I Z A

IZA DP No. 9274

Low Wage Returns to Schooling in a Developing Country: Evidence from a Major Policy Reform in Turkey

Abdurrahman Aydemir
Murat G. Kırdar

August 2015

Low Wage Returns to Schooling in a Developing Country: Evidence from a Major Policy Reform in Turkey

Abdurrahman Aydemir
Sabancı University
and IZA

Murat G. Kırdar
Boğaziçi University
and IZA

ABSTRACT

Low Wage Returns to Schooling in a Developing Country: Evidence from a Major Policy Reform in Turkey[*]

In this paper, we estimate the returns on schooling for young men and women in Turkey using the exogenous and substantial variation in schooling across birth-cohorts brought about by the 1997 reform of compulsory schooling. We estimate that among 18- to 26-year-olds, the return from an extra year of schooling is almost zero for men and 3.8 percent for women. The low level of these estimates contrasts starkly with those estimated for other developing countries. We identify several reasons why the returns on schooling are low and why they are higher for women in our context. In particular, the policy alters the schooling distributions of men and women differently, thus the average causal effect we estimate puts a higher weight on the causal effect of schooling at higher grade levels for women than for men.

Corresponding author:

Abdurrahman Aydemir
Faculty of Arts and Social Sciences
Sabancı University
Orhanli
Tuzla 34956 Istanbul
Turkey
E-mail: aaydemir@sabanciuniv.edu

---

1. Introduction

Few studies deal with the endogeneity of education in the estimation of returns to schooling in developing country contexts. These include Duflo (2001), who uses a major school construction policy as a source of exogenous change in schooling in Indonesia; Spohr (2003), who explores the impact of Taiwan's 1968 reform; and Fang et al. (2012), who study the 1986 implementation of compulsory education law in China. This paper contributes to this small literature by estimating the returns to schooling for both men and women in the developing country context of Turkey—using the extension of compulsory schooling from 5 to 8 years in 1997 as an instrumental variable.

A key distinguishing feature of this study is the plausible exogeneity and strength of our instrument. The timing of the implementation of this compulsory schooling reform was unexpected and took place over a short period of time. This led to a sharp increase in schooling levels across birth cohorts—which is the source of identification in this study—not only because of the long duration of the extension but also high dropout rates after the completion of compulsory schooling.

In 1997, Turkey instituted a major reform of its education system, which increased the duration of compulsory schooling from five to eight years.[1] According to national education statistics, during the 1996–97 school year, the year before the law changed, the net enrollment was 89.4% at the primary school level (grades 1 to 5), while enrollment at the secondary school level (grades 6 to 8) was 52.8%.[2] Thus, the 1997 reform potentially affected close to half of school-age children at the secondary school level. By the 2000–01 school year, four years after the law

---

[1] 5.1.1961 *tarih ve* 222 *Sayılı İlköğretim ve Eğitim Kanunu 2.Maddesi*.

[2] TUIK (Turkish Statistical Institute), Education Statistics

was enacted, the net enrollment ratio at the compulsory schooling stage (grades 1 to 8) increased to 95.3%. Such a drastic rise in the number of students could raise concerns about the quality of schooling. However, the government reacted very fast to expand the infrastructure of schooling. In fact, the results of certain international tests indicate that achievement of the affected cohorts vis-à-vis the unaffected indicate no deterioration in the quality of schooling.

The actual timing of the policy was driven by political developments and was independent of potential returns to education and macroeconomic conditions. Furthermore, there were no other simultaneous policies that could also affect schooling outcomes, as the timing of the policy was determined by political rather than economic events. Because the policy was implemented nationwide, there are no concerns about endogenous program placement or participation that arise in contexts in which identification comes from regional variation in policy implementation. The remarkable change in schooling outcomes resulting from the reform provides us a very strong instrument.[3] A related issue is that since Turkey's compulsory schooling reform affected a large fraction of the population, the local average treatment effect we estimate comes close to the average treatment effect as in Oreopoulos (2006).

The data we use in this study come from the Household Labor Force Surveys of Turkey for 2002 through 2013. Since our identification strategy is based on a comparison of birth cohorts— albeit one where the discontinuity across birth cohorts is very sharp and substantial and where there are no other contaminating policies—our key variable of interest, the policy variable, is invariant across birth cohorts. Moreover, the number of birth cohorts in our sample is relatively few at 20. Therefore, we use several alternative approaches (outlined in Cameron and Miller

---

[3] The importance of a strong source of exogenous variation is discussed by Bound et al. (1995), who show that asymptotic bias resulting from weak instruments pushes the 2SLS estimates toward the OLS estimate. This issue is further addressed by Staiger and Stock (1997).

[2015]) to estimate accurate standard errors. In addition to the standard cluster-robust estimates, these approaches include inference based on a T distribution with adjusted degrees-of-freedom, parametric Moulton correction of standard errors, nonparametric pairs cluster bootstrap, and wild cluster bootstrap.

We find that the 2SLS estimate among 18- to 26-year-old individuals is barely above zero for men and about 3.8 percent for women. These estimates are much smaller than the estimates reported by the previous studies in developing country settings. Duflo (2001) reports return estimates ranging from 6.8% to 10.6%. Spohr (2003) finds larger effects on females' workforce participation and total earnings, where annual earnings increase by 5.8% per additional year of schooling for men and by 16.7% for women. Fang et al. (2012) report overall returns to education of approximately 20% per year. However, the low returns we find in this study adds to a growing number of recent studies that report zero or low returns to schooling in developed country contexts (Pischke and von Wachter, 2008; Devereux and Hart, 2010; Stephens and Yang, 2014) but in a developing country context. Hence, this paper adds to the renewed discussion in the literature on returns to education.

While the previous studies in developing country contexts find larger returns for women than for men, there is little discussion of these differential returns. We provide a detailed account of how the reform affected the distribution of highest grade completed by gender (as in Acemoğlu and Angrist [2000]), which provides important insights related to this apparent difference in returns. In particular, the female sample for whom we estimate the returns to schooling include a larger fraction of those who are induced to complete high school as a result of the policy, compared to the male sample. Thus, the average causal effect that we estimate for women puts a higher weight on the returns to schooling at high school grade levels than the estimate for men, and returns to schooling are higher at high school grade levels than at lower grade levels.

In the next section, we provide background on several methodological challenges for estimating returns to schooling and discuss how the reform used in this study overcomes these estimation challenges. This is followed by a brief description of Turkey's compulsory schooling reform in Section 3. Section 4 discusses the data used in the analysis. The methodology is discussed in Section 5. Section 6 presents the findings, and Section 7 provides an interpretation of these findings. Section 8 concludes.

2. Related Literature on Methodology

In addition to being a plausibly clean and strong source of exogenous variation in schooling, the compulsory schooling reform in Turkey has a number of important features that relate to important methodological discussions in the literature regarding the comparison of the OLS and 2SLS estimates. Since the pioneering study of Griliches (1977), the literature on economic returns to schooling addresses the issue of how much of the reported return to schooling is due to the ability bias: does schooling really make people more productive, or is it because more able people are more likely to acquire higher levels of schooling? Since the 1990s, following the pioneering study of Angrist and Krueger (1991), many studies have used institutional features of schooling systems as a source of exogenous variation in schooling so that returns to schooling can be estimated consistently in terms of earnings.[4] Several studies in the literature use changes in compulsory schooling laws for identification. In the United Kingdom and Ireland, Harmon and Walker (1995) and Oreopoulos (2006) use an identification strategy that relies on a comparison of the birth cohorts affected by the policy with those not affected. In this

---

[4] Several studies in the literature also employ alternative instruments based on the distance to the nearest college (Kane and Rouse, 1995; Card, 1995; Connely and Uusitalo, 1997; Maluccio, 1997), family background variables (Card, 1995), and new school construction (Duflo, 2001) at various locations.

x

In the next section, we provide background on several methodological challenges for estimating returns to schooling and discuss how the reform used in this study overcomes these estimation challenges. This is followed by a brief description of Turkey's compulsory schooling reform in Section 3. Section 4 discusses the data used in the analysis. The methodology is discussed in Section 5. Section 6 presents the findings, and Section 7 provides an interpretation of these findings. Section 8 concludes.

2. Related Literature on Methodology

In addition to being a plausibly clean and strong source of exogenous variation in schooling, the compulsory schooling reform in Turkey has a number of important features that relate to important methodological discussions in the literature regarding the comparison of the OLS and 2SLS estimates. Since the pioneering study of Griliches (1977), the literature on economic returns to schooling addresses the issue of how much of the reported return to schooling is due to the ability bias: does schooling really make people more productive, or is it because more able people are more likely to acquire higher levels of schooling? Since the 1990s, following the pioneering study of Angrist and Krueger (1991), many studies have used institutional features of schooling systems as a source of exogenous variation in schooling so that returns to schooling can be estimated consistently in terms of earnings.[4] Several studies in the literature use changes in compulsory schooling laws for identification. In the United Kingdom and Ireland, Harmon and Walker (1995) and Oreopoulos (2006) use an identification strategy that relies on a comparison of the birth cohorts affected by the policy with those not affected. In this

---

[4] Several studies in the literature also employ alternative instruments based on the distance to the nearest college (Kane and Rouse, 1995; Card, 1995; Connely and Uusitalo, 1997; Maluccio, 1997), family background variables (Card, 1995), and new school construction (Duflo, 2001) at various locations.

paper, similar to Harman and Walker (1995) and Oreopoulos (2006), we compare the birth cohorts that are affected by the 1997 education policy in Turkey with those that are unaffected.

Many of the earlier studies in the literature find that 2SLS estimates of the returns to schooling are either higher than or not substantially different from OLS estimates. Such findings, however, are not consistent with the ability bias argument, which argues that since more able people acquire higher levels of schooling, the coefficient of schooling in the OLS estimates also captures the effect of unobserved ability. Bound and Jaeger (1996) identify other sources of bias, while Angrist and Krueger (1991) discuss the bias resulting from measurement errors that may lead to this apparent contradiction.

