Schooling Expansion and the Marriage Market: Evidence from Indonesia *

Danyan Zha‡

April 8, 2022

Abstract

This paper analyzes how education distribution affects the marriage market by exploiting a massive primary school construction program in Indonesia. Using variations across districts and across birth cohorts, I first show the program had an unintended consequence: it decreased secondary school attainment rate due to a crowding-out of teacher resources and a decrease in teacher quality in densely populated areas. Combining this variation and the large average spousal age gap (five years), I show that spousal age gap increases and never-married rate does not change for women when their secondary education attainment rate decreases holding their potential husbands' education distribution unchanged. The change in the spousal age gap suggests male education and female youth are complementary using a two-to-one dimensional OLG matching model with transferrable utilities.

JEL classifications: 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. I also thank So Yoon Ahn, Michael Best, Mingming Ma, Suresh Naidu, Rodrigo Soares, Jack Willis and participants at the Applied Microeconomics Theory Colloquium, the Development Colloquium at Columbia University, the 2018 North American Summer Meeting of the Econometric Society at UC Davis, the 2018 European Summer Meeting of the Econometric Society at University of Cologne, the SEA 88th Annual Meetings at Washington, DC, the 2019 AASLE conference at Singapore, and the 2019 Shanghai Empirical Micro conference for helpful discussions and comments. All errors are my own.

†School of Economics, Shanghai University of Finance and Economics.

‡Contact: zhadanyan@mail.shufe.edu.cn

1 Introduction

Human capital is key to individuals' lifetime outcomes and countries' economic growth [1], which provides a rationale for world-wide schooling expansion, especially in low- and middle-income countries in the past three decades (World Bank, 2018). Many papers have documented the positive effect of these education policies on individual outcomes including education, wage, income, wealth and health (Malamud et al., 2018; Jürges et al., 2011). Careful evaluations are necessary to guide government policies and international organizations to find the most cost-effective policies[2]. However, the potential existence of externalities makes program evaluation very hard. A program that targets one particular population may affect another population through information transmission, resource allocation, general equilibrium effect via labor market or marriage market, and other channels. Untreated individuals may benefit from a program if there is social positive externality, or may be affected negatively if the resources available to them decrease due to the program.

This paper re-evaluates the INPRES SD (*Presidential Instruction, Sekolah Dasar*) program in Indonesia, the largest and most successful primary school construction program so far. I find that surprisingly, building primary school has an unintended consequence on secondary education. It has a negative impact on the secondary school attainment rate[3] for both men and women in more densely populated districts due to a crowding-out of teacher resources and a decrease in primary school teacher quality. Furthermore, combining this variation and the large spousal age gap (5 years) in Indonesia, I find that female spousal age gap[4] increases and female never-married rate does not change due to the program in densely

[1] A large literature in economic growth (see Mankiw et al., 1992; Young, 1994,9; Barro and i Martin, 1995) documents the importance of endogenous human capital to economic growth.

[2] See (Glewwe and Kremer, 2006; Glewwe et al., 2013; McEwan, 2015; Glewwe and Muralidharan, 2016) for a good review of research on the effectiveness of education policies.

[3] The secondary school attainment rate is defined as the percentage of people completing at least secondary school for a given birth cohort in one district. Similarly, the primary school attainment rate is defined as the percentage of people who complete primary school or above.

[4] Spousal age gap is defined as husband's age minus wife's age, and spousal education gap is defined as husband's years of schooling minus wife's year of schooling, while never-married is defined as no marriage before the survey data.

populated districts when female secondary school attainment is negatively impacted by the program while holding potential husbands' education distribution unchanged, but in sparsely populated districts, the program has no effect on female spousal age-gap or never-married rate.

When oil price increased in 1972, the Indonesian government has experienced a tremendous revenue increase, which facilitated one of the largest education expansion programs in the world: INPRES SD. Approximately one primary school was built per 500 primary school-aged children between 1973/74 and 1978/79. This creates potential variation in the education distribution across districts and birth cohorts. Using both variations, my identification strategy is difference-in-differences similar to other papers studying this program (Duflo, 2001; Akresh, Halim, and Kleemans, Akresh et al.; Mazumder et al., 2019; Ashraf et al., 2020; Mazumder et al., 2021). One difference comes from the difference in construction intensity across districts, defined as the average number of schools constructed per 1000 children between 1972/73 and 1978/79; the other difference comes from individual birth cohorts. INPRES SD is a program that targets equality, and hence, more schools were built in regions where there was initially a larger number of un-enrolled school-aged children. Children attend primary school between ages 7 and 12. Therefore children aged 13 or older in 1974 should have been out of school when more primary schools were constructed and hence should have not been affected by the primary school construction program.

This paper builds upon earlier studies studying the effect of the same schooling expansion program (Duflo, 2001; Akresh, Halim, and Kleemans, Akresh et al.; Mazumder et al., 2019; Ashraf et al., 2020; Mazumder et al., 2021) and provides surprising results not mentioned in the previous literature. Consistent with previous findings, there is a positive effect on primary school attainment rate for average Indonesian men but not for average Indonesian women. Moreover, I find a negative effect on secondary school attainment rate for women in the full-sample analysis. As suggested in Duflo (2001), the program may have different effects in sparsely populated and densely populated districts. Exploring the heterogeneity of the effects depending on population densities of the districts considered, I find that in sparsely populated districts, the school construction program has a positive effect on primary school attainment rate for both men and women, a positive effect on secondary school attainment rate for men but not for women; in densely populated districts, the school construction program does not affect primary school attainment rate but has a negative effect on secondary school attainment rate for both men and women.

Two potential mechanisms leading to the negative result on secondary school attainment are further investigated: (1) a decrease in secondary school quality due to teacher resources being crowded out; (2) a decrease in primary school teacher quality due to the massive construction scale. Empirical data supports both mechanisms. Consistent with the first mechanism, this paper finds that the total number of teachers and the average number of teachers in secondary school increase less in the districts where more primary schools are constructed since the launch of the school construction program. Moreover, the negative effect on teacher availability in secondary education in future years only exists in densely populated districts, not in sparsely populated districts. Consistent with the second mechanism, using the education level of primary school teachers in the censuses as a proxy for primary school quality, it is shown that teacher education increases less in districts where more primary schools are constructed. Moreover, the effect is stronger in densely populated districts. In summary, the negative result on secondary school attainment rate could be attributed to both a crowding out of teacher resources in secondary education and a decrease in primary school teacher quality because of the massive scale of the primary school construction.

In the last part, this paper analyzes the effect of this program on individuals' marital outcomes. Matching theories suggest that individuals' marital outcomes depend on various marriage market conditions including the characteristic distributions of both sides: men and women. Hence, a woman's marital outcome can be affected by both her own education *and* that of others including other women and all men. The large spousal age gap (5 years) in Indonesia and the school construction program help to create a setting where only women's education are affected while keeping their potential husbands' education unchanged: for the first few cohorts of women who were impacted by the school construction program, the education level of their potential husbands was minimally impacted. Therefore, by comparing these female cohorts with the older cohorts who were not impacted by the program, we are able to observe how female marital outcomes respond to the change in the female education distribution while holding the male education distribution unchanged. This paper finds that for the first several cohorts of women in densely populated districts, the program has a negative effect on their education, a positive effect on their spousal age gap and no effect on the average never-married rate. For the first several cohorts of women in sparsely populated districts, the program does not affect their average education, spousal age gap and never-married rate.

To better understand the result on spousal age gap, this paper employs a theoretical framework to understand how female marriage age reacts to the change in the education distributions of men and women across cohorts. To incorporate marriage age as a choice for women, I build a two-period OLG model in which women can choose to seek partners either in the first period or the second, but men all marry in the second period. In any given year, the marriage market unfolds as in Choo and 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 the marriage market should be equalized between choosing to marry in the first period or the second. The model predicts that female marriage age choice does not depend on the education distribution of men or women if there is no interaction between male education and female youth in generating marital surplus. However, suppose that women marrying at a young age is "good" for marital surplus; then, in the case in which there is complementarity between men's education and women's youth and there is also complementarity between men's education and women's education, an increase in the proportion of educated women would increase female marriage age (i.e. a decrease in the percentage of females marrying in first period) and decrease the spousal age gap. Combined with the empirical finding that spousal age gap increases in densely populated districts where female education decreases after ruling out the effect of own education change, this suggests that male education and female youth are complementary in our context.

This paper is related to several strands of literature. Firstly, it is closely related to other papers that have studied the effect of INPRES SD program since the seminal paper Duflo (2001), which focuses on men and finds a positive effect on male years of schooling and their wages in 1995 using the Indonesian 1995 inter-census (SUPAS). Using the same dataset, Ashraf et al. (2020) looks at women and finds that the program increases years of schooling for women, but only for the ethnicities that practice bride price. With the same dataset, Breierova and Duflo (2004) shows that female education is improved, further taking advantage of the spousal age gap, the paper shows that both maternal and paternal education are important in reducing child mortality, while female education is a stronger determinant of marriage age and fertility. Using the fifth wave of the Indonesian Family Life Survey (IFLS, 2014), Bharati et al. (2018) shows that the school construction increases schooling for individuals who experienced negative shock (low rainfall) in the first year of life but not those who didn't experience the adverse rainfall shock, partly due to deteriorating school infrastructure and increased competition. Using the National Socioeconomic Survey (SUSENAS) in 2016 and multiple waves of IFLS, Akresh, Halim, and Kleemans (Akresh et al.) and Mazumder et al. (2019) examine the long-term and intergenerational effect of this program on individuals' work choice, household behavior, and children's education and health outcomes. Using multiple waves of Indonesian Village census (Potensi Desa), Martinez-Bravo (2017) shows that local public good provision increases in the villages where the education of the village heads increase due to this program.

Secondly, this paper contributes to a growing literature studying the impact of education reform on marriage market. Using the Indonesian 2010 census data, Dominguez (2014) uses structural estimation method to show that an increase in the primary school graduates increase the single rate and decrease the marital utility of the primary school graduates in Indonesia. Hener and Wilson (2018) studies a compulsory reform in UK and finds that women decrease the marital age gap to avoid marrying less-qualified men. André and Dupraz (2018) studies school construction in Cameroon and finds that education increases the likelihood of being in a polygamous union for both men and women.

Thirdly, this paper provides evidence for the existence of another type of externality when implementing a large-scale social intervention. Miguel and Kremer (2004) shows a large positive externality of deworming for untreated children in the treatment and neighboring schools. Bobonis and Finan (2009) finds a positive externality of the PROGRESA program for program-illegible children's secondary school participation in the program communities. Castro and Esposito (2018) finds a negative externality of the bonus paid to incentivize teachers on nearby rural schools.

Finally, this paper complements the literature studying the effect of teacher availability and quality on student learning outcomes by showing an unintended consequence of another education expansion program. Many papers have shown that higher teacher quality matters more to students' achievement than other education input including class size and school infrastructure (Rivkin et al., 2005; Hanushek, 2011). Andrabi et al. (2013) shows that more girls' secondary schools increase the local supply of skilled women that enlarges the pool of potential teachers which facilitates educational access of local children.

The paper proceeds as follows:

Section II discusses briefly the conceptual framework and the model's predictions. Section III introduces the background of the school construction program, the education system and the marriage market in Indonesia. Section IV demonstrates the data and empirical strategy. Section V shows the results on educational outcomes and explores the two mechanisms. Section VI explores the impact on marriage market. Section VII offers some discussions and Section VIII concludes.

2 Conceptual framework

In this section, I develop a two-period OLG matching model with Transferable Utility (TU) to study how a change in the education distribution across birth cohorts may affect marriage market outcomes, in particular, female marriage age. There are several important features:

- Individuals get utility from participating in the marriage market.
- Individuals' education affect the marital surplus, for both men and women.
- Individuals' age play an asymmetric role for men and women. Women's age matters but not men's in the surplus function. Much research has documented that female youth is more important than male youth in the marriage market, this could be due either to the fundamental difference of female age and male age in the household production function related to fertility, or due to a stronger male preference for youth related beauty. (Low, 2017; Siow, 1998; Edlund, 2006; Dessy and Djebbari, 2010; Zhang, 2018; Arunachalam and Naidu, 2006)
- Women are allowed to choose to participate in the marriage market either early or late. However, a woman who participated in period 1 cannot enter into the marriage market in period 2, whether she remains married or single. This can be rationalized as the existence of a stigma associated with women who have tried to seek partners in an early period.
- Each marriage market is modeled as a matching model with TU with idiosyncratic random preference draws. The existence of random preference draws allows the existence of couples of all types with respect to male education, female education and female age, which suits the reality more compared to the static model. In each marriage market, women differ in both education and age, while men only differ in education.

Two-period OLG

There is an infinite number of periods, r=1,2.... At the beginning of each period, a unit mass of men and a unit mass of women enter the economy. Assume people can only make marriage decisions in the first two periods, therefore the problem is simplified to a two-period OLG problem. Furthermore, to focus on female marriage age decision, I assume that women choose whether they want to seek partners in period 1 (when they are young) or delay this process to period 2 (when they are old). Men always seek partners in period 2. Individuals differ in their education type, L or H. In the model, let's focus on the utilities individuals obtain from the marriage market.

Marriage market at one period

In any period, the marriage market unfolds as in Choo and Siow (2006). The matching equilibrium (who is married with whom) is determined by the population of each type of men and women, and the marital surplus they create together if they form a couple. Since women can choose to participate in one of the two periods, hence in any period, there are at most four types of women: Low education and Young (L_1), Low education and Old (L_2), High education and Young (H_1), and High education and Old (H_2). Men only participate in period 2, hence there are two types of men in any period: Low education (L) and High education (H). Martial surplus depends on both of the couple's deterministic types and their idiosyncratic random preference draws. Detailed model set-up is illustrated in the appendix.

Stationary equilibrium with OLG

Before participating in any marriage market, the strategic choice for each woman in the model is to choose when to enter into the marriage market, given the predetermined education distribution of women and men, denoted by (E_f, E_m). For a woman with education e, if she chooses to enter in period 2 instead of period 1, this increases the expected marital return of all women in period 1 marriage market and decreases the expected return of all women in period 2 marriage market. [5] In a stationary equilibrium, the percentage of women who choose to wait until period 2 equates women's expected returns in the two marriage markets. Denote the percentage of women with education e who choose to seek partners in period 1 (or period 2) as q_e^1 (or q_e^2), assume $e \in \{L, H\}$. Of course, $q_e^1 + q_e^2 = 1, \forall e$.

We say the marriage market with distribution of female types and male types as (G_x, G_y)

[5] As proved in Galichon and Salanié (2017), an addition of one woman hurts all women and benefits all men; an addition of one man hurts all men and benefits all women.

is the induced marriage market of a strategy vector $\mathbf{q}$ if the distribution of female types (four) and male types (two) in the marriage market is (G_x, G_y) when women adopt strategy $\mathbf{q}$. Note that for male distribution, $G_y = G_m, \forall \mathbf{q}$.

Definition 1. *Strategy vector* $\mathbf{q} = \{q_H^1, q_H^2, q_L^1, q_L^2\}$ *forms a stationary equilibrium if* $u_{H_1} = u_{H_2}$ *and* $u_{L_1} = u_{L_2}$ *in the induced marriage market, where* u_{e_1} *(*u_{e_2}*) is the expected marriage payoff of women with education* e *who choose to enter the marriage market in period 1 (period 2).*

Denote $\Phi_{xy} = \alpha_{xy} + \gamma_{xy}$. We have woman's type $x \in \{L_1, L_2, H_1, H_2\}$, man's type $y \in \{L, H\}$.

Proposition 1. *There exists a unique stationary equilibrium, and the equilibrium strategy* $\mathbf{q}$ *satisfy:*

$$\min(\Phi_{L_1L} - \Phi_{L_2L}, \Phi_{L_1H} - \Phi_{L_2H}) \leq \ln(\frac{q_L^1}{q_L^2}) \leq \max(\Phi_{L_1L} - \Phi_{L_2L}, \Phi_{L_1H} - \Phi_{L_2H})$$

$$\min(\Phi_{H_1L} - \Phi_{H_2L}, \Phi_{H_1H} - \Phi_{H_2H}) \leq \ln(\frac{q_H^1}{q_H^2}) \leq \max(\Phi_{H_1L} - \Phi_{H_2L}, \Phi_{H_1H} - \Phi_{H_2H})$$

Proof. See the online appendix. □

Intuitively, the equilibrium percentage of women who decide to participate in period 1 depends on the marital surplus difference between marrying in period 1 and period 2 given any partner type. The larger the difference, the higher the percentage of women seeking partners in period 1.

One corollary of Proposition 1 is that the equilibrium strategy $\mathbf{q}$ satisfies the following conditions: $0 < q_e^1 < 1, 0 < q_e^2 < 1, \forall e \in \{L, H\}$. In equilibrium, it will never happen that all women of the same education type choose to participate in period 1 or period 2, as long as the surplus Φ terms are bounded. Intuitively, if all women of one education type choose to participate in period 1, a woman could benefit by choosing to participate in period 2, which makes her the only older woman with that education. The scarcity of this type would earn large marital returns for the woman. Since the support of Gumbel distribution is $\mathbb{R}$, the potential return could be large enough such that being the only one of older type in period 2 is more rewarded than participating in period 1 no matter how large the surplus difference $\Phi_{e_1y} - \Phi_{e_2y}$ is as long as it is finite. [6]

Proposition 2. *If given education type $e \in \{L, H\}$, $\Phi_{e_1H} - \Phi_{e_2H} = \Phi_{e_1L} - \Phi_{e_2L}$, then q_e^1, q_e^2 are uniquely pinned down by:*

$$q_e^1 = \frac{\exp(\Phi_{e_1L})}{\exp(\Phi_{e_1L}) + \exp(\Phi_{e_2L})}, q_e^2 = \frac{\exp(\Phi_{e_2L})}{\exp(\Phi_{e_1L}) + \exp(\Phi_{e_2L})}$$

Proof. See the online appendix. □

$\Phi_{e_1H} - \Phi_{e_2H} = \Phi_{e_1L} - \Phi_{e_2L}$ indicates that the gain of female youth in surplus is independent of men's education. [7] This means that male education and female youth don't interact in the marital surplus, hence the marginal contribution of female youth in the surplus doesn't depend on their partner's education type either. In a matching model, individuals' marital gain come from their marginal contributions to the surplus. In this case, women get all the benefit (or cost) of female youth if they choose to participate in period 1. Their choice of marriage market is fully pinned down by this difference in marital surplus independent of the education distribution of both sides.

Comparative statics

School construction would lead to a dynamic change in the population education. However, unlike in Bhaskar (2015), the current model doesn't focus on the transitory period, which

[6] One can also understand this in terms of the probability of singlehood. In the model, single probability has one-to-one correspondence with the expected utility: the lower the single probability, the higher the expected marital return. For a woman who is the only one of older type in period 2, she would almost for sure get married since the men who have very large draws for this particular older type would compete fiercely among themselves and want to marry her.

[7] It can depend on female education, e. For example, the return of female youth is larger for less educated women than more educated women, or the other way around. The empirical observation that less educated women marry earlier supports the case that the gain is larger for less educated women.

is of less interest in this paper. I will concentrate instead on how the stationary equilibrium changes in response to the change in population education. For simplicity, let's assume male population and female population are equal. Without loss of generality, I can also normalize the population of each side to 1 since the model has constant returns to scale. Let us analyze how female marriage age decision would change when the education distribution of men or women changes, respectively.

Proposition 3. *Denote female education distribution as* $G_f = (n_L, 1 - n_L)$ *and male education distribution as* $G_m = (m_L, 1 - m_L)$.

Keeping n *constant,* $\forall y \in \{L, H\}$, *a decrease in* m_L *would*

- *increase* q_e^1, *if* $\Phi_{e_1 H} - \Phi_{e_2 H} > \Phi_{e_1 L} - \Phi_{e_2 L}$;
- *decrease* q_e^1, *if* $\Phi_{e_1 H} - \Phi_{e_2 H} < \Phi_{e_1 L} - \Phi_{e_2 L}$.

Proof. See the online appendix. □

If the percentage of more-educated men increases, the equilibrium percentage of women marrying in period 1 increases if male education and female youth are complementary [8] in the marital surplus; the equilibrium percentage of women marrying in period 1 decreases if instead male education and female maturity are complementary in the marital surplus. Notice that whether the marital surplus is super-modular in male education and female education does not matter.

