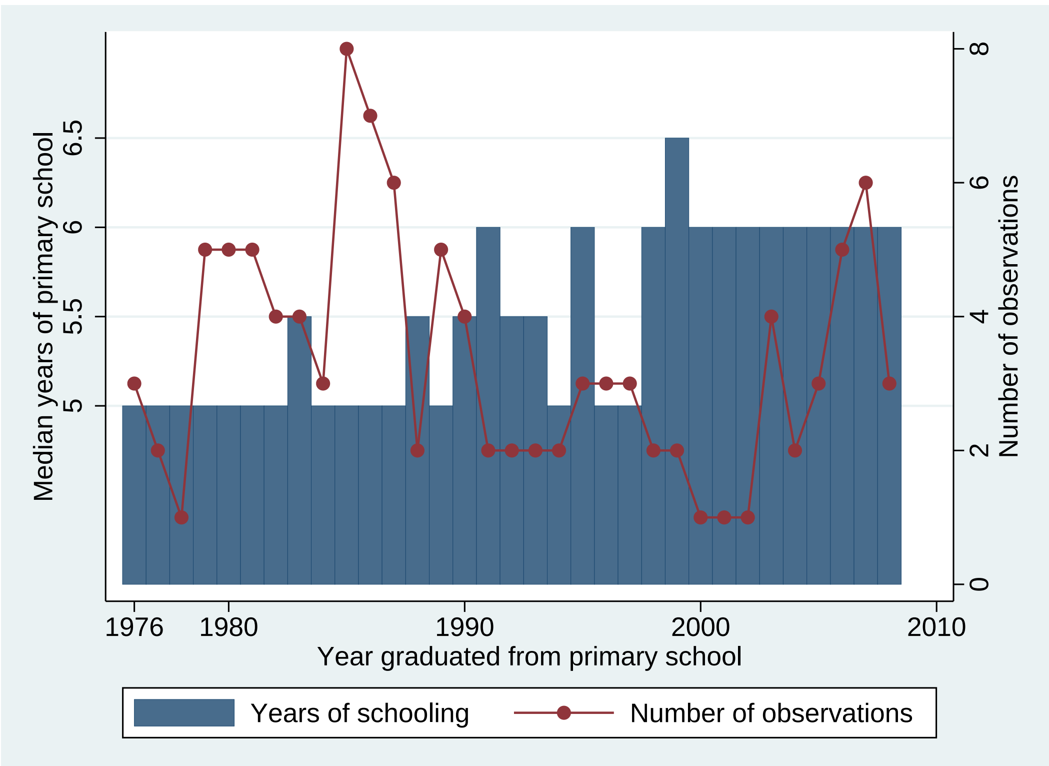
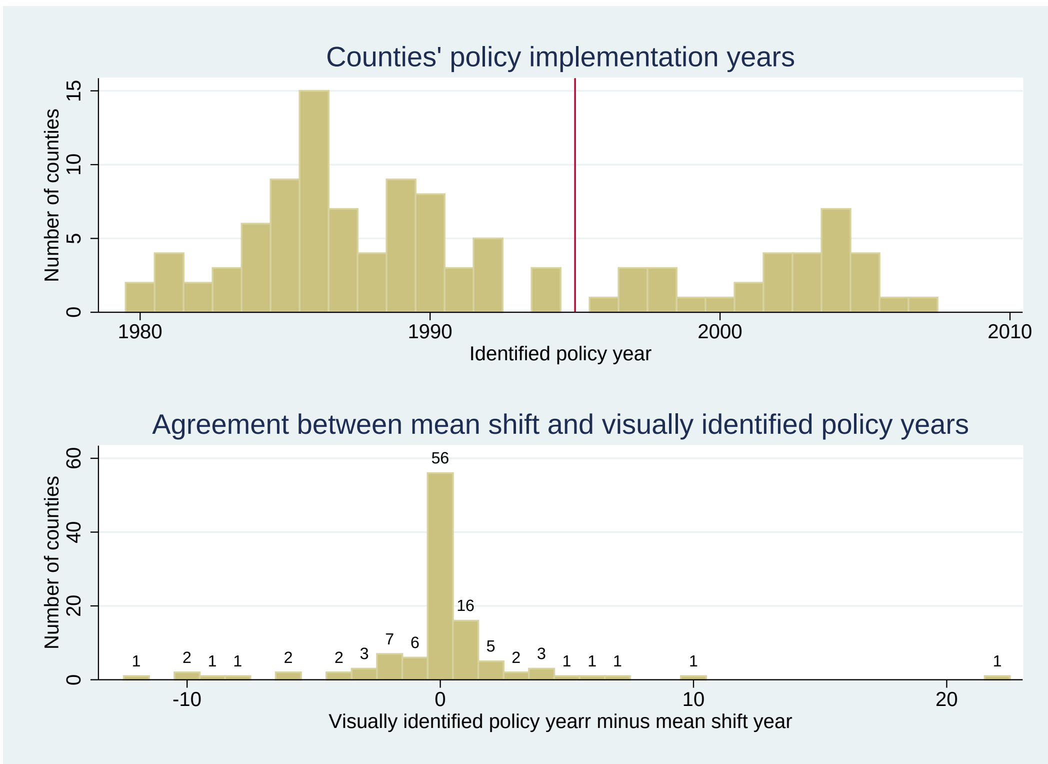


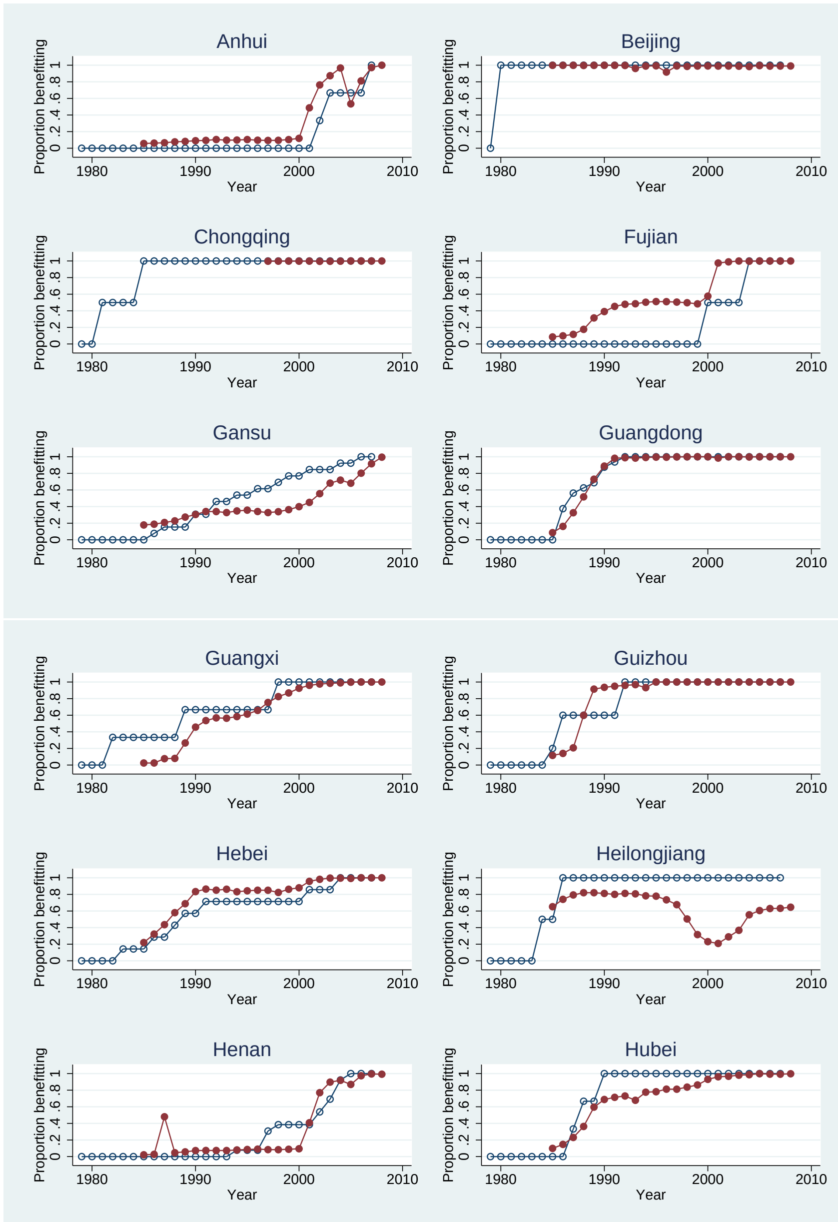
Figure 6: Implementation of policy across time



The vertical line in the top panel gives the cutoff between "early" and "late" implementers.

Figure 7: National statistics and province-level cdfs of policy implementation

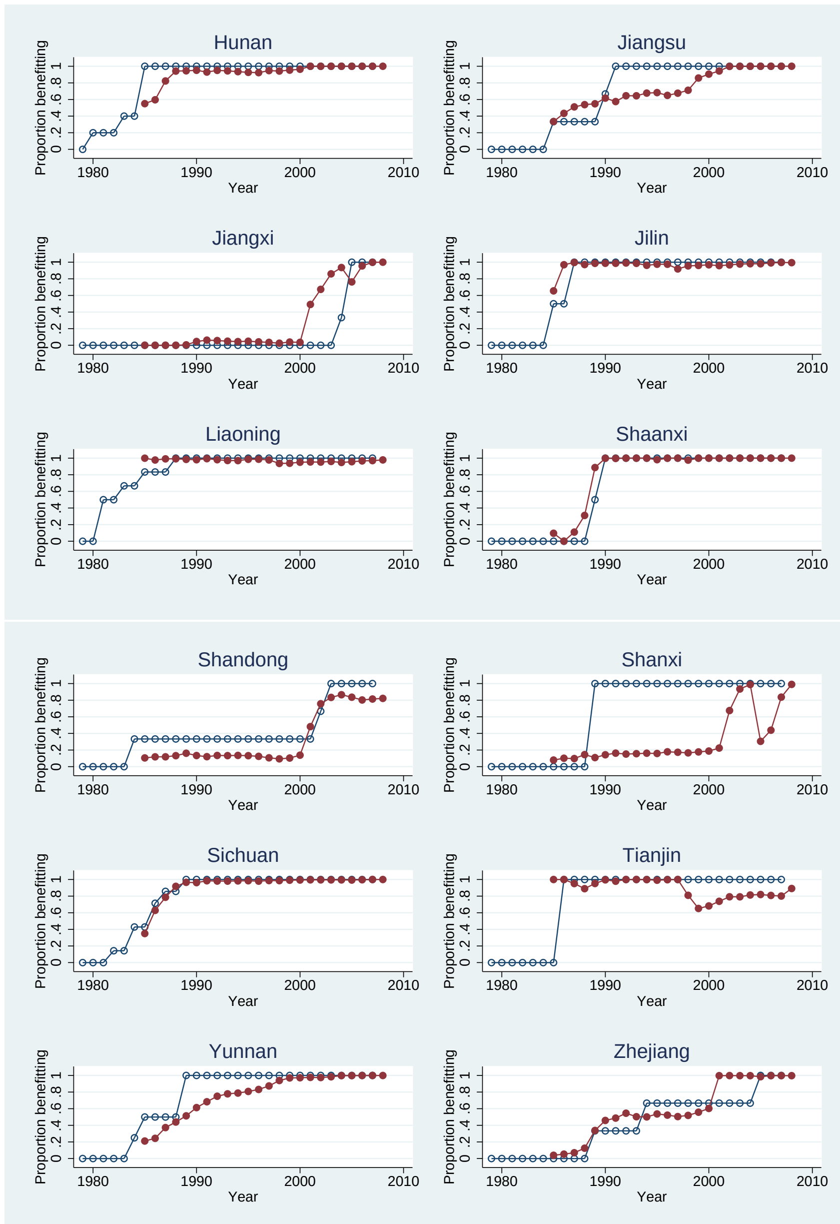
First 12 provinces



Note: hollow circles represent the cdf of proportion of treated counties in a given province by year according to our data. Solid circles are from publicly available, Chinese Ministry of Education statistics on the number of schoolchildren benefiting from six year education in a given province by year. Measurement error and policy experimentation cause these to trend downward in a few cases.

Figure 8: National statistics and province-level cdfs of policy implementation

Second 12 provinces



Note: hollow circles represent the cdf of proportion of treated counties in a given province by year according to our data. Solid circles are from publicly available, Chinese Ministry of Education statistics on the number of schoolchildren benefiting from six year education in a given province by year. Measurement error and policy experimentation cause these to trend downward in a few cases.