Another prominent explanation for the lower OLS estimates compared to the IV estimates is heterogeneity in the returns to education. Imbens and Angrist (1994) show that the 2SLS estimates give the effect for the set of compliers—those who are induced to change their schooling choice as a result of an exogenous source of variation in schooling. However, the set of compliers may not be representative of the whole population. The estimates derived from the changes in compulsory school leaving age provide local treatment effects (LATE) for compliers. Oreopoulos (2006) notes that as the fraction of the population affected by the policy change increases, the LATE estimate approaches the average treatment effect (ATE) as the marginal individual affected by the policy change becomes increasingly similar to the average individual in the population. Using a much larger set of compliers, Oreopoulos (2006) finds that the 2SLS estimates in the U.K. context are not substantially different from those found by Angrist and Krueger. Devereux and Hart (2010), however, find much lower returns in the same context. Similarly, Pischke and von Wachter (2008) report much lower IV estimates (zero returns) than the OLS estimates in the German context, where a large fraction of the population was affected by compulsory school law change. Thus, there is mixed evidence from these studies of developed countries about the IV and OLS estimates of the returns to education. Our

estimates in a developing country context contribute to the ongoing discussion in the literature about the relative size of the IV and OLS estimates. Similar to these studies, Turkey's compulsory schooling reform affects a large fraction of the population, which brings the local average treatment effect estimate closer to the average treatment effect.

In both the British and German contexts, compulsory school law changes involve a one-year increase, increasing the schooling leaving age from 14 to 15 in the former case and adding a compulsory ninth grade for students in the latter case. If there are nonlinearities in the return to schooling, estimating the return via a policy-induced increase at a given schooling level may lead to differences between the estimated local average treatment effect and the average treatment effect. For example, while Pischke and von Wachter find zero or low returns to an additional year of schooling resulting from the compulsory school extension from eighth to ninth grade, other IV estimates from Germany (Ichino and Winter-Ebmer, 1999, 2004; Becker and Sienbern-Thomas, 2001) are much larger, around 10 percent. Pischke and von Wachter argue that this difference in results may be due to instruments that consider education differences at different points—for example, differences in grades 10 and above in the Becker and Siebern-Thomas study. They argue that the additional year of schooling explored in their study, which they estimated to have a zero return in terms of earnings, may not be equipping students with any labor market-relevant skills. A unique feature of the Turkish context is that not only does the policy affect a large fraction of school-age children, it also induces higher schooling at different school levels. Individuals who would have left school by the end of grade 5, 6, or 7 in the absence of the policy are now required to complete eight years of schooling. Moreover, there are spillover effects beyond grade 8, and some students continue schooling up to grade 9, 10, or 11 as a result of the policy change. To the extent that such nonlinearities exist in the returns to education, the presence of policy effects at different schooling levels also brings our local treatment effects closer to the average treatment effect.

3. Turkey's Compulsory Schooling Reform

Prior to the 1997 reform in Turkey, compulsory schooling consisted of five years of primary education followed by three years of lower secondary education and three years of high-school education. With the reform, the first eight years of education were redefined as primary education, again followed by three years of high school. This policy was implemented nationwide, and its timing was closely related to the political developments of the time. The major motivation for the new policy was to restrict religious education.[5] Unlike some major schooling and health reforms in other contexts that are made possible by better-than-average economic conditions, the reform in Turkey, which was primarily politically motivated, was independent of the macroeconomic context, and therefore did not coincide with other policies that would have a bearing on schooling outcomes.

Prior to 1997, the dropout rate after the completion of five years of compulsory education was very high. According to the enrollment ratios reported by national education statistics, in the 1996–1997 school year, one year before the policy change took effect, almost 40 percent of students dropped out of school immediately after earning a primary school degree. The extension of compulsory school from five to eight years resulted in a sharp increase in the number of students in grades 6 to 8. In fact, the number of students enrolled in grades 1 to 8 increased by around 16 percent—from slightly over 9 million students to close to 10.5 million— between the 1996–1997 and 2000–2001 school years. The Ministry of Education responded to

---

[5] This was achieved primarily in two ways. First, the middle school levels (grades 6 to 8) of *"Imam Hatip schools*," which provide a combination of religious and secular education, were closed with the introduction of eight years of compulsory education. These schools still provided education at the high-school level. Second, students who enrolled in *Quranic schools,* which provide only religious education, immediately upon the completion of compulsory school had to delay their attendance for three years.

these increases with a number of measures to ensure that all students were schooled without a significant decline in education quality. These measures included construction of new schools, increase of capacity in existing schools, hiring of new teachers, bussing of students, and construction of boarding schools for those in rural areas.

The number of classrooms in urban areas increased from 169,202 in the 1997–1998 school year to 218,093 in the 2001–2002 school year, whereas in rural areas, the number increased from 148,516 in the 1996–1997 school year to 170,046 in the 2002–2003 school year. The number of students bussed to schools free of charge increased from around 127,000 in the 1996–1997 school year to about 622,000 in the 1999–2000 school year. Similarly, the number of students enrolled in boarding schools increased from around 34,000 in the 1996–1997 school year to 282,000 in the 2000–2001 school year. These large-scale investments were made possible by a large increase in the Ministry of Education's budget, which as a share of the total public investment budget increased from 15 percent in 1997 to 37.3 percent in 1998 and remained around 30 percent until 2000 (see Kirdar et al. [2014] for a detailed account of these changes). As a result, the student-to-class ratio, which rose from 28.6 in the 1997–1998 school year to 31.2 in the 1999–2000 school year, fell to 28.3 in the 2001–2002 school year and continued to decline as a result of investment in education. The student-to-teacher ratio remained steadily around 30 in the initial years of the policy change and fell below 28 by early 2000. While the newly hired teachers lacked the experience of existing teachers, they had more exposure to newer computer-assisted teaching techniques. Thus, the effect of new teacher hiring on instruction quality is unclear. Moreover, that student-to-class and student-to-teacher ratios either remained constant or declined following the reform.

These changes in the schooling infrastructure and the substantial increase in the student population with the policy might raise concerns about a deterioration in the quality of schooling. However, evidence from the TIMSS 1999 and 2007 international tests that cover grade 8

students suggests the opposite. These tests aim to uncover the trends in mathematics and science achievement over time and reflect the quality and content of instruction in these subjects. The cohort of 8 graders in Turkey who participated in TIMSS 1999 were not affected by the extension of the compulsory school reform whereas those that participated in TIMSS 2007 were affected by the reform. Between 1999 and 2007, the mathematics achievement score of Turkey increased by 3 points (from 429 to 432) while the average score among all participating countries declined by 37 points (from 487 to 450) (MEB, 2007). When we examine the changes by gender, we see that both males and females improved their mathematics scores by 3 points in this period. The improvement in science achievement among Turkish students was more pronounced, increasing by 21 points between 1999 and 2007 (from 433 to 454) while the average score among all participating countries decreased by 23 points (from 488 to 465). Over this period, the average science score increased by 26 points for girls while it increased by 18 points for boys.[6] Thus, these results suggest that there was no deterioration in the quality of schooling following the compulsory school reform as captured by these international achievement tests.

Next, we present the change in school attainment with the policy for the total population using micro-level data from Turkish Household Labor Force Surveys. Figure 1 displays the fraction of children who complete grade 8 by birth year for men and women. The figure shows a significant discontinuity between the 1985 and 1988 birth-cohorts as well as non-trivial time trends both before and after the discontinuity. The fraction of men who completed grade 8 increases by about 18.6 percentage points—from 71.9 to 90.5 percent—between the 1985 birth-

---

[6] As a result of these changes, the gap between Turkey and all other participating countries fell from 58 to 18 points in the average mathematics score and from 55 to 9 points in the average science score between 1999 and 2007.

cohort and the 1988 birth-cohort. This is in comparison to the 6-percentage-point increase from the 1982 birth-cohort to the 1985 birth-cohort and the 5-percentage-point increase from the 1979 birth-cohort to the 1982 birth-cohort. The jump from the policy in the fraction completing grade 8 is even more substantial for women: it rises from 49.7 percent to 74.8 percent—by about 25 percentage points—between the 1985 and 1988 birth-cohorts. This is in comparison to the 4.8-percentage-point increase between the 1982 and 1985 birth-cohorts and the 5.6 percentage point rise between the 1979 and 1982 birth cohorts.

Figure 2, which presents the fractions of birth cohorts completing at least high school, shows a significant spillover effect of the policy on high-school completion rates. The fraction earning at least a high-school degree increases by 6.1 percentage points from the 1985 birth-cohort to the 1988 birth-cohort (from 53.0 to 59.1 percent). This is in comparison to a 4-percentage-point increase between the 1982 and 1985 birth-cohorts and a 1.4-percentage-point increase between the 1979 and 1982 birth-cohorts. For women, these spillover effects of the policy on high-school completion are even more striking. While there is a rise of 2.7 percentage points in the fraction completing high school between the 1979 and 1982 birth-cohorts and a rise of 2.4 percentage points between the 1982 and 1985 birth-cohorts, the rise is 9.6 percentage points between the 1985 and 1988 birth-cohorts. A similar examination of college graduation rates does not indicate such a spillover effect of the policy—although the data after the cutoff are quite sparse. Overall, the policy has been extremely effective in increasing schooling among males and females, not only at the newly mandated grades 6 to 8 but also at the high-school grade levels.

4. Data

The data used in the analysis come from the 2002 to 2013 Turkish Household Labor Force Surveys (HLFS). HLFS is a nationally representative survey of individuals in Turkey. We use

11

all rounds of HLFS data that include wage and employment information for individuals aged 15 and over. The data report the age, level of educational attainment, hours of work for the reference week, and earnings of individuals during the past month, including bonus payments and premiums.

The sample is restricted to full-time workers, and it therefore excludes workers with fewer than 30 hours worked during the reference. Since the policy change had significant spillover effects on high-school completion, we restrict our sample to those who are at least 18 years old in the survey year in order to avoid excluding individuals who are affected by the policy but who are still at school. In addition, for reasons explained in detail in the next section, our primary sample includes those who are 26 years old or younger in the survey year. There is no information in the data about the earnings of self-employed individuals; thus, the analysis is restricted to wage and salary earners only. This restriction is relatively parsimonious in our context because as the primary sample includes only young workers, the fractions of wage earners are high: about 89 percent for men and 95 percent for women. We also trim the top 1 percent and bottom 1 percent of the wage distribution to exclude outlier observations. The sample size for our primary sample including 18 to 26-year-old individuals is 134,113 for men and 58,835 for women.

The HLFS surveys do not include information on completed years of schooling, but includes information on the highest completed level of schooling (illiterate; literate but no degree; primary school [5 years, earlier system]; lower secondary school; general, vocational, or technical high school; four-year university). Therefore, we refer to the Turkish Demographic and Health Surveys, which include information on both completed degrees and completed years of schooling, to find the distribution of years of schooling conditional on the highest completed

schooling level. Using this information, we generate the mean years of schooling for each of the highest completed schooling levels in the HLFS.[7]

The birth year of individuals, which determines exposure to the policy, is generated by subtracting the reported age from the survey year. We define a policy dummy, which is equal to one for those born in 1987 or later and zero for earlier birth cohorts. This policy dummy is used as an instrument for schooling.

