IFS

Institute for Fiscal Studies

Britta Augsburg
Juan P. Baquero
Sanghmitra Gautam
Paul Rodriguez-Lesmes

21/44 Working paper

Sanitation and marriage markets in India: evidence from the total sanitation campaign

Economic and Social Research Council

Sanitation and Marriage Markets in India: Evidence from the Total Sanitation Campaign*

Britta Augsburg, Juan P. Baquero, Sanghmitra Gautam‡, Paul Rodriguez-Lesmes

November 30th, 2021

Click here for an updated version

Abstract

We analyse the marriage decisions of men and women in rural India, focusing on the added attractiveness of sanitation within the marital living arrangement. We demonstrate that the Government's Total Sanitation Campaign (TSC) changed marriage market outcomes for both men and women. To decompose the overall policy impact on the marriage market equilibrium, we develop a simple matching model. The model is identified and estimated using data from the Indian Human Development household survey (IHDS) and quasi-random variation from the TSC. Decompositions reveal that (i) cohorts within TSC exposed markets experienced a shift in marital gains both across matches and within a given match, which is characterised by a marked gender asymmetry, and that (ii) TSC exposure led to a decline in women's effective control over resources, reflected in the surplus division.

Keywords: Empirical matching, Marriage markets, Sanitation, Sorting

JEL Classification: C78, D13, J12, O18

*We thank Marcus Berliant, George-Levi Gayle, SangMok Lee, Bruce Petersen, Robert Pollak, Yaniv Stopnitzky, and seminar participants at Work, Families, and Public Policy seminar WUSTL, WEAI conference 2020, ADBI-IFS Sanitation and Development conference for valuable comments. Gautam gratefully acknowledges research support grant from the Weidenbaum Center at Washington University in St Louis. Augsburg gratefully acknowledge support of the Economic and Social Research Council (ESRC) Center for the Microeconomic Analysis of Public Policy at the IFS (grant reference ES/M010147/1 and of the ESRC grant reference ES/S01571X/1. Rodriguez-Lesmes acknowledges financial support from Alianza EFI-Colombia Científica grant with code 60185 and FP44842-220-2018. Preliminary, please do not cite or redistribute without permission. **Augsburg**: Institute for Fiscal Studies (e-mail: britta_a@ifs.org.uk); **Baquero**: School of Economics, Universidad del Rosario (e-mail: juanp.baquero@urosario.edu.co); ‡*Corresponding author*: **Gautam**: Department of Economics, Washington University in St Louis, MSC 1208-228-308, One Brookings Drive, St. Louis, MO 63130, US (e-mail: sanghmitra.gautam@wustl.edu); **Rodriguez-Lesmes**: School of Economics, Universidad del Rosario (e-mail: paul.rodriguez@urosario.edu.co).

1 Introduction

Many social programs and policy interventions have focused on women in the expectation that women derive larger benefits from them. The female focus, which is for example used for promotion of cooking stoves (Miller and Mobarak, 2013), is especially evident in the context of Water Sanitation and Hygiene (WASH) interventions where various programs aimed at promoting uptake of improved practices and facilities focus both directly and indirectly on women. A notable example of such policies is the now popular 'No Toilet, No Bride' program implemented by the state government of Haryana, India.[1] The advertisement campaign promotes a narrative which encourages families with marriage-age girls to 'ask' prospective suitors to provide, and hence if necessary construct, a latrine in the new home of the bride – thereby defining sanitation as a living arrangement. The emphasis on gender is also true for the Total Sanitation Campaign (TSC), an earlier version of the Swatch Bharat Mission, the current Government of India (GoI)'s flagship sanitation policy.[2]

While there is substantial evidence that the TSC and other sanitation interventions have increased sanitation take-up (Arnold et al., 2010; Barnard et al., 2013a; Clasen et al., 2014; Hammer and Spears, 2016; Patil et al., 2014; Pattanayak et al., 2009; Stopnitzky, 2017)[3], little is known with respect to the implications on the marriage market and impact on women.[4] If sanitation can be considered a type of living arrangement for a couple, policies aimed at increasing sanitation coverage have the potential to modify marriage market outcomes, even if they do not explicitly intertwine sanitation take-up with marriage markets. *This paper focuses on quantifying the added marital value of having sanitation within the living arrangement, whether it entails a gain or a loss on the marriage market and, if so, for whom.*

Our objective to quantify the added marital gain from sanitation on the marriage market necessarily intersects with two fundamental questions in the study of marriage markets: who marries whom? and how does the market clear? This inter-linkage poses notable challenges to the objective at hand. In particular, the determinants of marital sorting among different living arrangements cannot solely be attributed to taste for sanitation which may correlate with marital preferences and vary across the wealth distribution. In addition, the presence of general equilibrium effects and two-sided unobserved heterogeneity in a bilateral matching process raises non-trivial challenges in our ability to quantify the marital value of sanitation and the impact of sanitation policies that interact with the marriage market. A central contribution of this paper is to explicitly model the marriage market equilibrium in order to decompose the total effect of sanitation programs, in our case the GoI's TSC, and thereby quantify the marital gains associated with sanitation on the marriage market.

[1] Since its initial inception the Haryana program has become a model in the use of promotional campaigns aimed at men but with a focus on women (Cavill et al., 2018; Radtke, 2018).

[2] The TSC was implemented between 1999 and 2012. Approx 6%-15% of the program's overall budget was allocated to gender focused campaigns in 2011 (CBGA, 2011).

[3] A more recent evaluation of the GoI's approach is conducted by Andres et al. (2020).

[4] A notable exception is Stopnitzky (2017), which we discuss below.

Focusing on the marriage market and the marriage decisions of men and women, we contribute to the literature on marital sorting and the importance of sanitation within the marriage market in several notable ways.[5] First, we exploit quasi-random variation from the TSC that generated exogenous variation in the level of sanitation across India's districts and over time. This is important since sanitation may be correlated with observable (e.g wealth) and unobservable (e.g. marital preferences) characteristics that are relevant for matching decision and are otherwise difficult to disentangle.Therefore, a regression analysis of marriage rates on the sanitation status of the spouse that does not account for endogeneity does not inform us about the added gain from sanitation within the marital living arrangement. Exploiting district and time variation in TSC exposure, coupled with multiple observations of the marriage market, allows us to overcome these identification challenges.

The empirical analysis uses three different data sources (1) the Indian Human Development Survey (IHDS) waves 2004 (Desai et al., 2010) and 2011 (Desai et al., 2018), a household-level panel which provides rich micro-data on household composition, including individual demographics and social economic indicators; (2) 2001 and 2011 census data at the district level; and (3) TSC performance monitoring data aggregated at the district level by WSP (2011). This information allows us to exploit district and time variation from the TSC which we show increased sanitation ownership by 6.6 percentage points (ppts) among households with marriage eligible children and generated an exogenous increase in the composition of households with sanitation. Moreover, we show that exposure to TSC increased the probability of marriage for men and women from poorer households by 3.8 ppts and 6.5 ppts respectively.

