



Structural Equation Approach to Modeling Social Norms in Women's Education: A Case Study of India

Tanu Gupta¹ · A. Ganesh Kumar²

Accepted: 17 October 2024 / Published online: 9 November 2024
© The Author(s), under exclusive licence to The Indian Econometric Society 2024

Abstract

Studies on women's education proxying social norms through indicators of social practice overlook three important characteristics of norms, viz., they are latent, multifaceted, and influenced by exogenous factors. To address this, we propose using the MIMIC model in a structural equation framework. This approach allows incorporating multiple social practices each of which could be imperfect manifestations of an underlying norm. Besides, it allows us to identify exogenous factors that could bring about a change in the norms. Applying it to India, we find that norms adversely affect women's education and that the mother's education has a norm-breaking effect while the father's education is norm-binding.

Keywords Social norms · Women's education · Structural equation · MIMIC model · India

JEL Classification C30 · I21 · I24 · Z1

Introduction

The role of social norms in shaping women's outcomes has been a subject of growing interest. While previous research has highlighted the importance of economic factors in determining women's outcomes, there is an increasing recognition of the influence of cultural and social norms, particularly in societies where traditional values are deeply entrenched (Jayachandran 2015). However, measuring social norms is not easy because of its latent nature. This paper uses a new approach, namely, the

✉ Tanu Gupta
tanug@isid.ac.in

A. Ganesh Kumar
agk@igidr.ac.in

¹ Indian Statistical Institute (ISI), New Delhi, India

² Indira Gandhi Institute of Development Research (IGIDR), Mumbai, India

Multiple Indicators and Multiple Causes (MIMIC) model, to measure the latent construct of norms and examine its impact on women's education in India. While the literature on education has extensively examined supply- and demand-side factors, such as expected returns from schooling, household income, and parental education (Psacharopoulos and Yang 1991; Wilson 2001; Connelly and Zheng 2003; Nakajima et al 2018, Drèze and Kingdon 2001), the role of social norms has received increasing attention (Dostie and Jayaraman 2006; Lahiri and Self 2007; Gueye et al. 2018).

Social norms are the beliefs, ideologies, and informal rules rooted in the history, traditions, and culture of a society that describes how an individual should behave in a group, community, or society as a whole. Such informal rules could be different based on gender, caste, or religion. Norms determine individuals' behavior by shaping their views about expectations, aspirations, and identities and thereby exert a powerful influence on their choices and outcomes. These norms, which prescribe gender roles and expectations, can create identity-based incentives or disincentives for educational attainment, particularly for women. In this context, norms can either constrain or promote educational opportunities based on how they align with the socially constructed identities of women within a community (Akerlof and Kranton 2000).

A few recent economic studies have analyzed the effect of various social norms on educational outcomes for different countries empirically by constructing different indicators. For instance, Bertrand and Pan (2013) examine how social norms around masculinity and femininity affect boys' and girls' behavior in schools in the United States, contributing to the gender gap in educational outcomes. Field and Ambrus (2008) study the impact of norms around early marriage on women's education in Bangladesh by using the 'age of menarche' as an instrument to measure the incidence of early marriage, whereas Maertens (2013) surveys parents in India about the 'ideal age of marriage' for their children. Ashraf et al. (2020) focus on norms relating to dowry in Indonesia and Zambia, and use Ethnographic Atlas database to categorize the ethnic groups based on practices related to bride price. Dyson and Moore (1983) measure norms relating to women's honor and purity in India by the percentage of women practicing *purdah*. Sundaram and Vanneman (2008) use log odds for women who have migrated from their birthplace to the log odds for men as the measure of norms relating to patrilocal exogamy in India, while Rammohan and Vu (2018) use the proportion of women who are not living in their natal district as the measure. Rammohan and Robertson (2012) examine the role of kinship and patrilineal norms on women's education outcomes in Indonesia by capturing inheritance and post-marital residence practices at the community level. Dostie and Jayaraman (2006) and Gueye et al. (2018) construct indicators based on caste composition in the village to capture the caste-specific norms at the village level for rural India and rural Senegal, respectively.

All these studies typically estimate a linear regression model of women's education outcomes on their specific indicator of 'norms', and a few control variables. These studies, however, suffer methodological shortcomings as they have overlooked three important aspects related to norms. First, we do not observe norms directly, but we may be able to observe only practices that are manifestations of the norms. Past studies have considered observed social practices related to norms (such

as wearing *purdah*, post-marital migration of women) where data on such practices exists, or have used other non-behavioral information (such as population composition, perceptions of ideal age of marriage) as proxies for norms rather than the norms themselves in their analysis. Using such information, however, may not be appropriate if these practices or non-behavioral data are imperfect measures of the underlying norms. That is, there may not be a one-to-one correspondence between norms and behavioral indicators, and their use in the regression analysis creates an errors-in-variable problem.

Second, the use of behavioral indicators is even more problematic because norms are not monolithic but are often multifaceted. That is, they manifest as more than one social practice. By using a single behavioral indicator, existing studies have focused only on one dimension and overlooked the multifaceted nature of norms. Such an analysis involves a strong assumption that all the underlying behavioral indicators are highly correlated with each other. However, when the multiple behavioral indicators are not strongly correlated, then the use of a single behavioral indicator could be misleading and result in biased estimates. This could be even more problematic when that indicator is not strongly correlated with the outcome variables of interest.¹ Third, norms are influenced by many other observable factors. Various factors, such as parents' education and policy intervention, may play a significant role in determining norms.

Against this background, this paper takes a new approach to model the relationship between social norms and women's education outcomes that address the above-mentioned concerns in conceptualizing social norms. 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 latent variables such as black economy, health, efficiency, fairness, economic performance, women's empowerment and so on (Van de Ven and Van Der Gaag 1982; Titman and Wessels 1988; Richards and Jeffrey 2000; Alañón and Gómez-Antonio 2005; Zereyesus et al. 2017; Ballon 2018; Omura 2019). However, to the best of our knowledge, this modeling framework has not been used to study social norms despite its advantages.

The MIMIC model allows for measuring latent constructs (like social norms) that are not directly observable, using multiple indicators to capture the underlying concept more accurately. It accounts for measurement errors in the indicators, providing more reliable estimates of the relationship between the latent construct and the outcome variables. It can incorporate multiple causes and multiple indicators of the latent variable, capturing its multidimensional nature. Additionally, MIMIC models also allow for the inclusion of exogenous variables that can explain changes in the

¹ For instance, it is conceivable that some 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 just 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 may be forced into child marriage, but nevertheless, they complete their education after marriage. On the other hand, certain families may not evince any such practices but may not educate their girl child.

latent construct, facilitating a more nuanced understanding of how these exogenous variables contribute to the underlying latent variable and how this latent variable, in turn, influences the outcome variable of interest.

We demonstrate the use of the MIMIC framework for India using data from the Indian Human Development Survey-II (IHDS-II) conducted in 2011–2012 (Desai et al. 2015). This is a nationally representative multi-topic survey that provides detailed individual and household-level information on education, gender relations, marriage practices, occupation, economic status, health, fertility, landholding, social capital, and social identity. A unique feature of this survey is that it provides information on the multifaceted behavioral aspects of norms that can be used directly to model norms as a latent variable.

We find that the norms are indeed multifaceted and have a significant adverse effect on women's educational attainment. The effect is stronger in the rural region than in the urban region. The effect of norms varies across social groups. For instance, the effect is more pronounced among Other Backward Caste (OBC), followed by *Dalit* and *Brahmin*. However, we do not find any significant effect of norms among Forward Caste, *Adivasi*, and Muslim.