Figure 9: First stage - proportion of population affected by policy

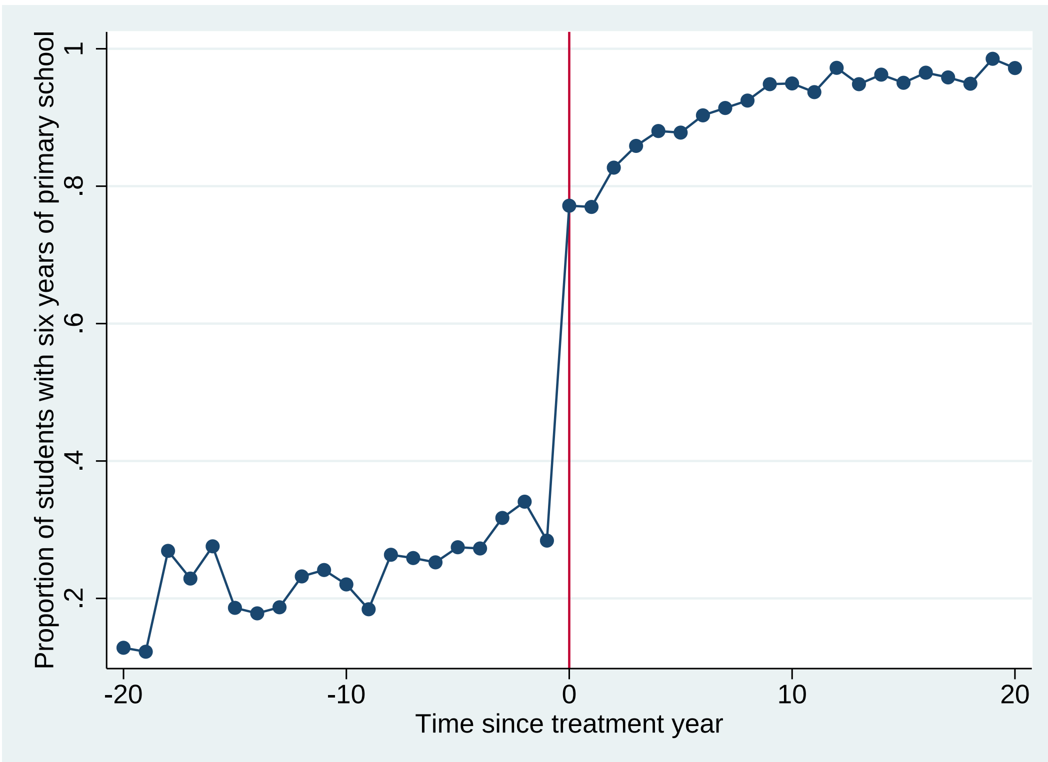
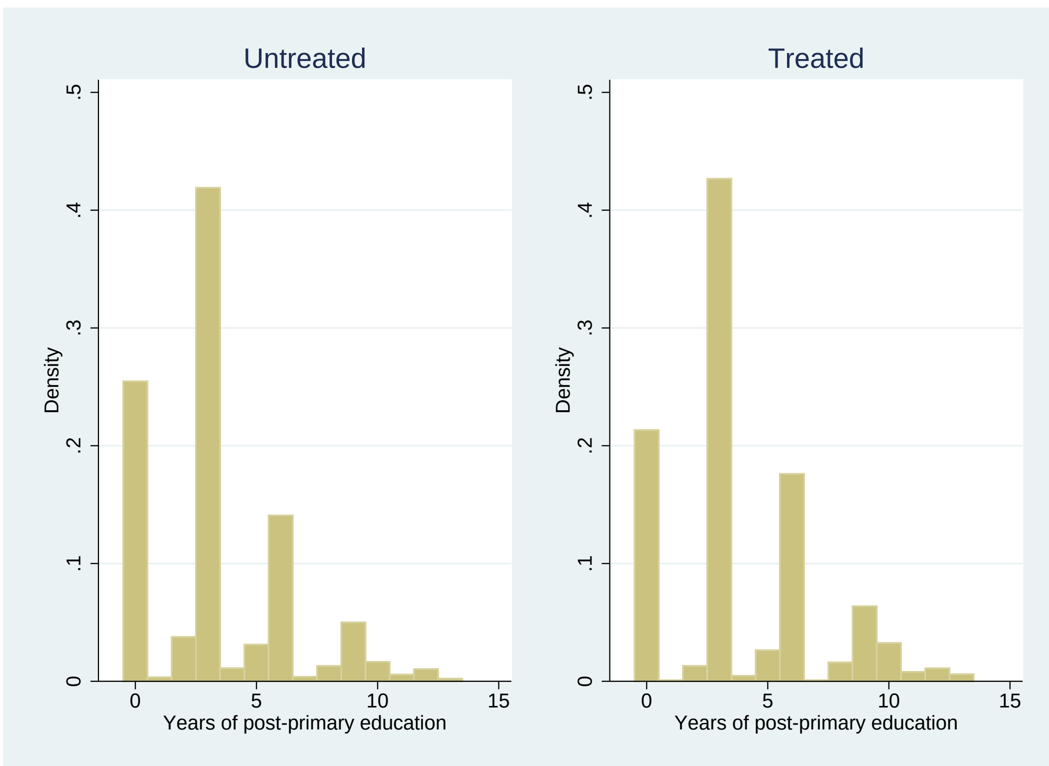


Figure 10: Distribution of post-primary education by treatment



This histogram includes only data from non-migrants, counties with clarity = 1 or 2, and individuals graduating from primary school within five years of the policy implementation year

Figure 11: Educational attainment before and after policy implementation

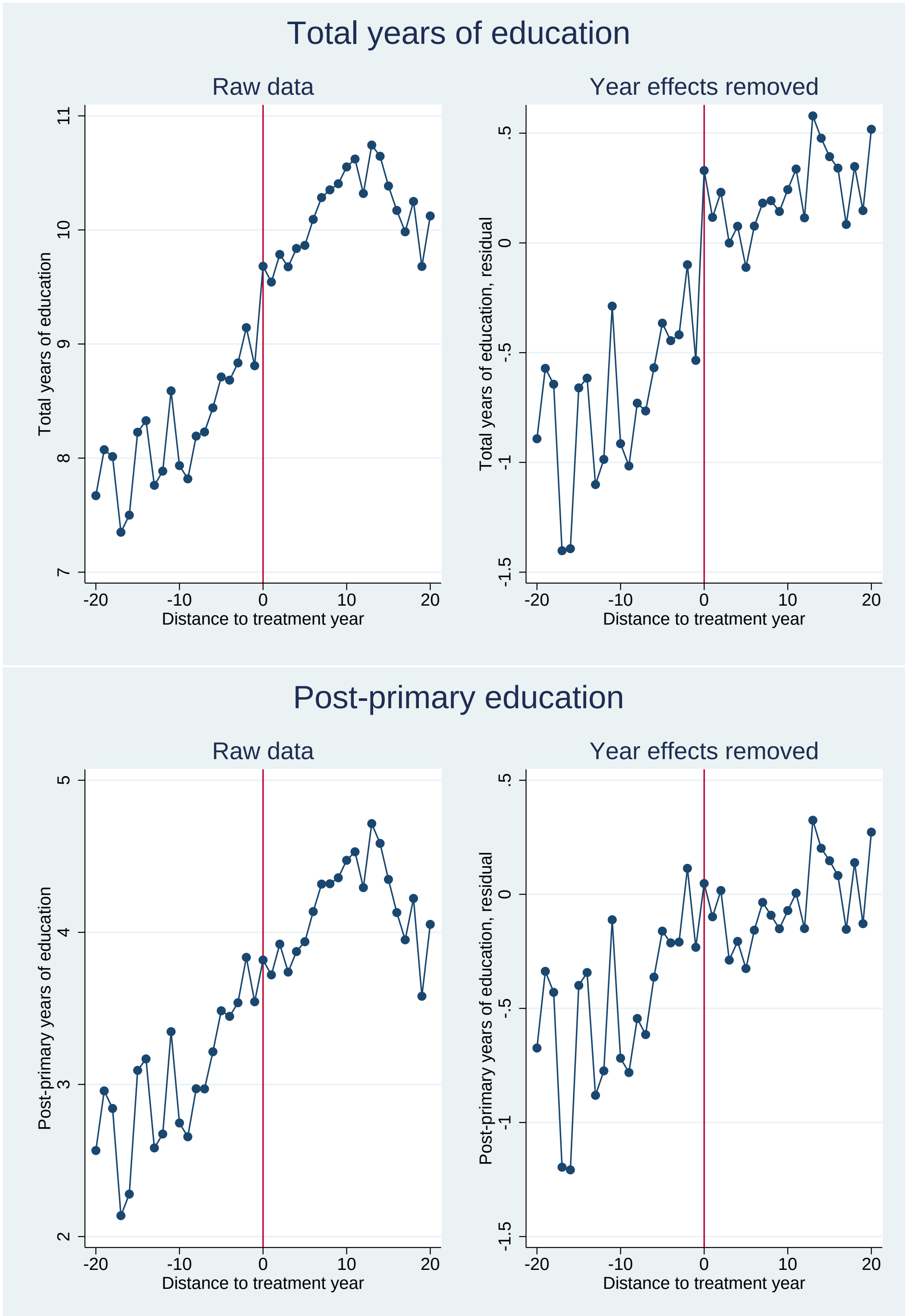
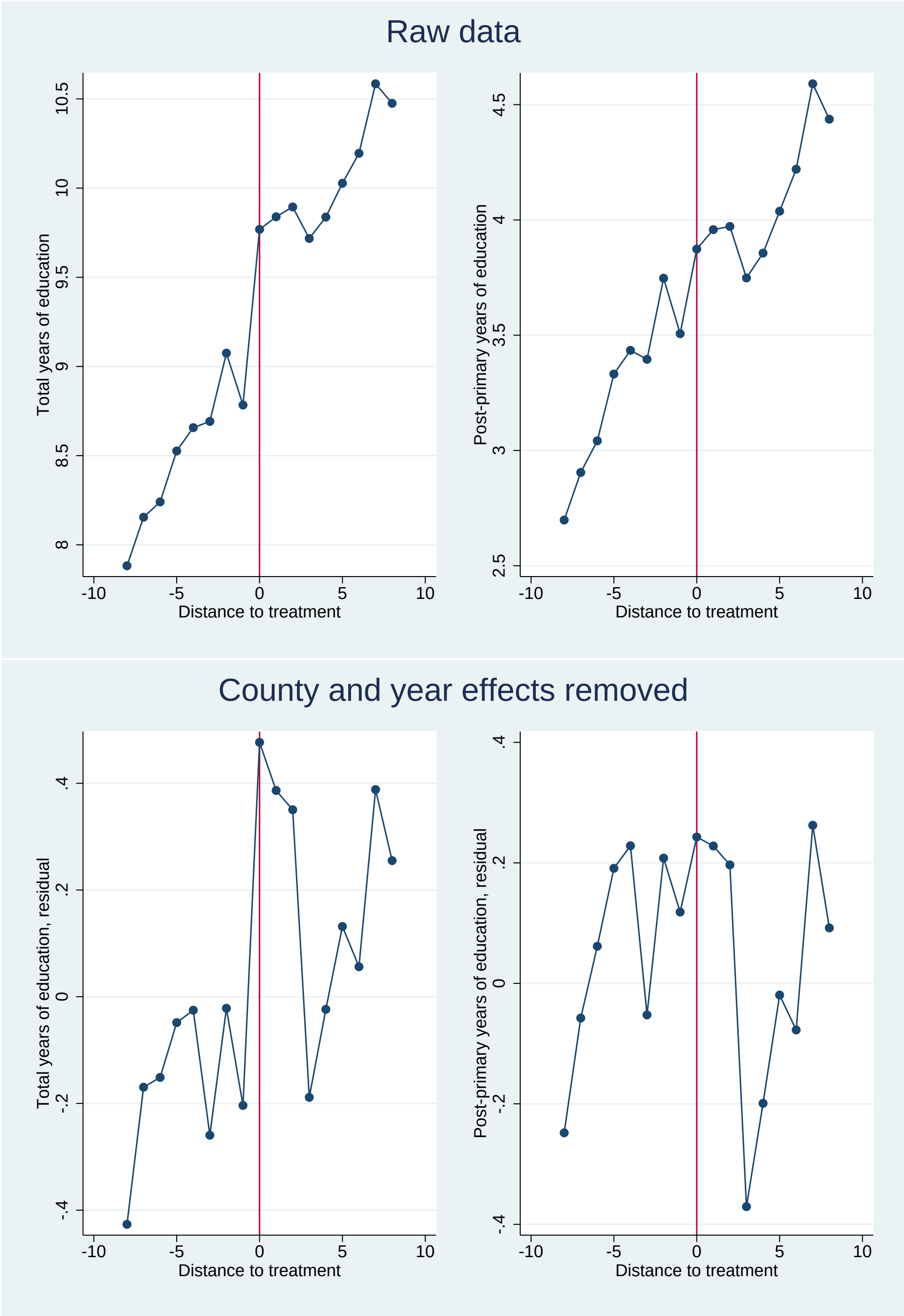


Figure 12: Balanced panel



These graphs include only observations from those 60 counties which have data in each of the 17 distance to treatment year bins between -8 and 8.

