Schooling Expansion and Marriage Market:
Evidence from Indonesia *

Danyan Zha†

November 15, 2018

Click here for the latest version.

Abstract

This paper analyzes how education distribution affects the marriage market, by exploiting a massive primary school construction program in Indonesia in the late 1970s as a quasi-natural experiment. Using the variation across regions in the number of schools constructed and the variation across birth cohorts, I show that in densely populated areas, primary school construction did not affect primary school attainment rate. More surprisingly, the program decreased secondary school attainment rate for both men and women due to the quality deterioration in secondary education. With this change in the education distribution as a source of variation, I find a woman marries earlier and spousal age gap increases when average education of other women decreases. I then rationalize this finding by developing a two-to-one dimensional matching model with Transferable Utilities in an OLG framework, in which the marital surplus allows complementarity between men’s education and both characteristics of women: education and younger age.

Keywords: Marriage market; Female marriage age; Spousal age gap; Schooling expansion

JEL classifications: D10, I24, I25, I28, J12, O12, O15

---

*I am very grateful to Pierre-André Chiappori, Bernard Salanié, and Cristian Kiki Pop-Eleches for their continuing guidance and support. I am indebted to Esther Duflo for sharing the INPRES school construction intensity data with me. All errors are my own.

†Department of Economics, Columbia University. Contact: dz2266@columbia.edu
1 Introduction

Recent decades have witnessed world-wide schooling expansion, especially in low- and middle-income countries (Bank, 2018). There is a large literature studying the impact of these policies on wage, income, and health outcomes of individuals. However, few researchers have investigated the impact on marital outcomes, an important dimension of individuals’ life. How should marriage market respond to changes in education distribution? Matching theories suggest that individuals’ marital outcomes depend on the marriage market conditions including the characteristic distributions of men and women. Hence a schooling expansion policy may affect an individual’s marriage outcome via changing his/her education and changing the education of others.

In this paper, I exploit the setting of primary school construction in Indonesia in the late 1970s as a quasi-natural experiment to answer this question: how a market level shock to education distribution affects marital outcomes, in particular, female marriage age and spousal age gap. I find a woman marry earlier and the spousal age gap increases when fewer women graduate from secondary school in her birth cohort and the education distribution of their potential husbands doesn’t change. To explain the empirical finding, I then develop a two-to-one dimensional matching model with Transferable Utilities (TU) in an overlapping generation (OLG) framework where men differ in education, and women differ in both education and age. I then show that the empirical finding is consistent with an assumption of complementarity between men’s education and women’s education, and a complementarity between men’s education and women’s young age in the marital surplus a couple can make.

The massive magnitude of the primary school construction program in Indonesia makes it a good setting to answer the question by providing a large exogenous shock to the education distribution. Since oil price increase in 1972, Indonesian government has experienced a huge revenue increase. This facilitated one of the largest education expansion program in the world: INPRES Sekolah Dasar. About one primary school was built per 500 primary school age kids between 1972/73 and 1978/79. This creates potential variation in education distribution across regions and across birth cohorts.

Using both variations, my identification strategy is difference-in-difference similar to other pa-
pers studying this program. (Duflo, 2001; Ashraf et al., 2016) One difference comes from the difference in construction intensity across regions, defined as the average number of schools constructed per 1000 kids between 1972/73 and 1978/79; the other difference comes from individual birth cohorts. INPRES SD is a program aiming for equality, hence more schools were built at regions where there were initially a larger number of un-enrolled school age kids. Children go to primary school between age 7 and 12. Therefore children who were of age 13 or larger at 1974 wouldn’t be impacted by the primary school construction. Kids who were younger than 13 would be impacted by this program.

The first part of my paper focuses on the impact of the school construction program on individuals’ education, which is my source of variation on marriage market outcomes. This part is building upon earlier studies using the same schooling expansion program (Duflo, 2001; Ashraf et al., 2016). I replicate some of their findings but also find some surprising ones that were not mentioned in previous literature. Consistent with previous findings, there is a positive effect on primary school attainment rate\(^1\) for men but not for women. However, I find a negative effect on secondary school attainment rate for women. As suggested in Duflo (2001), the program may have different effect in sparsely populated and densely populated areas. Exploring the heterogeneity effect depending on population density of the regions, I find that in sparsely populated regions, the school construction program has a positive effect on both primary school attainment rate and secondary school attainment rate for men but not for women; in densely populated regions, the school construction program doesn’t have a effect on primary school attainment rate but has a negative effect on secondary school attainment rate for both men and women.

I then investigate two potential mechanisms leading to the negative secondary school attainment result: (1) decrease in secondary school quality due to resources crowding out; (2) decrease in primary school quality due to the massive construction. The analysis supports the first mechanism. Building primary schools increases total demand for teachers in the region, which may affect the availability of teachers in secondary education. Moreover, demand for teachers can be more competitive in densely populated regions than sparsely populated regions since it’s earlier to relocate

\(^1\)Primary school attainment rate is defined as the percentage of people who complete primary school or above. Similarly, secondary school attainment rate is defined as the percentage of people who complete secondary school or above.
for teachers in the former area. I show that total number of teachers and average number of teachers in secondary school increase less in regions with more primary schools constructed after the launch of the school construction program. The negative effect on teacher availability in secondary education in future years only exists in densely populated regions, but not in sparsely populated regions. At the same time, building primary schools rapidly could also potentially decrease the quality of primary education. Using the education level of primary school teachers in the censuses as a proxy for school quality, I show that teacher education increases less in regions with more primary schools constructed. However, I don’t find a heterogeneity effect between sparsely and densely populated regions. In summary, the negative result on secondary school attainment rate is mostly due to the crowding out of teacher resources in secondary education because of the primary school construction.

The second part of my paper brings in a theoretical framework understanding how female marriage age reacts to the change in education distribution of men and women across cohorts. To incorporate marriage age as a choice for women, I build a two period OLG model where women can choose to seek partners either at the first period or the second, but men all marry at the second period. In any given year, the marriage market unfolds as in Choo & Siow (2006), where the marital surplus generated by a couple depends on their types and some idiosyncratic draws modeled by random vectors. Women differ in two dimensions: education and age while men only differ in one dimension: education. In a stationary equilibrium, a woman’s expected return from marriage market should be equalized between choosing to marry at the first period or the second.

How the percentage of women choosing to marry at the first period changes with respect to education distributions of men and women will depend on the assumptions on how male education interacts with female age and female education in the martial surplus. If there is no interaction between male education and female age in the marital surplus, female marriage age choice doesn’t depend on the education distribution of men or women. Intuitively, individuals’ gain from marriage comes from his/her marginal contribution to the martial surplus. Hence in this case, women will fully capture the contribution of their age in the marital surplus to their own utilities. Therefore the marriage age decision would be fully determined by how young age and old age contribute differently in the marital surplus.
Suppose women's young age is "good" in marital surplus, then in the case where there is complementarity between men’s education and women’s young age, the model predicts that an increase in male education would decrease female marriage age (i.e. increase the percentage of women marrying at first period). If there is also complementarity between men’s education and women’s education, an increase in average female education would have an opposite effect: increase female marriage age (i.e. decrease the percentage of females marrying at first period). Intuitively, an increase in female education create a relative shortage of educated males when there is complementarity between men’s education and women’s education. Hence it would have an opposite effect compared to the increase in female education.

In the last part of this paper, I examine the impact of the school construction program on female first marriage age and spousal age gap. Average spousal age gap is 5 years in Indonesia, hence for the first few cohorts of women who were impacted by the school construction program, the education of their potential husbands was minimally impacted. Therefore by comparing these female cohorts with the older cohorts who were not impacted by the program, I could see how female marriage age reacts to the change in female education distribution holding male education distribution the same. In sparsely populated regions where there is no effect on female education, as expected I do not observe any effect on female first marriage age and spousal age gap. In densely populated regions where there is a negative effect on secondary school attainment rate for women, I find a decrease in average female marriage age and an increase in spousal age gap. This is consistent with the model prediction when women’s young age is "good" and there exists both complementarity between men’s education and women’s education, and complementarity between men’s education and women’s young age in marital surplus.

I then proceed to quantify the effect by first estimating the impact of education distribution change on female first marriage age using the school construction program as an instrument variable for the percentage of female secondary school graduates and found 10 percentage points decrease in secondary school attainment rate leads to a decrease in average female marriage age by 1.1 years and an increase in average spousal age gap by 0.35 years. After removing the impact of individual education, average female marriage age decreases by 0.7 years and spousal age gap increases by 0.2 years.
This paper is related to several distinct literatures. The modeling approach in this paper is built on previous research studying marriage age using overlapping generation (OLG) model (Bhaskar, 2015; Iyigun & Lafortune, 2016; Zhang, 2018) and a model of matching with transferable utility with separable idiosyncratic preferences in marital surplus (Choo & Siow, 2006; Chiappori et al., 2017; Galichon & Salanié, 2015). In these OLG papers, some only focus on age (Bhaskar, 2015), the others study individuals’ education decision and marriage age decision simultaneously. Having an endogenous education is an attractive feature by itself, but it’s less attractive in answering the current research question that how marriage age responds to a change in education distribution in the marriage market.

My paper contributes to a growing literature studying the impact of education reform on marriage market. Hener & Wilson (2018) studies a compulsory reform in UK and finds that females decrease marital age gap to avoid marrying less qualified men. André & Dupraz (2018) studies school construction in Cameroon and finds that education increases the likelihood to be in a polygamous union for both males and females. Unlike both papers, mine analyzed the effect via a general equilibrium framework.

My paper complements a large literature on the impact of marriage market conditions on individuals’ outcomes. Most of the existing literature focus on sex ratio in the marriage market. (e.g. Abramitzky et al., 2011; Angrist, 2002; Charles & Luoh, 2010) I focus on a distinct yet equally important dimension of marriage market condition: education distribution of males and females.

Lastly, the paper also adds to the large literature studying Indonesia INPRES program. (Duflot, 2001; Breierova & Duflot, 2004; Ashraf et al., 2016; Dominguez, 2014). Previous papers only mentioned the negative effect on secondary education, this paper complements them by providing evidence for plausible mechanisms.

