The Fertility Effect of Education: Regression Discontinuity for Counts and Exponential Models*

Kamhon Kan
Institute of Economics
Academia Sinica
Taipei, Taiwan, R.O.C.

Myoung-jae Lee
Department of Economics
Korea University
Seoul, South Korea;

Research School of Economics
Australian National University
Canberra, Australia

Abstract

Negative relations between women's schooling and fertility have been observed by social scientists, leading some of them to propose education as a policy tool for family planning. Both fertility and schooling are, however, choice variables that may be affected by some common factors. This would then make schooling endogenous, and consequently, the observed negative relationships might not be causal. We take advantage of a compulsory education law change to estimate the impact of female education on fertility, using regression discontinuity estimators motivated by exponential regression models appropriate for counts or non-negative response variables. Our data utilizes the entire 1990 Population Census of Taiwan. We find that married women's schooling has no impact on their number of children and the age at first-birth.

Keywords: Fertility, Education, Compulsory education reform, Regression discontinuity estimators, Counts and non-negative variables

JEL codes: J13, C21

*We thanks participants in the *2010 Asian Conference on Applied Micro-Economics/Econometrics* for comments and suggestions. We also thank Ruei-Hua Wang and Chai-Ling Lee for able research assistance.

1 Introduction

It has been speculated that education has effects on social and economic outcomes, and these effects are important concerns to policy-makers and social scientists. Economists' emphasis has traditionally been on the effect of education on economic growth (Lucas 1988 and Romer 1990), and cross-country income disparity was attributed to differences in human capital (e.g., Mankiw, Romer and Weil 1992). Another strand of economic research examined the private return to education and found that the return per year is around 6–10% (Card 1999). One worker's education was also found to have a spillover effect on other workers' productivity (Acemoglu and Angrist 2000, Moretti 2004, Iranzo and Peri 2009, and Mas and Moretti 2009).

More recently, economists explored education's impact on non-market outcomes. It was found that education has a positive effect on health (Lleras-Muney 2005 and Oreopoulos 2007), a negative effect on the participation in crime (Lochner and Moretti 2004), and a positive effect on civic engagement and attitude (Dee 2004, and Milligan, Moretti and Oreopoulos 2004). Non-market outcomes were also found to be related to parental education, and beneficial effects were found (see Currie and Moretti 2003, and Oreopoulos, Page and Stevens 2006). For a recent survey on the impact of education on individual non-market outcomes, see Grossman (2006).

In the relevant literature, an important issue is the effects of education on demographic changes, which play a central role in economic and social developments (Caldwell 1980, Portes 2006, and Cohen 2008). Underlying the linkage between education and demographic changes is a negative relationship between a woman's years of schooling and the number of children she has. There are several channels through which female schooling leads to a decline in fertility. For example, education raises a woman's value of time due to an increase in market wage and opportunities, education enables women to make better use of contraceptive devices, and better educated women have more autonomy and bargaining power in fertility decisions. See, e.g., Michael (1973), Rosenzweig and Seiver (1982) and Rosenzweig and Schultz (1985).

There are numerous studies documenting a negative relationship between female schooling and fertility (e.g., Sander 1992, Schultz 1997, Lam and Duryea 1999, Lloyd, Kaufman and Hewett 2000, Axinn and Barber 2001, and Sathar *et al.* 2003). However, many of these estimates are not "causal", as the endogeneity of schooling is not accounted for. A woman's fertility choice and schooling choice are jointly determined, and are part of her lifetime time allocation plan (Becker 1981). Even if the two choices are not simultaneously determined, they are likely to be affected by common unobserved factors, such as parental influences and social norms. Without taking into account the simultaneity and unobserved factors between fertility and schooling decisions, most early studies are not informative about the causal relationship. In view of this problem, recent studies used instrumental variables for schooling; e.g., León (2004), Black *et al.* (2008), Osili and Long (2008), Monstad *et al.* (2008) and McCrary and Royer (2010). These studies reviewed in the following obtained mixed results.

León (2004) took advantage of an exogenous variation in schooling years created by compulsory education laws and child labor laws in the U.S. Using the 1950–1990 U.S. census data, it was found that schooling years has a coefficient of -0.33, implying that a woman with three more years of schooling will have one less child. Black *et al.* (2008) looked at the effect of changes in women's education (induced by compulsory education legislations) on the timing of fertility in the U.S. and Norway. The compulsory education legislations were found to have negative and statistically significant effects on having the first-birth before the age of 18, 19, and 20 (but not 16 and 17) in the U.S. and before the age of 18 and 20 (but not 16, 17, and 19) in Norway. These effects are not large in magnitude though.

Osili and Long (2008) also used an education reform to instrument education. Exploiting the introduction of universal primary education in Nigeria in 1976 and using survey data, it was found that increasing female education by one year reduces early fertility (i.e., children born before age 25) by 0.6 births. The Norwegian extension of compulsory education was also exploited by Monstad *et al.* (2008). Using two-stage least squares with the timing of the reform in each municipality as an instrument, Monstad *et al.* (2008) obtained results indicating that female education has no impact on the number of children, but a positive effect on the age of first-birth.

Policies on age-at-school-entry were used by McCrary and Royer (2010) to identify the effect of female education on fertility and infant health. Based on natality data of mothers 23 or younger in Texas (1989–2001) and California (1989–2002), the study found that age-at-school-entry policies have a statistically significant effect on education. Four percent of California young women and six percent of Texas young women's high school completion were affected by age-at-school-entry policies: they would complete high school if their birth date within the birth year falls before the school entry date and would fail to do so otherwise. However, the age-at-school-entry policies did not have any effect on the probability of motherhood and a very small effect on infant health. This indicates that there is no causal relationship between education and fertility outcomes.

All these papers used compulsory education laws/policies to identify the effect of education on fertility to obtain rather mixed results. It seems that the effect of female education on fertility depends on the social and economic context of the population under study. In this paper, we investigate the causal effect of schooling on fertility using data from Taiwan. Our outcomes of interest are a married woman's number of children and age at first-birth. We take advantage of a jump in schooling caused by a change in Taiwan's compulsory education law in 1968, which extended compulsory education from 6 years to 9 years.

For our empirical analysis, we propose three new regression discontinuity (RD) estimators to incorporate the fact that our outcome variables are integer-valued (the number of children) or at least non-negative. The first estimator is a semiparametric ratio estimator that assumes an exponential regression function, but leaves the error distribution unspecified. The second is a nonparametric ratio estimator. The third is a modified nonparametric ratio estimator, taking into account the discreteness of the "running variable" age. The discreteness of the running variable arises from the lack of exact date of birth in our data. In addition to these RD estimators, we also use an instrumental variable estimator (IVE) for exponential models developed by Mullahy (1997), which is interesting in view of the well-known nexus between RD estimators and IVE.

We use the entire 1990 Population Census of Taiwan, not part of the Census. Putting our findings in advance, contrary to the negative relationship obtained by most studies in the literature, we find that schooling does not have an impact on the number of children. Also, education does not have an indirect deterring effect on fertility by delaying the age of first-birth or lowering the likelihood for a woman to marry.

The rest of this paper is organized as follows. Section 2 describes Taiwan's education system and the historical background of fertility in Taiwan. Section 3 elaborates on our RD methodologies. Section 4 describes our data. Section 5 reports the empirical results. Section 6 gives concluding remarks.

2 Background

(a) Taiwan's Compulsory Education

Taiwan's first compulsory education law was implemented in 1943 when Taiwan was a colony of Japan. The expansion of education in colonial Taiwan was for the purpose of assimilation, and also to produce more skilled workers; see Tsurumi (1979). Under this law, children between 6–12 of age have to receive six years of free education. In the sixties, after a period of rapid economic growth, the government realized the need to upgrade the industrial foundations. A part of the strategies to achieve the goal was improving human resources through training and education. This prompted Taiwan's compulsory education law change from six years to nine in 1968.

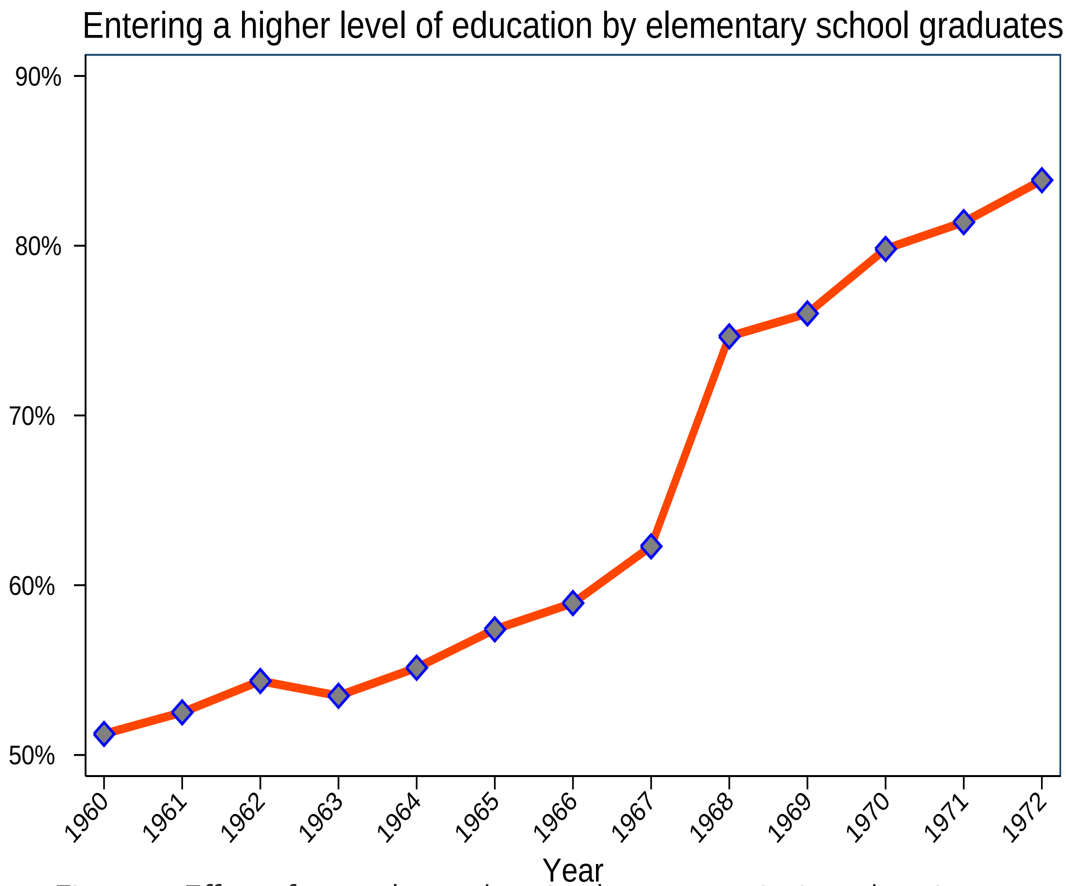

Figure 1: Effect of compulsory education laws on continuing education

In anticipation of the junior high school enrollment increase induced by the new compulsory education law, the government began constructing junior high schools and expanded teacher training starting from 1965. In 1968, 140 new junior high schools were opened and the number of junior high school teachers per elementary school graduate increased by 30 percent from 1967 to 1968. According to Hsieh and Clark (2000), this increased the number of junior high schools for every thousand elementary school graduate from 0.8 to 1.4 between 1967 and 1968.

The extension of compulsory education from six to nine years was one of the most dramatic reforms in modern Taiwan. Before 1968, elementary school education was compulsory and its enrollment rate reached almost 100%. However, restricted by the limited number of junior high schools, only 60% of elementary school graduates continued their education in the mid 1960s. Due to the 1968 compulsory education law change and the expansion of the junior high school capacity, the junior high school enrollment rate jumped from 62.3% to 74.7% between 1967 and 1968. Afterwards, the enrollment rate increased gradually and reached 83.9% in 1972. See Figure 1.