Figure 13: Treatment effect estimates - total years of education

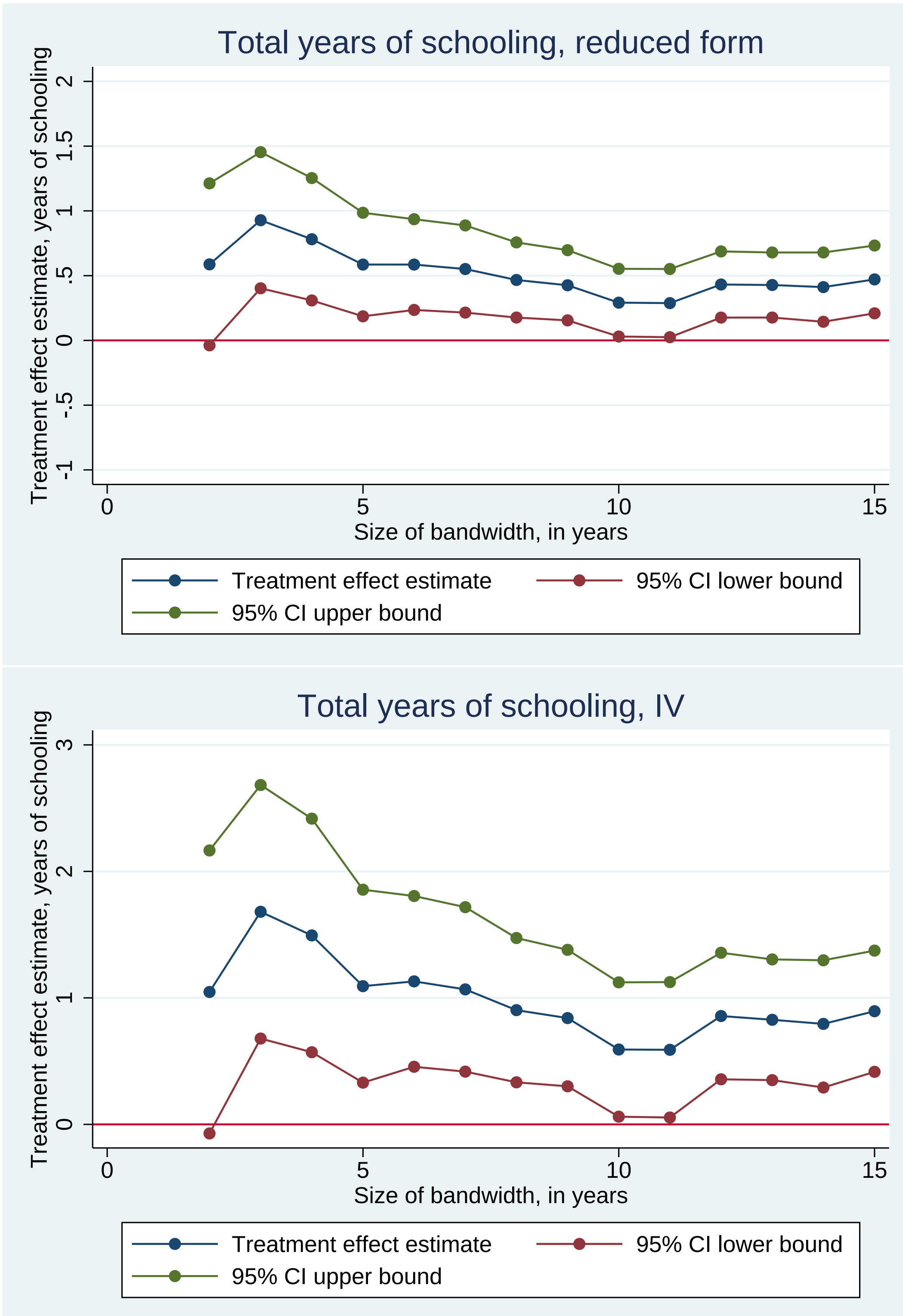
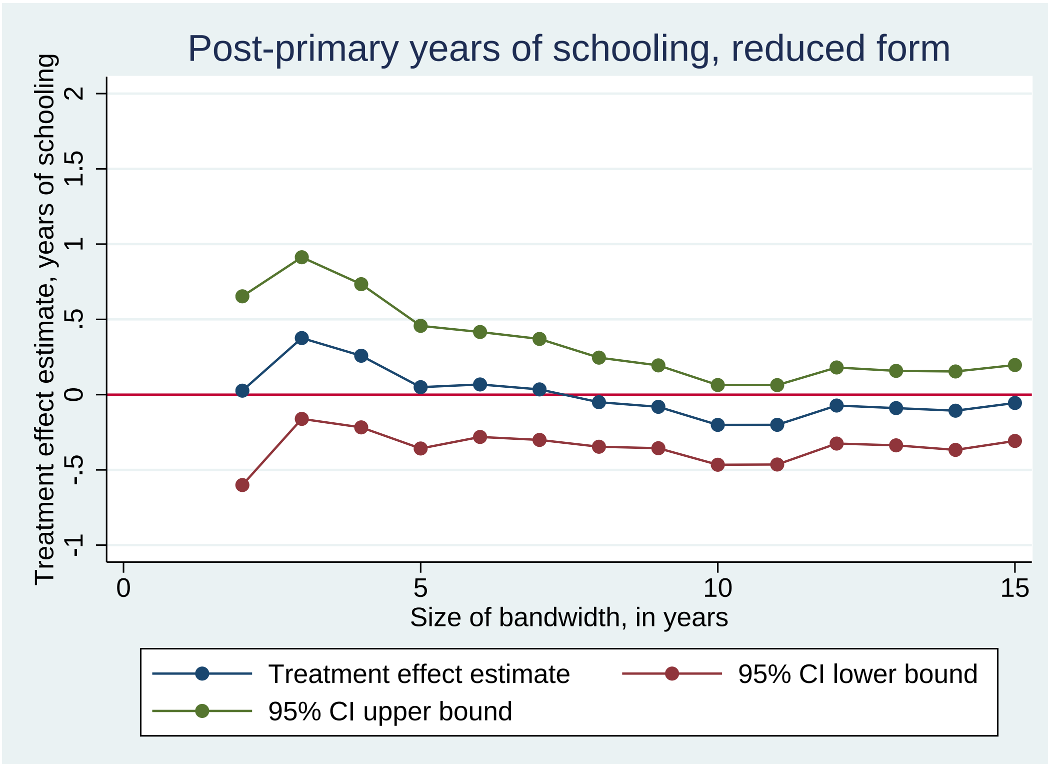