The rest of the paper proceeds as follows. Next section discusses the model and its predictions. Section 3 describes the school construction program and background information on Indonesia. Section 4 describes the data and my identification strategy. Main results on education and marriage outcomes are presented in Section 5 and 6. Section 7 concludes.
2 Model

In this section, I develop a two-period Overlapping Generation (OLG) matching model with transferable utilities (TU) to study how an education distribution change across birth cohorts may affect marriage market outcomes. One key feature of the model is that it allows women to choose their marriage age to enter a marriage market that offers a higher marital return. In each marriage market, it is a two to one dimensional matching, women differ in both education and age, while men only differ in education. The detailed model set up is as follows:

There is an infinite number of discrete periods. At the beginning of each period, a unit mass of men and a unit mass of women enter the economy. Individuals differ in their education type, L or H. Women choose whether they want to seek partners at period 1 when they are young or delay this process to period 2 when they are old, and men always marry at period 2.

2.1 The marriage market

In the marriage market at period \( t \), overlapping generations of men and women meet and bargain over the division of their marriage surplus to reach a stable equilibrium. Since men always marry at period 2, there are two types of men in one marriage market: L and H. Women can choose to marry at either period 1 and period 2, hence there are at most four types of women in one marriage market: \( L^0, L^+, H^0, H^+ \), where \( \{L, H\} \) indicate education and \( \{0, +\} \) indicate whether women marry in period 1 when they are young or delay marriage to period 2 when they are old.

**Martial surplus in TU framework with heterogeneities.** When man \( i \) and woman \( j \) meet and marry as a couple, there exists a martial surplus \( S_{ij} \) where they can choose how to divide within themselves. The marital surplus has two parts: one deterministic part (\( \Phi \)) depending on the types, i.e., education of man \( i \), the education and age of woman \( j \); another stochastic part that captures everything else that’s not observable by researchers but by individuals themselves: \( \varepsilon_{ij} \) for man \( i \), \( \eta_{ij} \) for woman \( j \).

**Deterministic part** The deterministic part depends on man \( i \)’s education, woman \( j \)’s education and age, denote \( \phi_{xy}^o \) as the deterministic surplus where \( x \) is male education, \( y \) is female education.
and $a$ is female age. Moreover, in our model here:

$$x \in \{L, H\}, y \in \{L, H\}, a \in \{0, +\}$$

A higher value in $x$ or $y$ indicates higher education; $a = 0$ indicates women who marry at period 1 of their lifetime, $a = +$ indicates women who delay marriage to period 2 of their lifetime. Denote

$$\delta_y^a = \phi_{Hy} - \phi_{Ly}.$$

A strict super-modularity in education would imply:

$$\delta_H^a > \delta_L^a, \forall a \in \{0, +\}$$

A strict super-modularity in men’s education and women’s young age would imply:

$$\delta_y^0 > \delta_y^+, \forall y \in \{L, H\}$$

**Stochastic part** In addition to the deterministic part depending on types, the stochastic part captures everything else in the marital surplus and is modeled as random vectors. Following Choo & Siow (2006), let’s assume the unobserved part of the marital surplus is individual-specific and only depends on the spouse’s type. For instance, woman $j$ has a vector of marital preferences

$$\eta_j = (\eta_{0j}, \eta_{Lj}, \eta_{Hj})$$

where $\eta_{xj}$ denotes the idiosyncratic utility woman $j$ derives from marrying a spouse with type $x$ (and, $\eta_{0j}$ denotes utility woman $j$ derives from being single). Similarly, man $i$’s idiosyncratic marital utilities are described by the vector

$$\varepsilon_i = (\varepsilon_{0i}, \varepsilon_{iL}, \varepsilon_{iL}^+, \varepsilon_{iH}, \varepsilon_{iH}^+)$$

**Assumption (Gumbel):** The random terms $\eta_{xj}$, $\varepsilon_{iy}^0$ follow independent Gumbel distributions $G(-k, 1)$, with $k \simeq 0.5772$ the Euler constant.
Finally, let’s assume that the two parts are additively separable. To be more precise, the total marital surplus $S_{ij}$ generated by a match between man $i$ and woman $j$ is the sum of the two components.

$$S_{ij} = \phi^a_{xy} + (\varepsilon^a_{iy} - \varepsilon_{i0}) + (\eta_{xj} - \eta_{0j})$$

**Stable matching and expected marital return.**

Given previous marital surplus and any marginals that indicate the mass of men and women of each type, I can solve the expected return each type can obtain from entering one particular marriage market. With Gumbel assumptions, I can further express the expected marital return with single rate of one type.

Denote $(G_m, G_w)$ as the distribution of male types and female types in one marriage market, we know that the expected return for each type should depend on $(G_m, G_w)$.

$$u_x(G_m, G_w), v_y(G_m, G_w)$$

### 2.2 Stationary Equilibrium with OLG

Let’s now analyze the stationary equilibrium when women choose whether they enter the marriage market at period 1 or period 2. Define $q^0_y$ (resp. $q^+_y$) as the percentage of women with education $y$ that choose to marry at period 1 (period 2). I say that strategies $(q_L, q_H)$ induce the marriage market $(G_m, G_w)$ if the distributions of female types is $G_w$ when women with education $y$ choose strategy $q_y$.

**Definition.** Stationary equilibrium $q_y$ is a stationary equilibrium strategy if $v^0_y = v^+_y$ where:

- $v^0_y$ is the expected marriage payoff of woman with education $y$ who chooses to enter the marriage market at period 1 in the induced marriage market $(G_m, G_w)$;
- $v^+_y$ is the expected marriage payoff of woman with education $y$ who chooses to enter the marriage market at period 2 in the induced marriage market $(G_m, G_w)$.

In equilibrium, women should be indifferent between marrying in period 1 or delaying marriage to period 2 since they are free to choose.
Proposition 1. With the Gumbel assumption, there exists a unique stationary equilibrium and the equilibrium strategies \( q_y \) satisfy:

\[
\min(\exp(\frac{\phi^{0}\_y - \phi^{+}\_1y}{2}), \exp(\frac{\phi^{0}\_y - \phi^{+}\_2y}{2})) \leq \frac{q^0_y}{q^+_y} \leq \max(\exp(\frac{\phi^{0}\_y - \phi^{+}\_1y}{2}), \exp(\frac{\phi^{0}\_y - \phi^{+}\_2y}{2})), \forall y \in \{L, H\}
\]

Proof. See the appendix

Intuitively, the equilibrium percentage of women who decide to marry at period 1 depends on the marital surplus difference of marrying at period 1 and period 2.

Proposition 2. If given type \( y' \), \( \delta^+_y = \delta^0_y \), then \( q^*_y \) is uniquely pinned down by:

\[
q^+_y = \frac{\exp(\phi^+_L y')}{\exp(\phi^+_L y') + \exp(\phi^0_L y')} m_y, \quad q^0_y = \frac{\exp(\phi^0_L y')}{\exp(\phi^+_L y') + \exp(\phi^0_L y')} m_y
\]

where \( m_y \) is the percentage of women with education level \( y \) each period.

Proof. See the appendix

If there doesn’t exist complementarity between men’s education and women’s young age for women with education \( y' \), then the equilibrium strategy in terms of how many marry at period 1 for this education level is fully pinned down by the contribution of young age in marital surplus, and won’t depend on education distributions of men.

2.3 Comparative Statics

An education expansion policy would affect the education distribution of both men and women. Let’s analyze how female marriage age decision would change when the education distribution of men or women changes, respectively.

Proposition 3. Denote male education distribution as \( n = (1 - n_H, n_H) \) and female education distribution as \( m = (1 - m_H, m_H) \).

Keeping \( m \) constant, \( \forall y \in \{L, H\} \), an increase in \( n_H \) would

- increase \( \frac{q^0_y}{q^+_y} \) if \( \delta^0_y > \delta^+_y \)
- decrease \( \frac{q^0_y}{q^+_y} \) if \( \delta^0_y < \delta^+_y \)
If the percentage of higher educated men increases, the equilibrium percentage of women marrying at period 1 would increase if there is complementarity between men’s education and women’s young age in marital surplus; the equilibrium percentage of women marrying at period 1 would decrease if there is instead complementarity between male higher education and female old age in marital surplus. Notice that whether the marital surplus is super-modular in education or not doesn’t matter.

Proof. See the appendix.

Proposition 4. Denote male education distribution as \( n = (1 - n_H, n_H) \) and female education distribution as \( m = (1 - m_H, m_H) \)
Further assume (1) super-modularity in men’s education and women’s education, (2) \( n_H, m_H \) are both bounded away from 0 and 1 (i.e. there exists \( \delta_n, \delta^+_n, \delta_m, \delta^+_m \) such that \( \delta_n < n_H < \delta^+_n, \delta_m < m_H < \delta^+_m \), then:
Keeping \( n \) constant, \( \forall y \in \{L, H\} \), an increase in \( m_H \) would

- decrease \( q^0_y \) if \( \delta^+_y > \delta_y^0 \)
- increase \( q^0_y \) if \( \delta_y^0 < \delta_y^+ \)

Notice here the prediction is opposite to Proposition 3. If the percentage of higher educated women increases, super-modularity in education would imply a relative shortage of higher educated men, which creates smaller incentive for women to marry younger if male high education and female young marriage age are complementary in marital surplus.

Proof. See the appendix

3 Background

3.1 INPRES Primary School Construction Program in Indonesia

Indonesian government has always desired for widening of educational opportunity since its independence at 1945. However, due to financial difficulties and political conflict, Indonesia remained backward relative to neighboring countries and to countries at similar levels of income at the early
As late as the 1971 population census, only 62% of primary school-age children (ages 7-12 inclusive) were enrolled in any kind of school, while only 54% appeared on the rolls of public and private schools reporting to the Ministry of Education. (see Snodgrass, 1984). Due to oil production increase and the first OPEC-engineered price rise in 1972-1973 which raised government revenue in a way that people haven’t expected, a primary school construction aid program (Program Bantuan Pembangunan Sekolah Dasar), known as INPRES Sekolah Dasar and more informally as INPRES SD, was inaugurated in 1973. Besides school construction, the government also provides textbooks and teacher training to make sure the building is for education purpose. By 1983, nearly all Indonesian children at least began primary school, while the percentage of 7-12 year olds enrolled exceeded 90%. INPRES SD has been a successful case study for education policies in developing countries.

Between 1973/74 and 1978/79, 62,000 primary schools were scheduled to build. Each school consists of three classrooms and each classroom has one teacher and can accommodate 40 pupils. The allocation rule every year is as follows: (a) make sure each district (kecamatan in Indonesian, one level below regency, two levels below province) allocated at least one school and each province at least 50, (b) the remainder were distributed according to the estimated population of unenrolled 7-12 year old kids. This creates variation in the construction intensity for my empirical analysis.