Our second contribution is to develop and structurally estimate a friction-less matching model (Choo and Siow, 2006b) of the marriage market, where sanitation is modelled as type of living arrangement and individuals match on both observables and unobervables. In the model individuals on either side of the market, local to district and caste, are defined and matched over a discrete number of wealth types. To allow for composition effects and differential unobserved marital preference, men and women simultaneously choose their spouse *and* living arrangement. In other words, prospective spouses decide not only whether to marry, and whom to marry, but also the type of living arrangement they want to share. Importantly, in the proposed model we allow the marital surplus that derives from sanitation to vary over the wealth distribution through a match-specific component termed "love-for-hygiene". The resultant marital surplus function, specific to the living arrangement and the match type of the couple, is empirically tractable and allows for sorting on unobserved characteristics. We extend the framework to multiple markets and exploit variation over time (before and after TSC exposure), and across geographical areas with different populations and marriage market conditions, to achieve model identification.

We use the model to estimate the total effect of the TSC on marital behavior of individuals living

[5]Several other programs are known to have effects on marriage markets. For instance, the introduction of divorce laws (Fernández and Wong, 2017; Greenwood et al., 2016; Rasul, 2006; Reynoso, 2018), incarceration of offenders (Charles and Luoh, 2010), compulsory schooling (Hener and Wilson, 2018), and school construction (André and Dupraz, 2019; Zha, 2019).

in districts with high TSC exposure in rural India. Specifically, the model estimates allow us to (i) decompose and quantify the overall policy impact on marital sorting, gains to marriage, and division of gains among partners, and to (ii) identify the overall importance of sanitation on the marriage market. This is achieved by exploiting both observed matching patterns, under different living arrangements, and associated marital gains over time and across markets under specific assumptions.

In this paper, we focus on TSC impacts on marriage markets which in our opinion is of first order importance. Intuitively, what makes sanitation uptake in marriage markets relevant is the magnitude of the unobserved heterogeneity that may be indicative of underlying preferences for marriage and sanitation. Simply put, if there was no such variation, there would be no impact on sorting and thereby no role for the marriage markets in the uptake of sanitation. To the best of our knowledge the importance of sanitation in the marriage market has not been explicitly modelled. Stopnitzky (2017), who evaluates the No Toilet, No Bride', relies on a triple-difference specification, comparing the sanitation status of households with and without boys of marriageable age, living in the policy implementing state or a comparable, Northern Indian states, before and after program implementation. Analysing heterogeneity in impacts by scarcity of women, he infers that impacts are driven by markets where women are scarce. However, instead of modelling the marriage market equilibrium, the paper relies on an empirical proxy i.e., sex ratios, to draw inference on marriage market outcomes from sanitation take-up regressions.

The use of sex ratios as an empirical proxy for marriage market conditions has been widely popular in the literature.[6] Nevertheless, the use of sex ratios in marriage rate regressions has a key limitation: it ignores the availability of alternatives. By allowing for spousal alternatives the Choo and Siow (2006b) marriage matching function encapsulates both the general equilibrium and the heterogeneous policy effect thereby capturing the overall equilibrium market response. While our paper, as Stopnitzky (2017), show that changes in marriage market conditions may affect sanitation take-up, the extent to which underlying market clearing transfers adjust is unknown. The estimates from the structural model help address this question and allow us to decompose the overall marriage market response to TSC exposure. Overall, we argue that a careful analysis of the marriage market is an integral component of the overall policy impact on existing marriages, and important in order to infer the determinants of well-being of men and women.[7]

Lastly, through model simulations we show that the general equilibrium effects induced by the TSC had unintended consequences both across different sub-markets defined by wealth types but also within a match among spouses. We find that access to sanitation makes it more attractive to be in a marriage for both men and women, inducing some to marry. We find in particular that the final match is rearranged, so that men of low and median wealth levels are more likely to marry women of low wealth. The new sorting pattern results in a higher match surplus for couples where men are wealthier than women, and in a lower match surplus for most remaining cases. Moreover, we show through our simulations that

[6] Abramitzky et al. (2011), Angrist (2002), Charles and Luoh (2010), Gupta (2014).

[7] The importance of marriage markets and changes in sorting was first emphasised by Lundberg and Pollak (1996).

these new sorting patterns would have increased low wealth expected surplus in areas with no shortage of women, whereas it would have unambiguously *de*creases the expected surplus in areas with shortage of women. These latter findings effectively imply a decline in women's control over residual household resources. Instead of preventing the unintended consequences, the presence of less women per men ended up exacerbating the loss of control of resources. Therefore, scarcity per se does not imply that female-targeted programs necessarily result in better outcomes for women without additional costs.

The rest of the article is organized as follows. Section (2) describes the data and characteristics of the Indian marriage market, which make it a particularly apt context for our analysis. Section (3) describes the TSC and presents the effect of the policy on sorting patterns and sanitation take-up among households with marriage eligible children. The theoretical framework is developed and estimated in Section (4). Our main empirical findings are presented and discussed in Section (5). Section (6) concludes.

2 The Context

A number of distinctive features of the Indian marriage market are conducive for our analysis. We discuss in turn how the convention of marriages taking place within a physical distance provides us with non-overlapping and multiple observations, and we use wealth as a matching characteristic, which allows us to quantify the added value from sanitation vis-a-vis other marital assets. We further describe other important features of the market that determine our sample, that we control for in our analysis, and that we exploit for heterogeneity analysis.

2.1 Multiple marriage markets

The Indian marriage market is characterized by the practice of patrilocal or virilocal residence by virtue of which a bride almost always moves from her natal home to her groom's home. (Desai and Andrist, 2010). In line, we see in the IHDS that not more than 12% of women grew up in the same village as their husband.[8] Importantly though, spouses are chosen from geographically close areas - the couples' villages of origin are on average a distance of 3 hours away with locally common transportation (see Appendix Table 10). Only 4% of total female migration (including for education and other non-marriage purposes) is inter-state, and 9.8% inter-district (but intra-state) (Kone et al., 2018), implying that the vast majority of brides' destinations are intra-district. This feature, as most recently highlighted for the Indian context by Beauchamp et al. (2017), provides us with not just one but many distinct marriage markets for our analysis, and generates overidentifying restrictions on the model parameters.

[8]The practice is closely linked with endogamy (marriage within one's own caste group although outside one's own 'gotra' - clan), which we will discuss further below.

2.2 Assortative matching by wealth

A second important feature of the Indian marriage market are the strong patterns of positive assortative mating observed in the matching process. We focus on wealth. Economic indicators, such as land and income, have been shown to be important attributes in the decision whom to marry in the Indian context (Borker et al., 2018,1; Rao, 1993).[9]

Our proxy for wealth is based on self-reported asset ownership using principal component analysis (PCA).[10], but will take different approaches to construct pre-marital wealth for the wife and the husband, driven by the practice of patrilocal exogamy, i.e. the bride to move into the groom's family's house. This almost universal practice implies that we can proxy the groom's pre-marital wealth by using information on asset ownership as reported for the household the couple resides in. For the bride's pre-marital wealth we instead use a predicted asset score which is based on how observable characteristics of single women (age, education, literacy, English knowledge, caste and state) predict the asset index of their families. Figure 1 shows the distribution of the asset index by year for both genders. In line with significant economic growth over this period, we observe a rightward shift in the distribution between 2004 and 2011. In our model, we consider a discrete version of this variables based on the terciles defines over gender and year. The vertical lines of the line indicates the limits of these categories.