The variation across workers in annual earnings reflects the differences in the hourly earnings, which is a measure of their productivity, as well as the number of hours they work in a year. The annual hours of work could differ across educational groups because workers who are more educated may be more likely to be employed at a given time or more likely to work for longer hours (Card [1999] provides evidence for the latter in the U.S. context). Therefore, the primary measure of earnings used in this study is earnings per hour. Using information on monthly income and hours of work, we compute the hourly wages (monthly income/(weekly hours*4.3)), and we use the log hourly wages as our dependent variable.

5. Identification Method and Estimation

Our goal is to establish a causal link between schooling and wages in the Turkish context. The well-known problem in establishing this causal link is the endogeneity of schooling, which is brought about by omitted variables such as ability, motivation, and parental connections. To

---

[7]According to the 2008 Turkish Demographic and Health Survey, the mean years of schooling for people with a certain degree could be well approximated by the minimum years of schooling required for that degree: five years for a primary school degree, eight years for a secondary school degree, 11 years for a high-school degree, and 15 years for a four-year college degree. The average years of schooling is very close to zero for illiterate people and close to two years for those who are literate but have no degree.

solve this problem, we need a source of exogenous variation in schooling that is correlated with schooling but not correlated with other determinants of wages. More precisely, such an instrument needs to be independent of wages conditional on covariates and have no direct effect on wages other than through its effect in the first stage.

In our study, this instrument is the compulsory schooling reform in Turkey. In particular, we exploit the variation in schooling across birth-cohorts induced by this institutional change. This instrument satisfies the above-mentioned conditions for a valid instrument. First, the timing of the reform has nothing to do with wages because its timing was the result of political developments. The secular government, which came to power just before the 1997 reform, saw this reform as a means to curtail religious education. Second, there is no reason to expect birth year to be correlated with ability, motivation, or parental connections. Thus, we would not expect a direct effect of birth year on wages other than through its effect on schooling.

The education reform of 1997 went into effect for the 1997–1998 school year. It covered students who finished grade 4 or a lower grade at the end of the 1996–1997 school year—those who did not hold a primary education diploma at the beginning of the 1997–1998 school year. Thus, students who started school in September 1993 or afterward were bound by the policy, whereas students who started school in September 1992 or earlier were exempt from it. Most children in Turkey start school at the age of six. Conditional on this school start age, children who were born before September 1986 were not affected by this policy. However, some who were born before September 1986 would be bound by the policy if they started school later than the normal age, which is frequent in Turkey. Similarly, some who were born after September 1986 would be exempt from the policy if they started school earlier than the normal age (and thus had already earned a primary school diploma by the time of the policy). In addition, late implementation of the policy in certain areas (e.g., rural areas in which bussing schemes had to be established) could mean that some children who were born within a year after September

1986 were not affected by the policy. Owing to this uncertainty in the treatment for the 1986 and 1987 birth-cohorts, we check the sensitivity of dropping them from the sample. Note that this uncertainty can also be seen in Figures 1 and 2.

We use two-stage least squares estimation to find the causal link between schooling and wages as follows:

$$s_i = \alpha_0 + \alpha_1 D_i + X'\theta + u_i \tag{1}$$

$$\log w_i = \beta_0 + \beta_1 s_i + X'\delta + v_i \tag{2}$$

Here, s denotes the years of schooling, w denotes the wage rate, and D is a dummy variable for the policy. The key parameter of interest is β_1, which denotes the percent change in wages when years of schooling are raised by one. Covariates are shown by X; these include dummies for each age and each calendar year, and a dummy for urban status.[8]

Table 1 indicates whether or not a certain age group at a certain year in our data is influenced by the education reform by identifying the birth-cohort it belongs to. Only ages 16 through 26 include both birth-cohorts that are affected by the policy and those that are not. We also place a restriction on the minimum age, excluding ages below 18, because the policy had significant spillover effects on high-school grades, as shown in Figure 2. By setting 18 as the minimum age, we restrict our sample to those who are beyond the high-school age. Thus, our main sample, which we call sample A, covers ages 18 through 26. We also define an alternative sample, sample B, which excludes the 1986 and 1987 birth-cohorts due to the fuzziness in the treatment status of these birth cohorts (as can be seen in Figure 1). Sample B is restricted to 18 to 25-

---

[8] In an alternative specification, we also include dummies for 26 NUTS-2 level regions. We do not include them in our main specification because region information is missing in the 2002 and 2003 data.

year-olds because there is no variation in the policy variable for the 26-year-olds when the 1987 birth-cohort is dropped (as can be seen in Table 1). We estimate the returns to schooling for a limited age group with our sample; however, this is appropriate because the earnings–schooling relationship changes with age—which is an implication of the Mincer model (see, e.g., Heckman et al. [2007]) and which can be seen empirically in Figure A1 in the Appendix. As Imbens and Angrist (1994) show, the 2SLS estimates have a local average treatment effect (LATE) interpretation; in other words, they give the treatment effect for those whose behavior is changed by the instrument, who are called *compliers*. This is important when there is heterogeneity in the returns schooling across various subpopulations because, as LATE is for the compliers only, it may be quite different from the average treatment effect (ATE) for the total population.[9] As discussed in Oreopoulos (2006), the larger the set of compliers, the closer LATE is to ATE. In fact, as in Oreopoulos (2006), our set of compliers is large, and thus, our estimated returns on schooling are closer to the average treatment effect than are those in many other studies.[10]

Finally, we put much effort in obtaining accurate standard errors. In our data, each observation is not independent due to the group structure—the policy variable does not vary within birth-year clusters. In this case, failure to account for within-cluster error correlation could significantly overstate estimator precision, which is well known since Moulton (1986).

---

[9] Another assumption of the LATE framework is monotonicity: the policy does not decrease individuals' years of education. We cannot test this using our data because we do not know the exact years of schooling. The fraction that does not have a primary school diploma increases, but this is likely due to dropouts after completion of grade 5, 6, or 7 after the policy was implemented, which does not mean that the policy decreases years of schooling.

[10] A difference between our context and that of Oreopoulos (2006) is that in our context, there are those who do not accept the treatment even though they are bound by it (*defiers*). Therefore, our LATE is not equal to the treatment on the non-treated.

Therefore, we use cluster-robust standard errors.[11] However, when the number of clusters is small, even clustered standard errors might not be good enough (Angrist and Pischke, 2013; Cameron et al., 2008) because the asymptotic distribution for clustered standard errors is based on a large number of clusters. We have 20 clusters in our main sample; however since the clusters are unbalanced in terms of the sample size, the effective number of clusters could be fewer as illustrated by MacKinnon and Webb (2013). Hence, in order to address the few clusters issue, we use a number of approaches—outlined in Cameron and Miller (2015)—in addition to the standard cluster-robust estimates, which include: i) inference based on a T distribution with adjusted degrees-of-freedom, ii) parametric Moulton correction of standard errors,[12] iii) nonparametric pairs cluster bootstrap,[13] iv) wild cluster bootstrap.[14]

---

[11] This is implemented using the cluster option in Stata. Cameron and Miller (2015) provide a helpful guide for cluster-robust inference.

[12] Moulton-factor adjusted standard errors for OLS and 2SLS are calculated using the Stata program available at *http://economics.mit.edu/faculty/angrist/data1/data/anglavy99*.

[13] In pairs cluster bootstrapping, in each draw, one randomly obtains G clusters with replacement from the original sample of G clusters and calculates the Wald test statistic with this resample. This method is applicable to 2SLS as well as to OLS in Stata. We perform 999 replications in all bootstraps.

[14] Cameron et al. (2008) and Cameron and Miller (2015) implement wild-cluster bootstrap for the few-clusters problem. The key difference of wild cluster bootstrap from pairs cluster bootstrap is that in wild cluster bootstrap the set of control variables is fixed but new pseudo-residuals are generated in each resample. Various distributions (e.g. Rademacher, Liu, and Mammen) can be used in generating the new pseudo-residuals. Cameron and Miller find that wild cluster bootstrap performs much better than the standard cluster-robust inference (as well as inference based on alternative techniques) in Monte-Carlo rejection rates of the null hypothesis. However, Cameron and Miller implement wild cluster bootstrap only in the OLS setting. Davidson and McKinnon (2012) develop wild cluster bootstrap methods for IV; however, their main concern is the weak-instrument problem and they do not allow for clustering. Gelbach et al. (2009) apply a variant of the wild-cluster bootstrap in Cameron et

6. Results

6.1 First-Stage Results: Policy Effects on Schooling Outcomes

We first examine how the policy changes schooling attainment for the individuals in our sample of wage earners. For this purpose, Figure 3 illustrates the mean completed years of schooling by birth year so that we can examine any potential discontinuity around the 1986–1987 cutoff. This is given for four different values of age within the age range of our sample. For ages 18 and 20, there is a remarkable jump around the cutoff for both men and women. For instance, for 18-year-old men, while the average years of schooling is around seven before the cutoff, it is above eight after the cutoff. There are also important discontinuities for 22- and 24-year-old individuals around the cutoff, particularly for men; however, these are relatively less visible because of the wider range of the *y*-scale in these graphs. Another prominent feature in Figure 3 is the rise in the average years of schooling for women as age increases. This occurs because college graduates begin to enter the labor force as age increases. This is much more obvious for women than for men because the share of college graduates in the workforce is much higher for women than for men in Turkey.

Table 2 presents the first-stage results for men and women for the two separate samples defined earlier. Overall, the estimates confirm that our instrument is highly relevant. In addition, the exclusion of the 1986 and 1987 birth-cohorts changes the coefficients substantially. According to the estimates with sample (A), the policy increases schooling by about 0.59 years for men and 0.44 years for women. However, when the 1986 and 1987 birth-cohorts are excluded, this increase becomes almost one year for men and more than 0.6 years for women. In a parallel

---

al. (2008) to an IV setting; in fact, for wild-cluster bootstrap we use the program by Gelbach at *http://gelbach.law.upenn.edu/ado/wildbs_new.ado*.

fashion, the F-statistics are much higher when the transition birth-cohorts are excluded; however, when they are included, the F-statistics are still much higher than the recommended levels.