Figure 14: Treatment effect estimates - post-primary education



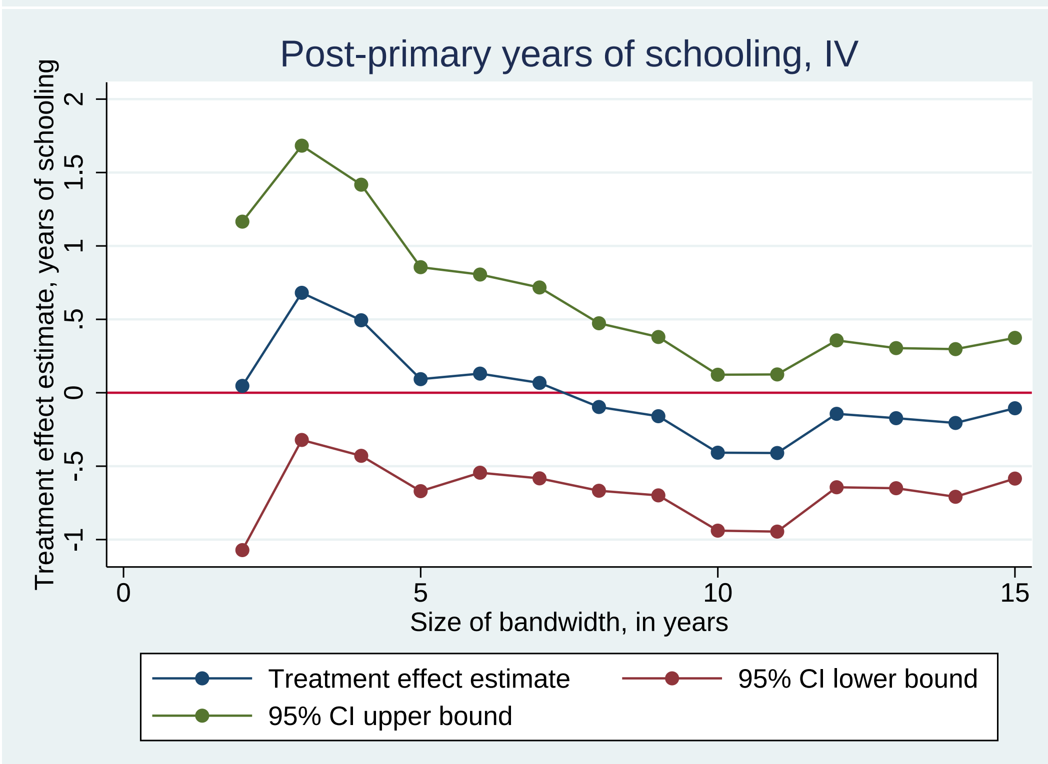


Figure 15: Graduation probability, IV, middle school

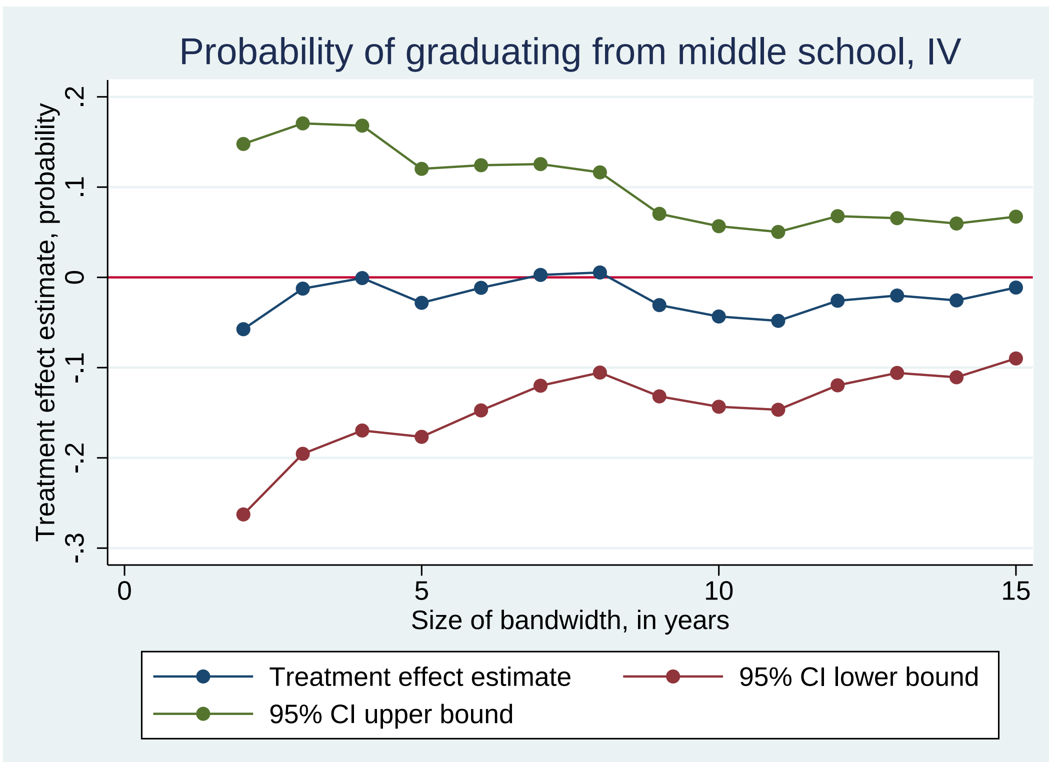


Figure 16: Graduation probability, IV, high school

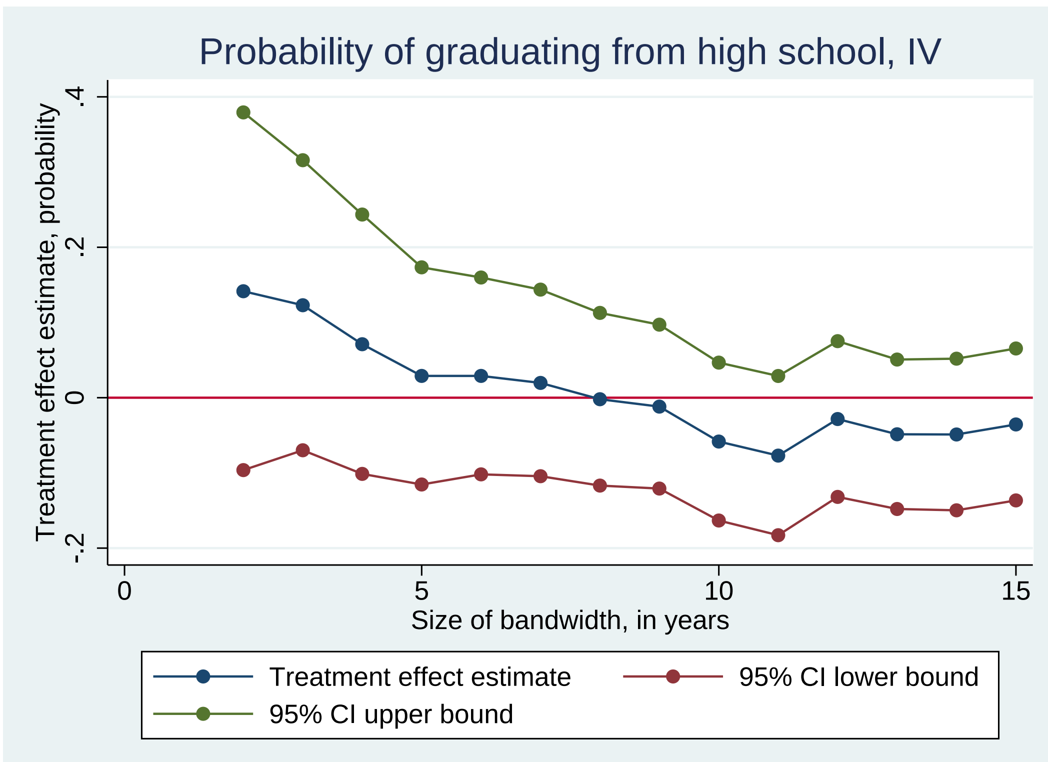


Figure 17: Graduation probability, IV, post-secondary school

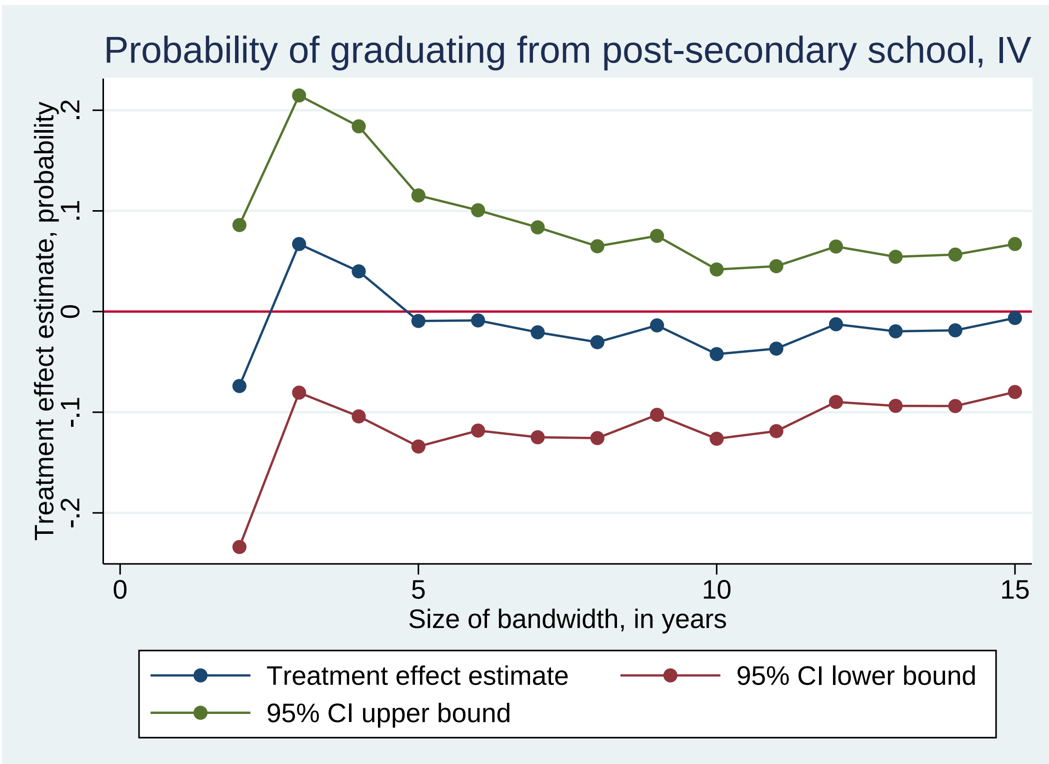


Figure 18: IV results for ever having attended a given level of school

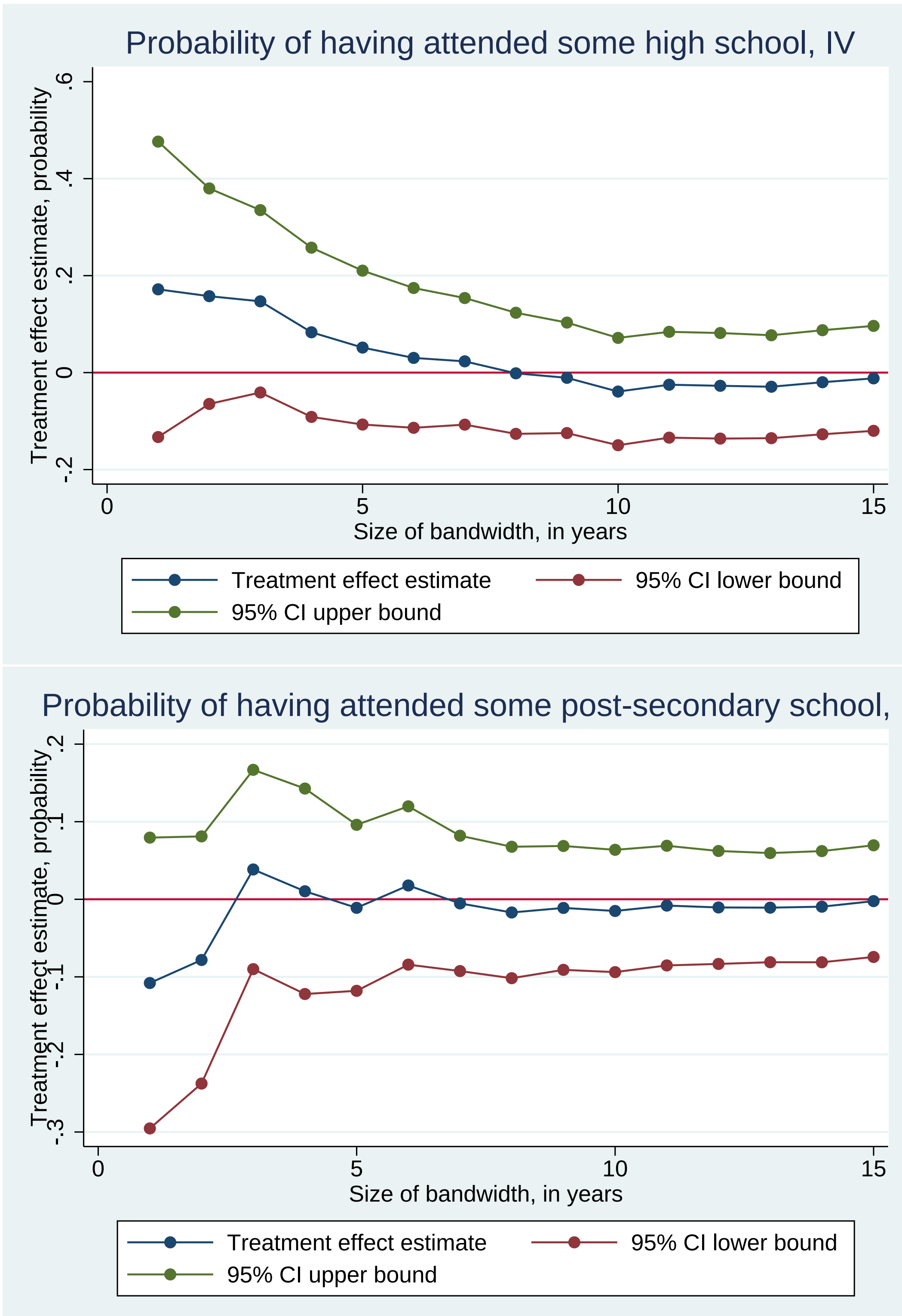


Figure 19: Investigating heterogeneous treatment effects: gender

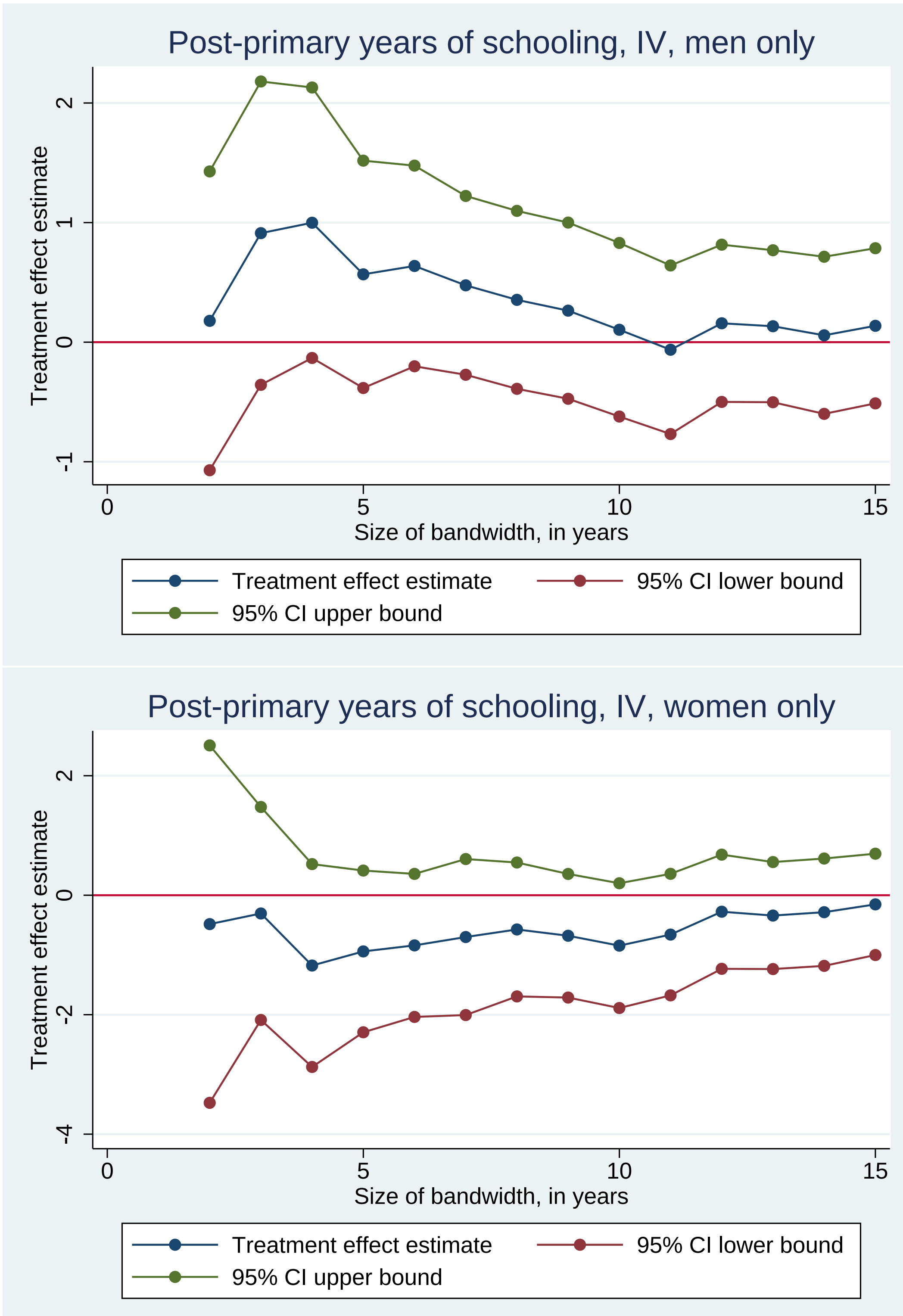


Figure 20: Investigating heterogeneous treatment effects: rural/urban residence

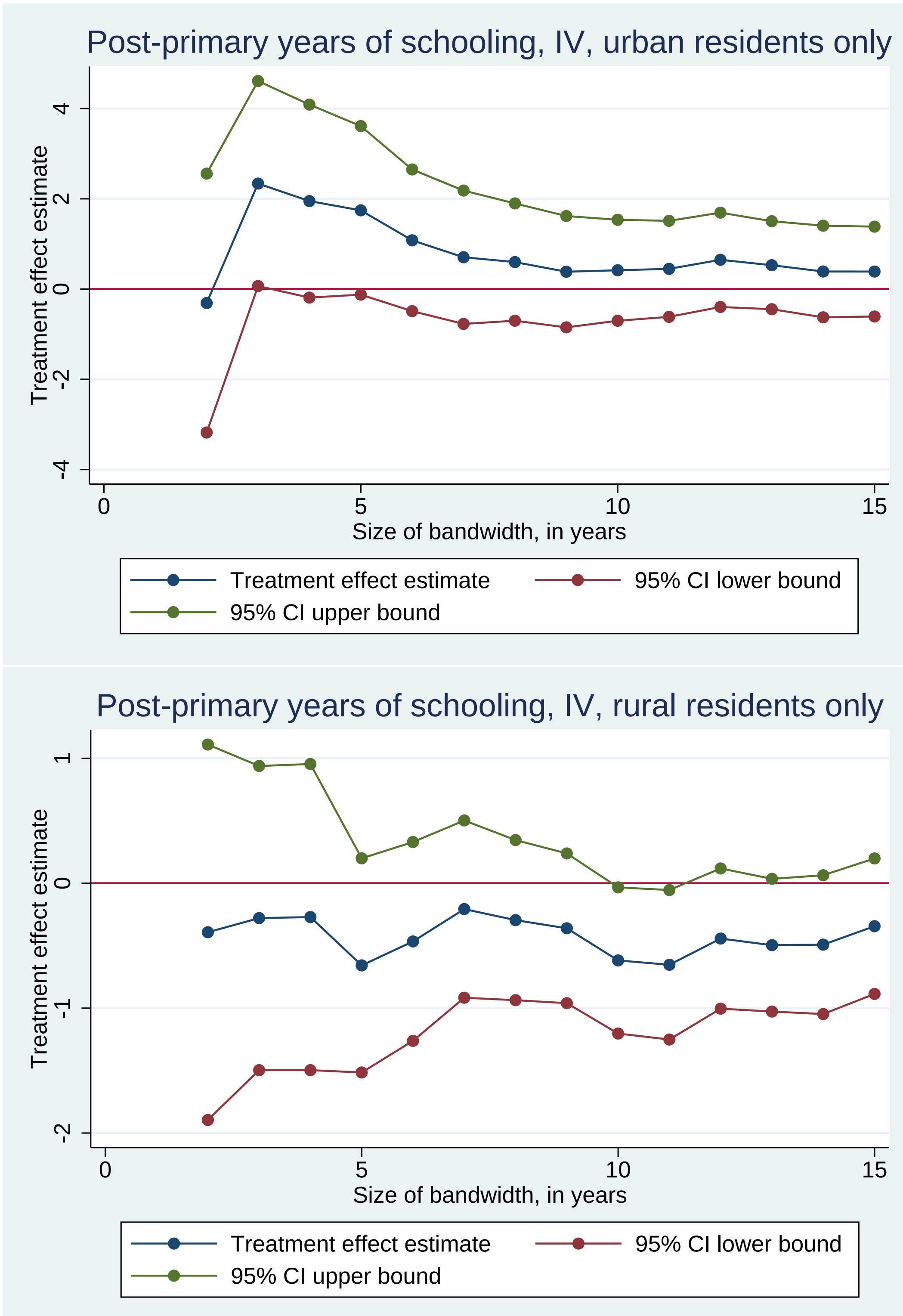


Figure 21: Investigating heterogeneous treatment effects: western/non-western provinces