We calculate the Pearson's correlation coefficients to characterise the degree of marital matching along wealth. We find a correlation coefficient of 0.679 between the groom's and bride's wealth index in 2005 and of 0.676 in 2011. A value of 1 would indicate perfect correlation, implying that these statistics confirm a high degree of matching along wealth in the Indian context. If we consider instead the discrete variable (low, medium, high wealth), Kruskal's Gamma correlation coefficients present similar results: a rank correlation of 0.736 in 2005 and 0.724 in 2011.[11]

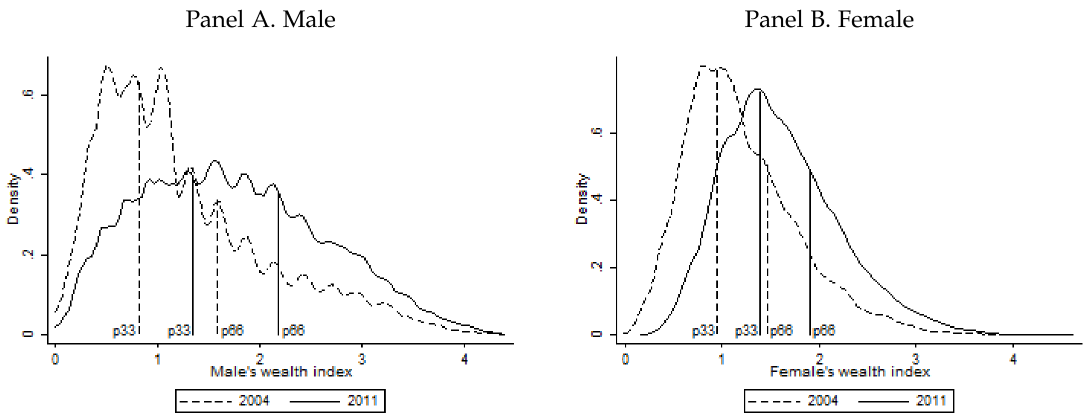
We take into account in our analysis that wealth and sanitation ownership tend to co-vary. The correlation coefficient of 0.49 in 2005 shows that there is considerable heterogeneity in sanitation ownership at different levels of wealth. Lastly, it is interesting to note that the coefficient of 0.49 is less than the correlation with owning a colour-TV (0.76) or a refrigerator (0.74), but more than owing a pressure-cooker (0.46).

[9] Assortative matching based on characteristics such as income, education, physical characteristics, and age have been similarly established in other contexts (see for example Becker and Becker (2009); Chiappori et al. (2018,1,1); Choo and Siow (2006b); Grossbard (1993); Hitsch et al. (2010); Pencavel (1998).

[10] The assets included in the component are: bicycle, sewing machine, generator set, grinder, motorcycle, TV, air cooler, clock, electric fan, chair/table, cot, telephone, cell phone, refrigerator, pressure cooker, car, Air conditioner, washing machine, computer, credit card, two clothes, footwear.

[11] Figure (10) in the appendix presents a graphical representation of the matching patterns from the perspective of women.

Figure 1: Wealth index distribution by gender and over time



Note. Own calculations using data from the IHDS waves 2004 and 2011. The wealth index, defined at household level, is the sum of self-reported assets at the household, which include: bicycle, sewing machine, generator set, grinder, motorcycle, TV, air cooler, clock, electric fan, chair/table, cot, telephone, cell phone, refrigerator, pressure cooker, car, Air conditioner, washing machine, computer, credit card, two clothes, footwear. For males it corresponds to the household information, while for female its is the prediction based on observed characteristics: age, education, literacy, English knowledge, caste and state. The prediction comes from a model for single women where the dependent variable is the wealth index of their family.

2.3 Other marriage market features shaping our analysis

Caste, age and education

Our analysis will include a set of controls to ensure we account for additional characteristics known to shape the Indian marriage market.

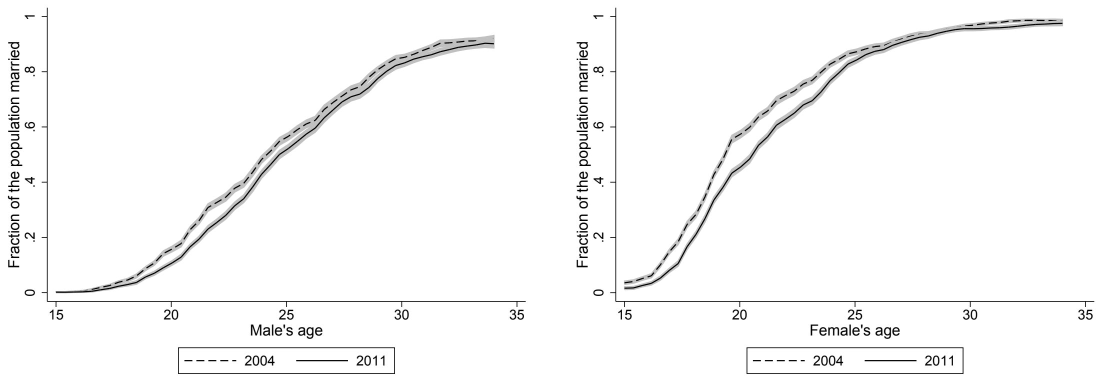
First off is the longstanding practice of caste endogamy, marriages taking place within a particular social group, determined here by caste. Over the period 2004-2011, less than 6% of marriages took place across caste (Table 10). This share has remained stable despite strong economic growth. Banerjee et al. (2013) show that underlying this statistic is a strong preference and low cost to marry someone of the same caste, and recent genetic analyses have established that these patterns of endogamous marriage have been in place for over 2,000 years (Moorjani et al., 2013).

Second, both age and education are known to be characteristics that determine marriage matches. Adams-Prassl and Andrew (2019) for example highlight how marriage market returns provide the primary motivation for investing in a daughter's college education in India, in line with other literature that demonstrates positive returns to (female) education on the marriage market (Attanasio and Kaufmann, 2017; Lafortune, 2013). The median age of brides was 18 years in 2011 (one year higher than in 2004, Appendix Table 10) and by the time women reach their late twenties, their likelihood of being married approaches 1 (Figure 2).[12] Men on the contrary get married at a later age and their likelihood of marriage

[12] The existence of unofficial age limits by which women are expected to be married (Jaggi, 2001), is at least partly driven by the quality of a match deteriorating quickly with a girl's age of leaving school (Adams-Prassl and Andrew, 2019).

levels off at just over 90% when they reach their mid-thirties.[13] The fact that marriages are near universal and rarely reversed leads us to focus our analysis on households of all single and married females (males) aged 15 to 34 at the survey time, e.g. excluding those that are unlikely active in the market.

Figure 2: Fraction of respondents who are married by age, gender, and year of survey



Note. Own calculations using data from the IHDS waves 2004 and 2011. Sample includes all married men and women 15 to 34 years.