A stable matching maximizes the total social surplus in a TU framework (Shapley and Shubik, 1971). When male education and female youth are complementary, the social surplus is larger if we pair more educated men with younger women. Hence when there is a decrease

[8] There are at least four ways to interpret the complementarity between male education and female youth. For example, (1) all men prefer female youth and more educated men value female youth more than less educated men. (2) All women prefer more educated men and younger women value male education more than older women. (3) All men dislike female youth and more educated men dislike female youth less than less educated men. (4) All women dislike more educated men and younger women dislike more educated men less than older women. Of course, the first and second seem to be more plausible than the last two.

in m_L, the existence of more educated men would induce more women to marry in period 1 to take advantage of the higher social surplus. Vice versa.

Proposition 4. *Denote female education distribution as $N_f = (n_L, 1 - n_L)$ and male education distribution as $N_m = (m_L, 1 - m_L)$.*

Further assume super-modularity in men's education and women's education: holding m constant, $\forall e \in \{L, H\}$, a decrease in n_L would

- *decrease q_e^1, if $\Phi_{e_1 H} - \Phi_{e_2 H} > \Phi_{e_1 L} - \Phi_{e_2 L}$*
- *increase q_e^1, if $\Phi_{e_1 H} - \Phi_{e_2 H} < \Phi_{e_1 L} - \Phi_{e_2 L}$*

Proof. See the online appendix. □

A change in female education distribution affects the equilibrium female choice by affecting the potential gain of female youth via affecting the potential distribution of men a woman can marry. If n_L decreases, for a given woman, other women are more educated. They are more likely to marry with more educated men due to the complementarity in education. Therefore, on the market, more educated men are more scarce, which will discourage all women from participating in period 1 as predicted in Proposition 3 if male education and female youth are complementary.

3 Background

3.1 INPRES Primary School Construction Program in Indonesia

The Indonesian government has consistently sought to broaden educational opportunity since the country's independence in 1945. However, due to financial difficulties and political conflict, in the country's early years, Indonesia remained backward relative to neighboring countries and to countries with similar levels of income. As late as the 1971 population census, only 62% of primary school-aged 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 increased oil production and the first OPEC-engineered price rise in 1972-1973, which unexpectedly raised government revenue, 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.

Between 1973/74 and 1978/79, 62,000 primary schools were scheduled to be built. 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) ensure that each subdistrict(*kecamatan*, one level below the district(*kabupaten*) and two levels below the province level) was 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 children. This creates variation in the construction intensity exploited in the empirical analysis.

In addition to school construction, the government also provided textbooks and teacher training to ensure that the buildings were used for education purposes. Moreover, the primary school fee was abolished in 1977. By 1983, nearly all Indonesian children had at least begun to enroll in primary school, while the percentage of 7-12 year olds enrolled exceeded 90%. INPRES SD has been a successful case of education policies in developing countries.

3.2 Education System in Indonesia

In Indonesia, the education system consists of six years of primary school (*sekolah dasar, SD*), three years of junior secondary school (*sekolah menengah pertama, SMP*) and three years of senior secondary school (*sekolah menengah atas, SMA*), followed by various kinds of higher education. Children generally begin primary school at age 7. Two ministries are responsible for managing the education system, with 84 percent of schools being under the Ministry of National Education and the remaining 16 percent being under the Ministry of Religious Affairs. In the 2000 census, 86.1 percent of the population was registered as Muslim in Indonesia, but only 15 percent of school-aged kids attended religious schools. (Frederick and Worden, 1993)

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

In the 1980s, although all children began primary school, only approximately 62% of pupils entering primary school eventually graduated from grade 6. Transition between primary school and junior secondary school was low, at approximately 60%. (see Jones and Hagul, 2001, table 1, figure 2). Transition between junior secondary and senior secondary was also low: 53%. However, the survival rates of junior secondary school and senior secondary school were fairly high in Indonesia, at more than 90%. (see Tan and Mingat, 1992, table4.5, table 4.6, Table A.1)

3.3 Teacher in Indonesia

Teachers used to be of high quality and the profession used to be regarded as highly prestigious before early 1970s. However, with rapid school construction, there were not enough trained teachers and teachers were prepared in rush, which diluted the teacher quality between 1970s and 1980s in Indonesia.(Jalal et al., 2009) Figure A.1 plotted the number of newly appointed primary school teachers between 1974 and 1998. The number of new hires kept increasing since 1974 and followed a similar trend with the funding in the last column of Table A1. Lots of effort have been spent by Indonesian government to upgrade the teacher profile including the implementation of Law No. 14/2005 on Teachers and Lecturers, known as the Teacher Law which contains certification requirements for teachers. (World Bank, 2016)

Teacher salary increases on average 6.5% from primary school to junior secondary school, and increases on average 15% from junior secondary school to senior secondary school in 2004/05. [9] Compared to others with similar education levels, teachers with high education are paid less, while teachers with low education are overpaid.

Before the decentralized Education Law 20/2003, teacher hiring was very centralized, as well as the delivery of other public services. Central government agencies, the Ministry of National Education (MONE) and the Ministry Religious Affairs (MORA), were responsible for hiring teachers and paying salaries. Public teachers have always been trained by centrally accredited teacher training institutions through public examinations. In the 1970s, primary school teachers were prepared in the teacher education school called *Sekolah Pendidikan Guru* (SPG) after completing junior secondary school. Junior secondary school teachers were prepared in the institutes and faculties of teacher education (IKIP/FKIP) with Diploma 1 qualification after completing senior secondary school. Senior secondary school teachers were prepared in the institutes and faculties of teacher education (IKIP/FKIP) with Diploma 2 qualification after completing senior secondary school. (See Jalal et al., 2009, Table 1.11)

3.4 The Marriage Market in Indonesia

Marriage traditions differ in Indonesia's hundreds of different ethnolinguistic groups. However, under the influence of national policies, certain commonalities also emerge (Frederick and Worden, 1993). With more than 87% population as Muslim (according to 2010 census), polygamy is legal. However, only 2% of marriage is polygamous(Jones, 1994). Arranged marriages still exist, but the percentage is decreasing. Most marriages require the consent of the children, especially for the groom's family (Malhotra, 1991). In Indonesia, average female marriage age is about 19. It is low but similar to other southeastern Asian countries.

Divorce rate used to be very high in the 1940s due to the prevalence of early arranged

[9] In 2004/05, salary for primary school ranges from 2,733 to 3,941 (in US Dollars), ranges from 2,913 to 4,281 for junior secondary school, and ranges from 3,373 to 4,756 for senior secondary school. (Jalal et al., 2009, Table 1.5)

marriages and the liberal attitudes towards divorce, however, it has been decreasing to around 2 per 100 marriages in 1990s from 5 per 100 marriages in 1940s using survey data(Heaton et al., 2001). Fertility rate has also been declining from 5.0 in 1970 to 2.3 in 2009 with increasing education, decreasing child mortality and a family-planning program; there is no evidence for son preference in Indonesia (Frederick and Worden, 1993).

4 Data and Empirical Strategy

4.1 Data

Indonesian Census Data. For the main analysis, I use information from the 10% sample of the Indonesian Population Census 2010 downloaded from IPUMS International. [10] The 2010 census is representative of the whole country. Moreover, the birthplace (district level) of individuals is recorded, which can be used to proxy for their exposure to the primary school construction program when they were of primary school age. Education, current martial status and current spousal information is also available. Duflo(2001) uses the 1995 inter-census (SUPAS) for her analysis. Since I am interested in marriage market outcomes, to avoid truncation problems, i.e. young men and women who are single in the survey year may marry in future years, I choose the latest censuses available from IPUMS.

Table 2 shows the summary statistics for the main sample used in the empirical analysis: individuals born between 1957 and 1972. Men are more educated than women. 87% of men and 80% of women have at least primary school degree. People born in densely populated districts are more educated than those born in sparsely populated districts. In terms of marital outcomes, there are about 3% of individuals who have never married when they are surveyed in 2010, the spousal age gap (defined as husband's age minus wife's age) is about 4.7 years, which is much larger than the gap of 2 years observed in the US. Spousal education gap (defined as husband's years of schooling minus wife's years of schooling) is around 0.5

[10] Minnesota Population Center. Integrated Public Use Microdata Series, International: Version 7.1 [dataset]. Minneapolis, MN: IPUMS, 2018. https://doi.org/10.18128/D020.V7.1. The researchers would like to acknowledge the statistical agency that originally produced the data: Statistics Indonesia.

years.

Table 3 shows the summary statistics of marital outcomes by education. The never-married rate is lowest for individuals with just primary school degree for both men and women. Spousal age gap is similar for men with different education levels, around 4.5 years. However for women, the spousal age gap decreases as female education increases. Spousal education gap increases for men as education increases and decreases for women as education increases. Descriptive statistics for marital outcomes by education are similar across sparsely and densely populated districts.

School Construction Data. The number of schools planned to be constructed across districts is collected in Duflo (2001). *Intensity* is defined as the average number of primary schools planned to be constructed between 1973 and 1978 (inclusive) per 1000 children aged 5-14 at the district level in the 1971 census. From Table 2, we can see that on average, 1.94 primary schools were built per 1000 children across the country. 0.48 more schools per 1000 kids were built in sparsely populated districts than densely populated districts. Sparsely populated districts refer to those districts whose population density in 1971 (defined as population in 1971 divided by the area of that district in 1971) is below median density (defined as the density of the district of birth for the median person in the sample, which is 497 inhabitants per square kilometer.) Accordingly, densely populated districts refer to the rest of districts with density larger than the median density.

Link District Code between Censuses. Indonesia has experienced a substantial increase in the number of districts (*Pemekaran Daerah*) since the enactment of Law No.22 of 1999 concerning districtal autonomy. The number of districts increased from 271 in 1971 to 304 in 1995, to 437 in 2005, and to 494 in 2010. To tackle this issue, I use the GIS shape-files provided by IPUMS across census years to link districts of birth in 2010 back to the district of birth variable in 1995 to assign the proper program intensity to each individual. Since most of this expansion is in the form of dividing existing regencies into several small regencies, I can link the majority of the regencies. [11]

School Quality Data. The number of schools and teachers at different levels is available from the Ministry of Education, which is also collected in the original dataset used in Duflo (2001). As for teacher quality, I adapt the method used in Behrman and Birdsall (1983); Bharati et al. (2018): calculating the percentage of teachers (self-report) who complete secondary school or some college across districts in the Indonesian censuses of 1971, 1980, and 1990 and the inter-censuses of 1976 and 1985.

4.2 Identification Strategy

Education. To analyze how the education distribution was impacted across districts and birth cohorts, my empirical strategy is difference-in-differences, as used in Duflo (2001). One difference comes from the school construction intensity, defined as the average number of primary schools built between 1973 and 1978 in one district per 1000 children aged 5 $\sim$ 14 in 1971. The other difference comes from birth cohorts. In Indonesia, children attend primary school at ages 7 $\sim$12. Those aged 13 or above in 1974 would not have been impacted by the program because they were already out of primary school. For those aged less than or equal to 12 in 1974, the younger they were, the more exposed they were to this school construction program.

To estimate the effect of the school construction on younger cohorts, the following regression can be run:

$$y_{ijk} = \alpha_j + \beta_k + (P_j T_i)\gamma_1 + (\boldsymbol{C_j} d_k)\delta_k + \varepsilon_{ijk} \tag{1}$$

where y_{ijk} is the outcome variable of individual i of birth cohort k in district j, α_j denotes the district fixed effect, and β_k denotes the birth cohort fixed effect. T_i is a dummy that indicates whether the individual was born in the younger cohorts (1968 $\sim$ 1972). P_j is the school construction intensity in district j. ε_{ijk} is the error term. C_j represents other district-specific variables including the number of children in the district of birth in 1971, the school

[11] Between 1995 and 2010, I can link all regencies except Sarmi regency (9419) in Papua Province.

enrollment rate in the population in 1971 and the allocation of the water and sanitation program as in Duflo (2001). d_k is a dummy that indicates whether the individual i is of birth cohort k. Controlling the interaction effect between C_j and birth cohort dummies can avoid estimate bias from mean reversion or omitted programs. The coefficient of interest is γ_1, which captures the average effect of building one primary school per 1000 children on the treated young cohorts.

The regression can also be generalized to the following specification:

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

where d_{kl} is a dummy that indicates whether birth cohort k individuals are of age l in 1974 (year-of-birth dummy).

In this specification, the coefficients γ_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 γ_l. A valid identification strategy would require that $\gamma_l = 0$ if $l > 12$, i.e., the variation in the outcome variable is not correlated with the primary school available starting in 1974 for the children who were already out of primary school in 1974. I should expect that for $l \leq 12$, $\gamma_l > 0$, and that γ_l decreases with l, in other words, a higher impact on the younger generation.

Marriage Market. Different from education, martial outcomes are an equilibrium phenomenon as shown in the conceptual framwork: not only do they depend on individuals' own characteristics, they also depend on others'. Take a woman as an example, a change in her own education will affect her marital outcomes, so does a change in other women's education and a change in men's education.

Empirically, the large positive spousal age gap in my sample provides a novel setting in which only the female education distribution in the marriage market changes, while that of men does not. Because women marry older husbands, for the first few cohorts of women whose education is impacted, their potential husbands are older than they are and would not have been impacted by the program. The larger the average spousal age gap is, the more birth cohorts of women I can attribute to the experiment in which only the women's, not the men's, education distribution changes in the marriage market. To capture this idea, the regression specification is modified as following:

$$y_{ijk} = \alpha_j + \beta_k + (P_j T_i)\gamma_1 + (P_j T_i^p)\gamma_2 + (\boldsymbol{C_j} d_k)\delta_k + \varepsilon_{ijk} \tag{3}$$

T_i equals to 1 for the birth cohorts between 1968 and 1972 with full exposure to the school construction program, T_i^p equals to 1 for the birth cohorts between 1963 and 1967 with partial exposure to the program. Therefore, γ_2 in the women's regression captures the effect when women's education changes but not their potential husbands.

However, the same logic can not be extended to the analysis of male marital outcomes, γ_2 in the men's regression reflects both the impact of a change in male education and a change in their potential wives who are younger and would have been affected by the program.

In the current context, though there is no setting in which I can disentangle the effect of a change in a woman's own education and the effect of a change in other women's education, I rely on a rough back-of-envelope calculation to disentangle the two effects quantitatively as shown in Section 6.3.

5 Empirical Results on Education

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

5.1 Whole Sample

Table 4 presents the difference-in-differences results of primary school construction on educational outcomes for men and women. Post indicates the young cohorts born between 1968 and 1972 (age 2-6 in 1974) while the control group consist of individuals born between 1957 and 1962 (age 12-17 in 1974). In terms of years of schooling, the effect of school construction is not significantly different from zero, which is a surprising result given the positive results found in the literature (Duflo, 2001; Akresh, Halim, and Kleemans, Akresh et al.; Mazumder et al., 2021). I'll discuss several potential reasons behind this discrepancy later.

Column (2)-(4) break down years of schooling to education attainment: wether the individual completes primary school, junior secondary school or senior secondary school. These variables directly come from the census data, while years of schooling is inputed by the author. Building primary school does increase the primary school attainment rate for men but not women. In terms of secondary school attainment, men are not affected while women are negatively affected: women are 1.2 percentage less likely to complete junior secondary school and are 0.8 percentage less likely to complete senior secondary school when one more primary school is constructed in their birth of district per 1000 kids. However, this is not to say that people's secondary school attainment decreases but indicates that building one more primary school may slow the increase of secondary school attainment compared to the districts where fewer primary schools were constructed in the program.

Results in the Literature and Reasons Behind the Difference:

Duflo (2001) used the 1995 Indonesian intercensal survey (SUPAS) and found that one more primary school construction increased male years of schooling by 0.19 years. Breaking down into two subsamples by population density in 1971, Duflo (2001) found that "the program had no effect in densely populated districts, and a large effect in sparsely populated districts". Using the same data as in Duflo (2001), Ashraf et al. (2020) found the program had no effect on female primary school attainment rate in the whole sample but increased female primary school attainment rate by 2.5 percentage points for those ethnicities with bride price practice.

Using the Indonesian National Socioeconomic Survey conducted in 2016 (Susenas 2016), Akresh, Halim, and Kleemans (Akresh et al.) found that the program increased male years of schooling by 0.27 years and female years of schooling by 0.23 years. They also found the program had significantly positive results for male attainment rates of primary school, junior secondary school and senior secondary school, positive results for female attainment rate of primary school but not junior or senior secondary school. Using data from the Indonesian Family Life Survey (IFLS), Mazumder et al. (2019) and Mazumder et al. (2021) found that one primary school constructed increased male primary school attainment rate by 3.2 percentage points and female primary school attainment rate by 5.3 percentage points.

Previous papers including the current one use the same variation but different datasets surveyed at different years. There are several reasons why the results are not consistent with each other. First is the representativeness of the datasets: though the 1995 Indonesian intercensal survey is representative of Indonesian population excluding four provinces, the sample design involves multistage random sampling, observations may bear different weights. Therefore, findings using 1995 intercensal data do not necessarily indicate the average effect for Indonesian population even using the weight information since weighted estimates may not necessarily solve the problem according to Solon et al. (2015) if the effects are heterogeneous. IFLS data covered about 300 out of nearly 500 districts in Indonesia in 2010s, and has a relatively small sample. Both the Indonesian 2010 census and Susenas 2016 are representative of Indonesian population. The representativeness of different datasets may explain partial discrepancy between results in this paper and the papers using SUPAS 1995 and IFLS data (Duflo, 2001; Breierova and Duflo, 2004; Ashraf et al., 2020; Mazumder et al., 2019,0). Second is the difference of the survey years. Later survey years may suffer (1) sample attrition due to mortality and migration abroad and (2) larger measurement errors in linking districts across years given the large increase in the number of districts in Indonesian between 1970s and 2010s. Sample attrition could come from either less educated individuals don't survive to the later survey years or more educated individuals emigrate to other countries as time passes by, however, whether an overestimate or underestimate of the effect should be expected depends on the difference between the attrition rates across treated and control cohorts and also across different districts of various school construction intensities. At the same time, the measurement errors in linking districts may bias estimates towards zero for results using later surveys.

Event-study figures and the validity test of assumption:.

The assumption underlying the difference-in-differences strategy is that different districts should have parallel trends for the education outcomes across birth cohorts without the school construction program. This could be tested showing the estimates of γ from the specification of Equation 5 for the old cohorts who should have not been affected by the program. Figure 1 shows the event-study graphs for the four education outcomes: years of schooling, primary school attainment dummy, junior secondary school attainment dummy and senior secondary school attainment dummy. For simplicity, the coefficients are grouped by three birth cohorts. From the figure, we can see that for men, estimates of γ are not significantly different from zero for all four educational outcomes in the old cohorts, and as suggested in Table 4, the program increases the primary school attainment rate for men of younger birth cohorts. For women, estimates of γ are not significantly different from zero for junior secondary school attainment and senior secondary school attainment rate. However, for primary school attainment rate, a decreasing trend is observed across cohorts who shouldn't have been affected by the program for women. To understand this decreasing trend better, the top panel of Figure A.2 shows the estimates of γ by each birth cohort using 2010 census, it is shown that the trend is mostly driven by the two oldest cohorts of age 23-24 at 1974, while the estimates for the other birth cohorts are not significantly from zero and there is no trend. This could be due to differential mortality selection in districts with different school construction intensities: those districts with higher construction intensity tend to be poorer, hence those less educated individuals are even less likely to survive to the survey year 2010 compared to those in the districts with lower construction intensity, this could explain the large positive estimate for age 23-24 at 1974. As a robustness check, the bottom panel of Figure A.2 replicates the estimation using 2000 census, which may suffer less severe mortality selection problem, we can see that estimates of γ are not significantly different from zero for the old cohorts except the oldest cohort of age 24 at 1974 and are smaller.

5.2 Heterogeneity Results on Education

As suggested in Duflo (2001), the effect of this program may differ depending on the population density of the districts. In this section, I repeat the previous exercise on two subsamples divided by population density: sparsely populated districts with densities below the medium density and densely populated districts with densities above the medium. Population density is calculated as the population in 1971 divided by the area of each district in 1971, and the median density is defined as the density for the district of birth for the median person in the sample, which is 497 inhabitants per square kilometer. There are 184 districts in the sparsely populated subsample, and the weighted average number of schools constructed per 1000 children is 2.18. There are 90 districts in the densely populated subsample, and the weighted average number of schools constructed per 1000 children is 1.70, which is lower than that in the sparsely populated subsample, as shown in Table 2.