6.2 A Closer Look at Changes in Schooling Distributions by Gender

The first-stage results indicate a stronger effect of the policy on the schooling of men than on that of women because our sample includes only individuals who are wage earners. This is different from the impact of the policy on the schooling attainment of the total population by gender suggested in Figures 1 and 2. A well-known fact about employment and gender in Turkey is that women's employment probability increases tremendously as education increases, whereas there is no such difference across education groups for men (Dayıoğlu and Kırdar, 2009).

In order to understand this issue better, we examine the cumulative distribution function for education (specifically, one minus the cumulative distribution function) in our sample by gender. Figure 4 displays how this cumulative distribution function changes because of the policy; for this purpose, we compare the 18 to 26-year-olds in the 1984 and 1985 birth-cohorts in our sample (who are not affected by the policy) with the same age group in the 1988 and 1989 birth-cohorts (who are affected). Note that we take only these birth-cohorts immediately around the cutoff in order to minimize the potential effect of a time trend. As can be seen from the figure, the fraction of men who finish eight or more years of schooling increases by 24.72 percentage points, whereas the fraction of men who finish high school or a higher level of schooling increases by 2.97 percentage points. Among women, the respective percentage-point changes are 14.78 and 5.34. In other words, the ratio of the rise in the fraction earning a high-school degree or higher to the rise in the fraction earning a primary school degree or higher is much higher among women (0.36) than for men (0.12).

Put differently, in our sample of wage earners, the policy pushes a higher fraction of men to finish eight years of schooling because compared to women, a larger fraction of male wage earners include individuals with a primary school degree. On the other hand, a larger fraction of female wage earners have a high-school degree or higher; therefore, they are more likely to include compliers who are induced to finish high school by the reform. Thus, within the LATE interpretation of our estimates, a much higher fraction of female compliers are drawn from among high-school graduates. Therefore, potential differences in the returns to schooling across grade levels could make a substantial difference in the returns to schooling estimates for men and women; this issue is critical in understanding the differences in returns to schooling by gender.

6.3 Second Stage Results

Before we present our 2SLS estimation results on the effect of schooling on wages, we examine the effect of the reform—that is, our instrument—on wages. Figure 5 displays the mean log wages by birth year for men and women. There are no obvious discontinuities in wages around the cutoff birth-cohorts of 1986 and 1987. While there seems to be a jump for 18-year-old women, overall, the figure suggests that wages are not substantially affected by the policy. Given the remarkable effect of the policy on schooling outcomes presented in Figure 3, this result suggests that the effect of schooling on wages in Turkey is small.

Table 3 presents the OLS and 2SLS estimates of returns on schooling for men and women. Estimates with different methods of calculating the standard errors are given in panel (A). Our benchmark case is the cluster-robust standard errors, given as the second set of estimates in panel (A). However, we first present the White heteroskedasticity-robust standard errors, which do not take clustering into consideration, to gauge the potential importance of clustering. We carry out the parametric Moulton correction in the third set of estimates. Finally, in the last set

of estimates, we conduct non-parametric pairs-cluster bootstrap. In panel (B) of Table 3, we carry out certain methods which do not produce standard error estimates but instead provide p-values for the rejection of the null hypothesis (that the returns to schooling is zero). The first two of these are based on the cluster-robust estimates in panel (A) but instead use different rejection methods: T distribution with G-1 and G-2 degrees of freedom. The next three p-values come from wild cluster bootstrap estimations. In each one, a different distribution (Rademacher, Liu, and Mammen) is used in generating the auxiliary wild random variable; symmetric tests of the null hypothesis are carried out in all. Using the same set of wild-cluster bootstrap t-values, equaltailed tests of the null hypothesis are carried out in the last three rows.[15]

For men, the OLS estimate of returns to schooling is 0.05 and very precisely estimated. However, the 2SLS coefficient of returns to schooling for men is both statistically and economically insignificant; in fact, it is almost zero. Therefore, we do not carry out the alternative methods of Wald tests for men, which we present in panel (B) for the 2SLS estimates for women. The OLS estimate of returns to schooling (0.063) is higher for women than for men, and it is very precisely estimated. However, unlike for men, the 2SLS coefficient is statistically significant. The level of significance is at the 5-percent level with clustered standard errors—which also holds with T distribution with adjusted degrees of freedom as well, as can be seen from the first two rows of panel (B).[16] For the wild-cluster bootstrap case, the level of significance is at the 5-percent level, regardless of the distribution of the wild random variable,

_____

[15] Suppose that w(1), w(2),…,w(999) are the ordered values of the simulated t-ratios in bootstrap. The symmetric p-value is equal to the proportion of times that |w|>|w(i)|, i = 1,…,999, where w denotes the t-ratio with the original sample. In the equaltailed test of 5-percent level, for instance, the lower 2.5 percentile and the upper 97.5 percentile of the ordered w(1), w(2),…,w(999) are taken, then it is tested whether the t-ratio with the original sample falls in this interval.

[16] Stata uses a standard normal distribution for p-values of the *ivregress* command with cluster option.

when symmetric tests of the null hypothesis are carried out; when equaltailed tests of the null hypothesis are done, the level of significance is at the 5-percent level with two of the distributions for the wild random variable and at the 10-percent level with the remaining one. Also with the parametric Moulton correction, the statistical significance is at the 10-percent level. Only with the pairs-cluster bootstrap is the statistical significance not at the conventional levels; however, Cameron and Miller (2015) show that pairs-cluster bootstrap performs very poorly with few clusters. In terms of economic significance, an additional year of schooling increases women's wages by 3.8 percent.

Overall, our 2SLS estimate of the returns to schooling is barely above zero for men and 3.8 percent for women among 18 to 26-year-old individuals.[17] These estimates are significantly lower than those estimated for other developing countries (e.g., Duflo [2001] for Indonesia; Spohr [2003] for Taiwan; Fang et al. [2012] for China), but are in fact similar to those estimated for some European countries (e.g., Devereux and Hart [2010] for the United Kingdom; Pischke and von Wachter [2008] for Germany). In addition, the 2SLS coefficient of the returns on schooling is much lower than the OLS coefficient for both young men and women. However, this comparison may be misleading because since the 2SLS estimate is for the set of compliers only, the OLS and 2SLS estimates calculate average causal effects by assigning very different weights to causal effects conditional on schooling level. In particular, the OLS estimates assign higher weights to college graduates, as there are very few of them among the compliers of the 2SLS estimates. The next section investigates this issue further by imposing restrictions on the sample by schooling level in the estimation of average causal effects.

---

[17] With the inclusion of dummies for the 26 NUTS-2 level regions, where the data include years 2004 to 2013 only, the 2SLS estimate is zero for men and 3.0 percent for women (and statistically significant).

6.4 Second-Stage Results with Restrictions on Educational Attainment

We have already demonstrated that the policy affects schooling outcomes in different grade levels, both because of the length of the extension itself and because of the spillover effects of the policy on high-school grade levels. Moreover, the resulting changes in schooling distributions are quite different between men and women, as illustrated in Section 5.3. Therefore, in this section, we try to understand the variation in the returns to schooling at different schooling levels by placing upper bounds on educational attainment. The results are displayed in Table 6.

First, we exclude college graduates. Our analysis of schooling outcomes suggests that the policy has no effect on college graduation, although the data are relatively sparse for age groups that include college graduates. When the policy does not have a spillover effect on college graduation, the set of compliers in the sample does not change with the exclusion of college graduates from the sample. However, since college graduates are part of the control group when they are in the sample, to the degree that college graduates are different from others in terms of unobserved covariates, their inclusion or exclusion could have an important bearing on the estimated coefficients.

As can be seen from panel (A) in Table 4, with the exclusion of college graduates from the sample, the 2SLS coefficient for men increases to 0.014 and becomes statistically significant at the 10-percent level. The 2SLS coefficient for women rises to 0.071 (from 0.038 in Table 3), as well as its statistical significance (which is at the 1-percent level now). In other words, with an improved covariate balance, the 2SLS coefficient for women is substantially higher, and that for men is somewhat higher. At the same time, the OLS estimates with samples excluding college graduates are much lower for both men and women (0.027 for men as compared to 0.05 in Table 3, and 0.036 for women as compared to 0.063 in Table 3). In addition, the 2SLS

estimates now exceed the OLS estimates for women, whereas the 2SLS estimates are still lower for men.

Next, we exclude high-school graduates as well as college graduates. In this restricted sample, the compliers now include only those who are induced to finish grade 8 as a result of the policy. Note that these compliers include not only those who finish grade 8 with the policy but who would otherwise only finish grade 5, but also those who finish grade 8 with the policy but who would otherwise only finish any grade level from 0 to 7.[18] Hence, the average causal effect here is a weighted average of the average causal effect at one to eight years of schooling; nonetheless, most of the weight comes from grades 6 to 8. As can be seen in panel (B) of Table 4, with this sample, there is no evidence that returns to schooling are different from zero either for men or for women. At the same time, the magnitude of the 2SLS coefficient is still substantially larger for women than for men.

Overall, Table 4 indicates that the returns to an extra year of schooling for high-school grade levels are higher than the average returns to an extra year schooling for grade levels 1 to 8 both for men and for women—but particularly for women. This finding, along with the difference between genders in the change in schooling distribution (which was illustrated in Section 5.3) is critical to interpreting the difference in the estimated returns to schooling by gender.

6.5 Policy Effects on Wage Employment

The validity of our instrument in establishing the causal link between schooling and wages requires that the instrument have no effect on wages other than through its effect on schooling.

---

[18] Although the policy made grades 6 to 8 compulsory, with the falling costs of enrollment in all primary school grade levels via bussing and the construction of boarding schools, enrollment in grades 1 through 8 could change. In fact, Kirdar et al. (2014) provide evidence for rising enrollment in grades 1 through 3 in certain areas.

However, the policy could also change the wage-earner status of individuals, thereby affecting the distribution of observed wages. To understand this issue, we examine the effects of the policy on wage employment outcomes in this section. Here, wage employment is defined as full-time employment as a wage or salary worker, as in the analysis of wages.

