The impact of women's age at marriage on own and spousal labor market outcomes in India: causation or selection?

Dhamija, Gaurav and Roychowdhury, Punarjit

Shiv Nadar University, Indian Institute of Management, Indore

5 May 2018

The impact of women's age at marriage on own and spousal labor market outcomes in India: causation or selection?*

Gaurav Dhamija†

Shiv Nadar University

Punarjit Roychowdhury‡

Indian Institute of Management, Indore

May 5, 2018

Abstract

The labor market impacts of women's age at marriage have recently received significant attention from social scientists. The focus of this literature, however, has been the developed world and almost nothing is known about how a delay in marriage affects labor market prospects of women in developing countries. This paper addresses this gap in the existing literature by providing the first comprehensive assessment of the relationship between women's age at marriage and own as well as spousal labor market outcomes specifically in context of a developing country. Using nationally representative household data from India, we find evidence of positive effects of women's age at marriage on their own and their spouses' labor market outcomes. To examine whether these effects are causal or arise due to selection into marriage, we use an instrumental variables-based empirical strategy that utilizes variation in age at menarche to obtain exogenous variation in women's age at marriage. Our results indicate that the positive effects of age at marriage of women on own as well spousal labor market outcomes are not causal and arise purely due to selection. The results are robust to addressing biases due to nonrandom selection of individuals into labor force. Our findings shed new light on theories of labor market in developing countries specifically through the lens of marriage.

JEL: J12, J16, J22, J31, O12
Keywords: Age at Marriage, India, Instrumental Variables, Labor Market Outcomes, Selection, Women.

---

*The authors are very grateful to Farzana Afridi, Tarun Jain and Daniel Millimet for useful comments and discussions, and to Le Wang for sharing his Stata codes for the test of instrumental validity in selection models.

†Department of Economics, Shiv Nadar University. email: gd343@snu.edu.in. NH91, Tehsil Dadri, Greater Noida, Uttar Pradesh 201314, India.

‡Corresponding Author. Department of Economics, Indian Institute of Management Indore. email: punarjitroychowdhury@gmail.com. Prabandh Shikhar, Rau-Pithampur Road, Indore, Madhya Pradesh 453556, India.

1 Introduction

The labor market impacts of women's age at marriage have recently received considerable attention from economists and demographers. The focus of this literature, however, has almost entirely been the developed world, and more specifically the United States (see for e.g. Loughran and Zissimopoulos 2004; Dahl 2010; Wang and Wang 2017). The present study contributes to this literature by providing the first comprehensive assessment of the relationship between women's age at marriage and own and spousal labor market outcomes specifically in context of a *developing country*. In particular, we examine whether women's age at marriage affects their own as well as their spouses' labor market outcomes in India, and if there exists a relationship, whether that arises due to *selection* into marriage or has a *causal* component.

While over the last few years, a small body of literature has emerged that looks at the impact of women's age at marriage on various socioeconomic outcomes including women's schooling and their children's health and education using data from developing countries like Bangladesh, India and Uganda (see for e.g. Field and Ambrus 2008; Sekhri and Debnath 2014; Sunder 2016; Chari et al. 2017), none of the studies in this literature focus on the labor market impacts of a delay in women's marriage. Our findings, thus, we believe, are likely to shed new light on theories of labor market in emerging economies specifically through the lens of marriage. Additionally, our findings are likely to be useful for understanding the efficacy of existing policies that aim at delaying women's age at marriage in developing countries, as well as for developing more effective ones, given that early marriage of women is an issue of deep concern in these countries.[1]

[1] The mean marital age of women in India was 19.3 years according to the 2011 Census data. Moreover, an article in the *The Wire* (June 1, 2016) states that in India as many as 102 million girls (30% of the female population) were married before 18 in 2011 even though the Prohibition of Child Marriage Act states that a girl in India cannot marry before she turns 18. In Bangladesh 46.1% of women between the ages of 15 and 19 were married between 2003 and 2005. Corresponding figures for some other poor countries for the same period were: 42% in Chad, 32.9% in Malawi, 50.4% in Mali, 38.2% in Mozambique and 31.7% in Nigeria (data from Demographic and Health surveys). For the developed countries, the average age of marriage for women is much higher. For instance, the mean age of marriage of women in the US was 26.9 years in 2011 (Pew Research Foundation, 2011), for Germany it was 30.9 years, and for Sweden 33.3 years.

Several theories are put forth to hypothesize the possible relationship between marital age and labor market outcomes of women. First, early marriage might interrupt the accumulation of formal education and labor market skills for women leading to lower productivity and earnings in the labor market. Marriage comes with family responsibilities that often impede the pursuit of formal education for women. For example, Field and Ambrus (2008) find that an additional year of marriage delay leads to an increase in schooling by 0.22 years for women in Bangladesh. As such, marriage delay would be beneficial in the labor market for women by allowing them to accumulate more formal education and labor market skills. This is referred to as the formal education hypothesis.

Second, fertility is a significant part of a traditional marriage. To the extent that fertility affects a women's labor market outcomes and marriage precedes child-bearing, delayed marriage could lead to delayed fertility and hence better labor market outcomes for women (Wang and Wang 2017). In fact, Lundberg and Rose (2002) note that, fertility can affect a women's wages through two channels. On the one hand, there is the specialization effect due to the increased value of a woman's time relative to that of her spouse's, which in turn would lead women to focus on home production. On the other hand, fertility could increase value of both parents' time at home on child care, which could again have a negative effect on women's labor market earnings.[2]

Third, according to Loughran and Zissimpoulos (2004), quality job matches between employers and job-seekers are generally achieved only after several job changes which often involve migration. Gladden (1999) shows that women tend to sacrifice their career for their spouses (more than men do when making migration decisions), and finds that women's earnings decrease following a move. This theory, thus, suggests that marriage might hamper mobility for working women, and hence it is likely that marital delay will be beneficial for women.

Finally, Bergstrom and Bagnoli (1993) propose a theory based on asymmetric informa-

[2] As Wang and Wang note (2017), this is consistent with the motherhood penalty commonly found in the literature.

tion. According to this theory, information on the earnings capabilities of individuals is available only at later stages of the life cycle. Consequently, parents of girls with high earnings potential or those parents who are more focused about their daughter's career might be more likely to be interested in delaying marriages (perhaps due to potential interruption of accumulation of both labor market skills and formal education from early marriages) until their daughter's earnings potential is fully revealed. This line of reasoning suggests that marital delay is positively associated with earnings for women. Note, however, the positive relationship that is predicted by this theory is due to *self-selection* instead of a *causal* effect.

In addition to having impacts on own labor market outcomes, women's age at marriage can potentially also have spillover effects on their spouses' labor market outcome. For instance, if a marriage delay leads a woman to accumulate more formal education or better labor market skills, this might have a positive impact on her husband's labor market outcomes as well. As argued by Benham (1974), this could be because a wife's education is likely to be a substitute for her husband's own formal education as she offers advice and information, and help her husband acquire specific skills (e.g. coping with change). Additionally, an educated wife, by investing in her husbands' human capital through better physical and mental health (i.e., an educated wife can efficiently monitor her spouse's health and reduce his participation in risky activities), and by contributing to "peripheral tasks" such as entertaining (Jepsen 2005) could also positively impact the labor market outcomes of their husbands. Put more succinctly, "a wife's education may be an input to her husband's productivity, either directly or by allowing him to specialize in market work" (Lefgren and MacIntyre 2006, p. 802-803), which according to Choi et al. (2008), is likely to take the form of an increase work hours and higher earnings for married men.

However, of course, as for the women themselves, the positive effect of women's age at marriage on their husband's labor market outcomes may not necessarily be causal; instead it may arise due to selection. This may be because men of higher ability may want to marry more educated women (Welch 1974; Liu and Zhang 1999; Lefgren and McIntyre 2006; Huang et al. 2009), and hence might prefer marrying older women since they are likely to be more educated.

As noted previously, empirical studies that examine the link between marriage timing and labor market outcomes of women, and specifically attempt to understand whether the relationship is causal or arises due to self selection pertains only to the United States. Loughran and Zissimopoulos (2004) using data from the National Longitudinal Survey of the Youth 1979 (NLSY79) find that delaying marriage causes hourly wages of women to increase by nearly four percent for each year they delay in the US. They employ panel data methods that exploit longitudinal variation in wages and marriage timing. They argue that their results are consistent with the mobility hypothesis. Dahl (2010) employs an instrumental variables based approach to address the endogeneity of the timing of first marriage and examines the impact of early marriage on poverty for US women. He finds a positive impact of early marriage on poverty which suggests a positive relationship between late marriage and earnings for women. Finally, in a recent study Wang and Wang (2017), using 1980 US Census data and employing an identification strategy that is similar to that used by Dahl (2010), find a positive causal impact of marriage delay on wages for women. Further, they provide strong evidence that the positive causal effects are almost exclusively through increased education.[3] Whether the relationship between marriage timing of women and their labor market outcomes holds specifically in a developing country (which is likely to differ in significant ways from a developed country in terms of population characteristics like income, education, marital age, family structure, etc.), and whether this relationship is causal, are questions which have remained unexplored to date. The potential spousal spillover effect of women's age at marriage has also not been looked at in the extant literature.

We use data from the Indian Human Development Survey of 2012 (Desai et al. 2015), and focus on labor market outcomes of women and their spouses including hourly earnings, annual wage earnings, and work days per year. The main empirical challenge in identifying

[3] Wang and Wang (2017) also carry out their analysis for men. They find that men's age at marriage also has a strong positive causal effect on their wages.

the causal effect of age at marriage on labor market outcomes is that marriage age may be endogenous. To address this issue, we employ the empirical strategy proposed by Field and Ambrus (2008), who instrument women's age at marriage by their age at menarche. As noted by Sekhri and Debnath (2017), variation in the age at menarche generates a quasi-random difference in the age at which a girl enters the marriage market. This instrument is motivated by the observation that has been made by sociologists and anthropologists that parents become extremely anxious to get their daughters married once they have reached menarche, partly to avert any unwanted pregnancies (Caldwell et al. 1983; Srinivas 1984).

Our results are interesting. The ordinary least squares (OLS) results for the full sample indicate that a year of delayed marriage of women is associated with significant beneficial effects for their own labor market performance as well as spouses' labor market performance. However, when we use the instrumental variable (IV) approach we find that the age at marriage of women does not have a significant impact on their own as well as their spousal labor market outcomes. Further, we find that the magnitude of the IV estimates are smaller than the OLS estimates. Thus, our results support the positive selection hypothesis and indicate that there does not exist a causal relationship between marriage delay of women and their own as well as their spouses' labor market outcomes in India. This suggests that the labor market premium due to women's marital delay that has been documented in the previous literature is likely to be only a developed country phenomenon and unlikely to be relevant for developing countries.