Furthermore, we find that the education of parents plays a significant role in shaping the norms around women's education. The education of both parents also has a direct positive effect on women's 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 women through the change in norms by making them more favorable or unfavorable for women's education, and the other is the more direct effect, which is the effect of parents' unobserved family background. We find that the education of the father has a norm-binding effect, whereas the education of the mother has a norm-breaking effect on women's education. However, the total effect of the education of both parents is positive, with the effect of the mother's education being much larger.

The rest of the paper is organized as follows. “[Data and Descriptive Statistics](#)” describes the data descriptive statistics. “[Empirical Model](#)” describes the modeling of the effect of the norms on women's education in the MIMIC framework and presents the complete model, along with its advantages. Results are presented in “[Results](#)”, and the last section concludes.

Data and Descriptive Statistics

Data

Average Years of Schooling

We use data from the Indian Human Development Survey-II (IHDS-2) 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 island territories Andaman & Nicobar and Lakshadweep. The surveyed households are spread across 1503 villages and 971 urban neighborhoods. The rural sample was drawn

using stratified random sampling. In urban regions, the towns and cities were first 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 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 allow us to identify and capture the norms directly at the household level that could influence the educational outcomes for women.

Our sample consists of married women aged 15–49 years old. We use the ‘eligible women’ module, which records detailed information on her education, fertility, birth history, gender relations, health beliefs, and natal family’s education history. The primary dependent variable is the Average Years of Schooling (AYS), defined as the number of years of education a woman has completed at the time of the survey.

To model women’s educational attainment, we consider a range of individual- and household-level characteristics as well as the social norms prevailing within the household. At the household level, we include variables such as the education level of the father and mother, the proportion of brothers and sisters among all the siblings (separately), the logarithm of per-capita household expenditure, land ownership, and a count of the number of assets owned by the household.² We proxy a household’s income by its per-capita consumption expenditure as the dataset does not report income.

The education of the father and mother has been included as explanatory variables to capture the effects 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, which may impact the educational outcomes of all the children, especially for the girls, who are usually discriminated against the boys while making such investment decisions. However, not only the number but also the composition of the siblings may have a differential effect on women’s education. The explanatory variables, such as the proportion of brothers and sisters among all the siblings, would capture these differential effects. Finally, we control for variations in educational attainment among women across different states, regions, religions, and caste groups to account for broader socio-cultural and geographic factors that could influence educational outcomes.

Social Norms

The social norm is a latent explanatory variable denoted as *Norm**. We use the MIMIC approach to measure this latent variable. The MIMIC approach requires both cause and indicator variables to measure any latent variable. IHDS-2 contains a few questions that help us to identify various social practices or behaviors that are the manifestation of norms. These questions will serve as indicators for latent

² The dataset does not report the value of the assets.

variable norms. The detailed questions, the construction of social and behavioral indicators, and the norms that they represent are reported in Table 1.

Indicators

Norms around women's marriage play a significant role in shaping women's education, particularly in societies where traditional views on gender roles and family responsibilities are strong and deeply entrenched. These conservative norms often manifest in various marriage practices followed in the community. IHDS-2 asks questions associated with these marriage practices, and we examine specific indicators to measure the 'norm around marriage' in women's communities.³

For instance, the indicator *Exogamy* captures whether women are allowed to marry within their natal village, reflecting the 'norm of exogamy'. The practice of cross-cousin marriage (*Cousin*) is another important indicator, capturing a key aspect of the kinship norm related to marriage. Marriages between blood relatives often occur in communities with strong adherence to traditional cultural norms. Additionally, the practice of restriction on widow remarriage (*Widow*) is indicative of the community's attitudes towards women's marriage and its level of conservatism. The literature highlights that the communities where marriage norms are such that the practice of village exogamy, cross-cousin marriage, and restrictions on widow marriage are being followed are also the regions where attitudes towards women's education may not be encouraging (Dyson and Moore 1983). Thus, these three indicators serve as measures of the 'norm around women's marriage' prevalent in a given community.

Furthermore, the norm of purity often dictates that women and girls should be kept away from situations where their chastity might be "compromised." As a result, families may put restrictions on women's mobility and limit their interactions with males outside the family. To capture the 'norm of honor and purity', we use the indicator *Veil*, which measures the prevalence of the practice of *purdahlburqa* within a household.

In cultures where traditional gender roles are strongly enforced, education for women may be deemed unnecessary or less important. The norm that a woman's primary role is as a caregiver and homemaker often leads to the perception that formal education is irrelevant for girls, especially beyond basic literacy. To capture the 'norm around traditional gender-specific roles', we use the indicator *Gender role*. In families where marriages between blood relatives are common, they often emphasize traditional gender roles. These families may place a higher value on a woman's role as a wife and mother and follow strong patriarchal structures. The indicator '*Hus Related*', which measures whether a given woman's husband is related to her by blood, helps in identifying the underlying norm of cultural conservatism.

Finally, norms that give parents control over their daughters' education and marriage decisions often prioritize marriage over education. A woman's say in her own marriage decisions (measured by indicator *Marriage Say*), along with whether her husband belongs to the same caste as her (indicator *Hus Caste*), captures the underlying norm of parents' control.

³ The information on these practices is reported in the education and health modules under the section of 'marriage practices' in IHDS-2 questionnaire.

Table 1 Definitions of indicators used in the MIMIC model

Indicator	Survey question	Norm	Coding
Exogamy	Is it permissible to marry a girl in her natal village in your community/ <i>jati</i> ?	Marriage	Yes = 0; No = 1
Cousin	Is it permissible to marry a girl to her cousin in your community/ <i>jati</i> ?	Marriage	No = 0; Yes = 1
Widow	Is it permissible for a widow to remarry in your community/ <i>jati</i> ?	Marriage	Yes = 0; No = 1
Veil	Do you practice <i>ghungat/burkha/purdah/pallu</i> ?	Honor and Purity	No = 0; Yes = 1
Gender role	When your family takes the main meal, do women usually eat with the men or eat first by themselves? Or do men eat first?	Traditional gender role	Eat together or Varies or Women first = 0; Men first = 1
Hus related	Are you related to your husband by blood?	Traditional gender role	No relation or Not permitted = 0; Uncle, Cousin or Others = 1
Marriage say	Who chose your husband? Did you have any say in choosing your husband?	Parents' control	Respondent/her parents and she has some say = 0, Parents/ Others and No say = 1
Hus caste	Does your husband's family have the same caste as your natal family?	Parents' control	No = 0, Yes = 1

Source: Authors' tabulation

Causes

The advantage of the MIMIC framework is that it allows the inclusion of the variables that might influence norms. These variables are referred to as cause variables. In this paper, we include various household-level characteristics like the education of parents, the composition of siblings, exposure to media by male and female household members, the number of elderly males and females in the household, and land ownership. Education is considered to be a powerful tool that can influence norms, as it shapes individuals' perspectives and attitudes toward gender roles and the value of education. In many households, particularly in developing countries, the decision to invest in girls' education is often made by their parents. Therefore, it is reasonable to believe that the education of the previous generation (parents) may play a significant role in shaping the strength of norms,⁴ either reinforcing or challenging traditional views on women's education.⁵

The composition of siblings can also affect norms around education. For example, households with a higher proportion of male siblings may adhere more strongly to traditional gender roles, where educational resources are prioritized for boys over girls. On the other hand, households with a higher proportion of female siblings might face pressure to conform to societal expectations of early marriage and domestic roles, potentially limiting educational opportunities for girls.