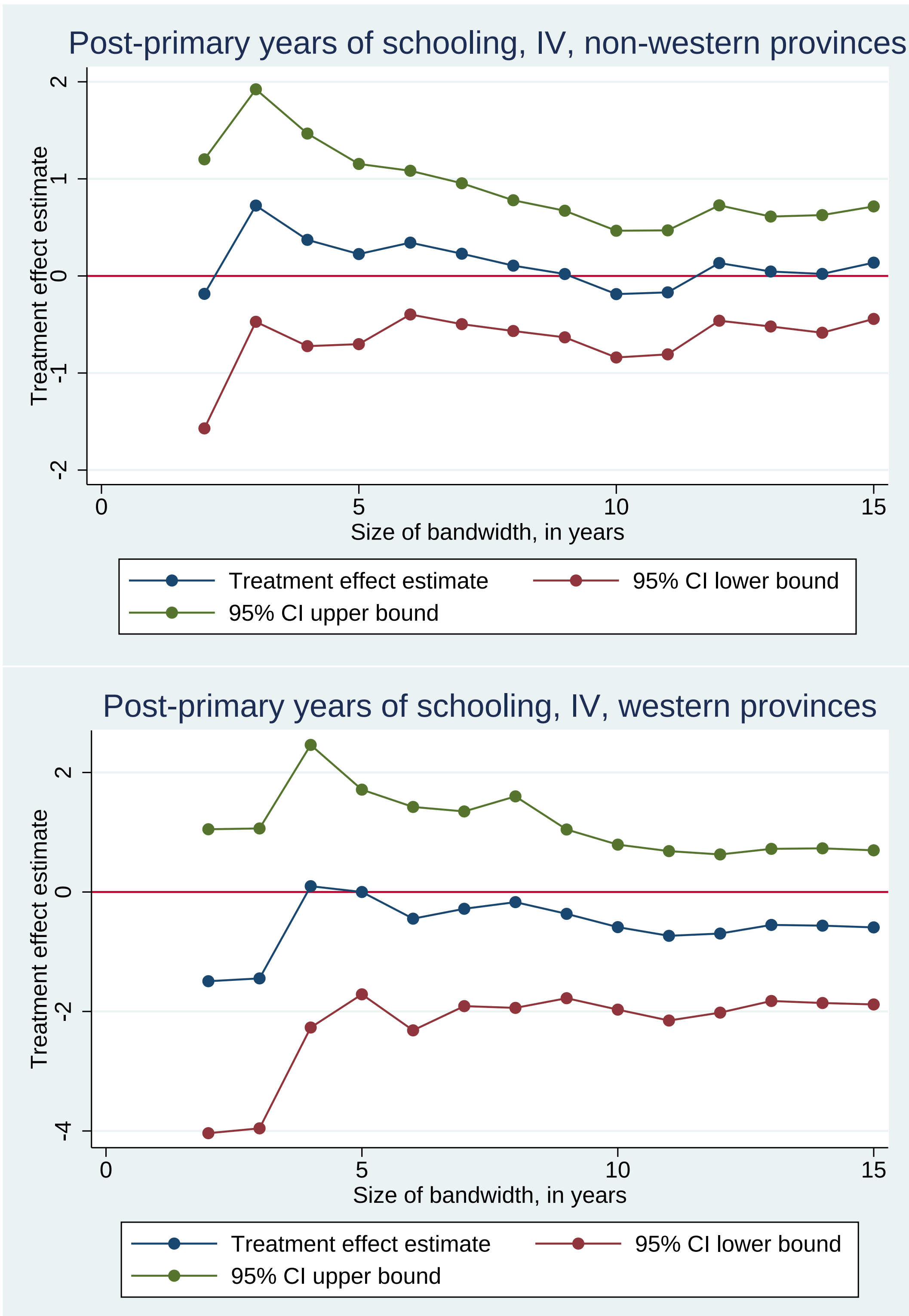
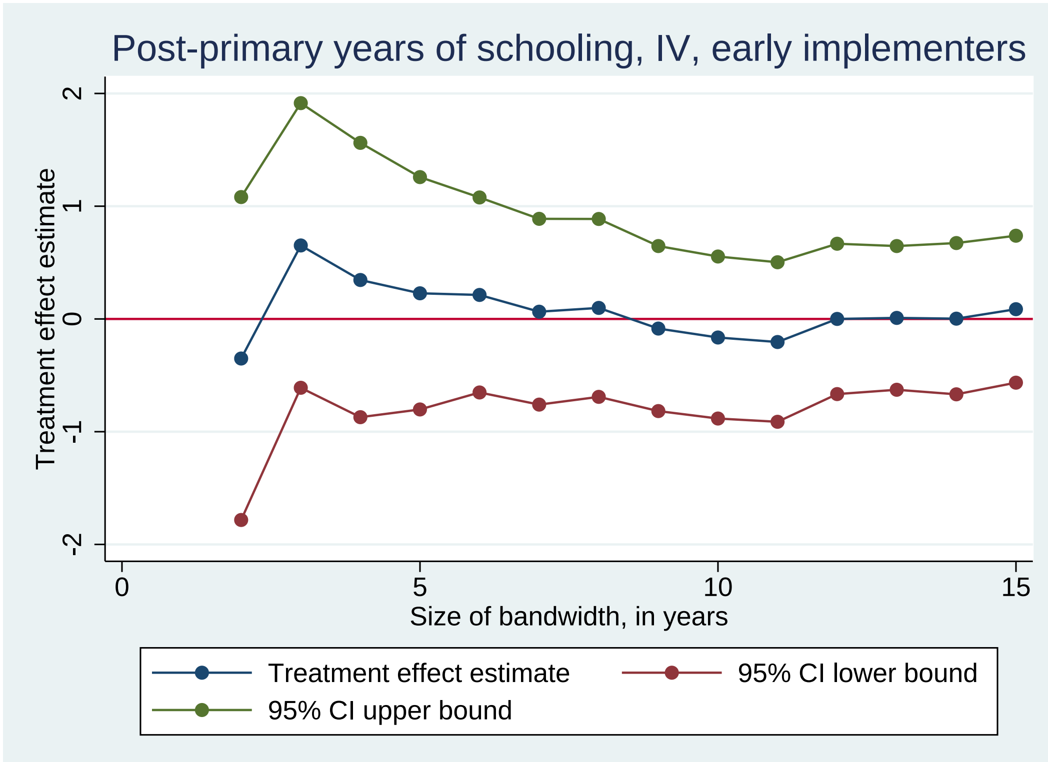


Figure 22: Investigating heterogeneous treatment effects: early and late implementers