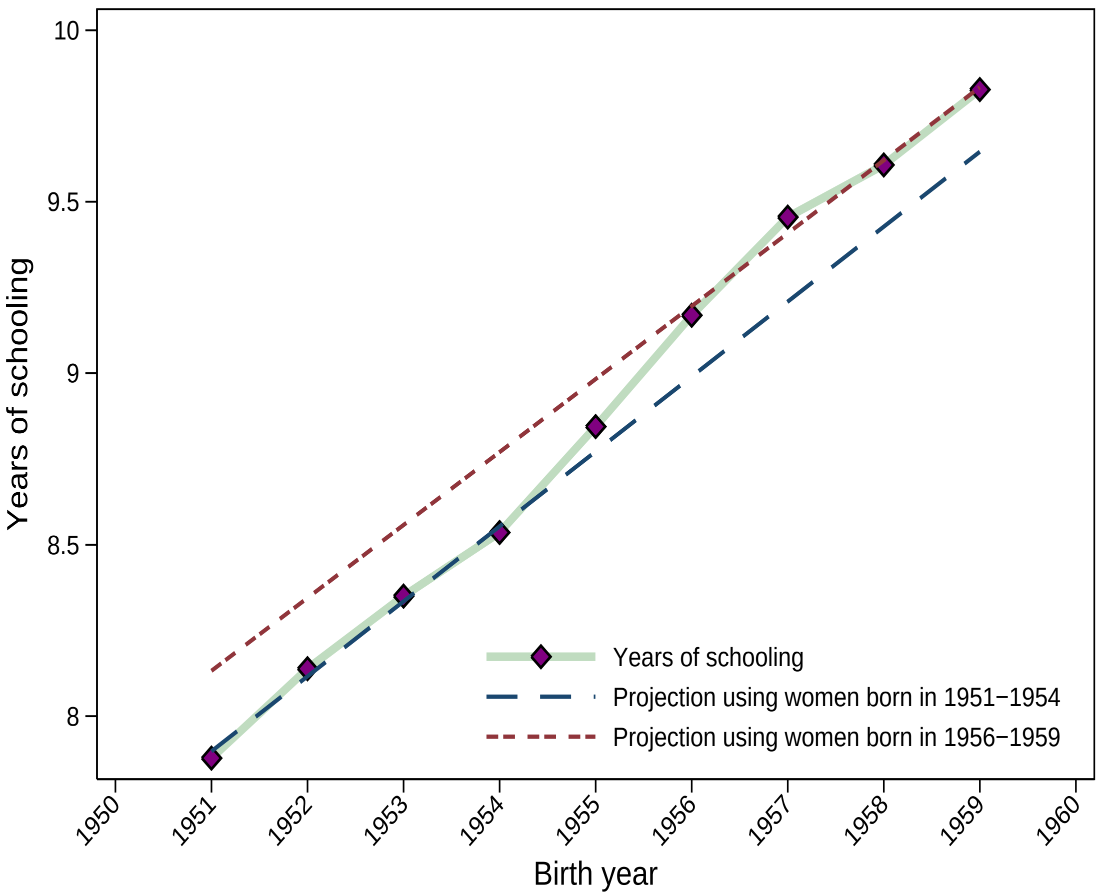

Figure 2: Effect of compulsory education law on years of schooling

Figure 2 shows the average years of schooling for women aged 31–39 in 1990 (i.e., born in 1951–1959) based on the Taiwan 1990 population census. Comparing the projected years of schooling for women aged 31–34 (born in 1956–1959, the cohorts affected by the education law change) and those aged 36–39 (born in 1951–1954, the cohorts not affected), there is a marked increase in schooling years, which is, however, not greater than 0.25 years. Thus, while the effect is clear, it is not substantial. Note that some women in the 1955 birth cohort (age 35 in 1990) were affected by the law and some were not, because the Taiwanese academic year starts in September 1. This cohort will not be used in our empirical analysis, as only birth year (not the exact date) is known in our data.

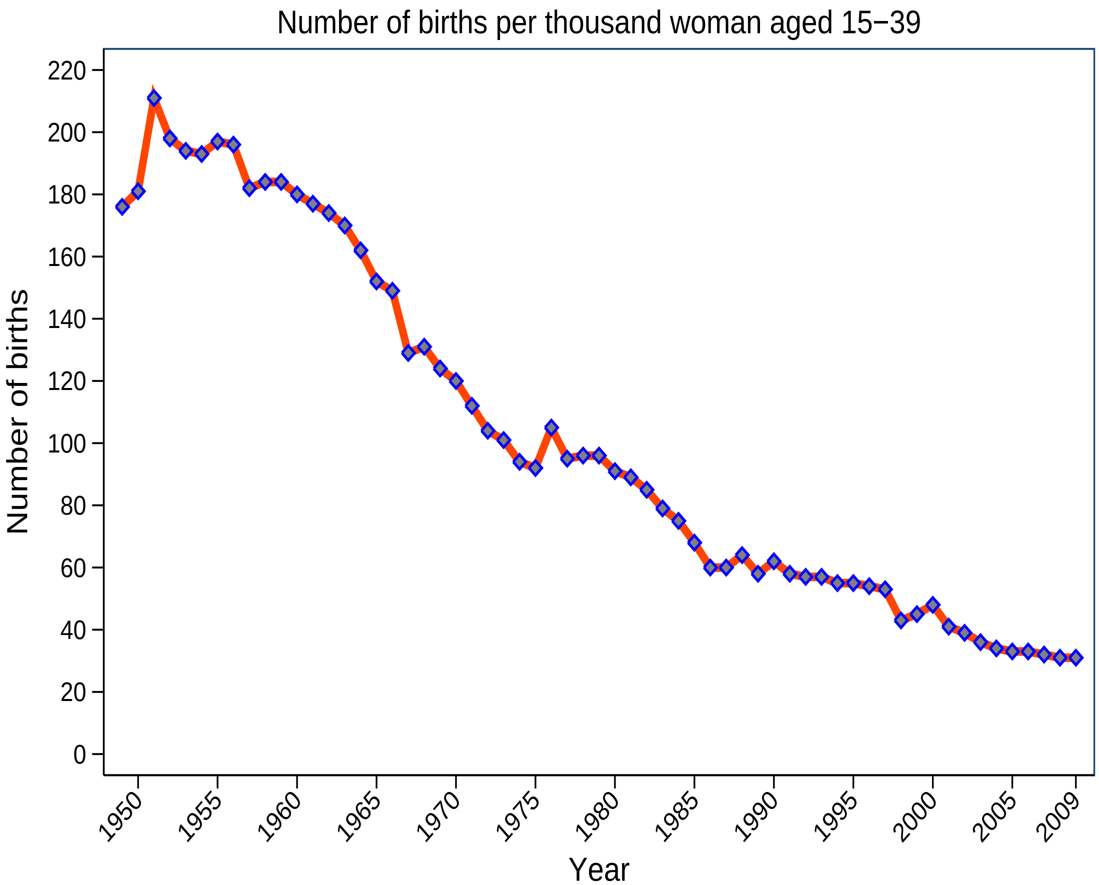

Figure 3: Number of births per 1,000 for women aged 15–39

(b) Fertility in Taiwan

The birth rate in Taiwan was very high after the second world war. See Figure 3, where the data come from the Department of Household Registration in the Interior Ministry. The total fertility rate (i.e., the number of births per 1000 thousand women aged 15–39) reached 211 in 1951. The post-war baby boom ended in 1957 when total fertility rate dropped to 182, but the total fertility rate was still very high. A clear downward trend in Taiwan's total fertility rate can be seen in the early 1960's. The total fertility rate declined very rapidly starting from the early 1960's and dropped to 101 in 1973. The rate of decline in the Taiwan's total fertility rate flattened after 1985. It fell to a meager 31 in 2008 and 2009.

The rapid decline in the Taiwan's fertility in the 1950's and 1960's may have to do with the fast growth of its economy. The real national income doubled between 1951 and 1961, and doubled again between 1961 and 1968, and the real growth in per capita GDP was about 5% in the 1960's. Another reason for the rapid decline in fertility is the unusually successful family planning program. The Taiwan's family planning program was implemented in 1964. Intrauterine

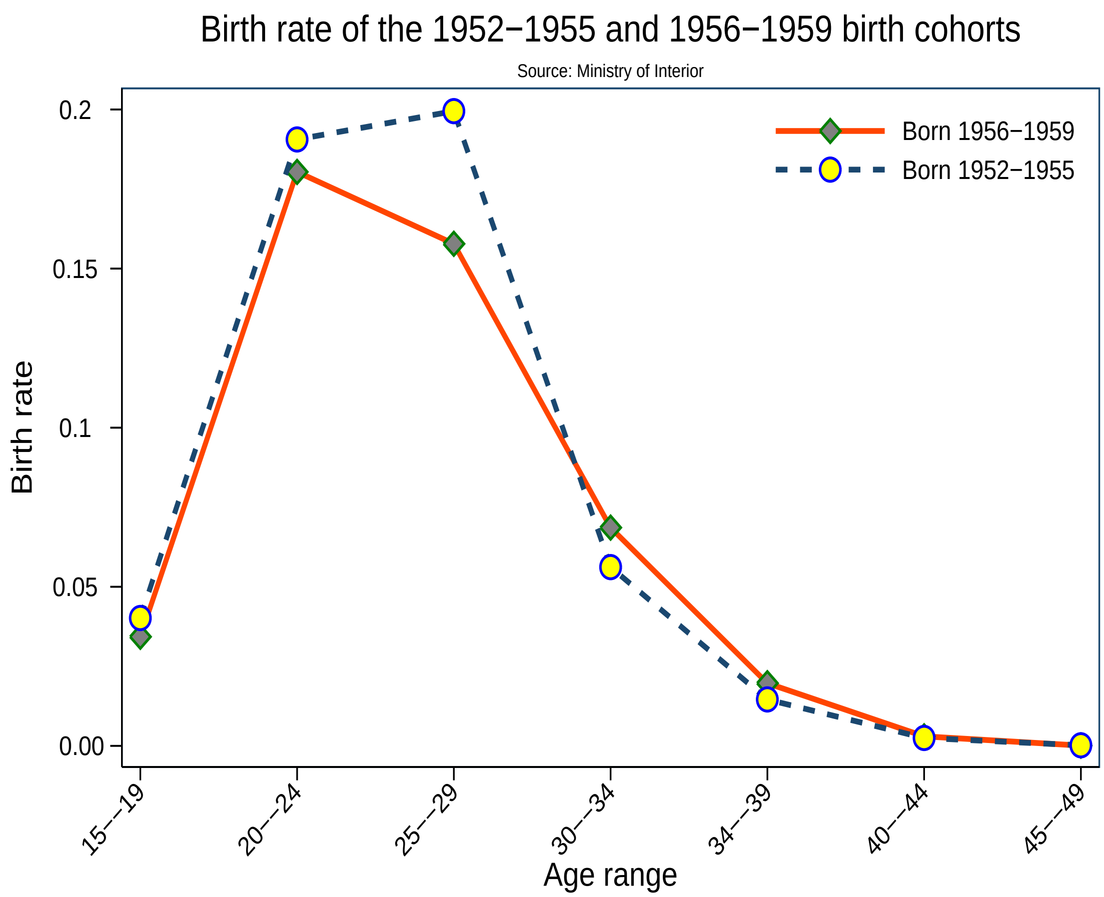

Figure 4: Fertility of cohorts affected and not affected by compulsory education laws

contraceptive devices (IUD's) were the major contraceptive method promoted by the government, followed by oral contraceptives and condoms. According to Chow (1974), the prevalence rate of the IUD's was 2.64% in 1964, and it climbed to 18.70% in 1971.

An interesting finding of Chow (1965) about the acceptors of IUD's is that married women with more education were more likely to accept them. Among the married women with more than senior high school education, 5.30% of them had an insertion, and only 4.14%, 2.47%, and 2.75% for married women with junior high school, primary school and no formal education, respectively. These findings suggest that the 1968 extension of the compulsory education, which lifted the average level of education in Taiwan, may partly account for the success of the Taiwan's family planning program.

To have a rough idea on the impact of the 1968 compulsory education law change on fertility outcomes, we examine Figure 4. Using aggregate data obtained from the Interior Ministry, Figure 4 compares the total fertility rate of women born during 1956–1959 (who were affected by the 1968 law change) with that of women born during 1952–1955 (who were not affected).

The comparison shows that the younger cohort (born 1956–1959) had a lower total fertility rate when they were 20–24 and 25–29 years old. Their total fertility rate picked up a little when they were 30–34 and 35–40. Overall, according to Figure 4 the total fertility rate of the 1956–1959 birth cohort is lower than that of the 1952–1955 birth cohort. This suggests that the 1968 compulsory education law change might have an impact on fertility outcomes.