Exposure to media by male and female household members can influence norms by exposing individuals to different perspectives, modern values, and the benefits of education, potentially leading to more progressive attitudes toward women's education. Media exposure can serve as a channel for disseminating information and ideas that challenge traditional norms, thus making them more favorable toward gender equality in education.

The presence of elderly males and females in the household might reinforce traditional norms, as older generations often hold more conservative views regarding gender roles and education. These individuals might exert influence over household decisions, including those related to the education of girls, based on their adherence to long-standing cultural norms.

Land ownership, as a proxy for wealth and economic stability, can also influence norms. Richer households often have more conservative views and stricter norms around women's mobility due to their greater emphasis on maintaining traditional values and social status. In many societies, wealthier families may prioritize preserving their lineage and social standing, which can lead to more

⁴ We do not claim that parents' education is not affected by "norms" in the areas. However, they are predetermined in our sample. In addition, we also include fixed effects for state, region, caste, religion, and year of birth to control for differences in norms across location and time that could determine parents' education.

⁵ It is conceivable that policy intervention can also have an impact on norms. For instance, laws mandating compulsory education up to a certain level can be a norm-breaker. In the absence of such laws, governments can also offer incentives for educating girls through cash transfers, fee subsidies, and providing free meals, and bicycles, and such programs can also influence the strength of norms. As the dataset that we use does not provide any information on such transfers, we are unable to account for this in the structural equation.

rigid adherence to gender roles. For instance, in these households, there may be a stronger preference for women to fulfill traditional roles as homemakers and caregivers, leading to less emphasis on formal education for women.

Additionally, different religions and caste groups might have different norms and traditional beliefs. Similarly, rural regions have different norms compared to urban regions. We, therefore, also control for religion, caste, region, and state dummies to account for these variations in norms across different demographic and geographic contexts.

Descriptive Statistics

From the descriptive statistics presented in Table 2, we find that the average years of schooling for women in India is only 5.4 years, which is equivalent to completing the primary level of education. While the average years of schooling for both the parents are below primary (5th grade); the average year of schooling for the father (3.6 years) is two years higher than that of the mother (1.5 years).

Figure 1 depicts the differences in the women's average years of schooling across different social groups in rural and urban regions. The entire sample has been divided into six social groups based on caste and religion. These six social groups are *Brahmin*, Forward Caste (FC), Other Backward Caste (OBC), *Dalit*, *Adivasi*, and Muslim. It is interesting to note that women in urban areas have significantly higher years of schooling than women in rural areas across all social groups. On average, there is a significant difference of 3.1 years of schooling between urban women and rural women. We observe from Fig. 1 that *Brahmin* women have the highest years of education, followed by Forward Caste, OBC, *Adivasi*, *Dalit*, and Muslim. This holds for both rural and urban areas, with the exception that *Adivasi* have lower years of schooling than *Dalit* and Muslim in rural areas.

We find that 46.6% and 62.2% of the women live in communities where it is *not* permissible to marry a girl in their natal village and to their cousin, whereas 34.1% of the women belong to communities where widow remarriage is *not* allowed (see Table 2). It is also noteworthy that approximately 59% of the women belong to households where they follow the custom of *purdah/burqa*. Furthermore, we find that in 41.3% of cases, women had *no* say at all in marriage decisions, whereas 94.6% of women reported having husbands of their own caste.

Table 3 presents the Tetrachoric correlation matrix between the indicator variables.⁶ For instance, the coefficient of -0.82 in entry (2, 1) shows a strong, significant inverse correlation between indicators *Exogamy* and *Cousin*. Whereas the coefficient in entry (3, 1) is 0.02 and significant, this shows a weak positive correlation between indicators *Exogamy* and *Widow*. Similarly, entry (3, 2) is -0.04 and significant, showing a weak negative correlation between indicators *Cousin* and *Widow*. Although all these three indicators are the manifested practices that are associated with the norm around women's marriage, the correlation among them is

⁶ Tetrachoric correlation is used to measure the correlation between two binary or dichotomous variables. This correlation between binary or dichotomous variables is a special case of Polychoric correlation.

Table 2 Variables description and descriptive statistics

Variable	Description	Mean	S.D
<i>Dependent variable</i>			
AYS	Women's years of schooling in single years	5.381	4.936
<i>Explanatory variables</i>			
FEDU	Father's education in single years	3.580	4.559
MEDU	Mother's education in single years	1.547	3.107
BRO	Proportion (%) of brothers among all siblings	0.396	0.201
SIS	Proportion (%) of sisters among all siblings	0.350	0.218
MELDER	Presence of elderly males in household	0.194	0.405
FELDER	Presence of elderly females in household	0.234	0.434
MMEDIA	Male exposure to media in household	0.842	0.364
FMEDIA	Female exposure to media in household	0.830	0.375
Consumption	Logarithm of per-capita household monthly expenditure	9.881	0.667
Assets	Number of assets in the household	15.84	6.463
Land	Dummy if household has landholding	0.462	0.499
<i>Indicators</i>			
Exogamy	Indicator whether <i>not</i> permissible to marry daughter in the natal village	0.466	0.499
Cousin	Indicator whether permissible to marry daughter with a cousin	0.377	0.485
Widow	Indicator whether <i>not</i> permissible widow remarriage	0.341	0.474
Veil	Indicator whether women practice <i>purdah/pallulburkha</i>	0.594	0.491
Gender role	Indicator of whether women eat a meal after men	0.262	0.440
Hus related	Indicator if husband is related by blood	0.0726	0.260
Marriage say	Indicator if the woman did not have any say in her marriage decision	0.413	0.492
Hus caste	Indicator if the husband has the same caste as hers	0.946	0.225
	Observations	34,067	

The table shows the mean and standard deviation of the key variables for women aged 15–49 years. The variable proportion of brothers (sisters) for a woman is calculated using the number of brothers (sisters) among all siblings divided by the total number of siblings (including her)

Source: Authors' estimates based on data from IHDS-2

not the same and is in different directions. This shows that even if individuals adhere to some specific norm, they may choose *not* to follow all the manifested practices. Similarly, indicators, *Hus Caste* and *Marriage Say*, manifest the norm around parents' control. However, the magnitude of the correlation between these indicators is very low (11.8).

Empirical Model

Average Years of Schooling

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

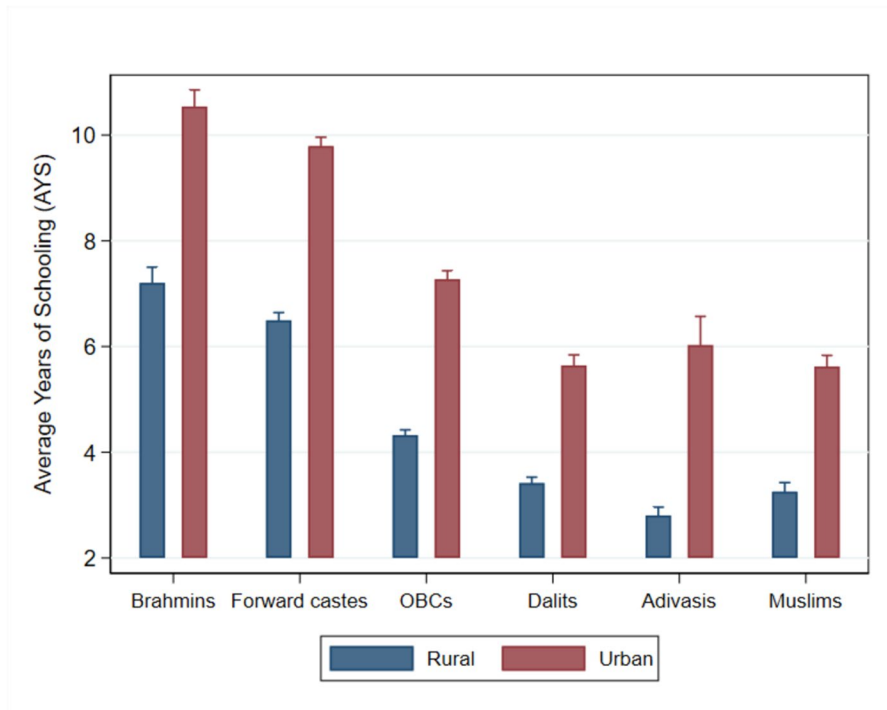


Fig. 1 Average years of schooling for women across different categories. This figure shows the average years of schooling for women in the estimable sample (between the ages of 15 and 49 years) for different social groups in rural and urban regions in India. Source: Authors' estimates based on data from IHDS-2