3.2 Education system in Indonesia

In Indonesia, education contains six years of primary school (sekolah dasar, SD), three years of middle school (sekolah menengah pertama, SMP) and three years of high school (sekolah menengah atas, SMA), followed by various kinds of higher education. Children mostly begin primary school at age 7. Two ministries are responsible for managing the education system, with 84 percent of schools under the Ministry of National Education and the remaining 16 percent under the Ministry of Religious Affairs. In 2000 census, although 86.1 percent of the Indonesian population is registered as Muslim, only 15 percent of school-age individuals attended religious schools. (Library of congress).

INPRES 1973 initiated Indonesia’s program of compulsory education, but six year compulsory education for primary school age children (7-12 age group) was not fully implemented until 1984. On May 1994, nine year compulsory education for 7 to 15 age group was launched. 92% pupils
were enrolled in public schools for primary education, and 50% pupils were enrolled in public school for secondary education. Indonesian government focused more on primary education compared to secondary level. In 1985, among public spending on education, 62% goes to primary education while 27% goes to secondary. (see Tan & Mingat, 1992, table 3.1, table 6.5)

In 1980s, though all children began primary school, only about 62% of pupils entering primary school actually graduated from grade 6. Transition between primary school and junior secondary school was also small, about 60%. (see Jones & Hagul, 2001, table 1, figure 2). Transition between junior secondary and senior secondary was also small: 53%. However, the survival rate in junior secondary school and senior secondary school is pretty high in Indonesia, more than 90%. (see Tan & Mingat, 1992, table4.5, table 4.6, Table A.1)

4 Data and Empirical Strategy

4.1 Data

Indonesian Census Data. For the main analysis, I use information from the 10% sample Indonesian Population Census 2010 and the 0.51% sample Indonesian Intercensal Population Survey (SUPAS) 2005 downloaded from IPUMS International. The two censuses were designed to be representative of the whole country. Moreover, birth place (regency level) of an individual is recorded in both censuses, which can be used to proxy their exposure to primary school construction program when they were at primary school age. Education, current martial status and current spousal information is also available in both censuses. however, only Census 2005 records detailed lifetime martial outcomes such as first marriage age and number of marriages, moreover, only females were surveyed on those questions. For male marriage age, I can proxy their marriage age using the wife’s information if their wife is in her first marriage.

Duflo(2001) uses SUPAS 1995 for her analysis. Since I am interested in marriage market outcomes, to avoid truncation problem, i.e. young males and females who are single in the survey year may marry in future years; I choose the latest censuses available from IPUMS.

Table 1 displays the descriptive statistics of individuals’ education using 2010 census for the old cohort born between 1950 and 1961 (not exposed to the program) and young cohort born between 1962 and 1972 (exposed to the program). The number of individuals with secondary school degree is pretty small, even for the younger cohort. An increase in education is observed for the younger cohort. Males are more educated than females. For married couples, on average, husbands are 5 years older than wives, this is much larger than a gap of 2 years observed in US.

Figure 1 displays the matching patterns with respect to education for the observed married couples in which wives of age 40-50 in 2010.

INPRES Data. I borrow information on the number of schools planned to be constructed across regencies from Duflo(2001). Intensity is defined as the average number of primary schools planned to be constructed between 1973 and 1978 (inclusive) per 1000 kids with age 5-14 at the regency in 1971 census. INPRES school construction started at 1973, the last year of Repelita II (the second five year development plan), and continued through Repelita III(1974-1978) and Repelita IV (1979-1983).

Link regency code between censuses. Indonesia has experienced massive expansion of regions (Pemekaran Daerah) since the enactment of Law No.22 of 1999 concerning Regional Autonomy. The number of regencies increased from 271 in 1971 to 304 in 1995 to 437 in 2005 and to 494 in 2010. Hence I use the GIS shapefiles provided by IPUMS across census years to link the birth place regencies in 2005 and 2010 back to the birth place regency variable in 1995 to assign the proper program intensity to each individual. Since most expansion is dividing existing regency into several small regencies, I can link most of the regencies.

School intensity data is available for 290 unique regencies in Duflo(2001), which were coded using 1995 labels. There were 304 regencies in 1995, the lost 14 regencies were in East Timor, which became part of Indonesia as the 27th province in 1976.

School infrastructure data.
4.2 Identification Strategy

**Education.** To analyze how education distribution was impacted across regencies and across birth cohorts, our empirical strategy is difference-in-difference as used in citeDuflo2001. One difference comes from the school construction intensity, defined as the average number of primary schools built between 1973 and 1978 in one regency per 1000 children with age $5 \sim 14$ at 1971. The other difference comes from birth cohorts. In Indonesia, children go to primary school at age $7 \sim 12$. Those with age 13 or above at 1974 wouldn’t be impacted by the program because they would be already out of primary school. Those with age less or equal to 12 at 1974, the younger they were, the more they were exposed to this school construction program.

The quantitative effect of the school construction program on individuals born at birth cohort $k$ and regency $j$ can be estimated in the following specification:

$$y_{j k} = \alpha_j + \beta_k + \sum_{l=2}^{12} (P_j d_{kl}) \gamma_l + \sum_{l=14}^{21} (P_j d_{kl}) \gamma_l + \sum_{l=2}^{21} (C_j d_{kl}) \delta_l + \varepsilon_{jk}$$

where $y_{j k}$ is the percentage of individuals completing primary school (secondary school) born in regency $j$ and at birth cohort $k$, $d_{kl}$ is a dummy that indicates whether birth cohort $k$ individuals are age $l$ in 1974 (year-of-birth dummy). $\alpha_j$ denotes regency fixed effect, $\beta_k$ denotes birth cohort fixed effect. $P_j$ is the school construction intensity in regency $j$. $\varepsilon_{jk}$ is the error term. $C_j$ is other region-specific variables.

The coefficients $\gamma_l$ are the coefficients of interest. They represent the effect of one additional primary school constructed on the dependent variable for individuals of age $l$ in 1974. There is a testable restriction on coefficients $\gamma_l$. A valid identification strategy would require that $\gamma_l = 0$ if $l > 13$, i.e., the variation in the outcome variable is not correlated with the primary school available starting at 1974 for the kids who were already out of primary school at 1974. I should expect for $l \leq 12$, $\gamma_l > 0$. Moreover $\gamma_l$ decreases with $l$, indicating a higher impact on younger generation.

**Marriage market** One difficulty in empirical analysis on marriage market is that both men and women are potentially affected by the school construction at the same time, hence I may not be able to identify which side drives my results. However, the large positive spousal age gap in my
sample provides me a novel setting where only female education distribution changes, but not male education distribution in the marriage market. Because women marry older husbands, then for the first few cohorts of women whose education is impacted, their potential husbands are older and would not be impacted by the program. The larger the average spousal age gap norm is, the more birth cohorts of women I can attribute to the experiment where only female education changes but not their potential husbands’ in the marriage market.

The reduced form regression specification on marriage market outcomes for women is the same as previous specification for education.

5 Empirical Results on Education

In this section, I present my empirical results on education, which is my source of variation for marriage market outcomes. I first show the results on whole sample, then show the results on two subsamples depending on population density. Lastly, I provide further evidence for the mechanisms behind the different results on subsamples.

5.1 First Stage Effect on Education for the Whole Sample

First, I examine the impact on education for the whole sample so that I can introduce shifts in the distribution of education distribution for our later analysis of marriage market outcomes. In Figure 3, I plotted $\gamma_l$ when the dependent variable is the percentage of individuals who completes at least primary school for males (or females), i.e., the effect of one additional primary school constructed per 1000 kids on primary school attainment rate for males (or females) with age $l$ at 1974. For simplicity, I combined three birth cohorts together on the graph.

Two important results stand out from Figure 3. First, $\gamma_l$ is not significantly different from 0 for $l$ larger than 13 for both men and women. This gives us confidence in the identification assumption: birth cohort trend in primary school attainment rate doesn’t differ across regions with different school construction intensity. Secondly, $\gamma_l$ is positive for males with age $l \leq 12$ at 1974, indicating a positive effect on primary school attainment rate for men; $\gamma_l$ is zero for women except the youngest cohorts, with age $l \leq 3$, indicating a lagging effect on female primary school attainment rate. Both results are consistent with previous findings in Duflo (2001) and Ashraf et al. (2016).
Difference-in-difference estimates are provided in column (1)-(3) in Table 2. Following Duflo (2001), the sample includes individuals born between 1950 and 1961 who are larger than 12 at 1974, and individuals born between 1968 and 1972, who are smaller than 7 at 1974. Post indicates individuals born between 1968 and 1972. Column (3) suggests that one additional school increases male primary school attainment rate by 0.6 percentage point. This is smaller than the estimate in Figure 2 in Duflo (2001) where it’s shown that about 1.5% more individuals had at least 6 years of schooling between high program regions (where on average 2.44 schools were built) and low program regencies (where on average 1.54 schools are built). My estimate is smaller; one potential reason is that I include more controls compared to Equation (4) in Duflo (2001).

In Figure 4, I plotted the coefficients of the interactions of age in 1974 and program intensity for completing secondary school. Surprisingly, I found a negative impact on secondary school attainment rate, especially for women. This is surprising because if anything I should expect positive spillover effects from primary school completion to secondary school completion. This finding was also mentioned for men in Duflo (2001) but not discussed much there. Difference-in-difference estimates in column (4)-(6) in Table 2 suggests one additional school built decreases females’ secondary school attainment rate by 0.53 percentage point.

### 5.2 Heterogeneity Results on Education

More insight into the effect of the program can be obtained by examining its impact on different types of regions. In this section, I repeated the previous exercise on two subsamples divided by population density: sparsely populated regions with density smaller than medium density and densely populated regions with density larger than medium density. Population density is calculated as population in 1971 census divided by the area of each region in 1971. The median density (the density for the region of birth for the median person in the weighted sample) is 470 habitants per square kilometer. There are 183 regions in sparsely populated subsample and the average number of schools constructed per 1000 kids is 2.1. There are 91 regions in densely populated subsample and the average number of schools constructed per 1000 kids is 1.67, a little lower than the sparsely populated subsample.