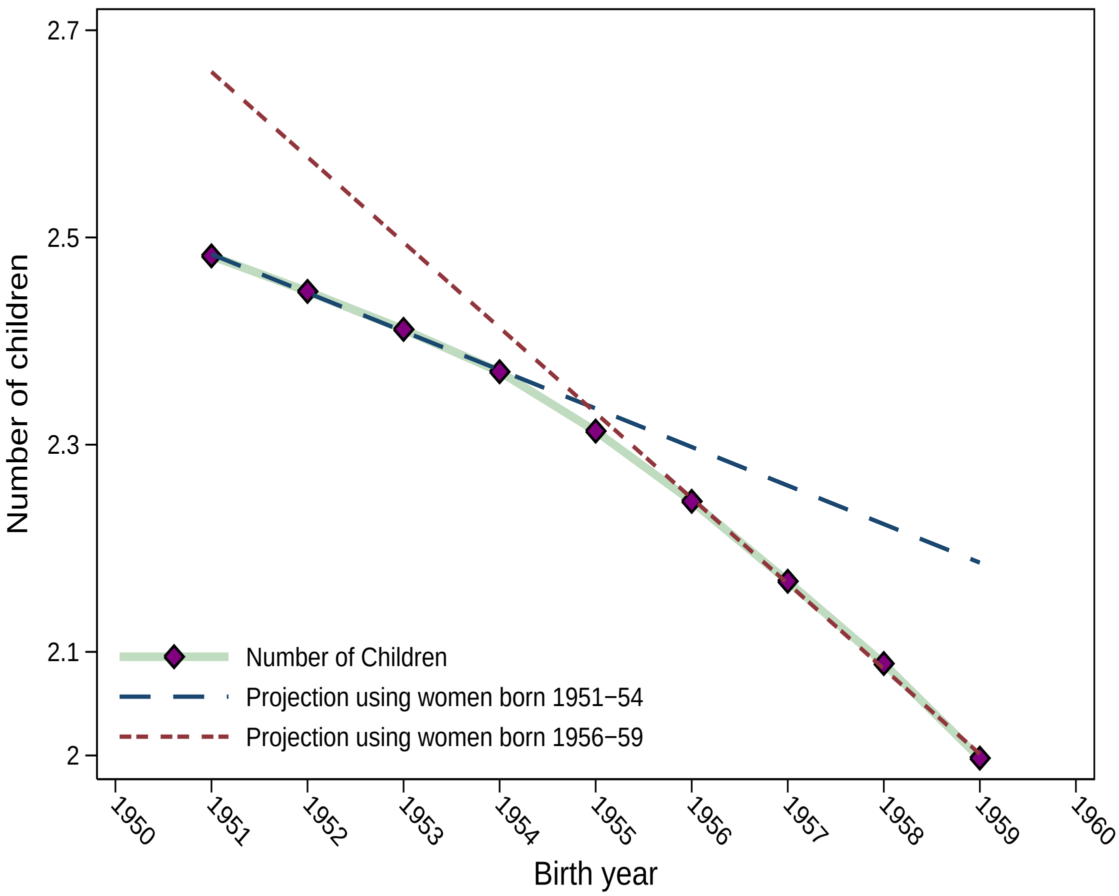

Figure 5: Number of children for married women in the 1990 Census of Taiwan

Figure 5 shows the average number of children for married women in the 1990 Population Census of Taiwan. There is a downward trend in fertility with respect to women's birth year. The trend is steeper for women born during 1956–1959, who were affected by the 1968 law change. The relationship between the number of children and birth year in the figure may represent three effects. The first is the cohort effect, reflecting differences in fertility behavior in different birth cohorts. The second is the effect of "incomplete fertility", as the young married women in our sample might not have completed their fertility at the time of the data collection, which would then exaggerate the negative effect of schooling because the younger cohort (control group) would have lower numbers of children than when they completed fertility. The third is the possible education effect.

In order to isolate the desired education effect, we have to remove the cohort effect and deal with the incomplete fertility problem. Between these two, we will deal with only the cohort effect, because, as it turns out, no significant negative effect of schooling is found despite the possible incomplete fertility problem that would exaggerate the negative effect.

3 Methodologies

One of the contributions of this study is the proposal of a set of new regression discontinuity estimators. These estimators are nonparametric or semiparametric in nature. They account for the endogeneity of the key explanatory variable of interest (education in our case) and the non-negativity of the outcome variable (fertility). The first estimator is a semiparametric ratio estimator—semiparametric in the sense that the regression functions are specified, but not the distributions of the error terms. The second is a nonparametric ratio estimator motivated by exponential models for non-negative (integer) responses. The third is a modified version of the second to correct for a bias due to the discreteness of the running variable (i.e., age or birth year in our case).

Moreover, we will compare our results with those obtained from the IVE of Mullahy (1997). The IVE arises naturally because it is known in the RD literature that a nonparametric ratio estimator has a numerically equivalent IVE version. Although we do not have any proof for such an equivalence in our case for non-negative (integer) responses, it will be still of interest to compare the ratio estimators with the IVE. These estimators require different assumptions to be consistent. Whereas the ratio estimators are “indirect estimators” using the reduced form of the outcome equation, the IVE estimates the structural form directly. This section details the estimators.

3.1 Semiparametric Ratio Estimator

3.1.1 Identification and Estimation

Let $1[A] = 1$ if A holds and 0 otherwise. For integer-valued fertility y, schooling s and age a, suppose

$$E(y|s,a) = \exp\{\beta_s s + m(a)\}, \tag{1}$$

$$s = \alpha_d d + m_s(a) + \varepsilon, \quad \text{where } d = 1[a \le 35], \tag{2}$$

β_s and $\alpha_d \neq 0$ are unknown parameters, $m(a)$ and $m_s(a)$ are unknown continuous functions of a, and ε is an error with $E(\varepsilon|a)=0$ and the density $f(\varepsilon|a)$ *of* $\varepsilon|a$ is continuous in a. Although we use $d = 1[a \leq 35]$, the cohort with $a = 35$ will not be used, as there is an ambiguity on whether individuals with $a = 35$ were treated or not due to the lack of exact birth date in our data.

Substituting (2) into (1), we obtain

$$E(y|a,\varepsilon) = \exp\left\{\alpha_d \beta_s d + m(a) + \beta_s m_s(a) + \beta_s \varepsilon\right\}. \tag{3}$$

Since (3) contains the unobserved ε, integrate out ε in $E(y|a,\varepsilon)$, we get $E(y|a)$:

$$\begin{aligned}\int E(y|a,\varepsilon) f(\varepsilon|a)d\varepsilon &= \exp\left\{\alpha_d\beta_s d + m(a) + \beta_s m_s(a)\right\}\int \exp(\beta_s\varepsilon) f(\varepsilon|a)d\varepsilon,\\ &= \exp\left\{\gamma_d d + m_y(a)\right\}, \qquad (4)\\ &\quad \text{where } \gamma_d \equiv \alpha_d\beta_s, \text{ and } m_y(a) \equiv m(a) + \beta_s m_s(a) + \ln\int \exp(\beta_s\varepsilon) f(\varepsilon|a)d\varepsilon.\end{aligned}$$

This is a reduced form for y whereas (1) is the structural form. The main fact is that β_d is identified by γ_d/α_d using the reduced form (4) and the structural form (1).

Typically in RD estimation, a local neighborhood of the break point 35 is specified as in

$$\delta = 1[a \in 35 \pm h] = 1\left[a \in [35-h, 35+h]\right],$$

where h is a bandwidth, say $h = 1$ or 2, as the running variable a is discrete. Then, with $m_s(a)$ and $m_y(a)$ specified, β_s can be estimated in two stages using the conditional moment conditions

$$E\Big[y - \exp\left\{\gamma_d d - m_y(\alpha,\gamma)\right\}\Big|a\Big] = 0 \quad \text{and} \quad E(\varepsilon|a) = 0.$$

For instance, suppose polynomial functions are used as in

$$m_s(a;\alpha) \equiv \alpha_0 + \alpha_1 a \quad \text{and} \quad m_y(a;\gamma) \equiv \gamma_0 + \gamma_1 a + \gamma_2 a^2,$$

where the α's and γ's are parameters to be estimated. Then $\alpha \equiv (\alpha_d, \alpha_0, \alpha_1)'$ can be estimated by the least square estimator (LSE) of δy on $\delta(d, 1, a)$, and $\gamma \equiv (\gamma_d, \gamma_0, \gamma_1, \gamma_2)'$ can be estimated by the generalized method of moments (GMM) using a moment condition such as

$$\sum_i \delta\{y_i - \exp(\gamma_d d_i - \gamma_0 - \gamma_1 a_i - \gamma_2 a_i^2)\}(d_i, 1, a_i, a_i^2)'.$$

Hence β_s can be estimated by the ratio of estimates of the parameters γ and α, i.e., $\widehat{\beta}_s = \widehat{\gamma}_d / \widehat{\alpha}_d$. Instead of attaching δ to the LSE or GMM minimand, we may simply select the observations with $\delta = 1$ to pretend that the sampling has been done already from the $\delta = 1$ subpopulation.

Strictly speaking, the neighborhood radius h should converge to 0 in RD, in which case the convergence rate of $\widehat{\gamma}_d/\widehat{\alpha}_d$ with $h \longrightarrow 0^+$ is not the usual $N^{-1/2}$. But typically in practice, once h is chosen to select the local observations, the ensuing asymptotic inference proceeds with the chosen fixed h using the asymptotic distribution theory under the usual $N^{-1/2}$ rate. This amounts to a 'fixed neighborhood' approach. We will also proceed with this wide-spread practice, not just because of its "convenience", but also because of the necessity: our RD running variable a in our data is discrete, and thus we cannot let $h \longrightarrow 0^+$.

Putting the requisite local identification conditions for $\gamma_d/\alpha_d = \beta_d$ together, we have

$$\begin{aligned}&\text{for all } a \in 35 \pm h,\ E(y|a) = \exp\left\{\gamma_d d + m_y(a;\gamma)\right\},\ s = \alpha_d d + m_s(a;\alpha) + \varepsilon,\\ &m_y(a;\gamma) \text{ and } m_s(a;\alpha) \text{ are parametric functions continuous in } a, \qquad \qquad (C_1)\\ &E\left[y - \exp\left\{\gamma_d d - m_y(\alpha,\gamma)\right\}\middle| a\right] = 0 \text{ and } E(\varepsilon|a) = 0.\end{aligned}$$

3.1.2 Asymptotic Distribution

For asymptotic inference, confidence intervals (CI) may be drawn using nonparametric bootstrap; i.e., resample from the data with replacement, and then use the quantiles of the bootstrap pseudo estimates to construct CI's. Instead of bootstrap CI's, we will show how to obtain the asymptotic distribution analytically in the following. To simplify the presentation, we will ignore δ by presuming that only the observations with $\delta = 1$ have been selected already. Since a ratio estimator can be unstable when the denominator is close to zero, bootstrap inference may be preferred.

Let w denote the regressors in the s equation; $w = (d, 1, a)'$ for the linear approximation of $m_s(a)$, and $w = (d, 1, a, a^2)'$ for the quadratic approximation. The LSE of s on w is done to obtain an estimate $\widehat{\alpha}$ of α, and the first element of $\widehat{\alpha}$ is denoted $\widehat{\alpha}_d$. Then

$$\sqrt{N}(\widehat{\alpha} - \alpha) = \frac{1}{\sqrt{N}}\sum_i \eta_{si}, \quad \text{where} \quad \eta_{si} \equiv \left(\frac{1}{N}\sum_i w_i w_i\right)^{-1} w_i \varepsilon_i,$$

where η_{si} is an 'influence function' for $\widehat{\alpha}$, and its first component η_{s1i} is the influence function for $\widehat{\alpha}_d$.

Let z denote the regressors in the reduced form (4) with the coefficient vector γ; with the quadratic approximation for $m_y(a;\gamma)$, $z = (d, 1, a, a^2)'$. Suppose that 'Poisson pseudo MLE' is used, which is the GMM with the moment condition

$$E\left[\left\{y - \exp(z'\gamma)\right\}z\right] = 0.$$

The estimator can be implemented by maximizing the usual Poisson MLE log-likelihood function for γ (omitting the unnecessary constant $-\ln(y_i!)$):

$$\sum_i \{y_i(z_i'\gamma) - \exp(z_i'\gamma)\}.$$

But the asymptotic variance takes the "sandwich" form

$$E^{-1}\left\{zz'\exp(z'\gamma)\right\} \cdot E\left[\{y - \exp(z'\gamma)\}^2 zz'\right] \cdot E^{-1}\left\{zz'\exp(z'\gamma)\right\}$$

instead of the usual MLE asymptotic variance, where $E^{-1}(\cdot) \equiv \left\{E(\cdot)\right\}^{-1}$.

The estimator $\widehat{\gamma}$ satisfies

$$\sqrt{N}(\widehat{\gamma} - \gamma) = \frac{1}{\sqrt{N}}\sum_i \eta_{yi} + o_p(1), \quad \text{where } \eta_{yi} \equiv \left\{\frac{1}{N}\sum_i z_i z_i' \exp(z_i'\gamma)\right\}^{-1} z_i\{y_i - \exp(z_i'\gamma)\} + o_p(1).$$

Let η_{y1i} be the first component of η_{yi}. Observe, with '$\rightsquigarrow$' denoting convergence in law,

$$\begin{aligned}
&\sqrt{N}\left(\frac{\widehat{\gamma}_d}{\widehat{\alpha}_d} - \frac{\gamma_d}{\alpha_d}\right) = \sqrt{N}\left(\frac{\widehat{\gamma}_d}{\widehat{\alpha}_d} - \frac{\widehat{\gamma}_d}{\alpha_d} + \frac{\widehat{\gamma}_d}{\alpha_d} - \frac{\gamma_d}{\alpha_d}\right) = \sqrt{N}\widehat{\gamma}_d\left(\frac{1}{\widehat{\alpha}_d} - \frac{1}{\alpha_d}\right) + \sqrt{N}\frac{1}{\alpha_d}\left(\widehat{\gamma}_d - \gamma_d\right) \\
&= -\frac{\widehat{\gamma}_d}{\widehat{\alpha}_d^2}\sqrt{N}(\widehat{\alpha}_d - \alpha_d) + \frac{1}{\alpha_d}\sqrt{N}(\widehat{\gamma}_d - \gamma_d) + o_p(1) = \frac{1}{\sqrt{N}}\sum_i\left(-\frac{\widehat{\gamma}_d}{\widehat{\alpha}_d^2}\eta_{s1i} + \frac{1}{\alpha_d}\eta_{y1i}\right) + o_p(1) \\
&\rightsquigarrow N(0, V_1), \quad \text{where } V_1 \equiv E(\eta_i\eta_i') \text{ and } \eta_i \equiv \frac{1}{\alpha_d}\eta_{y1i} - \frac{\gamma_d}{\alpha_d^2}\eta_{s1i}.
\end{aligned}$$

The variance V_1 can be consistently estimated by $N^{-1}\sum_i \widehat{\eta}_i\widehat{\eta}_i'$, where $\widehat{\eta}_i$ is η_i with its parameters replaced by the corresponding estimators.