Table 5 presents the estimates of the effect of the policy on wage employment status by gender while controlling for age and year effects. As can be seen from the table, the policy has no effect on the wage employment status of men. On the other hand, the policy increases the wage employment probability of women. According to our preferred set of estimates in column (4), the wage employment probability increases by 1.8 percentage points for women as a result of the policy. This is a substantial increase given that the wage employment probability for women in this age bracket is just above 14 percent. Appendix Table A2 shows the robustness of this inference on the policy effect on women's wage employment using different methods of estimating the standard errors. While the finding in Table 5 is robust to Moulton correction and to using t distributions with alternative degrees of freedom, the results with wild cluster bootstrap is mixed. We fail to reject the null hypothesis with symmetric tests for all three different ways of generating the wild random variable, whereas we reject the null hypothesis with equaltailed tests at the 10-percent level with two of the three distributions used in generating the auxiliary wild random variable.[19]

These findings imply that we need to be careful in interpreting the findings for women because the education reform not only increases the education of women who would be employed regardless of the policy, but it also pushes other women who would otherwise not be employed

_____

[19] We do not present the results of pairs cluster bootstrap because several replications cannot be estimated. Cameron and Miller (2015) also report a similar problem in their replications with pairs cluster bootstrap, which they attribute to the lack of variation in treatment in the bootstrap samples.

into wage employment. Consequently, the pool of employed women who are affected by the policy might be different in ways that pertain to their wages from employed women who are not affected by the policy. However, the women who are new entrants to employment with the policy would be those who were on the margin of employment before the policy; therefore, they are not likely to be substantially different from the average employed woman before the policy in terms of potential wages.

7. Interpretation

In this section, we examine potential causes of the low estimated returns to schooling in Turkey as well as the reasons for the higher estimated returns for women. The low returns to schooling in Turkey might be surprising given that several previous studies report high returns in developing countries (Patrinos and Psacharopoulos, 2010), including some that use plausibly exogenous sources of variation in schooling to estimate the causal link (Duflo, 2001). There are three key reasons for the low returns on schooling estimate in our study.

First, for the grade levels at which the policy directly affects individuals (grades 6 through 8), the wage–schooling locus is very flat, as can be seen in Figure A1 in the Appendix, where the wage premium from completing various levels of schooling is calculated by regressing (for men and women separately) the logarithm of the wage rate on dummies for educational attainment, age, year, and urban status.[20] The sample is restricted to 18 to 26-year-old individuals who were born before 1986 in order to limit the sample to individuals who are not affected by the policy. For men, a secondary school diploma increases the wage rate by 6%, compared to 14% for a high-school diploma and 49% for a college degree. For women, a secondary school diploma

---

[20] Here, the inclusion of cohort controls makes virtually no change.

increases the wage rate by 6%, compared to 21% for a high-school diploma and 48% for a college degree (these denote marginal increases from the lower education category). Obviously, the 6-percent increase in wages both for men and for women resulting from the completion of secondary school is not the causal effect of additional three years of schooling. Nevertheless, the very low premiums from completing secondary school suggest that the causal returns to schooling might not be high either. In fact, our findings in the previous section reveal that with a sample restricted to individuals with less than a high-school degree (where compliers include only those who are induced to complete grade 8 [secondary school under the old system] as a result of the policy), there is no evidence that returns to schooling are different from zero for either men or women.

Second, when we compare above the earnings of the groups with five and with eight years of schooling before the policy based on Figure A1, both the productivity effects of three more years of schooling and the sheepskin effect of earning a secondary school diploma have effects. On the other hand, in the policy experiment we conduct, the individuals who are induced to complete eight years of schooling with the policy but who would otherwise complete five years hold a primary school diploma both before and after the policy; it is only the years of schooling that is different for them. Therefore, we measure only the productivity effects of three more years of schooling.[21]

Third, our returns on schooling estimate is for young individuals: 18 to 26-year-old workers. Since returns on schooling increase by age in the typical age-earnings profile, we observe them

---

[21] For only those who are induced to complete eight years of schooling with the policy but would not finish five years of schooling without it and those who are induced to finish high school with the policy but who would not otherwise, we measure both the productivity and sheepskin effects of schooling. However, since these groups are smaller than the former group, we are mostly measuring the productivity effects of additional schooling.

in the part of profile in which the returns to schooling are lower. In fact, as can be seen in Figure A2 in the Appendix, for both men and women, the age-earnings profile varies much less across different education groups in the beginning of the work lifecycle, and there is less variation across men than across women.

As reviewed earlier, the compulsory schooling reform did not have an obvious negative impact on standard measures of schooling quality, such as student-to-teacher, student-to-classroom ratios or international test scores. However, it is important to note that the low returns to primary and secondary education in Turkey could also reflect the quality of schooling if school quality was already low before the reform. While school quality is a potential factor for low returns, for Indonesia, where schooling quality is worse than that in Turkey according to international math and science tests,[22] Duflo (2001) finds quite high returns to schooling. Moreover, for England and Germany, where the results of international tests indicate a better quality of schooling, Devereux and Hart (2010) and Pischke and von Wachter (2008), respectively, find quite low returns. Therefore, rather than the quality of schooling in terms of improving the cognitive skills of students, it might be the labor market relevance of the skills acquired in the grade levels at which students change their behavior as a result of a particular policy, which determines the returns to schooling, as Pischke and von Wachter (2008) argue.

Our other key finding is that the return on schooling is higher for women than for men. This finding is consistent with the implication of Figure A1 in the Appendix. In the age-earnings profiles by gender, the variation in wages across education groups is wider for women than for men among 18 to 26-year-old individuals. A potential reason for higher estimated returns in Turkey for women is that since the female labor force participation rate in Turkey is much lower, women in the labor market form a much more select group in terms of average ability

---

[22] National Center for Education Statistics, Trends in International Mathematics and Science Study.

conditional on years of schooling than men who are in the labor market. Nonetheless, an investigation of the characteristics of the compliers among men and women gives us a much stronger argument as to why there are gender differences in the returns to schooling in Turkey.

The difference in the estimated average causal effects between men and women could be the result of differential weightings of the causal effects conditional on years of schooling. As shown earlier, the policy increased the completion of not only grade 8 but also high school. Moreover, as displayed in Figure 4, the rise in the fraction completing high school or a higher schooling level compared to the rise in the fraction completing grade 8 or a higher schooling level is much higher among women than men. In other words, a much larger fraction of female compliers complete high-school grade levels. At the same time, our findings on the returns to schooling with restrictions on educational attainment in Section 6.4 suggest that the returns to schooling for an extra year in high-school grade levels are higher than the average returns for an extra year in primary school level. In essence, the returns to schooling are higher for high-school grade levels than for earlier grade levels and the female compliers include a higher share of those who completed high-school grade levels, resulting in a higher returns to schooling estimate for women in our sample.

8. Conclusion

In this paper, we estimated the causal wage returns on schooling for young men and women in Turkey and thus contribute to the small body of literature on the causal monetary returns of education in developing countries. For this purpose, we used the exogenous variation in schooling brought about by the 1997 reform in compulsory schooling. The features of this reform make it very close to being an ideal instrument. The reform brought about a tremendous

change in schooling outcomes because of both the length of the extension and the high fraction of students who drop out after completing compulsory education.

Our empirical analysis fails to reject that the returns to schooling for 18 to 26-year-old men are zero despite our large sample size; in fact, the coefficient estimates are close to zero. On the other hand, for women of the same age group, we estimate that an extra year of schooling increases wages by 3.8 percent. An important feature of these local average treatment effect estimates is that they come close to the average treatment effect estimates because the policy affects the schooling outcomes of a large section of the population.

Our estimates of the returns to schooling are relatively low for three primary reasons. First, in the grade levels at which the policy changed the schooling distribution, the returns to schooling are in fact low. Our estimates suggest that in grades 6 through 8, the average returns to schooling are close to zero for both men and women; they are higher for the high-school grade levels, particularly for women. Second, while the policy mandated three more years of schooling, a new diploma was not awarded for its completion. Therefore, there were no new sheepskin effects, only the productivity effects of extra schooling for many compliers. Third, our sample includes young individuals for whom the differences across education groups in the age-earnings profile are typically smaller.

The returns to schooling are higher for women than for men in our context because the policy changes the schooling distributions of men and women differently. Our estimate, which is an average of the causal returns at different grade levels, puts more weight on high-school grade levels for women than for men, and the returns to an extra year of schooling for high-school grade levels are higher than the returns to an extra year for primary education. In other words, female compliers are more likely to come from those who are induced to finish high school as a result of the policy. This is because women's employment probability increases substantially

with education. Thus, the women in our sample of wage earners have higher education on average than the men in our sample.

In our analysis, conditional on schooling attainment, women are a more select group because of the much lower employment rate for women than for men in Turkey. This would be another reason for the higher returns to schooling for women if selection into employment is associated with ability. This issue also highlights a potential limitation of our estimate for women: we estimate the returns to schooling for a select group who are wage earners. However, this group is actually relevant for our purpose even though it is a selected group because the women who are on the margin of becoming wage earners would be similar to the women in our sample. Another potential contaminating factor in our returns to schooling estimate for women is that the policy also increases the probability of wage employment among women. This would be problematic if the women who become wage earners as a result of the policy are different from those who are already wage earners in ways related to their earnings potential. However, again, since the new entrants to employment with the policy were on the margin of employment before the policy, they are not likely to be substantially different from those who were already employed before the policy in terms of earnings potential.

References

Acemoğlu, D. and J. Angrist (2000) "How large are human-capital externalities? Evidence from compulsory schooling laws" In: Bernanke, B.S. and K. Rogoff, eds., NBER Macroeconomics Annual, Vol. 15. (Cambridge, Massachusetts), pp. 9-+.

Angrist, J. and A. Krueger (1991) "Does compulsory school attendance affect schooling and earnings?" Quarterly Journal of Economics 106: 979-1014.

Angrist J. and Pischke, J.S. (2009). "Mostly Harmless Econometrics". Princeton, NJ: Princeton University Press.

Becker, S., and F. Siebern-Thomas (2001) "Returns to education in Germany: A variable treatment intensity approach." EUI working paper ECO 2001/09

Bell, R.M. and McCaffrey, D.F. (2002): "Bias Reduction in Standard Errors for Linear Regression with Multistage Samples", Survey Methodology 28, 169-181.

Bound, J., Jaeger, D., and R. Baker (1995) "Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variables is weak." Journal of the American Statistical Association 90: 443-450.

Bound, J., and D.A. Jaeger (1996) "On the validity of season of birth as an instrument in wage equations: A comment on Angrist and Krueger's 'Does compulsory school attendance affect schooling and earnings?' " Working paper no. 5835.(NBER, Cambridge, MA).

Cameron C., Gelbach J. and D. Miller (2008). "Bootstrap-Based Improvements for Inference with Clustered Errors," Review of Economics and Statistics 90: 414-427.