$$\begin{aligned}
 AYS_i = & \alpha_0 + \alpha_1 Norm^* + \alpha_2 FEDU_i + \alpha_3 MEDU_i + \alpha_4 BRO_i + \alpha_5 SIS_i + \alpha_6 Consumption_i \\
 & + \alpha_7 Assets_i + \alpha_8 Land_i + \alpha_9 Rural_i + \delta_c + \gamma_r + \mu_t + \tau_s + \epsilon_i
 \end{aligned} \quad (1)$$

where AYS_i denotes the average years of schooling of a given woman i . $FEDU_i$ and $MEDU_i$ represent the father's and mother's education, respectively, whereas BRO_i and SIS_i represent the proportion of brothers and sisters among all siblings. $Consumption_i$ is the logarithm of per-capita consumption expenditure, and $Assets_i$ indicates the number of assets owned by the household. $Land_i$ is an indicator variable if the household owns the land. Additionally, we account for heterogeneity in women's education across different regions, castes, and religions by including rural dummy ($Rural$) and religion (γ_r) and caste (δ_c) fixed-effects.

One important concern with this estimation strategy is the potential correlation between state-level unobserved factors and social norms, which could influence women's education. Some states in India might have introduced certain policies at the state level that directly or indirectly affect women's education. To account for this, we include state-fixed effects (τ_s), which controls for time-invariant unobserved heterogeneity across states. We also include year-of-birth fixed effects (μ_t) to control for differences in women's education levels due to differences in age. Additionally, these

Table 3 Tetrachoric correlation between the indicators

	Exogamy	Cousin	Widow	Veil	Gender role	Hus related	marriage say	Hus caste
Indicator 1: Exogamy	1							
Indicator 2: Cousin	-0.8158*	1						
Indicator 3: Widow	0.0220*	-0.0379*	1					
Indicator 4: Veil	0.2946*	-0.3087*	-0.1128*	1				
Indicator 5: Gender role	0.1807*	-0.1755*	0.0033	0.3154*	1			
Indicator 6: Hus Related	-0.2487*	0.4762*	0.1370*	-0.2627*	-0.1514*	1		
Indicator 7: Marriage Say	0.3245*	-0.2810*	-0.1579*	0.4651*	0.2857*	-0.2508*	1	
Indicator 8: Hus Caste	0.0477*	-0.0199	0.0462*	0.008	0.0724*	0.1081*	0.1175*	1

This table shows the tetrachoric correlation matrix between the indicators used in the measurement model. The details on this indicator are given in Table 1. * represents the significance at 5% level

Source: Authors' estimates based on data from IHDS-2

year-of-birth fixed effects help to account for women's exposure to varying education policies over time, ensuring that our estimates capture the true effect of social norms on women's education.

Norm in MIMIC Framework

Given that social norms are unobservable, we model them as latent variables using the MIMIC model given by Jöreskog and Goldberger (1975). In the MIMIC model, the latent variable is linearly determined by a set of observed exogenous factors and, in turn, determines a set of observed endogenous indicators. The MIMIC model, therefore, consists of two sub-models: one is the Structural equation model, and the other is the Measurement model. The Structural equation model defines the relationship between the latent variable and the set of exogenous causal variables and is written as

$$y^* = \lambda'x + \nu \quad (2)$$

where $x = (x_1, \dots, x_k)'$ is a vector of k observed variables that potentially influence the latent variable y^* , therefore called cause variables. $\lambda' = (\lambda_1, \dots, \lambda_k)$ is a vector of k coefficients showing the marginal effects of these variables on latent variable y^* . ν is the white noise disturbance.

The Measurement model links latent variable y^* to a set of observed endogenous variables and is written as

$$y = \delta y^* + u \quad (3)$$

where $y = (y_1, \dots, y_m)$ is a set of m indicators reflecting latent variable y^* , and the elements of vector $\delta = (\delta_1, \dots, \delta_m)$ represent the factor-loading coefficients. Each indicator in vector y is an imperfect measure of the latent variable y^* , hence, an error term is added to each equation. $u = (u_1, \dots, u_m)$ is a vector of m error terms. The error terms of Eqs. (2) and (3) are assumed to be unrelated.

Given this MIMIC framework, we specify the structural model for the latent variable $Norm^*$ as follows:

$$\begin{aligned} Norm^* = & \beta_1 FEDU_i + \beta_2 MEDU_i + \beta_3 BRO_i + \beta_4 SIS_i + \beta_5 MELDER_i + \beta_6 FELDER_i + \beta_7 MMEDIA_i \\ & + \beta_8 FMEDIA_i + \beta_9 Land_i + \beta_{10} Rural_i + \delta_c + \gamma_r + \mu_t + \tau_s + \nu_i \end{aligned} \quad (4)$$

where $FEDU_i$ and $MEDU_i$ denote the years of schooling for the father and mother of the woman i , respectively. $MELDER_i$ and $FELDER_i$ represent the presence of elderly male and female members in the household, respectively. $MMEDIA_i$ and $FMEDIA_i$ represent male and female media exposure in the household, respectively. $Land_i$ measures the household's land ownership. Since the norms are community-driven practices, different social groups might follow different social norms. To control for these differences, we include religion (γ_r) and caste (δ_c) fixed-effects and region dummies ($Rural_i$). Additionally, the social norms could be different across different states; hence, we include state-level fixed effects (τ_s). We also include year-of-birth fixed effects (μ_t) to control for differences in norms among women born in

different years. In the terminology of the MIMIC model, the set of these variables denotes the *cause* variables (predictors) of the latent variable $Norm^*$.

Following the MIMIC framework, the measurement model for $Norm^*$ can be written as:

$$\mathbf{y} = \delta Norm^* + \mathbf{u} \quad (5)$$

where $\mathbf{y} = (\text{Exogamy}, \text{Cousin}, \text{Widow}, \text{Veil}, \text{Gender role}, \text{Hus Relation}, \text{Marriage Say}, \text{Hus Caste})'$ denotes the vector of indicators of social behavior that are manifestations of norms as explained in “Average Years of Schooling”. δ is the vector of factor-loading coefficients that measure how much variation in the indicators is explained by the latent variable $Norm^*$.