3.2 Crude Nonparametric Ratio Estimator

3.2.1 Identification

In the reduced form (4), it holds that

$$\begin{aligned}
&\lim_{a\uparrow 35} E(y|a) \equiv E(y|35^-) = \exp(\alpha_d\beta_s d)\exp\left\{m_y(35)\right\}, \ \lim_{a\downarrow 35} E(y|a) \equiv E(y|35^+) = \exp\left\{m_y(35)\right\} \\
&\Longrightarrow \ \ln\frac{E(y|35^-)}{E(y|35^+)} = \alpha_d\beta_s.
\end{aligned}$$

Also from the model for s (2), we get

$$E(s|35^-) = \alpha_d + m_s(35), \quad E(s|35^+) = m_s(35) \quad \Longrightarrow \ E(s|35^-) - E(s|35^+) = \alpha_d.$$

This motivates a nonparametric ratio identification (the *desired ratio*):

$$\frac{\ln E(y|35^-) - \ln E(y|35^+)}{E(s|35^-) - E(s|35^+)} = \frac{\alpha_d\beta_s}{\alpha_d} = \beta_s.$$

That is, β_s can be identified by the left-hand side ratio, and thus can be estimated by replacing each of the four terms on the left-hand side with an nonparametric estimator. This ratio is in "log-linear" form.

As already mentioned, s and a are not continuous in our data, but integer-valued. Differently from y that takes mostly 0-3, the range of s is much larger, and thus we may still regard s as "continuous" to maintain the linear model for s. Later we will show that, if desired, an exponential model for s can be easily accommodated by replacing $E(s|35^-) - E(s|35^+)$ in the desired ratio with $\ln E(s|35^-) - \ln E(s|35^+)$. For a, however, there is no way to avoid the discontinuity problem as 35^- and 35^+ appear.

Suppose we use the women aged 34 as the treatment group and the those aged 36 as the control group. Observe now

$$\begin{aligned}
&E(y|34) = \exp\left\{\alpha_d \beta_s + m_y(34)\right\}, \quad E(y|36) = \exp\left\{m_y(36)\right\} \\
&\Longrightarrow \ln \frac{E(y|34)}{E(y|36)} = \alpha_d \beta_s + \ln m_y(34) - \ln m_y(36); \\
&E(s|34) - E(s|36) = \alpha_d + m_s(34) - m_s(36).
\end{aligned}$$

Therefore, assuming $\alpha_d + m_s(34) - m_s(36) \neq 0$, the *identified ratio* using ages 34 and 36 is

$$\frac{\ln E(y|34) - \ln E(y|36)}{E(s|34) - E(s|36)} = \frac{\alpha_d \beta_s + \ln m_y(34) - \ln m_y(36)}{\alpha_d + m_s(34) - m_s(36)}.$$

If we estimate the left-hand side nonparametrically to get a *crude nonparametric ratio estimator*, the estimator will be consistent for the right-hand side which deviates from the desired β_s. Nevertheless, $m_s(34) - m_s(36)$ has to do with the smooth change in schooling years over the two year span, and may be small compared with the magnitude of the break α_d, the effect of mandated compulsory education on schooling years. Also $m_y(34) - m_y(36)$ has to do with the smooth change in fertility over the two year span, and may be small compared with the break magnitude $\alpha_d \beta_s$. If so, the identified ratio using women aged 34 and 36 may be close enough to the desired ratio using 35^- and 35^+.

The exact bias for the crude nonparametric ratio estimator is

$$\begin{aligned}
&\frac{\alpha_d \beta_s + \ln m_y(34) - \ln m_y(36)}{\alpha_d + m_s(34) - m_s(36)} - \beta_s \\
&= \frac{\alpha_d \beta_s + \ln m_y(34) - \ln m_y(36) - \beta_s\{\alpha_d + m_s(34) - m_s(36)\}}{\alpha_d + m_s(34) - m_s(36)}, \\
&= \frac{\ln m_y(34) - \ln m_y(36) - \beta_s\{m_s(34) - m_s(36)\}}{\alpha_d + m_s(34) - m_s(36)}.
\end{aligned}$$

Hence the bias becomes zero iff

$$\begin{aligned}(i):\quad & m_y(34)=m_y(36),\quad \text{when}\ \ m_s(34)=m_s(36),\\ (ii):\quad & \beta_s=\frac{\ln m_y(34)-\ln m_y(36)}{m_s(34)-m_s(36)},\ \ \text{when}\ \ m_s(34)\neq m_s(36).\end{aligned}\qquad (C_2)$$

For (i), we may appeal to the smoothness of m_y and m_s, but (ii) involving β would hold only by a pure luck.

3.2.2 Estimation and Asymptotic Distribution

Although the above analysis was done with the differences of $\ln E(y|a)$ and $E(s|a)$, since both y and a are integer-valued in our data, it may be better to treat them in the same way; i.e., we may look at "linear-linear" or "log-log" ratios

$$\frac{E(y|35^-)-E(y|35^+)}{E(s|35^-)-E(s|35^+)}\quad \text{or}\quad \frac{\ln E(y|35^-)-\ln E(y|35^+)}{\ln E(s|35^-)-\ln E(s|35^+)}.$$

Which one is better depends on how y and s would have evolved over time (without break). If y grows at a constant rate, then an exponential model will be suitable, which leads to the ln transformation. If the increase in y is constant, then no ln transformation would be called for. In presenting the asymptotic distribution of the identified sample ratio in the following, we will use $T_y\{E(y|a)\}$ and $T_s\{E(s|a)\}$ for generality, where T_y and T_s are either the identity mapping or logarithm. This way, we can accommodate both exponential and linear models for y and s.

The identified sample ratios involving $T_y\{E(y|a)\}$ and $T_s\{E(s|a)\}$ are versions of nonparametric ratio estimators for treatment effects in the RD literature; see, e.g., Imbens and Lemieux (2008) and Lee and Lemieux (2010). A difference between our approaches and those in the RD literature is that we have a discrete running variable a, which simplifies the nonparametric estimation but complicates the identification issue as was shown above. A novelty of this paper is that ratio estimators motivated by the linear model in the RD literature are generalized for ratio estimators motivated by exponential models (or mixed models having both linear and exponential models). Also, to overcome the identification problem caused by a discrete running variable, we will later introduce a modified (i.e., bias-corrected) nonparametric ratio estimator.

For the asymptotic distribution of the crude nonparametric ratio estimator, denote the control and treatment group responses as

$$y_{0i},\ i=1,\dots,N_0\quad \text{and}\quad y_{1j},\ j=1,\dots,N_1,$$

where $N \equiv N_0 + N_1$. Let a_0 and a_1 be the control and treatment group ages (i.e., $a_0 = 36$ and $a_1 = 34$), respectively. For bigger age bands, we just have to redefine a_0 and a_1 as, e.g., $a_0 = \{36, 37\}$ and $a_1 = \{33, 34\}$.

Denote the sample versions of the numerator and denominator as, respectively, $\widehat{f}$ and $\widehat{g}$, and their probability limit as f and g. Then we have

$$
\begin{aligned}
\frac{\widehat{f}}{\widehat{g}} &\xrightarrow{p} \frac{f}{g}, \quad \text{where} \\
\widehat{f} &\equiv T_y(\widehat{\mu}_{y1}) - T_y(\widehat{\mu}_{y0}) \quad \text{and} \quad \widehat{g} \equiv T_s(\widehat{\mu}_{s1}) - T_s(\widehat{\mu}_{s0}), \\
f &\equiv T_y(\mu_{y1}) - T_y(\mu_{y0}) \quad \text{and} \quad g \equiv T_s(\mu_{s1}) - T_s(\mu_{s0}),
\end{aligned}
$$

and

$$
\begin{aligned}
\widehat{\mu}_{y0} &\equiv \frac{1}{N_0}\sum_i y_{0i} \xrightarrow{p} \mu_{y0} \equiv E(y|a_0) \quad \text{and} \quad \widehat{\mu}_{y1} \equiv \frac{1}{N_1}\sum_j y_{1j} \xrightarrow{p} \mu_{y1} \equiv E(y|a_1), \\
\widehat{\mu}_{s0} &\equiv \frac{1}{N_0}\sum_i s_{0i} \xrightarrow{p} \mu_{s0} \equiv E(s|a_0) \quad \text{and} \quad \widehat{\mu}_{s1} \equiv \frac{1}{N_1}\sum_j s_{1j} \xrightarrow{p} \mu_{s1} \equiv E(s|a_1).
\end{aligned}
$$

Under the null of no effect, we get $f = 0 \Longleftrightarrow \mu_{y0} = \mu_{y1}$. Note that $T_y'(c) = c$ when no ln transformation is used, and $T_y'(c) = 1/c$ when ln transformation is used.

The appendix shows that

$$
\sqrt{N}\left(\frac{\widehat{f}}{\widehat{g}} - \frac{f}{g}\right) \rightsquigarrow N(0, V_2),
$$

where V_2 is consistently estimable with

$$
\begin{aligned}
&\frac{N}{N_1}\cdot\frac{1}{N_1}\sum_j\left\{\frac{T_y'(\widehat{\mu}_{y1})}{\widehat{g}}(y_{1j} - \widehat{\mu}_{y1}) - \frac{\widehat{f}}{\widehat{g}^2}T_s'(\widehat{\mu}_{s1})(s_{1j} - \widehat{\mu}_{s1})\right\}^2 \\
&+\frac{N}{N_0}\cdot\frac{1}{N_0}\sum_i\left\{\frac{T_y'(\widehat{\mu}_{y0})}{\widehat{g}}(y_{0i} - \widehat{\mu}_{y0}) - \frac{\widehat{f}}{\widehat{g}^2}T_s'(\widehat{\mu}_{s0})(s_{0i} - \widehat{\mu}_{s0})\right\}^2.
\end{aligned}
$$

3.3 Modified Nonparametric Ratio Estimator

3.3.1 Identification

Instead of appealing to (C_2) including the negligibility of the "m function differences", we may be able to approximate $\ln m_y(34) - \ln m_y(36)$ and $m_s(34) - m_s(36)$ in (C_2) with some terms under certain conditions, which will then lead to a 'modified ratio'. This is explored in the following.

Rewrite the desired ratio with 35^- and 35^+ as

$$\frac{\ln E(y|35^-) - \ln E(y|35^+)}{E(s|35^-) - E(s|35^+)}$$
$$= \frac{\{\ln E(y|34) - \ln E(y|36)\} + \{\ln E(y|35^-) - \ln E(y|34)\} + \{\ln E(y|36) - \ln E(y|35^+)\}}{\{E(s|34) - E(s|36)\} + \{E(s|35^-) - E(s|34)\} + \{E(s|36) - E(s|35^+)\}}.$$

In the numerator, the first term uses the identified ages, the second term is the "proxy error" of using $E(y|34)$ instead of $E(y|35^-)$, and the third term is the proxy error of using $E(y|36)$ instead of $E(y|35^+)$.

Assume that *the differences involving* 35^- *and* 35^+ *are equal to the corresponding one year earlier/later differences within the same treatment/control group:*

$$\begin{aligned}
(i): &\quad \ln E(y|35^-) - \ln E(y|34) = \ln E(y|34) - \ln E(y|33),\\
(ii): &\quad \ln E(y|36) - \ln E(y|35^+) = \ln E(y|37) - \ln E(y|36),\\
(iii): &\quad E(s|35^-) - E(s|34) = E(s|34) - E(s|33),\\
(iv): &\quad E(s|36) - E(s|35^+) = E(s|37) - E(s|36).
\end{aligned} \tag{C$_3$}$$

Under (C_3), the preceding ratio becomes

$$\frac{\{\ln E(y|34) - \ln E(y|36)\} + \{\ln E(y|34) - \ln E(y|33)\} + \{\ln E(y|37) - \ln E(y|36)\}}{\{E(s|34) - E(s|36)\} + \{E(s|34) - E(s|33)\} + \{E(s|37) - E(s|36)\}}$$
$$= \frac{2\{\ln E(y|34) - \ln E(y|36)\} + \{\ln E(y|37) - \ln E(y|33)\}}{2\{E(s|34) - E(s|36)\} + \{E(s|37) - E(s|33)\}}.$$

Dividing both the numerator and denominator by 2, we get the desired equation under (C_3):

$$\frac{\ln E(y|35^-) - \ln E(y|35^+)}{E(s|35^-) - E(s|35^+)} = \frac{\{\ln E(y|34) - \ln E(y|36)\} + 0.5\{\ln E(y|37) - \ln E(y|33)\}}{\{E(s|34) - E(s|36)\} + 0.5\{E(s|37) - E(s|33)\}}.$$

Call the right-hand side the '*modified ratio*'. Since the age gap between 36 and 34 is 2 while the age gap between 37 and 33 is 4, it is natural that 0.5 appears in the modified ratio.