Cameron C. and D. Miller (2015). "A Practitioner's Guide to Cluster-Robust Inference," Journal of Human Resources, forthcoming.

Card, D. (1995) "Using geographic variation in college proximity to estimate the return to schooling." In: Louis N. Christofides, E. Kenneth Grant, and Robert Swidinsky, eds., Aspects of labour market behaviour: Essays in honour of John Vanderkamp (University of Toronto Press, Toronto, Canada) pp. 201-222.

Card, D. (1999) "The causal effect of education on earnings." in: O. Ashenfelter and D. Card, ed., Handbook of Labor Economics, Vol. 3 (Elsevier).

Conneely, K. and R. Uusitalo (1997) "Estimating heterogeneous treatment effects in the Becker schooling model." Unpublished discussion paper (Industrial Relations Section, Princeton University).

Davidson R. and J.G. MacKinnon (2010). "Wild bootstrap tests for IV regression," *Journal of Business and Economic Statistics* 28: 128-144.

Dayıoğlu, M. and M. Kırdar (2009) "Determinants of and trends in labor force participation of women in Turkey," State Planning Organization of the Republic of Turkey and World Bank, Welfare and Social Policy Analytical Work Program, Working Paper Number 5.

Devereux, P.J. and R.A. Hart (2010) "Forced to be rich? Returns to compulsory schooling in Britain." Economic Journal, December.

Devereux, P.J. and W. Fan (2011) "Earnings returns to the British education expansion," Economics of Education Review 30: 1153-1166.

Duflo, E. (2001) "Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment." American Economic Review 91(4): 795-813.

Fang, H., Eggleston K., Rizzo J.A., Rozelle S., and R. Zeckhauser (2012) "The returns to education in China: Evidence from the 1986 compulsory education law," NBER Working Paper 18189.

Gelbach J.B., Klick, J. and T. Stratman (2009). "Cheap Donuts and Expensive Broccoli: The Effect of Relative Prices on Obesity," mimeo.

Griliches, Z. (1977) "Estimating the returns to schooling—some econometric problems," Econometrica 45: 1-22.

Harmon, C. and I. Walker (1995) "Estimates of the economic return to schooling for the United Kingdom." American Economic Review 85: 1278-1286.

Heckman J.J., Lochner L.J., and P.E. Todd (2007) "Earnings functions, rates of return and treatment effects: The Mincer equation and beyond." in: *Handbook of Education Economics*, Vol. 1, ed. E. Hanushek and F. Welch, Elsevier.

Ichino, A. and R. Winter-Ebmer (1999)"Lower and upper bounds of returns to schooling: An exercise in IV estimation with different instruments," European Economic Review, Elsevier, 43(4-6): 889-901, April.

Ichino, Andrea and Rudolf Winter-Ebmer (2004) "The long-run educational cost of World War II: An application of local average treatment effect estimation." Journal of Labor Economics, University of Chicago Press 22(1): 57-86, January.

Imbens, G.W. and J.D. Angrist (1994) "Identification and estimation of local average treatment effects." Econometrica 62(2): 467-75.Kane, T. and C.E. Rouse (1995) "Labor market returns to two- and four-year colleges: Is a credit a credit and do degrees matter?" American Economic Review 85(3): 600-14.

Kırdar, M.G., Dayıoğlu-Tayfur, M., and I. Koc (2014) "Does longer compulsory education equalize schooling by gender and rural/urban residence?" IZA Discussion Papers 7939, Institute for the Study of Labor (IZA).

Maluccio, J. (1997) "Endogeneity of schooling in the wage function." Unpublished manuscript (Department of Economics, Yale University).

MacKinnon J.G. and M.D. Webb (2013) "Wild bootstrap inference for wildly different cluster sizes", QED Working Paper No. 1314.

MEB (2007), TIMSS Uluslararasi matematik ve fen eğilimleri araştirmasi, TIMMS 2007, Ulusal matematik ve fen raporu, 8. Siniflar (TIMSS National Report, 2007), Ankara.

Moulton B.R. (1986), "Random Group Effects and the Precision of Regression Estimates", Journal of Econometrics 32, 385-397.

National Center for Education Statistics. Trends in International mathematics and science study (TIMMS). http://nces.ed.gov/timss/table11_3.asp. Accessed August 2014.

Oreopoulos P. (2006) "Estimating average and local average treatment effects of education when compulsory schooling laws really matter." American Economic Review 96: 152-175.

Patrinos, H.A. and G. Psacharopoulos (2010) "Returns to education in developing Countries." In: Economics of Education (eds. D.J. Brewer and P.J. McEwan), pp. 44-51, Elsevier, San Diego.

Pischke, J.-S., and T. von Wachter (2008) "Zero returns to compulsory schooling in Germany: Evidence and interpretation." Review of Economics and Statistics 90(3): 592-598.

Spohr, C.A. (2003) "Formal schooling and workforce participation in a rapidly developing economy: Evidence from 'compulsory' junior high school in Taiwan." Journal of Development Economics 70(2): 291-327.

Staiger, D. and J.H. Stock (1997) "Instrumental variables regression with weak instruments." Econometrica 65: 557-586.

Stephens M. and D. Yang (2014) "Compulsory Education and the Benefits of Schooling," American Economic Review 104(6): 1777-1792.

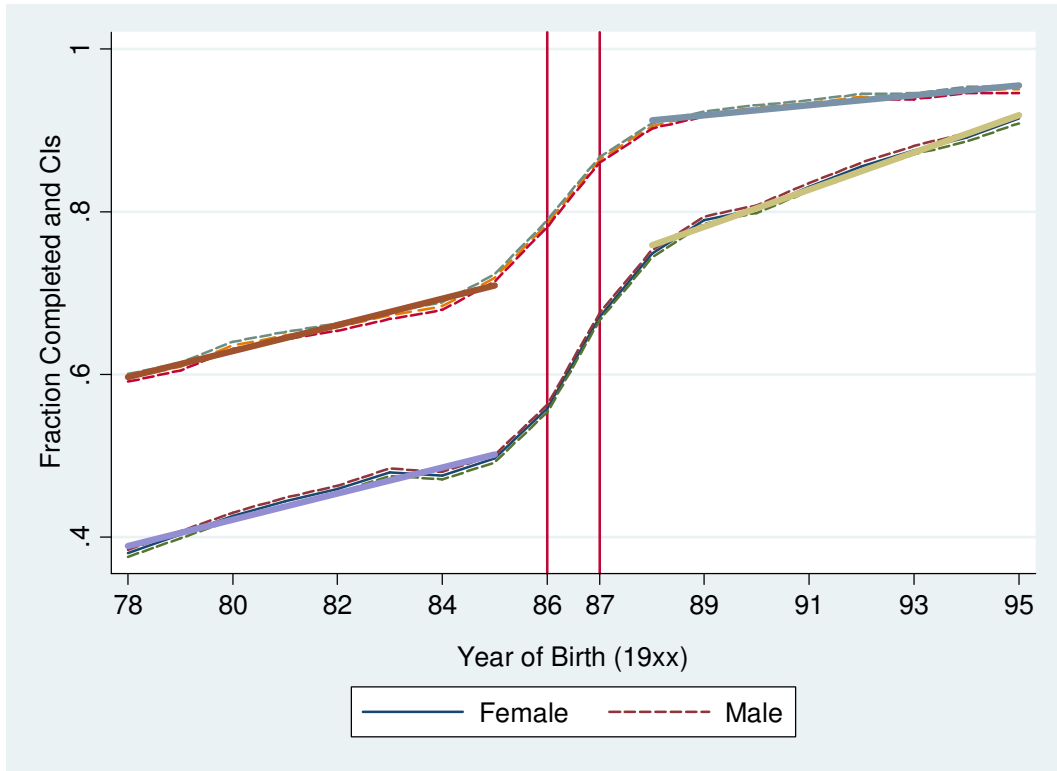**Figure 1: Fraction Completing at Least Grade 8**
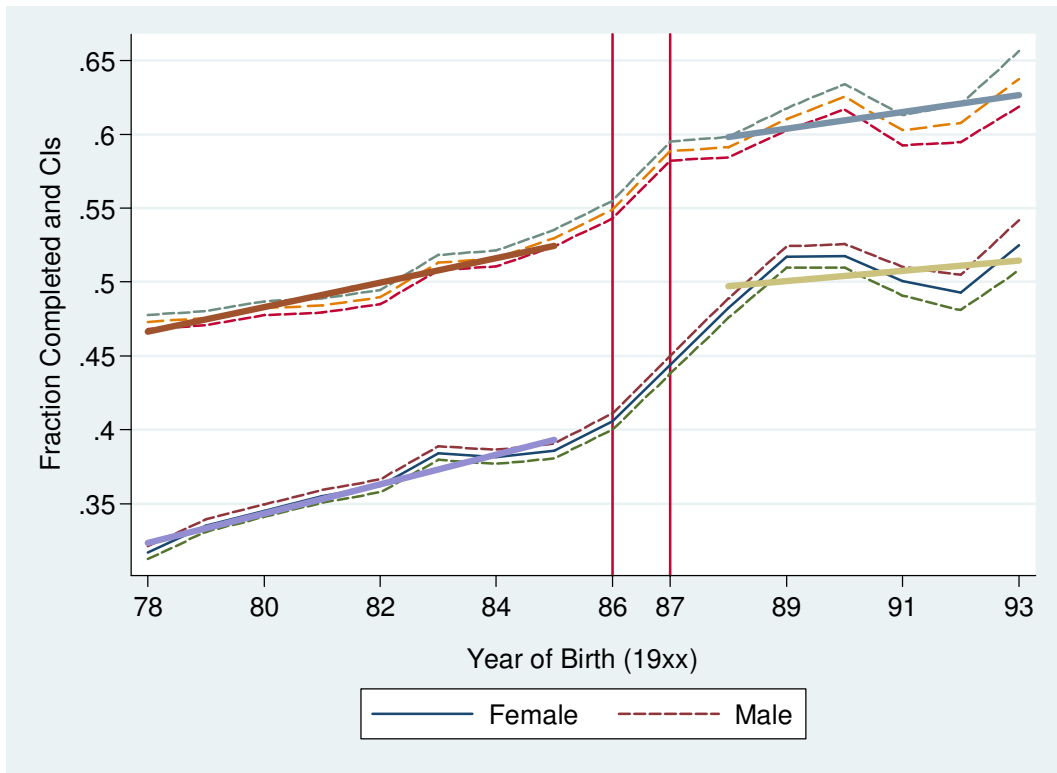
Figure 2: Fraction Completing at Least High School