Table 5 presents the difference-in-differences results for the two subsamples. In sparsely populated districts, one school constructed per 1000 children increases male years of schooling by 0.096 years and has no effect on female years of schooling. In terms of educational attainment, the program increases the primary school and junior secondary school attainment for men and increases the primary school attainment rate for women. In densely populated districts, on the contrary, school construction has a significantly negative effect on years of schooling for both men and women. One school constructed per 1000 children decreases male years of schooling by 0.093 years and female years of schooling by 0.17 years. In terms of educational attainment, the program has a negative effect on junior secondary school and senior secondary school attainment rate for both men and women.

5.3 Mechanism

It is surprising to find a negative effect on secondary educational attainment in densely populated districts, since if any, the primary school construction should have some positive spillover effects on secondary schools. Therefore, it is important to find the mechanisms behind this finding. Two possibilities are explored: (1) building primary schools crowds out resources available to secondary schools and deteriorates secondary school quality and (2) a sudden increase in primary school availability may decrease primary education quality and hence the quality of primary school graduates. Heterogeneity between sparsely populated and densely populated districts is further explored to show that both conjectures are plausible.

First Conjecture: Deterioration in Secondary Education Quality?

Teacher scarcity is always a challenge in Indonesia's education system. Building primary schools increases the aggregate demand for teachers. This could affect the availability of secondary school teachers if we consider a common pool of teacher hiring in primary and secondary school. To test this conjecture, I use the total number and average number of teachers per school in secondary education across districts in the years after the INPRES-SD program and check whether there is a differential change in districts where more primary schools were constructed. Specifically, I estimate the following model:

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

where j denotes district, and t denotes the survey year: 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 teachers in year t in district j. d_{tl} is a year dummy indicating whether $t = l$. α_j denotes the district fixed effect, β_j denotes the survey year fixed effect. P_j is the school construction intensity in district j. ε_{jk} is the error term. C_j represents other district-specific variables including the number of children in the district of birth in 1971, the school enrollment rate in the population in 1971 and the allocation of the water and sanitation program as in the regression for education outcomes.

Results are presented in column (1)-(3) in Table 6, the outcome variables are total number of teachers in secondary school, total number of secondary schools and the average number of teachers per school. Both junior secondary and senior secondary school are considered as secondary school in the analysis. The omitted baseline year in the table is 1973/74 ($t = 1$), which is the beginning year of the school construction program, therefore, the coefficient could be interpreted as whether districts with higher intensity of primary school construction have experienced differential change in teacher resources in secondary education in the later years compared to 1973/74. Unfortunately, I don't have information for years before 1973/74 in the data, hence it is impossible for me to test whether there is a pre-trend in previous years.

The negative coefficients in column (1) and column (3) suggest that in districts where more primary schools were constructed, a smaller increase is observed in the total number and the average number of teachers per school in secondary education in later years. For the treated cohorts in previous education analysis, the birth cohorts between 1968 and 1972, they were going to attend secondary school between year 1980 and 1984, and would be out of secondary school between year 1986 and 1990. Hence secondary education resources between 1980 and 1990 are especially relevant to them. In terms of primary education, it is not surprising to see positive effect in column (5) since more primary schools have been constructed. And column (4) shows a positive effect of the program on the total number of teachers which is consistent with the teacher crowding out story.

Moreover, since the negative effect on secondary school attainment is only observed in densely populated districts, this negative effect on the number of teachers in secondary school should also only exist in densely populated districts if this is the mechanism responsible. Figure 2 separately plots the coefficients before the interaction term of the year dummy and school construction intensity from the previous specification for sparsely and densely populated districts. A negative effect on total number of teachers and average number of teachers in secondary education is found for densely populated districts but not for sparsely populated districts. This confirms the conjecture that primary school construction increases the demand for teachers, which crowds out teacher resources available for secondary school education and leads to a negative effect on the secondary school attainment rate. Moreover, this phenomenon exists only in densely populated districts.

Second Conjecture: Deterioration in Primary Education Quality?

A second conjecture is that the deterioration in primary school quality may lead to a decrease in primary education quality which reduces student quality among primary school graduates and in turn 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 have been sacrificed (Jalal et al., 2009; Bharati et al., 2018).

The empirical specification is similar to the specification above in the first conjecture analysis. The outcome variables are the educational outcomes of primary school teachers in the census surveys across 1971, 1976, 1980, 1985 and 1990. The baseline year is 1971, before the school expansion program started.

Table 7 shows the coefficients of the interaction term between the census year fixed effect and school construction intensity for the three educational outcomes of teachers: years of schooling, a dummy variable that indicates the teachers have completed senior secondary school and a dummy variable that indicates the teachers have completed some post secondary education. Consistent with the results in Bharati et al. (2018), a negative (though not significant) impact of the program is found on teacher quality in terms of years of schooling and secondary school attainment in 1976, but not for later years. In terms of post-secondary education, results in column (3) suggest that a smaller decrease in the percentage of teachers who have some post-secondary education is observed in districts where more primary schools are constructed. Breaking the sample by population density, Figure 3 shows the coefficients for the three educational outcomes for the two subsamples. From the figures, we can see that there is no differential pattern between the sparsely and densely populated districts for the census years since 1980. However, for the census year 1976, for years of schooling and the dummy variable indicating completing some post-secondary education, a larger negative effect is observed for densely populated districts but not for sparsely populated districts. This suggests that the deterioration in primary education quality may be responsible for the negative impact on the secondary school attainment rate.

5.4 Summary.

Here is a summary of the results on educational outcomes.

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 districts, there is a positive effect on primary school attainment for both men and women, a positive effect on secondary school attainment rate for men but no effect for women.

Result 3: In densely populated districts, 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 districts, I should expect different results on marriage market outcomes in sparsely and densely populated districts.

6 Empirical Results On Marriage Market

In this section, I present the reduced form results on the effect of school construction on individuals' marital outcomes, more specifically, spousal age gap, spousal education gap and never-married rate. As discussed in the identification strategy, a modified specification as in Equation 3 is in use and the corresponding results are shown below.

6.1 Spousal Age gap and spousal education gap

Results are shown in Table 8 for years of schooling, spousal educational gap and spousal age gap for men and women in the whole sample and the two subsamples.

The most interesting estimate is the effect on women of partially treated cohorts, since we can interpret it as the impact of a change in women's education on their marital outcomes while keeping their potential husbands' education unchanged. Column (4)-(6) in Panel A suggest that in the whole sample, one additional school per 1000 kids constructed reduced women's years of schooling by 0.039 years, increased spousal education gap by 0.018 years and spousal age gap by 0.049 years. If we look at the two subsamples separately, as shown in Column (4)-(6) in Panel B, in sparsely populated districts: women are not significantly affected by the program, the school construction program has no effect on women's education, spousal education gap and spousal age gap. However, in densely populated districts, for women of partially treated cohorts, one additional school constructed has a negative effect on their years of schooling by 0.11 years, no significant impact on the spousal education gap and a positive effect on the spousal age gap by 0.075 years. This suggests that the spousal age gap increases when we decrease female education and hold male education unchanged in the marriage market.

For the women of fully exposed cohorts, both their education and their potential husbands' education could be affected by the program. The estimate should capture both effect. In densely populated districts, the program has a larger negative effect on female years of schooling for the fully exposed cohorts (0.17 years) compared than those cohorts of partially exposed (0.11 years) and also a stronger positive effect on female spousal age gap (0.14 years). Male years of schooling in the partially exposed cohorts are also affected, therefore we can not distinguish the roles of a change in female education in the fully exposed cohorts and of a change in male education in the partially exposed cohorts. In sparsely populated districts, the program has no significant effect on women's education and no effect on the partially treated cohorts of men either, hence we don't observe any effect of the program on the spousal education gap and spousal age gap for women.

The interpretations of results on male spousal education and age gap are complicated since it is hard to find good control groups in terms of marital outcomes given their potential wives are younger than them. Control groups in Table 8 are those born between 1957 and 1962, even though they should have not been affected by the program, their wives could have been partially affected since wives are younger. In sparsely populated districts, female education is unaffected by the program ex post, the results on male marital outcomes may suggest when male education increases holding their potential wives' education unchanged, their spousal age gap increases.

Figure 4 shows the estimates of the general specification in Equation 5 for female spousal age gap in sparsely and densely populated districts. Consistent with the findings in Table 8, we can see that the program has no effect on female spousal age gap in sparsely densely populated districts, but a positive effect in densely populated districts. Figure 5 replicates the exercise for female spousal education gap. Also consistent with Table 8, no effect is found for both sparsely and densely populated districts.

6.2 Never-Married Rate

Table 9 shows the results for never-married rate. Results for educational attainment are shown besides to help with the interpretation.

Let's start with the women of partially exposed cohorts. Column (4)-(6) suggest that the program has no effect on their never-married rate in both sparsely populated districts where female education is not affected and densely populated districts where female education is negatively impacted. Results in densely populated districts seem odd but Table 3 explains why: the never-married rate is the lowest for women with just primary school degree in densely populated districts, hence the effect of a decrease in primary school attainment rate and a decrease in secondary school attainment rate on never-married rate have the opposite sign and may cancel out each other. If we look at the women of fully exposed cohorts: in sparsely populated districts, the program has a positive effect on primary school attainment rate but no effect on never-married rate though there is no impact on men of partially exposed cohorts; in densely populated districts, the program has a negative effect on secondary school attainment rate and negative effect on never-married rate. However, male education for the partially exposed cohorts in densely populated districts are affected by the program, hence we are not sure whether the negative effect on never-married rate is due to a decrease in women's secondary attainment or a decrease in secondary attainment of their potential husbands' or both.

For men, in densely populated districts, we see a negative effect on secondary school attainment rate and a negative effect on never-married rate for both partially exposed cohorts and fully exposed cohorts. In sparsely populated districts, the program has positive effect on fully exposed cohorts but no effect on the never-married rate. Again, we should be cautious in the interpretation of these results since the potential wives of men in control group could have been affected by the program.

Figure 6 shows the estimates of the general specification in Equation 5 for female never-married rate in sparsely and densely populated districts. The program does not affect female never-married rate in either sparsely or densely populated districts.

6.3 Mechanism and Interpretation

Marriage choices could be affected by both a person's own characteristics and those of the others participating in the same marriage market, as shown in the conceptual framework. To interpret our findings in the marriage market, a discussion of both a change in a person's own education and a change in the education distribution of the others is necessary. More specifically, there are three types of changes which could affect the marriage market: a change in own education, a change in the education distribution of others on the same side and a change in the education distribution of the others on the other side.

To facilitate the discussion, Table 1 summarizes the effect of school construction program on education for both men and women of different birth cohorts in densely populated districts. We can see that by comparing the women of partially affected cohorts (i.e., women born between 1963 and 1967) and the women of not affected cohorts (i.e., women born between 1857 and 1962), we could mute the effect of a change in the education distribution of others on the other side. However, it's impossible to find two settings in our context to separate the effect of a change in own education and a change in the education distribution of others on the same side.

To separate the effect of a change in own education and a change in the education distribution of others on the same side, we need to apply a rough back-of-the-envelope calculation.

As shown in Table 3, spousal age gap is different for people of different education levels. On average, more educated women have smaller spousal age gaps. In densely populated districts, the average spousal age gap is 5.94 for women not completing primary school, is

Table 1: Summary of changes in education in densely populated areas

Gender	Birth Cohorts	Age at 1974	Effect on own education	Effect on the education of potential spouses
Women	1957-1962	17-12	null	null
Women	1963-1967	11-7	negative	null
Women	1968-1972	6-2	negative	negative
Men	1957-1962	17-12	null	negative
Men	1963-1967	11-7	negative	negative
Men	1968-1972	6-2	negative	out of sample

4.94 for women completing primary school but not senior secondary school and is 3.24 for women completing senior secondary school.

As shown in Table 9, on average, one more primary school constructed per 1000 children during the school construction program decreased the number of women completing primary school by 0.86 percentage points and the number of women completing senior secondary school by 0.73 percentage points.

Hence a rough back-of-the-envelope calculation would predict that if only own education were affected by the school construction program, one additional primary school built per 1000 children would increase the spousal age gap by about $0.73\% * (5.94 - 3.24) + (0.86\% - 0.73\%) * (5.94 - 4.94) = 0.021$ years. While the total effect is about 0.075 years as shown in Table 8, which is much larger than our rough estimate of the change brought by a pure change in own education. This implies that a decrease in other women's education holding own education and potential husbands' education distribution unchanged leads to an increase in the spousal age gap. Combining this finding and the predictions of the two-to-one dimensional OLG matching model with transferrable utilities in the conceptual framework, it suggests that male education and female youth are complementary in our context.

7 Discussions

7.1 Census of 2000

To check the robustness of the surprising results on secondary school attainment rate, I replicate the analysis using the 2000 census, in which the youngest cohort (born in 1972) was already 28. Results are shown in Table A2 and Table A3, similar to Table 4 and Table 5 in the main analysis using 2010 census data. As shown in Table A2, in the full sample analysis, the program does increase primary school attainment rate for men, but decrease secondary school attainment rate for women. By looking at the subsample analysis in Table A3, the negative effect on secondary school attainment mainly comes from the densely populated districts, confirming our finding using the 2010 census.

7.2 Unavailability of school construction data for future years

This program spanned 1973/74-1988/89, as shown in Table A1. However, the district-specific construction target is only available for the first six years (1973/74-1978/79), hence in the empirical analysis, the sample are limited to those individuals born at or before 1972 who were older than age 7 in 1979. For those born after 1972, I am unable to identify the primary school construction intensity they were exposed to at age 7[12]. Unobserving school construction data after 1979 would not invalidate the empirical analysis since we are comparing older cohorts who would not have been affected by the program with younger cohorts who were between age 2 and age 7 in 1974. All children in the treatment group were of age 7 and above in 1979, and would not have been affected by primary school construction after 1979. However, this would still create two issues. First, the effect of one additional school built may be overestimated if schools built after 1979 could have positive effect on current students who have enrolled before 1979. Second, this creates more difficulties to test the mechanism behind the negative effect in the secondary school attainment. If school

[12]For later years, the aggregate number of school construction targets was available. New primary school entrants kept increasing between 1973 and 1979, then fluctuated about 4.3 million (World Bank, 1989, page 16).

construction ended in 1979, we could test whether it is the short-run teacher shortage that contributed to the negative effect on the secondary school education by comparing the children who were exposed (to schools construction between 1973 and 1979) and those even younger children. If it was short-run shortage, we should expect the negative effect diminishes for the younger children.

7.3 Control group

In the main analysis, the control group consists of individuals who were between age 13 and 17 in 1974. This may not be appropriate if the secondary schools are immediately affected during the construction time of primary schools between 1973 and 1978, since these individuals would have been in secondary school during this period. If this is the case, then the best control group should be those we were at least 19 in 1974. This is actually an empirical question that can be tested using the event-study graph shown in Figure 1. In all four panels, the estimates for the cohorts of age 19-21 are not significantly different from cohorts of age 13-15. This suggests it is not the story that current secondary school teachers switch from secondary school to primary school, but that the high demand for primary school teachers affect the new supply of teachers in secondary school.

7.4 Spousal age gap

In the analysis of marital outcomes, I take advantage of the large spousal age gap (5 years) while spousal age gap is later considered as an endogenous variable that could be affected by the school construction program. This could be problematic if the change of spousal age gap would invalidate the benefit a large spousal age gap offers in the empirical setting. However, the spousal age gap remains pretty large for the sample in the main analysis across young and old birth cohorts.

8 Conclusion

This paper studies the INPRES primary school construction program in Indonesia and analyzes its impact on individuals' marital outcomes. I first show that the program has an unintended consequence on secondary school education. In densely populated districts, the secondary school attainment rate declines for both men and women due to a crowding out of teacher resources in secondary education and deterioration of teacher quality due to massive primary school construction. The program also affects individuals' marital outcomes by changing the education distribution of all participants in the marriage market. Taking advantage of the large spousal age gap, I show that female spousal age gap increases and never-married rate does not change when female secondary school attainment rate decreases holding their potential husbands' education unchanged in the densely populated districts. Applying a rough decomposition exercise and a matching model, the result also suggests that a woman marries earlier when average education of other women decreases holding their potential husbands education distribution unchanged. This finding suggests male education and female youth are complementary using a two-to-one dimensional OLG matching model with transferrable utilities.

The unintended consequence on secondary education found in this paper raises the importance of a careful evaluation of potential externalities of a large scale social intervention to policy makers and researchers. Moreover, this study 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 spousal age gap and never-married rate responds to a change in female education distribution has direct policy implications. When evaluating education policies with potential market-level impacts, we as researchers should consider both the direct effect on individuals and the indirect effect via changing market conditions.

Table 2: Summary Statistics

	All Sample		Sparsely Populated Districts		Densely Populated Districts	
	Men	Women	Men	Women	Men	Women
intensity	1.94	1.94	2.18	2.18	1.70	1.70
Education Attainment						
less than primary	0.13	0.20	0.16	0.23	0.11	0.17
primary	0.38	0.44	0.39	0.44	0.37	0.43
junior secondary	0.15	0.13	0.15	0.12	0.16	0.13
senior secondary	0.24	0.16	0.22	0.15	0.27	0.18
post secondary	0.09	0.07	0.08	0.06	0.10	0.08
Marriage Market						
never-married	0.03	0.03	0.03	0.03	0.03	0.02
spousal age gap	4.73	4.69	4.79	4.73	4.68	4.65
spousal education gap	0.47	0.47	0.46	0.45	0.48	0.49

Notes: This table presents summary statistics for the main sample in the empirical analysis, individuals born between 1957 and 1972. Intensity measure the average number of primary schools constructed per 1000 kids between 1973 and 1978. Spousal age gap is defined as husband's age minus wife's age, and spousal education gap is defined as husband's years of schooling minus wife's years of schooling. Sparsely populated districts refer to those districts whose population density in 1971 (defined as population in 1971 divided by the area of that district in 1971) is below median density (defined as the density of the district of birth for the median person in the sample, which is 497 inhabitants per square kilometer.) Accordingly, densely populated districts refer to the rest of districts with density larger than the median density.
Source: Indonesian Census 2010

Table 3: Summary Statistics on Marital Outcomes by Education

	All Sample			Sparsely Populated Districts			Densely Populated Districts		
Men By Education:	never-married rate	spousal age gap	spousal education gap	never-married rate	spousal age gap	spousal education gap	never-married rate	spousal age gap	spousal education gap
less than primary	0.04	4.45	-1.51	0.04	4.57	-1.42	0.03	4.28	-1.65
complete primary but not senior secondary	0.02	4.81	0.16	0.02	4.87	0.19	0.02	4.75	0.14
senior secondary and above	0.03	4.72	1.72	0.03	4.76	1.92	0.04	4.68	1.55
Women By Education:									
less than primary	0.03	5.82	1.84	0.03	5.73	1.80	0.03	5.94	1.89
complete primary but not senior secondary	0.02	4.96	0.44	0.02	4.98	0.39	0.01	4.94	0.49
senior secondary and above	0.04	3.17	-0.51	0.04	3.09	-0.74	0.05	3.24	-0.32

Notes: This table presents summary statistics of marital outcomes for men and women with different education levels for the same sample as in Table 2.
Source: Indonesian Census 2010

Table 4: Effect of School Construction on Education

All sample:		Indicator for Completing at least:		
	Years of Schooling	Primary School	Junior Secondary School	Senior Secondary School
Males:	(1)	(2)	(3)	(4)
Post × Intensity	0.019	0.0079*	-0.00012	-0.0036
	(0.034)	(0.0043)	(0.0054)	(0.0037)
Dep. var. mean	8.198	0.868	0.485	0.325
Observations	1,621,730	1,621,730	1,621,730	1,621,730
Clusters	274	274	274	274
Adjusted R-squared	0.161	0.116	0.143	0.116
Duflo Controls:	Yes	Yes	Yes	Yes
Females:				
Post × Intensity	-0.038	0.0032	-0.012**	-0.0084**
	(0.044)	(0.0055)	(0.0059)	(0.0039)
Dep. var. mean	7.150	0.801	0.369	0.236
Observations	1,583,837	1,583,837	1,583,837	1,583,837
Clusters	274	274	274	274
Adjusted R-squared	0.216	0.157	0.174	0.140
Duflo Controls:	Yes	Yes	Yes	Yes