Full Model

The complete model integrates the structural and measurement models and is referred to as a system of ‘Structural Equation Modeling (SEM)’.⁷ The whole system of equations can be written as follows:

Structural Model

$$AYS_i = \alpha_0 + \alpha_1 Norm^* + \alpha_2 FEDU_i + \alpha_3 MEDU_i + \alpha_4 BRO_i + \alpha_5 SIS_i + \alpha_6 Consumption_i \\ + \alpha_7 Assets_i + \alpha_8 Land_i + \alpha_9 Rural_i + \delta_c + \gamma_r + \mu_t + \tau_s + \epsilon_i$$

$$Norm^* = \beta_1 FEDU_i + \beta_2 MEDU_i + \beta_3 BRO_i + \beta_4 SIS_i + \beta_5 MELDER_i + \beta_6 FELDER_i \\ + \beta_7 MMEDIA_i + \beta_8 FMEDIA_i + \beta_9 Land_i + \beta_{10} Rural_i \\ + \delta_c + \gamma_r + \mu_t + \tau_s + v_i$$

Measurement Model:

$$\mathbf{y} = \delta Norm^* + \mathbf{u}$$

where $\mathbf{y} = (\text{Exogamy}, \text{Cousin}, \text{Widow}, \text{Veil}, \text{Gender role}, \text{Hus Relation}, \text{Marriage Say}, \text{Hus Caste})'$. In this system of equations, AYS_i^* is a latent variable that is perfectly measured by observed AYS_i and $Norm^*$ is another latent variable that is measured by the indicators specified in the measurement model. The whole model is pictorially described in Fig. 2.

One way of estimating the system is first to estimate the MIMIC model of latent variable $Norm^*$ (Eqs. (4) and (5)) using the Maximum Likelihood Estimation (MLE) method, obtain the empirical predictions of the latent variable from this model, and use these predicted values to estimate Eq. (1) by OLS. However, the estimates could be biased and inconsistent due to endogeneity issues.

The endogeneity could arise because the errors of Eqs. (1) and (4) could be correlated to each other due to omitted variables or the presence of observed

⁷ MIMIC model is a special case of Structural Equation Modeling (SEM) (Bollen 1989).

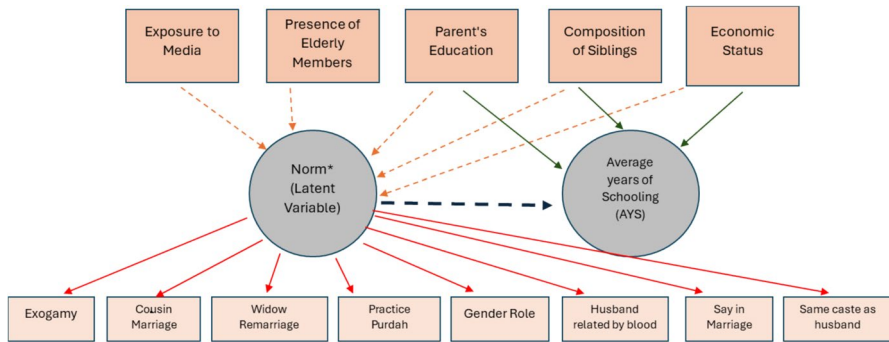


Fig. 2 Pictorial depiction. This figure shows the pictorial representation of the full/MIMIC model presented by Eqs. (1), (4), and (5). Source: Authors' construction

exogenous variables (like FEDU, MEDU, and so on) that are common to both equations. As a result, the endogenous variable $Norm^*$ becomes correlated to the disturbance term in Eq. (1), due to which single equation estimation will give inconsistent estimates. Therefore, the system of Eqs. (1), (4), and (5) requires joint estimation. Another advantage of joint estimation is that it helps in gaining efficiency.

Besides endogeneity, another important issue in the estimation of such a system of equations is 'identification'. The full system is identified if all the parameters of the model are identified. The MIMIC model is identified if two conditions hold (Jöreskog and Goldberger 1975). First, the number of exogenous cause variables is one or more ($k > 1$), and the number of endogenous indicators is two or more ($m > 2$). This condition holds in our case. Second, one of the coefficients of the indicators in the Measurement model is set to unity. This identifying restriction also helps in providing the scale for the latent variable, which is always indeterminate apriori.

We, therefore, set the coefficient of the indicator *Exogamy* equal to 1. Previous studies have shown that the practice of marriage exogamy is associated with an adherence to stricter traditional norms or unfavorable norms. Since the value of 1 of indicator *Exogamy* reflects traditional or conservative norms, the latent variable $Norm^*$ represents the overall degree of adherence to these traditional norms. Therefore, the higher score for latent variable $Norm^*$ reflects higher adherence to traditional norms, and a lower value indicates that the norms are relatively more favorable for women's education.

Advantages of MIMIC Model

We will begin our analysis by estimating a regression using indicators for social norms as explanatory variables directly, followed by estimating a regression using a Principal Component Analysis (PCA) score (created from all indicators) as an explanatory variable for $Norm$. Finally, we estimate the SEM where norms are measured in the MIMIC framework.

The MIMIC (Multiple Indicators Multiple Causes) model offers significant advantages over both the PCA (Principal Component Analysis) and the simple addition of indicators in regression analysis when measuring the impact of social norms. First, unlike PCA, which aggregates indicators into a single score based on their variance without considering the underlying theoretical framework, the MIMIC model explicitly models the relationship between latent variables (e.g., norms) and observed indicators while simultaneously accounting for the influence of external causes on the latent construct. This allows for a more nuanced understanding of how different indicators contribute to the underlying latent variable and how this latent variable, in turn, influences the outcome variable of interest. A simple regression with added indicators may lack theoretical coherence, as it does not account for the latent variable that the indicators are supposed to represent.

Second, in the MIMIC model, measurement error is explicitly accounted for by modeling the relationship between the latent variable and its indicators. This leads to more accurate and reliable estimates. PCA, on the other hand, does not account for measurement error in the same way. Similarly, in a regression with multiple indicators, each indicator may be subject to measurement error, potentially leading to a biased estimate. Additionally, by incorporating multiple causes and indicators, the MIMIC model reduces potential biases and provides a more accurate representation of the complex relationships between norms and outcomes. This approach also offers more flexibility and interpretability than simply adding binary indicators, as it captures the latent dimension of norms in a way that is grounded in the theoretical understanding of the construct rather than relying solely on the observed data's statistical properties.

Third, the MIMIC model directly models the causal relationships between the observed causes (exogenous variables) and the latent variable, providing a clearer understanding of how specific factors contribute to the latent construct. It allows for different indicators to have varying degrees of association with the latent variable, which can be tested and estimated within the model. This flexibility ensures that the latent variable captures the most relevant aspects of the underlying construct. PCA, by contrast, creates a linear combination of indicators that maximizes variance, but it treats all indicators equally in terms of their contribution to the principal components.

Results

Regression Estimates

We first present two sets of regression results—(a) Adding indicators for social norms as explanatory variables and (b) Generating scores from all indicators using Principal Component Analysis (PCA) and using the score as a measure of Norm. These results are presented in Table 4 in columns (1) and (2), respectively.

In column (1), we observe that certain traditional norms, such as the practice of wearing *purdah* (*Veil*), the expectation for women to serve meals (*Gender role*),

parents' involvement in decisions regarding marriage (*Marriage Say*), and husbands' related by blood (*Hus Related*), are significantly associated with lower years of schooling. For example, the negative and statistically significant coefficient on the variable *Veil* suggests that women who adhere to this norm are likely to have significantly fewer years of education. Similarly, the variable *Marriage Say* shows a strong negative effect, indicating that when women's marriage decisions are controlled by others, their educational attainment is notably reduced. On the other hand, variable *Hus Caste* shows positive but less significant association with education, implying that this practice might not uniformly discourage women's education. Additionally, women's years of schooling are uncorrelated with variables like *Exogamy*, *Cousin*, and *Widow*.