Figure 3: Mean Years of Schooling by Birth Year and Gender
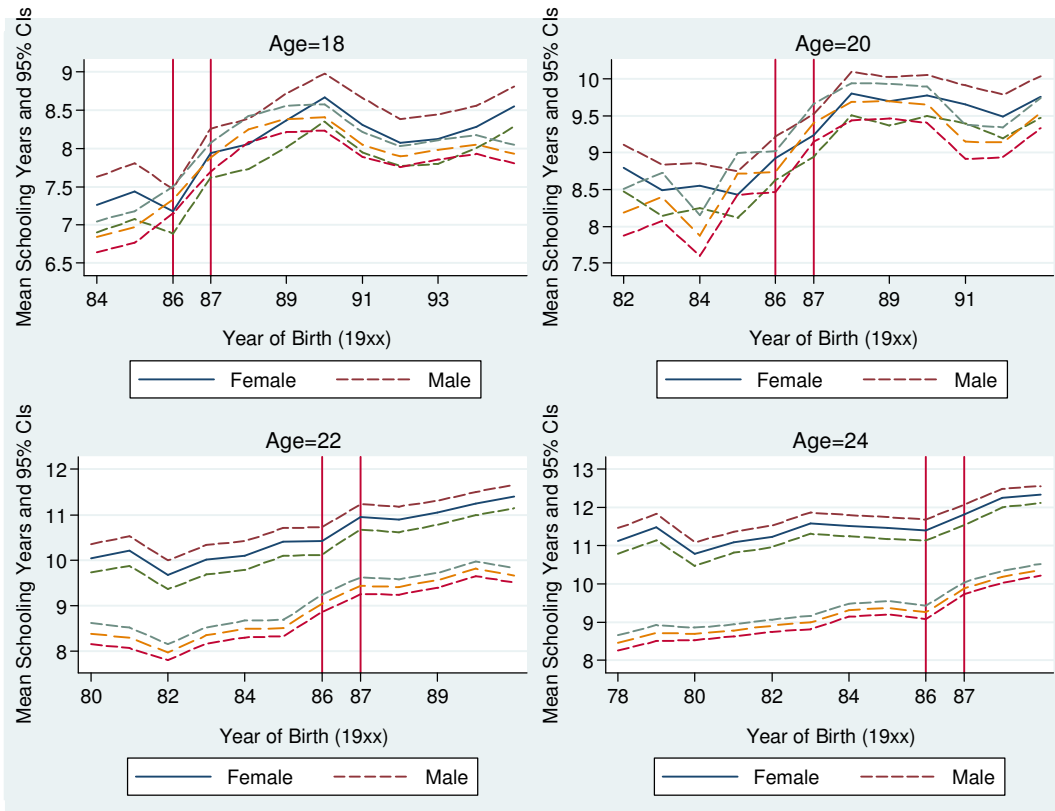
Notes: The sample includes fulltime wage earners.

Figure 4: Difference in Schooling Distributions before and after the Policy
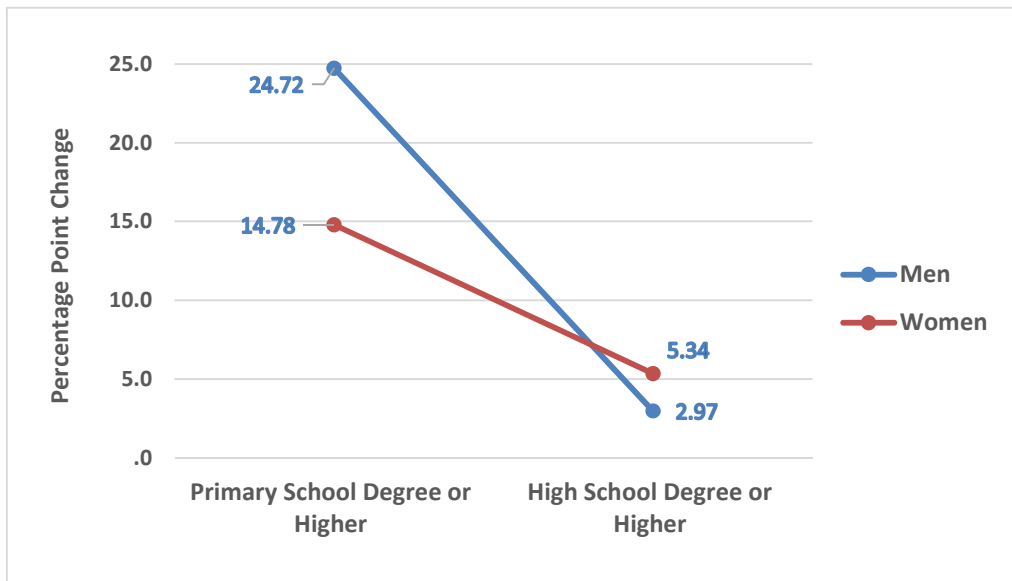
Notes: The distributions are one minus cumulative distributions. The sample includes 18 to 26-year-old fulltime wage earners. The distribution before the policy is for the 1984 and 1985 birth-cohorts, and that after the policy is for the 1988 and 1989 birth-cohorts.

Figure 5: Mean Log Wages by Birth Year and Gender
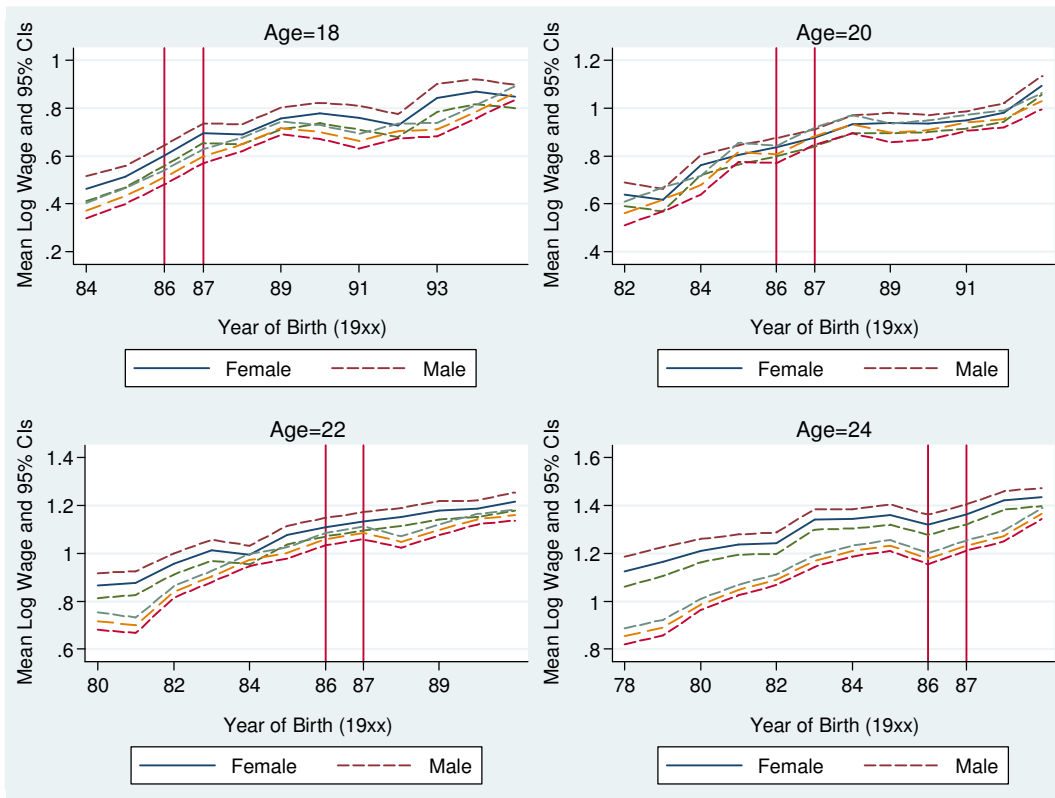
Notes: The sample includes fulltime wage earners.

Table 1: Birth Year (and Policy Dummy) by Age and Year

						Year						
	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Age												
15	*1987*	**1988**	**1989**	**1990**	**1991**	**1992**	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**
16	*1986*	*1987*	**1988**	**1989**	**1990**	**1991**	**1992**	**1993**	**1994**	**1995**	**1996**	**1997**
17	1985	*1986*	***1987***	**1988**	**1989**	**1990**	**1991**	**1992**	**1993**	**1994**	**1995**	**1996**
18	1984	1985	*1986*	***1987***	**1988**	**1989**	**1990**	**1991**	**1992**	**1993**	**1994**	**1995**
19	1983	1984	1985	*1986*	***1987***	**1988**	**1989**	**1990**	**1991**	**1992**	**1993**	**1994**
20	1982	1983	1984	1985	*1986*	***1987***	**1988**	**1989**	**1990**	**1991**	**1992**	**1993**
21	1981	1982	1983	1984	1985	*1986*	***1987***	**1988**	**1989**	**1990**	**1991**	**1992**
22	1980	1981	1982	1983	1984	1985	*1986*	***1987***	**1988**	**1989**	**1990**	**1991**
23	1979	1980	1981	1982	1983	1984	1985	*1986*	***1987***	**1988**	**1989**	**1990**
24	1978	1979	1980	1981	1982	1983	1984	1985	*1986*	***1987***	**1988**	**1989**
25	1977	1978	1979	1980	1981	1982	1983	1984	1985	*1986*	***1987***	**1988**
26	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	*1986*	***1987***
27	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	*1986*

Notes: (i) Year of birth is displayed in bold when the policy dummy takes a value of one. Transition years (1986 and 1987) are also displayed in italics. (ii) Only ages 16 to 26 include both birth cohorts that are affected by the policy and those that are not. (iii) Only ages 17 to 25 include both birth cohorts that are affected by the policy and those that are not when the 1986 and 1987 birth cohorts are excluded because of fuzzy policy status.

Table 2: First-Stage Results for Men and Women

	Men		Women	
Dependent Variable: Years of Schooling				
	Sample A	Sample B	Sample A	Sample B
	(1)	(2)	(3)	(4)
Policy Dummy	0.583***	0.955***	0.441***	0.615***
	[0.034]	[0.049]	[0.053]	[0.082]
Observations	134,113	91,756	58,835	41,562
R-squared	0.045	0.048	0.120	0.118
F-Statistic	17.8	54.1	21.0	67.3

Notes: Sample A includes 18 to 26-year-old individuals in 2002 to 2013 surveys, whereas sample B includes 18 to 25-year-old individuals in 2002 to 2013 surveys but excludes the 1986 and 1987 birth-cohorts. All models also include age and year dummies and a dummy for urban status. The F-statistic is adjusted for 20 clusters in columns (1) and (3) and for 17 clusters in columns (2) and (4). Statistically significant: *** 1 percent level; ** 5 percent level, * 10 percent level.

