WP-2020-032

Structural Equation Approach to Modelling Social Norms in Female Education: A Case Study of India

Tanu Gupta and A. Ganesh Kumar

Indira Gandhi Institute of Development Research, Mumbai
October 2020

Structural Equation Approach to Modelling Social Norms in Female Education: A Case Study of India

Tanu Gupta and A. Ganesh Kumar

Email(corresponding author): tanug@igidr.ac.in

Abstract

The literature on the relationship between social norms and female educational outcomes has ignored some important aspects. First, norms are unobservable; what we observe are practices that are manifestation of norms. Second, norms are not monolithic that can be measured using a single indicator. They are indicated by several indicators and each indicator is an imperfect measure of the underlying norm. Third, norms are dynamic and can be affected by various factors. This paper examines the relationship between social norms and female educational outcomes addressing these concerns. We estimate a MIMIC (Multiple-Indicator- andMultiple-Cause) model in structural equation framework using the Indian Human Development Survey (IHDS), 2011-12 data for different regions and social groups. We find that norms adversely affect educational attainment for females and the effect is stronger in rural region than urban. The effect is more pronounced among Brahmins in rural region, followed by OBCs, Forward Castes, Adivasis and Dalits, whereas it is significant only among OBCs and Dalits in urban region. While education of both parents has a positive impact on female education, the mother's education has a norm-breaking effect and father's education has a norm-binding effect.

Keywords: Social Norms, Female Education, Structural Equation, MIMIC model, India

JEL Code: C30, D91, I24, Z13

Structural Equation Approach to Modelling Social Norms in Female Education: A Case Study of India

Tanu Gupta* and A. Ganesh Kumar†

Abstract

The literature on the relationship between social norms and female educational outcomes has ignored some important aspects. First, norms are unobservable; what we observe are practices that are manifestation of norms. Second, norms are not monolithic that can be measured using a single indicator. They are indicated by several indicators and each indicator is an imperfect measure of the underlying norm. Third, norms are dynamic and can be affected by various factors. This paper examines the relationship between social norms and female educational outcomes addressing these concerns. We estimate a MIMIC (Multiple-Indicator- and-Multiple-Cause) model in structural equation framework using the Indian Human Development Survey (IHDS), 2011-12 data for different regions and social groups. We find that norms adversely affect educational attainment for females and the effect is stronger in rural region than urban. The effect is more pronounced among *Brahmins* in rural region, followed by OBCs, Forward Castes, *Adivasis* and *Dalits*, whereas it is significant only among OBCs and *Dalits* in urban region. While education of both parents has a positive impact on female education, the mother's education has a norm-breaking effect and father's education has a norm-binding effect.

Keywords: Social Norms, Female Education, Structural Equation, MIMIC model, India

JEL Classification: **C30, D91, I24, Z13**

*Doctoral Student, Indira Gandhi Institute of Development Research, Mumbai, 400065, India. Email: tanug@igidr.ac.in

†Professor, Indira Gandhi Institute of Development Research, Mumbai, 400065, India. Email: agk@igidr.ac.in

1 Introduction

Researchers often advocated that the most important and common channel through which women can be empowered is through greater education. Although the gender gap in education in developed countries is almost non-existent, the gap between the education of men and women is large among developing countries (Eswaran, 2014). From Table 1, it is evident that the gender gap in literacy in 2011 was around 15 per cent and the female literacy rate was just 49.7 per cent in the least developed countries. In terms of female literacy, the South-Asian region is performing worse than other regions, except for Sub-Saharan region[1]. Empirical evidence suggests that the gender gap in education in developing countries considerably reduces the economic growth (Klasen, 2002; Knowles et al., 2002). Schultz (2002) conjectures that in developing economies, especially where women are less educated than men, disproportionately higher investment in the women's education is equitable as well as efficient. Increasing women's education contributes to reductions in mortality and fertility levels (Murthi et al., 1995). Furthermore, the education of mother has inter-generational benefits; it enhances the human capital of their children by improving the educational and health outcomes of their children, particularly for girls (Thomas, 1994; Behrman et al., 1999).

The educational attainment for the woman is influenced by myriad economic, social and cultural factors. The education literature has recognised that the social norms play a significant role in shaping the educational outcomes for women (Gandhi Kingdon, 2002; Dostie and Jayaraman, 2006; Gueye et al., 2018).

The social norm is a very broad term; in general, norms are those informal rules that describe how an individual should behave in a group, community or society as a whole[2]. Individuals adhering to a specific norm follow the same set of practices. It is very difficult to quantify or measure norm. Few studies (Sundaram and Vanneman, 2008; Rammohan and Robertson, 2012; Maertens, 2013; Rammohan and Vu, 2018) have analysed the effect of the norms on educational outcomes, by capturing the different

[1]Even for India, the figures associated with female literacy are not very promising. According to Human Development Report 2019, the proportion of women (ages 25 and older) in India who have at least secondary education is 39.0 per cent in 2018. The corresponding figure for the men in India is 63.5 per cent.

[2]We use the words social norms and norms interchangeably.

dimensions of norms with different indicators. But these studies have ignored three important aspects related to norms. First, we do not observe norms directly but we observe practices that are manifestations of the norms. The previous studies measure the effect of indicators (of the norms) rather the norms. Second, norms are not monolithic. They are multi-faceted. Existing studies focusing only on one dimension of norm using a single indicator do not capture the multi-faceted aspect of norm. Moreover, all such indicators are an imperfect measure of norm. That is, there is no one-to-one correspondence between norm and indicators. Few families may evince certain practices that are associated with a specific norm and yet they would educate their girl child. For example, many girls in Muslim communities follow the practice of wearing *hijab* and yet, they are well-educated. Looking at these practices, one might expect that families following such practices may have conservative attitudes and prevent their girls from getting educated. Similarly, in few families, girls are forced into child marriage but nevertheless they complete their education after marriage. On the other hand, certain families may not evince such practices but may not even educate their girl child. Third, the norms are dynamic; i.e., they are not static but change over time, even if only slowly. Various factors may play a significant role in changing norms. To the best of our knowledge, there is hardly any study which examines the factors that can cause social norms to change.

Against this background, this paper takes a new approach to modelling the relationship between social norms and female education outcomes that addresses the above mentioned concerns in measuring social norm. Specifically, we measure social norm as a latent variable in a MIMIC (Multiple-Indicator- and- Multiple-Cause) framework given by Jöreskog and Goldberger (1975). The MIMIC model has been extensively used in diverse contexts to measure the latent variable (Lee et al., 1997; Richards and Jeffrey, 2000; Alañón* and Gómez-Antonio, 2005).

The MIMIC model allows us to explicitly recognize norm as an unobservable variable, that is not monolithic but manifests imperfectly through several social practices, and is dynamic being influenced by other factors. Briefly, the MIMIC model considers a set of observable variables as the *indicators* of a latent variable or as *causes* of a latent variable or both. The MIMIC model is a set of simultaneous equations wherein the latent variable is specified to determine a set of endogenous observable variables (*indicators*)

and is itself linearly determined by a set of exogenous observable variables (*causes*). In the context of this paper, norm is the latent variable, which is (imperfectly) reflected in several observable indicators (various social practices) while simultaneously various factors can cause the norm itself to change.

The norm thus estimated as a latent variable using the MIMIC model is then hypothesised to affect female educational outcomes. Given the simultaneity, we estimate the MIMIC model with the structural equation to examine the relationship between norms and female educational outcomes. An added advantage here is that this approach allows us to identify factors that may weaken or bind the social norms around education.

We demonstrate the use of MIMIC framework to model the relationship between social norms and female education outcomes using India as a case. We estimate the model using the Indian Human Development Survey (IHDS) 2011-12, data across different social groups in rural and urban regions. We find that the norms have a significant effect on females' educational attainment. The effect is stronger in the rural region as compared with the urban region. The effect of norms varies social groups within the rural and urban region. The effect is more pronounced among *Brahmins* in rural region, followed by Other Backward Castes (OBCs), Forward Castes, *Adivasis* and *Dalits*, whereas it is significant only among OBCs and *Dalits* in the urban region. We do not find any significant effect of norms among Muslims.