Column (2) provides a more aggregated perspective by using a PCA score to represent the combined influence of all the social norm indicators. The PCA score, labeled as "Norm (PCA)," captures the underlying commonality across these indicators, and its coefficient is negative and highly significant. This finding suggests that the cumulative presence of these traditional norms is strongly associated with a reduction in women's years of schooling. Interestingly, despite the aggregation, the coefficients on other variables, such as parental education, asset ownership, and consumption, remain largely consistent across both models. This reinforces the robustness of the relationship between social norms and women's education, highlighting the pervasive influence of conservative social practices on limiting educational opportunities for women.

Measurement Model

Table 5 presents Maximum Likelihood estimates for the complete model. We first discuss the results of the measurement model, which is laid down in Eq. (5). These estimates are presented in Panel A. The coefficients associated with each indicator show the factor loadings, which measure the amount of variability of each indicator explained by the latent variable $Norm^*$. The coefficient with indicator *Exogamy* is set equal to one as identifying restriction.

A positive factor loading means that as the latent variable increases, the observed variable also tends to increase, whereas a high factor loading (closer to 1) indicates that the observed variable strongly reflects the latent variable. We see that all the coefficients attached to each indicator are statistically significant, which shows that the latent variable $Norm^*$ is significantly correlated with these practices or behavior indicators. These results support a key point of our analysis that norms are indeed multifaceted and manifested in multiple practices or behavioral aspects.

Indicators like *Veil*, *Gender role*, and *Marriage Say* have strong positive loadings, meaning that as the traditional conservative norms (higher score of $Norm^*$) strengthen or norms become unfavorable, these practices become more prevalent. This suggests that more conservative norms are associated with a greater likelihood of practicing *purdah*, adhering to gender roles, and having marriage decisions controlled by families. The indicators *Cousin* and *Hus Related* have strong negative loadings, which implies that the norm becomes stronger (more conservative or

Table 4 Regression estimates for women's average years of schooling

	(1) Regression with binary indicators	(2) Regression with PCA score of indicators
<i>Dependent variable: average years of schooling (AYS)</i>		
Exogamy	−0.016 (0.049)	
Cousin	−0.051 (0.058)	
Widow	0.064 (0.040)	
Veil	−0.442*** (0.047)	
Gender role	−0.155*** (0.044)	
Hus related	−0.206*** (0.072)	
Marriage say	−0.850*** (0.044)	
Hus caste	0.153* (0.080)	
<i>Norm (PCA)</i>		−0.236*** (0.019)
Father's education	0.246*** (0.005)	0.249*** (0.005)
Mother's education	0.243*** (0.007)	0.251*** (0.008)
Brother (prop.)	−1.219*** (0.137)	−1.223*** (0.137)
Sister (prop.)	−0.947*** (0.126)	−0.955*** (0.127)
Assets	0.265*** (0.004)	0.271*** (0.004)
Land	0.035 (0.044)	0.010 (0.044)
Consumption	0.345*** (0.036)	0.352*** (0.036)
Rural dummy	−0.355*** (0.051)	−0.369*** (0.051)
Observations	34,067	34,067
R-squared	0.567	0.563

Column (1) shows the regression estimates by adding indicators as social norms. Column (2) shows the regression estimates by using PCA-generated scores for norms. Standard errors are reported in the parentheses. ***, ** and * represents the significance at 1%, 5% and 10% respectively. The variable description is given in Tables 1 and 2

traditional), the prevalence of cross-cousin marriage decreases, and the likelihood of having marriages between blood relatives decreases. Indicators like *Widow* and *Hus Caste* have opposite signs but have smaller effects, suggesting they are less strongly associated with the latent norm *Norm**.

It is interesting to note that the coefficients attached to each individual indicator differ in terms of magnitude, suggesting that norms do not have the same influence on each of them. This perhaps explains the weak correlations among these behavioral indicators seen earlier in Table 3. These findings give credence to our argument that individual behavioral indicators could be imperfect proxies for the underlying unobserved norm. This also means that all these indicators are not substitutes for each other, and using just one indicator can give misleading conclusions.

Table 5 Maximum-likelihood estimates for women's years of schooling and latent variable $Norm^*$

	(1)	(2)
Panel A: measurement model		
Indicators as dependent variables of $Norm^*$ (Eq. 5)		
Exogamy	1.000	
Cousin	− 1.145***	(0.010)
Widow	− 0.066***	(0.009)
Veil	0.443***	(0.010)
Gender role	0.287***	(0.008)
Hus related	− 0.186***	(0.005)
Marriage say	0.510***	(0.010)
Hus caste	0.020***	(0.004)
Panel B: structural equations	Equation (4)	Equation (1)
Dependent variables	$Norm^*$	AYS
$Norm^*$		− 1.025*** (0.169)
Father's education	0.002*** (0.000)	0.250*** (0.005)
Mothers education	− 0.004*** (0.001)	0.253*** (0.008)
Brother (prop.)	0.073*** (0.011)	− 1.229*** (0.138)
Sister (prop.)	0.059*** (0.010)	− 0.964*** (0.127)
Presence of elderly male	0.015*** (0.004)	
Presence of elderly female	0.001 (0.003)	
Male exposure to media	0.013** (0.005)	
Female exposure to media	− 0.008 (0.006)	
Assets		0.273*** (0.004)
Consumption		0.356*** (0.036)
Land	0.027*** (0.004)	0.003 (0.044)
Rural dummy	0.030*** (0.004)	− 0.374*** (0.051)
Observation	34,067	

Panel A of the table shows the estimates of Eq. (5); the coefficients are in Column 1, and standard errors are in parentheses in Column 2. Panel B shows the estimates of Eqs. (4) and (1) in columns 1 and 2, respectively. The coefficient of the constant has been suppressed. Standard errors are reported in the parentheses. ***, ** and * represents the significance at 1%, 5% and 10% respectively. The variable description is given in Tables 1 and 2

Structural Equation Model

Norm

In this section, we discuss the estimation results of structural Eqs. (1) and (4) presented in Panel B of Table 5. Column (1) shows the coefficients for Eq. (4), and Column (2) shows the same for Eq. (1). We will begin by discussing the results for the equation of the latent variable $Norm^*$ (Eq. (4)). The results reported in Column (1) show that the education of father and mother plays a significant role in shaping norms for women's education. The coefficient attached to the variable father's

education (FEDU) is positive and statistically significant, implying that an increase in the father's education is associated with a higher value of latent variable *Norm**. On the contrary, the coefficient attached to the variable mother's education (MEDU) is negative and statistically significant, implying that an increase in mother's education is associated with a lower score of *Norm**. These results suggest that an increase in the education of mothers has a norm-breaking effect by making it more favorable for women's education, whereas an increase in the education of fathers has a norm-binding effect by reinforcing/making the norms stricter for women's education.⁸

Interestingly, the absolute size of the education of mother's coefficient (0.004) is more than twice the absolute size of the father's education (0.002), which highlights the significant importance of mother's education in deciding the investment on education for their daughters. Even though the fathers' education has a norm-binding effect, the norm-breaking effect of the mothers' education is much larger.⁹

Furthermore, the positive coefficients for the proportions of brothers and sisters among all siblings suggest that in families with more siblings, traditional norms tend to be stronger. The effect of having sisters is slightly smaller compared to brothers, which reflects that pressure of traditional norms is much higher in families with more male children. Having more brothers makes norms more unfavorable for women's education.