Notes: This table displays results on the effect of primary school construction on years of schooling, education attainment (completing primary school and completing secondary school) for men and women. Following the strategy of Duflo (2001), the sample consists of individuals born between either 1957 and 1962 or 1968 and 1972. Post refers to the treated cohort, born between 1968 and 1972, while the untreated cohort was born between 1957 and 1962. Educational attainment data are taken from the Indonesian 2010 Census and years of schooling are inputed by the author. Intensity is the number of schools built in a district per 1,000 kids in the school-aged population. All columns include district fixed effect, birth year fixed effect, birth year interacted with number of children at 1971. Duflo Controls consist of birth year dummy interacted with number of children in 1971, with enrollment rate at 1971 and with water sanitization program. Standard errors are clustered at the birth place district level.
Significance levels: * 10%, ** 5%, *** 1%.
Source: Indonesian Census 2010

Table 5: Heterogeneity Results: Effect of School Construction on Education

	Men				Women			
		Indicator for Completing at least:				Indicator for Completing at least:		
	Years of schooling	Primary School	Junior High	Senior High	Years of schooling	Primary School	Junior High	Senior High
Panel A: Density < Medium:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post × Intensity	0.096**	0.011**	0.013**	0.0059	0.068	0.012**	0.0028	-0.0014
	(0.038)	(0.0051)	(0.0057)	(0.0042)	(0.045)	(0.0059)	(0.0061)	(0.0043)
Dep. var. mean	7.806	0.843	0.446	0.288	6.776	0.771	0.335	0.206
Observations	834,646	834,646	834,646	834,646	812,260	812,260	812,260	812,260
Clusters	184	184	184	184	184	184	184	184
Adjusted R-squared	0.139	0.131	0.119	0.086	0.197	0.168	0.148	0.109
Panel B: Density > Medium:								
Post × Intensity	-0.093**	0.0060	-0.026***	-0.021***	-0.17***	-0.0053	-0.039***	-0.019***
	(0.045)	(0.0063)	(0.0086)	(0.0060)	(0.055)	(0.0076)	(0.0099)	(0.0064)
Dep. var. mean	8.390	0.888	0.500	0.339	7.307	0.821	0.379	0.242
Observations	787,084	787,084	787,084	787,084	771,577	771,577	771,577	771,577
Clusters	90	90	90	90	90	90	90	90
Adjusted R-squared	0.176	0.087	0.165	0.141	0.230	0.137	0.198	0.166

Notes: This table repeats the exercise as in Table 4 in two subsamples: sparsely populated districts in Panel A and densely populated districts in Panel B. Population density is calculated as the population in 1971 divided by the area of each district in 1971, and the median density is defined as the density for the district of birth for the median person in the sample, which is 497 inhabitants per square kilometer. Standard errors are clustered at the birth place district level.
Significance levels: * 10%, ** 5%, *** 1%.
Source: Indonesian Census 2010

Table 6: Effect of School Construction on Number of Teachers in Secondary and Primary Education

	Secondary School			Primary School		
	total number of teachers	total number of schools	average number of teachers per school	total number of teachers	total number of schools	average number of teachers per school
	(1)	(2)	(3)	(4)	(5)	(6)
INPRES Intensity ×						
year=1978/79	-16.2*	0.58	-0.19*	24.1***	10.5***	-0.17***
	(8.36)	(0.38)	(0.11)	(9.19)	(2.00)	(0.064)
year=1983/84	-52.4*	0.79	-0.14	49.2*	14.8***	-0.073
	(31.5)	(1.16)	(0.18)	(27.1)	(4.26)	(0.11)
year=1988/89	-79.2	0.023	-0.22	73.0*	17.3***	-0.034
	(48.5)	(1.76)	(0.21)	(43.0)	(6.11)	(0.066)
year=1993/94	-74.8	0.27	-0.097	84.5	17.5**	-0.066
	(51.3)	(1.88)	(0.22)	(57.1)	(6.79)	(0.075)
year=1995/96	-68.8	0.73	-0.087	165.9**	17.4**	0.29*
	(59.0)	(2.26)	(0.19)	(68.2)	(7.04)	(0.17)
Dep. var. mean in 1973/74	557.954	36.412	14.742	1535.384	232.007	6.750
Dep. var. mean in 1995/96						
Observations	1,644	1,644	1,644	1,644	1,644	1,644
Clusters	274	274	274	274	274	274
Adjusted R-squared	0.914	0.927	0.913	0.930	0.964	0.789

Notes: This table displays the effect of school construction on the number of teachers and the average number of teachers per school in secondary and primary education in the future years. Baseline year is 1973/74. All columns include district fixed effect, survey year fixed effect, survey year interacted with number of children at 1971, survey year interacted with enrollment rate at 1971 and survey year interacted with water sanitization program. Standard errors are clustered at the district level.

Source: Indonesian Education Ministry

Table 7: Effect of School Construction on The Education of Primary School Teachers

	years of schooling (1)	senior secondary school (2)	post-secondary education (3)
year=1971	0	0	0
	(.)	(.)	(.)
INPRES Intensity ×			
year=1976	-0.18	-0.031	-0.045***
	(0.13)	(0.028)	(0.014)
year=1980	0.14	0.053*	-0.017*
	(0.12)	(0.028)	(0.0092)
year=1985	0.18	0.055*	-0.0070
	(0.13)	(0.028)	(0.010)
year=1990	0.15	0.048	-0.014
	(0.13)	(0.030)	(0.011)
Dep. var. mean in 1971	10.824	0.599	0.049
Dep. var. mean in 1990			
Observations	42,683	42,724	42,724
Clusters	266	266	266
Adjusted R-squared	0.062	0.074	0.043

Notes: This table displays the effect of school construction on the educational outcomes of primary school teachers across districts in different census years . Baseline year is 1971. All columns include district fixed effect, census year fixed effect, census year interacted with number of children at 1971, census year interacted with enrollment rate at 1971 and census year interacted with water sanitization program. Standard errors are clustered at the district level.

Source: Indonesian censuses of 1971, 1980, 1990; Inter-censuses of 1976 and 1985.

Table 8: Effect of School Construction on Spousal Age Gap

	Men			Women		
	Years of schooling	Spousal education gap	Spousal age gap	Years of schooling	Spousal education gap	Spousal age gap
Panel A: All sample:	(1)	(2)	(3)	(4)	(5)	(6)
Partial × Intensity	-0.020	-0.0087	-0.00020	-0.039*	0.018*	0.049**
	(0.022)	(0.0090)	(0.017)	(0.022)	(0.0095)	(0.019)
Post × Intensity	0.019	0.0021	0.019	-0.035	0.0019	0.047*
	(0.034)	(0.011)	(0.025)	(0.044)	(0.015)	(0.025)
Dep. var. mean	8.153	0.459	4.743	7.032	0.464	4.723
Observations	2,337,453	2,114,846	2,114,846	2,305,451	1,912,156	1,912,156
Clusters	274	274	274	274	274	274
Adjusted R-squared	0.153	0.011	0.023	0.201	0.014	0.022
Panel B: Density < Medium:						
Partial × Intensity	0.025	-0.0087	0.016	0.010	0.015	0.029
	(0.025)	(0.011)	(0.020)	(0.024)	(0.011)	(0.025)
Post × Intensity	0.096**	-0.0060	0.056**	0.071	-0.0040	-0.017
	(0.038)	(0.013)	(0.026)	(0.045)	(0.018)	(0.027)
Dep. var. mean	7.864	0.463	4.793	6.782	0.448	4.737
Observations	1,196,799	1,089,951	1,089,951	1,176,495	972,325	972,325
Clusters	184	184	184	184	184	184
Adjusted R-squared	0.131	0.015	0.030	0.183	0.019	0.027
Panel C: Density > Medium:						
Partial × Intensity	-0.082***	-0.014	-0.032	-0.11***	0.016	0.075***
	(0.029)	(0.017)	(0.029)	(0.025)	(0.015)	(0.027)
Post × Intensity	-0.093**	0.0036	-0.067	-0.17***	-0.0026	0.14***
	(0.045)	(0.018)	(0.041)	(0.055)	(0.021)	(0.041)
Dep. var. mean	8.455	0.454	4.689	7.293	0.482	4.709
Observations	1,140,654	1,024,895	1,024,895	1,128,956	939,831	939,831
Clusters	90	90	90	90	90	90
Adjusted R-squared	0.167	0.007	0.015	0.214	0.009	0.017

Notes: This table displays the effect of school construction on spousal education gap and spousal age gap using 2010 census. Spousal education gap is defined as husband's years of schooling minus wife's years of schooling, and the spousal age gap is defined as husband's age minus wife's age. The sample consists of individuals born between 1957 and 1972. Post refers to the fully treated cohort, born between 1968 and 1972, Partial refers to the partially treated cohorts born between 1963 and 1967, while the untreated cohorts were born between 1957 and 1962. All columns include district fixed effect, birth year fixed effect, birth year dummy interacted with number of children at 1971, with enrollment rate at 1971 and with water sanitization program. Standard errors are clustered at the birth place district level.
Significance levels: * 10%, ** 5%, *** 1%.
Source: Indonesian Census 2010

Table 9: Effect of School Construction on Never-Married Rate

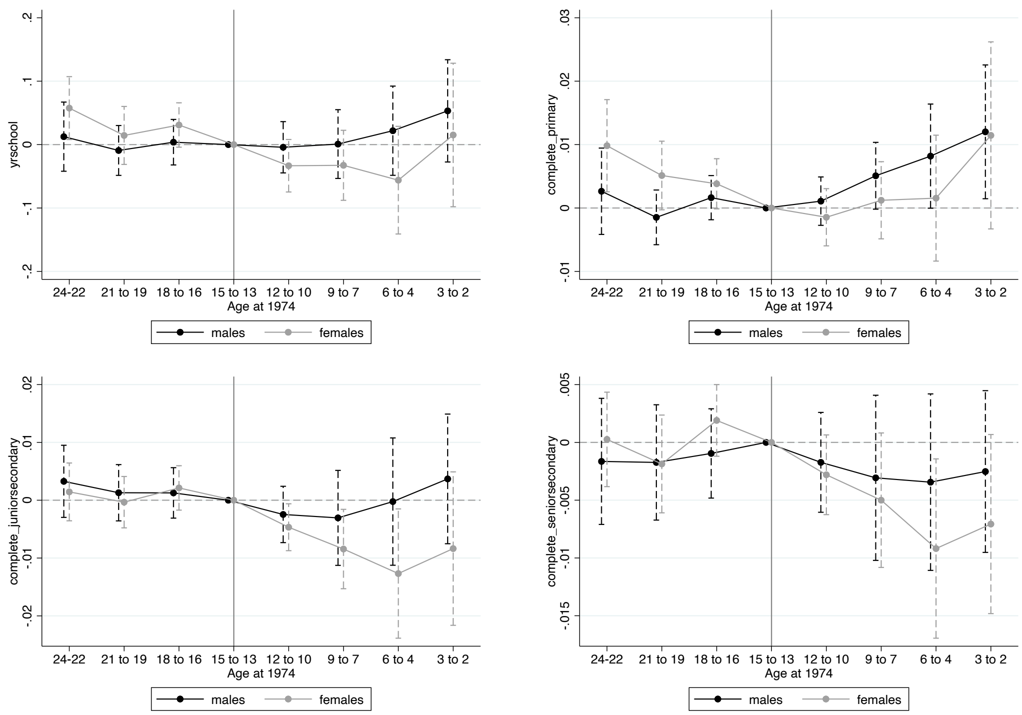
	Men			Women		
	complete primary	complete senior secondary	never-married	complete primary	complete senior secondary	never-married
Panel A: All sample:	(1)	(2)	(3)	(4)	(5)	(6)
Partial × Intensity	0.0017	-0.0036	-0.00084**	-0.0016	-0.0042*	-0.00045
	(0.0021)	(0.0031)	(0.00040)	(0.0024)	(0.0025)	(0.00051)
Post × Intensity	0.0080*	-0.0037	-0.0023***	0.0035	-0.0082**	-0.0017**
	(0.0044)	(0.0037)	(0.00088)	(0.0055)	(0.0038)	(0.00067)
Dep. var. mean	0.866	0.322	0.027	0.795	0.225	0.025
Observations	2,337,453	2,337,453	2,336,072	2,305,451	2,305,451	2,305,088
Clusters	274	274	274	274	274	274
Adjusted R-squared	0.109	0.114	0.014	0.145	0.134	0.021
Panel B: Density < Medium:						
Partial × Intensity	0.0040	0.0023	-0.00030	0.0039	-0.0022	-0.00027
	(0.0025)	(0.0038)	(0.00047)	(0.0027)	(0.0029)	(0.00056)
Post × Intensity	0.011**	0.0058	-0.00079	0.012**	-0.0013	-0.0011
	(0.0051)	(0.0042)	(0.00099)	(0.0059)	(0.0043)	(0.00072)
Dep. var. mean	0.844	0.297	0.027	0.771	0.208	0.026
Observations	1,196,799	1,196,799	1,196,044	1,176,495	1,176,495	1,176,332
Clusters	184	184	184	184	184	184
Adjusted R-squared	0.124	0.084	0.014	0.158	0.104	0.023
Panel C: Density > Medium:						
Partial × Intensity	0.00026	-0.015***	-0.0016**	-0.0086**	-0.0073*	-0.00051
	(0.0028)	(0.0049)	(0.00073)	(0.0034)	(0.0037)	(0.0010)
Post × Intensity	0.0059	-0.021***	-0.0045**	-0.0055	-0.019***	-0.0020*
	(0.0063)	(0.0060)	(0.0017)	(0.0076)	(0.0064)	(0.0012)
Dep. var. mean	0.889	0.348	0.028	0.820	0.242	0.023
Observations	1,140,654	1,140,654	1,140,028	1,128,956	1,128,956	1,128,756
Clusters	90	90	90	90	90	90
Adjusted R-squared	0.078	0.138	0.014	0.124	0.159	0.018

Notes: This table displays the effect of school construction on never-married rate using 2010 census. Never married is a dummy variable indicating the individual has never been married before the census survey time. The sample consists of individuals born between 1957 and 1972. Post refers to the fully treated cohort, born between 1968 and 1972, Partial refers to the partially treated cohorts born between 1963 and 1967, while the untreated cohorts were born between 1957 and 1962. All columns include district fixed effect, birth year fixed effect, birth year dummy interacted with number of children at 1971, with enrollment rate at 1971 and with water sanitization program. Standard errors are clustered at the birth place district level.

Significance levels: * 10%, ** 5%, *** 1%.

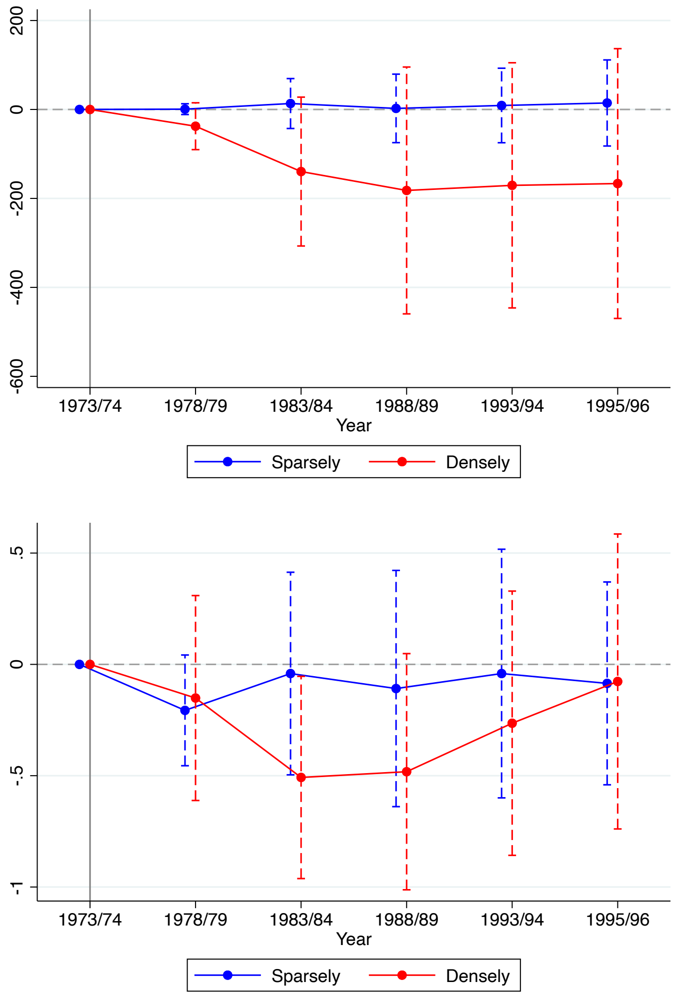
Source: Indonesian Census 2010

Figure 1: Coefficients of the Interactions: Birth cohort dummy* Program Intensity in Education Regression



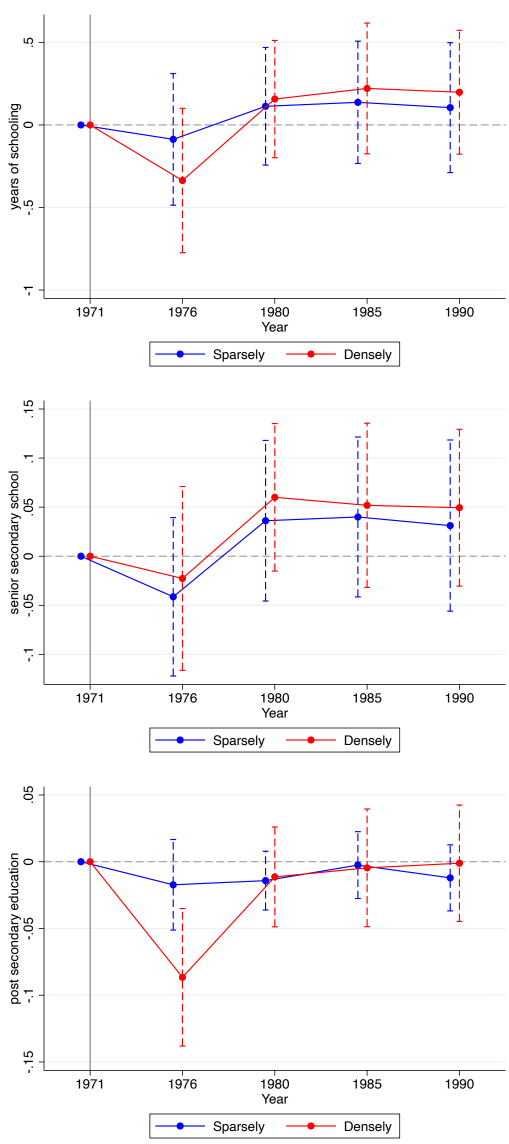
Note: This figure presents the event-study graph of the impact of the program on different birth cohorts for the full sample. The baseline cohorts are set as those we were of age 13-15 in 1974, who should have not been affected by the program. The four panels represent the results for the four educational outcomes, years of schooling, the dummy variable indicating attainment of primary school degree, attainment of junior secondary school degree and attainment of senior secondary school degree. Estimation results are shown separately for men and women using different lines. Controls include district fixed effect, birth year fixed effect, birth year dummy interacted with number of children at 1971, with enrollment rate at 1971 and with water sanitization program. Standard errors are clustered at the birth place district level. 95% confidence intervals are shown using dashed line.
Source: Indonesian 2010 Census.