Since the IV estimates capture only the effect for a subgroup of individuals who are likely impacted by the IV (i.e., the compliers), the comparison of full-sample OLS and IV estimates may not necessarily be a fair evaluation of the selection hypothesis. To address this concern, we reestimate our model for the potential set of compliers. Our instrument will not affect the women who were married before puberty. As such, we assume that the set of compliers consists of women who were married only after attaining menarche. When we restrict our analysis only to the potential compliers, the OLS coefficients are again larger than the IV coefficients indicating positive selection into marriage.

Our results might appear to be somewhat puzzling since it has been documented in the previous literature that a delay in women's marriage leads them to complete more schooling in developing countries (see for e.g. Field and Ambrus 2008; Sunder 2016). In fact, for our analytical sample as well, we find a positive link between marriage delay and women's schooling. As such, at least due to the formal schooling hypothesis, one would have expected to find a causal effect of a women's age at marriage on labor market outcomes. Despite a delay in marriage leading to more schooling for women, why does not it get translated into better labor market outcomes?

We argue that one potential explanation for this is as follows. It has been noted by Kingdon (1998), Kingdon and Unni (2001), and Kanjilal-Bhaduri and Pastore (2017) among many others that in India, labor market returns to education is low and insignificant for women with relatively low education. This could be due to: (1) low quality of primary education in India (Pratham, 2011), and/or (2) for labor market success, a threshold level of education might be necessary (for instance, completing college or vocational degree); below that, an extra year of schooling might not lead to better labor market outcomes. As it turns out, the majority of women in our sample have completed at most primary education. Thus, although women in our sample might complete more formal schooling due to a delay in marriage by a year, this might not be sufficiently productive to get translated into better labor market outcomes since most women in our sample would still belong to the lower end of the education distribution.

The rest of the paper unfolds as follows. In section 2 we discuss the dataset used. Section 3 presents the econometric model and empirical strategy. Results are presented in the section 4. The last section concludes.

2 Data

The data come from the Indian Human Development Survey (IHDS) 2012. IHDS 2012 is a nationally representative multitopic household survey conducted by the National Council for Applied Economic Research (NCAER) in New Delhi and University of Maryland (Desai et al. 2015). The survey was conducted between November 2011 and October 2012, covers 42,152 households located throughout India. The survey covered all the states and union territories of India except Andaman and Nicobar, and Lakshadweep. These two account for less than 0.05 percent of India's population. The data is publicly available from the Data Sharing for Demographic Research program of the Inter-university Consortium for Political and Social Research (ICPSR).[4] The sample was drawn using stratified random sampling.

The IHDS sampled ever-married women above the ages of 15 (one was randomly chosen from each surveyed household), who were then administered a separate health and education questionnaire that included questions on marriage and reproductive history, as well as questions on health investments. For the analysis of the effect of age at marriage on labor market outcomes of women, we restrict ourself to the women who participate in the labor force, have non-missing information on labor market earnings and labor supply, whose marital age is not less than 5 years and menarcheal age between 9 and 21 years,[5] have valid information on age, height, caste, family attributes like parental education and number of siblings, and place of residence (rural/urban), leaving us with 10,511 women. We will refer to this sample as the "women's sample".

For the analysis of the effect of marital age of women on spousal labor market outcomes, we however need not restrict ourselves to working women. In this case, our sample consists

[4] http://www.icpsr.umich.edu/icpsrweb/DSDR/studies/36151

[5] The normal menarcheal age is between 10 and 15 years. However, menarcheal age as low as 9 years is not unusual (see for e.g. https://timesofindia.indiatimes.com/city/goa/Girl-talk-Menarche-now-at-8-9-years/articleshow/34169175.cms). Similarly, menarcheal age above 15 years, and in fact, as high as 20-21 years is also not biologically impossible. Delayed puberty may be constitutional or due to pathologic causes (Blondell et al. 1999). Uundernourishment during childhood is, in fact, one major reason for delayed menarche. Also, intense physical activity during childhood may delay menarcheal age. In this context, based on a survey of dancers and atletes, Frisch et al. (1980) and Frisch et al. (1981) note that dancers and athletes who began their training at ages 9 or 10 years still had not menarche at ages 18–20 years.

of all those women whose spouses are working and spouses have valid labor market earnings and labor supply information, and spousal age is non-missing. Moreover, as in the women sample, the women included in this sample were married not before 5 years of age, their menarcheal age is between 9 and 21 years, and they have valid information on age, height, caste, family attributes like parental education and number of siblings, and place of residence. With all these restrictions in place, the sample consists of 21,718 women. We will refer to this sample as the "spousal sample".[6]

In our analysis, we specifically focus on three labor market outcomes of women and their spouses including hourly earnings,[7] annual wage earnings, and work days per year.[8] Table 1 provides descriptive statistics on the two analytical samples. The average hourly earnings of the women included in our analytical sample is Rs. 18.25, average annual wage earnings is Rs. 24,000, average number of work days per year is 205. The average hourly earnings of the working men (i.e., spouses of the women) is Rs. 33.12, average annual wage earnings is Rs. 66,900, average number of work days is 273 per year. The average age of marriage of women is 17.23 years for the women's sample and 17.93 years for the spousal sample. The average age at menarche is 13.88 years for the women's sample and 13.85 years for the spousal sample. Figures 1 and 2 graph the distribution of the age at marriage and age at menarche respectively for each of the two samples.

[6] Just to be clear, In the women's sample, we include working women for whom we have valid data on labor market outcomes irrespective of whether their husbands are working or not. In the spousal sample, we include working spouses for whom we have valid data on labor market outcomes irrespective of whether their wives are working or not.

[7] Hourly earnings include hourly wages, bonus and other in-cash or kind benefits.

[8] In the women's sample, around 46% of the women work as the agricultural labourers, 23% work in construction, 6% work as teachers and the rest work in other areas. In terms of employment by industry type, 48% of the women are engaged in agriculture, agriculture-related or mining sector, 9% are engaged in the manufacturing sector, 22% are engaged in the construction sector and the rest are engaged in other sectors. In the spousal sample, 21% of the men (i.e., spouses of the women) work as agricultural labourers, 28% work as construction workers, 6% work as drivers and the rest work in other areas. By type of industry, 22% percent of the men are engaged in agriculture, agriculture-related or mining sector, 14% are engaged in the manufacturing sector, 27% are engaged in the construction sector and remaining are engaged in other sectors.

3 Empirical strategy

3.1 Econometric model

To examine the impact of women's age at marriage on their own labor market outcomes, we begin by estimating the following econometric model:

$$y_i = \alpha + \beta MarriageAge_i + \gamma X_i + \varepsilon_i \tag{1}$$

where y_i denotes a labor market outcome of woman i, $MarriageAge_i$ denotes the woman's age at marriage, X_i denotes the vector of individual and household level controls, and ε_i is the idiosyncratic error term that includes unobserved attributes like ability.

To examine the impact of women's age at marriage on her spousal labor market outcomes, we estimate a model similar to the model given by Equation (1). However, y_i now denotes a labor market outcome of woman i's husband. Moreover, X_i includes additional controls for spousal characteristics. Henceforth, we will refer to Equation (1) as the "women's regression", and the version of the equation (1) which models spousal labor market outcomes, as the "spousal regression".

Our parameter of interest is the coefficient β. In the women's regression, β captures the effect of women's age at marriage on their own labor market outcomes. In the spousal regression, it captures the effect of women's age at marriage on the labor market outcomes of their spouses.

As most studies do, we exclude various determinants of labor market outcomes such as educational attainment from the estimation, as these variables are potentially endogenous variables that could be influenced by an individual's decision about her timing of first marriage (see for e.g. Wang and Wang 2017). That is, these variables themselves could be the reasons why age at first marriage affects individuals' wages and work effort. Given that we condition on only exogenous variables, the estimated coefficient of β should be interpreted as the *total* effect of age at marriage on labor market outcomes.

Note, for the women's regression as well as the spousal regression, ordinary least squares (OLS) estimate of $\beta > 0$ is consistent with the selection hypothesis as well as the causal mechanisms. We could have consistently estimated β via OLS estimation and interpreted it as causal effect of age of marriage on labor market outcomes if, conditioning on exogenous characteristics, age at marriage was uncorrelated with unobservable determinants of labor market outcomes such as wages and labor supply (or more formally, $\mathbb{E}[MarriageAge \cdot \varepsilon | X] = 0$). However, such assumption may be violated for several reasons. First, omitted variables may affect both the age at marriage of the women and their own and their spousal labor market outcomes. For instance, as noted previously, parents of girls with better labor market prospects may postpone their daughters' marriages until their earnings capabilities are fully revealed. Also, those parents who care a great deal about their daughter's career could postpone their daughters' marriages in order to let them pursue their careers. Both examples suggest that $\mathbb{E}[MarriageAge \cdot \varepsilon \mid X] \neq 0$ and more likely, $\mathbb{E}[MarriageAge \cdot \varepsilon \mid X] > 0$ in the women's regression. For the spousal regression, it is also likely that $\mathbb{E}[MarriageAge \cdot \varepsilon \mid X] > 0$. This may be because men of higher ability may want to marry more educated women, and hence might prefer marrying older women since they are likely to be more educated. As a result, OLS estimates would be biased in both the women's regression as well as the spousal regression.[9]

The second issue relates to the accuracy of the report of age of marriage. In the IHDS 2012 age at marriage was self reported. Inaccurate reports would generate measurement error in the explanatory variable. This could attenuate the estimates of the coefficient of interest. To address these concerns, we follow an instrument variable (IV) approach. We use age of menarche as an instrument for women's age at marriage. This instrument is motivated by the observation that has been made by sociologists and anthropologists that parents become extremely anxious to get their daughter married once she has reached menarche, partly to avert any unwanted pregnancies (Caldwell et al. 1983; Srinivas 1984; Chari et al. 2017).

[9] In principal, there might be other potential omitted variables which are not orthogonal to age of marriage of the women and might be correlated with the labor market outcomes considered.

As noted by Field and Ambrus (2008), a significant portion of the variation in timing of menarche is random, rendering it a good instrument for the age at marriage.[10] In what follows, we discuss our IV strategy in detail.