We find that the education of parents plays a significant role in shaping the norms around education. The education of father has a norm-binding effect, whereas the education of mother has a norm-breaking effect on female education. Besides this, the education of both parents also has a direct positive effect on females' years of schooling. In other words, the education of parents has two effects: one is the indirect effect, in which parents' education affects the education of female through the change in norms; and the other is the more direct effect, which is the effect of parents' unobserved family background. However, the total effect of the education of parents is positive but the effect of the education of father is much larger.

The paper is organised in six sections. Section 2 gives a brief review of the studies of norms on female education. Section 3 provides data and descriptive statistics.

Section 4 describes the methodology used to capture the effect of the norms on female education. Results are presented in section 5 and section 6 concludes.

2 Literature

Social norms refer to the set of informal rules which governs what an individual believes he/she should do. Social norms shape multiple dimensions of human life, including females' educational outcomes. Social norms in various forms tend to impose binding constraint for female education. For example, norms around honour, marriage and dowry cost; norms related to marriage exogamy[3], patrilocality[4] and patrilineality[5] ; and norms around traditional gender-specific roles, together induce the parents to put higher weight to boys' education as compared to girls' education and, thus have implications for female education.

Maertens (2013) finds that perceptions about the ideal age of marriage among parents in rural India lowers the educational aspirations for their daughters compared to their sons. This social norm of the "ideal age of marriage" significantly constrains the desired education for girls. However, in the case of boys, the educational aspirations are affected by the returns to higher education, not by the norm of "ideal age of marriage". Also, parents tend to marry off their daughter at an early age in the fear of honour and reputation (Bicchieri, 2016). As a result, girls are taken out of school at an early age. This norm of honour and reputation is so strong in many developing countries that it leads to the practice of child and adolescent marriage. Early marriage is associated with lower educational attainment among women (Field and Ambrus, 2008).

Norms around honour and purity restrict women's mobility and induce the social practice of *Purdah* for female seclusion. In the regions where the women's movement is rigidly controlled, women practice *purdah* and experience lower literacy rates (Dyson and Moore, 1983; Bano, 2012).

[3]Exogamy refers to the practice of marrying outside one's social group.

[4]Patrilocality refers to the system where the bride leaves her parents' house and stays with or near the groom's family after marriage.

[5]Patrilineality refers to the system of family lineage in which descent is traced through the paternal side.

Norms around dowry cost exhibits broadly two types of practices related to marriage payments: dowry and bride price[6]. The custom of marriage payments affects the net returns to education and, thereby influences the parents' decision regarding investment in girls' education. Ashraf et al. (2020) find a large positive effect of school construction program on girls' education in the ethnic groups with the custom of bride price in India. This is because an increase in girls' education raises the bride price and increases the incentive for the parents to invest more in their daughters' education. On the other hand, in the societies with dowry custom, where the bride's family make marriage payments to the groom's family, parents have to accumulate a dowry for daughters and may rather decide to save their money for future dowry than investing in education.

Furthermore, in the regions where the norms of marriage exogamy, patrilocality and patrilineality are followed, women may face poorer educational and health outcomes (Dyson and Moore, 1983). This is because parents view their sons as financial support in their old age . They believe that the returns to their daughter's education might go to their in-laws family after marriage; thereby they discount the benefits which accrue to them personally and invest less in their daughter's education (Lahiri and Self, 2007). Different measures have been used to capture the effect of the custom of patrilocal exogamy and patrineality on women's education (Sundaram and Vanneman, 2008; Rammohan and Robertson, 2012; Rammohan and Vu, 2018). Sundaram and Vanneman (2008) use log-odds for women who have migrated from their birth-place to the log odds for men, whereas Rammohan and Vu (2018) use the proportion of women who are not living in their natal district as the measure of patrilocal exogamy. Furthermore, Rammohan and Robertson (2012) examine the role of kinship norms on female educational outcomes in Indonesia by capturing the effect of inheritance and post-marital residence practices at community-level. All these studies find that practices of patrilocality and patrilineality are unambiguously associated with poor educational outcomes for women.

Norms associated with traditional gender-specific roles hold females in the household responsible for household chores and care-giving. As a result, women are often burdened with domestic work. Many anecdotal evidences show that young girls drop out

[6]Bride Price is where the marriage payments and transfers are made by the groom's family to the bride's family, and Dowry is where the transfers are made by bride's family to the groom's family at the time of marriage.

from the schools due to domestic workloads .

Critique of literature

While previous studies have examined the implication of norms for the female education, they have ignored three important aspects of the norms. First, norms are not directly observable but what we observe are the practices that are manifestations of norms. For example, the norm around honour and purity lead to the practice of *Purdah/burkha*. We observe only the practice of women wearing *Purdah/burkha*. We do not observe the underlying social norm that contributes to such practices. Hence, it is very difficult to quantify or measure the norm.

Second, existing studies focus only on one dimension of norm by using a single indicator at a time. Since norms are multi-faceted, the existing studies have ignored this aspect of social norms. Moreover, all such indicators are imperfect measures of the underlying norm. A practice that we consider as an indicator may or may not actually be a result of any social norm. For example, certain households, who follow *purdah* practice, may restrict a girls' mobility and not send their girls' to school because of the norm of honour and purity. However, certain families may restrict their girls' mobility due to security reasons, not because of the norm. While the cultural practice of *purdah* hints to the prevalence of unfavourable norm for women, the same practice may lead to a favourable change in the norm for women. For example, evidence from Bangladesh suggests that the cultural practice of girls wears *burka/purdah*, may rather result in increasing secondary schooling by shifting the norm towards gender parity. It is also possible that many families do not follow the practice of *burka/purdah* and yet they do not send their girls to the school. This implies that there is no one-to-one correspondence between indicators and social norms which affect female educational outcomes.

Third, norms are dynamic in nature. They are not static. For example, norm around reputation becomes stronger as soon as a girl reaches to adolescent age. This shows that the norms change over time, even if only slowly. There are various factors or tools that may cause the change in norms. To the best of our knowledge, there is hardly any study that examines the factors that may cause the change in norms. Education is considered to be a powerful tool for the change in norms. However, when we talk about

education, the decision of investment in education is taken by the parents. Hence it could be the education of the previous generation that matters for the change in norms around education. This implies that parents' education could play a significant role in changing norms.

This highlights that since norms are latent, the polylithic nature of norms cannot be captured using a single indicator. The presence of social norms is evinced in the multiple indicators, even though imperfectly. However, the norms may also change because of various factors at the same time. We need to take into account all these aspects while quantifying the effect of social norms on female education. We propose to use the MIMIC model to measure the social norm as a latent variable which has been described in section 4. Dyson and Moore (1983) assert that practice of village exogamy[7], cross-cousin marriage, restrictions on widow marriage, restrictions on women's behaviour and mobility, and the severance of the relationship between the women and the natal family, affect woman's social status and shape gender attitudes, which may have implications for woman's outcomes like literacy, mortality, health etc. Hence we use the indicators reflecting the presence of these practices, to measure the social norms which have been discussed in section 3.

3 Data and Descriptive Statistics

3.1 Data and Variable Construction

We use data from the Indian Human Development Survey-II (IHDS-II) conducted in 2011-12 (Desai et al., 2015). It is a nationally representative multi-topic survey of 42,152 households, covering all the states and union territories, except Andaman & Nicobars and Lakshadweep. The surveyed households are spread across 1503 villages and 971 urban neighbourhoods. The rural sample was drawn using stratified random sampling. In urban regions, firstly, the towns and cities were selected using stratified sampling and then the household sample was drawn using probability proportional to population (PPP) sampling. IHDS-II survey collects detailed individual- and household-level information

[7]Village exogamy refers to the system where girls are not allowed to marry outside the natal village.

on education, gender relations, marriage practices, occupation, economic status, health, fertility, landholding, social capital and social identity. The advantage of using this survey is that it contains a few questions that could allow us to identify and capture the norms directly at household-level that could influence the educational outcomes for females.

We use the two modules to collect the information on females: (a) eligible women module and (b) household roster. The 'eligible women'[8] module collects the information on ever-married women of the household who belongs to the age of 15-49 years. It collects the information on their education, family education, fertility, birth history, gender relations and health beliefs. Besides eligible women, the survey contains information on other women (majorly never-married) in the household who were not interviewed. We collect the information for these women from the household roster, the details of which are given below. By combining the information on ever-married and never-married women, we get a sample consisting 42,322 women belonging to age 6-59 years for whom the complete information is available.