Sex Ratio

Considerable attention has been paid to the fact that India is characterized by great disparity in sex ratios across the country, ranging from equally balanced to only two women for every three men in the North of the country (Figure 3). Research in contexts other than India suggest that such gender imbalances tend to lead to improved marriage outcomes for the scarcer sex (Abramitzky et al. (2011); Angrist (2002); Charles and Luoh (2010)).[14] We take sex ratios as a given market characteristic that varies across districts, and analyse whether the differences across districts affect our findings. To do so, we conduct heterogeneity analysis by high and low district ratios. We define high sex ratio as a district where there are at least 999 women per 1000 men aged 15-34; nearly 20% of the sample, both in 2004 and 2011.[15]

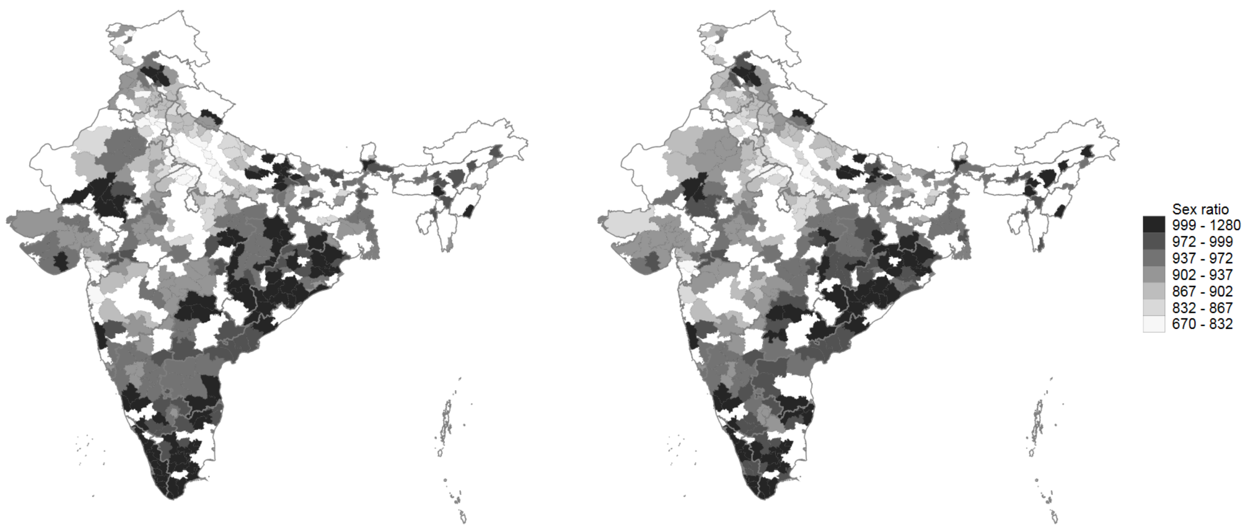
[13]Gupta (2014) discusses how male marriage rates show greater variation by region and over time than female marriage rates and depend on the availability of brides.

[14]While we are not aware of any study making use of exogenous variation in sex ratios in India, studies analysing the co-existence of sex selection and dowries in India (Basu (1999); Borker et al. (2019); Das Gupta et al. (2003); Murthi et al. (1995)) provide evidence in support of a link between sex selection and marriage markets.

[15]Other aspect of consideration is the presence of dowries. Appendix Table 10 shows that more than a third of households of the sample report that 'gifts' at marriage are usual in their context. Yet, the illegality of dowry in the Indian context makes accurate measurement a challenge. Moreover, Botticini and Siow (2003) warns against interpreting dowries as the sole form of transfer that clears the marriage market. As such for our analysis, the dowry response can be thought to be encapsulated within the market clearing transfer.

Figure 3: District-level sex ratio

A. 2001 B. 2011



Notes: Source: Indian census data, 2001 and 2011. Sample includes 15-34 year-olds.

3 Sanitation policy and marriage markets

3.1 The Total Sanitation Campaign

The Indian government has devoted large amounts of resources to make India "open defecation free" through its flagship Swachh Bharat Mission (Clean India) policy introduced in 2014. The policy follows similar previous national policies, going back to the Central Rural Sanitation Program (CRSP) in 1986, replaced by the Total Sanitation Campaign (TSC) in 1999 and the Nirmal Bharat Abhiyan (NBA) in 2012, which in 2014 evolved into the current Swachh Bharat Mission.

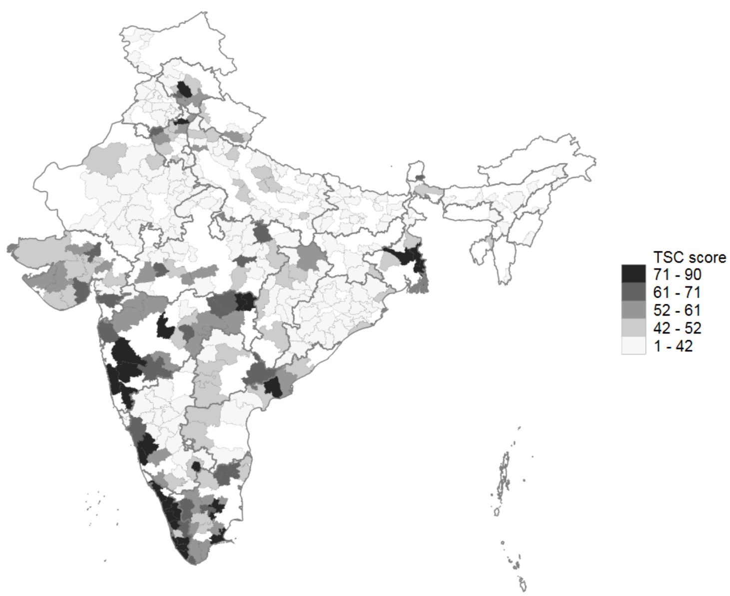
The TSC continued a focus on constructing new private household latrines with subsidies for vulnerable households, but added Information, Education and Communication (IEC) activities around sanitation and introduced in 2003 the provision of a financial award to communities that achieve Open Defecation Free (ODF) status, the Nirmal Gram Puraskar (NGP), or Clean Village Prize.[16] The state and central governments had a facilitating role in the implementation of the policy, that took the form of framing, enabling, providing financial and capacity-building support, and monitoring progress (WSP, 2011). The TSC is known for the importance of local leaders for the actual implementation of the program, due to

[16] The NBA program further introduced the concept of community-led total sanitation (CLTS), as part of the education and communication activities on sanitation. The SBM expanded to urban areas, among other changes.

elements like the NGP.

A strong upward trend in rural sanitation coverage has been documented over the course of program implementation with, however, significant disparities across states and districts (WSP, 2011). The study illustrates the unequal performance of TSC among Indian districts through a global monitoring index ("The Grand Score"), which weights performance based on eight key indicators.[17] Figure 4 shows the geographic variation of this indicator. 54 districts (16% of the IHDS sample) receive a score of 61 or above (the two darkest colors in the map), which we use as a cut-off for high-performing districts.