The modified ratio may look like approximating

$$\ln m_y(34) - \ln m_y(36) \quad \text{by} \quad 0.5\{\ln E(y|37) - \ln E(y|33)\}, \quad \text{and}$$
$$m_s(34) - m_s(36) \quad \text{by} \quad 0.5\{E(s|37) - E(s|33)\}.$$

Although this provides a helpful intuition, it is not exactly true because m_y and m_s are for smooth changes, whereas $\ln E(y|37) - \ln E(y|33)$ and $E(s|37) - E(s|33)$ include breaks. To see why we ended up with the approximation expressions like these, note that the numerator in the above ratio expression with 2 has the main term $2\{\ln E(y|34) - \ln E(y|36)\}$ with the break doubled due to 2, one break of which gets then cancelled by one break in the modifying term $\{\ln E(y|37) - \ln E(y|33)\}$.

3.3.2 Estimation and Asymptotic Distribution

As for the asymptotic distribution of the modified nonparametric ratio estimator, we need four groups. This can be understood by recalling the ages 33 ("post–treatment"), 34 (treatment), 36 (control), and 37 ("pre-control"). If we use {33, 34} as the treatment group ages and {36, 37} as the control group ages, then 32 is a post-treatment age and 38 is a pre-control age. Certainly, if desired, multiple post-treatment and pre-control ages may be used as well.

Denote the group responses with

$$
\begin{aligned}
y_{00i} \text{ (pre-control)}, \ i &= 1, \dots, N_{00}, \quad y_{01j} \text{ (control)}, \ j = 1, \dots, N_{01}, \\
y_{10k} \text{ (treatment)}, \ k &= 1, \dots, N_{10}, \quad \text{and} \quad y_{11l} \text{ (post-treatment)}, \ l = 1, \dots, N_{11},
\end{aligned}
$$

where $N \equiv N_{00} + N_{01} + N_{10} + N_{11}$. Let a_{00}, a_{01}, a_{10} and a_{11} be, respectively, the pre-control, control, treatment, and post-treatment group ages.

Denote the sample versions of the numerator and denominator of the modified ratio as, respectively, $\widehat{p}$ and $\widehat{q}$, and their probability limits as p and q. That is, the modified nonparametric ratio estimator $\widehat{p}/\widehat{q}$ satisfies

$$
\begin{aligned}
\frac{\widehat{p}}{\widehat{q}} &\xrightarrow{p} \frac{p}{q}, \quad \text{where} \\
\widehat{p} &\equiv T_y(\widehat{\mu}_{y10}) - T_y(\widehat{\mu}_{y01}) + 0.5\{T_y(\widehat{\mu}_{y00}) - T_y(\widehat{\mu}_{y11})\}, \\
\widehat{q} &\equiv T_s(\widehat{\mu}_{s10}) - T_s(\widehat{\mu}_{s01}) + 0.5\{T_s(\widehat{\mu}_{s00}) - T_s(\widehat{\mu}_{s11})\}, \\
p &\equiv T_y(\mu_{y10}) - T_y(\mu_{y01}) + 0.5\{T_y(\mu_{y00}) - T_y(\mu_{y11})\}, \\
q &\equiv T_s(\mu_{s10}) - T_s(\mu_{s01}) + 0.5\{T_s(\mu_{s00}) - T_s(\mu_{s11})\},
\end{aligned}
$$

and

$$
\begin{aligned}
\widehat{\mu}_{y00} &\equiv \frac{1}{N_{00}} \sum_i y_{00i} \xrightarrow{p} \mu_{y00} \equiv E(y|a_{00}) \quad \text{and} \quad \widehat{\mu}_{y01} \equiv \frac{1}{N_{01}} \sum_j y_{01j} \xrightarrow{p} \mu_{y01} \equiv E(y|a_{01}), \\
\widehat{\mu}_{y10} &\equiv \frac{1}{N_{10}} \sum_k y_{10k} \xrightarrow{p} \mu_{y10} \equiv E(y|a_{10}) \quad \text{and} \quad \widehat{\mu}_{y11} \equiv \frac{1}{N_{11}} \sum_l y_{11l} \xrightarrow{p} \mu_{y11} \equiv E(y|a_{11});
\end{aligned}
$$

$\widehat{\mu}_{s00}$, $\widehat{\mu}_{s01}$, $\widehat{\mu}_{s10}$, and $\widehat{\mu}_{s11}$ as well as their population versions are analogously defined.

The appendix shows that

$$
\sqrt{N}\left(\frac{\widehat{p}}{\widehat{q}} - \frac{p}{q}\right) \rightsquigarrow N(0, V_3),
$$

where V_3 is consistently estimable by

$$\begin{aligned}
&\frac{N}{N_{10}}\cdot\frac{1}{N_{10}}\sum_k\left\{\frac{T_y'(\widehat{\mu}_{y10})}{\widehat{q}}(y_{10k}-\widehat{\mu}_{y10})-\frac{\widehat{p}T_s'(\widehat{\mu}_{s10})}{\widehat{q}^2}(s_{10k}-\widehat{\mu}_{s10})\right\}\\
&+\frac{N}{N_{01}}\cdot\frac{1}{N_{01}}\sum_j\left\{\frac{T_y'(\widehat{\mu}_{y01})}{\widehat{q}}(y_{01j}-\widehat{\mu}_{y01})-\frac{\widehat{p}T_s'(\widehat{\mu}_{s01})}{\widehat{q}^2}(s_{01j}-\widehat{\mu}_{s01})\right\}\\
&+0.25\frac{N}{N_{00}}\cdot\frac{1}{N_{00}}\sum_i\left\{\frac{T_y'(\widehat{\mu}_{y00})}{\widehat{q}}(y_{00i}-\widehat{\mu}_{y00})-\frac{\widehat{p}T_s'(\widehat{\mu}_{s00})}{\widehat{q}^2}(s_{00i}-\widehat{\mu}_{s00})\right\}\\
&+0.25\frac{N}{N_{11}}\cdot\frac{1}{N_{11}}\sum_l\left\{\frac{T_y'(\widehat{\mu}_{y11})}{\widehat{q}}(y_{11l}-\widehat{\mu}_{y11})-\frac{\widehat{p}T_s'(\widehat{\mu}_{s11})}{\widehat{q}^2}(s_{11l}-\widehat{\mu}_{s11})\right\}.
\end{aligned}$$

3.4 IVE for Structural Exponential Model

If a linear regression model for RD holds in the sense

$$E(y|a)=\beta_d E(s|a)+m(a),\ E(s|a)=\alpha_d d+m_s(a)\Longrightarrow E(y|a)=\alpha_d\beta_d d+\beta_d m_s(a)+m(a),$$

then, β_d is identified by the ratio

$$\frac{E(y|35^-)-E(y|35^+)}{E(s|35^-)-E(s|35^+)}.$$

In the RD literature, it is known that a kernel estimator for this ratio has a numerically equivalent IVE form with d instrumenting for s. In view of this equivalence, it would be interesting to see how close our nonparametric "log-linear" ratio estimator is to the IVE for exponential structural models for y proposed by Mullahy (1997). It will be seen later that the nonparametric "log-linear" ratio estimates are indeed quite close to the IVE results. In the following, we will explain the IVE of Mullahy (1997).

Consider an error term ξ such that

$$E(e^{\xi}|a)=\tau\ \forall a\ \ (\tau \text{ is an unknown constant}),$$

which holds if ξ is independent of a. Suppose

$$E(y|s,a,d,\xi)=E(y|s,a,\xi)=\exp\{\beta_s s+m(a;\beta_m)+\xi\},\qquad(5)$$

where $m(a;\beta_m)$ is a parametric function of a, say,

$$m(a;\beta_m)=\beta_0+\beta_1 a+\beta_2 a^2.$$

The first equality in (5) is the exclusion restriction that d is excluded from the structural for y once (s,a,ξ) are controlled for, and the second equality is the usual exponential model specification. Instead of the quadratic function for $m(a;\beta_m)$ that will be used in the rest of this section, a more general function such as cubic or quartic may be used so that the functional form restriction on $m(a;\beta_m)$ becomes less binding.

Define

$$v \equiv y - E(y|s,a,d,\xi) = y - E(y|s,a,\xi) = y - \exp(\beta_s s + \beta_0 + \beta_1 a + \beta_2 a^2 + \xi)$$
$$\Longrightarrow \; y = \exp(\beta_s s + \beta_0 + \beta_1 a + \beta_2 a^2)\cdot e^{\xi} + v, \quad \text{where } E(v|s,a,d,\xi) = 0 \text{ by construction.}$$

Rewrite this y equation as

$$y = \exp(\beta_s s + \beta_0^* + \beta_1 a + \beta_2 a^2)\cdot \frac{e^{\xi}}{\tau} + v, \quad \text{where } \beta_0^* \equiv \beta_0 + \ln\tau. \tag{6}$$

To simplify notations, further define

$$\beta \equiv (\beta_s, \beta_0^*, \beta_1, \beta_2)', \quad x = (s,1,a,a^2)' \quad \text{and} \quad z = (d,1,a,a^2)'.$$

Divide both sides of (6) by $\exp(x'\beta)$ and subtract 1 to get

$$y\exp(-x'\beta) - 1 = \frac{e^{\xi}}{\tau} - 1 + v\exp(-x'\beta).$$

Multiply this equation by z and take $E(\cdot)$ to get

$$E\Big[z\{y\exp(-x'\beta) - 1\}\Big] = 0. \tag{7}$$

Equation (7) holds because

$$E\left\{z\Big(\frac{e^{\xi}}{\tau} - 1\Big)\right\} = E\left\{z\cdot E\Big(\frac{e^{\xi}}{\tau} - 1\Big|z\Big)\right\} = 0 \quad \text{and}$$
$$E\Big\{zv\exp(-x'\beta)\Big\} = E\Big[z\cdot E\{v\exp(-x'\beta)\big|s,z\}\Big] = 0.$$

Hence GMM can be applied to estimate β using the unconditional moment condition (7).

In the IVE/GMM, essentially d is an instrument for s as can be seen in comparing x and z. The exclusion restriction for d holds by the assumption $E(y|s,a,d,\xi) = E(y|s,a,\xi)$, and the inclusion restriction is implicit in using d in z for s in x. The condition that the instrument d should be orthogonal to the y model error term appears in $E(e^{\xi}|a) = \tau$, as ξ is the y model error term that may be correlated with s.

Turning to the localization with $\delta = 1$, the localized moment condition is

$$E\Big[\delta\cdot z\{y\exp(-x'\beta) - 1\}\Big] = 0.$$

Recall that $E\left[z\{y\exp(-x'\beta)-1\}\right]=0$ was derived from

$$E(e^{\xi}|a)=\tau \quad \text{and} \quad E(y|s,a,d,\xi)=E(y|s,a,\xi)=\exp\left\{\beta_d s+m(a;\beta_m)+\xi\right\} \quad \forall a.$$

What is needed for the localized moment condition is thus

$$E(e^{\xi}|a)=\tau \text{ and } E(y|s,a,d,\xi)=E(y|s,a,\xi)=\exp\left\{\beta_d s+m(a;\beta_m)+\xi\right\} \quad \forall a\in 35\pm h. \tag{C$_4$}$$

Condition (C_4) is much weaker than the preceding "global" condition that should hold for all ages.

Omitting δ again, rewrite the local GMM moment condition as

$$E\left[z\left\{y\exp(-x'\beta)-1\right\}\right]=E\{\psi(\beta)\}=0 \quad \text{where} \quad \psi(\beta)\equiv z\left\{y\exp(-x'\beta)-1\right\}.$$

With $W\equiv E\{\psi(\beta)\psi(\beta)'\}$ and $\psi_b=\partial\psi(\beta)/\partial b$, as well known, the GMM $\widehat{\beta}$ has

$$\sqrt{N}(\widehat{\beta}-\beta)\rightsquigarrow N\left\{0,(E\psi_b W^{-1}E\psi_{b'})^{-1}\right\}.$$

For the moment condition, the inverse of the asymptotic variance is

$$E\psi_b W^{-1}E\psi_{b'}=E\left\{yxz'\exp(-x'\beta)\right\}\cdot E^{-1}\left[zz'\left\{y\exp(-x'\beta)-1\right\}^2\right]\cdot E\left\{zx'y\exp(-x'\beta)\right\}.$$

4 Data

Our data come from the entire 1990 Population Census of Taiwan. With the largest age band $h=3$, there are totally 1,021,204 women (both married or unmarried, and may or may not be the household head or household head's spouse) born in 1952–54 and 1956–58 in the data. Using a larger age band would increase the efficiency of the estimators, but only at the expense of increasing the bias due to the treatment and control groups differing in aspects other than age. Since our data is already large, efficiency is of lesser concern.