In Figure 5 and Figure 6, I plotted the coefficients on education $\gamma_l$ for both sparsely populated and densely populated subsamples. And the difference-in-difference estimates are shown in Table 3.
As shown in Figure 5, in sparsely populated areas, the program increased primary school attainment rate (top) and secondary attainment rate (bottom) for men but didn’t affect female education. Difference-in-difference estimates are provided in Panel A of Table 3. For males, one additional school constructed per 1000 kids increased the percentage of completing primary school by 1 percentage point and the percentage of completing secondary school by 0.69 percentage point.

As shown in Figure 6, in densely populated areas, the program didn’t affect primary school attainment rate (top) but decreased secondary school attainment rate (bottom) for both men and women. DD estimates in Panel B of Table 3 suggests an additional school built per 1000 kids decreased the secondary school attainment rate by 2.3 percentage points for both men and women.

This heterogeneity effect is consistent with the finding in Duflo (2001) that the program increased years of schooling in sparsely populated areas but not in densely populated areas for males. Duflo (2001) interpreted this as evidence that the program increased male education mainly through decreasing average school distance. This could explain the difference in the results on primary school attainment rate across the two subsamples, but had no explanatory power in the negative result on secondary school attainment rate in densely populated regions.

5.3 Mechanism

In this section, I explore further the surprising negative effect on secondary school attainment rate in densely populated regions. There are at least two conjectures: (1) building primary schools crowds out resources available to secondary school and deteriorate secondary school quality (2) a sudden increase in primary school availability may decrease primary education quality and hence the quality of primary school graduates. Then I explore heterogeneity in sparsely and densely populated regions and show that the first conjecture is more plausible.

**Secondary education quality deterioration?** Teacher scarcity is always a challenge in education system in Indonesia. (need reference) Building primary schools increases the aggregate demand for teachers. This could affect the availability of secondary school teachers. To test this conjecture, I use the total number and average number of teachers per school in secondary education across regions in the years after INPRES-SD program and to check whether there is a differential change in regions with more primary schools constructed. More specifically, I estimate the following
specification:

\[ y_{jt} = \alpha_j + \beta_t + \sum_{l=2}^{6} (P_j d_{tl}) \gamma_l + \sum_{l=2}^{6} (C_j d_{tl}) \delta_l + \varepsilon_{jt} \]

where \( j \) denotes region, \( t \) denotes the survey year, where 1 indicates year 1973/74, 2 indicates year 1978/79, 3 indicates year 1983/84, 4 indicates year 1988/89, 5 indicates year 1993/94, and 6 indicates 1995/96. \( y_{jt} \) indicates the total or average number of secondary school teacher at year \( t \) in region \( j \). \( d_{tl} \) is a year dummy indicating whether \( t = l \). \( \alpha_j \) denotes regency fixed effect, \( \beta_j \) denotes year fixed effect. \( P_j \) is the school construction intensity in regency \( j \). \( \varepsilon_{jk} \) is the error term. \( C_j \) is other region-specific variables. Baseline year is 1973/74 (\( t = 1 \)).

Results are presented in Table 4. Omitted baseline year is 1973/74. Negative coefficients in column 1 and column 2 suggest that in regions with more schools constructed, a smaller increase is observed for the total number and the average number of teachers per school in secondary education in later years. Re-assuringly, Column (3) shows a positive effect of the program on the total number of teachers in primary school education, which is consistent with the teacher crowding out story.

Moreover, since negative effect on secondary school completion is only observed in densely populated regions, this negative effect on number of teachers in secondary school should also only exist in densely populated regions if this is the mechanism. Figure 7 plots the coefficients before the interaction term of year dummy and school construction intensity of previous specification for sparsely and densely populated regions respectively. A negative effect on average number of teachers in secondary education is found for densely populated regions, but not for sparsely populated regions. This confirms my conjecture that primary school construction increases demand for teachers and this crowds out teacher resource available in secondary school education, which leads to a negative effect on secondary school attainment rate. Moreover, this only exists in densely populated regions.

**Primary education quality deterioration?** A second conjecture is that the deterioration in primary school quality leads to a decrease in student quality in primary school graduates and this induces a lower secondary school attainment rate. To meet the surge in demand for teachers created by the school expansion, primary school teacher quality may be sacrificed. Jalal et al. (2009); Bharati et al. (2018) To proxy teacher quality measure in primary school, I adapted the

Empirical specification is similar as before. Outcome variable is the percentage of primary school teachers finishing secondary school (or some college) in one regency in that census year. Baseline year is 1971, before the school expansion program started.

Figure 8 shows the coefficients of the interaction term between year fixed effect and school construction intensity, for sparsely and densely populated regions separately, for the two proxies: the percentage of teachers completing secondary school (top) and completing some college (bottom). Like Bharati et al. (2018), I found a negative impact on teachers’ quality in 1976 of the program, but not for later years. However, I didn’t find different patterns between sparsely and densely populated regions. This suggests that primary education deterioration is not the main reason of the negative impact of secondary school attainment rate.

5.4 Summary.

Here is a summary of the results on Education.

**Result 1:** The program has a positive effect on primary school attainment rate for men and a surprising negative effect on secondary school attainment rate for women.

**Result 2:** In sparsely populated regions, there is a positive effect on primary school attainment and secondary school attainment rate for men but zero effect for women.

**Result 3:** In densely populated regions, for both men and women, there is no effect on primary school attainment rate, but negative effect on secondary school attainment rate.

In light of the different effects on education in sparsely and densely populated regions, I should expect different results on marriage market outcomes in sparsely and densely populated regions. Moreover, I should expect zero effect on female marriage age or spousal age gap in sparsely populated regions since female education is not impacted.

6 Results on marriage market outcomes

In this section, I present my empirical results on female marriage age and spousal age gap. I first show reduced-form event study results on the impact of school construction on female marriage
age and spousal age gap for the treated female cohorts, for sparsely and densely populated regions separately. I then provide the 2SLS estimate on how female marriage age and spousal age gap changes with respect to female’s education distribution using the school construction program as an instrument variable for education distribution.

6.1 Reduced form results

The empirical specification for reduced form results is the same as previous specification for education results. However, I dropped the first birth group because of the concern of attrition in my female sample due to death of husbands.

Figure 9 presents the coefficients of the interaction between birth cohort dummy and the school construction intensity on female first marriage age (top) and spousal age gap (bottom) by females age group at 1974 in sparsely populated regions. All coefficients of the interaction between birth cohort dummy and the school construction intensity are not significantly different from zero. This is expected since female education wasn’t affected much by the school construction program in sparsely populated regions. Results for densely populated regions are presented in Figure 10. The top panel shows a negative effect on female first marriage age for one additional primary school built in the regions. Correspondingly, the bottom panel shows a positive effect on spousal age gap.

Difference-in-difference estimates are presented in Table 5. The sample includes women born between 1953 and 1961 who are larger than 12 at 1974, and women born between 1965 and 1970. Post indicates women born between 1965 and 1970. Column (1) and (2) shows the estimates for sparsely populated regions. Neither female first marriage age nor spousal age gap was impacted. Column (3) and (4) present the estimates for densely populated regions and suggest that one additional school constructed decreases the average female first marriage age by 0.25 years and increases the spousal age gap by 0.075 years.

6.2 2SLS estimate

Since I don’t have first stage results for female education in sparsely populated regions. In this subsection, I’ll focus on densely populated regions. Consider the following equation that characterizes how own education and education distribution may affect an individual’s choice of marriage
age and the spousal age gap:

\[ y_{ijk} = \alpha_j + \beta_k + D_{ijk}c + E_{jk}b + \nu_{ijk} \]

where \( \alpha_j \) is region fixed effect, \( \beta_k \) is birth cohort fixed effect. \( y_{ijk} \) denotes the marriage age or spousal age gap of female \( i \) born at year \( k \) in region \( j \), \( D_{ijk} \) is a dummy variable denoting whether female \( j \) completes secondary school or not, and \( E_{jk} \) denotes the female secondary school attainment rate of birth cohort \( k \) at region \( j \).

Coefficient of interest is \( b \), indicating the impact of an increase in the proportion of educated females on female marriage age and the spousal age gap. However, ordinary least-squares (OLS) estimates of this equation may lead to biased estimates if there is correlation between \( E_{jk} \) with \( \nu_{ijk} \) or there is correlation between \( D_{ijk} \) and \( \nu_{ijk} \). Unobserved individual characteristics such as ability or family attitudes could affect both her education attainment and marriage decisions, leading to correlation between \( D_{ijk} \) and \( \nu_{ijk} \). Unobserved region cohort specific characteristics such as a construction of entertainment facilities or a promotion of family planning policies could affect the education attainment and marriage decisions of a few cohorts in the region, leading to correlation between \( E_{jk} \) with \( \nu_{ijk} \).

To tackle the issue, let’s take the average across individuals \( i \) given birth cohort \( k \) and region \( j \):

\[ \bar{y}_{jk} = \alpha_j + \beta_k + E_{jk}(b + c) + \bar{\nu}_{jk} \]

The school construction program provides a good instrument variable for \( E_{jk} \), hence I can get a valid estimate of \( b + c \). OLS and 2SLS estimates of this specification are shown in Panel A of Table 6 for female first marriage age and spousal age gap. The IV estimate for female first marriage age, though imprecisely estimated, indicates that increasing the percentage of female secondary graduates by 10 percentage points would increase the average female marriage age by 1.09 years. The IV estimate for female spousal age gap indicates that increasing the percentage of female secondary graduates by 10 percentage points would decrease the average spousal age gap by 0.35 years.

Separate the effect from own education and education distribution. From previous spec-
ification, we know that:

\[ E(y_{ijk}|D_{ijk} = 0) = \alpha_i + \beta_k + E_{jk}b \]
\[ E(y_{ijk}|D_{ijk} = 1) = \alpha_i + \beta_k + c + E_{jk}b \]

Hence \( c = E(y_{ijk}|D = 1) - E(y_{ijk}|D = 0) \), which can be empirically estimated as the difference of outcome variable conditional on education level. From summary statistics, we know that the difference in first marriage age between female secondary graduates and female primary graduates is 4 years, while the difference in spousal age gap between secondary graduates and primary graduates is (-1.5) years. Compare it with previous estimates, this indicates that controlling an individual’s education, increasing the percentage of female secondary graduates by 10 percentage points in her birth cohort would increase her first marriage age by 0.69 years and decrease the spousal age gap by 0.2 years.