The presence of elderly male members significantly increases the conservatism of norms, whereas the presence of elderly females does not have a statistically significant effect on the norms. This suggests that older male family members might reinforce traditional norms. Male exposure to media is associated with a slight increase in conservatism of norms, which might indicate that media consumption among men could reinforce traditional views. However, female exposure to media does not significantly affect the latent norm, suggesting that media exposure among women might not play a strong role in shaping norms. At last, land ownership is significantly associated with more conservative norms. This reflects that richer households might follow more conservative norms.

In sum, we find that various factors, such as parents' education, sibling composition, presence of elderly males, and land ownership, significantly influence the strength of traditional norms, making norms more unfavorable for women's education.

⁸ There could be one possible concern that the parents' education could be correlated with the marriage market or other factors that operate at the household level, as a result this may lead to biased estimates. We checked for the correlation between these two variables, and we found that the correlation is 0.58, which is not very high. However, this correlation drops to 0.41 when the sample is restricted to father's education being strictly positive. To further rule out multicollinearity issue, we carried out two separate estimations of the model with only the father's education and mother's education considered separately. These results (available with the authors) showed that the sign of the father and mother education coefficients did not change in these individual estimations, and they remained significant.

⁹ Another possible concern could be that variable mothers' education could have less variation as the majority of the mothers have zero years of education. We carried out separate estimations of the model for two sub-samples: (a) where the mother's education is zero, and (b) the mother's education is positive. These results (available with the authors) showed that the results in the two sub-samples are qualitatively similar and are in line with the main results of the full sample.

Average Years of Schooling

We next move to estimates presented in Column (2) of Table 5, which reports the estimates for Eq. (1), where the dependent variable is ‘Average Years of Schooling (AYS)’ for women. Here, the key variable of interest is the latent variable $Norm^*$. As explained in “Full Model”, the *higher* score for latent variable $Norm^*$ reflects *higher* adherence to traditional norms.

Table 5 shows that the coefficient attached to $Norm^*$ is statistically significant and negative. The negative coefficient of -1.025 suggests that for each unit increase in the $Norm^*$, women’s years of schooling are expected to decrease by around 1 year. This implies that as the prevalence or intensity of traditional norms increases, women’s educational attainment decreases. That is, norms have an adverse impact on the educational outcomes of women. This result implies that communities or households with stronger traditional norms (e.g., those that enforce practices like exogamy, cousin marriage, restrictions on widow remarriage, wearing *purdah*, traditional gender roles, and control of women’s decisions) are associated with lower educational attainment for women.

Interestingly, we find that fathers’ and mothers’ education directly also affects women’s education, and we call this the ‘direct effect’. The estimates reported in Column (2) for both explanatory variables $FEDU$ and $MEDU$ are positive, statistically significant, and of similar magnitude. This implies that an additional year of increase in the education level of fathers and mothers increases the average years of schooling for women by approximately 0.25 years each. Presumably, this operates through the unobserved family background.

In sum, the education of both parents has two effects; one is the ‘direct’ effect, and the second is the ‘indirect’ effect, which works through the changing the strength of norms. Though the direct effect of both parents’ education is similar, the total effect on father’s education is slightly lower than mother’s education. Specifically, the total effect of an additional year of schooling of father and mother is 0.25 and 0.26 years, respectively. That is, the education of mothers has a slightly larger positive effect on women’s education, as an increase in mother’s education makes the norms more favorable for daughter’s education.

Furthermore, we find that the number of siblings is negatively associated with the women’s years of schooling. This could be because if there are more children in the family, the family resources get distributed, leaving fewer resources for each child. As a result, this could adversely affect women’s education. Furthermore, it is also interesting to note that the composition of siblings also has implications for women’s education. In particular, having more brothers has a more negative effect on a woman’s education, than having more sisters. Like parents’ education, composition of siblings also has two effects on education: direct effect and indirect effect through norms. Though the total effect of having siblings is negative, having more brothers (−1.3) has larger negative effect on women’s education than having more sisters (−1.0).

These results demonstrate another advantage of using the MIMIC framework for modeling norms. It has helped in disentangling the total effect of parents’ education and composition of siblings on women’s education separately. Finally, the economic

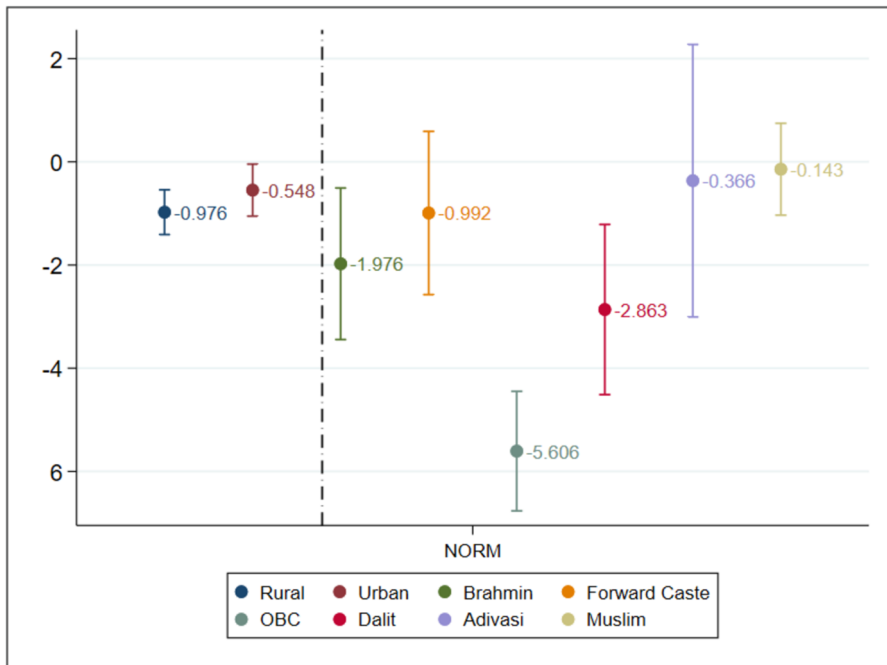


Fig. 3 Heterogeneity in the effect of Norm on women's education. The figure shows the estimated coefficients on the latent variable *Norm** from the model described in Eqs. (1), (4), and (5) estimated separately for the rural and urban regions and different social groups. The horizontal line with a cap around the coefficients represents the 95% confidence interval. Source: Authors' estimates based on data from IHDS-2

status of the household captured in terms of their consumption expenditure level and asset holding has a strong positive effect on women's education. Rural women have fewer years of schooling compared to urban women.

Heterogeneity by Place of Residence and Social Groups

India also 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 influences across social groups. As seen earlier in Fig. 1, the years of education show substantial variation across population groups, such as between rural and urban areas and across social categories. In this section, we examine heterogeneity in the impact of norms on women's average years of schooling. Figure 3 presents the Maximum Likelihood estimates for years of schooling for women, disaggregated by rural and urban regions and social groups. Here, the social groups are defined based on religion and caste.

These results in Fig. 3 indicate that the effect of norms differs significantly across these regions. Specifically, unfavorable norms around education have a more pronounced negative impact on women's years of schooling in rural areas compared to

urban areas. A unit increase in *Norm** decreases women's years of schooling by 0.98 and 0.55 years in rural and urban areas, respectively. These findings corroborate the widely held view that urbanization contributes to the weakening of the norms.