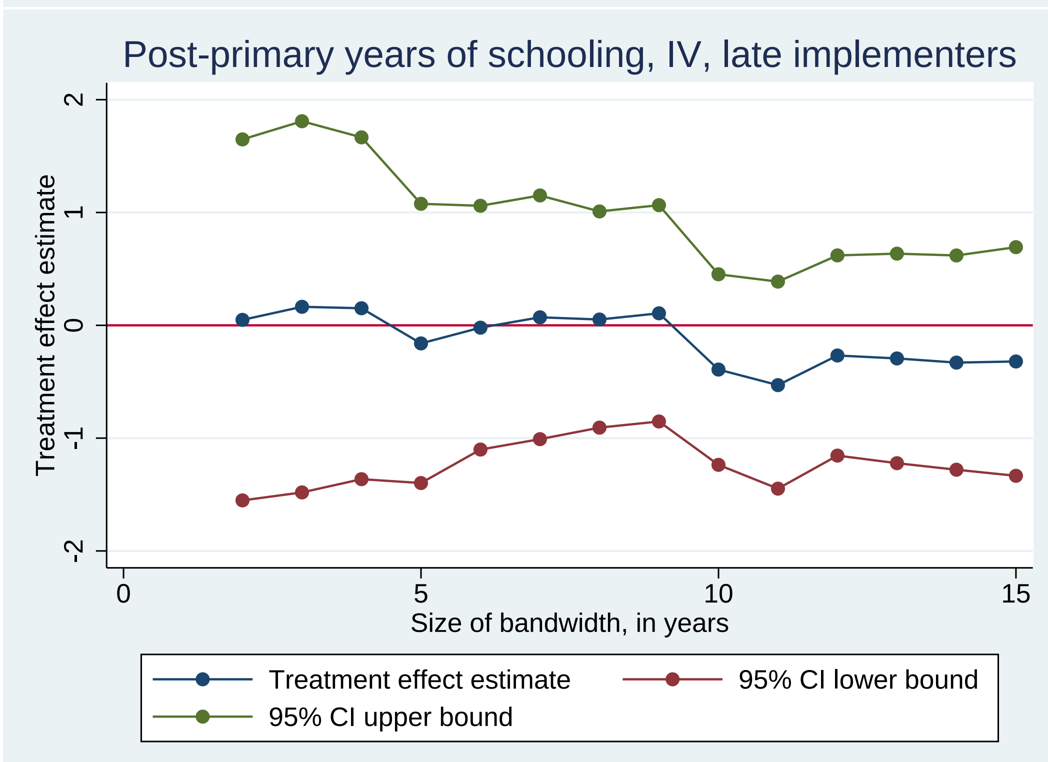


Figure 23: Mean years of education over time and completion of middle school

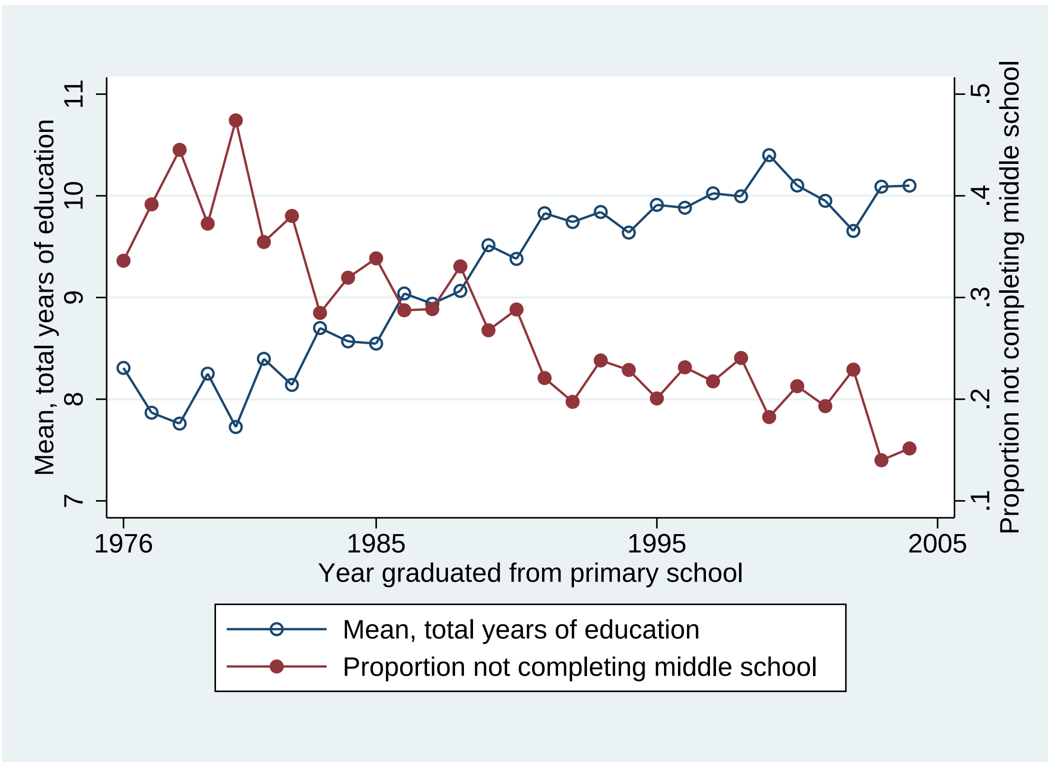


Figure 24: Proportion of students advancing between school levels, national data

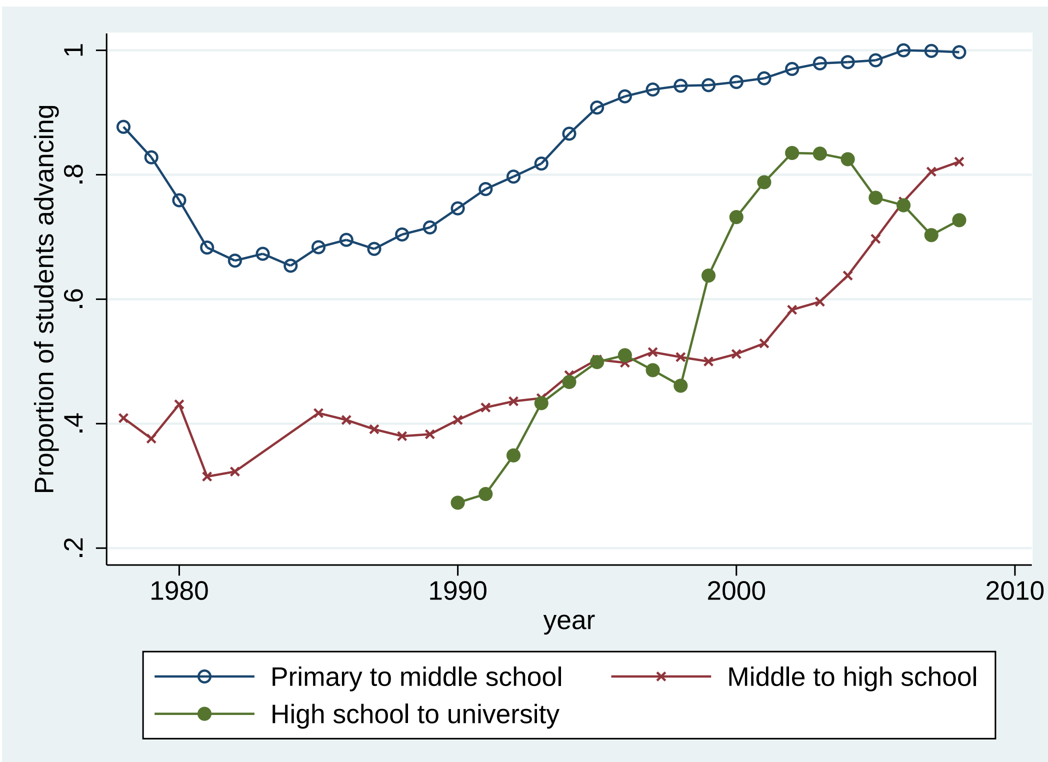
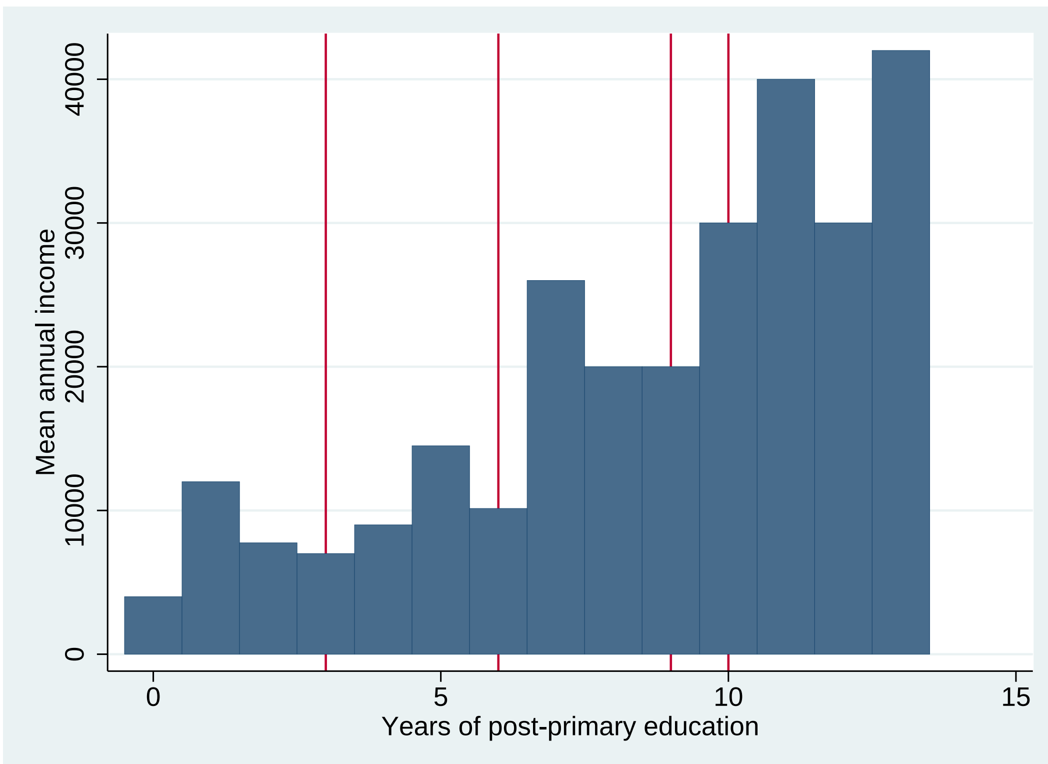


Figure 25: Non-agricultural income and educational attainment



Vertical lines are placed at degree attainment years. The first line is when a middle school degree is attained, the second is when a high school degree is attained, the third is technical school and the fourth university.

Table 1: Anecdotes on implementation of six year primary education from across China

Region	Strategy of Changing 5-Year to 6-Year Primary Education	Source
Beijing	In 1969, the length of primary education was shortened to five years. Starting from September 1st, 1980, the length of primary education was prolonged to six years, with the fifth-grade students continuing to be in the old system and students of other grades entering into the new system.	Beijing General Education Chronicle (Part I)
Xinle County, Hebei Province	In 1967, the length of primary education was shortened to five years. Starting from August 1985, first-grade students and one half of second-grade students entered into the new six year system, while the rest of students remained in the old system.	Xinle Educational Chronicle
Nanjing City, Jiangsu Province	In 1969, the length of primary education was shortened to five years. Starting from 1982, the length of primary education was prolonged to six years in urban districts and children started primary education at the age of six. By contrast, the length of primary education remained to be five years until 1999 in five other counties (Jiangning, Jiangpu, Liuhe, Gaochun, and Lishui).	Nanjing Educational Chronicle
Wuyi County, Zhejiang Province	In 1984, first-grade students entered into the new six year system in half of primary schools, with the other half following the next year. In 1987, however, all primary schools were required to resume the five year system. In September 1999, all primary schools were restored to the new system of six years. The final cohort under the old system graduated in June 2004.	Wuyi Educational Chronicle
Dongying District, Shandong Province	In 1997, the compulsory education system changed from the 5-3 (years of primary school-years of middle school) to the 5-4 system. In 2003, first-grade students entered into the new 6-3 system.	Dongying District Chronicle (1998-2005)
Xishui County, Hubei Province	In 1986, the first-grade students entered into the new 6-3 system in the primary schools located in county seats, while other primary schools remained in the old 5-3 system. In 1987, the first-grade students in the primary schools located in the township entered into the new 6-3 system. In 1991, first-grade students in the remaining primary schools entered into the new 6-3 system.	Xishui Educational Chronicle (1986-2006)

Table 2: Reasons for exclusion

Reason for exclusion	Number of counties
Shanghai (extended middle school instead)	18
No variation - consistently 5 years	9
No variation - consistently 6 years	12
No clear pattern	9
Too few observations	2
Total	50

Table 3: Village/community level characteristics

	(1) Non compliers	(2) Early implementers	(3) Late implementers
Western province	0.0625	0.286	0.0286
	(0.246)	(0.455)	(0.169)
Number of primary schools	0.626	0.771	0.910
	(0.405)	(0.504)	(0.351)
Number of medical centers	1.919	1.921	2.058
	(1.217)	(1.280)	(1.348)
Number of small stores	27.70	35.31	17.85
	(62.08)	(47.12)	(17.55)
Number of residents	3816.5	4893.3	3310.4
	(2897.9)	(3795.7)	(2427.6)
Household size	3.414	3.660	4.710
	(0.739)	(0.943)	(2.495)
Year electricity connected	1974.8	1978.5	1978.1
	(10.95)	(10.05)	(7.615)
Year first major road built	1990.6	1985.9	1985.6
	(11.60)	(13.06)	(13.33)
Year running water provided	1996.0	1993.6	1995.7
	(9.656)	(9.749)	(11.79)
Proportion of households on welfare	0.0938	0.0971	0.121
	(0.0694)	(0.0922)	(0.0833)
Observations	32	77	35

mean coefficients; sd in parentheses

Table 4: Individual-level characteristics

	(1) Non compliers	(2) Early implementers	(3) Late implementers
Western province	0.0596	0.246	0.0409
	(0.237)	(0.431)	(0.198)
Proportion female	0.497	0.482	0.454
	(0.500)	(0.500)	(0.498)
Proportion in urban area	0.333	0.422	0.213
	(0.472)	(0.494)	(0.409)
Age at time of survey	33.39	33.83	33.30
	(9.565)	(8.840)	(9.288)
Years of education	9.356	9.437	8.309
	(2.955)	(3.096)	(2.683)
Years of post-primary education	3.861	3.813	3.087
	(2.917)	(2.994)	(2.596)
Graduated primary school	0.760	0.809	0.820
	(0.427)	(0.393)	(0.384)
Graduated middle school	0.735	0.726	0.673
	(0.441)	(0.446)	(0.469)
Years of middle school	2.479	2.351	2.224
	(1.260)	(1.290)	(1.415)
Graduated high school	0.321	0.336	0.230
	(0.467)	(0.472)	(0.421)
Years of high school	1.001	1.019	0.722
	(1.457)	(1.429)	(1.324)
Graduated university	0.0365	0.0293	0.0130
	(0.187)	(0.169)	(0.113)
Year of policy	.	1986.9	2001.8
	(.)	(3.241)	(3.001)
Observations	2331	5846	2300

mean coefficients; sd in parentheses

Table 5: Hazard model results - predicting time of policy implementation

	(1)
VARIABLES	Year of policy implementation
Western province	1.272***
	(0.329)
Number of primary schools	-0.783**
	(0.319)
Number of medical centers	-0.147
	(0.117)
Number of small stores	-0.00587
	(0.00389)
Population, in 1000's	0.157***
	(0.0462)
Household size	-0.183**
	(0.0721)
Year electricity connected	0.00628
	(0.0140)
Year first major road built	-0.00107
	(0.0101)
Year running water provided	-0.0167
	(0.0111)
(mean) provcd	-0.00405
	(0.0103)
Observations	1,144

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note that a positive coefficient here signifies increased hazard and thus earlier policy implementation

Table 6: Baseline characteristics of treated and untreated individuals

	(1) Untreated, Entire Sample	(2) Treated, Entire Sample	(3) Untreated, 5 year band	(4) Treated, 5 year band
Female	0.454	0.497	0.468	0.474
	(0.498)	(0.500)	(0.499)	(0.499)
Urban resident	0.321	0.410	0.352	0.418
	(0.467)	(0.492)	(0.478)	(0.493)
Urban household registry	0.280	0.343	0.323	0.363
	(0.449)	(0.475)	(0.468)	(0.481)
Ethnic minority	0.0468	0.105	0.0605	0.0898
	(0.211)	(0.306)	(0.239)	(0.286)
Observations	4339	3807	1142	1441

mean coefficients; sd in parentheses

Table 7: Outcomes for treated and untreated individuals

	(1) Untreated, Entire Sample	(2) Treated, Entire Sample	(3) Untreated, 5 year band	(4) Treated, 5 year band
Total years of education	8.270	10.20	8.875	9.769
	(2.747)	(2.987)	(2.935)	(3.018)
Years of post-primary education	3.062	4.217	3.597	3.876
	(2.703)	(2.973)	(2.913)	(3.000)
Graduated middle school	0.723	0.824	0.771	0.793
	(0.448)	(0.381)	(0.421)	(0.405)
Graduated high school	0.251	0.432	0.318	0.357
	(0.433)	(0.495)	(0.466)	(0.479)
Graduated post-secondary school	0.0931	0.158	0.128	0.149
	(0.291)	(0.365)	(0.334)	(0.356)
Ever attended high school	0.230	0.398	0.292	0.325
	(0.421)	(0.489)	(0.455)	(0.469)
Ever attended college/tech school	0.0537	0.118	0.0753	0.0930
	(0.225)	(0.322)	(0.264)	(0.291)
Income, rough estimate	11032.6	10562.7	10847.9	12311.4
	(11573.2)	(12634.8)	(11487.3)	(13336.7)
Observations	4339	3807	1142	1441

mean coefficients; sd in parentheses

Table 8: “Naive” regressions, first stage

First stage results

Dependent variable: probability of completing 6 years of primary education

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	Naive	Controling for distance to treatment	Excluding unclear	Year fixed effects	Year + county fixed effects	With quadratic and cubic distance to treatment	As in previous column plus controls
Treatment	0.671***	0.534***	0.583***	0.510***	0.507***	0.464***	0.461***
	(0.0145)	(0.0230)	(0.0240)	(0.0244)	(0.0248)	(0.0298)	(0.0311)
Time since treatment year		0.00757***	0.00641***	0.00805***	0.00721***	0.0127***	0.0117***
		(0.000879)	(0.00108)	(0.00120)	(0.000822)	(0.00205)	(0.00242)
Time since treatment year, squared						7.32e-05	6.95e-05
						(6.74e-05)	(6.40e-05)
Time since treatment year, cubic						-5.92e-06*	-5.94e-06*
						(3.42e-06)	(3.43e-06)
Female							-0.0170*
							(0.00903)
Urban resident							-0.0226
							(0.0149)
Urban household registry							-0.00817
							(0.0147)
Ethnic minority							0.0213
							(0.0204)
Western province							0.0544***
							(0.0154)
Late implementer							0.0129
							(0.0240)
Constant	0.236***	0.311***	0.286***	0.344***	0.334***	0.353***	0.277***
	(0.00936)	(0.0131)	(0.0142)	(0.0322)	(0.0348)	(0.0359)	(0.0235)
Observations	8,146	8,146	6,321	8,146	8,146	8,146	7,940
R-squared	0.453	0.463	0.502	0.466	0.486	0.487	0.488

Robust standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 9: “Naive” regressions, reduced form, total years of education

Reduced form results
Dependent variable: total years of education attained

VARIABLES	(1) Naive	(2) Controling for distance to treatment	(3) Excluding unclear	(4) Year fixed effects	(5) Year + county fixed effects	(6) With quadratic and cubic distance to treatment	(7) As in previous column plus controls
Treatment	1.926***	0.904***	1.029***	0.780***	0.666***	0.464***	0.403***
	(0.154)	(0.142)	(0.163)	(0.138)	(0.131)	(0.146)	(0.134)
Time since treatment year		0.0562***	0.0570***	0.0571***	0.145***	0.172***	0.0995***
		(0.00774)	(0.00884)	(0.0143)	(0.00476)	(0.0106)	(0.0107)
Time since treatment year, squared						-0.000605	-0.000284
						(0.000406)	(0.000365)
Time since treatment year, cubic						-5.67e-05***	-6.37e-05***
						(2.05e-05)	(1.90e-05)
Female							-0.389***
							(0.0648)
Urban resident							0.550***
							(0.136)
Urban household registry							2.570***
							(0.133)
Ethnic minority							-0.0862
							(0.182)
Western province							1.047***
							(0.105)
Late implementer							1.122***
							(0.151)
Constant	8.270***	8.832***	8.816***	8.916***	9.778***	9.919***	7.385***
	(0.113)	(0.125)	(0.142)	(0.305)	(0.240)	(0.245)	(0.141)
Observations	8,146	8,146	6,321	8,146	8,146	8,146	7,940
R-squared	0.101	0.116	0.123	0.122	0.279	0.280	0.403