6.3 Interpretation

My findings on marriage market are consistent with the model when there is complementarity between more education of husbands and younger age of wives in marital surplus. A decrease in the percentage of female secondary graduates creates a relative abundance of secondary graduate men, which would encourage more women to marry earlier. In the Indonesian setting, the regions with more school constructed experienced a smaller increase in female secondary graduates, which creates a relative abundance of male secondary graduates in the marriage market, and this encourages even more women to marry earlier.

7 Conclusion

In this paper, I show that women adjust their marriage age to the education of other women in the local marriage market. Exploiting a massive school construction program in the late 1970s in Indonesia, I first document an empirical finding that in densely populated regions, the secondary school attainment rate goes down for both men and women due to crowding out of teacher resources in secondary education from primary school construction. I then analyze the first marriage age of the first few cohorts of women who were exposed to the school construction program. Since spousal age gap is on average 5 years, these women’s potential husbands’ education are minimally
impacted. I find that women decrease their first marriage age when there is a decrease in average secondary school attainment rate of other women in the same cohort. To explain this, I build a two to one dimensional matching model embedding female choice of marriage age into a two period OLG framework, and show that if in marital surplus, (1) a younger wife is ”good”, (2) there exists complementarity between men’s education and women’s education, (3) there exists complementarity between men’s education and women’s young age, then women will decrease their marriage age in response to a decrease in other women’s education. Intuitively, when the education of other women decreases, they tend to marry less educated men due to the complementarity between education, this creates abundance in more educated men. Due to the complementarity between men’s education and women’s young age, the abundance in more educated men induce women to marry younger.

This project is a step toward further understanding the effect of market conditions on individuals’ marriage decisions and outcomes. Education expansion policies have been observed around the world. The empirical finding that female marriage age responds to other womens’ education has direct policy implications. When evaluating education policies with potential market-level impact, we as researchers should consider both the direct effect on individuals and the indirect effect via changing market conditions.
References


(Dissertation, Yale University)


### Table 1: Summary Statistics

<table>
<thead>
<tr>
<th></th>
<th>Old Cohorts Born between 1950 and 1961</th>
<th>Young Cohorts Born between 1962 and 1974</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Males</td>
<td>Females</td>
</tr>
<tr>
<td><strong>Education Attainment</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Some School</td>
<td>0.22</td>
<td>0.33</td>
</tr>
<tr>
<td>Primary School</td>
<td>0.56</td>
<td>0.54</td>
</tr>
<tr>
<td>Secondary School</td>
<td>0.17</td>
<td>0.11</td>
</tr>
<tr>
<td>University or above</td>
<td>0.05</td>
<td>0.02</td>
</tr>
<tr>
<td><strong>Marriage outcomes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never Married (=1)</td>
<td>0.01</td>
<td>0.02</td>
</tr>
<tr>
<td>Separated (=1)</td>
<td>0.01</td>
<td>0.04</td>
</tr>
<tr>
<td>Husband Age minus Wife Age</td>
<td>5.83</td>
<td>4.84</td>
</tr>
<tr>
<td><strong>Married Couples</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Husband More Educated (=1)</td>
<td>0.26</td>
<td>0.27</td>
</tr>
<tr>
<td>Same Education (=1)</td>
<td>0.62</td>
<td>0.63</td>
</tr>
<tr>
<td>Wife More Educated (=1)</td>
<td>0.13</td>
<td>0.11</td>
</tr>
<tr>
<td>Years of Schooling Gap (Husband’s minus Wife’s)</td>
<td>0.57</td>
<td>0.68</td>
</tr>
<tr>
<td>Observations</td>
<td>1,275,648</td>
<td>1,231,961</td>
</tr>
</tbody>
</table>

Source: Indonesian Census 2010.
Figure 1: Marriage Frequencies (left) and Marriage Proportions (right) by Education for Females
Figure 2: Primary School Construction Intensity Map

Geographic Distribution of Number of Schools Built Between 1973 and 1978 in Indonesia

<table>
<thead>
<tr>
<th>Number of schools built per 1000 kids</th>
<th>0.69 - 1.21</th>
<th>1.21 - 1.39</th>
<th>1.39 - 1.62</th>
<th>1.62 - 1.80</th>
<th>1.80 - 2.06</th>
<th>2.06 - 2.37</th>
<th>2.37 - 2.54</th>
<th>2.54 - 2.83</th>
<th>2.83 - 3.53</th>
<th>3.53 - 8.60</th>
</tr>
</thead>
</table>

Figure 3: Coefficients of the interactions: Age in 1974 * Program intensity in the region of birth in the education equation for Completing Primary school

Note: This figure reports estimates of the effect of school construction on primary school completion for 3-year cohorts separately for males and females born in one regency. Dependent variable is the percentage of individuals completing primary school when observed in 2010. The x-axis reports the age range (in 1974) for each cohort and the y-axis reports the estimated coefficient, which can be interpreted as the effect of one additional primary school built per 1000 kids on primary school attainment rate in that regency. The sample consists of individuals born between 1950 and 1972 observed in 2010 Indonesian census. The vertical line indicates the youngest cohort that did not receive any treatment from school construction, since they were out of primary school at 1974, when the first round of constructed primary schools became available. Confidence intervals of 95% were plotted. The figure shows zero effect for individuals older than 13 at 1974, but an increasing positive effect for males younger than 13. For females, the effect is smaller.
Figure 4: Coefficients of the interactions: Age in 1974 * Program intensity in the region of birth in the education equation for Completing Secondary school

Note: This figure is built like Figure 3 but considers secondary school attainment rate. It reports estimates of the effect of school construction on secondary school completion for 3-year cohorts separately for males and females born in one regency. Dependent variable is the percentage of individuals completing secondary school when observed in 2010. The x-axis reports the age range (in 1974) for each cohort and the y-axis reports the estimated coefficient, which can be interpreted as the effect of one additional primary school built per 1000 kids on primary school attainment rate in that regency. The sample consists of individuals born between 1950 and 1972 observed in 2010 Indonesian census. The vertical line indicates the youngest cohort that did not receive any treatment from school construction, since they were out of primary school at 1974, when the first round of constructed primary schools became available. Confidence intervals of 95% were plotted.
Figure 5: Effect of Program on Completing Primary School (top) and Completing Secondary School (bottom) in Sparsely Populated Areas

Note: This figure is similar to Figure 3 and Figure 4 but focuses on a subgroup: sparsely populated regions. It reports estimates of the effect of school construction on primary school completion (top) and secondary school completion (bottom) for 3-year cohorts separately for males and females in this subgroup. Sparsely populated regions are defined as those regions with population density smaller than the weighted medium density in 1971. This figure shows that:
Figure 6: Effect of Program on Completing Primary School (top) and Completing Secondary School (bottom) in Densely Populated Areas

Note: This figure is built like Figure 5 but focuses on the other subgroup: densely populated regions. It reports estimates of the effect of school construction on primary school completion (top) and secondary school completion (bottom) for 3-year cohorts separately for males and females in this subgroup. Densely populated regions are defined as those regions with population density larger than the weighted medium density in 1971. This figure shows that:
Figure 7: Coefficients of the interactions: Census year * Program intensity in the regency in Average number of secondary school teachers equation

Note: This figure reports estimates of the effect of school construction on average number of teachers in secondary school across different years in sparsely populated areas and densely populated areas. The baseline year is 1973/74. The data was provided by Indonesian Education Ministry and was collected in Duflo (2001). The dependent variable was the average number of teachers in secondary school across different regencies. This figure supports the argument that the negative effect on secondary school attainment is due to teacher resource crowding out in densely populated regions because of primary school construction.
Figure 8: Effect of Program on Primary teacher education

Note: This figure reports estimates of the effect of school construction on the education level of primary school teacher in the two subsamples: sparsely and densely populated regions. Dependent variable for the top panel is a dummy indicating the teacher completes secondary school, for the bottom panel is a dummy indicating the teacher has some post-secondary education. The baseline year is 1971. Primary school teacher information for each region is obtained from identifying those individuals who claim their occupation is primary school teacher in the census year.
Figure 9: Effect of Program on female first marriage age (top) and spousal age gap (bottom) in Sparsely Populated Areas

Note: This figure reports estimates of the effect of school construction on female first marriage age (top) and spousal age gap (bottom) for 3-year cohorts females in sparsely populated areas. The x-axis reports the age range (in 1974) for each cohort and the y-axis reports the estimated coefficient, which can be interpreted as the effect of one additional primary school built per 1000 kids on primary school attainment rate in that regency.
Figure 10: Effect of Program on female first marriage age (top) and spousal age gap (bottom) in Densely Populated Areas

Note: This figure reports estimates of the effect of school construction on female first marriage age (top) and spousal age gap (bottom) for 3-year cohorts females in densely populated areas. The x-axis reports the age range (in 1974) for each cohort and the y-axis reports the estimated coefficient, which can be interpreted as the effect of one additional primary school built per 1000 kids on primary school attainment rate in that regency.
Table 2: Effect of School Construction on Education

<table>
<thead>
<tr>
<th></th>
<th>Indicator for Completing at least:</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Primary School</td>
<td>Secondary School</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All sample:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Males:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post × Intensity</td>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.022***</td>
<td>0.0079*</td>
<td>0.0062**</td>
<td>-0.0077***</td>
<td>-0.0012</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0060)</td>
<td>(0.0042)</td>
<td>(0.0025)</td>
<td>(0.0033)</td>
<td>(0.0030)</td>
</tr>
<tr>
<td>Dep. var. mean</td>
<td></td>
<td>0.847</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td></td>
<td>6509</td>
<td>6302</td>
<td>6302</td>
<td>6509</td>
<td>6302</td>
</tr>
<tr>
<td>Clusters</td>
<td></td>
<td>283</td>
<td>274</td>
<td>274</td>
<td>283</td>
<td>274</td>
</tr>
<tr>
<td>Adjusted R-squared</td>
<td></td>
<td>0.917</td>
<td>0.951</td>
<td>0.974</td>
<td>0.961</td>
<td>0.959</td>
</tr>
<tr>
<td>Duflo Controls:</td>
<td></td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Log-linear Trend:</td>
<td></td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Females:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post × Intensity</td>
<td></td>
<td>(4)</td>
<td>(5)</td>
<td>(6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.020***</td>
<td>0.0027</td>
<td>0.0041</td>
<td>-0.021***</td>
<td>-0.0064*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0065)</td>
<td>(0.0052)</td>
<td>(0.0034)</td>
<td>(0.0049)</td>
<td>(0.0033)</td>
</tr>
<tr>
<td>Dep. var. mean</td>
<td></td>
<td>0.766</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td></td>
<td>6509</td>
<td>6302</td>
<td>6302</td>
<td>6509</td>
<td>6302</td>
</tr>
<tr>
<td>Clusters</td>
<td></td>
<td>283</td>
<td>274</td>
<td>274</td>
<td>283</td>
<td>274</td>
</tr>
<tr>
<td>Adjusted R-squared</td>
<td></td>
<td>0.934</td>
<td>0.956</td>
<td>0.979</td>
<td>0.949</td>
<td>0.960</td>
</tr>
<tr>
<td>Duflo Controls:</td>
<td></td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Log-linear Trend:</td>
<td></td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