We aim to examine how norms affect the educational outcomes for females in India. We measure the educational outcome by the females' Average Years of Schooling (AYS). Hence we use the dependent variable as the completed level of education at the time of the survey measured in single years, which ranges between zero (illiterate) and sixteen (above Bachelors). One problem in using average years of schooling as the dependent variable is that the completed years of schooling is unknown for the girls who are still going into the school at the time of the survey. One way to solve this problem is to standardise the years of schooling with the females' age. Another way of dealing with this issue is to include the females's age as one of the explanatory variables (Duraisamy, 1992). In this paper, we follow the latter approach.

The females' years of schooling can be determined by various individual-level, household-level and school-level characteristics, apart from the prevalent social norm. We include the individual-level variables like her age and marital status; household-level variables like the education level of father and mother, proportion of brothers and

[8]The eligible women refer to the ever-married women in the household who ages between 15 and 49 years at the time of survey. A separate module is used to interview these women which is known as 'eligible women' module. A maximum of two such women were interviewed in each household.

sisters among all the siblings (separately), the highest level of education among brothers and sisters (separately), the logarithm of per-capita household expenditure and number of household's assets; and Norm as explanatory variables for females' average years of schooling. The data set does not contain any information on school-level characteristics for those who have completed their schooling, hence we cannot control for school-level variables.

The marital status of the given woman has been coded as a binary variable. It takes the value one if a woman is married, separated or divorced; and zero if she is never married. The household's income is one of the important determinants of females' years of schooling. We proxy the household's income by its consumption expenditure per-capita. The household's asset is given by the total number of assets owned by a household.

The education of father and mother has been included as explanatory variables to capture the effects of unobserved family background, information, income levels etc. It is because the household's consumption expenditure may not fully capture the true income effect of the household income, the education of father and mother could supplement it by capturing the effect of unobserved family background. Given household resources, a higher number of children (dependents) would imply a lower amount of resources available for each of them. Lower per-capita resources may in such a case impact the educational outcomes of all the children, especially for the girls who are discriminated against the boys while making such investment decisions. The explanatory variables such as the proportion of brothers and sisters among all the siblings would capture this effect. Furthermore, in a household where the son-preference exists and resources are constrained, parents would invest first in their son's education and then in their daughter's education. This implies that not only the number but also the composition of the siblings may affect women's education. The educational outcomes are also influenced by the peer effects. However, for a child, the closest peer group in his/her neighbourhood is his/her siblings. Hence, the explanatory variables like the education of brother and sister capture the peer effect, attitude towards schooling etc. These variables will not just capture the peer-effect but particularly, the gender-specific peer effect.

The information on variables such as the education of parents and siblings and

the composition of siblings is directly available for the 'eligible' ever-married women who were interviewed in the 'eligible women' module. The 'eligible women' module records the information on the parent's education (both father and mother separately), the number of brothers and sisters and their education levels separately[9]. Whereas this information is not directly available for the women (majorly unmarried) who were not interviewed in the eligible women's questionnaire. However, we can retrieve the same information for those women who are co-residing with their parents and siblings (in their natal house) at the time of the survey from the household roster. The household roster collects the information on all the household members. If an individual is residing with his/her parents in the same household, then the *"ID of the father"* and *"ID of the mother"* is recorded. We use the variables *"ID of the father"* and *"ID of the mother"* from the household roster to match the given girl or woman to her father and mother directly if parents are co-residing with the girl or woman. Other household members with the same parents' ID in the household roster are identified as the siblings of the given woman. After mapping the parents and the siblings to the given woman, we collect the information on the above explanatory variables from the household roster.

Our key variable of interest is the *Norm*. Identification of norm is not simple, as they are not observable. IHDS-II contains a few questions that help us to identify such practices within which these norms may be manifested. These questions are:

(i) Is it permissible to marry a girl in her natal village in your community/*jati*? (*Village*)

(ii) Is it permissible to marry a girl to her cousin in your community/*jati*? (*Cousin*)

(iii) Is it permissible for a widow to remarry in your community/*jati*? (*Widow*)

(iv) Do you practice ghungat/burkha/purdah/pallu? (*Veil*)

(v) When your family takes the main meal, do women usually eat with the men or eat first by themselves? Or do men eat first? (*Meal*)

[9]If a woman has more than one brother (sister), then the highest level of education among all the brothers (sisters) is recorded.

Each variable has been coded as a binary variable; the details of which are given in Appendix A1. We have considered these five questions to identify the norms associated around education, which will be used as indicators to measure the underlying norm[10]. Question (i)-(iii) capture the marriage practices prevalent in females' communities[11] and hence capture the norm around marriage. Moreover, the indicator *Village* captures the norm of village exogamy. The practice of cross-cousin marriage captures an important aspect of kinship norm. The practice of restriction on widow remarriage captures the attitudes towards females in their respective communities. It captures the norm of conservatism towards females. The indicator showing the practice of *purdah/burkha* captures the underlying norm of honour and purity. The indicator *Meal* captures the norm around traditional gender-specific roles. All these indicators have been constructed at the household level. The responses to the indicators (i)-(iii) are given at the household level. However, questions (iv)-(v) were posed to the 'eligible women' of the household. We use this information and construct the indicators at the household level[12].

3.2 Descriptive Statistics

Figure 1 depicts the average years of schooling for both males and females between the ages of 6-59 years. The gap between the average years of schooling between males and females increases as we move to higher age cohort. It is interesting to note that females aged 18-59 years have not even completed higher secondary schooling (10th grade). The highest year of schooling for females is only around 9 years.

From the descriptive statistics presented in Table 2, we find that the average years of schooling for females in India is only 5.3 years, which is equivalent to completing

[10]The 'eligible women' module has three more questions which could be used as indicators to measure the underlying social norm. These questions are as follows: (a) Would you consider being financially supported by your daughter? (b) Would you consider living with your daughter when you get old? (c) Do any members of your natal family live close enough for you to visit them and come back the same day? Questions (a) and (b) capture norms around traditional gender-specific roles. Question (c) capture the norm of patrilocal exogamy. These indicators capture the norm at individual-level but this information is not available for all the women in the sample. The sample size reduces drastically on the inclusion of these indicators in the analysis because of the missing information.

[11]These questions have been asked in the education and health module under the section of 'marriage practices'.

[12]There are few households where two women have been interviewed for the 'eligible women' questionnaire. In such case, the information from the younger women in the household has been utilised.

the primary level of education. It is evident from Table 2 that there is a male bias in terms of educational outcomes. Even though the average years of schooling for both the parents are below primary (5th grade); the average year of schooling for the father (4.0 years) is 2 years higher than that of the mother (1.8 years). The highest level of education among brothers is 7.6 years. However, the highest level of education among sisters is 5.8 years. This shows that even though the level of the educational attainment of males and females has increased over the generations within the same family, but the gap between their educational outcomes has not declined much.

We find that 48.6 per cent and 62.3 per cent of the females live in the communities where it is *not* permissible to marry a girl in their natal village and to their cousin, respectively (see Table 2). It is also noteworthy that approximately 64 per cent of the females belong to the household where women follow the custom of *purdah*/*burkha*.

Figure 2 depicts the differences in the females' average years of schooling across different social groups for rural and urban region separately. The entire sample has been divided into six social groups based on caste and religion. These six social groups are *Brahmins*, Forward Castes (FCs), Other Backward Castes (OBCs), *Dalits*, *Adivasis* and Muslims. It is interesting to note that females in urban areas have significantly higher years of schooling than females in rural areas across all social groups. On average, there is a significant difference of 2.6 years of schooling between urban females and rural females[13]. We observe from Figure 2 that *Brahmins* have highest years of schooling, followed by Forward Castes, OBCs, *Dalits*, *Adivasis* and Muslims. This holds for both rural and urban areas, with an exception that *Adivasis* have higher years of schooling than *Dalits* in urban areas.

[13]The average years of schooling for females ages between 6-59 years in the urban and rural area is 7.3 and 4.7 respectively.