Table 3: Estimates of Returns to Schooling

A) Estimates with different ways of calculating the standard errors

	Men		Women	
	OLS	2SLS	OLS	2SLS
1) White Heteroskedasticity-Robust SE	0.050***	0.005	0.063***	0.038**
	[0.000]	[0.008]	[0.001]	[0.016]
2) Cluster-Robust SE	0.050***	0.005	0.063***	0.038**
	[0.002]	[0.014]	[0.003]	[0.015]
3) Moulton-Factor Adjusted SE	0.050***	0.005	0.063***	0.038*
	[0.000]	[0.019]	[0.001]	[0.022]
4) Pairs Cluster Bootstrap	0.050***	0.005	0.063***	0.038
	[0.002]	[1.092]	[0.003]	[0.029]
Observations	134,113	134,113	58,835	58,835
R-squared	0.277	0.199	0.338	0.316

B) p-values with different rejection methods: 2SLS estimates for women

1) Cluster on birth-year, T(G-1) for critical value	0.022
2) Cluster on birth-year, T(G-2) for critical value	0.022
3) Wild cluster bootstrap: wild random variable: Liu, symmetric test	0.030
4) Wild cluster bootstrap: wild random variable: Rademacher, symmetric test	0.035
5) Wild cluster bootstrap: wild random variable: Mammen, symmetric test	0.061
6) Wild cluster bootstrap: wild random variable: Liu, equaltailed test	reject at 5% level
7) Wild cluster bootstrap: wild random variable: Rademacher, equaltailed test	reject at 5% level
8) Wild cluster bootstrap: wild random variable: Mammen, equaltailed test	reject at 5% level

Notes: All models also include dummies for calendar year and age and a dummy for urban status. In panel (A2), clustering is done at the birth-year level. The number of clusters (G) is 20. In cluster-robust standard errors in panel (A2), Stata uses a standard normal distribution for p-values with "ivregress" command. Statistically significant: *** 1 percent level; ** 5 percent level, * 10 percent level.

Table 4: Estimates of Returns to Schooling by Education Level

	Men		Women	
	OLS	2SLS	OLS	2SLS
A) Less Than College				
Years of Schooling	0.027***	0.014*	0.036***	0.071***
	[0.002]	[0.007]	[0.003]	[0.026]
Observations	116,933	116,933	41,665	41,665
R-squared	0.189	0.182	0.165	0.119
B) Less Than High School				
Years of Schooling	0.013***	-0.003	0.014***	0.031
	[0.002]	[0.008]	[0.003]	[0.059]
Observations	70,410	70,410	18,455	18,455
R-squared	0.178	0.173	0.114	0.106

Notes: All models also include dummies for calendar year and age and a dummy for urban status. Standard errors are clustered at the year-of-birth level. The number of clusters is 20. Statistically significant: *** 1 percent level; ** 5 percent level, * 10 percent level.

Table 5: Policy Effects on Wage Employment

	Men		Women	
	Sample A	Sample B	Sample A	Sample B
	(1)	(2)	(3)	(4)
Policy	0.000	0.001	0.008	0.018*
	[0.007]	[0.010]	[0.007]	[0.009]
Observations	353,740	258,149	408,999	299,238
R-squared	0.083	0.071	0.022	0.021

Notes: Sample A includes all 18- to 26-year-old individuals, whereas sample B excludes the 1986 and 1987 birth-cohorts and includes 18- to 25-year-old individuals (because when the 1987 birth-cohort is excluded, there is no variation in the policy variable for 26 year-olds). The reason for the exclusion of the 1986 and 1987 birth cohorts is that the policy dummy variable is fuzzy for these birth-cohorts. Wage employment takes a value of 1 for wage earners who are fulltime workers and 0 otherwise. All models include dummies for each year and age and a dummy for urban status. Standard errors are clustered at birth-year level. Statistically significant: *** 1 percent level; ** 5 percent level, * 10 percent level.
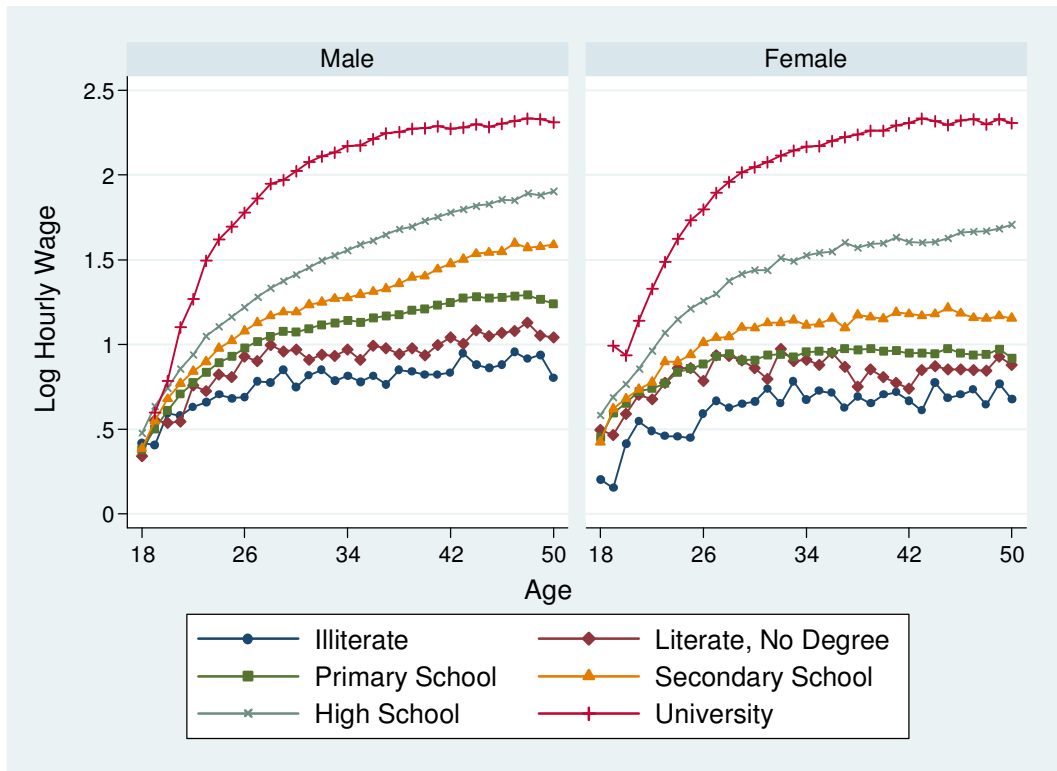
APPENDIX

Figure A1: Wage-Schooling Locus by Gender
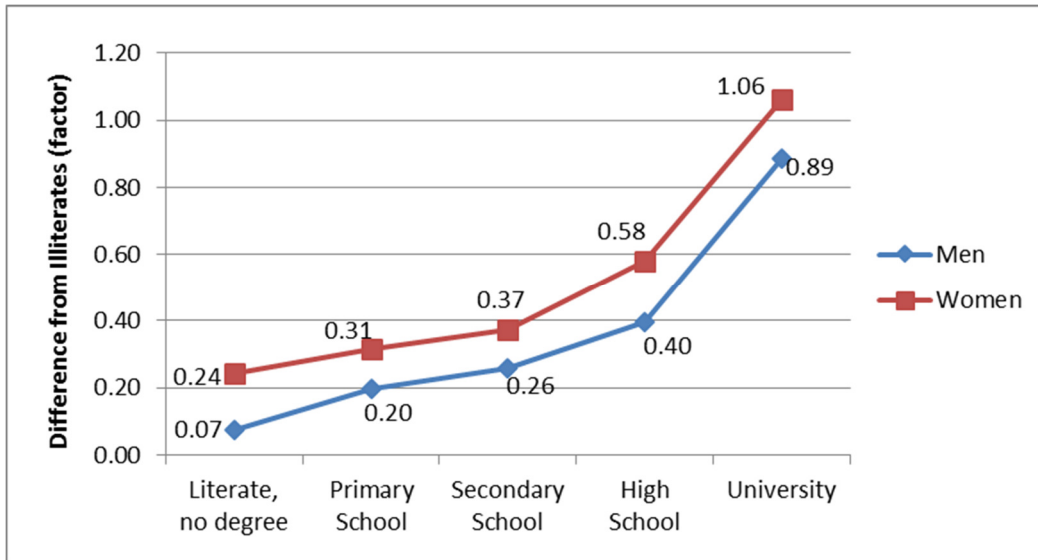
Notes: The data include 18 to 26-year-old fulltime wage earners who were born before 1986 (i.e., those who were not affected by the policy).

Figure A2: Age Profile of Hourly Wages by Gender

Notes: The data include fulltime wage earners who were born before 1986 (i.e., those who were not affected by the policy).

Table A1: Robustness Check - Policy Effects on Wage Employment

A) OLS estimates with different ways of calculating the standard errors			
	Robust SE	Clustered SE	Moulton Correction
Policy	0.018***	0.018*	0.018**
	[0.003]	[0.009]	[0.009]
Observations	299,238	299,238	299,238
R-squared	0.021	0.021	0.021

B) p-values of the null hypothesis (slope=0) with different rejection methods	
Cluster on birth-year, T(G-1) for critical value	0.057
Cluster on birth-year, T(G-2) for critical value	0.059
Wild cluster bootstrap: wild random variable: Liu, symmetric test	0.144
Wild cluster bootstrap: wild random variable: Rademacher, symmetric test	0.235
Wild cluster bootstrap: wild random variable: Mammen, symmetric test	0.201
Wild cluster bootstrap: wild random variable: Liu, equaltailed test	reject at 0.1
Wild cluster bootstrap: wild random variable: Rademacher, equaltailed test	fail to reject at 0.1
Wild cluster bootstrap: wild random variable: Mammen, equaltailed test	reject at 0.1

Notes: The sample includes all 18 to 25-year-olds in the 2002-2013 surveys excluding the 1986 and 1987 birth cohorts. All models also include dummies for calendar year and age and a dummy for urban status. Clustering is done at the year-of-birth level. The number of clusters (G) is 17. Statistically significant: *** 1 percent level; ** 5 percent level, * 10 percent level.