Notes: This table displays results on the effect of school building on education attainment (completing primary school and completing secondary school) for males and females. Following the strategy of Duflo (2001), the sample consists of individuals born between either 1968 and 1972 or 1950 and 1961. Post refers to the treated cohort, born between 1968 and 1972, while the untreated cohort was born between 1950 and 1961. Educational attainment data are taken from the Indonesian 2010 Census. Intensity is the number of schools built in a region per 1,000 kids in the school-aged population. All columns include district fixed effect, school year fixed effect, school year interacted with number of children at 1971. Duflo Controls consist of school year interacted with enrollment rate at 1971 and school year interacted with water sanitization program. Standard errors are clustered at the birthplace district level. Significance levels: * 10%, ** 5%, *** 1%.
Source: Indonesian Census 2010
Table 3: Heterogeneity Effect of School Construction on Education

<table>
<thead>
<tr>
<th>Panel A: Density &lt; Medium:</th>
<th>Indicator for Completing at least:</th>
<th>Primary School</th>
<th>Secondary School</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(1) Male</td>
<td>(2) Female</td>
</tr>
<tr>
<td>Post × Intensity</td>
<td>0.010**</td>
<td>0.0066</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0051)</td>
<td>(0.0062)</td>
<td></td>
</tr>
<tr>
<td>Dep. var. mean</td>
<td>0.820</td>
<td>0.736</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>4209</td>
<td>4209</td>
<td></td>
</tr>
<tr>
<td>Clusters</td>
<td>183</td>
<td>183</td>
<td></td>
</tr>
<tr>
<td>Adjusted R-squared</td>
<td>0.949</td>
<td>0.952</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: Density &gt; Medium:</th>
<th></th>
<th>(3) Male</th>
<th>(4) Female</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0.0069**</td>
<td>-0.00066</td>
</tr>
<tr>
<td></td>
<td>(0.0032)</td>
<td>(0.0037)</td>
<td></td>
</tr>
<tr>
<td>Post × Intensity</td>
<td>0.0054</td>
<td>-0.0060</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0066)</td>
<td>(0.0073)</td>
<td></td>
</tr>
<tr>
<td>Dep. var. mean</td>
<td>0.873</td>
<td>0.795</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>2093</td>
<td>2093</td>
<td></td>
</tr>
<tr>
<td>Clusters</td>
<td>91</td>
<td>91</td>
<td></td>
</tr>
<tr>
<td>Adjusted R-squared</td>
<td>0.957</td>
<td>0.967</td>
<td></td>
</tr>
<tr>
<td>Duflo Controls:</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Log-linear trend:</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
</tbody>
</table>

Notes: This table is similar to Table 2 and displays the heterogeneity effect of school building on education attainment in sparsely and densely populated regions. All columns include district fixed effect, school year fixed effect, school year interacted with number of children at 1971. Duflo Controls consist of school year interacted with enrollment rate at 1971 and school year interacted with water sanitation program. Standard errors are clustered at the birthplace district level. Significance levels: * 10%, ** 5%, *** 1%. Source: Indonesian Census 2010
Table 4: Effect of School Construction on Number of Teachers in Secondary and Primary Education

<table>
<thead>
<tr>
<th></th>
<th>Secondary School</th>
<th>Primary School</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1) Total number</td>
<td>(2) Average number</td>
</tr>
<tr>
<td>INPRES Intensity × year=1978/79</td>
<td>-13.9 (8.48)</td>
<td>-0.19* (0.11)</td>
</tr>
<tr>
<td>year=1983/84</td>
<td>-43.5 (31.8)</td>
<td>-0.14 (0.18)</td>
</tr>
<tr>
<td>year=1988/89</td>
<td>-65.1 (49.1)</td>
<td>-0.20 (0.21)</td>
</tr>
<tr>
<td>year=1993/94</td>
<td>-59.3 (52.0)</td>
<td>-0.086 (0.22)</td>
</tr>
<tr>
<td>year=1995/96</td>
<td>-51.7 (60.0)</td>
<td>-0.074 (0.19)</td>
</tr>
</tbody>
</table>

Dep. var. mean in 1973/74 555.996 14.723 1529.996 6.762
Dep. var. mean in 1995/96 2583.821 22.989 4207.180 8.345
Observations 1656.000 1656.000 1664.000 1664.000
R-squared 0.928 0.929 0.942 0.829
Duflo Controls: Yes Yes Yes Yes

Notes: This table displays the effect of school construction on the number of teachers in secondary and primary education in the future years. Baseline year is 1973/74. All columns include district fixed effect, school year fixed effect, school year interacted with number of children at 1971. Duflo Controls consist of school year interacted with enrollment rate at 1971 and school year interacted with water sanitation program. Standard errors are clustered at the birthplace district level. Significance levels: * 10%, ** 5%, *** 1%.
Source: Indonesian Education Ministry
Table 5: Reduced-form Effect of School Construction on Female Marriage Outcomes

<table>
<thead>
<tr>
<th>Panel A:</th>
<th>Density &lt; Medium</th>
<th>Density &gt; Medium</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female First Marriage Age:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post × Intensity</td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Observations</td>
<td>2664</td>
<td>2664</td>
</tr>
<tr>
<td>Clusters</td>
<td>183</td>
<td>183</td>
</tr>
<tr>
<td>Adjusted R-squared</td>
<td>0.673</td>
<td>0.702</td>
</tr>
<tr>
<td>Clusters</td>
<td>183</td>
<td>183</td>
</tr>
<tr>
<td>Adjusted R-squared</td>
<td>0.673</td>
<td>0.702</td>
</tr>
<tr>
<td>Panel B:</td>
<td>Spousal age gap</td>
<td></td>
</tr>
<tr>
<td>Post × Intensity</td>
<td>0.030</td>
<td>0.057</td>
</tr>
<tr>
<td>Dep. var. mean</td>
<td>4.838</td>
<td>4.776</td>
</tr>
<tr>
<td>Observations</td>
<td>2745</td>
<td>2745</td>
</tr>
<tr>
<td>Clusters</td>
<td>183</td>
<td>183</td>
</tr>
<tr>
<td>Adjusted R-squared</td>
<td>0.858</td>
<td>0.879</td>
</tr>
</tbody>
</table>

Notes: This table displays the reduced-form effect of school construction on female first marriage age (top) and spousal age gap (bottom) in sparsely populated regions (left) and densely populated regions (right). All columns include district fixed effect, school year fixed effect, school year interacted with number of children at 1971. Duflo Controls consist of school year interacted with enrollment rate at 1971 and school year interacted with water sanitation program. Standard errors are clustered at the birthplace district level. Significance levels: * 10%, ** 5%, *** 1%.

Source: Indonesian SUPAS 2005, Indonesian Census 2010
Table 6: Results of Female Education Distribution on Female Marriage Outcomes

<table>
<thead>
<tr>
<th>Panel A: OLS and IV</th>
<th>Female first marriage age</th>
<th>Spousal age gap</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1) OLS</td>
<td>(2) IV</td>
</tr>
<tr>
<td>Percentage of females with secondary degree</td>
<td>1.92</td>
<td>10.9*</td>
</tr>
<tr>
<td></td>
<td>(1.37)</td>
<td>(6.53)</td>
</tr>
<tr>
<td>First Stage F statistics</td>
<td>12.929</td>
<td>12.929</td>
</tr>
<tr>
<td>Dep. var. mean</td>
<td>19.647</td>
<td>4.550</td>
</tr>
<tr>
<td>Observations</td>
<td>1365</td>
<td>1365</td>
</tr>
<tr>
<td>Clusters</td>
<td>91</td>
<td>91</td>
</tr>
<tr>
<td>Adjusted R-squared</td>
<td>0.763</td>
<td>0.754</td>
</tr>
<tr>
<td>Duflo Controls:</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Log-linear trend:</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>(3) OLS</th>
<th>(4) IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage of females with secondary degree</td>
<td>-2.71***</td>
<td>-3.49*</td>
</tr>
<tr>
<td></td>
<td>(0.38)</td>
<td>(2.03)</td>
</tr>
</tbody>
</table>

Notes: This table displays the OLS and IV estimates of the effect of female education distribution on marriage market outcomes. All columns include district fixed effect, school year fixed effect, school year interacted with number of children at 1971. Duflo Controls consist of school year interacted with enrollment rate at 1971 and school year interacted with water sanitation program. Standard errors are clustered at the birthplace district level. Significance levels: * 10%, ** 5%, *** 1%.