4 Empirical Model

To examine the impact of norms on the females' educational attainment, we specify the following econometric model:

$$\begin{aligned} AYS_i = \alpha_0 + \alpha_1 Norm^* + \alpha_2 Age_i + \alpha_3 Age_i^2 + \alpha_4 FEDU_i + \alpha_5 MEDU_i \\ + \alpha_6 BRO_i + \alpha_7 BRO_i * BEDU_i + \alpha_8 SIS_i * SEDU_i + \alpha_9 EverMarried_i \\ + \alpha_{10} Consumption_i + \alpha_{11} Assets_i + \bar{\boldsymbol{X}}_{\boldsymbol{i}}{}' \boldsymbol{\gamma} + \epsilon_i \end{aligned} \tag{1}$$

where AYS_i denotes the average years of schooling of female i, Age_i and Age_i^2 denotes the current age and its square of female i, $FEDU_i$ denotes father's education, $MEDU_i$ denotes mother's education, BRO_i denotes the proportion of brothers among all siblings, $BEDU_i$ denotes the highest level of education among all brothers, SIS_i denotes the proportion of sisters among all siblings, $SEDU_i$ denotes the highest level of education among all sisters, $EverMarried_i$ denotes the current marital status of the female i; and $Consumption_i$ and $Assets_i$ denotes the log of per-capita consumption expenditure and number of assets of the household of female i. Following the Chamberlain-Mundlak approach to unobserved effects, we include a vector of $\bar{\boldsymbol{X}}_{\boldsymbol{i}}$ as additional sets of controls which consists the mean of all explanatory variables at the PSU (primary sampling unit) level to control for the community-level unobserved effects[14].

As discussed earlier, the norm is unobservable; hence we model it as a latent variable using the MIMIC model given by Jöreskog and Goldberger (1975). In the MIMIC model, the latent variable is linearly determined by a set of exogenous observed factors and determines a set of endogenous observed indicators. The MIMIC model consists of two sub-models: one is structural equation model and the other is measurement model. The structural equation model defines the relationship between the latent variable and the set of exogenous causal variables and is represented by

$$y^* = \boldsymbol{\lambda}' \boldsymbol{x} + \nu \tag{2}$$

where $\boldsymbol{x} = (x_1, ..., x_k)'$ is a column vector of k variables that are the possible cause of

[14]For more details, refer Wooldridge (2010).

the latent variable y^*. $\boldsymbol{\lambda}' = (\lambda_1, ..., \lambda_k)$ is a row vector of k coefficients showing the marginal effects of cause variables on latent variable y^*. ν is the white noise disturbance with mean zero and variance σ_ν^2.

The measurement model describes how latent variables are indicated or measured by observed endogenous variables and is represented by

$$\boldsymbol{y} = \boldsymbol{\delta} y^* + \boldsymbol{u} \tag{3}$$

where $\boldsymbol{y} = (y_1, ..., y_m)$ is a set of m indicators and the elements of vector $\boldsymbol{\delta} = (\delta_1, ..., \delta_m)$ represent the factor-loading coefficients. Each indicator in vector $\boldsymbol{y}$ is an imperfect measure of the latent variable y^*, hence an error term is added with each equation. $\boldsymbol{u} = (u_1, ..., u_m)$ is a column vector of m mutually independent error terms with zero means and their diagonal covariance matrix is represented by Θ_u. The error terms of the equations (2) and (3) are unrelated to each other.

Given this, the equation for the Norm can be written as:

$$Norm^* = \beta_1 FEDU_i + \beta_2 FEDU_i^2 + \beta_3 MEDU_i + \beta_4 MEDU_i^2 + \nu \tag{4}$$

$$\boldsymbol{y} = \boldsymbol{\delta} Norm^* + \boldsymbol{u} \tag{5}$$

where $Norm^*$ is the latent construct of variable Norm. $FEDU_i$ and $FEDU_i^2$ denotes the years of schooling for father of female i and its square respectively. Similarly $MEDU_i$ and $MEDU_i^2$ denotes the years of schooling for mother of female i and its square. In the terminology of MIMIC model, the set of these variables denotes the *cause* variables of the latent variable $Norm^*$. $\boldsymbol{y} = (Village, Cousin, Widow, Veil, Meal)'$ denotes the vector of indicators used to measure the latent variable $Norm^*$. $\boldsymbol{\delta}$ are the factor-loading coefficients that measure the how much of the variation in the latent variable is explained by the indicators. The path diagram for the equation of norm is given in Figure 3.

The whole system of equations (1), (4) and (5) is of our interest and it can be

interpreted as a system of 'Structural Equation Modelling' (SEM)[15]. The SEM has two sub-models: structural model and measurement model[16] (Bollen, 1989). Equations (1) and (4) are the structural equations[17] as they define the relation between structural variables $Norm^*$ and AYS[18]. The set of equations (5) forms the measurement model as it specifies how the latent construct $Norm^*$ is indicated or manifested in observed variables, $\boldsymbol{y}$.

One way of estimating equation (1) is to first estimate the MIMIC model of latent variable $Norm^*$ (equations (4)-(5)) using Maximum Likelihood (ML) method, and then obtain the empirical predictions of the latent variable from this model and lastly estimate the equation (1) to examine the female's educational outcomes. However, the estimates could be biased and inconsistent due to endogeneity issues.

The endogeneity could arise because the errors of equations (1) and (4) could be correlated to each other due to omitted variables or due to presence of exogenous variables (like FEDU, MEDU) which are common in both equations (see Figure 4). As a result, the endogenous variable $Norm^*$ become correlated to the disturbance term of equation (1), hence the single equation estimation will give inconsistent estimates. Therefore, the system of equations (1), (4) and (5) requires joint estimation. Another advantage of the joint estimation is that it helps in gaining efficiency. The estimators will be more efficient i.e. will have lower standard errors.

Besides endogeneity, another important issue in the estimation of such a model is 'identification'. Identification is the key element of all structural equation models. The whole model is identified if all the parameters of the model are identified. In the case of MIMIC model, one of the coefficients of the indicators needs to be set to unity[19] (Jöreskog and Goldberger, 1975). Hence, we set the coefficient of indicator *Village* equal to 1.

[15] MIMIC model is a special case of *Structural Equation Modelling (SEM).*

[16] A structural model defines the causal relationships between latent variables, between observed variables, or between latent and observed variables. A measurement model describes how latent variables are indicated or measured by observed variables.

[17] Structural equations refers to the set of equations that examines the relationship between latent variables.

[18] We can think about AYS as a latent which is directly observable and perfectly measurable, which means that $AYS = 1.0 * AYS^*$.

[19] The MIMIC model is identified if two things hold. First, the number of exogenous cause variables is one or more ($k \geqslant 1$) and the number of endogenous indicators is two or more ($m \geqslant 2$). Second, one of the coefficients of the indicators is set to unity. This also helps in providing the scale for the latent variable.

The parameters are estimated using a Maximum Likelihood (ML) method following the structural equations approach wherein the objective of the estimation procedure is to find the estimates for the model parameter vector $\boldsymbol{\theta}$[20] such that the predicted covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta})$[21] (i.e. $\hat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$) is as close as possible to sample covariance matrix $\mathbf{S}$ of the observed variables (causes and indicators).

5 Results

5.1 OLS estimates

Table 3 shows the OLS (ordinary least squares) estimates of equation (1), with the "indicators" as additional explanatory variables. We find the practice of village exogamy (*Village*) is negatively associated with females' years of schooling. This means that in the communities, where the girls are *not* allowed to marry outside their villages, females face poor educational outcomes. Restriction on widow marriage (*Widow*) is positively associated with females' years of schooling which imply that in communities where widow marriage is restricted, the average years of schooling for females are higher. However, the practice of restriction on cross-cousin marriage (*Cousin*) is uncorrelated with the education of women.

Three key observations emerge from the OLS estimates. First, even though these indicators or practices represent the manifestation of the same underlying norm for education, each indicator variable does not have a similar effect (or in the same direction) on females' years of schooling. This implies that a simple aggregation of these indicators would not be appropriate. Second, since these indicators are the measure of the underlying phenomenon, they could be highly collinear to each other. This could lead to the problem of multicollinearity among indicators. As a result, a few associated t-statistics may turn out to be insignificant. This could increase the probability of type-2 error. Third, since these indicators are imperfect measure of the same latent variable,

[20] Parameter vector $\boldsymbol{\theta}$ includes the elements of the vectors $\boldsymbol{\lambda}$ and $\boldsymbol{\delta}$; and the elements of the covariance matrix of error terms $\boldsymbol{u}$ and ν.