Figure 4: Grand Score TSC implementation performance by district



Note. Data source: WSP (2011). The "Grand Score" is an index that weights performance based on eight key TSC performance indicators, covering inputs, outputs, process and outcomes and each indicator is allocated a maximum score. Scores add up to a maximum of 100 and are available at the district level.

3.2 Toilet take-up in marriage markets

Several studies have evaluated this policy for, including several randomised control trials (Clasen et al. (2014), Hammer and Spears (2016), Patil et al. (2014), Pattanayak et al. (2009), Stopnitzky (2017) Barnard

[17] The eight indicators cover inputs, outputs, process and outcomes and each indicator is allocated a maximum score. The first indicator is an input (% of TSC budget spent) for which the maximum score is 5; the next two are outputs (% of household toilet targets achieved, 15 scores, and % of school sanitation target achieved, 10 scores); there are three process indicators with maximal 10 scores for each (financial efficiency (cost per NGP community), average population per community (gram panchayat), success rate of NGP applications; and two outcomes indicators (number of NGP panchayats, 30 scores, and % of NGP panchayats, 10 scores.

et al. (2013b)), establishing its effectiveness in increasing toilet ownership, with impacts that are said to be high compared to many other evaluated sanitation policies Garn et al. (2017).[18]

In this section we explore whether the effectiveness of the TSC policy in increasing household toilet ownership varies according to characteristics of the marriage market. We exploit the variation on implementation across districts described above. Specifically, we focus on cross-Indian households that are likely to enter/be in the marriage market between 2004/05 and 2011, and consider differential impacts by underlying market conditions, as indicated by the district-level sex ratio. We use a repeated cross-section of all households where the household head has at least one son or daughter aged 15-34.[19] We estimate the following difference-in-difference specification:

$$Toilet_{it} = \pi_0 dT_t + \pi_1 TSC_i + \pi_2 TSC_i \cdot dT_t + \mathbf{Z}_{it}\Lambda + e_{it}, \tag{1}$$

where $Toilet_{it}$ is toilet adoption in the household i in survey wave t. We define $dT = 1$ for survey wave 2011, and $dT = 0$ for 2004/05; $TSC = 1$ for households located in districts with a high TSC performance, which corresponds to a high grand score above or equal 61, and 0 otherwise. The matrix $\mathbf{Z}$ represent the control variables: age and marital status of the oldest marriageable female (male); the wealth index of the female's (male's); household size; education level of the household head (no education [base], primary, incomplete secondary, secondary, and above secondary); caste (Brahmin [base], High caste, Other backward caste, Dalit, Adivasi, Muslim, Sikh - Jain); an indicator variable for the rural zone; and fixed effects at the state level. e_{it} corresponds to the unobserved component. Standard errors are clustered at the district level. The coefficient of interest is π_2.

Table 1 shows the results, considering separately household where the oldest member within the marriage age range is male (Columns 1–3) and where it is female (Columns 4–6). The impact of high TSC-performance is $dT \times TSC$. We find that for households where the eldest individual in the age range 15-34 is a male, toilet ownership increases by 6.57 percentage points (ppts) due to living in a high-performing TSC district (column 1). The result is significant at the 5% level. For females, the estimated coefficient is smaller at 3.85, but statistically insignificant (column 4). In Table 12 in the appendix, we show that these results are robust to alternative definitions of shortage and TSC exposure. These impacts hide heterogeneity with larger impacts where women are scarce (low sex ratio, Columns 3 and 6), and even negative (significant at the 10% level) where women are abundant. In low sex ration districts, households where the eldest marriageable child of the household head is male, we find a 13.8 ppts increase in toilet ownership, and where it is female, one of 7.7 ppts. These estimates are significantly different from each other, suggesting

[18] The awareness creation approach used in the program, Community-Led Total Sanitation (CLTS) was shown effective in increasing sanitation investment in for example Tanzania (Briceno et al., 2017), Mali (Pickering et al., 2015), Nigeria (Abramovsky et al., 2019), Ghana (Crocker et al., 2016), Ethiopia (Crocker, 2016). A study by Guiteras et al. (2015) demonstrates effectiveness of CLTS with subsidy provision in the context of Bangladesh.

[19] Appendix Table 11 demonstrates that results are in line when all households with a 15-34 year-old member are included in the sample.

that households with a marriageable boy tend to react more to the TSC campaign. These results are in line with Stopnitzky (2017) in the context of the 'No Toilet No Bride' campaign in Haryana. The author found that households with men of the marriageable age were more likely to take-up sanitation when exposed to the campaign than households with women of similar ages.

Table 1: TSC impact on Sanitation for households with an oldest marriageable son (daughter)

	Oldest marriageable son household			Oldest marriageable daughter household		
	(1)	(2)	(3)	(4)	(5)	(6)
Variables	All	High sex ratio	Low sex ratio	All	High sex ratio	Low sex ratio
dT	-0.0345***	-0.0326	-0.0363***	-0.0220*	-0.0118	-0.0253*
	(0.0112)	(0.0334)	(0.0119)	(0.0129)	(0.0287)	(0.0147)
TSC	0.0547	0.150**	0.0143	0.0871**	0.110	0.0601
	(0.0411)	(0.0625)	(0.0480)	(0.0391)	(0.0735)	(0.0459)
$dT \times TSC$	0.0657**	-0.0849*	0.138***	0.0385	-0.0344	0.0769**
	(0.0333)	(0.0439)	(0.0372)	(0.0296)	(0.0527)	(0.0368)
Observations	27,993	5,784	22,209	10,437	2,619	7,818
R^2	0.416	0.482	0.407	0.404	0.460	0.396

Notes: Own calculations using data from the IHDS waves 2004 ($dT = 0$) and 2011 ($dT = 1$). The sample consists of households of all single and married females (males) aged 15 to 34 at the survey time. Households are classified according to the gender of the eldest son (columns 1 to 3) or daughter (columns 4 to 6) or the household head within such age range. $TSC = 1$ correspond to a grand score for implementation of 61 or above, and $TSC = 0$ to those with a score of 61 or above. The grand score per district is taken from WSP (2011). The high sex ratio corresponds to districts with at least 999 women per 1000 men in the age range 15-34 (columns 2 and 5), while the low sex ratio is districts with less than 999 (columns 3 and 6). District level sex ratio information was computed using data from the population census 2001 and 2011. Apart from the coefficients presented in the table, as controls, we consider the age and marital status of the individual for whom the household is in the sample; the wealth index of the household; household size; education level of the household head (no education [base], primary, incomplete secondary, secondary, and above secondary); caste (Brahmin [base], High caste, Other backward caste, Dalit, Adivasi, Muslim, Sikh - Jain); an indicator variable for the rural zone; and fixed effects at the state level. Clustered at the district level standard errors in parentheses. Significance: *** p<0.01, ** p<0.05, * p<0.1

In Table 2 we repeat the analysis by whether a household falls within a high- or low-wealth category, based on the median of the wealth index distribution for each wave. We find that the impact on toilet ownership is higher for households at the top of the wealth distribution, living in areas with a scarcity of women. For high-wealth households with a boy in line to get married the impact is 20.0 ppts compared to 12.1% for low-wealth households with a groom-to-be; for high-wealth households with a girl in marriageable age the impact is 11.6% compared to an insignificant 2.85% for low-wealth households with a bride-to-be.