Figure 3 further shows that the effect of the norm on women's years of schooling is highest among OBC, followed by *Dalit* and *Brahmin*. However, we do not find any significant effect of norms on women's education among Forward Caste, *Adi-vasi*, and Muslim.

Conclusion

This paper attempts to explore the relationship between norms and women's education by measuring the norm as a latent variable instead of using individual behavioral indicators as proxies for norms. We show that the MIMIC model is an appropriate framework for quantifying the effect of norms when the dataset used in the analysis contains information on social practices / behavioral indicators that are rooted in norms. The MIMIC model allows us to address three important shortcomings in the existing literature on the relationship between social norms and women's educational outcomes, viz., (i) that norms are unobservable; (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 can be affected by various exogenous factors.

We have used data from the Indian Human Development Survey-II (IHDS-2) conducted in 2011–2012, a nationally representative multi-topic survey that provides detailed individual- and household-level information on education, gender relations, marriage practices, occupation, economic status, health, fertility, landholding, social capital, and social identity. Besides, this dataset provides information on the multifaceted behavioral aspects of norms that can be used directly to measure norms as a latent variable using the MIMIC framework.

Our results show that norms, measured as a latent variable, have a significant effect on negative women's educational attainment. The effect is more pronounced in the rural region, and it is significant only amongst OBC, *Dalit*, and *Brahmin*.

One methodological advantage of the MIMIC model is that it allows us to estimate the strength with which norms influence individual behavioral indicators. We find that norms are indeed multifaceted, and they have a statistically significant influence on several indicators. However, this influence is not uniform across various social practices. This finding reinforces our argument that individual indicators of social practice are likely to be only imperfect proxies for the underlying latent norms, and using them as a proxy for the latent norms is methodologically inappropriate.

A second methodological advantage of the MIMIC model is that it allows us to identify factors that weaken or bind the norms around education. We find that the father's education has a norm-binding effect on women's education, whereas mother's education has a norm-breaking effect on women's education. Additionally, the total effect of the mother's education on women's education is positive and much larger than that of the father's education. This result, viz., the identification of factors

that cause social norms to change, is, we believe, probably unique to our study. An important policy implication of this result is that increasing female education has an inter-generational virtuous effect since increasing female education makes the norms weaker for the next generation. By educating girls, the effect of norms that do not favor educating girls can be weakened over time.

The MIMIC model relies on the accurate measurement of latent variables like norms. However, norms are complex and multifaceted. One potential limitation of the above methodology could be that the indicators that we have used to consider may not fully capture the nuances of norms. This could be one potential reason for not observing a significant effect of norms with some social groups like *Adivasi* and Muslims. These groups might have diverse cultural practices, beliefs, and values that are not fully captured by the model. If the indicators used to measure norms are not as relevant or do not align with the specific values and practices of these groups, the model may fail to detect significant effects.

Acknowledgements This paper is based on a chapter of the first author's doctoral thesis. The research was carried out during the first author's tenure as Ph.D. Scholar at IGIDR and has not been financed by any external body.

Data Availability The paper uses the IHDS data, which is available in the public domain. IHDS-2 data is available here: <https://www.icpsr.umich.edu/web/DSDR/studies/36151>.

Declarations

Conflict of interest The authors have no conflict of interest to declare. I certify that the submission is original work and is not under review at any other publication.

References

- Akerlof, G.A., and R.E. Kranton. 2000. Economics and identity. *The Quarterly Journal of Economics* 115 (3): 715–753.
- Alañón, A., and M. Gómez-Antonio. 2005. Estimating the size of the shadow economy in Spain: A structural model with latent variables. *Applied Economics* 37 (9): 1011–1025.
- Ashraf, N., N. Bau, N. Nunn, and A. Voena. 2020. Bride price and female education. *Journal of Political Economy* 128 (2): 591–641.
- Ballon, P. 2018. A structural equation model of female empowerment. *The Journal of Development Studies* 54 (8): 1303–1320.
- Bertrand, M., and J. Pan. 2013. The trouble with boys: Social influences and the gender gap in disruptive behavior. *American Economic Journal: Applied Economics* 5 (1): 32–64.
- Bollen, K.A. 1989. *Structural equations with latent variables*. New York: Wiley.
- Connelly, R., and Z. Zheng. 2003. Determinants of school enrollment and completion of 10 to 18 year olds in China. *Economics of Education Review* 22 (4): 379–388.
- 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 R. Jayaraman. 2006. Determinants of school enrollment in Indian villages. *Economic Development and Cultural Change* 54 (2): 405–421.
- Drèze, J., and G.G. Kingdon. 2001. School participation in rural India. *Review of Development Economics* 5 (1): 1–24.
- Dyson, T., and M. Moore. 1983. On kinship structure, female autonomy, and demographic behavior in India. *Population and Development Review* 9: 35–60.

- Field, E., and A. Ambrus. 2008. Early marriage, age of menarche, and female schooling attainment in Bangladesh. *Journal of Political Economy* 116 (5): 881–930.
- Gueye, A.S., M. Audibert, and V. Delaunay. 2018. Can social groups impact schooling decisions? Evidence from castes in rural Senegal. *World Development* 110: 307–323.
- Jayachandran, S. 2015. The roots of gender inequality in developing countries. *Annual Review of Economics* 7 (1): 63–88.
- Jöreskog, K.G., and A.S. Goldberger. 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.
- Lahiri, S., and S. Self. 2007. Gender bias in education: The role of inter-household externality, dowry and other social institutions. *Review of Development Economics* 11 (4): 591–606.
- Maertens, A. 2013. Social norms and aspirations: Age of marriage and education in rural India. *World Development* 47: 1–15.
- Nakajima, M., Y. Kijima, and K. Otsuka. 2018. Is the learning crisis responsible for school dropout? A longitudinal study of Andhra Pradesh, India. *International Journal of Educational Development* 62: 245–253.
- Omura, M. 2019. Why can't I keep my surname? The fairness and welfare of the Japanese legal system. *Feminist Economics* 25 (3): 171–200.
- Psacharopoulos, G., and H. Yang. 1991. Educational attainment among Venezuelan youth: An analysis of its determinants. *International Journal of Educational Development* 11 (4): 289–294.
- Rammohan, A., and P. Robertson. 2012. Do kinship norms influence female education? *Evidence from Indonesia. Oxford Development Studies* 40 (3): 283–304.
- Rammohan, A., and P. Vu. 2018. Gender inequality in education and kinship norms in India. *Feminist Economics* 24 (1): 142–167.
- Richards, T.J., and S.R. Jeffrey. 2000. Efficiency and economic performance: An application of the MIMIC model. *Journal of Agricultural and Resource Economics* 25 (1): 232–251.
- Sundaram, A., and R. Vanneman. 2008. Gender differentials in literacy in India: The intriguing relationship with women's labor force participation. *World Development* 36 (1): 128–143.
- Titman, S., and R. Wessels. 1988. The determinants of capital structure choice. *The Journal of Finance* 43 (1): 1–19.
- Van de Ven, W.P., and J. Van Der Gaag. 1982. Health as an unobservable: A MIMIC-model of demand for health care. *Journal of Health Economics* 1 (2): 157–183.
- Wilson, K. 2001. The determinants of educational attainment: Modeling and estimating the human capital model and education production functions. *Southern Economic Journal*, 67 (3): 518–551.
- Zereyesus, Y.A., V. Amanor-Boadu, K.L. Ross, and A. Shanoyan. 2017. Does women's empowerment in agriculture matter for children's health status? Insights from Northern Ghana. *Social Indicators Research* 132 (3): 1265–1280.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.