[21] It is the covariance matrix of observed variables $(\boldsymbol{y}\ \boldsymbol{x})$ written as a function of model parameters $\boldsymbol{\Sigma}(\boldsymbol{\theta})$.

hence they suffer from the problem of measurement error. Hence the OLS estimation with all these indicators may give biased and inconsistent estimates. Most importantly, norms are dynamic in nature; which imply that there are various structural factors which may affect the strength of 'norm'. It is difficult to distinguish the effect of such structural factors on norm. Given this, it would be more appropriate to model the social norm as latent variable in the MIMIC model framework.

5.2 Maximum-Likelihood estimates

Table 4 presents the Maximum Likelihood (ML) estimates of the model described in section 4. For the model identification, the coefficient of indicator *Village* is set equal to one. This identifying restriction helps not only in identifying the model parameters but also helps in providing the scale to the latent variable. The scale of the latent variable is always indeterminate a priori, but it is set by the identifying restrictions. Previous studies have shown that marriage exogamy is associated with an unfavourable norm for education. Thus the higher value of the latent variable $Norm^*$ measures more unfavourable norms towards education. In other words, a higher value of $Norm^*$ indicates the presence of stricter norms for education and a lower value indicates that the norms are relatively more favourable for education.

We will first discuss the results of the measurement model presented in panel B of Table 4. The coefficient with indicator *Village* is set equal to one as identifying restriction. The coefficient attached to each indicator in the measurement model measures the amount of variability of each indicator explained by the latent variable $Norm^*$. All the coefficients of the indicators are significant and are in the same direction, which shows that the social norm significantly impact these practices.

Moving to the results of structural equations presented in panel A, it would be interesting to see the factors that lead to the change in norms towards education. The education of father and mother has two effects on education; one is indirect and other is direct. The results reported in column (4) shows that the education of father and mother plays a significant role in shaping the norms. The positive sign with father's education implies that an increase in the education of father makes the norms stricter for education.

The negative sign with mother's education shows that an increase in the education of mother makes the norms more favourable for education. In other words, the education of mother has a norm-breaking effect whereas the education of father has a norm-binding effect on the educational outcomes for females. Two observations are noteworthy. First, the absolute size of the education of father's coefficient (0.012) is approximately half the absolute size of the mother's education (0.030). Second, the years of schooling of the mother has a U-shape effect on education norms and its maximal liberal effect reaches at 15 years.

The education of father and mother education also has a direct positive and significant effect on women's education. An additional year of increase in the education level of father and mother increases the women's education by 0.11 and 0.05 years respectively. However, the total effect of these two variables is positive and significant. The total effect of an additional year of schooling of father and mother is 0.10 and 0.06 years respectively. This implies that parents' education is an important determinant for women's years of schooling. However, the education of father has a much larger effect on women's education.

Moving to the results on average years of schooling, our key variable of interest is the norm. The coefficient of the norm is statistically significant and negative. This shows that as we move towards the more unfavourable norm, females' years of schooling decreases. This shows that the prevalent norms have a detrimental effect on the female educational outcomes.

The number of siblings is negatively associated with the females' years of schooling. If there are more children in the family then the family resources get distributed, leaving fewer resources for each child. As a result, this could adversely affect women's education. Furthermore, the siblings' composition also has implications for females' education. Having more number of sisters has a more negative effect on her education, than having more number of brothers. This could be because when parents have more daughters then they would save more for their daughters' marriage and investing less in their education. However, it is interesting to note that not only the composition of siblings but the education of siblings also influences female education. The education of brother and sister has a differential effect on females' years of schooling. An increase in the education

of sister has a larger effect on females' years of schooling than the education of brother.

5.3 Heterogeneity by place of residence and social groups

Table 5 presents the estimates for years of schooling for women by rural and urban region separately. These results show that the effect of norms varies across regions. Norms in the rural region appear to be stronger than in urban regions. We find that unfavourable norm around education has a larger negative impact on years for schooling for women in rural areas, compared to women in urban areas. It corroborates with the fact that urbanisation contributes to the weakening of the norm.

India has a history of the presence of norms based on caste, religion, gender and ethnicity. Different social groups may follow different norms and hence norms could have differential influence across different social groups. Table 6 presents the estimates for years of schooling for females by different social groups. We find that the effect of the norm on females' years of schooling is highest among *Brahmins*, followed by OBCs, *Dalits* and Forward castes. However, we do not find any significant effect of norms on females' education among *Adivasis* and Muslims.

Figure 5 presents the estimates for years of schooling for women for different social groups in the rural area and urban area separately. We find that in rural areas, the norms have maximum influence on *Brahmins*, followed by OBCs, Forward Castes, *Adivasis* and *Dalits*. However in the urban areas, the effect of norms is more pronounced only for OBCs and *Dalits*. We do not find any significant effect of norms among Muslims.

6 Conclusion

This paper attempts to explore the relationship between norms and female education by measuring the norm as a latent variable. We show that the MIMIC model is an appropriate framework for quantifying the effect of the norm. The MIMIC model allows us to address three important shortcomings in the existing literature on the relationship between social norms and female educational outcomes, viz., (i) that norms are unob-

servable; (ii) they are not monolithic but are indicated by several indicators of social practices each of which is an imperfect measure of the underlying norm; and (iii) norms are dynamic and can be affected by various factors. The last of these, viz., the identification of factors that cause social norms to change is, we believe, probably unique to our study. It has allowed us to identify factors that weaken or bind the social norms around education.

Using the Indian Human Development Survey (IHDS) 2011-12 data, we find that the norms have a significant effect on females' educational attainment. The effect is more pronounced in rural region across all social groups, whereas it is significant only amongst OBCs and *Dalits* in urban region. However, we do not find any significant effect of norms among Muslims. We also find that father's education has a norm-binding effect on female education, whereas mother's education has a norm-breaking effect on female education. However, the total effect of father's education on female education is positive and much larger than total effect of mother's education. An important policy implication drawn from this analysis is that increasing female education has a virtuous effect; increasing female education makes the norms weaker for the next-generation. By educating girls, the effect of norms that do not favour educating girls can be weakened.

References

Alañón*, A. and Gómez-Antonio, M. (2005). Estimating the size of the shadow economy in Spain: A structural model with latent variables. *Applied Economics*, 37(9):1011–1025.

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

Bano, S. (2012). Reform and identity: Purdah in muslim women's education in Aligarh in the early twentieth century. In *Proceedings of the Indian History Congress*, volume 73, pages 607–614.

Behrman, J. R., Foster, A. D., Rosenweig, M. R., and Vashishtha, P. (1999). Women's schooling, home teaching, and economic growth. *Journal of Political Economy*, 107(4):682–714.

Bicchieri, C. (2016). *Norms in the wild: How to diagnose, measure, and change social norms*. Oxford University Press, New York.

Bollen, K. A. (1989). *Structural Equations with Latent Variables*. Wiley, New York.

Desai, S., Vanneman, R., and National Council of Applied Economic Research, N. D. (2015). India Human Development Survey-II (IHDS-II), 2011-12. *ICPSR36151-v2. Ann Arbor, MI: Inter-university Consortium for Political and Social Research.*

Dostie, B. and Jayaraman, R. (2006). Determinants of school enrollment in Indian villages. *Economic Development and Cultural Change*, 54(2):405–421.

Duraisamy, P. (1992). Gender, intrafamily allocation of resources and child schooling in South India. Discussion Paper No.675, Economic Growth Centre, Yale University.

Dyson, T. and Moore, M. (1983). On kinship structure, female autonomy, and demographic behavior in India. *Population and Development Review*, 9(1):35–60.

Eswaran, M. (2014). *Why gender matters in economics*. Princeton University Press, New Jersey.

Field, E. and Ambrus, A. (2008). Early marriage, age of menarche, and female schooling attainment in Bangladesh. *Journal of Political Economy*, 116(5):881–930.

Gandhi Kingdon, G. (2002). The gender gap in educational attainment in India: How much can be explained? *Journal of Development Studies*, 39(2):25–53.