Table 2: TSC impact on Sanitation for Households with an oldest marriageable son (daughter) by wealth and sex ratio

	Oldest marriageable son household				Oldest marriageable daughter household			
	High sex ratio		Low sex ratio		High sex ratio		Low sex ratio	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variables	High wealth	Low wealth	High wealth	Low wealth	High wealth	Low wealth	High wealth	Low wealth
dT	-0.0155	-0.0522	-0.0557***	-0.0300**	-0.0416	-0.0651	-0.0134	-0.0637**
	(0.0335)	(0.0436)	(0.0183)	(0.0128)	(0.0351)	(0.0416)	(0.0186)	(0.0247)
TSC	0.159**	0.102	-0.0592	0.0711	0.102	0.0982	-0.0119	0.183**
	(0.0659)	(0.0764)	(0.0515)	(0.0631)	(0.0750)	(0.0764)	(0.0499)	(0.0711)
$dT \times TSC$	-0.0713	-0.0431	0.200***	0.121*	-0.00346	-0.0601	0.116**	0.0285
	(0.0437)	(0.0669)	(0.0451)	(0.0627)	(0.0511)	(0.0940)	(0.0495)	(0.0457)
Observations	3,294	2,490	12,349	9,860	1,655	964	5,173	2,645
R^2	0.373	0.337	0.273	0.295	0.386	0.344	0.327	0.332

Notes: Own calculations using data from the IHDS waves 2004 ($dT = 0$) and 2011 ($dT = 1$). The sample consists of households of all single and married females (males) aged 15 to 34 at the survey time. Households are classified according to the gender of the eldest son (columns 1 to 4) or daughter (columns 5 to 8) or the household head within such age range. $TSC = 1$ correspond to a grand score for implementation of 61 or above, and $TSC = 0$ to those with a score of 61 or above. The grand score per district is taken from WSP (2011). The high sex ratio corresponds to districts with at least 999 women per 1000 men in the age range 15-34 (columns 1, 2, 5 and 6), while the low sex ratio is districts with less than 999 (columns 3, 4, 7 and 8). District level sex ratio information was computed using data from the population census 2001 and 2011. High wealth correspond to individuals whose asset index is above the 50 percentile of the entire country distribution per wave (columns 1, 3, 5 and 7). Those households below such cutoff are classified as low wealth (columns 2, 4, 6 and 8). Apart from the coefficients presented in the table, as controls, we consider the age and marital status of the individual for whom the household is in the sample; the wealth index of the household; household size; education level of the household head (no education [base], primary, incomplete secondary, secondary, and above secondary); caste (Brahmin [base], High caste, Other backward caste, Dalit, Adivasi, Muslim, Sikh - Jain); an indicator variable for the rural zone; and fixed effects at the state level. Clustered at the district level standard errors in parentheses. Significance: *** p<0.01, ** p<0.05, * p<0.1

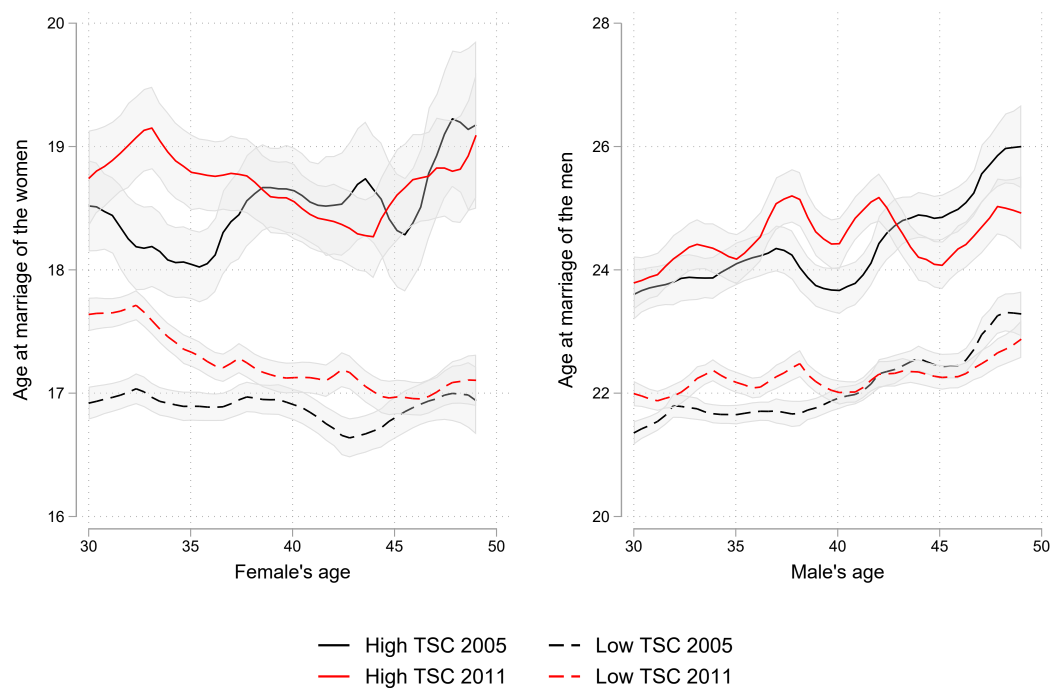
An important take-away from the analysis presented in this section is that the TSC program likely impacted the marriage market equilibrium: the extend of the uptake is a function of marriage market characteristics and attributes such as the scarcity of women and wealth. In addition Tables (1) and (2) provide indirect evidence on the importance of incorporating heterogeneous policy effects of the TSC in our analysis of the marriage market equilibrium. We unpack this link in Section (5), based on the structural model presented in Section (4). Before doing so, we once more turn back to marital sorting.

3.3 Parallel trends

Our identification strategy relies on district variation of TSC implementation. There are of course concerns that unobserved characteristics could explain both the speed in TSC implementation, and changes in the marriage market, making it important to assess the parallel trends assumption our approach relies on.

We do so in the vein of Spears and Lamba (2016), who use IHDS data to assess TSC impact on children's cognitive skills. The authors similarly do not have information prior to the program implementation, but observe another educational outcome (literacy) in a cross-sectional set-up In line, we observe the marriage market outcome 'age age at marriage' for older cohorts of the marriage market. Figure 5 shows that for both districts with high and low exposure to TSC (in both 2005 and 2011), the mean age of marriage for men and women has been historically different, but these differences have been stable over time.[20] Thus, we consider parallel trends to be a valid assumption on this setup.

Figure 5: Age at marriage by age for individuals not in the marriage market, by TSC implementation group and IHDS wave



Note. Local polynomial smoother using an Epanechnikov kernel.