3.2 Instrumental variable strategy

The IV approach involves estimating a two stage model which is specified as follows:

$$MarriageAge_i = \lambda + \delta MenarcheAge_i + \kappa X_i + \eta_i \tag{2}$$

$$y_i = \alpha + \beta MarriageAge_i + \gamma X_i + \varepsilon_i \tag{3}$$

The first stage is given by the equation (2), and equation (3) is the structural equation. The women's age at marriage, $MarriageAge_i$, is instrumented by $MenarcheAge_i$, their age at menarche, and y_i are the women's and their spouses' labor market outcomes of interest. As above, X_i denotes a vector of individual and household level controls such as the woman's age, height, family attributes like her father's and mother's years of schooling, number of siblings, place of residence (urban/rural), caste and district fixed effects. Note, in the spousal regression, the vector X_i also includes the husband's age as an additional control.

We use a standard two stage estimation procedure (i.e., two stage least squares (TSLS)) and cluster standard errors at the district level.

3.3 Validity of the instrumental variable

In this section, we perform several checks to test the validity of the instrumental variable. First, we examine whether age at menarche predicts age at marriage which is the endogenous regressor. In line with the findings of Field and Ambrus (2008) in context of Bangladesh,

[10] Studies of twins have found that random genetic variation is the single largest source of variations in menarche (see for e.g. Kaprio et al., 1995)

and that of Sekhri and Debnath (2017) and Chari et al. (2017) in context of India, we find that age at menarche significantly predicts age at marriage. The results from the regression of women's age at marriage are presented in Table 2. Panel A presents the results obtained using the sample of working women (i.e., the women's sample), and panel B presents the results obtained using the sample of women with working spouses (i.e., the spousal sample). Column (1) reports the coefficient of age at menarche without additional controls. The value of the coefficient is 0.243, and it is statistically significant at 1% level of significance, the F-Statistic for the regression model is 14.87. The corresponding coefficient value of age at menarche and the F-statistic for the spousal sample are 0.147 and 9.4 respectively. Again the coefficient is statistically significant at 1% level of significance. These results eliminate concerns about 'weak instruments' Additionally, Figure 3 also presents the kernel density estimate of women's age at marriage by menarcheal age groups (early and late menarche)[11] revealing that the distributions of women's age at marriage is positively related to age at menarche.

Next, we examine the potential threats to the validity of this instrument. Medical literature suggests that severe malnutrition in early childhood might result in delayed onset of menarche (Sekhri and Debnath, 2017). Exposure to severe malnutrition could potentially also affect long term health of the women (for e.g. Stathopolu et al. (2003) note that acute malnutrition could result in stunting) and this consequently could affect their labor market prospects. This could undermine our instrument. We examine this correlation in our sample. Figure 4 shows average adult heights by age at menarche among the women in our sample. We do not observe any evidence of significant correlation between adult height and age at menarche.

As argued by Field and Ambrus (2008), abrupt changes in diet might also affects maturation. Sekhri and Debnath (2017) in this context note that, agriculture and agriculture-related activities, that employ majority of the Indians, are highly weather dependent. Extreme

[11] The early menarche group consists of those women who attained menarche at the age of 14 or earlier. The late menarche group consists of those women who attained menarche after the age of 14.

weather conditions (e.g. droughts, floods, etc.) in the women's birth year might lead to loss in household income resulting in transitory but severe malnutrition. Therefore, females born during these extreme weather events may experience delayed age at menarche as they are more likely to be malnourished. We control for this possibility in our first stage regression. In column (2), we add age of the women to account for extreme weather events at the time of birth. Additionally, in this regression we also include controls for women's caste affiliation in column (2). The point estimates and standard errors are similar across columns (1) and (2) in both Panels A as well as B.

Next, we include adult height in the regression in column (3) as a proxy for acute malnutrition in childhood. As noted by Chari et al. (2017), if height is a sufficient statistic for health investments and if undernutrition that affects menarche is also is severe enough to result in stunting, then conditioning on height is likely to eliminate any confounding factor related to health investments that affect both menarche and marriage conditions. We find that inclusion of height as an additional control changes the point estimates and the standard errors only slightly. Even if height is not a sufficient statistic for health, since it is closely related to health (Strauss and Thomas 1989), the fact that controlling for height has very small effects on our results suggests that they are not driven by unobserved health inputs that also affect age at menarche. We condition all subsequent results on adult height, women's age and caste affiliation.

It is thought that strenuous physical labor during early childhood can adversely affect health of children and lead to a delay in menarche (Pellerin-Massicotte et al. 1997). Thus women who end up marrying late may also be less healthy, and this could have a direct effect on her labor market prospects. However, as argued by Sekhri and Debnath, the children who work in India are not involved in hard physical work such as construction. They note that detailed data on child labor collected from northern India show that more than 99 % of working girls of age 6 to 14 are engaged in domestic work while 0.001 % of them work for wage (Basu et al. 2010). As such, strenuous physical labor during early childhood is unlikely to render our instrument endogenous. Nevertheless, to address this concern we include controls for women's father's and mother's educational attainment (i.e., years of schooling) as well as the number of siblings of the women and reestimate the first stage equation. We argue that these family characteristics are likely to serve as good proxies for economic status of women's natal family. Consequently, these family characteristics are likely to be correlated with a woman's age at marriage as well be determinants of whether the woman worked strenuously as a child or not. As evident from the results reported in Column (4), the inclusion of the women's natal family characteristics as additional controls does not change the point estimates of the coefficient of age at menarche significantly.

Age at menarche might also be potentially endogenous due to geographical factors such as temperature, rainfall, altitude, etc. (Field and Ambrus 2008; Chari et al. 2017). To address this issue, we control for place of residence (whether the household resides in an urban or a rural locality) and use district fixed effects to account for spatial variation in exposure to environmental factors that affect menarche. Note, we are able to control for district of residence of the married woman, and not her natal district since we do not have any information about the location of her natal family. This, however, is not likely to be a problem because in India most marriages occur within the same district, so the district of residence of the married woman is also likely also her natal district (Fulford 2015). The results of the specification that include geographic controls is presented in Column (5). The coefficient of age at menarche is still highly statistically significant.[12]

Another concern is measurement error in the age at menarche. While this is possible since it was self-reported by respondents at the time of the survey, Garg et al. (2001) and Sharma et al. (2006) note that menarche is a major event for girls in India, and girls of both low and high caste report knowing little or nothing about menstruation before it began, but afterwards learning of taboos about eating and mobility during menstrual periods. These

[12] Note, for all the spousal regressions reported in Panel B, except for that reported in column (1), spousal age is included as an additional control variable since we use this variable as an exogenous covariate in the second stage of the spousal regressions.

changes in lifestyle imply that respondents are likely to recall its timing with fair degree of accuracy (Chari et al. 2017).[13] Furthermore, the distribution of reported age at menarche (Figure 2) does not show any heaping at key ages (e.g. school leaving ages) that might be suggestive of significant recall error.

The final concern that we need to address is whether our instrument is exogenous conditional on the fact that we are not controlling for education. One might be tempted to argue that a woman's educational attainment as measured by her years of schooling, is correlated with her age at menarche. More specifically, menarche itself might be a barrier to schooling. If this is the case, then leaving out education from the set of control variables will violate the condition that $\mathbb{E}[MenarcheAge \cdot \varepsilon | X] = 0$, and the IV regressions will not yield consistent estimates of the parameters of interest.

To address the issue of endogeneity due to omission of schooling from our model, we do the following. First, we plot the average years of schooling of women by different menarcheal age in Figure 5. We find no evidence of an upward trend in the relationship between schooling and age at menarche. Second, we present the kernel density estimate of women's years of schooling by terciles of menarcheal age in Figure 6. The figure reveals that the population distributions, and not just averages, are remarkably similar across all subsamples. This is not what we would have expected to find if menarcheal age was correlated with years of schooling. Third, we explore the relationship between years of schooling of women, age at menarche, and marriage age using a regression framework. Results are reported in Table 3. We find that age at menarche has a positive and significant impact on years of schooling when we do not control for age at marriage. However, when we control for age at marriage, menarcheal age no longer significantly affects educational attainment of women. This suggests that conditional on age at marriage, menarcheal age does not have an effect on educational attainment. Thus, all the evidences suggest that not controlling for educational attainment

[13] Ellis (2004, 921) based on a survey also note, "both adolescent girls and adult women are generally willing and able to report accurately on their ages at menarche...and retrospective reports may be more reliable than those obtained during puberty".

of women are unlikely to confound our analysis. Our result, is in fact, consistent with Field and Ambrus' (2008) finding that menarcheal age has no direct impact on women's schooling in Bangladesh.[14]

4 Results

4.1 OLS results

The OLS estimates of the effect of women's age at marriage on own and spousal labor market outcomes are presented in Table 4. While these estimates are not causal, nevertheless they are likely to serve as useful benchmarks with which we would be able to compare our IV estimates, in turn, allowing us to distinguish causality from correlation due to selection into marriage.

Examining the results in Panel A of Table 4, we first find that women's age at marriage is positively associated with all own labor market outcomes that we have considered namely hourly earnings (column 1), annual wage earnings (column 2), and work days per year (column 3). More specifically, our results indicate that a year of delayed marriage increases women's hourly earnings by 2%, annual wage earnings by 3%, and work days per year by roughly 2. These results imply that marital delay is beneficial for women in terms of labor market outcomes.

Panel B of Table 4 presents the associations between women's age at marriage and labor market outcomes of their spouses. We find that, a year of delayed marriage for women has a positive effect on their spouses' labor market performance. Specifically, our results indicate that a year of delayed marriage of women increases their spouses' hourly earnings by 1%,

[14] Note, Sekhri and Debnath (2014) and Chari et al. (2017) also implicitly assume that age of menarche is not correlated with women's education. Both the papers investigate investigate the impact of marital age of the mother on child health and education outcomes. Marital age is instrumented by menarcheal age, but mother's education is not controlled for. Given that mother's education is conjectured to a determinant of child outcomes, mother's education becomes of the part of the error term in the second stage regression, which must be assumed to be uncorrelated to menarcheal age, for their second stage parameter estimates to be consistent.

annual wage earnings by 2%, and work days per year by 1. These results imply that women's age at marriage has a positive spillover effect on their spouses' labor market outcomes.

These results appear to be consistent with causal hypotheses such as the formal education hypothesis and mobility argument in Loughran and Zissimpoulos (2004). These results, however, could also arise due to the selection hypothesis (Bergstrom and Bagnoli 1993). To examine whether the observed relationship between women's age at marriage and their own and their spouses' labor market is causal, we use the IV approach.