Gueye, A. S., Audibert, M., and Delaunay, V. (2018). Can social groups impact schooling decisions? evidence from castes in rural Senegal. *World Development*, 110:307–323.

Jöreskog, K. G. and Goldberger, A. S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association*, 70(351):631–639.

Klasen, S. (2002). Low schooling for girls, slower growth for all? cross-country evidence on the effect of gender inequality in education on economic development. *The World Bank Economic Review*, 16(3):345–373.

Knowles, S., Lorgelly, P. K., and Owen, P. D. (2002). Are educational gender gaps a brake on economic development? some cross-country empirical evidence. *Oxford Economic Papers*, 54(1):118–149.

Lahiri, S. and Self, S. (2007). Gender bias in education: The role of inter-household externality, dowry and other social institutions. *Review of Development Economics*, 11(4):591–606.

Lee, J. S., Bailey, M. J., Jeong, J., Mauldin, P. D., and Weintraub, W. S. (1997). A study on the cost-effectiveness of coronary revascularization: Introducing the simultaneous mimic health status model. *Health Economics*, 6(6):613–623.

Maertens, A. (2013). Social norms and aspirations: Age of marriage and education in rural India. *World Development*, 47:1–15.

Murthi, M., Guio, A.-C., and Dreze, J. (1995). Mortality, fertility, and gender bias in India: A district-level analysis. *Population and Development Review*, 21(4):745–782.

Rammohan, A. and Robertson, P. (2012). Do kinship norms influence female education? evidence from Indonesia. *Oxford Development Studies*, 40(3):283–304.

Rammohan, A. and Vu, P. (2018). Gender inequality in education and kinship norms in India. *Feminist Economics*, 24(1):142–167.

Richards, T. J. and Jeffrey, S. R. (2000). Efficiency and economic performance: An application of the MIMIC model. *Journal of Agricultural and Resource Economics*, 25(1):232–251.

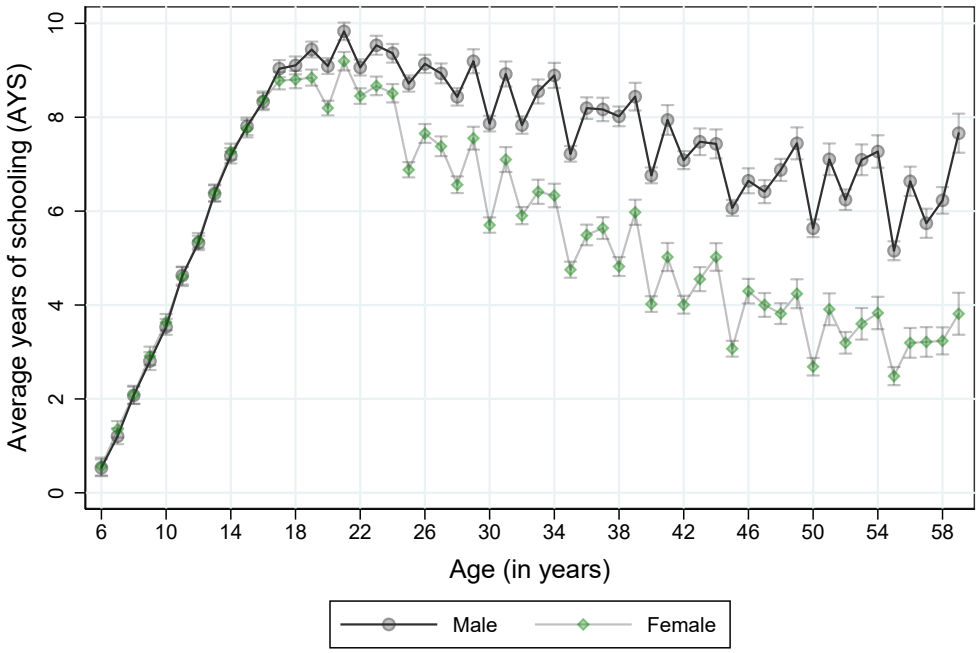
Schultz, T. P. (2002). Why governments should invest more to educate girls. *World Development*, 30(2):207–225.

Sundaram, A. and Vanneman, R. (2008). Gender differentials in literacy in India: The intriguing relationship with women's labor force participation. *World Development*, 36(1):128–143.

Thomas, D. (1994). Like father, like son; like mother, like daughter: Parental resources and child height. *Journal of Human Resources*, 29(4):950–988.

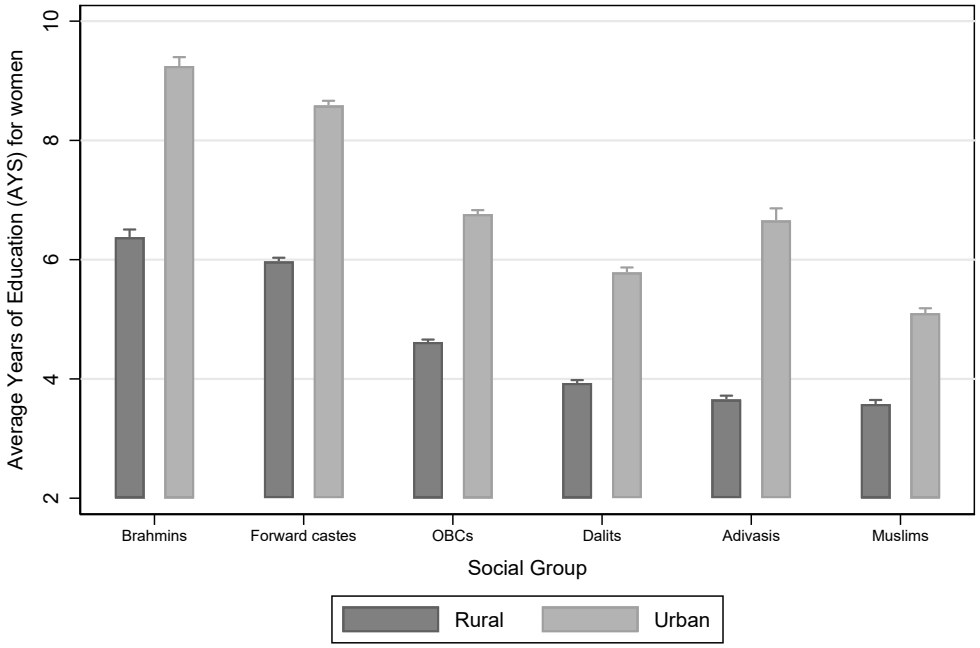
Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press, Cambridge.

Figure 1: Years of schooling for males and females



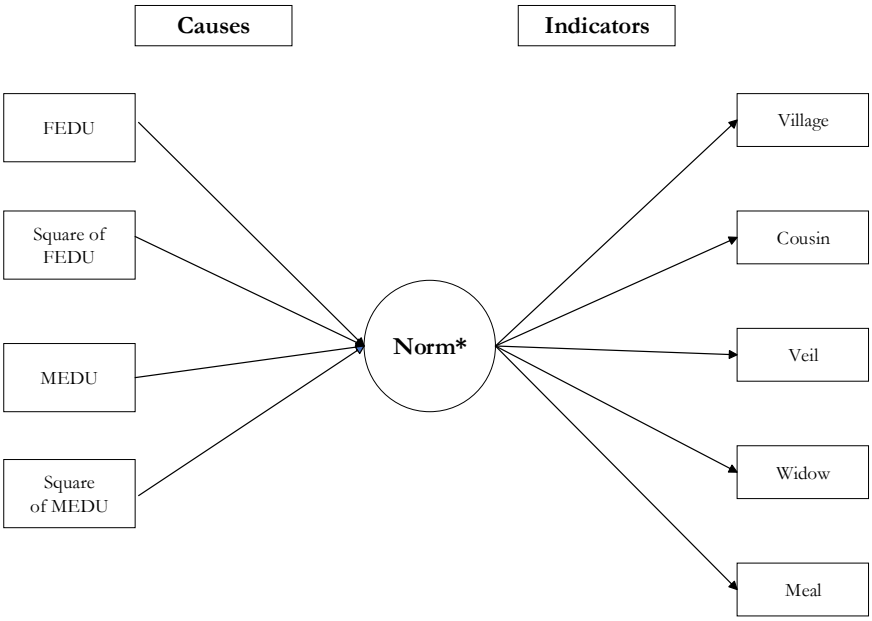
Note: This figure shows the average years of schooling for males and females between the ages of 6 to 59 years in India.