Figure 2: Coefficients of the Interactions: Survey Year * Program Intensity
in Total (Top) and Average (Bottom) Number of Secondary School Teachers Equation



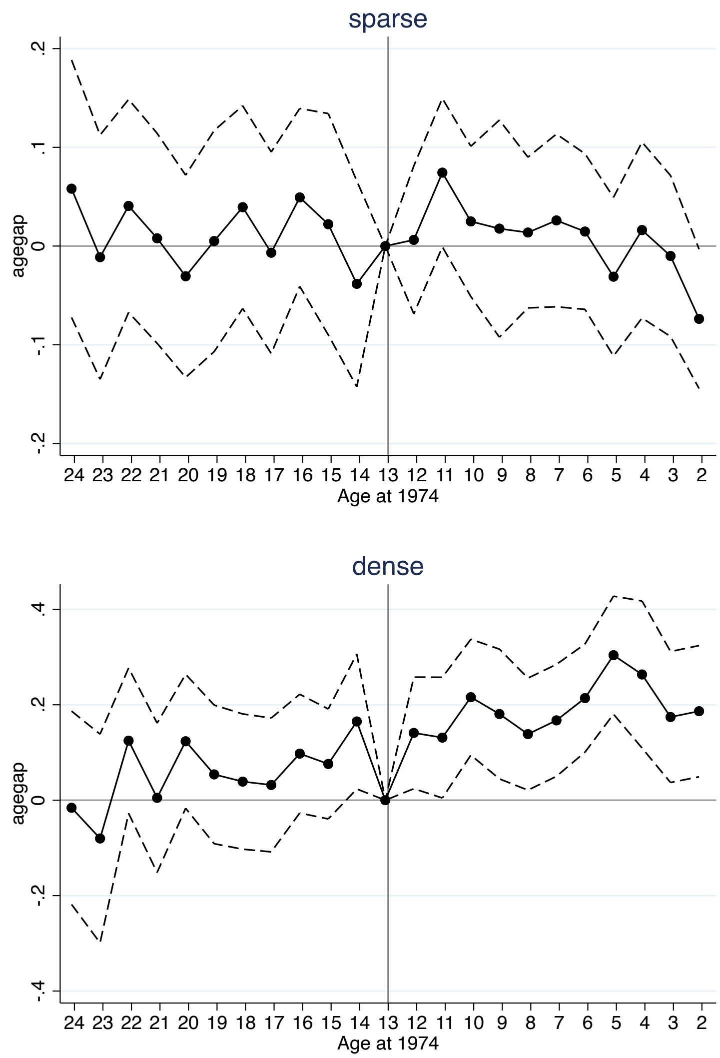
Note: This figure reports estimates of the effect of school construction on total number of teachers and average number of teachers per school in secondary school across different years in sparsely populated areas and densely populated areas. The baseline year is 1973/74. The data was provided by Indonesian Education Ministry and was collected in Duflo (2001). The dependent variable was the average number of teachers in secondary school across different districts. Controls include district fixed effect, year fixed effect, year dummy interacted with number of children at 1971, with enrollment rate at 1971 and with water sanitization program. Standard errors are clustered at the birth place district level. 95% confidence intervals are shown using dashed line.This figure supports the argument that the negative effect on secondary school attainment is due to teacher resource crowding out in densely populated districts because of primary school construction.
Source: Indonesian Education Ministry

Figure 3: Coefficients of the Interactions: Census Year * Program Intensity in Primary School Teacher Education Regression



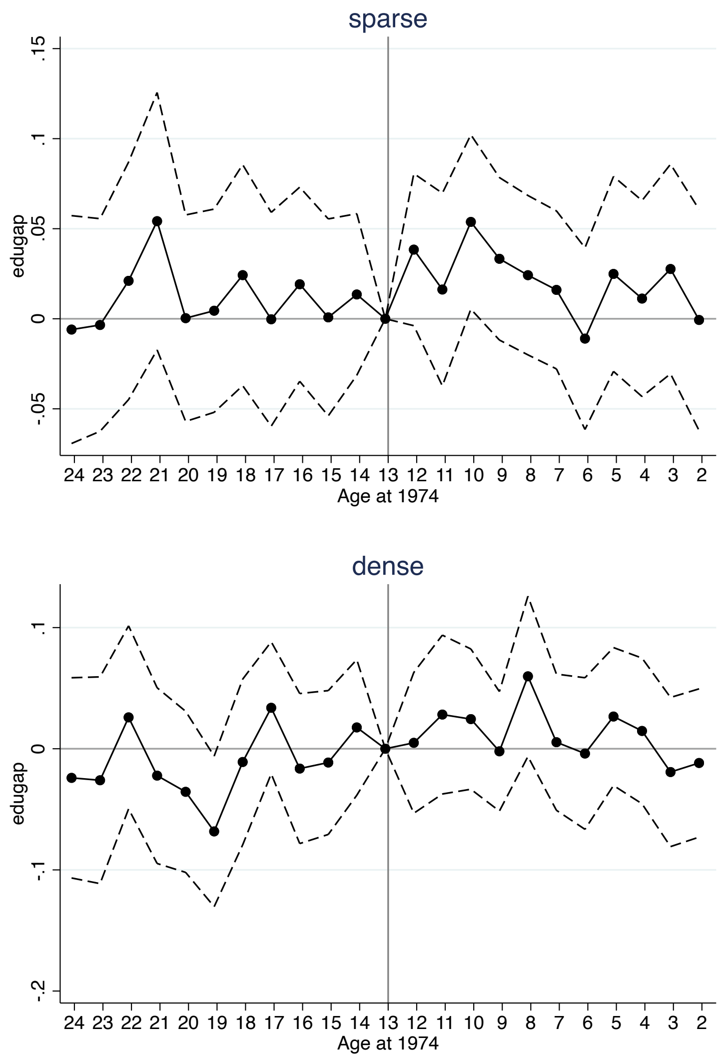
Note: This figure reports estimates of the effect of school construction on three educational outcomes of the primary school teachers across different census years in sparsely populated districts and densely populated districts. Results for years of schooling are shown in top panel, whether completing senior secondary school in medium panel and whether completing some post-secondary education in bottom panel. The baseline year is 1971. Controls include district fixed effect, year fixed effect, year dummy interacted with number of children at 1971, with enrollment rate at 1971 and with water sanitization program. Standard errors are clustered at the birth place district level. 95% confidence intervals are shown using dashed line. This figure suggests that the negative effect on secondary school attainment could be partially due to a decrease in the quality of primary school teachers due to a surge demand from primary school construction. Data in use is the same as in Table 7.
Source: Indonesian censuses of 1971, 1980, 1990; Inter-censuses of 1976 and 1985

Figure 4: Coefficients of the Interactions: Birth Cohort Dummy * Program Intensity in Female Spousal Age Gap Regression



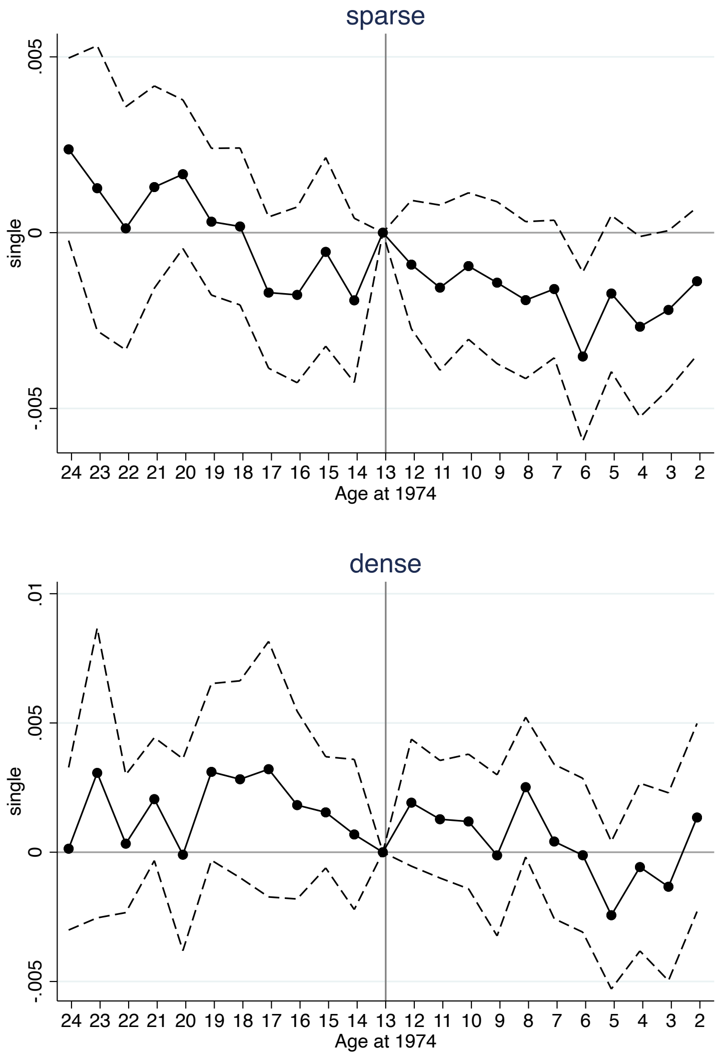
Note: This figure reports estimates of the effect of school construction on female spousal age gap in sparsely populated districts (top) and densely populated districts (bottom). Controls include district fixed effect, birth year fixed effect, birth year dummy interacted with number of children at 1971, with enrollment rate at 1971 and with water sanitization program. Standard errors are clustered at the birth place district level. 95% confidence intervals are shown using dashed line.
Source: Indonesian 2010 Census.

Figure 5: Coefficients of the Interactions: Census Year * Program Intensity in Female Spousal Education Gap Regression



Note: This figure reports estimates of the effect of school construction on female spousal education gap in sparsely populated districts (top) and densely populated districts (bottom). Controls include district fixed effect, birth year fixed effect, birth year dummy interacted with number of children at 1971, with enrollment rate at 1971 and with water sanitization program. Standard errors are clustered at the birth place district level. 95% confidence intervals are shown using dashed line.
Source: Indonesian 2010 Census.

Figure 6: Coefficients of the Interactions: Census Year * Program Intensity in Female Never-Married Rate Regression



Note: This figure reports estimates of the effect of school construction on never-married rate in sparsely populated districts (top) and densely populated districts (bottom). Controls include district fixed effect, birth year fixed effect, birth year dummy interacted with number of children at 1971, with enrollment rate at 1971 and with water sanitization program. Standard errors are clustered at the birth place district level. 95% confidence intervals are shown using dashed line.
Source: Indonesian 2010 Census.

References

Akresh, R., D. Halim, and M. Kleemans. Long-term and intergenerational effects of education: Evidence from school construction in indonesia. *Economic Journal*. Conditionally Accepted.

Andrabi, T., J. Das, and A. I. Khwaja (2013). Students today, teachers tomorrow: Identifying constraints on the provision of education. *Journal of Public Economics 100*, 1 – 14.

André, P. and Y. Dupraz (2018). Education and polygamy: Evidence from cameroon. Technical report.

Arunachalam, R. and S. Naidu (2006). The price of fertility: marriage markets and family planning in bangladesh. *University of California, Berkeley*.

Ashraf, N., N. Bau, N. Nunn, and A. Voena (2020). Bride price and female education. *Journal of Political Economy 128*(2), 591–641.

Barro, R. and X. S. i Martin (1995). *Economic Growth.* New York: McGraw-Hill.

Behrman, J. R. and N. Birdsall (1983). The quality of schooling: quantity alone is misleading. *The American Economic Review 73*(5), 928–946.

Bharati, T., S. Chin, and D. Jung (2018). Recovery from an early life shock through improved access to schools: Evidence from indonesia.

Bhaskar, V. (2015). The demographic transition and the position of women: A marriage market perspective.

Bobonis, G. J. and F. Finan (2009). Neighborhood peer effects in secondary school enrollment decisions. *The Review of Economics and Statistics 91*(4), 695–716.

Breierova, L. and E. Duflo (2004). The impact of education on fertility and child mortality: Do fathers really matter less than mothers? Technical report, National bureau of economic research.

Castro, J. F. and B. Esposito (2018). The effect of bonuses on teacher behavior: A story with spillovers. Technical report, Discussion Paper 104, Peruvian Economic Association, August.

Chiappori, P.-A., B. Salanié, and Y. Weiss (2017). Partner choice, investment in children, and the marital college premium. *American Economic Review 107*(8), 2109–67.

Choo, E. and A. Siow (2006). Who marries whom and why. *Journal of political Economy 114*(1), 175–201.

Decker, C., E. H. Lieb, R. J. McCann, and B. K. Stephens (2013). Unique equilibria and substitution effects in a stochastic model of the marriage market. *Journal of Economic Theory 148*(2), 778–792.

Dessy, S. and H. Djebbari (2010). High-powered careers and marriage: can women have it all? *The BE Journal of Economic Analysis & Policy 10*(1).

Dominguez, C. (2014). Aggregate effects on the marriage market of a big increase in educational attainment. Dissertation, Yale University.

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

Edlund, L. (2006). Marriage: past, present, future? *CESifo Economic Studies 52*(4), 621–639.

Frederick, W. H. and R. L. Worden (1993). *Indonesia: A country study*, Volume 550. Washington, DC: Federal Research Division, Library of Congress.

Galichon, A. and B. Salanié (2017). The econometrics and some properties of separable matching models. *American Economic Review 107*(5), 251–55.

Galichon, A. and B. Salanié (2021, 12). Cupid's invisible hand: Social surplus and identification in matching models. *The Review of Economic Studies*. rdab090.

Glewwe, P., E. A. Hanushek, S. Humpage, and R. Ravina (2013). School resources and educational outcomes in developing countries: A review of the literature from 1990 to 2010. *Education Policy in Developing Countries 4*(14,972), 13.

Glewwe, P. and M. Kremer (2006). Schools, teachers, and education outcomes in developing countries. *Handbook of the Economics of Education 2*, 945–1017.

Glewwe, P. and K. Muralidharan (2016). Improving education outcomes in developing countries: Evidence, knowledge gaps, and policy implications. In *Handbook of the Economics of Education*, Volume 5, pp. 653–743. Elsevier.

Hanushek, E. A. (2011). The economic value of higher teacher quality. *Economics of Education review 30*(3), 466–479.

Heaton, T. B., M. Cammack, and L. Young (2001). Why is the divorce rate declining in indonesia? *Journal of Marriage and Family 63*(2), 480–490.

Hener, T. and T. Wilson (2018). Marital age gaps and educational homogamy-evidence from a compulsory schooling reform in the uk. Technical report, Ifo Working Paper.

Jalal, F., M. Samani, M. C. Chang, R. Stevenson, A. B. Ragatz, and S. D. Negara (2009). *Teacher certification in Indonesia: A strategy for teacher quality improvement.* Departemen Pendidikan Nasional, Republik Indonesia.

Jones, G. W. (1994). Marriage and divorce in Islamic South-East Asia.

Jones, G. W. and P. Hagul (2001). Schooling in indonesia: crisis-related and longer-term issues. *Bulletin of Indonesian Economic Studies 37*(2), 207–231.

Jürges, H., S. Reinhold, and M. Salm (2011). Does schooling affect health behavior? evidence from the educational expansion in western germany. *Economics of Education Review 30*(5), 862–872.

Low, C. (2017). A "reproductive capital" model of marriage market matching. *Manuscript, Wharton School of Business*.

Malamud, O., A. Mitrut, and C. Pop-Eleches (2018). The effect of education on mortality and health: Evidence from a schooling expansion in romania. Technical report, National Bureau of Economic Research.

Malhotra, A. (1991). Gender and changing generational relations: Spouse choice in Indonesia. *Demography 28*(4), 549–570.

Mankiw, N. G., D. Romer, and D. N. Weil (1992). A contribution to the empirics of economic growth. *The quarterly journal of economics 107*(2), 407–437.

Martinez-Bravo, M. (2017). The local political economy effects of school construction in indonesia. *American Economic Journal: Applied Economics 9*(2), 256–89.

Mazumder, B., M. Rosales-Rueda, and M. Triyana (2019). Intergenerational human capital spillovers: Indonesia's school construction and its effects on the next generation. In *AEA Papers and Proceedings*, Volume 109, pp. 243–49.

Mazumder, B., M. Triyana, and M. Fernanda Rosales (2021). Social interventions, health and wellbeing: The long-term and intergenerational effects of a school construction program. *Journal of Human Resources*.

McEwan, P. J. (2015). Improving learning in primary schools of developing countries: A meta-analysis of randomized experiments. *Review of Educational Research 85*(3), 353–394.

Miguel, E. and M. Kremer (2004). Worms: identifying impacts on education and health in the presence of treatment externalities. *Econometrica 72*(1), 159–217.

Rivkin, S. G., E. A. Hanushek, and J. F. Kain (2005). Teachers, schools, and academic achievement. *Econometrica 73*(2), 417–458.

Shapley, L. S. and M. Shubik (1971). The assignment game i: The core. *International Journal of game theory 1*(1), 111–130.

Siow, A. (1998). Differntial fecundity, markets, and gender roles. *Journal of Political Economy 106*(2), 334–354.

Snodgrass, D. (1984, October). *Development Program Implementation Studies No.5: Inpres Sekolah Dasar*. Cambridge, MA.: Harvard Institute for International Development.

Snodgrass, D., L. Hutagalung, and S. Dasar (1980). *Inpres Sekolah Dasar: An Analytical Study*. Economics and Human Resources Research Center, Faculty of Economics,Padjadjaran University.

Solon, G., S. J. Haider, and J. M. Wooldridge (2015). What are we weighting for? *Journal of Human resources 50*(2), 301–316.

Tan, J. P. and A. Mingat (1992). *Education in Asia : a comparative study of cost and financing (English).* World Bank regional and sectoral studies. Washington, DC : The WorldBank.

World Bank (1989). *Indonesia - Basic education study (English)*. Washington, DC: World Bank.

World Bank (2016, January). *Indonesia Teacher Certification and Beyond*. World Bank, Jakarta.

World Bank (2018). *World Development Report, LEARNING to Realize Education's Promise*.

Young, A. (1994). Lessons from the east asian nics: a contrarian view. *European economic review 38*(3-4), 964–973.

Young, A. (1995). The tyranny of numbers: confronting the statistical realities of the east asian growth experience. *The Quarterly Journal of Economics 110*(3), 641–680.

Zhang, H. (2018, October). Human Capital Investments, Differential Fecundity, and the Marriage Market. Working Papers 2018-7, Michigan State University, Department of Economics.

Appendix

Table A1: Inpres Sekolah Dassar

Primary School Investment Program 1973/74-1988/89							
Financial year	New Primary schools	New Class-rooms for existing primary schools	Primary schools to be rehabili-tated	Primary princi-pal and teacher housing	Primary school books (mln)	Primary school sport kits	Total alloca-tion (billions of current rupiah)
1973/74	6,000	-	-	-	6.6	-	17.2
1974/75	6,000	-	-	-	6.9	-	19.7
1975/76	10,000	-	10,000	-	7.3	-	49.9
1976/77	10,000	-	16,000	-	8.6	-	57.3
1977/78	15,000	-	15,000	-	7.3	-	85.0
1978/79	15,000	15,000	15,000	-	8.5	-	111.8
1979/80	10,000	15,000	15,000	5,000	12.5	-	155.8
1980/81	14,000	20,000	20,000	7,500	14.0	-	249.8
1981/82	15,000	25,000	25,000	9,500	15.0	-	374.5
1982/83	22,600	35,000	25,000	20,000	30.0	50,000	267.4
1983/84	13,140	15,700	21,000	50,000	32.0	96,000	549.3
1984/85	2,200	12,500	31,000	60,000	32.0	157,799	526.1
1985/86	3,200	12,500	31,000	60,000	32.0	157,799	526.1
1986/87	2,200	10,000	95,000	44,070	32.6	120,000	495.9
1987/88	660	2,200	157,500	2,400	22.9	-	100.8
1988/89	250	1,350	6,000	2,650	18.5	-	112.5

Note: For the first time in 1980/81 new first-phase units were started while second-phase units were still being added to first-phase units built in the preceding year. The 1980/81 targets were 4,000 first-phase units and 10,000 second-phase units.(Snodgrass et al., 1980, table 2)
Original source: Republik Indonesia,"Nota Keuangan dan Rancangan Anggaran Pendapatan dan Belanja Negara, Tahun 1988/1989"
Source: Annex Table 4 of World Bank (1989), page 109.