4.2 IV results

We next turn to the IV results in Table 5. As in case of OLS, we present the effects of women's age at marriage on their own labor market outcomes in Panel A. The effects of women's age at marriage on their spouses' labor market outcomes are presented in Panel B. We find that the magnitude of the IV estimates of the effect of women's age at marriage on their labor market outcomes are much smaller compared to the OLS estimates. For example, the effect of a one year delay in women's age at marriage on their hourly earnings is now only 0.5%. The corresponding OLS figure was 2%. Similarly, the effect of a year delay in women's annual wage earnings is now 1%, which according to the OLS estimates was 3%. More importantly, in sharp contrast to our OLS results, we now find that IV estimates of the coefficients of women's age at marriage is not statistically significant in any of the regressions that are reported in Panel A. These results indicate that a delay in marriage of women by a year has no significant causal impact on their own labor market outcomes.

Panel B of Table 5 reveals results that are similar to that of Panel A. Specifically, we find no evidence that a delay in women's marriage by a year has any causal impact on the labor market outcomes of their spouses. In terms of the magnitudes of the estimates of the effect of women's age at marriage on spousal labor market outcomes, the effects are now actually negative although statistically insignificant.

Overall, thus the IV results indicate that a delay in marriage of women by a year has no significant causal impact on own or spousal labor market outcomes. The OLS results were, thus, arising due to positive selection into marriage.

4.3 Comparison of OLS and IV based on compliers: selection vs. causal mechanisms

The comparison of our OLS and IV estimates indicates that our results are consistent with the selection hypothesis. However, as noted by Wang and Wang (2017), the comparison of full-sample OLS and IV estimates may not necessarily be a fair evaluation of the hypothesis. As pointed out in Imbens and Angrist (1994), the IV estimates capture only the effect for a subgroup of individuals who are likely impacted by the IV (i.e., the compliers). The IV estimates for the complier sample might actually be higher than the IV estimates for the full sample.

In the present study, the set of compliers will not consist of women who were married before they attained maturation. For these women, age at menarche would not have affected their marital age. Consequently, we assume that our complier subsample excludes women who were married before menarche.[15] [16] We re-estimate the OLS and IV models for these women. The OLS and IV results for the potential compliers are reported in Tables 6 and 7 respectively. We continue to find that the OLS estimates are larger than the IV estimates. Moreover, as before, the OLS estimates are statistically significant whereas the IV estimates are not. This again indicates that our results are consistent with the selection hypothesis.

[15] The average age at marriage of these women is 17.91 years for the women's sample and 18.41 years for the spousal sample.

[16] It might be that the complier set excludes not only women who were married before maturation but also those who were married *much* after maturation (say, those who were married after teenage or after early adulthood). Consequently, for checking sensitivity of our results, we assume that our complier subsample excludes women who were married before menarche or were married ten years after menarche. The results —reported in the Appendix—indicate no substantial difference in the our findings with respect to that obtained based on the baseline complier sample.

4.4 Addressing biases due to selection into labor force

To this point, we have considered only labor market outcomes of women and their spouses who are working in our empirical analysis. We have excluded nonworking women (from the women's regression) and men (from the spousal regression) since we do not observe wages and labor supply for these nonworkers. Given that non-working individuals systematically differ from working individuals, our analysis could be biased. This might especially be relevant for our analysis since women in India have remarkably low rates of employment and labor force participation.[17] In fact, in accordance to Becker's (1973) theory of specialization, if women drop out of the labor market, ignoring this could potentially lead to underestimation of the beneficial effects of delayed marriage for women. In principle, thus, this might be one reason why we observe statistically insignificant causal effects of delayed marriage on labor market outcomes of women for the full sample and the complier sample. In what follows, we examine whether the observed results and patterns are robust to addressing the selection. Our approach closely follows that used by Wang and Wang (2017).

4.4.1 Selection models and validity of exclusion restriction

To address the selection issue, consider the extended system of Equations (2) and (3) in the presence of endogeneity.

$$MarriageAge_i = \lambda + \delta MenarcheAge_i + \kappa X_i + \eta_i \quad (4)$$

$$y_i = \alpha + \beta MarriageAge_i + \gamma X_i + \varepsilon_i \quad (5)$$

$$S_i = \mathbb{I}(\delta Z_i - \mu_i \geq 0) \quad (6)$$

[17] National Sample Survey (NSS) data for India show that labour force participation rates of women aged 25-54 (including primary and subsidiary status) have stagnated at about 26-28% in urban areas, and fallen substantially from 57% to 44% in rural areas, between 1987 and 2011 (see https://www.livemint.com/Opinion/vgO1ynMV6UMDnF6kW5Z3VJ/Low-stagnating-female-labourforce-participation-in-India.html)

where S is an indicator and equal to one if one participates in the labor market and zero otherwise; and Z is a vector of exogenous characteristics, which can include a variable not in the set of X. Equation (6) indicates that a woman or a man decides to participate in the labor market when $\delta Z - \mu \geq 0$. The model can be identified under typical assumptions for IV and selection models and estimated via a variant of the conventional Heckman model (see Wooldridge (2010, p. 809) for details).

As is noted by Wang and Wang (2017) and many others, Heckman type of selection models do not perform quite well even though identification can be achieved through distributional assumption without an exclusion restriction (i.e., $Z = X$). To address this concern, we include an exclusion restriction—spousal earnings (Z)—to aid identification. Specifically, Z equals one if spousal earnings is greater than median income and zero otherwise. This choice of exclusion restrictions for the labor supply equation (particularly for that by females) is a popular one in the literature, and similar variables have been used in the previous literature (e.g. Buchinsky 2001; Chang 2011; Martins 2001). Below, we present our evidence supporting this choice.

To assess the validity of our exclusion restriction, we present two sets of results in Table 8 based on the full sample as well as the complier sample. The first set is concerned with the strength of empirical relationship between our exclusion restriction and labor force participation decision of the women and their spouses. The literature has generally found strong evidence that spousal income influences a woman's decision to participate in the labor market (e.g. Mroz 1987; Zabel 1993). We present the marginal effects of spousal income on a woman's probabilities of labor force participation. Consistent with the literature, our first-stage results show that spousal income indeed has a negative and statistically significant effect on labor force participation rates among women. Specifically, having a spouse who earns more than median income can reduce female labor force participation rate by roughly 12% (Panel A, Column 1 of Table 8). Similar results are also obtained for the spouses of women. Specifically, having a spouse who earns more than median income can reduce male labor force participation rate by roughly 3% (Panel B, Column 1 of Table 8).

The second set of results is concerned with the independence of an exclusion restriction; the exclusion restriction must be independent of potential labor market outcomes (or conditional on X). Such assumption may be violated if spousal earnings has any direct effects on individual wages or labor supply, or is indirectly related with individuals wages or labor supply through other channels. As noted by Wang and Wang (2017), one possibility is selection into marriage based on unobservable determinants of individual labor market outcomes, which implies potential non-zero correlation between spousal income and the error term as well. To formally test whether this assumption (along with the monotonicity assumption) is violated, we conduct a formal test based on a novel method proposed in Huber and Mellace (2014). They show that under our model assumptions, the following inequalities hold:

$$\begin{aligned} \mathbb{E}[y|z &= 1, S = 1, y_i \leq y_q] \leq \mathbb{E}[y|z = 0, S = 1] \\ &\leq \mathbb{E}[y|z = 1, S = 1, y_i \geq y_{1-q}] \end{aligned}$$

where y_q the qth conditional quantile in the conditional outcome distribution given $Z = 1$ and $S = 1$. Such inequalities imply the following null hypothesis:

$$\begin{aligned} \mathbb{E}[y|z &= 1, S = 1, y_i \leq y_q] - \mathbb{E}[y|z = 0, S = 1] \leq 0 \\ \mathbb{E}[y|z &= 0, S = 1] - \mathbb{E}[y|z = 1, S = 1, y_i \geq y_{1-q}] \leq 0 \end{aligned}$$

Huber and Mellace (2014) propose a test procedure to verify these inequalities. A negative test statistic with a large p value indicates that the IV is valid. The results for the full sample are presented in column (2) of panels A and B of Table 8. We fail to reject the validity of our exclusion restriction, strongly in favor of the use of the presence of spousal income as an exclusion restriction for the selection equation.

Columns (3) and (4) in Panels A and B of Table 8 replicate regressions reported in columns (1) and (2) respectively for the complier sample. As evident, these results are in line with the results for the full sample. These results, while not necessarily definitive, do increase our confidence in the identification assumption used in our analysis.

4.4.2 Results addressing selection

We now turn to actual estimates addressing the selection issue. We repeat all of our analysis addressing the selection issue. The results for the full sample and complier sample are presented in Table 9. As we can see, all of our baseline results continue to hold. Not only do we find similar patterns in our estimates; we generally find estimates to be remarkably similar in magnitudes as well. Specifically, we again find a statistically insignificant effect of women's delayed marriage on hourly earnings, annual wage earnings, and work days per year among women and their spouses for the full sample and the complier sample. Since the IV estimates continue to be statistically insignificant even after correcting for selection bias, we conclude that there does not exist a causal relationship between women's age at marriage and their own and their spousal labor market outcomes.

5 Discussion and Conclusion

In this paper, we examine the relationship between women's age at marriage and their own and their spouses' labor market outcomes using nationally representative household data from India. We find evidence of positive effects of age at marriage of women on their own as well their spouses' labor market outcomes. To examine whether these effects are causal (i.e., the effect arises due to more schooling as a result of marriage delay for example) or the effects arise due to selection into marriage, we use an IV based empirical strategy that utilizes variation in age at menarche to obtain exogenous variation in the age at marriage.

Our results indicate that the positive effects of age at marriage of women on their own as well their spouses' labor market outcomes arise due to selection into marriage. Our findings are robust to (1) dropping women from our sample who are not likely to be affected by our instrument, and (2) addressing biases due to nonrandom selection into labor force. Our findings might appear to be somewhat puzzling since it has been documented in the previous literature that a delay in women's marriage leads to more schooling in developing countries (see for e.g. Field and Ambrus 2008). In fact, for our analytical sample as well, we show that there exists a positive link between marriage delay and women's schooling (see Table A2 in the Appendix). As such, at least due to the formal schooling hypothesis, one would have expected to find a causal effect of a women's age at marriage on labor market outcomes.