Figure 2: Average years of schooling for females



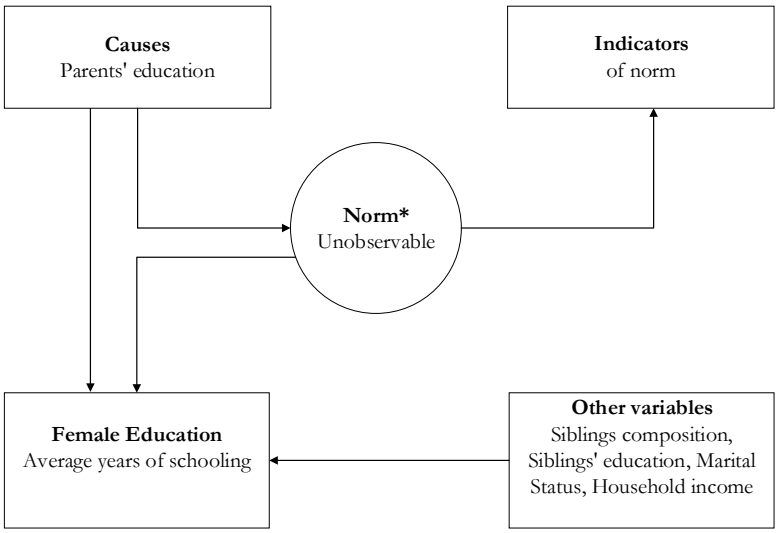
Note: This figure shows the average years of schooling for females (between the ages of 6 to 59 years) for different social categories by rural and urban region in India.

Figure 3: A pictorial description of the MIMIC model for latent variable Norm



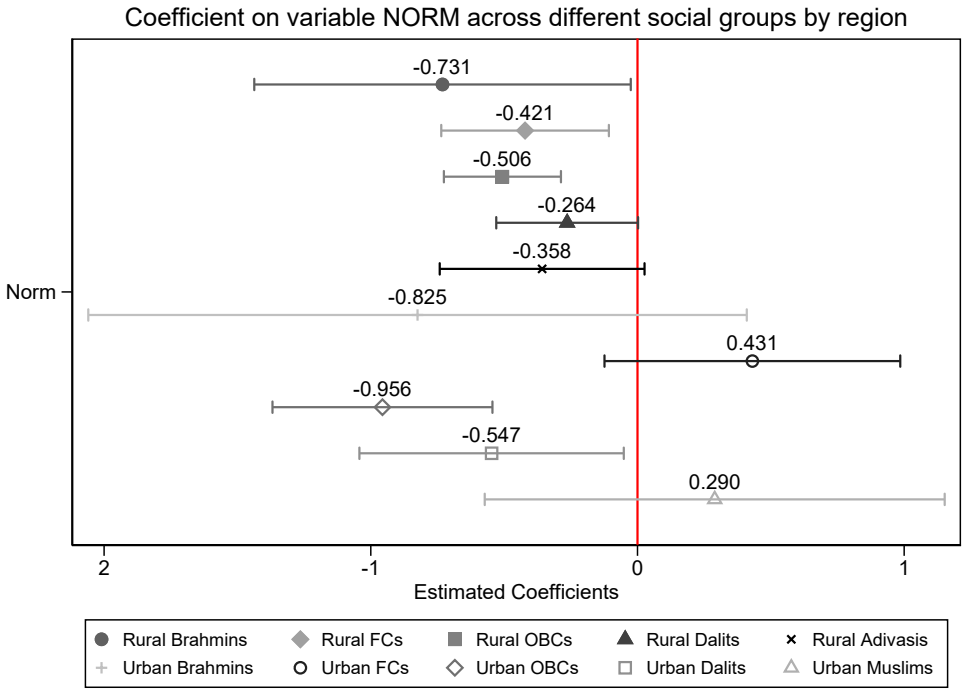
Note: The figure depicts the MIMIC model used for measuring social norms. The variables in the rectangular box represent the observed variables and the variable in the circle represents the latent variable. The variables on the left side in the rectangular box are the causes of the latent variable and the variables on the right side in the rectangular box are the indicators of the same latent variable. The description of the variables is given in Table 2.

Figure 4: A pictorial description of the Complete Model



Note: The figure depicts the complete model used for analysing the effect of norms on females' average years of education. The variables in the rectangular box represent the observed variables and the variable in the circle represents the latent variable. The description of the variables is given in Table 2.

Figure 5: Heterogeneity in the effect of Norm on female educational outcomes



Note: The figure shows the estimated coefficients on the latent variable Norm from the model described in equations (1), (4) and (5) estimated for different social groups separately by region. The horizontal line with cap around the coefficients represents the 95% confidence intervals.

Table 1: Literary Rate in selected regions

	2011		2018	
Region/Income group	Males	Females	Males	Females
Least developed countries: UN classification	65.96	49.71	71.80	58.10
South Asia	75.84	56.84	80.31	63.73
India	78.88	59.28	82.37	65.79
Bangladesh	62.48	55.12	76.67	71.18
Nepal	71.71	48.84	78.59	59.72
Pakistan	66.99	41.98	71.12	46.47
Sri Lanka	92.58	89.96	92.77	90.80
East Asia & Pacific	96.28	91.85	97.25	93.94
China	97.48	92.71	98.47	95.16
Europe & Central Asia	98.58	97.26	98.83	97.88
Middle East & North Africa	84.29	69.60	85.07	72.34
Sub-Saharan Africa	69.60	53.08	72.55	58.80
Latin America & Caribbean	92.86	91.64	94.31	93.45

Note: The table shows the literacy rate of adults aged 15 years or older. Figures presented for Pakistan and Sri Lanka for year 2011 correspond to year 2010. Figures presented for China for the year 2018 correspond to year 2017.
Source: Education Statistics, The World Bank Database

Table 2: Variables Description and Descriptive Statistics

Variables	Description	Mean	Standard deviation
Dependent Variable			
AYS	Women's years of schooling in single years	5.263	4.678
Explanatory variables			
FEDU	Father's education in single years	3.984	4.625
MEDU	Mother's education in single years	1.823	3.326
BRO	Proportion (%) of brothers among all siblings	0.386	0.157
BEDU	Highest level of education among all brothers in single years	7.630	4.946
SIS	Proportion (%) of sisters among all siblings	0.383	0.166
SEDU	Highest level of education among all sisters in single years	5.827	4.964
Age	Age of the female in single years	30.310	13.230
Ever Married	Dummy whether the given female is married, divorced or separated	0.742	0.438
Consumption	Logarithm of per-capita household expenditure	9.816	0.662
Assets	Number of assets in the household	15.440	6.431
Indicators			
Village	Indicator whether not permissible to marry daughter in natal village	0.486	0.500
Cousin	Indicator whether not permissible to marry with cousin	0.623	0.485
Widow	Indicator whether not permissible widow remarriage	0.327	0.469
Veil	Indicator whether women practice *purdah/pallu/burkha*	0.640	0.480
Meal	Indicator whether women eat meal after men	0.281	0.449
	Observations	42322	

Note: The table shows the mean and standard deviation of the key variables for women aged 6-59 years. For calculating proportion of brothers (sisters) for female is calculated using number of brothers (sisters) among all siblings divided by total number of siblings (including her).

Table 3: OLS estimates for women's years of schooling

Dependent Variable: AYS		
Indicators		
Village	-0.138***	(0.037)
Cousin	-0.007	(0.038)
Widow	0.092***	(0.031)
Veil	-0.291***	(0.034)
Meal	0.003	(0.034)
Other Explanatory Variables		
Age	0.339***	(0.008)
Age Squared	-0.006***	(0.000)
FEDU	0.112***	(0.005)
MEDU	0.051***	(0.006)
BRO	-6.174***	(0.197)
BRO*BEDU	0.407***	(0.009)
SIS	-6.883***	(0.178)
SIS*SEDU	0.653***	(0.009)
Married	-2.165***	(0.069)
Consumption	0.358***	(0.034)
Assets	0.159***	(0.004)
Observations	42322	

Note: The table shows the OLS estimates of equation (1), along with five indicators as additional explanatory variables. The dependent variable is female's years of schooling ages between 6-59 years. The description of the variables is given in Table 2. The coefficient of the constant has been suppressed. The regression specification controls for the mean values of all the explanatory variables created at the PSU (primary sampling unit) level to account for the unobserved community-level effects, based on the Chamberlin-Mundlak approach. Standard errors are reported in the parenthesis. ***, ** and * represents the significance at 1%, 5% and 10% respectively.