Table A2: Effect of School Construction on Education using the 2000 Census

All sample:		Indicator for Completing at least:		
	Years of Schooling	Primary School	Junior Secondary School	Senior Secondary School
Males:	(1)	(2)	(3)	(4)
Post × Intensity	0.0023	0.0060*	-0.0047	-0.0041
	(0.031)	(0.0036)	(0.0050)	(0.0037)
Dep. var. mean	7.656	0.853	0.459	0.309
Observations	1,605,910	1,605,910	1,605,910	1,605,910
Clusters	274	274	274	274
Adjusted R-squared	0.163	0.103	0.147	0.120
Duflo Controls:	Yes	Yes	Yes	Yes
Females:				
Post × Intensity	-0.051	0.0015	-0.012**	-0.0066*
	(0.035)	(0.0039)	(0.0052)	(0.0038)
Dep. var. mean	6.651	0.794	0.351	0.223
Observations	1,579,697	1,579,697	1,579,697	1,579,697
Clusters	274	274	274	274
Adjusted R-squared	0.203	0.140	0.170	0.137
Duflo Controls:	Yes	Yes	Yes	Yes

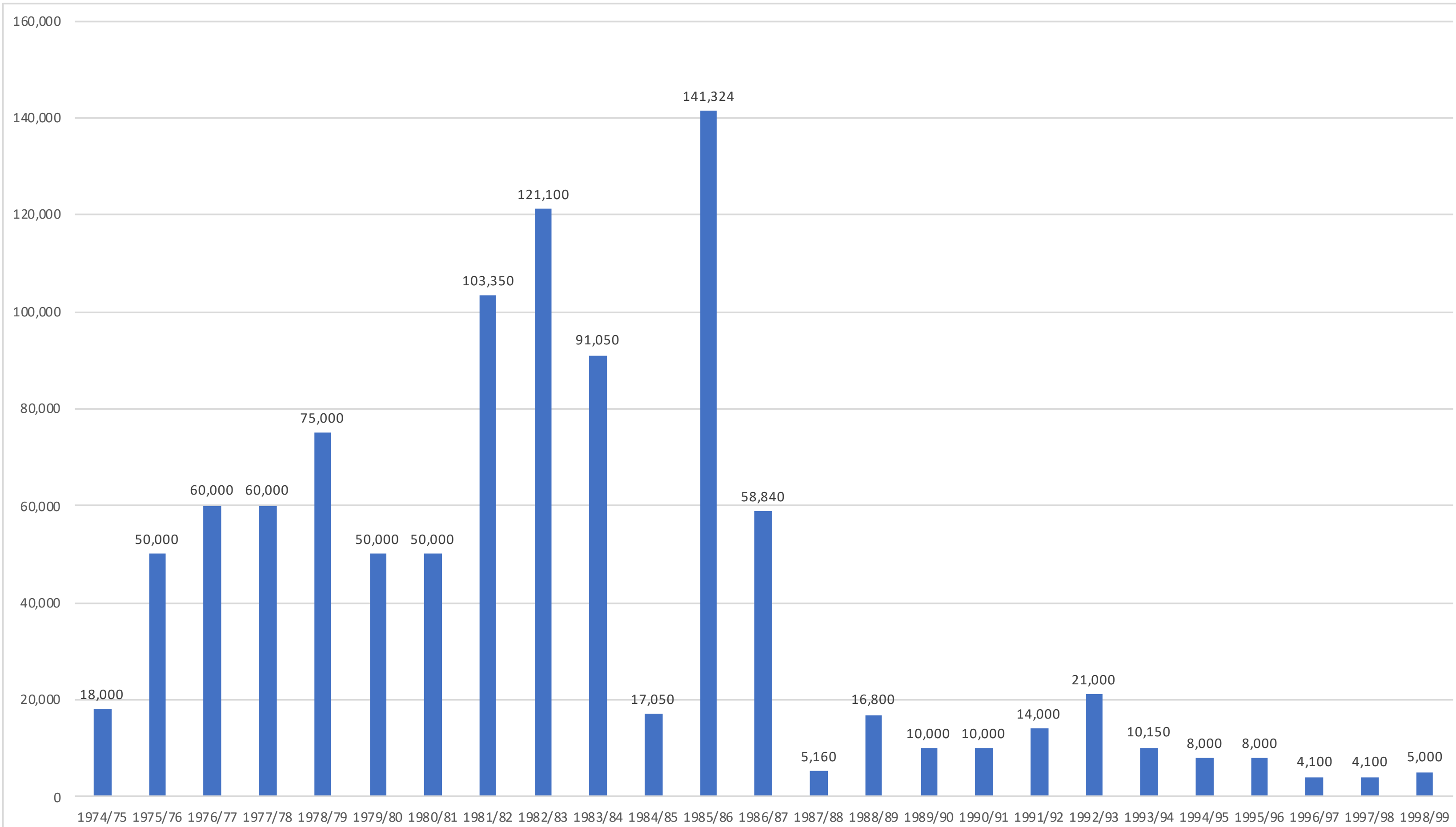
Notes: This table repeats the exercise as in Table 4 using the census data of 2000 as a robustness check. Standard errors are clustered at the birth place district level. The sample consists of individuals born between either 1957 and 1962 or 1968 and 1972. Post refers to the treated cohort, born between 1968 and 1972, while the untreated cohort was born between 1957 and 1962. Educational attainment data are taken from the Indonesian 2000 Census and years of schooling are inputed by the author. Intensity is the number of schools built in a district per 1,000 kids in the school-aged population. All columns include district fixed effect, birth year fixed effect, birth year interacted with number of children at 1971. Duflo Controls consist of birth year dummy interacted with number of children in 1971, with enrollment rate at 1971 and with water sanitization program. Standard errors are clustered at the birth place district level.
Significance levels: * 10%, ** 5%, *** 1%.
Source: Indonesian Census 2000

Table A3: Heterogeneity Results: Effect of School Construction on Education: 2000 Census

	Men				Women			
		Indicator for Completing at least:				Indicator for Completing at least:		
	Years of schooling	Primary School	Junior High	Senior High	Years of schooling	Primary School	Junior High	Senior High
Panel A: Density < Medium:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post × Intensity	0.074*	0.0068	0.0088	0.0044	0.034	0.0064	0.00035	-0.00082
	(0.041)	(0.0045)	(0.0058)	(0.0046)	(0.038)	(0.0045)	(0.0058)	(0.0046)
Dep. var. mean	7.203	0.827	0.417	0.268	6.236	0.765	0.315	0.192
Observations	808,179	808,179	808,179	808,179	795,307	795,307	795,307	795,307
Clusters	184	184	184	184	184	184	184	184
Adjusted R-squared	0.123	0.092	0.114	0.083	0.165	0.130	0.137	0.101
Panel B: Density > Medium:								
Post × Intensity	-0.078*	0.0082	-0.024***	-0.015***	-0.13**	-0.00068	-0.029***	-0.013**
	(0.043)	(0.0063)	(0.0078)	(0.0052)	(0.052)	(0.0070)	(0.0085)	(0.0056)
Dep. var. mean	7.875	0.871	0.475	0.323	6.829	0.812	0.361	0.231
Observations	797,731	797,731	797,731	797,731	784,390	784,390	784,390	784,390
Clusters	90	90	90	90	90	90	90	90
Adjusted R-squared	0.195	0.109	0.174	0.149	0.235	0.147	0.199	0.166

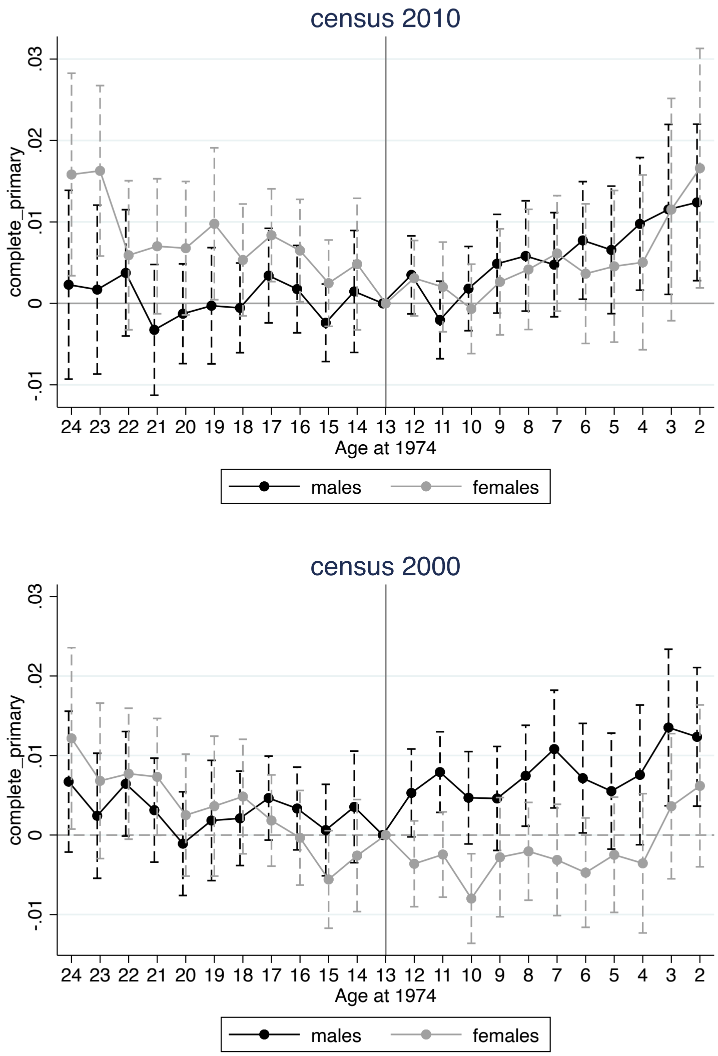
Notes: This table repeats the exercise as in Table 5 using the census of 2000. Standard errors are clustered at the birth place district level. Population density is calculated as the population in 1971 divided by the area of each district in 1971, and the median density is defined as the density for the district of birth for the median person in the sample, which is 497 inhabitants per square kilometer.
Significance levels: * 10%, ** 5%, *** 1%.
Source: Indonesian Census 2000

Figure A.1: Number of newly appointed primary school teachers,1974-1998



Original source:Ministry of National Education, Indonesia, 2005
Source:Jalal et al. (2009), Figure 1.1

Figure A.2: Coefficients of the Interactions: Birth cohort dummy * Program Intensity in Primary School Attainment Regression using data from Census 2010 (top) and Census 2000 (bottom)



Note: This figure shows the event-study graph for the estimate in primary school attainment regression for men and women separately using Indonesia 2010 census (top) and 2000 census (bottom) to explore the concern of mortality selection for older cohorts. Controls include district fixed effect, birth year fixed effect, birth year dummy interacted with number of children at 1971, with enrollment rate at 1971 and with water sanitization program. Standard errors are clustered at the birth place district level. 95% confidence intervals are shown using dashed line.

For Online Publication:

Appendix

A Marriage market in any one period

I will first discuss how marriage market unfolds given women's marriage timing choices in any given period.

Individual types

Women can choose to participate in one of the two periods, hence in any period, there are at most four types of women: Low education and Young (L_1), Low education and Old (L_2), High education and Young (H_1), and High education and Old (H_2). Men only participate in period 2, hence there are two types of men in any period: Low education (L) and High education (H).

Utilities and matching surplus

Denote x as the type of women and X as the type set, i.e. $x \in X = \{L_1, L_2, H_1, H_2\}$. Similarly, denote y as the type of men and Y as the type set, i.e. $y \in Y = \{L, H\}$. To include the possibility of being single, denote $X_0 = X \cup \emptyset$, $Y_0 = Y \cup \emptyset$. Suppose that a woman i with type x and a man j with type y form a couple. I assume their lifetime utilities are as following:

$$\text{woman i's utility: } u_{ij} = \alpha_{xy} + \tau_{ij} + \varepsilon_{iy}$$

$$\text{man j's utility: } v_{ij} = \gamma_{xy} - \tau_{ij} + \eta_{xj}$$

α_{xy}, γ_{xy} indicate the systematic part of the utility each individual gets from the marriage depending on their types. τ_{ij} represents the transfer between i and j, which is going to be determined in equilibrium. [13] $\varepsilon_{iy}, \eta_{xj}$ represent the individuals' idiosyncratic tastes in partner types. Notice they only depend on the partners' types.

[13] τ can be either positive or negative.

For individual singles, their utilities will be:

$$u_{i\emptyset} = \alpha_{x\emptyset} + \varepsilon_{i\emptyset}$$

$$v_{\emptyset j} = \gamma_{\emptyset y} + \eta_{\emptyset j}$$

Without loss of generality, we can normalize $\alpha_{x\emptyset} = 0$ and $\gamma_{\emptyset j} = 0$. [14] Then α_{xy} and γ_{xy} can be interpreted as the net systematic gain from marriage.

There are three important assumptions underlying my specification of individuals'utility: [15]

- There exists a transfer technology among a couple to transfer their utilities one to one without loss, which is the basic feature of a matching model with TU.
- Both transfer and the random taste terms are additive to the systematic part.
- The random terms are individual specific but only depend on the partner's type.

This utility specification may seem restrictive, but it allows for "matching on unobservables" and allows model tractability. What it rules out is the "chemistry" term between two individuals conditional on their types, i.e., some unobserved preferences of one individual towards some unobserved characteristic of one partner.

Stable Matching

Given the population and type distribution,G_x, G_y in a marriage market, a matching is defined as a measure μ on set $X \times Y$ and a set of payoffs $\{u_i, v_j, i \in I, j \in J\}$ such that

[14] Because we can always define $\tilde{\alpha}_{xy} = \alpha_{xy} - \alpha_{x\emptyset}; \tilde{\gamma}_{xy} = \gamma_{xy} - \gamma_{\emptyset y}$, as the systematic utility surplus an individual obtain from marriage compared to being single.

[15] This is the "Separability" assumption in Galichon and Salanié (2021). As noted in that paper, what matters in the model is the surplus a couple can jointly achieve, i.e. $\alpha_{xy} + \gamma_{xy} + \varepsilon_{iy} + \eta_{xj}$ in our case here. How we attribute this surplus to male preference or female preference doesn't matter. For example, it can be the case that women don't have any random taste for men and their utilities without any transfer is α_{xy}. Men's utilities are $\gamma_{xy} + \varepsilon_{iy} + \eta_{xj}$, indicating that man j not only has a random draw η_{xj} depending on women's type, but also has own-type specific random taste for a particular woman i, represented by ε_{iy}. The solution to the model is the same no matter how we interpret the joint surplus into people's preference. The same assumption is also imposed in Choo and Siow (2006) and Chiappori et al. (2017).

$u_i + v_j = \alpha_{xy} + \gamma_{xy} + \varepsilon_{iy} + \eta_{xj}$ for any matched couple (i, j). In other words, a matching specifies who marries with whom and how each mathched couple divides the surplus. Notice that the female type distribution G_x is endogenously determined by female marriage timing choices and the exogenous type distribution, denoted as $E_f = (n_L, n_H)$. And the male type distribution G_y is the same as the exogenous type distribution, denoted as $E_m = (m_L, m_H)$.

In a stable matching, there are two requirements:

- (Individual rationality) Any matched individual is weakly better off than being single.

$$u_i \geq \varepsilon_{i0}, v_j \geq \eta_{0j}, \forall i \in I, j \in J$$

- (No blocking pair) There doesn't exist any two individuals, woman i and man j, who are currently not matched to each other but would both rather match to each other compared with their current condition.

$$u_i + v_j \geq \alpha_{xy} + \gamma_{xy} + \varepsilon_{iy} + \eta_{xj}, \forall i \in I, j \in J$$

Therefore, in any stable matching and given equilibrium transfers τ_{ij}, the following conditions hold true:

$$\text{Woman i chooses } j^*(i) : j^*(i) = \max_{j \in J_0} u_{ij}$$

$$\text{Man j chooses } i^*(j) : i^*(j) = \max_{i \in I_0} v_{ij}$$

where J_0 represent all men and the possibility of being single, I_0 represent all women and the possibility of being single.

Lemma 1. *For any stable matching, there exists two vectors U^{xy} and V^{xy} such that:*

(i) Woman i of type x achieves utility:

$$\tilde{u}_i = \max_{y \in Y_0} (U^{xy} + \varepsilon_{iy})$$

and she matches some man whose type y achieves the maximum;

(ii) Man j of type y achieves utility:

$$\tilde{v_j} = \max_{x \in X_0}(V^{xy} + \eta_{xj})$$

and he matches some woman whose type x achieves the maximum.

(iii) If there exist women of type x matched with men of type y at equilibrium, then

$$U^{xy} + V^{xy} = \alpha_{xy} + \gamma_{xy}$$

This lemma has been proved in Chiappori et al. (2017); Galichon and Salanié (2021). I'll write a short version of the proof in the appendix. With TU, the additive structure and type-specific heterogeneity, this two-sided matching problem is simplified to a one-sided discrete choice problem.

<u>Solutions with Gumbel distribution</u>

If we further assume Gumbel distribution for ε, η, a closed form solution of the stable matching and the expected utilities of each type can be derived. From now on, let's assume the random terms $\varepsilon_{iy}, \eta_{xj}$ follow independent Gumbel distributions $G(-k, 1)$, with $k \simeq 0.5772$ being the Euler constant. With the properties of the Gumbel distribution and Lemma 1, for a given woman i of type x,

$$\begin{aligned}\mu_{y|x} &:= \text{Pr (Woman i (of type x) matched with a man of type y)} \\ &= \frac{exp(U^{xy})}{1 + \sum_{y \in Y} exp(U^{xy})}\end{aligned}$$

$$\begin{aligned}\mu_{\emptyset|x} &:= \text{Pr (Woman i (of type x) is single)} \\ &= \frac{1}{1 + \sum_{y \in Y} exp(U^{xy})}\end{aligned}$$

Therefore,

$$\frac{\mu_{y|x}}{\mu_{\emptyset|x}} = exp(U^{xy}), \forall x \in X$$

Similar logic applies to the other side: men:

$$\frac{\mu_{x|y}}{\mu_{\emptyset|y}} = exp(V^{xy}), \forall y \in Y$$

Denote n_x, m_y as the population of each type. Note that n_x depends on women's participation choices. Denote μ_{xy} as the mass of matched couples between woman of type x and man type y, note that $\mu_{xy} = \mu_{yx}$ by construction since it is a one-to-one match; denote μ_{x0} as the mass of single women of type x, μ_{0y} as the mass of single men of type y; then we have:

$$\frac{\mu_{xy}^2}{\mu_{x0}\mu_{0y}} = exp(U^{xy} + V^{xy}) = exp(\alpha_{xy} + \gamma_{xy})$$

Denote $\Phi_{xy} = \alpha_{xy} + \gamma_{xy}$. Then given Φ_{xy}, the previous equation provides a matching function between the mass of any couple type and the probabilities of singlehood. With the following feasibility constraints, we can construct a system of equations with $|X| + |Y|$ unknowns (probabilities of singlehood for each type) and $|X| + |Y|$ equations. Decker et al. (2013) shows the existence and uniqueness of the solution to this system.

$$\mu_{x0} + \mu_{xL} + \mu_{xH} = n_x, \forall x \in \{L_1, L_2, H_1, H_2\}$$

$$\mu_{0y} + \mu_{L_1y} + \mu_{L2y} + \mu_{H_1y} + \mu_{H_2y} = m_y, \forall y \in \{L, H\}$$

Moreover, we can recover the expected utilities each type gets from participating in this marriage market. With the properties of Gumbel distributions,

$$u_x := E[\tilde{u_i}] = E[\max_{y \in Y_0}(U^{xy} + \varepsilon_{iy})] = ln(1 + \sum_{y \in Y} exp(U^{xy})) = -ln(\mu_{\emptyset|x}) = -ln(\frac{\mu_{x0}}{n_x})$$

$$v_y := E[\tilde{v_j}] = E[\max_{x \in X_0}(V^{xy} + \eta_{xj})] ln(1 + \sum_{y \in Y} exp(U^{xy})) = -ln(\mu_{\emptyset|y}) = -ln(\frac{\mu_{0y}}{m_y})$$

In this case, the expected utility has one-to-one correspondence with the single rate in this case. The smaller the single rate is, the larger the expected utility is. [16]

[16] This is a specific property of the Gumbel distribution.

B Proof for Lemma 1

Proof. Denote $\tilde{u_i}, \tilde{v_j}$ the equilibrium utility individuals get. We know that if woman i and man j match in equilibrium, then $\tilde{u_i} + \tilde{v_j} = \alpha_{xy} + \gamma_{xy} + \varepsilon_{iy} + \eta_{xj}$.

For woman i of type x,

$$\begin{aligned}\tilde{u_i} &= \max_{j \in J} \{\alpha_{xy} + \gamma_{xy} + \varepsilon_{iy} + \eta_{xj} - \tilde{v_j}, \varepsilon_{i0}\} \\ &= \max_{y \in Y} \left\{ \max_{j \text{ where } j_i = y} (\alpha_{xy} + \gamma_{xy} + \eta_{xj} - \tilde{v_j}) + \varepsilon_{iy}, \varepsilon_{i0} \right\}\end{aligned}$$

Define $U^{xy} = \max_{j \text{ where } j_i = y}(\alpha_{xy} + \gamma_{xy} + \eta_{xj} - \tilde{v_j}), U^{x0} = 0$, then we get:

$$\tilde{u_i} = \max_{y \in Y_0}(U^{xy} + \varepsilon_{iy})$$

Moreover,

$$\tilde{u_i} \geq U^{xy} + \varepsilon_{iy}, \forall y \in Y_0$$

and it achieves equality when the set of women of type x matched with men of type y is nonempty.