We argue that one potential explanation for this apparently puzzling finding could be as follows. It has been noted by Kingdon (1998), Kingdon and Unni (2001), and Kanjilal-Bhaduri and Pastore (2017) among many others that in India, labor market returns to education is low and insignificant for women with relatively low education. This could be due to: (1) low quality of primary education in India (Pratham 2011),[18] and/or (2) for labor market success, a threshold level of education might be necessary (for instance, completing college or having a vocational degree); below that, an extra year of schooling might not lead to better labor market outcomes. As it turns out, 72% women in our sample have completed at most primary education (i.e., five years of formal schooling) and more than 90% have completed only secondary schooling (i.e., 10 years of formal schooling). Thus, although women in our sample might complete more formal schooling due to a delay in marriage by a year, this might not be sufficiently productive to get translated into better labor market outcomes since most women in our sample would still belong to the lower end of the education distribution.

Our findings thus suggest that complementing policies that seek to delay marriages of women in developing countries with educational policies that would augment the quality of primary schooling is likely to be useful. If this could be achieved, even a delay in marriage

[18] Pratham (2011) notes that 48% of Indian children in grade 5 could not read at grade 2 level, and nearly 58% could not solve a simple division problem.

by a year that might allow a woman to attain only one more year of primary schooling might be useful for her in the labor market. Additionally, policymakers perhaps might also think of designing policies that would incentivize parents to delay their daughters' marriages by such an extent that they are able to complete higher education (for e.g. complete college or finish 15 years of formal schooling), since our suggest that a marriage-delay policy that would cause women to complete an extra year of education is unlikely to be meaningful in terms of getting translated into better labor market prospects for women who only complete primary or secondary schooling.

References

[1] Basu, K., Das, S., and Dutta, B. (2010). "Child labor and household wealth: Theory and empirical evidence of an inverted-U," *Journal of Development Economics*, 91, 8-14.

[2] Becker, G.S. (1973). "A theory of marriage: part I," *Journal of Political Economy*, 81, 813–846.

[3] Benham, L. (1974). "Benefits of Women's Education within Marriage," *Journal of Political Economy*, 52, S57–S71.

[4] Bergstrom T., and Bagnoli M. (1993). "Courtship as a waiting game," *Journal of Political Economy*, 101, 185–202.

[5] Blondell, R.D., Foster, M.B., and Dave, K.C. (1999). "Disorders of Puberty," *American Family Physician*, 60, 209-218.

[6] Buchinsky, M. (2001). "Quantile regression with sample selection: estimating women's return to education in the U.S.," *Empirical Economics*, 26, 87–113.

[7] Caldwell, J., Reddy, P., Caldwell, P. (1983). "Causes of marriage change in South India," *Population Studies*, 37, 343–361.

[8] Chang, S.K. (2011). "Simulation estimation of two-tiered dynamic panel Tobit models with an application to the labor supply of married women," *Journal of Applied Econometrics*, 26, 854–871

[9] Chari, A.V., Heath, R., Maertens, A., and Fatima, F. (2017). "The causal effect of maternal age at marriage on child wellbeing: Evidence from India," *Journal of Development Economics*, 127, 42–55.

[10] Choi, H.-J., Joesch, J.M., and Lundberg, S. (2008). "Sons, daughters, wives, and the labour market outcomes of West German men," *Labour Economics*, 15, 795-811.

[11] Dahl G.B. (2010), "Early teen marriage and future poverty," *Demography*, 47, 689–718.

[12] Das Gupta, S., Mukherjee, S., and Singh, S. (2008). "Knot ready: Lessons from India on delaying marriage for girls," International Center for Research on Women (ICRW).

[13] Desai, S., Dubey, A., and Vanneman, R. (2015). India Human Development Survey-II (IHDS-II) (Computer file), University of Maryland and National Council of Applied Economic Research, New Delhi (producers), 2015. Ann Arbor, MI: Inter-university Consortium for Political and Social Research (distributor).

[14] Ellis, Bruce J. (2004). "Timing of Pubertal Maturation in Girls: An Integrated Life History Approach." *Psychological Bulletin*, 130, 920–58.

[15] Field, E., and Ambrus, A. (2008). "Early marriage, age of menarche, and female schooling attainment in Bangladesh," *Journal of Political Economy*, 116, 881-930.

[16] Frisch, R. E., Wyshak, G., and Vincent, L. (1980). "Delayed menarche and amenorrhea in ballet dancers," *New England Journal of Medicine*, 303, 17-19.

[17] Frisch, R. E., Von Gotz-Welbergen, A., McArthur, J. W., Albright, T., Witschi, J., Bullen, B., Burnholz, J., Reed, R. B., and Hermann, H. (1981). "Delayed menarche and

amenorrhoea of college athletes in relation to age of onset of training," *Journal of the American Medical Association*, 246,1559-1563.

[18] Fulford, S. L. (2015). "Marriage migration in India: Vast, varied, and misunderstood," Working Paper 820, Boston College.

[19] Garg, S., Sharma, N., and Sahay, R. (2001). "Socio-cultural aspects of menstruation in an urban slum in Delhi, India," *Reproductive Health Matters*, 9, 16–25.

[20] Gladden, T. (1999). "Labor market effects of family ties and migration," Unpublished Manuscript.

[21] Huang, C., Li, H., Liu, P.W., and Zhang, J. (2009). "Why Does Spousal Education Matter for Earnings? Assortative Mating and Cross-Productivity," *Journal of Labor Economics*, 27, 633–652.

[22] Huber, M., and Mellace, G. (2014). "Testing exclusion restrictions and additive separability in sample selection models," *Empirical Economics*, 2011–45.

[23] Imbens, G., and Angrist, J. (1994). "Identification and estimation of local average treatment effects," *Econometrica*, 62, 467–475.

[24] Jepsen, L.K. (2005). "The Relationship Between Wife's Education and Husband's Earnings: Evidence from 1960 to 2000," *Review of Economics of the Household*, 3, 197–214.

[25] Kanjilal-Bhaduri, S., and Pastore, F. (2017). "Returns to Education and Female Participation Nexus: Evidence from India," IZA Discussion Paper No. 11209.

[26] Kaprio, J., Rimpelä, A., Winter, T., Viken, R. J., Rimpelä, M., and Rose, R. J. (1995). "Common genetic influences on BMI and age at menarche," *Human Biology*, 739-753.

[27] Kingdon, G. (1998), "Does the Labour Market Explain Lower Female Schooling in India?", The Journal of Development Studies, Vol. 35, No. 1, pp. 39-65.

[28] Kingdon, G. and Unni, J. (2001), "Education and Women's Labour Market Outcome in India", Education Economics, Vol.9, No. 2, pp 173-195.

[29] Lefgren, L., and McIntyre, F. (2006). "The relationship between women's education and marriage outcomes," *Journal of Labor Economics*, 24, 787–830.

[30] Liu, P. W., and Zhang, J. (1999). "Assortative mating versus the cross-productivity effect," *Applied Economics Letters*, 6, 523–25.

[31] Loughran, D.S., and Zissimpoulos, J. (2004). "Are there gains to delaying marriage? The effect of age at firstmarriage on career development and wages," Unpublished Manuscript

[32] Lundberg, S.J, and Rose, E. (2002). "The effects of sons and daughters on men's labor supply and wages," *Review of Economics and Statistics*, 84, 251–268.

[33] Martins, M. (2001). "Parametric and semiparametric estimation of sample selection models: an empirical application to the female labour force in Portugal," *Journal of Applied Econometrics*, 16:23–39.

[34] Mroz, T. (1987). "The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions," *Econometrica*, 55, 765–799.

[35] Pellerin-Massicotte, J., Brisson, G.R., St.-Pierre, C., Rioux, P., and Rajotte, D. (1987). "Effect of Exercise on the Onset of Puberty, Gonadotropins, and Ovarian Inhibin," *Journal of Appllied Physiology*, Vol. 63, pp. 1165–1173.

[36] Pratham (2011). *Annual Status of Education Report (ASER)*. Available at www.asercentre.org.

[37] Sekhri, S., and Debnath, S. (2014). "Intergenerational consequences of early age marriages of girls: effect on children's human capital," *Journal of Development Studies*, 50, 1670–1686.

[38] Sharma, N., Vaid, S., and Manhas, A. (2006). "Age at Menarche in two caste groups (Brahmins and Rajputs) from rural areas of Jammu," *Anthropologist*, 8 (1), 55–57.

[39] Srinivas, M. (1984). *Some Reflections on Dowry*. Oxford University Press, New Delhi, India.

[40] Stathopolu, E., J. Antony H., and David C. (2003). "Difficulties with Age Estimation of Internet Images of Southeast Asians," *Child Abuse Review*, 12, 46–57.

[41] Strauss, J., and Thomas, D. (1989). "Health, nutrition, and economic development," *Journal of Economic Literature*, 36, 766–817.

[42] Sunder, N. (2016). "Marriage Age, Social Status and Intergenerational Effects in Uganda," Working Paper. Available at https://www.wider.unu.edu/sites/default/files/Events/PDF/Papers/Sunder.pdf

[43] Wang, C., and Wang, L. (2017). "Knot yet: minimum marriage age law, marriage delay, and earnings," *Journal of Population Economics*, 30, 771–804.

[44] Welch, F. (1974). "Benefits of women's education within marriage: Comment," *Journal of Political Economy*, 82, S72–S75.

[45] Wooldridge, J.M. (2010). *Econometric analysis of cross section and panel data*. MIT Press.

[46] Zabel J.E. (1993). "The relationship between hours of work and labor force participation in four models of labor supply behavior," *Journal of Labor Economics*, 11, 387–416.