Source: Indonesian SUPAS 2005, Indonesian Census 2010
8 Appendix: Proofs

To simplify notation, in the following proofs, 1 is used to indicate type L, while 2 is used to indicate type H. Let’s prove the following lemma first:

**Lemma 1.** In Choo-Siow framework with only two types for each side, denote the marginals as \( n = (x, 1-x) \), \( m = (y, 1-y) \), the surplus matrix as:

\[
\Phi = \begin{bmatrix}
\Phi_{11} & \Phi_{12} \\
\Phi_{21} & \Phi_{22}
\end{bmatrix}
\]

denote the mass of singles of males (females) in equilibrium as: \( \mu_{10}, \mu_{20}, \mu_{01}, \mu_{02} \) then:

(a) \[
\frac{\partial \mu_{10}}{\partial x} > 0, \quad \frac{\partial \mu_{20}}{\partial x} < 0
\]

(b) If the marital surplus function is super-modular, i.e., \( \Phi_{11} + \Phi_{22} > \Phi_{12} + \Phi_{21} \), then

(b1) \[
\frac{\partial \mu_{01}}{\partial x} > 0 \Rightarrow \frac{\partial \mu_{02}}{\partial x} > 0
\]

\[
\frac{\partial \mu_{02}}{\partial x} < 0 \Rightarrow \frac{\partial \mu_{01}}{\partial x} < 0
\]

(b2) Further assume \( x \) and \( y \) is bounded away from 0 and 1, i.e., there exists some \( \delta_x, \hat{\delta}_x, \delta_y, \hat{\delta}_y \), such that \( \delta_x < x < \hat{\delta}_x, \delta_y < y < \hat{\delta}_y \), then:

\[
\frac{\partial \mu_{01}}{\partial x} < 0
\]

**Proof.** Denote \( a = \exp(\frac{\Phi_{11}}{2}) \), \( b = \exp(\frac{\Phi_{12}}{2}) \), \( c = \exp(\frac{\Phi_{21}}{2}) \), \( d = \exp(\frac{\Phi_{22}}{2}) \);

denote \( s_{10} = \sqrt{\mu_{10}}, s_{20} = \sqrt{\mu_{20}}, s_{01} = \sqrt{\mu_{01}}, s_{02} = \sqrt{\mu_{02}} \);

denote \( D_{10} = \frac{\partial s_{10}}{\partial x}, D_{20} = \frac{\partial s_{20}}{\partial x}, D_{01} = \frac{\partial s_{01}}{\partial x}, D_{02} = \frac{\partial s_{02}}{\partial x} \)

Then we can rewrite the feasibility constraints with matching function information as:

\[
s_{10}^2 + s_{10}s_{01}a + s_{10}s_{02}b = x
\]

\[
s_{20}^2 + s_{20}s_{01}c + s_{20}s_{02}d = 1 - x
\]
\[
s_{01}^2 + s_{10}s_{01}a + s_{20}s_{01}c = y
\]
\[
s_{02}^2 + s_{10}s_{02}b + s_{20}s_{02}d = 1 - y
\]

In the four equations above, take derivative with respect to \( x \), we can get:

\[
(2s_{10} + as_{01} + bs_{02})D_{10} + s_{10}(aD_{01} + bD_{02}) = 1 \tag{1}
\]
\[
(2s_{20} + cs_{01} + ds_{02})D_{20} + s_{20}(cD_{01} + dD_{02}) = -1 \tag{2}
\]
\[
(2s_{01} + as_{10} + cs_{20})D_{01} + s_{01}(aD_{10} + cD_{20}) = 0 \tag{3}
\]
\[
(2s_{02} + bs_{10} + ds_{20})D_{02} + s_{02}(bD_{10} + dD_{20}) = 0 \tag{4}
\]

Hence we can express \( D_{10}, D_{02} \) using \( D_{10}, D_{02} \) from the (3) and (4):

\[
D_{01} = -\frac{s_{01}(aD_{10} + cD_{20})}{2s_{01} + as_{10} + cs_{20}} \tag{5}
\]
\[
D_{02} = -\frac{s_{02}(bD_{10} + dD_{20})}{2s_{02} + bs_{10} + ds_{20}} \tag{6}
\]

Plug in (1) and (2), we can get:

\[
(2s_{10} + as_{01} + bs_{02})\left(\frac{2s_{01}(2s_{01} + cs_{20}) + bs_{02}(2s_{02} + ds_{20})}{2s_{01} + as_{10} + cs_{20} + 2s_{02} + bs_{10} + ds_{20}}\right)D_{10} - \left(\frac{acs_{10}s_{01}}{2s_{01} + as_{10} + cs_{20} + 2s_{02} + bs_{10} + ds_{20}} + \frac{bds_{10}s_{02}}{2s_{02} + bs_{10} + ds_{20}}\right)D_{20} = 1 \tag{7}
\]
\[
(2s_{20} + cs_{01} + ds_{02})\left(\frac{cs_{01}(2s_{01} + as_{10}) + ds_{02}(2s_{02} + bs_{20})}{2s_{01} + as_{10} + cs_{20} + 2s_{02} + bs_{10} + ds_{20}}\right)D_{20} - \left(\frac{acs_{20}s_{01}}{2s_{01} + as_{10} + cs_{20} + 2s_{02} + bs_{10} + ds_{20}} + \frac{bds_{20}s_{02}}{2s_{02} + bs_{10} + ds_{20}}\right)D_{10} = -1 \tag{8}
\]

Add (7) and (8), we get:

\[
(2s_{10} + \frac{2as_{01}^2}{2s_{01} + as_{10} + cs_{20}} + \frac{2bs_{02}^2}{2s_{02} + bs_{10} + ds_{20}})D_{10} + (2s_{20} + \frac{2cs_{01}^2}{2s_{01} + as_{10} + cs_{20}} + \frac{2ds_{02}^2}{2s_{02} + bs_{10} + ds_{20}})D_{20} = 0 \tag{9}
\]

Hence \( D_{10}, D_{20} \) has negative signs. With equation (7), we know:

\[
D_{10} > 0, D_{20} < 0
\]
Proof for (a) is complete.

For part (b1) of the lemma, with super-modularity, we know:

\[ a \ast d > b \ast c \]

With \( D_{10} > 0 \):

\[ \frac{a}{c}D_{10} > \frac{b}{d}D_{10} \]

\[ \Rightarrow: \frac{a}{c}D_{10} + D_{20} > \frac{b}{d}D_{10} + D_{20} \]

Hence:

\[ aD_{10} + cD_{20} < 0 \quad \Rightarrow \quad bD_{10} + dD_{20} < 0 \]

\[ bD_{10} + dD_{20} > 0 \quad \Rightarrow \quad aD_{10} + cD_{20} > 0 \]

Recall equation (5) and (6), we have:

\[ \frac{\partial \mu_{01}}{\partial x} > 0 \Rightarrow \frac{\partial \mu_{02}}{\partial x} > 0 \]

\[ \frac{\partial \mu_{02}}{\partial x} < 0 \Rightarrow \frac{\partial \mu_{01}}{\partial x} < 0 \]

Proof for (b1) is complete.

Now let’s prove part (b2):

\[ \frac{\partial s_{01}}{\partial x} = \frac{D_{01}s_{02} - D_{02}s_{01}}{s_{02}^2} \]

Using equation (5) and (6),

\[ D_{01}s_{02} - D_{02}s_{01} = -\frac{s_{01}s_{02}(aD_{10} + cD_{20})}{2s_{01} + as_{10} + cs_{20}} + \frac{s_{01}s_{02}(bD_{10} + dD_{20})}{2s_{02} + bs_{10} + ds_{20}} \]

\[ = \frac{s_{01}s_{02}((b(2s_{01} + cs_{20}) - a(2s_{02} + ds_{20}))D_{10} + (d(2s_{01} + as_{10}) - c(2s_{02} + bs_{10}))D_{20})}{(2s_{01} + as_{10} + cs_{20})(2s_{02} + bs_{10} + ds_{20})} \]
It has the same sign as:

\[
[2bs_{01} - 2as_{02} + (bc - ad)s_{20}]D_{10} + [2ds_{01} - 2cs_{02} + (ad - bc)s_{10}]D_{20}
\]

\[
= 2s_{01}(bD_{10} + dD_{20}) - 2s_{02}(aD_{10} + cD_{20}) - (ad - bc)(D_{10}s_{20} - D_{20}s_{10})
\]

We know that \((ad - bc)(D_{10}s_{20} - D_{20}s_{10}) > 0\), since \(ad - bc > 0\), \(D_{10} > 0\), \(D_{20} < 0\).

According to (b1), there are only three cases:

(Case 1): \(aD_{10} + cD_{20} > 0\), \(bD_{10} + dD_{20} < 0\); it’s straightforward to show:

\[
\frac{\partial s_{01}}{\partial x} < 0
\]

(Case 2): \(aD_{10} + cD_{20} > 0\), \(bD_{10} + dD_{20} > 0\)

in this case, from equation (9), we know \(s_{10}D_{10} + s_{20}D_{20} < 0\), hence:

\[
\frac{a}{c} > \frac{b}{d} > \frac{s_{10}}{s_{20}}
\]

Since we know \(\frac{s_{10}}{s_{20}}\) increases with \(x\), to satisfy previous inequality, we know that \(x\) is also relatively small in this case.

There exists some \(\delta_x, \delta_y\) such that for \(x > \delta_x, y < \delta_y\),

\[
\frac{\partial s_{01}}{\partial x} < 0
\]

(Intuition: we need \(x\) to be away from 0 and \(y\) to be away from 1 to avoid large value of \(s_{01}\) and small value of \(s_{02}\).)

(Case 3): \(aD_{10} + cD_{20} < 0\), \(bD_{10} + dD_{20} < 0\)

in this case, from equation (9), we know \(s_{10}D_{10} + s_{20}D_{20} > 0\), hence:

\[
\frac{s_{10}}{s_{20}} > \frac{a}{c} > \frac{b}{d}
\]
\( x \) is relatively large in this case. There exists some \( \bar{\delta}_x, \delta_y \) such that for \( x < \bar{\delta}_x, y > \delta_y, \)
\[
\frac{\partial s_{01}}{\partial x} < 0
\]

(Intuition: we need \( x \) to be away from 1 and \( y \) to be away from 0 to avoid small value of \( s_{01} \) and large value of \( s_{01} \).)

Proof for part (b2) is complete. \( \square \)

Lemma 2. An extension of Lemma 1:
suppose there are two types on one side, and there are \( K > 2 \) types on the other side, denote the marginals as \( n = (x, 1 - x), m = (y_1, y_2, ..., y_K) \), where \( \sum_k y_k = 1 \). The surplus matrix as:
\[
\Phi = \begin{bmatrix}
\Phi_{11} & \Phi_{12} & \cdots & \Phi_{1K} \\
\Phi_{21} & \Phi_{22} & \cdots & \Phi_{2K}
\end{bmatrix}
\]
denote the mass of singles in equilibrium as: \( \mu_{10}, \mu_{20}, \mu_{0k} \) then:

(a) \[ \frac{\partial \mu_{10}}{\partial x} > 0, \frac{\partial \mu_{20}}{\partial x} < 0 \]

(b) \( \forall \) two types \( k_1, k_2 \), \( \Phi_{1k_1} + \Phi_{2k_2} > \Phi_{1k_2} + \Phi_{2k_1} \) further assume \( x, y_{k_1}, y_{k_2} \) are bounded away from 0 and 1, then: \( \frac{\mu_{20}}{\mu_{10}} \) decreases if we shift some mass from type \( k_1 \) to type \( k_2 \), i.e.:
\[
\mu_{20} \left( n, m = (\ldots, y_{k_1} - \Delta, y_{k_2} + \Delta, \ldots) \right) < \mu_{20} \left( n, m = (\ldots, y_{k_1}, y_{k_2}, \ldots) \right), \forall \Delta > 0
\]

The rest list the proofs for the propositions.