With similar notations, define $V^{xy} = \max_{i \text{ where } x_i = x}(\alpha_{xy} + \gamma_{xy} + \varepsilon_{iy} - \tilde{u_i}), V^{0y} = 0$, then:

$$\tilde{v_j} = \max_{x \in X_0}(V^{xy} + \eta_{xj})$$

$$\tilde{v_j} \geq V^{xy} + \eta_{xj}, \forall x \in X_0$$

and it achieves equality when the set of men of type y matched with women of type x is nonempty.

If there exist women of type x matched with men of type y,

$$\tilde{u_i} = U^{xy} + \varepsilon_{iy}$$

$$\tilde{v_j} = V^{xy} + \eta_{xj}$$

Hence $U^{xy} + V^{xy} = \alpha_{xy} + \gamma_{xy}$ $\square$

C An important lemma

To prove the propositions, I'll first establish an important lemma related to how probabilities of singlehood change related to the shift of marginals in types.

Lemma 2. *Assume the idiosyncratic tastes follow Gumbel distributions. Assume there are two types for each side, denote the female marginal as $n = (x, 1-x)$ and male marginal as $m = (y, 1-y)$, the surplus matrix as:*

$$\Phi = \begin{bmatrix} \Phi_{LL} & \Phi_{LH} \\ \Phi_{HL} & \Phi_{HH} \end{bmatrix}$$

denote the mass of singles of females (males) in equilibrium as: $\mu_{L0}, \mu_{H0}(\mu_{0L}, \mu_{0H})$ then:

(a)

$$\frac{\partial \mu_{L0}}{\partial x} > 0, \frac{\partial \mu_{H0}}{\partial x} < 0$$

(b) If the marital surplus function is super-modular, i.e., $\Phi_{LL} + \Phi_{HH} > \Phi_{LH} + \Phi_{HL}$, then

(b1)

$$\frac{\partial \mu_{0L}}{\partial x} > 0 \Rightarrow \frac{\partial \mu_{0H}}{\partial x} > 0$$

$$\frac{\partial \mu_{0H}}{\partial x} < 0 \Rightarrow \frac{\partial \mu_{0L}}{\partial x} < 0$$

(b2) There exists some $\underline{\delta_x}, \bar{\delta_x}, \underline{\delta_y}, \bar{\delta_y}$, such that if $\underline{\delta_x} < x < \bar{\delta_x}, \underline{\delta_y} < y < \bar{\delta_y}$, then:

$$\frac{\partial \frac{\mu_{0L}}{\mu_{0H}}}{\partial x} < 0$$

Proof. Denote $a = \exp(\frac{\Phi_{LL}}{2}), b = \exp(\frac{\Phi_{LH}}{2}), c = \exp(\frac{\Phi_{HL}}{2}), d = \exp(\frac{\Phi_{HH}}{2})$;

denote $s_{L0} = \sqrt{\mu_{L0}}, s_{H0} = \sqrt{\mu_{H0}}, s_{0L} = \sqrt{\mu_{0L}}, s_{0H} = \sqrt{\mu_{0H}}$;

denote $D_{L0} = \frac{\partial s_{L0}}{\partial x}, D_{H0} = \frac{\partial s_{H0}}{\partial x}, D_{0L} = \frac{\partial s_{0L}}{\partial x}, D_{0H} = \frac{\partial s_{0H}}{\partial x}$.

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

$$s_{L0}^2 + s_{L0}s_{0L}a + s_{L0}s_{0H}b = x$$

$$s_{H0}^2 + s_{H0}s_{0L}c + s_{H0}s_{0H}d = 1 - x$$

$$s_{0L}^2 + s_{L0}s_{0L}a + s_{H0}s_{0L}c = y$$

$$s_{0H}^2 + s_{L0}s_{0H}b + s_{H0}s_{0H}d = 1 - y$$

In the four equations above, taking the derivative with respect to x, we get:

$$(2s_{L0} + as_{0L} + bs_{0H})D_{L0} + s_{L0}(aD_{0L} + bD_{0H}) = 1 \tag{4}$$

$$(2s_{H0} + cs_{0L} + ds_{0H})D_{H0} + s_{H0}(cD_{0L} + dD_{0H}) = -1 \tag{5}$$

$$(2s_{0L} + as_{L0} + cs_{H0})D_{0L} + s_{0L}(aD_{L0} + cD_{H0}) = 0 \tag{6}$$

$$(2s_{0H} + bs_{L0} + ds_{H0})D_{0H} + s_{0H}(bD_{L0} + dD_{H0}) = 0 \tag{7}$$

Hence we can express D_{0L}, D_{0H} using D_{L0}, D_{H0} from Equation 6 and Equation 7:

$$D_{0L} = -\frac{s_{0L}(aD_{L0} + cD_{H0})}{2s_{0L} + as_{L0} + cs_{H0}} \tag{8}$$

$$D_{0H} = -\frac{s_{0H}(bD_{L0} + dD_{H0})}{2s_{0H} + bs_{L0} + ds_{H0}} \tag{9}$$

Plugging in Equation 4 and Equation 5, we get:

$$(2s_{L0}+\frac{as_{0L}(2s_{0L} + cs_{H0})}{2s_{0L} + as_{L0} + cs_{H0}}+\frac{bs_{0H}(2s_{0H} + ds_{H0})}{2s_{0H} + bs_{L0} + ds_{H0}})D_{L0}-(\frac{acs_{L0}s_{0L}}{2s_{0L} + as_{L0} + cs_{H0}}+\frac{bds_{L0}s_{0H}}{2s_{0H} + bs_{L0} + ds_{H0}})D_{H0} = 1 \tag{10}$$

$$(2s_{H0}+\frac{cs_{0L}(2s_{0L} + as_{L0})}{2s_{0L} + as_{L0} + cs_{H0}}+\frac{ds_{0H}(2s_{0H} + bs_{L0})}{2s_{0H} + bs_{L0} + ds_{H0}})D_{H0}-(\frac{acs_{H0}s_{0L}}{2s_{0L} + as_{L0} + cs_{H0}}+\frac{bds_{H0}s_{0H}}{2s_{0H} + bs_{L0} + ds_{H0}})D_{L0} = -1 \tag{11}$$

Add Equation 10 and Equation 11, we get:

$$(2s_{L0}+\frac{2as_{0L}^2}{2s_{0L}+as_{L0}+cs_{H0}}+\frac{2bs_{0H}^2}{2s_{0H}+bs_{L0}+ds_{H0}})D_{L0}+(2s_{H0}+\frac{2cs_{0L}^2}{2s_{0L}+as_{L0}+cs_{H0}}+\frac{2ds_{0H}^2}{2s_{0H}+bs_{L0}+ds_{H0}})D_{H0} \tag{12}$$

Hence D_{L0} and D_{H0} have opposite signs. With Equation 10, we know:

$$D_{L0} > 0, D_{H0} < 0$$

This completes the proof for (a).

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

$$a * d > b * c$$

Since $D_{L0} > 0$:

$$\frac{a}{c}D_{L0} > \frac{b}{d}D_{L0}$$

$$\Rightarrow: \frac{a}{c}D_{L0} + D_{H0} > \frac{b}{d}D_{L0} + D_{H0}$$

Hence:

$$aD_{L0} + cD_{H0} < 0 \quad \Rightarrow \quad bD_{L0} + dD_{H0} < 0$$

$$bD_{L0} + dD_{H0} > 0 \quad \Rightarrow \quad aD_{L0} + cD_{H0} > 0$$

Recall Equation 8 and Equation 9, we have:

$$\frac{\partial \mu_{0L}}{\partial x} > 0 \Rightarrow \frac{\partial \mu_{0H}}{\partial x} > 0$$

$$\frac{\partial \mu_{0H}}{\partial x} < 0 \Rightarrow \frac{\partial \mu_{0L}}{\partial x} < 0$$

Proof for (b1) is complete.

Now let's prove part (b2):

$$\frac{\partial \frac{s_{0L}}{s_{0H}}}{\partial x} = \frac{D_{0L}s_{0H} - D_{0H}s_{0L}}{s_{0H}^2}$$

Using Equation 8 and Equation 9,

$$D_{0L}s_{0H} - D_{0H}s_{0L} = -\frac{s_{0L}s_{0H}(aD_{L0} + cD_{H0})}{2s_{0L} + as_{L0} + cs_{H0}} + \frac{s_{0L}s_{0H}(bD_{L0} + dD_{H0})}{2s_{0H} + bs_{L0} + ds_{H0}}$$
$$= \frac{s_{0L}s_{0H}([b(2s_{0L} + cs_{H0}) - a(2s_{0H} + ds_{H0})]D_{L0} + [d(2s_{0L} + as_{L0}) - c(2s_{0H} + bs_{L0})]D_{H0})}{(2s_{0L} + as_{L0} + cs_{H0})(2s_{0H} + bs_{L0} + ds_{H0})}$$

It has the same sign as:

$$[2bs_{0L} - 2as_{0H} + (bc - ad)s_{H0}]D_{L0} + [2ds_{0L} - 2cs_{0H} + (ad - bc)s_{L0}]D_{H0}$$
$$= 2s_{0L}(bD_{L0} + dD_{H0}) - 2s_{0H}(aD_{L0} + cD_{H0}) - (ad - bc)(D_{L0}s_{H0} - D_{H0}s_{L0})$$

We know that $(ad - bc)(D_{L0}s_{H0} - D_{H0}s_{L0}) > 0$, since $ad - bc > 0, D_{L0} > 0, D_{H0} < 0$.

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

(Case 1): $aD_{L0} + cD_{H0} > 0, bD_{L0} + dD_{H0} < 0$; it is straightforward to show:

$$\frac{\partial \frac{s_{0L}}{s_{0H}}}{\partial x} < 0$$

(Case 2): $aD_{L0} + cD_{H0} > 0, bD_{L0} + dD_{H0} > 0$

in this case, from Equation 12, we know $s_{L0}D_{L0} + s_{H0}D_{H0} < 0$, hence:

$$\frac{a}{c} > \frac{b}{d} > \frac{s_{L0}}{s_{H0}}$$

Since we know $\frac{s_{L0}}{s_{H0}}$ increases with x, to satisfy previous inequality, we know that x is also relatively small in this case.

There exists some $\underline{\delta_x}, \bar{\delta_y}$ such that for $x > \underline{\delta_x}, y < \bar{\delta_y}$,

$$\frac{\partial \frac{s_{0L}}{s_{0H}}}{\partial x} < 0$$

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

(Case 3): $aD_{L0} + cD_{H0} < 0, bD_{L0} + dD_{H0} < 0$

in this case, from equation (9), we know $s_{L0}D_{L0} + s_{H0}D_{H0} > 0$, hence:

$$\frac{s_{L0}}{s_{H0}} > \frac{a}{c} > \frac{b}{d}$$

x is relatively large in this case. There exists some $\bar{\delta_x}, \underline{\delta_y}$ such that for $x < \bar{\delta_x}, y > \underline{\delta_y}$,

$$\frac{\partial \frac{s_{0L}}{s_{0H}}}{\partial x} < 0$$

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

Proof for part (b2) is complete. □

Lemma 3. *An extension of Lemma 2:*

Suppose there are two types on one side, and there are $K > 2$ types on the other side, denote the marginals as $n = (x_1, x_2, ..., x_K), m = (y, 1 - y)$, where $\sum_k x_k = r$, where r is a constant. The surplus matrix is:

$$\Phi = \begin{bmatrix} \Phi_{11} & \Phi_{12} \\ ... & ... \\ \Phi_{K1} & \Phi_{K2} \end{bmatrix}$$

denote the mass of singles in equilibrium as: $\mu_{k0}, \mu_{01}, \mu_{02}$ then:

(a)

$$\frac{\partial \mu_{01}}{\partial y} > 0, \frac{\partial \mu_{02}}{\partial y} < 0$$

(b) For any two types k_1, k_2, if we increase k_1 by decreasing k_2, then $\mu_{x_{k1}}$ increases and $\mu_{x_{k2}}$ decreases.

(c) For any two types k_1, k_2, if $\Phi_{k_1}+\Phi_{k_2 2} > \Phi_{k_2 1}+\Phi_{k_1 2}$, then there exist values $\underline{\delta_{x_1}}, \bar{\delta_{x_1}}, \underline{\delta_{x_2}}, \bar{\delta_{x_2}}, \underline{\delta_y}, \bar{\delta_y}$:

$$x_{k_1} \in (\underline{\delta_{x_1}}, \bar{\delta_{x_1}})$$

$$x_{k_2} \in (\underline{\delta_{x_2}}, \bar{\delta_{x_2}})$$

$$y \in (\underline{\delta_y}, \bar{\delta_y})$$

such that: $\frac{\mu_{01}}{\mu_{02}}$ decreases if we shift some mass from type k_2 to type k_1, i.e.:

$$\frac{\mu_{01}}{\mu_{02}}|_{(n=(\dots,x_{k_1}+\Delta,x_{k_2}-\Delta,\dots),m)} < \frac{\mu_{02}}{\mu_{01}}|_{(n=(\dots,x_{k_1},x_{k_2},\dots),m)}, \forall \Delta > 0$$

Proof. The proof is very similar to the proof of Lemma 2. WLOG, assume we shift the mass from type 2 to type 1 and denote $x_1 = x, x_2 = \gamma - x$, then $n = (x, \gamma - x, x_3, ..., x_k)$, and $m = (y, 1 - y)$. Denote $s_{i0} = \sqrt{\mu_{i0}}, s_{0j} = \sqrt{\mu_{0j}}$. First, write down the feasibility conditions:

$$\begin{aligned}
s_{10}^2 + s_{10}s_{01}\phi_1 + s_{10}s_{02}\widetilde{\phi}_1 &= x \\
s_{20}^2 + s_{20}s_{01}\phi_2 + s_{20}s_{02}\widetilde{\phi}_2 &= \gamma - x \\
&\vdots \\
s_{K0}^2 + s_{K0}s_{01}\phi_K + s_{K0}s_{02}\widetilde{\phi}_K &= x_K \\
s_{01}^2 + s_{10}s_{01}\phi_1 + s_{20}s_{01}\phi_2 + ... + s_{K0}s_{01}\phi_K &= y \\
s_{02}^2 + s_{10}s_{02}\widetilde{\phi_1} + s_{20}s_{02}\widetilde{\phi_2} + ... + s_{K0}s_{02}\widetilde{\phi}_K &= 1 - y
\end{aligned}$$

To prove part (a), let's take the derivative with respect to y for all $K + 2$ equations and denote $D_{i0} = \frac{\partial s_{i0}}{\partial y}, D_{0j} = \frac{\partial s_{0j}}{\partial y}$.

$$D_{10}(2s_{10} + \phi_1 s_{01} + \widetilde{\phi}_1 s_{02}) + s_{10}(\phi_1 D_{01} + \widetilde{\phi}_1 D_{02}) = 0 \quad (13)$$

$$D_{20}(2s_{20} + \phi_2 s_{01} + \widetilde{\phi}_2 s_{02}) + s_{20}(\phi_2 D_{01} + \widetilde{\phi}_2 D_{02}) = 0 \quad (14)$$

$$\vdots$$

$$D_{K0}(2s_{K0} + \phi_K s_{01} + \widetilde{\phi}_K s_{02}) + s_{K0}(\phi_K D_{01} + \widetilde{\phi}_K D_{02}) = 0 \quad (15)$$

$$D_{01}(2s_{01} + \phi_1 s_{10} + \phi_2 s_{20} + \cdots + \phi_K s_{K0}) + s_{01}(\phi_1 D_{10} + \phi_2 D_{20} + \cdots + \phi_K D_{K0}) = 1 \quad (16)$$

$$D_{02}(2s_{02} + \widetilde{\phi}_1 s_{10} + \widetilde{\phi}_2 s_{20} + \cdots + \widetilde{\phi}_K s_{K0}) + s_{02}(\widetilde{\phi}_1 D_{10} + \widetilde{\phi}_2 D_{20} + \cdots + \widetilde{\phi}_K D_{K0}) = -1 \quad (17)$$

We can rearrange Equation 13 - Equation 15 to express D_{k0} as a function of D_{01}, D_{02}

$$D_{k0} = -\frac{s_{k0}(\phi_k D_{01} + \widetilde{\phi}_k D_{02})}{2s_{k0} + \phi_k s_{01} + \widetilde{\phi}_k s_{02}}, \forall k = 1, 2, ..., K \quad (18)$$

We can substitute Equation 18 to Equation 16 and Equation 17:

$$D_{01}(2s_{01} + \sum_{k=1}^{K} \frac{\phi_k s_{k0}(2s_{k0} + \widetilde{\phi}_k s_{02})}{2s_{k0} + \phi_k s_{01} + \widetilde{\phi}_k s_{02}}) - D_{02} \sum_{k=1}^{K} \frac{s_{01}\phi_k s_{k0}\widetilde{\phi}_k}{2s_{k0} + \phi_k s_{01} + \widetilde{\phi}_k s_{02}} = 1 \quad (19)$$

$$D_{02}(2s_{02} + \sum_{k=1}^{K} \frac{\widetilde{\phi}_k s_{k0}(2s_{k0} + \phi_k s_{01})}{2s_{k0} + \phi_k s_{01} + \widetilde{\phi}_k s_{02}}) - D_{01} \sum_{k=1}^{K} \frac{s_{02}\widetilde{\phi}_k s_{k0}\phi_k}{2s_{k0} + \phi_k s_{01} + \widetilde{\phi}_k s_{02}} = -1 \quad (20)$$

Add Equation 19 and Equation 20,

$$D_{01}(2s_{01} + \sum_{k=1}^{K} \frac{\phi_k 2s_{k0}^2}{2s_{k0} + \phi_k s_{01} + \widetilde{\phi}_k s_{02}})) + D_{02}(2s_{02} + \sum_{k=1}^{K} \frac{\widetilde{\phi}_k 2s_{k0}^2}{2s_{k0} + \phi_k s_{01} + \widetilde{\phi}_k s_{02}}) = 0 \quad (21)$$

Therefore D_{01} and D_{02} should have negative signs. Moreover, with Equation 19, we know:

$$D_{01} > 0, D_{02} < 0$$

Part (a) is proved.