Figure 1. Distribution of women's age at marriage for the two samples

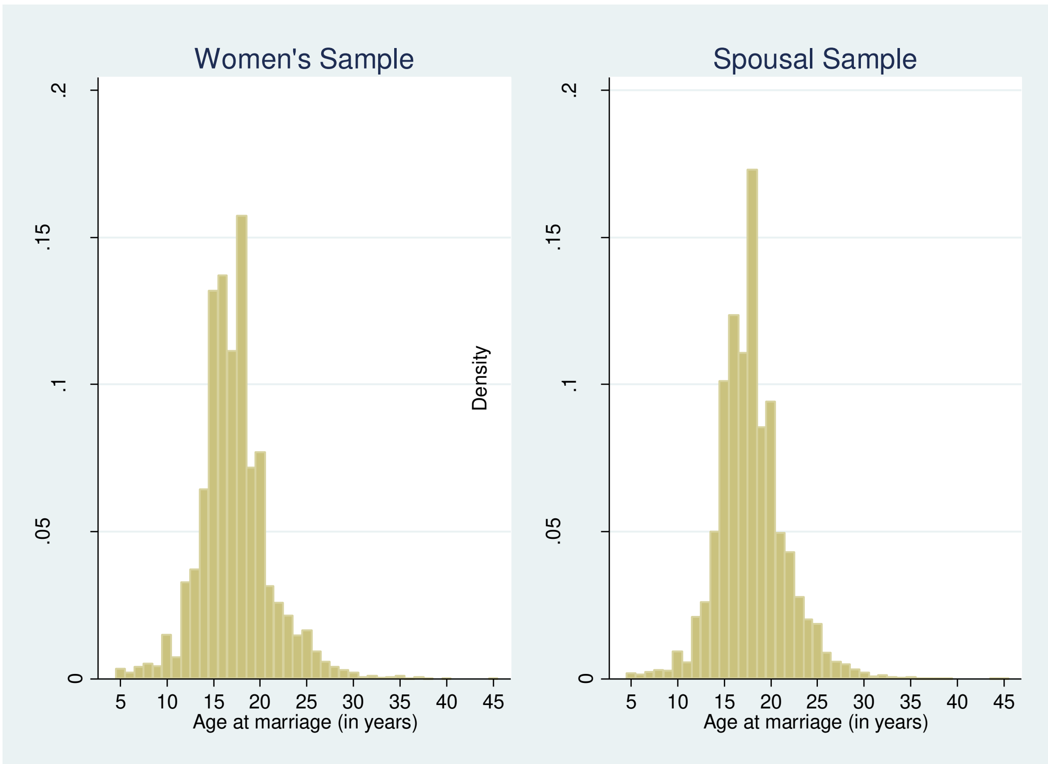

Figure 2. Distribution of age at menarche for the two samples

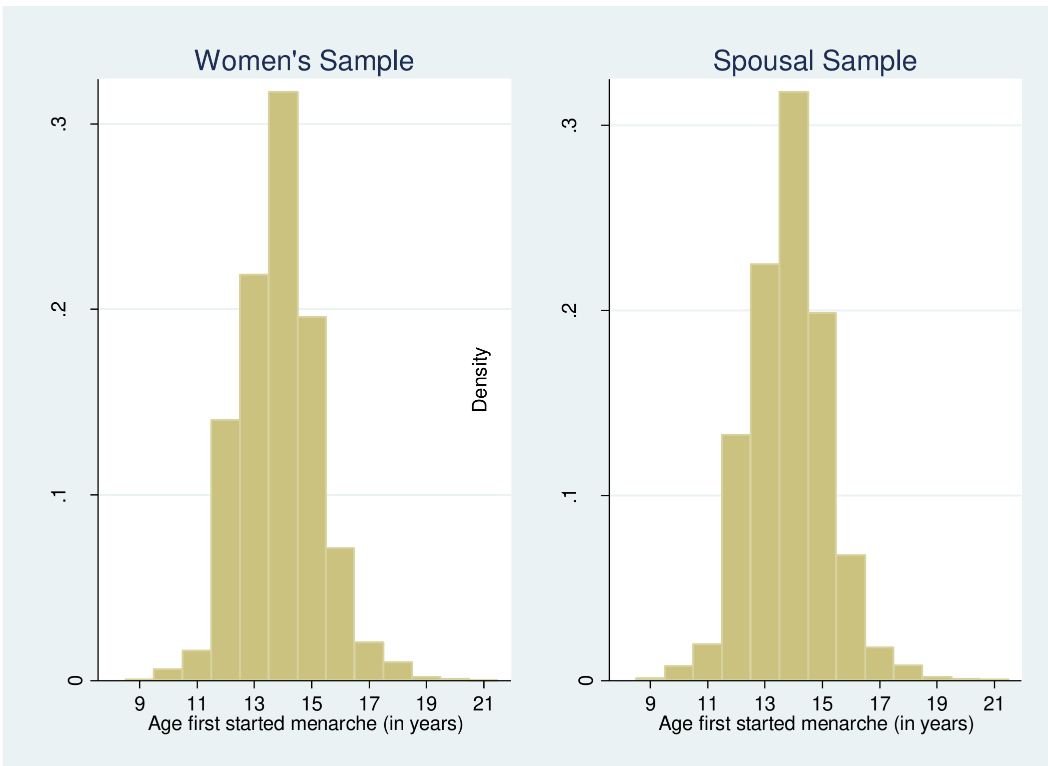

Figure 3. Distribution of women's age at marriage by age at menarche group for the two samples

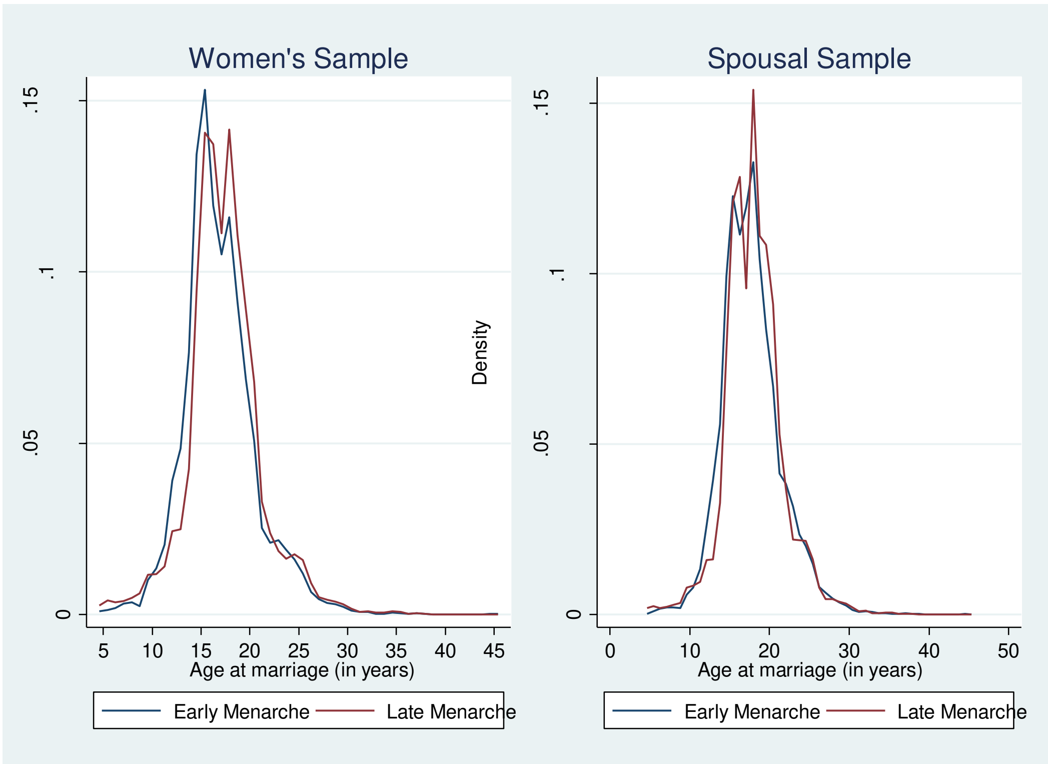

Figure 4. Relationship between women's average height and age at menarche for the two samples

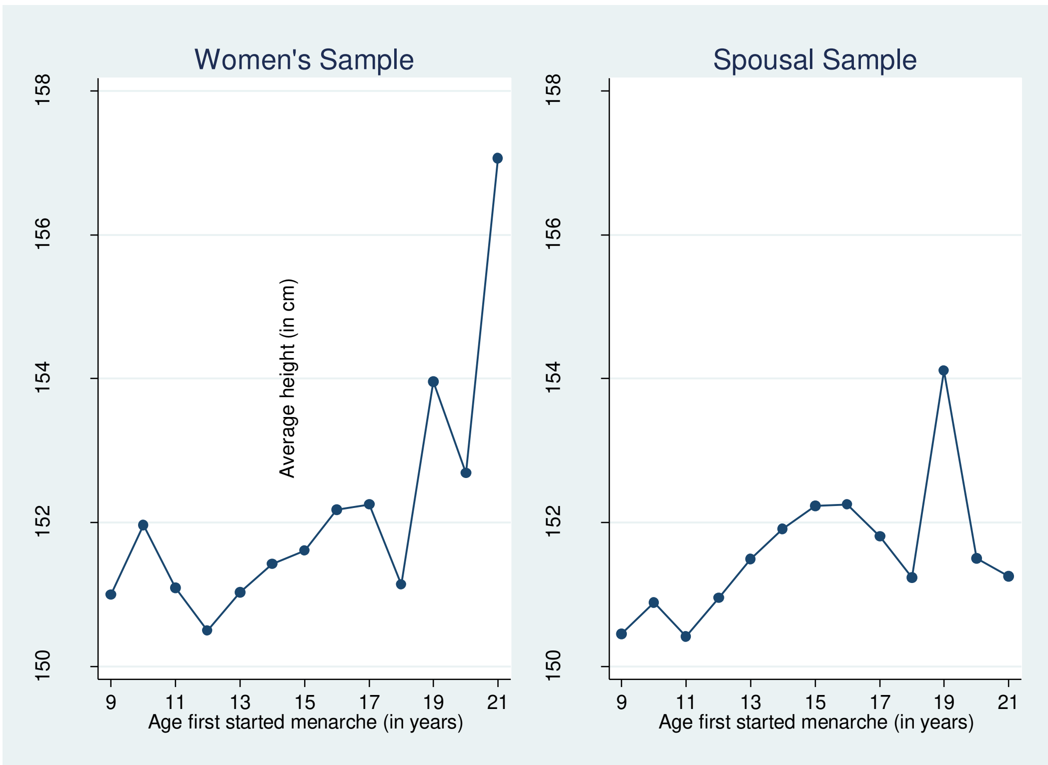

Figure 5. Relationship between women's average years of schooling and age at menarche group for the two samples