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 10: “Naive” regressions, reduced form, post-primary years of education

Reduced form results
Dependent variable: years of post-primary education attained

VARIABLES	(1) Naive	(2) Controling for distance to treatment	(3) Excluding unclear	(4) Year fixed effects	(5) Year + county fixed effects	(6) With quadratic and cubic distance to treatment	(7) As in previous column plus controls
Treatment	1.155***	0.305**	0.387**	0.208	0.0900	-0.0824	-0.136
	(0.152)	(0.137)	(0.159)	(0.134)	(0.128)	(0.142)	(0.128)
Time since treatment year		0.0467***	0.0486***	0.0474***	0.137***	0.160***	0.0863***
		(0.00789)	(0.00895)	(0.0144)	(0.00460)	(0.0102)	(0.0103)
Time since treatment year, squared						-0.000688*	-0.000339
						(0.000400)	(0.000356)
Time since treatment year, cubic						-5.35e-05***	-5.98e-05***
						(2.02e-05)	(1.85e-05)
Female							-0.345***
							(0.0617)
Urban resident							0.600***
							(0.138)
Urban household registry							2.580***
							(0.132)
Ethnic minority							-0.123
							(0.182)
Western province							1.007***
							(0.104)
Late implementer							1.086***
							(0.144)
Constant	3.062***	3.529***	3.535***	3.568***	4.413***	4.542***	2.058***
	(0.113)	(0.126)	(0.142)	(0.306)	(0.234)	(0.236)	(0.139)
Observations	8,146	8,146	6,321	8,146	8,146	8,146	7,940
R-squared	0.040	0.051	0.055	0.057	0.231	0.231	0.367

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 11: “Naive regressions”, instrumental variables, total years of education

IV results
Dependent variable: total years of education attained

VARIABLES	(1) Naive	(2) Controling for distance to treatment	(3) Excluding unclear	(4) Year fixed effects	(5) Year + county fixed effects	(6) With quadratic and cubic distance to treatment	(7) As in previous column plus controls
Had six years of primary school	2.869***	1.694***	1.766***	1.529***	1.315***	0.999***	0.874***
	(0.226)	(0.255)	(0.270)	(0.263)	(0.256)	(0.310)	(0.281)
Time since treatment year		0.0434***	0.0456***	0.0448***	0.0515***	0.0734***	0.0892***
		(0.00883)	(0.00982)	(0.0144)	(0.00445)	(0.0122)	(0.0125)
Time since treatment year, squared						-0.000678*	-0.000345
						(0.000405)	(0.000363)
Time since treatment year, cubic							-5.85e-05***
							(1.91e-05)
Female							-0.374***
							(0.0643)
Urban resident							0.570***
							(0.136)
Urban household registry							2.577***
							(0.132)
Ethnic minority							-0.105
							(0.185)
Western province							0.448**
							(0.185)
Late implementer							1.452***
							(0.149)
Constant	7.594***	8.304***	8.310***	8.542***	8.237***	8.408***	6.801***
	(0.138)	(0.170)	(0.187)	(0.290)	(0.257)	(0.260)	(0.225)
Observations	8,146	8,146	6,321	8,146	8,146	8,146	7,940
R-squared	0.060	0.115	0.123	0.125	0.286	0.290	0.414

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 12: “Naive regressions”, instrumental variables, post-primary years of education

IV results

Dependent variable: post-primary years of education attained

VARIABLES	(1) Naive	(2) Controling for distance to treatment	(3) Excluding unclear	(4) Year fixed effects	(5) Year + county fixed effects	(6) With quadratic and cubic distance to treatment	(7) As in previous column plus controls
Had six years of primary school	1.721***	0.572**	0.664**	0.407	0.178	-0.177	-0.296
	(0.226)	(0.255)	(0.271)	(0.262)	(0.253)	(0.306)	(0.280)
Time since treatment year		0.0424***	0.0444***	0.0442***	0.0540***	0.0785***	0.0897***
		(0.00891)	(0.0100)	(0.0145)	(0.00440)	(0.0120)	(0.0124)
Time since treatment year, squared						-0.000675*	-0.000319
						(0.000404)	(0.000359)
Time since treatment year, cubic						-5.45e-05***	-6.16e-05***
						(2.06e-05)	(1.89e-05)
Female							-0.350***
							(0.0622)
Urban resident							0.593***
							(0.138)
Urban household registry							2.577***
							(0.132)
Ethnic minority							-0.116
							(0.180)
Western province							0.488***
							(0.181)
Late implementer							1.367***
							(0.143)
Constant	2.657***	3.351***	3.344***	3.558***	3.244***	3.424***	1.863***
	(0.136)	(0.169)	(0.185)	(0.289)	(0.255)	(0.259)	(0.223)
Observations	8,146	8,146	6,321	8,146	8,146	8,146	7,940
R-squared		0.042	0.046	0.051	0.229	0.232	0.367

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 13: Restricted bandwidth regressions, reduced form, total years of education

	Dependent variable: total years of schooling						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	3 year	5 year	7 year	9 year	11 year	13 year	15 year
VARIABLES	bandwidth	bandwidth	bandwidth	bandwidth	bandwidth	bandwidth	bandwidth
Treatment	0.928***	0.586***	0.551***	0.426***	0.288**	0.428***	0.471***
	(0.268)	(0.204)	(0.172)	(0.138)	(0.134)	(0.128)	(0.134)
Time since treatment year	0.00534	0.0674	0.0624*	0.0953***	0.113***	0.101***	0.0951***
	(0.0912)	(0.0535)	(0.0340)	(0.0220)	(0.0163)	(0.0145)	(0.0111)
Female	-0.416***	-0.350***	-0.380***	-0.387***	-0.367***	-0.357***	-0.360***
	(0.118)	(0.0958)	(0.0895)	(0.0801)	(0.0694)	(0.0636)	(0.0681)
Urban resident	0.513*	0.478*	0.547**	0.510***	0.443***	0.446***	0.493***
	(0.300)	(0.276)	(0.217)	(0.182)	(0.169)	(0.159)	(0.150)
Urban household registry	2.931***	3.093***	2.925***	2.962***	2.934***	2.814***	2.719***
	(0.279)	(0.204)	(0.208)	(0.171)	(0.165)	(0.156)	(0.150)
Ethnic minority	-0.768	-0.320	-0.259	-0.117	-0.136	-0.139	-0.116
	(0.469)	(0.333)	(0.296)	(0.265)	(0.236)	(0.213)	(0.204)
Western province	-2.930***	-0.832**	0.0428	0.975***	-0.357***	0.913***	1.371***
	(0.611)	(0.338)	(0.209)	(0.256)	(0.125)	(0.139)	(0.113)
Late implementer	0.542	-1.782***	0.0359	1.415***	0.446**	1.463***	1.640***
	(0.642)	(0.670)	(0.262)	(0.250)	(0.222)	(0.190)	(0.159)
Constant	11.21***	8.701***	7.697***	6.526***	8.077***	7.250***	7.002***
	(1.094)	(0.466)	(0.418)	(0.297)	(0.362)	(0.198)	(0.170)
Observations	1,493	2,506	3,474	4,347	5,189	5,901	6,508
R-squared	0.439	0.418	0.409	0.404	0.407	0.403	0.401

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 14: Restricted bandwidth regressions, reduced form, post-primary years of education

	Dependent variable: years of post-primary education attained						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	3 year bandwidth	5 year bandwidth	7 year bandwidth	9 year bandwidth	11 year bandwidth	13 year bandwidth	15 year bandwidth
Treatment	0.376	0.0497	0.0347	-0.0808	-0.200	-0.0895	-0.0554
	(0.274)	(0.208)	(0.171)	(0.140)	(0.135)	(0.126)	(0.129)
Time since treatment year	-0.0107	0.0642	0.0557*	0.0825***	0.0942***	0.0865***	0.0817***
	(0.0865)	(0.0465)	(0.0325)	(0.0214)	(0.0161)	(0.0141)	(0.0104)
Female	-0.305**	-0.302***	-0.334***	-0.346***	-0.319***	-0.315***	-0.320***
	(0.117)	(0.0940)	(0.0865)	(0.0769)	(0.0673)	(0.0605)	(0.0648)
Urban resident	0.573*	0.540**	0.595***	0.553***	0.482***	0.491***	0.539***
	(0.294)	(0.266)	(0.216)	(0.183)	(0.169)	(0.159)	(0.152)
Urban household registry	2.933***	3.075***	2.920***	2.958***	2.934***	2.811***	2.717***
	(0.278)	(0.202)	(0.206)	(0.167)	(0.160)	(0.152)	(0.147)
Ethnic minority	-0.750	-0.309	-0.285	-0.129	-0.177	-0.173	-0.153
	(0.479)	(0.333)	(0.291)	(0.262)	(0.232)	(0.212)	(0.203)
Western province	-2.455***	-0.938***	-0.247	1.018***	-0.576***	1.024***	1.402***
	(0.635)	(0.336)	(0.191)	(0.255)	(0.120)	(0.138)	(0.110)
Late implementer	0.745	-1.610***	0.00496	1.631***	0.359	1.690***	1.756***
	(0.663)	(0.548)	(0.263)	(0.228)	(0.222)	(0.184)	(0.150)
Constant	5.027***	3.191***	2.490***	1.067***	2.888***	1.714***	1.568***
	(1.147)	(0.426)	(0.397)	(0.282)	(0.345)	(0.197)	(0.168)
Observations	1,493	2,506	3,474	4,347	5,189	5,901	6,508
R-squared	0.435	0.411	0.397	0.386	0.383	0.375	0.371

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 15: Restricted bandwidth regressions, instrumental variables, total years of education

	Dependent variable: total years of schooling						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	3 year bandwidth	5 year bandwidth	7 year bandwidth	9 year bandwidth	11 year bandwidth	13 year bandwidth	15 year bandwidth
Had six years of primary school	1.963***	1.279***	1.252***	0.969***	0.670**	0.968***	1.052***
	(0.595)	(0.454)	(0.389)	(0.316)	(0.310)	(0.284)	(0.288)
Time since treatment year	0.00124	0.0655	0.0541	0.0847***	0.102***	0.0877***	0.0825***
	(0.0820)	(0.0457)	(0.0342)	(0.0235)	(0.0190)	(0.0162)	(0.0127)
Female	-0.294**	-0.323***	-0.361***	-0.374***	-0.355***	-0.344***	-0.346***
	(0.119)	(0.0941)	(0.0880)	(0.0794)	(0.0687)	(0.0623)	(0.0668)
Urban resident	0.636**	0.522**	0.577***	0.531***	0.460***	0.472***	0.515***
	(0.294)	(0.263)	(0.213)	(0.180)	(0.168)	(0.157)	(0.150)
Urban household registry	2.879***	3.065***	2.920***	2.959***	2.928***	2.806***	2.718***
	(0.285)	(0.205)	(0.208)	(0.168)	(0.163)	(0.154)	(0.148)
Ethnic minority	-0.696	-0.300	-0.273	-0.128	-0.159	-0.166	-0.143
	(0.493)	(0.339)	(0.294)	(0.262)	(0.232)	(0.213)	(0.205)
Western province	-2.198***	1.977***	-0.301	0.998***	0.284	0.649***	1.446***
	(0.630)	(0.309)	(0.214)	(0.255)	(0.240)	(0.217)	(0.113)
Late implementer	0.815	1.691***	0.0510	1.651***	1.214***	1.705***	1.792***
	(0.663)	(0.541)	(0.266)	(0.257)	(0.225)	(0.202)	(0.166)
Constant	9.348***	4.938***	7.441***	6.103***	7.160***	6.733***	6.550***
	(1.210)	(0.544)	(0.439)	(0.365)	(0.324)	(0.295)	(0.259)
Observations	1,493	2,506	3,474	4,347	5,189	5,901	6,508
R-squared	0.427	0.428	0.417	0.416	0.418	0.415	0.413

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 16: Restricted bandwidth regressions, instrumental variables, post-primary years of education

	Dependent variable: years of post-primary education attained						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	3 year bandwidth	5 year bandwidth	7 year bandwidth	9 year bandwidth	11 year bandwidth	13 year bandwidth	15 year bandwidth
Had six years of primary school	0.795	0.109	0.0787	-0.184	-0.467	-0.202	-0.124
	(0.595)	(0.455)	(0.389)	(0.318)	(0.312)	(0.285)	(0.288)
Time since treatment year	-0.0124	0.0641	0.0552	0.0845***	0.101***	0.0893***	0.0832***
	(0.0843)	(0.0464)	(0.0344)	(0.0237)	(0.0192)	(0.0164)	(0.0125)
Female	-0.255**	-0.299***	-0.333***	-0.349***	-0.327***	-0.318***	-0.322***
	(0.119)	(0.0941)	(0.0866)	(0.0775)	(0.0675)	(0.0611)	(0.0653)
Urban resident	0.623**	0.544**	0.597***	0.549***	0.471***	0.486***	0.536***
	(0.297)	(0.264)	(0.216)	(0.183)	(0.171)	(0.160)	(0.153)
Urban household registry	2.912***	3.072***	2.920***	2.958***	2.938***	2.812***	2.717***
	(0.285)	(0.203)	(0.207)	(0.167)	(0.161)	(0.153)	(0.147)
Ethnic minority	-0.722	-0.307	-0.285	-0.127	-0.161	-0.167	-0.150
	(0.492)	(0.335)	(0.290)	(0.261)	(0.231)	(0.211)	(0.202)
Western province	-2.158***	1.932***	-0.269	1.013***	0.296	0.652***	1.393***
	(0.633)	(0.315)	(0.215)	(0.254)	(0.239)	(0.215)	(0.111)
Late implementer	0.855	1.599***	0.00591	1.586***	1.143***	1.639***	1.738***
	(0.681)	(0.552)	(0.264)	(0.257)	(0.224)	(0.201)	(0.162)
Constant	4.273***	-0.0430	2.474***	1.147***	2.208***	1.822***	1.621***
	(1.247)	(0.544)	(0.439)	(0.369)	(0.328)	(0.299)	(0.259)
Observations	1,493	2,506	3,474	4,347	5,189	5,901	6,508
R-squared	0.417	0.410	0.396	0.386	0.382	0.376	0.371