Proof for: Proposition 1

Proof. To prove the existence of stationary equilibrium, we need to show that there is a solution to the following equilibrium conditions given \( n_x, m_y, \Phi \):

\[
\mu_{x0} + \sqrt{\mu_{x0} \mu_{01}^0 \exp(\frac{\phi_{x1}^0}{2})} + \sqrt{\mu_{x0} \mu_{02}^+ \exp(\frac{\phi_{x2}^0}{2})} + \sqrt{\mu_{x0} \mu_{02}^+ \exp(\frac{\phi_{x2}^+}{2})} = n_x, \forall x \in \{1, 2\}
\]

\[ \tag{10} \]

\[
\mu_{0y}^0 + \sqrt{\mu_{10} \mu_{0y}^0 \exp(\frac{\phi_{0y}^0}{2})} + \sqrt{\mu_{20} \mu_{0y}^0 \exp(\frac{\phi_{2y}^0}{2})} = q_y^0, \forall y \in \{1, 2\}
\]

\[ \tag{11} \]
\[ \mu_{0y}^+ + \sqrt{\mu_{10}\mu_{0y}^+} \exp\left(\frac{\phi_{1y}^0}{2}\right) + \sqrt{\mu_{20}\mu_{0y}^+} \exp\left(\frac{\phi_{2y}^0}{2}\right) = q_y^+, \forall y \in \{1, 2\} \] (12)

\[ q_y^0 + q_y^+ = m_y, \forall y \in \{1, 2\} \] (13)

\[ \exp(-v_y^0) = \frac{\mu_{0y}}{q_y^0}, \forall y \in \{1, 2\} \] (14)

\[ \exp(-v_y^+) = \frac{\mu_{0y}^+}{q_y^+}, \forall y \in \{1, 2\} \] (15)

\[ v_y^0 = v_y^+, \forall y \in \{1, 2\} \] (16)

Equation (10)-(12) characterize the equilibrium conditions of marriage market stability for given \( q_y^0 \) strategy under the assumption of Gumbel distribution. Equation (14)-(15) characterize the expected marital utilities of females. Equation (13) comes from the definition of stationarity. Equation (16) guarantees that females are indifferent between choosing to marry at period 1 or period 2.

Re-arrange equation (11) and (12), we can get:

\[ \frac{\mu_{0y}}{q_y^0} + \sqrt{\mu_{10}} \frac{1}{q_y^0} \exp\left(\frac{\phi_{1y}^0}{2}\right) + \sqrt{\mu_{20}} \frac{1}{q_y^0} \exp\left(\frac{\phi_{2y}^0}{2}\right) = 1 \]

\[ \frac{\mu_{0y}^+}{q_y^+} + \sqrt{\mu_{10}} \frac{1}{q_y^+} \exp\left(\frac{\phi_{1y}^+}{2}\right) + \sqrt{\mu_{20}} \frac{1}{q_y^+} \exp\left(\frac{\phi_{2y}^+}{2}\right) = 1 \]

Combing with equation (14)-(16), we can get:

\[ \sqrt{\frac{q_y^0}{q_y^+}} = \frac{\sqrt{\mu_{10}} \exp\left(\frac{\phi_{1y}^0}{2}\right) + \sqrt{\mu_{20}} \exp\left(\frac{\phi_{2y}^0}{2}\right)}{\sqrt{\mu_{10}} \exp\left(\frac{\phi_{1y}^+}{2}\right) + \sqrt{\mu_{20}} \exp\left(\frac{\phi_{2y}^+}{2}\right)} \]

\[ = \exp\left(\frac{\phi_{1y}^0 - \phi_{1y}^+}{2}\right) \frac{\sqrt{\mu_{10}} + \sqrt{\mu_{20}} \exp\left(\frac{\phi_{2y}^0 - \phi_{12}^0}{2}\right)}{\sqrt{\mu_{10}} + \sqrt{\mu_{20}} \exp\left(\frac{\phi_{2y}^+ - \phi_{12}^+}{2}\right)} \]

\[ = \exp\left(\frac{\phi_{1y}^0 - \phi_{1y}^+}{2}\right) \frac{1 + \sqrt{\frac{\mu_{20}}{\mu_{10}}} \exp\left(\frac{\phi_{2y}^0 - \phi_{12}^0}{2}\right)}{1 + \sqrt{\frac{\mu_{20}}{\mu_{10}}} \exp\left(\frac{\phi_{2y}^+ - \phi_{12}^+}{2}\right)} \] (17)

There are three cases:

1. \( \phi_{2y}^0 - \phi_{1y}^0 = \phi_{2y}^+ - \phi_{1y}^+ \)
2. \( \phi_{2y}^0 - \phi_{1y}^0 > \phi_{2y}^+ - \phi_{1y}^+ \)

3. \( \phi_{2y}^0 - \phi_{1y}^0 < \phi_{2y}^+ - \phi_{1y}^+ \)

**Case one:** In the first case, we have:

\[
\sqrt{\frac{q_y^0}{q_y^+}} = \exp(\frac{\phi_{1y}^0 - \phi_{1y}^+}{2})
\] (18)

Hence equilibrium strategy \( q_y \) is pinned down by equation (18) and equation (13). Moreover, we know that given \( q_y \), equation (10)-(12) has unique equilibrium solution according to Decker et al. (2013). Hence stationary equilibrium exists in this case and is unique.

**Case two:** In the second case, \( \frac{q_y^0}{q_y^+} \) is an increasing function of \( \frac{\mu_{20}}{\mu_{10}} \) in equation (18). Moreover, according to Lemma 2, we know that when \( \phi_{2y}^0 - \phi_{1y}^0 > \phi_{2y}^+ - \phi_{1y}^+ \) indicating there is a complementarity between male High type and female marrying at period 0, an increase in \( \frac{q_y^0}{q_y^+} \) would lead to a decrease in \( \frac{\mu_{20}}{\mu_{10}} \) from equation (10)-(13).

Moreover, from equation (17), we know that:

\[
\sqrt{\frac{q_y^0}{q_y^+}} \to \exp(\frac{\phi_{1y}^0 - \phi_{1y}^+}{2}), \text{ as } \frac{\mu_{20}}{\mu_{10}} \to 0
\]

\[
\sqrt{\frac{q_y^0}{q_y^+}} \to \exp(\frac{\phi_{2y}^0 - \phi_{2y}^+}{2}), \text{ as } \frac{\mu_{20}}{\mu_{10}} \to +\infty
\]

While from equation (10)-(13), we know \( \frac{\mu_{20}}{\mu_{10}} \) is bounded by finite positive number when \( \exp(\frac{\phi_{1y}^0 - \phi_{1y}^+}{2}) \leq \sqrt{\frac{q_y^0}{q_y^+}} \leq \exp(\frac{\phi_{2y}^0 - \phi_{2y}^+}{2}) \).

Hence equilibrium exists and is unique.

**Case three:** In the third case, \( \frac{q_y^0}{q_y^+} \) is a decreasing function of \( \frac{\mu_{20}}{\mu_{10}} \) in equation (17). Moreover, according to Lemma 2, we know that when \( \phi_{2y}^0 - \phi_{1y}^0 < \phi_{2y}^+ - \phi_{1y}^+ \) indicating there is a complementarity between male L type and female marrying at period 0, an increase in \( \frac{q_y^0}{q_y^+} \) would lead to an increase in \( \frac{\mu_{20}}{\mu_{10}} \) from equation (10)-(13). Applying the same logic as in case two, equilibrium exists and is unique.
Moreover, we know that equilibrium strategy satisfies:

\[
\min(\exp(\frac{\phi^0_{1y} - \phi^+_{1y}}{2}), \exp(\frac{\phi^0_{2y} - \phi^+_{2y}}{2})) \leq \frac{q^0_y}{q^+_y} \leq \max(\exp(\frac{\phi^0_{1y} - \phi^+_{1y}}{2}), \exp(\frac{\phi^0_{2y} - \phi^+_{2y}}{2}))
\]

\[\blacksquare\]

Proof for: Proposition 2

Proof. Since \( \delta^+_{y'} = \delta^0_{y'}, \phi^+_{2y'} - \phi^+_{1y'} = \phi^0_{2y'} - \phi^0_{1y'} \). This is our first case in the previous proof of Proposition 1. Hence from equation (17), we know:

\[
\frac{q^+_y}{q^0_y} = \exp(\phi^+_{1y'} - \phi^0_{1y'})
\]

with \( q^0_y + q^+_y = m_y \), I have:

\[
q^+_y = \frac{\exp(\phi^+_{1y'})}{\exp(\phi^+_{1y'}) + \exp(\phi^0_{1y'})} m_y, \quad q^0_y = \frac{\exp(\phi^0_{1y'})}{\exp(\phi^+_{1y'}) + \exp(\phi^0_{1y'})} m_y
\]

\[\blacksquare\]

Proof for: Proposition 3 and Proposition 4

Proof. From the proof of proposition 1, we know that equilibrium strategy is pinned down by both equation (17) and equations (10)-(13). Hence how equilibrium strategies change depend on whether \( \delta^+_{y'} \) or \( \delta^+_{y'} < \delta^0_{y'} \) and how \( \frac{\mu_{20}}{\mu_{10}} \) changes in equilibrium.

Let’s first prove Proposition 3, according to Lemma 2 result (a), an increase in \( n_H \) would increase \( \mu_{20} \) and decrease \( \mu_{10} \), which increases \( \frac{\mu_{20}}{\mu_{10}} \) given any strategy \( q_y \), hence an increase in \( n_H \) would

- increase \( \frac{q^0_y}{q^+_y} \) if \( \delta^0_y > \delta^+_{y} \)
- decrease \( \frac{q^0_y}{q^+_y} \) if \( \delta^0_y < \delta^+_{y} \)

Then let’s prove Proposition 4, under the assumptions, according to Lemma 2 (b), an increase in \( m_H \) would decrease \( \frac{\mu_{20}}{\mu_{10}} \) given any strategy \( q_y \), hence an increase in \( m_H \) would

- decrease \( \frac{q^0_y}{q^+_y} \) if \( \delta^0_y > \delta^+_{y} \)
• increase $\frac{q_0}{q_y}$ if $\delta_y^0 < \delta_y^+$