Now let's prove part (b) Let me abuse the use of the notation D_{i0} and D_{0j}. For the proof of part (b), denote $D_{i0} = \frac{\partial s_{i0}}{\partial x}, D_{0j} = \frac{\partial s_{0j}}{\partial x}$. Let's take the derivative with respect to x for all $K+2$ feasibility equations:

$$D_{10}(2s_{10} + \phi_1 s_{01} + \widetilde{\phi}_1 s_{02}) + s_{10}(\phi_1 D_{01} + \widetilde{\phi}_1 D_{02}) = 1 \quad (22)$$

$$D_{20}(2s_{20} + \phi_2 s_{01} + \widetilde{\phi}_2 s_{02}) + s_{20}(\phi_2 D_{01} + \widetilde{\phi}_2 D_{02}) = -1 \quad (23)$$

$$D_{30}(2s_{30} + \phi_3 s_{01} + \widetilde{\phi}_3 s_{02}) + s_{30}(\phi_3 D_{01} + \widetilde{\phi}_3 D_{02}) = 0 \quad (24)$$

$$\vdots$$

$$D_{K0}(2s_{K0} + \phi_K s_{01} + \widetilde{\phi}_K s_{02}) + s_{K0}(\phi_K D_{01} + \widetilde{\phi}_K D_{02}) = 0 \quad (25)$$

$$D_{01}(2s_{01} + \phi_1 s_{10} + \phi_2 s_{20} + \cdots + \phi_K s_{K0}) + s_{01}(\phi_1 D_{10} + \phi_2 D_{20} + \cdots + \phi_K D_{K0}) = 0 \quad (26)$$

$$D_{02}(2s_{02} + \widetilde{\phi}_1 s_{10} + \widetilde{\phi}_2 s_{20} + \cdots + \widetilde{\phi}_K s_{K0}) + s_{02}(\widetilde{\phi}_1 D_{10} + \widetilde{\phi}_2 D_{20} + \cdots + \widetilde{\phi}_K D_{K0}) = 0 \quad (27)$$

Rearrange Equation 24 - Equation 25 to express D_{k0} as a function of D_{01}, D_{02} for $k > 2$:

$$D_{k0} = -\frac{s_{k0}(\phi_k D_{01} + \widetilde{\phi}_k D_{02})}{2s_{k0} + \phi_k s_{01} + \widetilde{\phi}_k s_{02}}, \forall k = 3, ..., K \quad (28)$$

Substitute Equation 28 to Equation 26 and Equation 27:

$$\begin{aligned} D_{01}(2s_{01} + \phi_1 s_{10} + \phi_2 s_{20} + \sum_{k=3}^{K} \frac{\phi_k s_{k0}(2s_{k0} + \widetilde{\phi}_k s_{02})}{2s_{k0} + \phi_k s_{01} + \widetilde{\phi}_k s_{02}}) \\ -D_{02} \sum_{k=3}^{K} \frac{s_{01}\phi_k s_{k0}\widetilde{\phi}_k}{2s_{k0} + \phi_k s_{01} + \widetilde{\phi}_k s_{02}} \\ +s_{01}(\phi_1 D_{10} + \phi_2 D_{20}) = 0 \end{aligned} \quad (29)$$

$$\begin{aligned} D_{02}(2s_{02} + \widetilde{\phi}_1 s_{10} + \widetilde{\phi}_2 s_{20} + \sum_{k=3}^{K} \frac{\widetilde{\phi}_k s_{k0}(2s_{k0} + \phi_k s_{01})}{2s_{k0} + \phi_k s_{01} + \widetilde{\phi}_k s_{02}}) \\ -D_{01} \sum_{k=3}^{K} \frac{s_{02}\widetilde{\phi}_k s_{k0}\phi_k}{2s_{k0} + \phi_k s_{01} + \widetilde{\phi}_k s_{02}} \\ +s_{02}(\widetilde{\phi}_1 D_{10} + \widetilde{\phi}_2 D_{20}) = 0 \end{aligned} \quad (30)$$

Then (Equation 22 + Equation 23)- (Equation 29 + Equation 30) gives us:

$$D_{10}2s_{10} + D_{20}2s_{20} - D_{01}(2s_{01} + \sum_{k=3}^{K} \frac{2\phi_k s_{k0}^2}{2s_{k0} + \phi_k s_{01} + \widetilde{\phi}_k s_{02}}) - D_{02}(2s_{02} + \sum_{k=3}^{K} \frac{2\widetilde{\phi}_k s_{k0}^2}{2s_{k0} + \phi_k s_{01} + \widetilde{\phi}_k s_{02}}) = 0 \tag{31}$$

Moreover, from Equation 29 and Equation 30, we can express D_{01} and D_{02} as a linear combination of D_{10} and D_{20}. Denote We can also show that the coefficents are all negative. Combing Equation 31, D_{10} and D_{20} should have negative signs. Therefore $D_{10} > 0, D_{20} < 0$. Part (b) is proved.

Now let's prove part (c). Let's follow the notation of the proof for part (b): $D_{i0} = \frac{\partial s_{i0}}{\partial x}, D_{0j} = \frac{\partial s_{0j}}{\partial x}$. Rearrange Equation 22 - Equation 25 to express D_{k0} as a function of D_{01}, D_{02}:

$$D_{10} = \frac{1 - s_{10}(\phi_1 D_{01} + \widetilde{\phi}_1 D_{02})}{2s_{10} + \phi_1 s_{01} + \widetilde{\phi}_1 s_{02}} \tag{32}$$

$$D_{20} = \frac{-1 - s_{20}(\phi_2 D_{01} + \widetilde{\phi}_2 D_{02})}{2s_{20} + \phi_2 s_{01} + \widetilde{\phi}_2 s_{02}} \tag{33}$$

$$D_{k0} = -\frac{s_{k0}(\phi_k D_{01} + \widetilde{\phi}_k D_{02})}{2s_{k0} + \phi_k s_{01} + \widetilde{\phi}_k s_{02}}, \forall k = 3, ..., K \tag{34}$$

Substitute Equation 32 - Equation 34 to Equation 26 and Equation 27:

$$\begin{aligned} D_{01}(2s_{01} + \sum_{k=1}^{K} \frac{\phi_k s_{k0}(2s_{k0} + \widetilde{\phi}_k s_{02})}{2s_{k0} + \phi_k s_{01} + \widetilde{\phi}_k s_{02}}) - D_{02} \sum_{k=1}^{K} \frac{s_{01}\phi_k s_{k0}\widetilde{\phi}_k}{2s_{k0} + \phi_k s_{01} + \widetilde{\phi}_k s_{02}} \\ = s_{01}(\frac{\phi_2}{2s_{20} + \phi_2 s_{01} + \widetilde{\phi}_2 s_{02}} - \frac{\phi_1}{2s_{10} + \phi_1 s_{01} + \widetilde{\phi}_1 s_{02}}) \end{aligned} \tag{35}$$

$$\begin{aligned} D_{02}(2s_{02} + \sum_{k=1}^{K} \frac{\widetilde{\phi}_k s_{k0}(2s_{k0} + \phi_k s_{01})}{2s_{k0} + \phi_k s_{01} + \widetilde{\phi}_k s_{02}}) - D_{01} \sum_{k=1}^{K} \frac{s_{02}\widetilde{\phi}_k s_{k0}\phi_k}{2s_{k0} + \phi_k s_{01} + \widetilde{\phi}_k s_{02}} \\ = s_{02}(\frac{\widetilde{\phi}_2}{2s_{20} + \phi_2 s_{01} + \widetilde{\phi}_2 s_{02}} - \frac{\widetilde{\phi}_1}{2s_{10} + \phi_1 s_{01} + \widetilde{\phi}_1 s_{02}}) \end{aligned} \tag{36}$$

Denote:

$$A = 2\frac{s_{01}}{s_{02}} + \sum_{k=1}^{K} \frac{\phi_k s_{k0} 2^{\frac{s_{k0}}{s_{02}}}}{2s_{k0} + \phi_k s_{01} + \widetilde{\phi}_k s_{02}}$$

$$B = \sum_{k=1}^{K} \frac{\widetilde{\phi}_k s_{k0} \phi_k}{2s_{k0} + \phi_k s_{01} + \widetilde{\phi}_k s_{02}}$$

$$C = 2\frac{s_{02}}{s_{01}} + \sum_{k=1}^{K} \frac{\widetilde{\phi}_k s_{k0} 2^{\frac{s_{k0}}{s_{01}}}}{2s_{k0} + \phi_k s_{01} + \widetilde{\phi}_k s_{02}}$$

$$F = s_{01}\left(\frac{\phi_2}{2s_{20} + \phi_2 s_{01} + \widetilde{\phi}_2 s_{02}} - \frac{\phi_1}{2s_{10} + \phi_1 s_{01} + \widetilde{\phi}_1 s_{02}}\right) \tag{37}$$

$$G = s_{02}\left(\frac{\widetilde{\phi}_2}{2s_{20} + \phi_2 s_{01} + \widetilde{\phi}_2 s_{02}} - \frac{\widetilde{\phi}_1}{2s_{10} + \phi_1 s_{01} + \widetilde{\phi}_1 s_{02}}\right) \tag{38}$$

we know $A > 0, B > 0, C > 0$, moreover:

$$D_{01}s_{02}(A + B) - D_{02}s_{01}B = F \tag{39}$$

$$D_{02}s_{01}(C + B) - D_{01}s_{02}B = G \tag{40}$$

Therefore:

$$D_{01}s_{02} - D_{02}s_{01} < 0 \iff C * F - A * G < 0$$

One sufficient condition for $CF - AG < 0$ is that $F < 0, G > 0$. One sufficient condition for $F < 0, G > 0$ when $\frac{\widetilde{\phi}_2}{\widetilde{\phi}_1} > \frac{\phi_2}{\phi_1}$ is that:

$$\frac{\phi_2}{\phi_1} < \frac{s_{20}}{s_{10}} < \frac{\widetilde{\phi}_2}{\widetilde{\phi}_1}$$

since we can arrange Equation 37 and Equation 38:

$$F = s_{01}\left(\frac{1}{\frac{2s_{20}}{\phi_2} + s_{01} + \frac{\widetilde{\phi}_2}{\phi_2} s_{02}} - \frac{1}{2\frac{s_{10}}{\phi_1} + s_{01} + \frac{\widetilde{\phi}_1}{\phi_1} s_{02}}\right) \tag{41}$$

$$G = s_{02}(\frac{\widetilde{\phi}_2}{2s_{20} + \phi_2 s_{01} + \widetilde{\phi}_2 s_{02}} - \frac{\widetilde{\phi}_1}{2s_{10} + \phi_1 s_{01} + \widetilde{\phi}_1 s_{02}}) \tag{42}$$

Hence there exists $\underline{\delta_{x_1}}, \bar{\delta_{x_1}}, \underline{\delta_{x_2}}, \bar{\delta_{x_2}}$, when

$$x \in (\underline{\delta_{x_1}}, \bar{\delta_{x_1}})$$

$$(\gamma - x) \in (\underline{\delta_{x_2}}, \bar{\delta_{x_2}})$$

we have:

$$\frac{\partial \frac{s_{01}}{s_{02}}}{\partial x} < 0$$

$\square$

D Proof for the Propositions

Proof for Proposition 1

Proof. To prove the existence of a stationary equilibrium, we need to show that there is a solution to the following equilibrium conditions given G_f, G_m, Φ, denote $G_f = (n_L, n_H), G_m = (m_L, m_H)$:

$$\mu_{0y} + \sqrt{\mu_{L_1 0}\mu_{0y}} \exp(\frac{\Phi_{L_1 y}}{2}) + \sqrt{\mu_{L_2 0}\mu_{0y}} \exp(\frac{\Phi_{L_2 y}}{2}) + \sqrt{\mu_{H_1 0}\mu_{0y}} \exp(\frac{\Phi_{H_1 y}}{2}) + \sqrt{\mu_{H_2 0}\mu_{0y}} \exp(\frac{\Phi_{H_2 y}}{2}) = m_y, \forall y \in \{L, H\} \tag{43}$$

$$\mu_{e_1 0} + \sqrt{\mu_{e_1 0}\mu_{0L}} \exp(\frac{\Phi_{e_1 L}}{2}) + \sqrt{\mu_{e_1 0}\mu_{0H}} \exp(\frac{\Phi_{e_1 H}}{2}) = q_e^1 * n_e, \forall e \in \{L, H\} \tag{44}$$

$$\mu_{e_2 0} + \sqrt{\mu_{e_2 0}\mu_{0L}} \exp(\frac{\Phi_{e_2 L}}{2}) + \sqrt{\mu_{e_2 0}\mu_{0H}} \exp(\frac{\Phi_{e_2 H}}{2}) = q_e^2 * n_e, \forall e \in \{L, H\} \tag{45}$$

$$q_e^1 + q_e^2 = 1, \forall e \in \{L, H\} \tag{46}$$

$$\exp(-u_{e_1}) = \frac{\mu_{e_1 0}}{q_e^1 * n_e}, \forall e \in \{L, H\} \tag{47}$$

$$\exp(-u_{e_2}) = \frac{\mu_{e_2 0}}{q_e^2 * n_e}, \forall e \in \{L, H\} \tag{48}$$

$$u_{e_1} = u_{e_2}, \forall e \in \{L, H\} \tag{49}$$

Equation 43-Equation 45 characterize the equilibrium conditions of marriage market stability for given **q** strategy under the assumption of Gumbel distribution. Equation 47-Equation 48 characterize the expected marital utilities of females. Equation 46 comes from the property of stationarity. Equation 49 guarantees that women are indifferent between choosing to marry at period 1 or period 2.

Re-arrange Equation 44 and Equation 45 , we can get:

$$\frac{\mu_{e_1 0}}{q_e^1} + \sqrt{\mu_{0L}}\sqrt{\frac{\mu_{e_1 0}}{q_e^1}}\frac{1}{\sqrt{q_e^1}} \exp(\frac{\Phi_{e_1 L}}{2}) + \sqrt{\mu_{0H}}\sqrt{\frac{\mu_{e_1 0}}{q_e^1}}\frac{1}{\sqrt{q_e^1}} \exp(\frac{\Phi_{e_1 H}}{2}) = n_e$$

$$\frac{\mu_{e_2 0}}{q_e^2} + \sqrt{\mu_{0L}}\sqrt{\frac{\mu_{e_2 0}}{q_e^2}}\frac{1}{\sqrt{q_e^2}}\exp(\frac{\Phi_{e_2 L}}{2}) + \sqrt{\mu_{0H}}\sqrt{\frac{\mu_{e_2 0}}{q_e^2}}\frac{1}{\sqrt{q_e^2}}\exp(\frac{\Phi_{e_2 H}}{2}) = n_e$$

Combining with Equation 47-Equation 49, we can get:

$$\begin{aligned}\sqrt{\frac{q_e^1}{q_e^2}} &= \frac{\sqrt{\mu_{0L}}\exp(\frac{\Phi_{e_1 L}}{2}) + \sqrt{\mu_{0H}}\exp(\frac{\Phi_{e_1 H}}{2})}{\sqrt{\mu_{0L}}\exp(\frac{\Phi_{e_2 L}}{2}) + \sqrt{\mu_{0H}}\exp(\frac{\Phi_{e_2 H}}{2})} \\ &= \exp(\frac{\Phi_{e_1 L} - \Phi_{e_2 L}}{2})\frac{\sqrt{\mu_{0L}} + \sqrt{\mu_{0H}}\exp(\frac{\Phi_{e_1 H} - \Phi_{e_1 L}}{2})}{\sqrt{\mu_{0L}} + \sqrt{\mu_{0H}}\exp(\frac{\Phi_{e_2 H} - \Phi_{e_2 L}}{2})} \\ &= \exp(\frac{\Phi_{e_1 L} - \Phi_{e_2 L}}{2})\frac{1 + \sqrt{\frac{\mu_{0H}}{\mu_{0L}}}\exp(\frac{\Phi_{e_1 H} - \Phi_{e_1 L}}{2})}{1 + \sqrt{\frac{\mu_{0H}}{\mu_{0L}}}\exp(\frac{\Phi_{e_2 H} - \Phi_{e_2 L}}{2})}\end{aligned} \tag{50}$$

There are three cases:

1. $\Phi_{e_1 H} - \Phi_{e_1 L} = \Phi_{e_2 H} - \Phi_{e_2 L}$
2. $\Phi_{e_1 H} - \Phi_{e_1 L} > \Phi_{e_2 H} - \Phi_{e_2 L}$
3. $\Phi_{e_1 H} - \Phi_{e_1 L} < \Phi_{e_2 H} - \Phi_{e_2 L}$

Case one: In the first case, we have:

$$\sqrt{\frac{q_e^1}{q_e^2}} = \exp(\frac{\Phi_{e_1 L} - \Phi_{e_2 L}}{2}) \tag{51}$$

Hence equilibrium strategy $\mathbf{q}$ is pinned down by Equation 51 and Equation 46. Moreover, we know that given $\mathbf{q}$, Equation 43-Equation 45 has a 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, $\frac{q_e^1}{q_e^2}$ is an increasing function of $\frac{\mu_{0H}}{\mu_{0L}}$ in Equation 50. Moreover, according to Lemma 3, we know that when $\Phi_{e_1 H} - \Phi_{e_1 L} > \Phi_{e_2 H} - \Phi_{e_2 L}$ indicating there is a complementarity between male High type and female marrying at period 1, an increase in $\frac{q_e^1}{q_e^2}$ would lead to a decrease in $\frac{\mu_{0H}}{\mu_{0L}}$ from Equation 43-Equation 46.

Moreover, from Equation 50, we know that:

$$\sqrt{\frac{q_e^1}{q_e^2}} \to \exp(\frac{\Phi_{e_1 L} - \Phi_{e_2 L}}{2}), \text{ as } \frac{\mu_{0H}}{\mu_{0L}} \to 0$$

$$\sqrt{\frac{q_e^1}{q_e^2}} \to \exp(\frac{\Phi_{e_1H}-\Phi_{e_2H}}{2}), \text{ as } \frac{\mu_{0H}}{\mu_{0L}} \to +\infty$$

While from Equation 43 - Equation 46, we know $\frac{\mu_{0H}}{\mu_{0L}}$ is bounded by finite positive number when $\exp(\frac{\Phi_{e_1L}-\Phi_{e_2L}}{2}) \leq \sqrt{\frac{q_e^1}{q_e^2}} \leq \exp(\frac{\Phi_{e_1H}-\Phi_{e_2H}}{2})$.

Hence equilibrium exists and is unique.

Case three: In the third case, $\frac{q_e^1}{q_e^2}$ is a decreasing function of $\frac{\mu_{0H}}{\mu_{0L}}$ in Equation 50. Moreover, according to Lemma 3, we know that when $\Phi_{e_1H} - \Phi_{e_1L} < \Phi_{e_2H} - \Phi_{e_2L}$ indicating there is a complementarity between male L type and female marrying at period 1, an increase in $\frac{q_e^1}{q_e^2}$ would lead to an increase in $\frac{\mu_{0H}}{\mu_{0L}}$ from Equation 43 - Equation 46. Applying the same logic as in case two, equilibrium exists and is unique.

Moreover, we know that equilibrium strategy satisfies:

$$\min(\Phi_{L_1L}-\Phi_{L_2L}, \Phi_{L_1H}-\Phi_{L_2}) \leq \ln(\frac{q_L^1}{q_L^2}) \leq \max(\Phi_{L_1L}-\Phi_{L_2L}, \Phi_{L_1H}-\Phi_{L_2H})$$

$$\min(\Phi_{H_1L}-\Phi_{H_2L}, \Phi_{H_1H}-\Phi_{H_2}) \leq \ln(\frac{q_H^1}{q_H^2}) \leq \max(\Phi_{H_1L}-\Phi_{H_2L}, \Phi_{H_1H}-\Phi_{H_2H})$$

□

Proof for Proposition 2

Proof. This is our first case in the previous proof of Proposition 1. Hence from Equation 50 equation (17), we know:

$$\frac{q_e^2}{q_e^1} = \exp(\Phi_{e_2L} - \Phi_{e_1L})$$

with $q_e^1 + q_e^2 = 1$, we have:

$$q_e^2 = \frac{\exp(\Phi_{e_2L})}{\exp(\Phi_{e_2L}) + \exp(\Phi_{e_1L})}, \quad q_e^1 = \frac{\exp(\Phi_{e_1L})}{\exp(\Phi_{e_2L}) + \exp(\Phi_{e_1L})}$$

□

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 50 and Equation 43-Equation 46. Hence how equilibrium strategies change depend on whether $\Phi_{e_1H} - \Phi_{e_2H} > \Phi_{e_1L} - \Phi_{e_2L}$ or $\Phi_{e_1H} - \Phi_{e_2H} < \Phi_{e_1L} - \Phi_{e_2L}$, and how $\frac{\mu_{0H}}{\mu_{0L}}$ changes in equilibrium.

Let's first prove Proposition 3, according to Lemma 3 result (a), an increase in m_H would increase μ_{0H} and decrease μ_{0L}, which increases $\frac{\mu_{0H}}{\mu_{0L}}$ given any strategy q_y, hence an increase in m_H would

- increase q_e^1, if $\Phi_{e_1H} - \Phi_{e_2H} > \Phi_{e_1L} - \Phi_{e_2L}$
- decrease q_e^1, if $\Phi_{e_1H} - \Phi_{e_2H} < \Phi_{e_1L} - \Phi_{e_2L}$

Then let's prove Proposition 4, according to Lemma 3(b), an increase in n_H would decrease $\frac{\mu_{0H}}{\mu_{0L}}$ given any strategy q_y if the following condition holds:

$$\exp(\frac{\Phi_{e_1H} - \Phi_{e_2H}}{2}) \leq \sqrt{\frac{\mu_{e_10}}{\mu_{e_20}}} \leq \exp(\frac{\Phi_{e_1L} - \Phi_{e_2L}}{2})$$

Moreover, we know that:

$$\frac{\mu_{e_10}}{\mu_{e_20}} = \frac{q_e^1}{q_e^2}$$

and

$$\exp(\frac{\Phi_{e_1H} - \Phi_{e_2H}}{2}) \leq \sqrt{\frac{q_e^1}{q_e^2}} \leq \exp(\frac{\Phi_{e_1L} - \Phi_{e_2L}}{2})$$

from Equation 50. Therefore the condition always holds in the neighborhood of the equilibrium. Hence an increase in n_H would

- decrease q_e^1, if $\Phi_{e_1H} - \Phi_{e_2H} > \Phi_{e_1L} - \Phi_{e_2L}$
- increase q_e^1, if $\Phi_{e_1H} - \Phi_{e_2H} < \Phi_{e_1L} - \Phi_{e_2L}$

□