The 1990 Population Census does not contain information on the number of children ever born to a woman. Individuals were asked about their relationship to the head of the household where they resided at the time of data collection. With this information, we could identify a woman's total number of children living in the household only if she was the household head or the spouse of the head. Thus, we use observations for the women who were the head or the spouse of the head. A woman's "number of children" in our data is her total number of children residing in the household, not necessarily the total number of children ever born to her. Due to

Table 1: Descriptive statistics†

	Married Women					
	(1) Born in 1952	(2) Born in 1953	(3) Born in 1954	(4) Born in 1956	(5) Born in 1957	(6) Born in 1958
Number of children	2.542	2.505	2.463	2.336	2.260	2.180
	(1.142)	(1.124)	(1.102)	(1.073)	(1.057)	(1.049)
Age at first-birth	23.950	23.968	23.924	23.842	23.794	23.691
(women with at least one child)	(3.802)	(3.773)	(3.737)	(3.660)	(3.621)	(3.551)
Years of schooling	7.911	8.140	8.332	8.981	9.267	9.420
	(4.112)	(4.080)	(4.016)	(3.891)	(3.765)	(3.650)
Junior high school graduate	0.418	0.446	0.472	0.586	0.635	0.668
(1 if yes, 0 otherwise)	(0.493)	(0.497)	(0.499)	(0.493)	(0.481)	(0.471)
Observation	123911	124008	127154	131149	120693	121632

†Standard deviation in parentheses.

this limitation, we use the 1990 Population Census instead of the 2000 Population Census that was also available: in the latter, the children were likely to have left their parents' home.

In the census, there were 808,028 women who were born during 1952–54 and 1956–58, and were the head or head's spouse. This is, among the 1,021,204 women in the 1990 Population Census, 213,176 were not the head or spouse of the head of the household. Among these women, 20,361 were never married, 26,185 were divorcees and 12,070 were widows. In our analysis, we only look at women who were currently married. For these currently married women, we drop 865 women who gave birth when they were 12 or younger, because it is likely that there are coding errors for these observations. As already mentioned, we also do not use the women born in 1955, as it is not clear whether they are treated or control subjects—the exact birth date was not released for confidentiality.

In our sample, there are 123911, 124008, and 127154 married (exclusive of divorcees and widows) women born in 1952, 1953, 1954; there are 131149, 120693, and 121632 married women born in 1956, 1957, 1958. In total, there are 748,547 married women in our sample. Descriptive statistics are in Table 1. The table indicates that married women affected by the law change (born in 1956–58) received more years of schooling than those not affected (born in 1952–54). There is also a higher proportion of high school graduates among the married women born in 1956–58 relative to those born in 1952–1954. This suggests that the 1968 compulsory education law change was effective in increasing schooling.

The Table 1 also shows that the married women born in 1952–54 have more children than those born in 1956–58. However, the first-birth age of the married women affected by the law change and those not affected is almost the same. The married women born in 1952–54 gave birth to their first child at the age of 24.0, 24.0 and 23.9, while those born in 1956–58 did it at the age of 23.8, 23.8 and 23.7, respectively.

5 Empirical Results

5.1 Main Findings for Impact of Schooling on Fertility

Before examining the main results, we take a look at the effect of the compulsory education law change on schooling in Table 2. The results are obtained by OLS and a quadratic function of age is used to control for the cohort effect. The estimation is done using the subsamples of married women born in 1952–58 ($h = 3$) or 1953–57 ($h = 2$). We did not use the subsample of women born in 1954–56 ($h = 1$) because two birth cohorts are not enough to estimate $m(a)$ to remove the cohort effect; the age band with $h = 1$ yields only two sample averages whereas the OLS has three or more parameters.

Our identification strategy takes advantage of an exogenous change in schooling created by the 1968 compulsory education law change in Taiwan affecting individuals born after September 1955. For this, we need the law change to have enough explanatory power for schooling. As reported in Table 2, the F-statistic for the dummy variable indicating the timing of the law change is quite large. In the two samples, the F-statistics are 26.0 and 109.7, respectively.

Table 2 shows that the law change boosted schooling by around 0.2 years: 0.17 and 0.23 years for the 1953–57 and 1952–58 cohorts, respectively. These estimates imply that the positive effect of the law change on schooling is gradual. This is probably because school capacity takes time to expand. These figures are modest compared with the 0.19–0.28 years increase due to Germany's compulsory ninth grade completion requirement (Pischke and von Wachter 2008), 0.236, 0.260 and 0.458 years increase due to the U.S. school-leaving-age requirements of 14, 15 and 16 years, respectively (Oreopoulos 2007), 0.405 and 0.643 years increase due to Canada's school-leaving-age requirements of 14 and 15 years, and 0.425 years increase due to UK's school-leaving-age requirement of 15 years (Oreopoulos 2007).

The main estimation results are in Tables 3. In Table 3 we report the estimation results for the samples of married women born in 1954–56 ($h = 1$) using the crude nonparametric ratio and modified nonparametric ratio estimators, and those born in 1953–57 ($h = 2$) and 1952–58 ($h = 3$) using the semiparametric and instrumental variable (IV) estimators. These two estimators need $h \geq 2$ because they have three or more parameters.

Table 2: Effect of compulsory education law change on years of schooling

	Dependent variable: Years of schooling	
	(1) Women born in 1953–57	(2) Women born in 1952–58
Age	-1.351***	-0.243**
	(0.260)	(0.096)
Age Square	0.016***	0.000
	(0.004)	(0.001)
Reform	0.171***	0.233***
	(0.035)	(0.024)
Constant	36.380***	16.573***
	(4.548)	(1.686)
R^2	0.013	0.021
F-statistic of "Reform"	26.036	109.736
Observation	503004	748547

† $***$, $**$, and $*$, denote statistically significance at the 1%, 5%, and 10% levels. Standard errors in parentheses. "Reform" is a dummy variable indicating the timing of the compulsory education reform. Women born in 1955 are excluded.

Table 3 reports the estimates of β_s, which equals $\left[\partial \exp\left(\beta s + m(a)\right)/\partial s\right] \big/ \exp\left(\beta_s s + m(a)\right)$, and $\beta_s \times 100$ represents the percentage change in the number of births caused by a one year increase in schooling. The results in column (1) of Table 3 for the crude nonparametric ratio estimator for women born in 1954 and 1956 suggest that an additional year of education has a -8.2% effect on fertility, and this effect is statistically significant at conventional levels. However, the estimate is potentially biased due to its failure to account for the discreteness of the running variable (i.e., the difference in fertility of women born in 1954 and 1956). The modified nonparametric ratio estimator corrects for this bias. Its estimate in column (2) suggests that schooling years reduces fertility only by 1.6%, which is statistically insignificant though. This suggests that schooling does not have an impact on fertility, and the large and statisitically significant effect obtained by the crude nonparametric ratio estimator seems to arise from age, not schooling, effect.

Statistically insignificant estimates of schooling's effect on fertility are also obtained by the semiparametric estimator and IVE for the women born in 1953–57 ($h = 2$) and 1952–1958 ($h = 3$), as reported in columns (3) and (4) of Table 3. Notice that the magnitude of the estimates obtained by the modified nonparametric ratio, semiparametric ratio, and IV estimators are quite similar, and this is assuring.

Table 3: Effect of years of schooling on married women's number of children†

Sample	Dependent variable: Number of children				
	(1)	(2)	(3)	(4)	(5)
	Crude Nonparametric Ratio Estimator	Modified Nonparametric Ratio Estimator	Semiparametric Ratio Estimator	IV Estimator	Nonlinear LS (endogeneity not accounted for)
Women Born in 1954–56 [251831 Observations]	-0.082*** (0.016)	-0.016 (0.048)			
Women Born in 1953–57 [503004 Observations]			-0.016 (0.023)	-0.018 (0.024)	-0.039*** (1.5e-04)
Women Born in 1952–58 [748547 Observations]			-0.012 (0.011)	-0.012 (0.012)	-0.039*** (1.3e-04)

† ∗∗∗,∗∗, and ∗, denote statistically significance at the 1%, 5%, and 10% levels. Standard errors in parentheses. Women born in 1955 are excluded.

Table 4: Effect of schooling on fertility of married women with one or more children†

Sample	Dependent variable: Number of children				
	(1)	(2)	(3)	(4)	(5)
	Crude Nonparametric Ratio Estimator	Modified Nonparametric Ratio Estimator	Semiparametric Ratio Estimator	IV Estimator	Nonlinear LS (endogeneity not accounted for)
Women Born in 1954–56 [236349 Observations]	-0.071*** (0.014)	0.006 (0.056)			
Women Born in 1953–57 [473111 Observations]			0.006 (0.022)	0.007 (0.025)	-0.033*** (1.3e-04)
Women Born in 1952–58 [703328 Observations]			-0.008 (0.010)	-0.008 (0.010)	-0.033*** (1.0e-04)

† ∗∗∗,∗∗, and ∗, denote statistically significance at the 1%, 5%, and 10% levels. Standard errors in parentheses. Women born in 1955 are excluded.

5.2 Various Comparisons and Sensitivity Checks

(a) Schooling Assumed Exogenous

The modified nonparametric ratio, semiparametric ratio, and IV estimation results are quite different from those obtained without allowing for schooling endogeneity. The results in column (5) of Table 3 are based on an exponential model

$$y = \exp(\beta_s s + \beta_{a1} a + \beta_{a2} a^2) + error, \tag{8}$$

which is estimated by the nonlinear least squares. The estimates for β_s are -0.039 in all samples of birth cohorts, and their standard errors are very small. The large difference in the estimates with and without allowing for schooling endogeneity suggests that schooling is endogenous in

Table 5: Effect of schooling on age at first-birth of married women†

	Dependent variable: Age at first-birth				
Sample	(1)	(2)	(3)	(4)	(5)
	Crude Nonparametric Ratio Estimator	Modified Nonparametric Ratio Estimator	Semiparametric Ratio Estimator	IV Estimator	Nonlinear LS (endogeneity not accounted for)
Women Born in 1954–56 [236349 Observations]	-0.005 (0.076)	0.003 (0.678)			
Women Born in 1953–57 [473111 Observations]			0.003 (0.009)	0.003 (0.008)	0.018*** (5.1e-05)
Women Born in 1952–58 [703328 Observations]			0.001 (0.004)	0.001 (0.004)	0.018*** (4.2e-05)

† $***$, $**$, and $*$, denote statistically significance at the 1%, 5%, and 10% levels. Standard errors in parentheses. Women born in 1955 are excluded.

Table 6: Effect of schooling on fertility of married women with no more than junior high school education†

	Dependent variable: Number of children				
Sample	(1)	(2)	(3)	(4)	(5)
	Crude Nonparametric Ratio Estimator	Modified Nonparametric Ratio Estimator	Semiparametric Ratio Estimator	IV Estimator	Nonlinear LS (endogeneity not accounted for)
Women Born in 1954–56 [152402 Observations]	-0.074*** (0.015)	0.011 (0.172)			
Women Born in 1953–57 [315962 Observations]			0.011 (0.065)	0.027 (0.236)	-0.021*** (2.9e-04)
Women Born in 1952–58 [471849 Observations]			0.007 (0.024)	0.009 (0.031)	-0.020*** (2.4e-04)

† $***$, $**$, and $*$, denote statistically significance at the 1%, 5%, and 10% levels. Standard errors in parentheses. Women born in 1955 are excluded.

the fertility equation.

(b) Effect on the Subpopulation with One Child or More

Although we found no effect of schooling on fertility, an effect may be present for those women with at least one child, i.e., those who have "decided" to have children. The results for the women with one child or more are in Table 4, which are similar to the results without excluding the women with no children, i.e., all estimates of the impact of schooling on fertility are statistically insignificant at conventional levels. By contrast, nonlinear least square estimates without accounting for endogeneity are negative, large in magnitude and statistically significant. The results in Table 4 again suggest that schooling has no effect on fertility.

Table 7: Effect of schooling on whether married†

	Dependent variable: Whether ever married or not	
	(1)	(2)
	Women born in 1953–57	Women born in 1952–58
Years of schooling	0.069 (0.066)	0.054* (0.029)
Age	0.629*** (0.097)	0.419*** (0.039)
Age square	-0.008*** (0.001)	-0.005*** (0.001)
Constant	-11.605*** (2.054)	-7.723*** (0.653)
ρ	-0.534*** (0.219)	-0.484*** (0.096)
Marginal effect of years of schooling	0.012 (0.015)	0.009 (0.006)
First-stage's F-statistic	23.53	90.30
Observation	681250	1021204

† ρ denotes the correlation coefficient of the error terms in the schooling and ever married models. $***$,$**$, and $*$, denote statistically significance at the 1%, 5%, and 10% levels. Standard errors in parentheses. Women born in 1955 are excluded.

(c) Schooling Effect on First-Birth Age

We next examine whether schooling has any impact on the age at first-birth. The results are in Table 5. Again, all the estimates suggest that schooling does not have a statistically significant impact on the age at first-birth. This is in sharp contrast with the results in column (5) of Table 5 where schooling is assumed to be exogenous and thus nonlinear least square for the exponential model (8) is used. The large difference between the results with and without endogeneity accounted for suggests that schooling is endogenous in the first-birth age equation. The results with exogenous schooling are similar to those obtained by previous studies when schooling endogeneity is not accounted for, e.g., Rindfuss, Morgan, and Offutt (1996) who used the 1980, 1985, and 1990 Current Population Surveys of the U.S., and Derose and Kravdal (2007) who used survey data from the sub-Saharan Africa.