Table 4: ML estimates for women's years of schooling and norm (N=42322)

Panel A: Structural Equations		
Dependent Variable	**Norm**	**AYS**
Norm		-0.263***
		(0.047)
Age		0.340***
		(0.008)
Age Squared		-0.006***
		(0.000)
FEDU	0.012***	0.114***
	(0.001)	(0.005)
Square of FEDU	0.000**	
	(0.000)	
MEDU	-0.030***	0.052***
	(0.002)	(0.006)
Square of MEDU	0.001***	
	(0.000)	
BRO		-6.219***
		(0.197)
BRO*BEDU		0.407***
		(0.009)
SIS		-6.921***
		(0.178)
SIS*SEDU		0.654***
		(0.009)
Ever Married		-2.171***
		(0.069)
Consumption		0.361***
		(0.034)
Assets		0.159***
		(0.004)
Panel B: Measurement Model		
Indicators		
Village	1.000	
Cousin	0.959***	(0.016)
Widow	0.028***	(0.007)
Veil	0.308***	(0.007)
Meal	0.186***	(0.007)

Note: The table shows the estimates of equations (1), (4) and (5). The description of the variables is given in Table 2. The coefficient of the constant has been suppressed. The regression control for the mean values of all the explanatory variables created at the PSU (primary sampling unit) level to account for the unobserved community-level effects, based on the Chamberlin-Mundlak approach. The coefficients in the measurement model show the factor-loadings to the latent variable Norm. Standard errors are reported in the parenthesis. ***, ** and * represents the significance at 1%, 5% and 10% respectively.

Table 5: Estimates for women's years of schooling and norm by region

	(1) Rural	(2) Urban
Dependent Variable: AYS		
Norm	-0.248***	-0.210**
	(0.054)	(0.100)
FEDU	0.117***	0.099***
	(0.006)	(0.008)
MEDU	0.035***	0.080***
	(0.008)	(0.009)
Age	0.332***	0.396***
	(0.010)	(0.015)
Age Squared	-0.006***	-0.006***
	(0.000)	(0.000)
BRO	-5.669***	-7.412***
	(0.233)	(0.360)
BRO*BEDU	0.351***	0.507***
	(0.011)	(0.017)
SIS	-6.149***	-8.589***
	(0.210)	(0.328)
SIS*SEDU	0.592***	0.748***
	(0.011)	(0.016)
Ever Married	-2.248***	-2.097***
	(0.082)	(0.124)
Consumption	0.290***	0.482***
	(0.040)	(0.061)
Assets	0.160***	0.160***
	(0.005)	(0.008)
Dependent Variable: Norm		
FEDU	0.012***	0.014***
	(0.002)	(0.002)
Square of FEDU	0.001***	-0.000
	(0.000)	(0.000)
MEDU	-0.037***	-0.010***
	(0.003)	(0.002)
Square of MEDU	0.002***	0.001***
	(0.000)	(0.000)
Observations	28497	13825

Note: The table shows the estimates of equations (1), (4) and (5) for rural and urban region separately.The estimates are based on the all the five indicators. The results of the measurement model have been suppressed. The description of the all the variable is given in Table 2. The regression specification in each column controls for the mean values of all the explanatory variables created at the PSU (primary sampling unit) level to account for the unobserved community-level effects, based on the Chamberlin-Mundlak approach. Standard errors are reported in the parenthesis. ***, ** and * represents the significance at 1%, 5% and 10% respectively.

Table 6: Estimates for women's years of schooling and norm by social groups

	(1) Brahmins	(2) Forward Castes	(3) OBCs	(4) *Dalits*	(5) *Adivasis*	(6) Muslims
Dependent Variable: AYS						
Norm	-0.823***	-0.349**	-0.685***	-0.419***	-0.227	-0.163
	(0.314)	(0.143)	(0.098)	(0.121)	(0.196)	(0.253)
FEDU	0.156***	0.105***	0.119***	0.094***	0.094***	0.101***
	(0.019)	(0.011)	(0.008)	(0.010)	(0.018)	(0.011)
MEDU	0.049**	0.067***	0.045***	0.051***	-0.000	0.039**
	(0.021)	(0.013)	(0.011)	(0.014)	(0.026)	(0.016)
Age	0.383***	0.288***	0.316***	0.390***	0.286***	0.413***
	(0.042)	(0.023)	(0.014)	(0.017)	(0.027)	(0.018)
Age Squared	-0.006***	-0.005***	-0.006***	-0.007***	-0.005***	-0.007***
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
BRO	-7.858***	-7.651***	-6.266***	-5.605***	-5.017***	-5.050***
	(1.065)	(0.520)	(0.334)	(0.418)	(0.634)	(0.511)
BRO*BEDU	0.466***	0.501***	0.398***	0.365***	0.330***	0.376***
	(0.053)	(0.026)	(0.016)	(0.019)	(0.029)	(0.021)
SIS	-9.817***	-8.227***	-7.042***	-6.052***	-5.158***	-5.812***
	(0.900)	(0.446)	(0.297)	(0.377)	(0.583)	(0.494)
SIS*SEDU	0.811***	0.706***	0.652***	0.617***	0.583***	0.565***
	(0.043)	(0.023)	(0.016)	(0.019)	(0.032)	(0.023)
Ever Married	-1.831***	-1.287***	-2.136***	-2.657***	-2.411***	-2.606***
	(0.346)	(0.192)	(0.124)	(0.141)	(0.226)	(0.153)
Consumption	0.598***	0.225***	0.239***	0.532***	0.600***	0.265***
	(0.143)	(0.082)	(0.056)	(0.075)	(0.120)	(0.089)
Assets	0.133***	0.174***	0.161***	0.149***	0.158***	0.145***
	(0.020)	(0.011)	(0.007)	(0.009)	(0.015)	(0.011)
Dependent Variable: Norm						
FEDU	0.008	0.006*	0.014***	0.019***	0.006	0.001
	(0.005)	(0.003)	(0.002)	(0.003)	(0.006)	(0.002)
Square of FEDU	0.000	0.001**	0.000	-0.000	-0.000	-0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
MEDU	-0.024***	-0.019***	-0.039***	-0.028***	-0.049***	0.007***
	(0.006)	(0.004)	(0.003)	(0.005)	(0.010)	(0.002)
Square of MEDU	0.001	0.001**	0.002***	0.001	0.004***	-0.000**
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)
Observations	2012	7064	14195	9257	3449	6345

Note: The table shows the estimates of equations (1), (4) and (5) for different social groups separately. The estimates are based on the all the five indicators. The results of the measurement model have been suppressed. The description of the all the variable is given in Table 2. The regression specification in each column controls for the mean values of all the explanatory variables created at the PSU (primary sampling unit) level to account for the unobserved community-level effects, based on the Chamberlin-Mundlak approach. OBCs stand for Other Backward Classes. Standard errors are reported in the parenthesis. ***, ** and * represents the significance at 1%, 5% and 10% respectively.

Appendix

Table A1: Definitions of Indicators used in the Measurement Model and their coding

Norm	Indicator	Description	Coding
Village Exogamy	Village	Is it permissible to marry a girl in her natal village in your community/*jati*?	Yes = 0 No = 1
Kinship	Cousin	Is it permissible to marry a girl to her cousin in your community/*jati*?	Yes = 0 No = 1
Marriage and Conservatism	Widow	Is it permissible for a widow to remarry in your community/*jati*?	Yes = 0 No = 1
Honour and Purity	Veil	Do you practice ghungat/burkha/purdah/pallu?	No = 0 Yes = 1
Traditional gender role	Meal	When your family takes the main meal, do women or usually eat with the men eat first by themselves? Or do men eat first?	Eat together or Varies or Women first = 0 Men first = 0