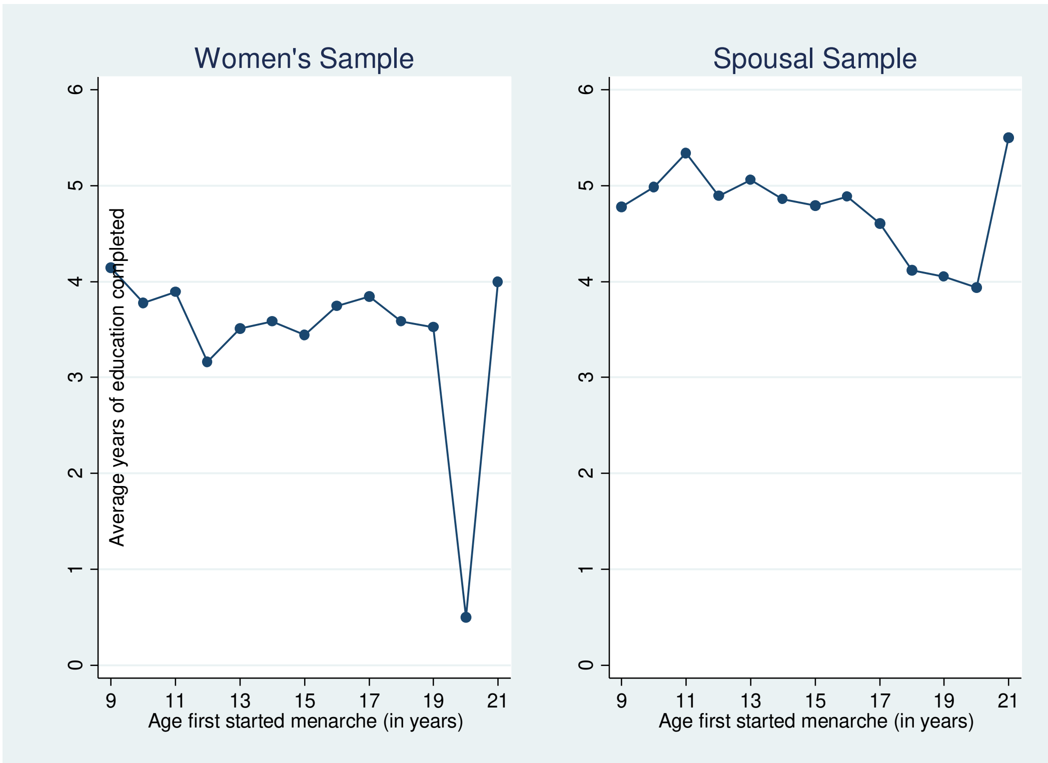

Figure 6. Kernel density estimates of women's years of schooling by terciles of age at menarche

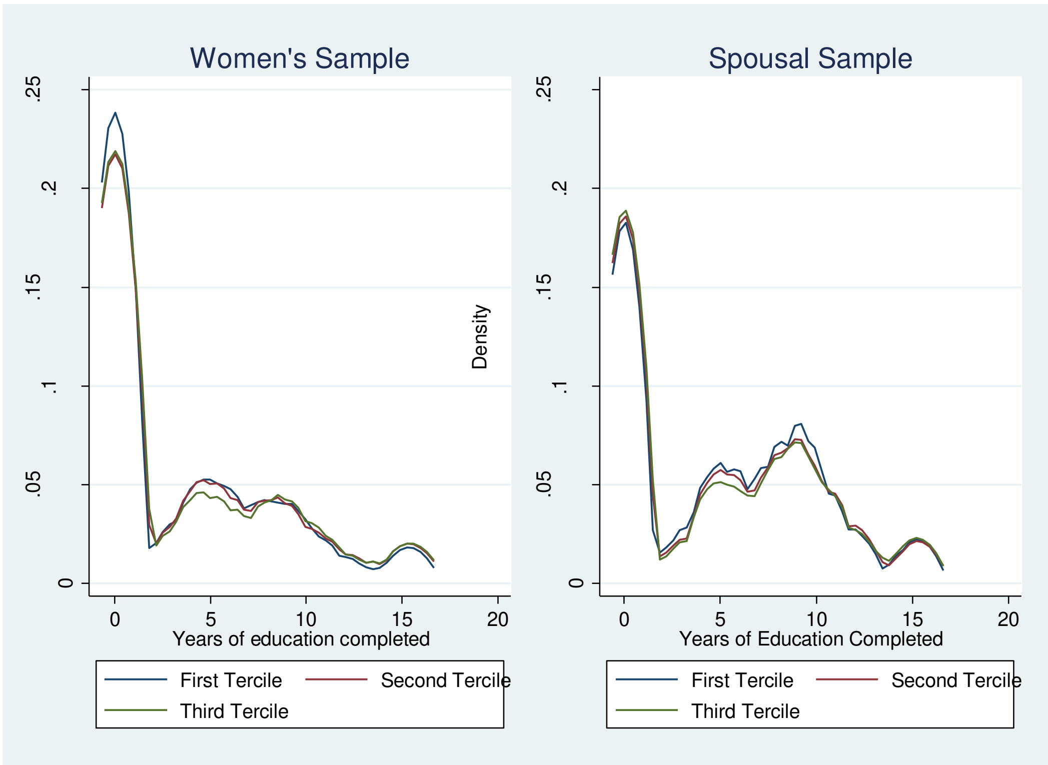

Table 1. Summary Statistics

	Women's Sample		Spousal Sample	
	Mean	SD	Mean	SD
Women's Labor Market Outcomes				
Hourly earnings (in Rs.)	18.25	24.40		
Annual wage earnings (in Rs.)	23977.56	50282.33		
Work days per year	205.29	103.85		
Spousal Labor Market Outcomes				
Hourly earnings (in Rs.)			33.12	41.16
Annual wage earnings (in Rs.)			66885.67	88398.96
Work days per year			273.22	82.13
Women's characteristics				
Age at marriage	17.23	3.76	17.93	3.62
Age at Menarche	13.88	1.40	13.85	1.39
Age	37.64	8.90	35.14	9.32
Spousal age	42.06	9.72	40.25	10.10
Height (in cm)	151.32	6.55	151.73	6.56
Father's years of schooling	2.09	3.76	3.11	4.34
Mother's years of schooling	0.85	2.42	1.36	2.94
Number of Siblings	3.77	1.97	3.81	1.98
Place of Residence (=1 if Urban)	0.21	0.41	0.34	0.47
N	10,511		21,718	

Notes: In subsequent regressions, women (spousal) sample is used for examining the impact of women's age at marriage on women's (spousal) labor market outcomes. Women (Spousal) sample consists of working as well as non-working spouses (women). As such we do not report the mean and standard deviations of labor market outcomes of spouses (women) in the women (spousal) sample. In the women sample, spousal age is available for 9,262 observations, and hence the mean and standard deviation of spousal age is computed based on these observations. We, however, do not use this variable in subsequent regressions based on the women sample.

Table 2. OLS estimates of the effect of age at menarche on women's age at marriage

Panel A: Women's Sample					
	[1]	[2]	[3]	[4]	[5]
Age at Menarche	0.243***	0.233***	0.220***	0.215***	0.446***
	(0.063)	(0.060)	(0.060)	(0.052)	(0.039)
F-statistic	14.87	32.06	30.69	84.00	130.00
R^2	0.008	0.053	0.059	0.149	0.361
Observations	10,511	10,511	10,511	10,511	10,511
Panel B: Spousal Sample					
	[1]	[2]	[3]	[4]	[5]
Age at Menarche	0.147***	0.164***	0.151***	0.154***	0.366***
	(0.048)	(0.043)	(0.043)	(0.037)	(0.024)
F-statistic	9.40	33.03	33.62	96.35	225.03
R^2	0.003	0.043	0.050	0.144	0.335
Observations	21,718	21,718	21,718	21,718	21,718

Notes: Estimation via OLS. The outcome variable is women's age at marriage. Regressions reported in columns (1) of Panels A and B, do not include any controls. In column (2) regressions we include women's age and caste affiliation as controls. In column (3) regressions the control variables are women's age, caste affiliation, and height. In column (4), controls include women's age, caste affiliation, height, father's years of schooling, mother's years of schooling, and number of siblings. In column (5), we include district fixed effects in addition to all controls used. For regressions reported in columns (2) through (5) in Panel B, we also include spousal age as an additional control. Standard errors reported in the parentheses are clustered at the district level. ***p < 0.01, **p <0.05, *p < 0.1.

Table 3. OLS estimates of the effect of age at menarche and women's age at marriage on women's years of schooling

Panel A: Women's sample		
	[1]	[2]
Age at Menarche	0.104***	-0.012
	(0.028)	(0.027)
Age at Marriage		0.260***
		(0.016)
R^2	0.539	0.567
Observations	10,511	10,511

Panel B: Spousal sample		
	[1]	[2]
Age at Menarche	0.116***	0.011
	(0.023)	(0.022)
Age at Marriage		0.286***
		(0.012)
R^2	0.535	0.566
Observations	21,718	21,718

Notes: Estimation via OLS. The outcome variable is women's years of schooling. The regressions reported in Panel A control for women's age, caste affiliation, height, father's years of schooling, mother's years of schooling, number of siblings, and district fixed effects. The regressions reported in Panel B include control for spousal age in addition to all the controls used in the regressions reported in Panel A. Standard errors reported in the parentheses are clustered at the district level. ***p < 0.01, **p <0.05, *p < 0.1.

Table 4. OLS estimates of the effect of women's age at marriage on own and spousal labor market outcomes

Panel A: Women's Regression			
	[1]	[2]	[3]
	Hourly Earnings	Annual Wage Earnings	Work Days Per Year
Age at Marriage	0.017***	0.033***	1.655***
	(0.002)	(0.004)	(0.352)
R^2	0.350	0.436	0.278
Observations	10,511	10,511	10,511
Panel B: Spousal Regression			
	[1]	[2]	[3]
	Hourly Earnings	Annual Wage Earnings	Work Days Per Year
Age at Marriage	0.011***	0.020***	0.873***
	(0.002)	(0.003)	(0.181)
R^2	0.372	0.449	0.240
Observations	21,718	21,718	21,718

Notes: Estimation via OLS. The regressions reported in Panel A control for women's age, caste affiliation, height, father's years of schooling, mother's years of schooling, number of siblings, and district fixed effects. The regressions reported in Panel B include control for spousal age in addition to all the controls used in the regressions reported in Panel A. Standard errors reported in the parentheses are clustered at the district level. ***p < 0.01, **p <0.05, *p < 0.1.