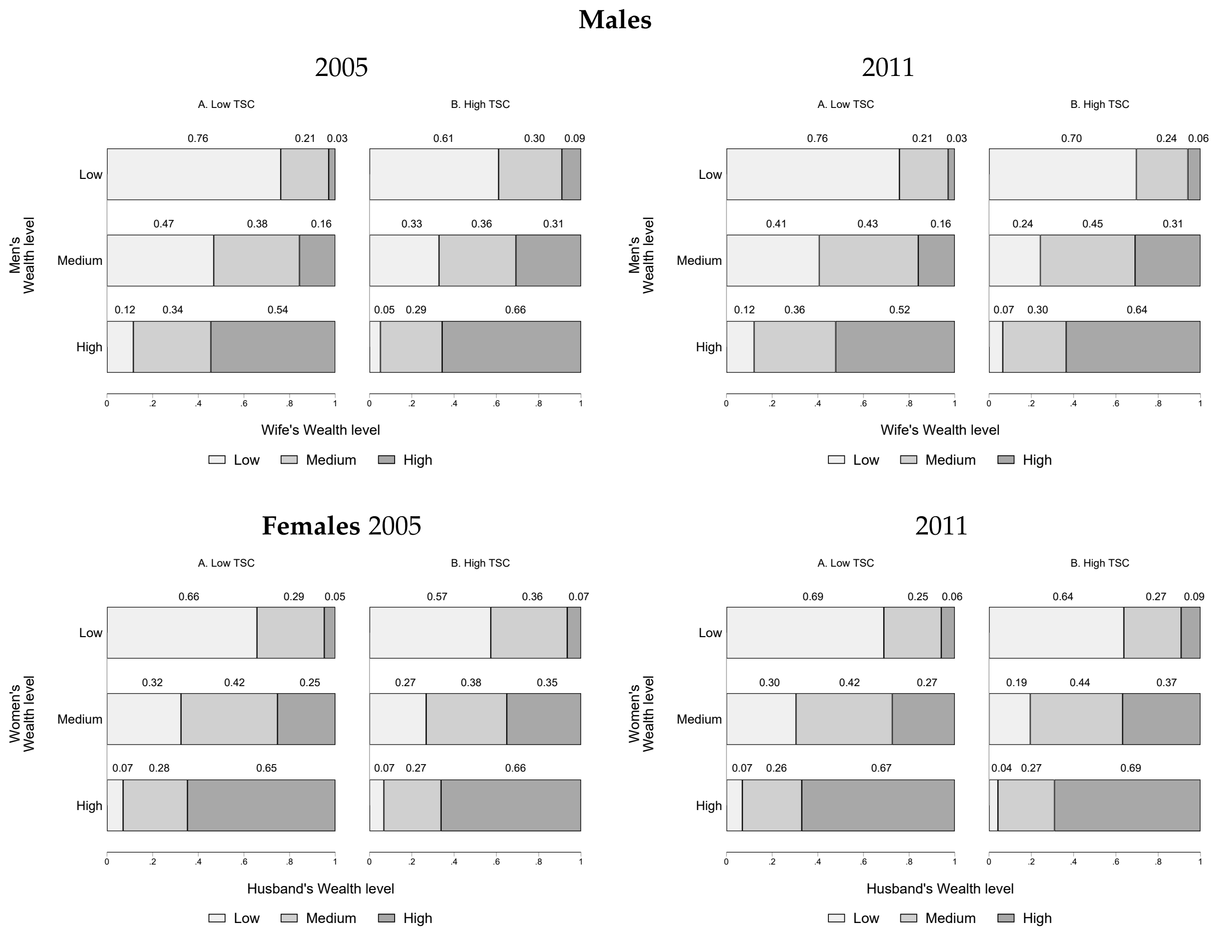
3.4 Marital Sorting

We discussed above the high degree of assortative matching along wealth in the Indian marriage market, which we present here graphically and by high and low TSC performance (Figure 6). The figure clearly shows that spouses tend to marry partners of the same wealth category. For example, in 2005, 76% of men with low pre-marital wealth living in low-performing TSC areas were matched with a woman whose

[20]Age in the graphs is restricted to be above 30 years of age, as it is highly likely that these individuals are already out of the marriage market by 2011. The top restriction at 50 years of age is due to questionnaire design which includes the exact age-at-marriage question for women (men's variable is derived from women's answers).

pre-marital wealth also fell into the low-wealth category; in high-performing TSC areas the percentage was 61. In 2011, the percentage of low-wealth matches in low-performing TSC areas remains at 76% and increases from 61% to 70% in high-performing TSC areas. Similar types of patterns, with same-wealth levels dominating matches are found for women, both in high- and low-performing TSC areas.

Figure 6: Marriage patterns of women and men over time across TSC exposure groups



Notes: Own calculations using data from the IHDS 2005 and 2011.

Following the strategy described in the previous subsections, other outcomes of the marriage market are -up to the moment- not changing due to the program. This is the case of age of marriage or age gaps (see Table 15 in the appendix). We cannot rule out that those outcomes change once the entire cohort is above age 35. However, these findings suggest that the focus of the analysis should be on the sorting patterns rather than on other potential behaviours such as the timing of marriage (in terms of age).

While these marriage rates provide indirect evidence of the importance of marital matching along wealth in both high and low-performing TSC areas, they do not take into account general equilibrium effects such as the availability of spousal alternatives and, in our context, changes in the number of types of matching opportunities in the market – notably, the large increase in the number of households with sanitation, particularly in high-performing TSC areas. Upon taking general equilibrium effects into account, we are not only able to quantify how individual marital outcomes are related to individual and match-specific characteristics, e.g. wealth and sanitation, but also any subsequent market equilibrium response in terms of sorting which are non-trivial. In such cases any reduced form analysis in the context of evaluating marriage market outcomes would conflate individual and market effects as individual of different types interact within a marriage market.[21] This issue is independent of the non-overlapping nature of the marriage markets and would persist within each geographic and caste isolated market cluster.

4 Theoretical Framework

To decompose the overall impact of TSC on marriage market outcomes and rationalize the sorting patterns observed in the data we employ a simple matching model.[22] Specifically, the structural model helps us to quantify the resulting impacts on marital gain and division of surplus, both of which are determined from the marriage market equilibrium and are thereby endogenous. For example, if TSC increases the gains from participation in the marriage market for both men and women of a given match type this would imply an increase in the total match specific surplus derived from that match. Moreover, the increased supply of sanitation, a potentially desirable marital living arrangement, on the marriage market may also generate changes in sorting along other dimensions. The impact on the surplus division, however, is less straightforward and would depend on the differential impacts by gender. In what follows, we employ a Choo and Siow (2006b) - henceforth, CS, marriage matching function which allows for general equilibrium effects i.e., spousal alternatives and incorporate unobserved heterogeneity using Chiappori, Salanié and Weiss (2017) - henceforth, CSW to quantify the added attractiveness of sanitation on the marriage market.

4.1 Matching and Gains from Marriage

We begin by describing the CS transferable utility model of a marriage market, where the unique Indian context allows for multiple observations of the relevant marriage market g defined over caste and region.[23] In particular we consider two large finite populations of males and females to be matched on the basis of their wealth type. Men and women within each group g differ only in their wealth defined by a wealth index $I(W_\kappa)$ and $J(W_\kappa)$ respectively. The distribution of wealth has a finite support K and each individual belongs to a finite set of wealth classes $W = \{W_1, ..., W_K\}$. Let $\mathbf{M}$ ($\mathbf{F}$) be the vector of available

[21] The limitations of using a regression framework on individual level data to account for general equilibrium effects are well documented. For example, Angrist (2002) finds the causal effect of changes in the sex ratio on the male marriage rate to be inconsistent with that found for the female marriage rate.