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 17: Restricted bandwidth regressions, instrumental variables, graduation from middle school

	Dependent variable: graduated middle school						
VARIABLES	(1) 3 year bandwidth	(2) 5 year bandwidth	(3) 7 year bandwidth	(4) 9 year bandwidth	(5) 11 year bandwidth	(6) 13 year bandwidth	(7) 15 year bandwidth
Had six years of primary school	-0.0125	-0.0282	0.00273	-0.0307	-0.0482	-0.0201	-0.0113
	(0.0934)	(0.0757)	(0.0626)	(0.0516)	(0.0503)	(0.0437)	(0.0401)
Time since treatment year	0.00642	0.00363	0.00596	0.00885**	0.00903**	0.00741***	0.00669***
	(0.0174)	(0.00674)	(0.00523)	(0.00395)	(0.00349)	(0.00279)	(0.00222)
Female	-0.0221	-0.0343**	-0.0424***	-0.0507***	-0.0477***	-0.0473***	-0.0483***
	(0.0196)	(0.0156)	(0.0138)	(0.0136)	(0.0124)	(0.0109)	(0.0112)
Urban resident	0.118***	0.126***	0.112***	0.103***	0.0974***	0.0971***	0.0953***
	(0.0426)	(0.0330)	(0.0265)	(0.0248)	(0.0223)	(0.0216)	(0.0209)
Urban household registry	0.221***	0.211***	0.206***	0.209***	0.211***	0.198***	0.195***
	(0.0334)	(0.0244)	(0.0231)	(0.0214)	(0.0207)	(0.0201)	(0.0195)
Ethnic minority	-0.111	-0.0536	-0.0494	-0.0216	-0.0468	-0.0522	-0.0476
	(0.0720)	(0.0501)	(0.0384)	(0.0364)	(0.0351)	(0.0359)	(0.0358)
Western province	-0.279**	0.501***	-0.0319	0.212***	0.200***	0.270***	0.258***
	(0.108)	(0.0467)	(0.0353)	(0.0364)	(0.0368)	(0.0369)	(0.0165)
Late implementer	0.0159	0.143**	-0.0609	0.252***	0.269***	0.416***	0.405***
	(0.124)	(0.0611)	(0.0416)	(0.0407)	(0.0323)	(0.0287)	(0.0271)
Constant	0.984***	0.252***	0.757***	0.528***	0.548***	0.498***	0.466***
	(0.234)	(0.0862)	(0.0603)	(0.0605)	(0.0547)	(0.0442)	(0.0362)
Observations	1,493	2,506	3,474	4,347	5,189	5,901	6,508
R-squared	0.226	0.214	0.198	0.184	0.189	0.182	0.181

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 18: Restricted bandwidth regressions, instrumental variables, graduation from high school

	Dependent variable: graduated high school						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	3 year bandwidth	5 year bandwidth	7 year bandwidth	9 year bandwidth	11 year bandwidth	13 year bandwidth	15 year bandwidth
Had six years of primary school	0.123	0.0290	0.0197	-0.0118	-0.0770	-0.0486	-0.0355
	(0.0984)	(0.0737)	(0.0633)	(0.0556)	(0.0540)	(0.0507)	(0.0515)
Time since treatment year	-0.0140	-0.00114	0.00325	0.00995**	0.0149***	0.0141***	0.0170***
	(0.0131)	(0.00740)	(0.00571)	(0.00423)	(0.00357)	(0.00322)	(0.00252)
Female	-0.0642***	-0.0615***	-0.0568***	-0.0512***	-0.0489***	-0.0444***	-0.0429***
	(0.0191)	(0.0152)	(0.0129)	(0.0116)	(0.0110)	(0.00952)	(0.00985)
Urban resident	0.0528	0.0333	0.0632*	0.0575**	0.0458	0.0468*	0.0581**
	(0.0475)	(0.0379)	(0.0330)	(0.0281)	(0.0280)	(0.0271)	(0.0253)
Urban household registry	0.453***	0.476***	0.427***	0.434***	0.437***	0.428***	0.414***
	(0.0513)	(0.0343)	(0.0332)	(0.0278)	(0.0266)	(0.0267)	(0.0253)
Ethnic minority	-0.133*	-0.0874**	-0.0515	-0.0268	-0.0208	-0.0204	-0.0199
	(0.0691)	(0.0409)	(0.0388)	(0.0344)	(0.0301)	(0.0260)	(0.0251)
Western province	-0.148	0.278***	0.0810*	0.0926**	-0.0125	0.0260	0.156***
	(0.128)	(0.0536)	(0.0427)	(0.0366)	(0.0334)	(0.0293)	(0.0207)
Late implementer	0.0613	-0.0745	-0.0130	0.0727	0.00765	0.0397	0.112***
	(0.0906)	(0.114)	(0.0364)	(0.0514)	(0.0392)	(0.0370)	(0.0287)
Constant	0.344*	-0.0624	0.124*	0.0520	0.237***	0.192***	0.154***
	(0.201)	(0.0817)	(0.0687)	(0.0649)	(0.0605)	(0.0569)	(0.0477)
Observations	1,493	2,506	3,474	4,347	5,189	5,901	6,508
R-squared	0.383	0.366	0.344	0.335	0.322	0.317	0.314

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 19: Restricted bandwidth regressions, instrumental variables, graduation from post-secondary school

	Dependent variable: graduated post-secondary school						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	3 year bandwidth	5 year bandwidth	7 year bandwidth	9 year bandwidth	11 year bandwidth	13 year bandwidth	15 year bandwidth
Had six years of primary school	0.0671	-0.00932	-0.0206	-0.0137	-0.0368	-0.0196	-0.00637
	(0.0753)	(0.0636)	(0.0532)	(0.0453)	(0.0418)	(0.0378)	(0.0375)
Time since treatment year	0.00161	0.0133**	0.00460	0.00252	0.00393	0.00318*	0.000794
	(0.00971)	(0.00561)	(0.00465)	(0.00334)	(0.00247)	(0.00192)	(0.00156)
Female	-0.00231	-0.0125	-0.0136	-0.0174**	-0.0141*	-0.0111	-0.0122*
	(0.0160)	(0.0115)	(0.00954)	(0.00867)	(0.00784)	(0.00711)	(0.00733)
Urban resident	0.0511	0.0266	0.0274	0.0377	0.0238	0.0266	0.0304
	(0.0348)	(0.0311)	(0.0263)	(0.0233)	(0.0213)	(0.0201)	(0.0194)
Urban household registry	0.260***	0.307***	0.300***	0.296***	0.297***	0.286***	0.275***
	(0.0358)	(0.0266)	(0.0259)	(0.0225)	(0.0218)	(0.0203)	(0.0193)
Ethnic minority	-0.0589	-0.0183	-0.0285	-0.0206	-0.00770	0.00584	0.00778
	(0.0518)	(0.0391)	(0.0345)	(0.0318)	(0.0290)	(0.0259)	(0.0235)
Western province	-0.189***	0.00142	-0.0674**	0.0941***	-0.0568*	-0.0625**	0.0447***
	(0.0650)	(0.0348)	(0.0278)	(0.0304)	(0.0297)	(0.0263)	(0.0139)
Late implementer	0.0268	0.337***	-0.0298	0.100***	-0.0330	-0.0421**	-0.0422***
	(0.0675)	(0.0592)	(0.0304)	(0.0313)	(0.0239)	(0.0195)	(0.0158)
Constant	0.172	-0.128*	0.0281	-0.127***	0.0460	0.0117	-0.00180
	(0.131)	(0.0661)	(0.0461)	(0.0445)	(0.0381)	(0.0334)	(0.0291)
Observations	1,493	2,506	3,474	4,347	5,189	5,901	6,508
R-squared	0.303	0.301	0.294	0.277	0.269	0.266	0.257

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 20: Restricted bandwidth regressions, instrumental variables, ever attended high school

	Dependent variable: attended high school						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	3 year bandwidth	5 year bandwidth	7 year bandwidth	9 year bandwidth	11 year bandwidth	13 year bandwidth	15 year bandwidth
Had six years of primary school	0.167*	0.0608	0.0393	0.0186	-0.0544	-0.0249	-0.0165
	(0.0910)	(0.0712)	(0.0623)	(0.0551)	(0.0532)	(0.0483)	(0.0492)
Time since treatment year	-0.0171	0.00527	0.00362	0.00844**	0.0141***	0.0131***	0.0159***
	(0.0122)	(0.00808)	(0.00551)	(0.00408)	(0.00349)	(0.00301)	(0.00240)
Female	-0.0629***	-0.0627***	-0.0608***	-0.0580***	-0.0542***	-0.0510***	-0.0489***
	(0.0204)	(0.0155)	(0.0129)	(0.0117)	(0.0109)	(0.00950)	(0.00964)
Urban resident	0.0441	0.0221	0.0419	0.0348	0.0353	0.0371	0.0465*
	(0.0496)	(0.0386)	(0.0341)	(0.0290)	(0.0277)	(0.0259)	(0.0239)
Urban household registry	0.423***	0.437***	0.388***	0.393***	0.389***	0.376***	0.367***
	(0.0490)	(0.0345)	(0.0330)	(0.0264)	(0.0249)	(0.0244)	(0.0235)
Ethnic minority	-0.120*	-0.0771*	-0.0449	-0.0199	-0.0175	-0.0204	-0.0227
	(0.0685)	(0.0446)	(0.0416)	(0.0374)	(0.0337)	(0.0291)	(0.0273)
Western province	-0.181	0.173***	0.0697*	0.0798**	-0.0207	0.0191	0.199***
	(0.140)	(0.0567)	(0.0414)	(0.0395)	(0.0370)	(0.0317)	(0.0208)
Late implementer	0.0890	0.0379	0.00538	0.0749	0.0136	0.0459	0.113***
	(0.0860)	(0.115)	(0.0400)	(0.0496)	(0.0397)	(0.0374)	(0.0288)
Constant	0.364*	-0.111	0.121*	0.0591	0.243***	0.203***	0.164***
	(0.217)	(0.0807)	(0.0681)	(0.0623)	(0.0580)	(0.0532)	(0.0466)
Observations	1,493	2,506	3,474	4,347	5,189	5,901	6,508
R-squared	0.340	0.325	0.298	0.290	0.280	0.274	0.275

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 21: Restricted bandwidth regressions, instrumental variables, graduation from post-secondary school

	Dependent variable: attended post-secondary school						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	3 year bandwidth	5 year bandwidth	7 year bandwidth	9 year bandwidth	11 year bandwidth	13 year bandwidth	15 year bandwidth
Had six years of primary school	0.0234	-0.0111	-0.0168	-0.0193	-0.0286	-0.0155	-0.00522
	(0.0630)	(0.0550)	(0.0451)	(0.0370)	(0.0370)	(0.0330)	(0.0327)
Time since treatment year	0.0101	0.0122**	0.00365	0.00354	0.00422**	0.00356**	0.00143
	(0.00894)	(0.00496)	(0.00348)	(0.00219)	(0.00188)	(0.00150)	(0.00138)
Female	-0.00236	-0.0106	-0.0147*	-0.0170**	-0.0134*	-0.0117*	-0.0101
	(0.0148)	(0.00984)	(0.00839)	(0.00797)	(0.00734)	(0.00679)	(0.00664)
Urban resident	0.0313	0.0205	0.0184	0.0192	0.00731	0.00448	0.00974
	(0.0259)	(0.0236)	(0.0197)	(0.0173)	(0.0159)	(0.0148)	(0.0143)
Urban household registry	0.154***	0.181***	0.190***	0.197***	0.203***	0.199***	0.193***
	(0.0288)	(0.0206)	(0.0229)	(0.0194)	(0.0187)	(0.0174)	(0.0169)
Ethnic minority	-0.0158	0.00667	-0.00568	-0.0142	-0.00814	-0.000579	0.00316
	(0.0419)	(0.0320)	(0.0264)	(0.0255)	(0.0212)	(0.0196)	(0.0177)
Western province	-0.232***	-0.0442*	-0.0363*	0.0694***	-0.0740***	-0.0763***	-0.0325***
	(0.0645)	(0.0252)	(0.0218)	(0.0239)	(0.0223)	(0.0201)	(0.0118)
Late implementer	0.00365	0.255***	-0.0202	0.0605***	-0.0530***	-0.0557***	-0.0531***
	(0.0544)	(0.0573)	(0.0247)	(0.0229)	(0.0201)	(0.0148)	(0.0123)
Constant	0.275**	-0.0294	0.0214	-0.0763**	0.0854**	0.0577*	0.0367
	(0.121)	(0.0611)	(0.0384)	(0.0370)	(0.0349)	(0.0299)	(0.0257)
Observations	1,493	2,506	3,474	4,347	5,189	5,901	6,508
R-squared	0.212	0.213	0.208	0.200	0.199	0.197	0.193

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1