(d) Effect on the Subpopulation with Low Education

Although we found no effect of schooling on married women's fertility so far, schooling might have some impact on the fertility of the women with low level of schooling. Hence we estimated the effect for the subpopulation with junior high school or less education. The results

Table 8: Effect of years of schooling on married women's number of children—Linear functional form†

	Dependent variable: Number of children				
Sample	(1)	(2)	(3)	(4)	(5)
	Crude Nonparametric Ratio Estimator	Modified Nonparametric Ratio Estimator	Semiparametric Ratio Estimator	IV Estimator	OLS
Women Born in 1954–56 [251831 Observations]	-0.196*** (0.038)	-0.050 (0.134)			
Women Born in 1953–57 [503004 Observations]			-0.050 (0.054)	-0.050 (0.054)	-0.093*** (3.7e-04)
Women Born in 1952–58 [748547 Observations]			-0.044 (0.027)	-0.044 (0.027)	-0.093*** (3.1e-04)

† ∗∗∗,∗∗, and ∗, denote statistically significance at the 1%, 5%, and 10% levels. Standard errors in parentheses. Women born in 1955 are excluded.

are in Table 6. The results in columns (2)–(4) of Table 6 for the modified nonparametric ratio, semiparametric estimator and IVE show that, when schooling endogeneity is accounted for, schooling does not have any statistically significant impact on fertility.

(e) Effect of Schooling on Marrying Ever

Even if schooling has no effect on fertility, it may still affect negatively women's decision to marry. If this is the case, then education still has a negative effect on fertility indirectly. To explore this, we applied probit to ever marrying or not, using all women born in 1952–58 ($N = 1,021,204$) or 1953–57 ($N = 681,250$)—again, the 1955 birth cohort is excluded. The proportion of women who were ever married is 0.92 for the 1952–1958 birth cohorts and 0.93 for 1953–1957. For regressors, schooling, age, and age squared are used along with a 'control function' to account for schooling endogeneity, where the control function is the residual from applying least squares to the linear model of schooling. The results are in Table 7. A one-year increase in schooling raises the likelihood of being ever married by 0.9 to 1.2 percentage points, which are small in magnitude and statistically insignificant though. This suggests that education does not have an impact on whether a woman is ever married or not.

Moreover, the estimates of the error correlation ρ are negative, large in magnitude, and statistically significant at conventional levels. Our estimates of ρ suggest that schooling is likely to be endogenous in the decision to be married, and there are omitted variables affecting schooling and the decision to be married in opposite directions. This accords with intuition.

(f) Linear Functional Form for Fertility

To check whether our results are robust to the exponential functional form, we estimate the semiparametric ratio, crude nonparametric ratio, and modified nonparametric ratio estimators, and IVE using a linear functional form for number of children and age at first-birth. While the whole set of results are available upon request, the results for the impact of education on number of children are reported in Tables 8. These results are similar to those reported in Tables 3–6. Thus, the results with a linear functional form again indicate that schooling does not have any impact on fertility outcomes.

6 Conclusions

This study investigates the impact of education on fertility. It has been widely believed that women's education leads to a reduction in fertility. This belief arises from the conjectures that education raises a woman's value of time due to an increase in market wage and opportunities, enables women to make better use of contraceptive devices, and enhances a woman's degree of autonomy and bargaining power in fertility decisions. This prompted some social scientists to propose investing in females' education as a means of achieving demographic transition and accelerating economic development. But many early studies did not account for the endogeneity of education properly. In contrast, recent studies were more cautious in handling the endogeneity problem, and some studies used 'natural experiments'. No consensus, however, emerged, with some studies finding negative effects while some others finding no effects.

We used an exogenous variation in schooling created by the 1968 extension of compulsory education from six to nine years in Taiwan to identify the effect of education on fertility. Since the law change applied only to the women born in 1955 or later, regression discontinuity (RD) design was appropriate for our study, with age (i.e., birth year) as the "running variable". RD overcomes the schooling endogeneity problem by using only local observations around the age threshold. Our "localized" data consisting a sample of 748,547 married women were derived from Taiwan's 1990 Population Census by extracting married women born in 1952–54 (the treatment group) and 1956–58 (the control group) around the threshold year 1955.

For our empirical analysis, we proposed three new regression discontinuity estimators that are appropriate for non-negative responses including counts. The first is a semiparametric ratio estimator which leaves the distributions of the error terms unspecified. The second is a nonparametric ratio estimator motivated by exponential models for non-negative integer responses. The third is a modified version of the second to correct for the "cohort" bias due to the discreteness of the running variable. The development of these regression discontinuity estimators is a methodological contribution of this paper. While these RD estimators are "indirect ratio estimators", we also applied an instrumental variable estimator for exponential models which is a direct estimator.

We found that the 1968 Taiwanese compulsory education extension raised the education of married women in the 1956–58 cohort by 0.23 years. As for our outcomes of interest, the modified nonparametric ratio, semiparametric ratio, and IV estimators yielded essentially the same result that there is no effect of schooling on fertility measured by the number of children or by first-birth age. Also there was no evidence that education lowers the probability of being ever married.

The null effect of education on fertility may have arisen from the fact that the family planning program has already set off a steep downward trend in fertility, with education having no further effect. Another possible reason for finding no effect is that the women in our treatment group did not have high levels of education: the schooling years in our sample was 9.2 years, and only 62.85% were junior high school graduates. Education may take a negative effect only if the treated attain higher levels of education such as college completion.

Another potential reason for us to find education to have no effect on fertility is that, while education may affect fertility via several channels, what we obtain from our estimation is a total effect. Because income is not controlled for in our estimation, our estimate embraces education's indirect effect through income: Women with more education are likely to be working and have higher income, and may be able to afford to have more children. This indirect positive effect of education through income may counteract its effects on fertility through other channels. Nevertheless, these are only our conjectures. Further investigations are needed to verify to what extent these conjectures are correct. Our conclusion remains that education does not affect fertility.

Appendix

A1. Asymptotic Normality of Crude Nonparametric Ratio Estimator

Define

$$\lambda_\nu \equiv \lim_{N\to\infty} \frac{N_\nu}{N}, \quad \nu = 0, 1.$$

Expand $\sqrt{N}(\widehat{f} - f)$ around (μ_{y1}, μ_{y0}) to get

$$\begin{aligned}
\sqrt{N}(\widehat{f} - f) &= T_y'(\mu_{y1}) \frac{\sqrt{N}}{N_1} \sum_j (y_{1j} - \mu_y) - T_y'(\mu_{y0}) \frac{\sqrt{N}}{N_0} \sum_i (y_{0i} - \mu_y) + o_p(1) \\
&= T_y'(\mu_{y1}) \frac{\sqrt{N}}{\sqrt{N_1}} \frac{1}{\sqrt{N_1}} \sum_j (y_{1j} - \mu_y) - T_y'(\mu_{y0}) \frac{\sqrt{N}}{\sqrt{N_0}} \frac{1}{\sqrt{N_0}} \sum_i (y_{0i} - \mu_y) + o_p(1).
\end{aligned}$$

Doing analogously,

$$\begin{aligned}
\sqrt{N}(\widehat{g} - g) &= T_s'(\mu_{s1}) \frac{1}{N_1} \sum_j (s_{1j} - \mu_{s1}) - T_s'(\mu_{s0}) \frac{1}{N_0} \sum_i (s_{0i} - \mu_{s0}) \\
&= T_s'(\mu_{s1}) \frac{\sqrt{N}}{\sqrt{N_1}} \frac{1}{\sqrt{N_1}} \sum_j (s_{1j} - \mu_{s1}) - T_s'(\mu_{s0}) \frac{\sqrt{N}}{\sqrt{N_0}} \frac{1}{\sqrt{N_0}} \sum_i (s_{0i} - \mu_{s0}) + o_p(1).
\end{aligned}$$

Observe

$$\begin{aligned}
\sqrt{N}\left(\frac{\widehat{f}}{\widehat{g}} - \frac{f}{g}\right) &= \sqrt{N} \frac{\widehat{f}g - f\widehat{g}}{\widehat{g}g} \\
&= \sqrt{N} \frac{(\widehat{f}g - fg) + (fg - f\widehat{g})}{\widehat{g}g} \\
&= \frac{1}{\widehat{g}} \sqrt{N}(\widehat{f} - f) - \frac{f}{\widehat{g}g} \sqrt{N}(\widehat{g} - g) \\
&= \frac{1}{g} \sqrt{N}(\widehat{f} - f) - \frac{f}{g^2} \sqrt{N}(\widehat{g} - g) + o_p(1).
\end{aligned}$$

Insert the above expressions of $\sqrt{N}(\widehat{f} - f)$ and $\sqrt{N}(\widehat{g} - g)$ into this to get

$$\begin{aligned}
\sqrt{N}\left(\frac{\widehat{f}}{\widehat{g}} - \frac{f}{g}\right) &= \frac{T_y'(\mu_{y1})}{g} \frac{\sqrt{N}}{\sqrt{N_1}} \frac{1}{\sqrt{N_1}} \sum_j (y_{1j} - \mu_y) - \frac{T_y'(\mu_{y0})}{g} \frac{\sqrt{N}}{\sqrt{N_0}} \frac{1}{\sqrt{N_0}} \sum_i (y_{0i} - \mu_y) \\
&\quad - \frac{f}{g^2} T_s'(\mu_{s1}) \frac{\sqrt{N}}{\sqrt{N_1}} \frac{1}{\sqrt{N_1}} \sum_j (s_{1j} - \mu_{s1}) + \frac{f}{g^2} T_s'(\mu_{s0}) \frac{\sqrt{N}}{\sqrt{N_0}} \frac{1}{\sqrt{N_0}} \sum_i (s_{0i} - \mu_{s0}) + o_p(1).
\end{aligned}$$

Separating the terms with $\sum_i$ from the terms with $\sum_j$, this becomes

$$\begin{aligned}
\sqrt{N}\left(\frac{\widehat{f}}{\widehat{g}} - \frac{f}{g}\right) &= \frac{\sqrt{N}}{\sqrt{N_1}} \frac{1}{\sqrt{N_1}} \sum_j \left\{ \frac{T_y'(\mu_{y1})}{g} (y_{1j} - \mu_y) - \frac{f}{g^2} T_s'(\mu_{s1})(s_{1j} - \mu_{s1}) \right\} \\
&\quad - \frac{\sqrt{N}}{\sqrt{N_0}} \frac{1}{\sqrt{N_0}} \sum_i \left\{ \frac{T_y'(\mu_{y0})}{g} (y_{0i} - \mu_y) - \frac{f}{g^2} T_s'(\mu_{s0})(s_{0i} - \mu_{s0}) \right\}.
\end{aligned}$$

The asymptotic distribution of this is normal with the variance

$$\lambda_1^{-1} E[\,\{\frac{T_y'(\mu_{y1})}{g}(y_1-\mu_y) - \frac{f}{g^2} T_s'(\mu_{s1})(s_1-\mu_{s1})\}^2 | d=1\,]$$
$$+\lambda_0^{-1} E[\,\{\frac{T_y'(\mu_{y0})}{g}(y_0-\mu_y) - \frac{f}{g^2} T_s'(\mu_{s0})(s_{0j}-\mu_{s0})\}^2\, | d=0\,].$$

A consistent estimator for this was shown in the main text.