Table 5. IV estimates of the effect of women's age at marriage on own and spousal labor market outcomes

Panel A: Women's Regression			
	[1]	[2]	[3]
	Hourly Earnings	Annual Wage Earnings	Work Days Per Year
Age at Marriage	0.005	0.015	-0.288
	(0.010)	(0.020)	(1.677)
R^2	0.347	0.434	0.275
First Stage F-statistic	130.00	130.00	130.00
	[p=0.000]	[p=0.000]	[p=0.000]
Kleibergen Paap rK LM statistic	65.07	65.07	65.07
	[p=0.000]	[p=0.000]	[p=0.000]
Observations	10,511	10,511	10,511

Panel B: Spousal Regression			
	[1]	[2]	[3]
	Hourly Earnings	Annual Wage Earnings	Work Days Per Year
Age at Marriage	-0.005	-0.007	-0.262
	(0.011)	(0.019)	(1.481)
R^2	0.368	0.445	0.239
First Stage F-statistic	225.03	225.03	225.03
	[p=0.000]	[p=0.000]	[p=0.000]
Kleibergen Paap rK LM statistic	109.13	109.13	109.13
	[p=0.000]	[p=0.000]	[p=0.000]
Observations	21,718	21,718	21,718

Notes: Estimation via TSLS. The regressions reported in Panel A control for women's age, caste affiliation, height, father's years of schooling, mother's years of schooling, number of siblings, and district fixed effects. The regressions reported in Panel B include control for spousal age in addition to all the controls used in the regressions reported in Panel A. Standard errors reported in the parentheses are clustered at the district level. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

Table 6. OLS estimates of the effect of age at marriage on labor market outcomes, Complier Subsample

Panel A: Women's Regression			
	[1]	[2]	[3]
	Hourly Earnings	Annual Wage Earnings	Work Days Per Year
Age at Marriage	0.023***	0.040***	2.070***
	(0.003)	(0.005)	(0.415)
R^2	0.364	0.438	0.283
Observations	9,362	9,362	9,362
Panel B: Spousal Regression			
	[1]	[2]	[3]
	Hourly Earnings	Annual Wage Earnings	Work Days Per Year
Age at Marriage	0.011***	0.020***	0.938***
	(0.002)	(0.003)	(0.199)
R^2	0.370	0.446	0.248
Observations	20,102	20,102	20,102

Notes: Estimation via OLS. The regressions reported in Panel A control for women's age, caste affiliation, height, father's years of schooling, mother's years of schooling, number of siblings, and district fixed effects. The regressions reported in Panel B include control for spousal age in addition to all the controls used in the regressions reported in Panel A. Standard errors reported in the parentheses are clustered at the district level. ***p < 0.01, **p <0.05, *p < 0.1.

Table 7. IV estimates of the effect of age at marriage on labor market outcomes, Complier subsample

Panel A: Women's Regression			
	[1]	[2]	[3]
	Hourly Earnings	Annual Wage Earnings	Work Days Per Year
Age at Marriage	0.005	0.021	0.072
	(0.009)	(0.017)	(1.429)
R^2	0.358	0.436	0.281
First stage F statistic	267.65	267.65	267.65
	[p=0.000]	[p=0.000]	[p=0.000]
Kleibergen Paap rK LM statistic	91.42	91.42	91.42
	[p=0.000]	[p=0.000]	[p=0.000]
Observations	9,362	9,362	9,362
Panel B: Spousal Regression			
	[1]	[2]	[3]
	Hourly Earnings	Annual Wage Earnings	Work Days Per Year
Age at Marriage	-0.001	-0.004	-0.07
	(0.010)	(0.016)	(1.281)
R^2	0.368	0.443	0.247
First stage F statistic	372.73	372.73	372.73
	[p=0.000]	[p=0.000]	[p=0.000]
Kleibergen Paap rK LM statistic	136.15	136.15	136.15
	[p=0.000]	[p=0.000]	[p=0.000]
Observations	20,102	20,102	20,102

Notes: Estimation via TSLS. The regressions reported in Panel A control for women's age, caste affiliation, height, father's years of schooling, mother's years of schooling, number of siblings, and district fixed effects. The regressions reported in Panel B include control for spousal age in addition to all the controls used in the regressions reported in Panel A. Standard errors reported in the parentheses are clustered at the district level. ***p < 0.01, **p <0.05, *p < 0.1.

Table 8. Validity tests of instruments in selection models

Panel A: Women's Regression				
	Full Sample		Compliers	
	[1]	[2]	[3]	[4]
	Marginal effects	Validity test	Marginal effects	Validity test
Husband's income	-0.122***	-3.547	-0.118***	-3.587
	(0.010)	[p=1.000]	(0.010)	[p=1.000]
Panel B: Spousal Regression				
	[1]	[2]	[1]	[2]
	Marginal effects	Validity test	Marginal effects	Validity test
Wife's income	-0.030***	-3.248	-0.033***	-3.257
	(0.003)	[p=1.000]	(0.003)	[p=1.000]

Notes: The regressions reported in Panel A control for women's age, caste affiliation, height, father's years of schooling, mother's years of schooling, number of siblings, and district fixed effects. The regressions reported in Panel B include control for spousal age in addition to all the controls used in the regressions reported in Panel A. Standard errors reported in the parentheses are clustered at the district level. ***p < 0.01, **p <0.05, *p < 0.1. The validity test of the IV is developed in Huber and Mellace (2011). The null hypothesis is that the IV is valid.

Table 9. Selection-bias corrected IV estimates of the effect of age at marriage on labor market outcomes

Panel A: Women's Regression

	Full Sample			Compliers		
	[1]	[2]	[3]	[4]	[5]	[6]
	Hourly Earnings	Annual Wage Earnings	Work Days Per Year	Hourly Earnings	Annual Wage Earnings	Work Days Per Year
Age at Marriage	0.005	0.014	-0.357	0.005	0.021	0.096
	(0.010)	(0.020)	(1.669)	(0.009)	(0.017)	(1.419)
R^2	0.351	0.441	0.282	0.362	0.444	0.289
First Stage F-statistic	130.05	130.05	130.05	268.41	268.41	268.41
	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]
Kleibergen Paap rK LM statistic	65.07	65.07	65.07	91.41	91.41	91.41
	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]
Observations	10,511	10,511	10,511	9,362	9,362	9,362

Panel B: Spousal Regression

	Full Sample			Compliers		
	[1]	[2]	[3]	[1]	[2]	[3]
	Hourly Earnings	Annual Wage Earnings	Work Days Per Year	Hourly Earnings	Annual Wage Earnings	Work Days Per Year
Age at Marriage	-0.005	-0.006	-2.199	-0.001	-0.005	-2.247
	(0.011)	(0.019)	(1.723)	(0.010)	(0.016)	(1.490)
R^2	0.375	0.454	0.212	0.374	0.452	0.210
First Stage F-statistic	227.69	227.69	227.69	372.57	372.57	372.57
	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]
Kleibergen Paap rK LM statistic	109.38	109.38	109.38	135.87	135.87	135.87
	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]
Observations	21,718	21,718	21,718	20,102	20,102	20,102

Notes: Estimation via TSLS. The regressions reported in Panel A control for women's age, caste affiliation, height, father's years of schooling, mother's years of schooling, number of siblings, and district fixed effects. The regressions reported in Panel B include control for spousal age in addition to all the controls used in the regressions reported in Panel A. Standard errors reported in the parentheses are clustered at the district level. ***p < 0.01, **p <0.05, *p < 0.1.

Appendix

Table A1. OLS estimates of the effect of age at marriage on labor market outcomes, Alternative Complier Subsample

Panel A: Women's Regression			
	[1]	[2]	[3]
	Hourly Earnings	Annual Wage Earnings	Work Days Per Year
Age at Marriage	0.016***	0.028***	1.676***
	(0.003)	(0.005)	(0.466)
R^2	0.329	0.420	0.276
Observations	8,861	8,861	8,861
Panel B: Spousal Regression			
	[1]	[2]	[3]
	Hourly Earnings	Annual Wage Earnings	Work Days Per Year
Age at Marriage	0.013***	0.021***	1.267***
	(0.002)	(0.004)	(0.236)
R^2	0.355	0.434	0.242
Observations	18,910	18,910	18,910

Notes: Estimation via OLS. The regressions reported in Panel A control for women's age, caste affiliation, height, father's years of schooling, mother's years of schooling, number of siblings, and district fixed effects. The regressions reported in Panel B include control for spousal age in addition to all the controls used in the regressions reported in Panel A. Standard errors reported in the parentheses are clustered at the district level. ***p < 0.01, **p <0.05, *p < 0.1.

Table A2. IV estimates of the effect of age at marriage on labor market outcomes, Alternative complier subsample

Panel A: Women's Regression			
	[1]	[2]	[3]
	Hourly Earnings	Annual Wage Earnings	Work Days Per Year
Age at Marriage	0.005	0.021	0.157
	(0.008)	(0.016)	(1.395)
R^2	0.327	0.419	0.275
First stage F statistic	417.13	417.13	417.13
	[p=0.000]	[p=0.000]	[p=0.000]
Kleibergen Paap rK LM statistic	102.11	102.11	102.11
	[p=0.000]	[p=0.000]	[p=0.000]
Observations	8,861	8,861	8,861
Panel B: Spousal Regression			
	[1]	[2]	[3]
	Hourly Earnings	Annual Wage Earnings	Work Days Per Year
Age at Marriage	0.002	0.001	0.206
	(0.008)	(0.014)	(1.140)
R^2	0.354	0.432	0.242
First stage F statistic	747.88	747.88	747.88
	[p=0.000]	[p=0.000]	[p=0.000]
Kleibergen Paap rK LM statistic	155.86	155.86	155.86
	[p=0.000]	[p=0.000]	[p=0.000]
Observations	18,910	18,910	18,910

Notes: Estimation via TSLS. The regressions reported in Panel A control for women's age, caste affiliation, height, father's years of schooling, mother's years of schooling, number of siblings, and district fixed effects. The regressions reported in Panel B include control for spousal age in addition to all the controls used in the regressions reported in Panel A. Standard errors reported in the parentheses are clustered at the district level. ***p < 0.01, **p <0.05, *p < 0.1.

Table A3. IV estimates of the effect of age at marriage on women's years of schooling

Panel A: Women's Regression		
	Full Sample	Complier Subsample
	[1]	[2]
Age at Marriage	0.233***	0.242***
	(.058)	(.051)
R^2	0.567	0.569
First Stage F-statistic	130.00	267.65
	[p=0.000]	[p=0.000]
Kleibergen Paap rK LM statistic	65.07	91.42
	[p=0.000]	[p=0.000]
Observations	10,511	9,362
Panel B: Spousal Regression		
	Full Sample	Complier Subsample
	[1]	[2]
Age at Marriage	0.316***	0.329***
	(.058)	(.051)
R^2	0.565	0.556
First Stage F-statistic	230.33	383.46
	[p=0.000]	[p=0.000]
Kleibergen Paap rK LM statistic	110.57	137.69
	[p=0.000]	[p=0.000]
Observations	21,718	20,102

Notes: Estimation via TSLS. The regressions reported in Panel A control for women's age, caste affiliation, height, father's years of schooling, mother's years of schooling, number of siblings, and district fixed effects. The regressions reported in Panel B include control for spousal age in addition to all the controls used in the regressions reported in Panel A. Standard errors reported in the parentheses are clustered at the district level. ***p < 0.01, **p <0.05, *p < 0.1.