A2. Asymptotic Normality of Modified Nonparametric Ratio Estimator

Define

$$\lambda_{jk} \equiv \lim_{N\to\infty} \frac{N_{jk}}{N}, \quad j,k=0,1.$$

Recall

$$\widehat{p} \equiv T_y(\widehat{\mu}_{y10}) - T_y(\widehat{\mu}_{y01}) + 0.5\{T_y(\widehat{\mu}_{y00}) - T_y(\widehat{\mu}_{y11})\}$$

and expand $\sqrt{N}(\widehat{p}-p)$ around $(\mu_{y10}, \mu_{y01}, \mu_{y00}, \mu_{y11})$ to get

$$\begin{aligned}\sqrt{N}(\widehat{p}-p) &= T_y'(\mu_{y10})\frac{\sqrt{N}}{\sqrt{N_{10}}}\frac{1}{\sqrt{N_{10}}}\sum_k (y_{10k}-\mu_{y10}) - T_y'(\mu_{y01})\frac{\sqrt{N}}{\sqrt{N_{01}}}\frac{1}{\sqrt{N_{01}}}\sum_j (y_{01j}-\mu_{y01}) \\ &\quad +0.5\left\{T_y'(\mu_{y00})\frac{\sqrt{N}}{\sqrt{N_{00}}}\frac{1}{\sqrt{N_{00}}}\sum_i (y_{00i}-\mu_{y00}) - T_y'(\mu_{y11})\frac{\sqrt{N}}{\sqrt{N_{11}}}\frac{1}{\sqrt{N_{11}}}\sum_l (y_{11l}-\mu_{y11})\right\} + o_p(1).\end{aligned}$$

Doing analogously,

$$\begin{aligned}\sqrt{N}(\widehat{q}-q) &= T_s'(\mu_{s10})\frac{\sqrt{N}}{\sqrt{N_{10}}}\frac{1}{\sqrt{N_{10}}}\sum_k (s_{10k}-\mu_{s10}) - T_s'(\mu_{s01})\frac{\sqrt{N}}{\sqrt{N_{01}}}\frac{1}{\sqrt{N_{01}}}\sum_j (s_{01j}-\mu_{s01}) \\ &\quad +0.5\left\{T_s'(\mu_{s00})\frac{\sqrt{N}}{\sqrt{N_{00}}}\frac{1}{\sqrt{N_{00}}}\sum_i (s_{00i}-\mu_{s00}) - T_s'(\mu_{s11})\frac{\sqrt{N}}{\sqrt{N_{11}}}\frac{1}{\sqrt{N_{11}}}\sum_l (s_{11l}-\mu_{s11})\right\} + o_p(1).\end{aligned}$$

Observe

$$\sqrt{N}\left(\frac{\widehat{p}}{\widehat{q}} - \frac{\widehat{p}}{q}\right) = \frac{1}{q}\sqrt{N}(\widehat{p}-p) - \frac{p}{q^2}\sqrt{N}(\widehat{q}-q) + o_p(1).$$

Insert the above expressions of $\sqrt{N}(\widehat{p}-p)$ and $\sqrt{N}(\widehat{q}-q)$ into this to get

$$\begin{aligned}\sqrt{N}\left(\frac{\widehat{p}}{\widehat{q}} - \frac{p}{q}\right) &= \frac{T_y'(\mu_{y10})}{q}\frac{\sqrt{N}}{\sqrt{N_{10}}}\frac{1}{\sqrt{N_{10}}}\sum_k (y_{10k}-\mu_{y10}) - \frac{T_y'(\mu_{y01})}{q}\frac{\sqrt{N}}{\sqrt{N_{01}}}\frac{1}{\sqrt{N_{01}}}\sum_j (y_{01j}-\mu_{y01}) \\ &\quad +0.5\left\{\frac{T_y'(\mu_{y00})}{q}\frac{\sqrt{N}}{\sqrt{N_{00}}}\frac{1}{\sqrt{N_{00}}}\sum_i (y_{00i}-\mu_{y00}) - \frac{T_y'(\mu_{y11})}{q}\frac{\sqrt{N}}{\sqrt{N_{11}}}\frac{1}{\sqrt{N_{11}}}\sum_l (y_{11l}-\mu_{y11})\right\} \\ &\quad -\left\{\frac{pT_s'(\mu_{s10})}{q^2}\frac{\sqrt{N}}{\sqrt{N_{10}}}\frac{1}{\sqrt{N_{10}}}\sum_k (s_{10k}-\mu_{s10}) - \frac{pT_s'(\mu_{s01})}{q^2}\frac{\sqrt{N}}{\sqrt{N_{01}}}\frac{1}{\sqrt{N_{01}}}\sum_j (s_{01j}-\mu_{s01})\right\} \\ &\quad -0.5\left\{\frac{pT_s'(\mu_{s00})}{q^2}\frac{\sqrt{N}}{\sqrt{N_{00}}}\frac{1}{\sqrt{N_{00}}}\sum_i (s_{00i}-\mu_{s00}) - \frac{pT_s'(\mu_{s11})}{q^2}\frac{\sqrt{N}}{\sqrt{N_{11}}}\frac{1}{\sqrt{N_{11}}}\sum_l (s_{11l}-\mu_{s11})\right\} + o_p(1).\end{aligned}$$

Collecting the terms with $\sum_i$, $\sum_j$, $\sum_k$ and $\sum_l$, this becomes

$$\begin{aligned}
\sqrt{N}\left(\frac{\widehat{p}}{\widehat{q}}-\frac{p}{q}\right) \quad = \quad & \frac{\sqrt{N}}{\sqrt{N_{10}}}\frac{1}{\sqrt{N_{10}}}\sum_k\left\{\frac{T_y'(\mu_{y10})}{q}(y_{10k}-\mu_{y10})-\frac{pT_s'(\mu_{s10})}{q^2}(s_{10k}-\mu_{s10})\right\} \\
& -\frac{\sqrt{N}}{\sqrt{N_{01}}}\frac{1}{\sqrt{N_{01}}}\sum_j\left\{\frac{T_y'(\mu_{y01})}{q}(y_{01j}-\mu_{y01})-\frac{pT_s'(\mu_{s01})}{q^2}(s_{01j}-\mu_{s01})\right\} \\
& +0.5\frac{\sqrt{N}}{\sqrt{N_{00}}}\frac{1}{\sqrt{N_{00}}}\sum_i\left\{\frac{T_y'(\mu_{y00})}{q}(y_{00i}-\mu_{y00})-\frac{pT_s'(\mu_{s00})}{q^2}(s_{00i}-\mu_{s00})\right\} \\
& -0.5\frac{\sqrt{N}}{\sqrt{N_{11}}}\frac{1}{\sqrt{N_{11}}}\sum_l\left\{\frac{T_y'(\mu_{y11})}{q}(y_{11l}-\mu_{y11})-\frac{pT_s'(\mu_{s11})}{q^2}(s_{11l}-\mu_{s11})\right\}+o_p(1).
\end{aligned}$$

The asymptotic distribution of this is normal with the variance consistently estimable with the matrix shown in the main text.

References

[1] Acemoglu, D. and J. Angrist (2000), "How Large are Human-Capital Externalities? Evidence from Compulsory-Schooling Laws" in: *NBER Macroeconomics Annual,* Volume 15, B.S. Bernanke and K. Rogoff (eds.), Boston: MIT Press, 9–59.

[2] Axinn, W.G. and J.S. Barber (2001), "Mass Education and Fertility Transition," *American Sociological Review,* 66, 481–505.

[3] Becker, G.S. (1981), *A Treatise on the Family,* Cambridge, MA: Harvard University Press.

[4] Black, S.E., P.J. Devereux, and K.G. Salvanes (2008), "Staying in the Classroom and out of the Maternity Ward? The Effect of Compulsory Schooling Laws on Teenage Births," *Economic Journal*, 118, 1025–1054.

[5] Caldwell, J.C. (1980), "Mass Education as a Determinant of the Timing of Fertility Decline," *Population and Development Review,* 6, 225–255.

[6] Card, D. (1999), "The Causal Effect of Education on Earnings," in *Handbook of Labor Economics,* Volume III, O. Ashenfelter and Card (eds.), Amsterdam: Elsevier.

[7] Chow, L.P. (1965), "A Programme to Control Fertility in Taiwan: Setting, Accomplishment and Evaluation *Population Studies,* 19, 155–166.

[8] Chow, L.P. (1974), "The Island-Wide Family Planning Programme in Taiwan: Analysis of the Accomplishments of the Past Eight Years," *Population Studies,* 28, 107–126.

[9] Cohen, J.E. (2008), "Make Secondary Education Universal," *Nature,* 456, 572–573.

[10] Currie, J. and E. Moretti (2003), "Mother's Education and the Intergenerational Transmission of Human Capital: Evidence from college Openings and Longitudinal data," *Quarterly Journal of Economics,* 118, 1495–1532.

[11] Dee, T.S. (2004) "Are There Civic Returns to Education?" *Journal of Public Economics,* 88, 1697–1720

[12] Derose, L.F. and Ø. Kravdal (2007), "Educational Reversals and First-Birth Timing in Sub-Saharan Africa: A Dynamic Multilevel Approach" *Demography,* 44, 59–77

[13] Grossman, M. (2006), "Education and Nonmarket Outcomes," in *Handbook of the Economics of Education,* Volume 1, 2006, E. Hanushek and F. Welch (ed.), Amsterdam: Elsevier, 577–633

[14] Hsieh, C.-T. and D.E. Clark (2000), "Schooling and Labor Market Impact of the 1968 Nine-Year Education Program in Taiwan," Working Paper 1225, Research Program in Development Studies, Woodrow Wilson School of Public and International Affairs, Princeton University.

[15] Imbens, G.W., and T. Lemieux (2008), "Regression discontinuity designs: A guide to practice," *Journal of Econometrics,* 142, 615–635.

[16] Iranzo, S. and G. Peri (2009), "Schooling Externalities, Technology, and Productivity: Theory and Evidence from U.S. States," *Review of Economics and Statistics,* 91, 420–431.

[17] Lee, D.S. and T. Lemieux (2010), "Regression discontinuity designs in economics," *Journal of Economic Literature,* 48, 281–355.

[18] Lam, D. and S. Duryea (1999), "Effects of Schooling on Fertility, Labor Supply, and Investments in Children, with Evidence from Brazil," *Journal of Human Resources,* 34, 160–192.

[19] León, A. (2004) "The Effect of Education on Fertility: Evidence from Compulsory Schooling Laws," Working Paper, University of Pittsburgh.

[20] Lleras-Muney, A. (2005), "The Relationship between Education and Adult Mortality in the United States," *Review of Economic Studies,* 72, 189-221

[21] Lloyd, C.B., C.E. Kaufman, and P. Hewett (2000), "The Spread of Primary Schooling in Sub-Saharan African: Implications for Fertility Change," *Population and Development Review,* 26, 483–515.

[22] Lochner, L. and E. Moretti (2004), "The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports," *American Economic Review,* 94, 155–189.

[23] Lucas, R. (1988), "On the Mechanics of Economic Development," *Journal of Monetary Economics*, 22, 3–42.

[24] Mankiw, N.G., D. Romer, and D.N. Weil (1992), "A contribution to the Empirics of Economic Growth," *Quarterly Journal of Economics*, 107, 407–437.

[25] Mas, A. and E. Moretti (2009), "Peers at Work," *American Economic Review,* 99, 112–145.

[26] McCrary, J. and H. Royer (2010), "The Effect of Female Education on Fertility and Infant Health: Evidence from School Entry Laws Using Exact Date of Birth," *American Economic Review*, 101, 158–195.

[27] Michael, R.T. (1973), "Education and the Derived Demand for Children," *Journal of Political Economy,* 81, pp. S128–S164

[28] Milligan, K., E. Moretti, and P. Oreopoulos (2004), "Does Education Improve Citizenship? Evidence from the United States and the United Kingdom," *Journal of Public Economics* 88, 1667–1695

[29] Monstad, K., C. Propper, K.G. Salvanes (2008), "Education and Fertility: Evidence from a Natural Experiment" *Scandinavian. Journal of Economics,* 110, 827–852.

[30] Moretti, E. (2004), "Estimating the Social Return to Higher Education: Evidence from Cross-Sectional and Longitudinal Data," *Journal of Econometrics,* 121, 175–212.

[31] Mullahy, J. (1997), "Instrumental-Variable Estimation of Count Data Models: Applications to Models of Cigarette Smoking Behavior," *Review of Economic Statistics,* 79, 586–593.

[32] Oreopoulos, P., M.E. Page, and A.H. Stevens (2006), "The Intergenerational Effects of Compulsory Schooling," *Journal of Labor Economics,* 24, 729–760.

[33] Oreopoulos, P. (2007), "Do Dropouts Drop out Too Soon? Wealth, Health and Happiness from Compulsory Schooling," *Journal of Public Economics,* 91, 2213–2229.

[34] Osili, U.O. and B.T. Long (2008), "Does Female Schooling Reduce Fertility? Evidence from Nigeria," *Journal of Development Economics*, 87, 57–75

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

[36] Portes, A. (2006), "Institutions and Development: A Conceptual Reanalysis," *Population and Development Review,* 32, 233–262.

[37] Rindfuss, R.R., S.P. Morgan, and K. Offutt (1996) "Education and the Changing Age Pattern of American Fertility: 1963–1989," *Demography,* 33, 277–290

[38] Romer, P.M. (1990), "Endogenous Technological Change," *Journal of Political Economy*, 98, S71–S102.

[39] Rosenzweig, M.R. and T.P. Schultz (1985), "The Demand for and Supply of Births: Fertility and its Life Cycle Consequences," *American Economic Review,* 75, 992–1015.

[40] Rosenzweig, M.R. and D.A. Seiver (1982), "Education and Contraceptive Choice: A Conditional Demand Framework," *International Economic Review,* 23, 171–198.

[41] Sander, W. (1992), "The Effect of Women's Schooling on Fertility," *Economic Letters*, 40, 229–233.

[42] Sathar, Z.A., C.B. Lloyd, C. Mete, and M. Haque (2003), "Schooling Opportunities for Girls as a Stimulus for Fertility Change in Rural Pakistan" *Economic Development and Cultural Change,* 51, 677–698.

[43] Schultz, T.P. (1997), "Demand for Children in Low Income Countries" *Handbook of population and family economics*, , Mark Rosenzweig and Oded Stark, (eds.), Amsterdam: Elsevier Press.

[44] Tsurumi, E.P. (I979) "Education and Assimilation Taiwan under Japanese Rule, 1895–1945," *Modern Asian Studies* 13, 617–641.