[22] This characterization is similar to that undertaken by Chiappori et al. (2017); Choo and Siow (2006b).

[23] We subsume all g subscript notation and return to this feature for the estimation of the model.

male (and female) types, the I^{th} and J^{th} element of which denoted the measure of different types of men m_I and women f_J respectively. Each man $i \in I$ and women $j \in J$ must chose a type of partner. The choice of remaining unmatched is captured by augmenting the type space for men and women, e.g. $I = \{0, 1, ..., I, ..., \mathcal{I}\}$, where choosing 0 by woman J corresponds to remaining single and vice versa for males.[24] The total marital gains generated by a match π_{ij} between $i \in I$ and $j \in J$ is given by

$$\pi_{ij} = \pi_{IJ} + \tilde{\varepsilon}_i^I(J) + \tilde{\eta}_j^J(I) \tag{2}$$

where in addition to the systematic match surplus π_{IJ}, a match between a type I male and type J is also subject to a vector of gender specific preference shocks. The preference heterogeneity components $\tilde{\varepsilon}$ and $tilde\eta$ are assumed to be additively separable in the gains function and depend only on the partner's type I (or J), not his or her exact identity i (or j). The matrix $\Pi = (\pi_{IJ})$ denotes aggregate surplus. The preference shocks often characterised as a love shock can be interpreted, in a parsimonious way, as the "quality" of the match (Chiappori, 2020). The systematic gains to marriage π_{IJ} in Equation (2) also referred to as the joint marital surplus will be the main object of interest in our empirical analysis.

4.2 Adding Sanitation to Gains from Marriage

Since we only observe the sanitation status together with the matched status of individuals we model sanitation as a type of living arrangement l, where $l \in \{T, NT\}$ denote living arrangement with and without sanitation, respectively.[25] In other words a living arrangement establishes types of marriages which individuals can choose to sort into and thereby allows us to incorporate sanitation without a priori assumptions on the underlying preferences.

To incorporate the dependence between match quality and living arrangement which generates sorting on unobservables we amend Equation (2) under specific assumptions on tastes for hygiene which amount to relaxing the strict separability in Equation (2).[26] Following CSW we assume that

$$\tilde{\varepsilon}_i^I(J) = A_I^J + \varepsilon_i^{IJ}$$
$$\tilde{\eta}_j^J(I) = B_J^I + \eta_j^{IJ}$$

where A_I^J and B_J^I are type-specific means reflecting preference for hygiene which may may vary across the wealth distribution, and ε_i^{IJ} and η_j^{IJ} are standard, type 1 extreme value with zero mean. Under this

[24] Female type space $J = \{0, 1, ..., J, ..., \mathcal{J}\}$

[25] Stopnitzky (2012) incorporates a Chiappori, Iyigun Weiss (2009) framework to analyse premarital investment response in sanitation take-up under transferable utility and gender differences in preferences. In contrast, our analysis focuses on quantifying the relative attractiveness of sanitation on the marriage market. Specifically, we model spousal and living arrangement as a simultaneous choice without a priori restrictions on the preference parameters across gender.

[26] Galichon and Salanie (2021) and Chiappori and Salanié (2016) provide a detailed overview of the challenge in incorporating unobserved heterogeneity in a two-sided matching model without additional distributions assumptions.

assumption Equation (2) becomes

$$\pi_{ij} = \underbrace{Z^{IJ} + A_I^J + B_J^I}_{\pi_{IJ}} + \varepsilon_i^I(J) + \eta_j^J(I) \tag{3}$$

where the systematic gains $\pi_{IJ} = Z^{IJ} + A_I^J + B_J^I$ can be reinterpreted to include monetary gains Z^{IJ} and individual type specific preference means capturing non-monetary gains from marriage. Even though A_I^J and B_J^I are not separately identified, the extension of martial gains function in Equation (3) and the variation across living arrangements, discussed in the next section, allows us to incorporate a match-specific love-for-hygiene component $\delta^{IJ} = A_I^J + B_J^I$. Intuitively, conditional on sorting into different living arrangements the distribution of marital preference will depend on these mean preferences for hygiene, driven by the selection operated on the underling preference for spouses and sanitation.

In addition, as discussed, the spousal marriage rates depicted in Figure (6) do not take into account changes in the number of available matching opportunities in the market post TSC exposure. To incorporate this policy induced change in the composition of available marriage types, we first follow the methodology introduced by Choo and Siow (2006b) to generate a marriage matching function (MMF) for each marriage types l in the next section.[27] Specifically, we incorporate marriage types or living arrangements which delivers an additional set of equilibrium relationships which define the marriage matching functions to calculate the gains to marriage. By allowing for different marriage types each individual can decide whether to marry in an arrangement with or without sanitation or to remain unmatched.

4.3 Marital Preferences over Partner Types and Living Arrangements

Let the payoff of a man i of wealth type I who matches with a woman of wealth type J in a living arrangement l with own idiosyncratic preferences ε_i^{IJ} be given by

$$U_{iIJ}^l = \tilde{u}_{IJ}^l - \tau_{IJ}^l + \varepsilon_i^{IJ} \quad l \in \{T, NT\} \tag{4}$$

where $\tilde{u}_{IJ}^l$ denotes the systematic gross return common to all males of type I matching to a female of type J in living arrangement l. For each type I man to match with a type J woman in living arrangement l, he must transfer to her a part of his utility that he values at τ_{IJ}^l. The idiosyncratic component of male marital preference measures the departure of each individual male i's payoff U_{iIJ}^l from the systematic component which is common to all male type I who marry females of type J. Similarly the payoff for a male i of type I who remains unmatched ($J = 0$) is given by:

$$U_{iI0} = \tilde{u}_{I0} + \varepsilon_i^{I0} \tag{5}$$

where $\tilde{u}_{I0}$ denotes the systematic gain that is common to all type I males from remaining unmatched. In the empirical specification we relax the assumption that men and women face similar gains from

[27] Choo and Siow (2006a) use the matching function to study marriage and cohabitation patterns in Canada.

remaining unmatched. Moreover, the inclusion of living arrangements, allows us to specify the joint choice of partner and living arrangement with or without sanitation. Specifically, the underlying decision that would govern each male type I behaviour maximizes the objective function given by Equation (6).

$$U_{iI} = \max_{J,l}\{U_{iI0}, U^{T}_{iI1}, ..., U^{T}_{iIJ}, ..., U^{T}_{iI\mathcal{J}}, U^{NT}_{iI1}, ..., U^{NT}_{iIJ}, ..., U^{NT}_{iI\mathcal{J}}\} \tag{6}$$

The random utility payoffs for women have a similar form. Specifically, the choice specific payoff of a woman j of wealth type J who marries a man of wealth type I includes own idiosyncratic preferences η^{IJ}_{j} along with systematic gross return $\tilde{v}^{l}_{IJ}$. The woman values the transfer as τ^{l}_{IJ} which maybe positive or negative.

$$V^{l}_{jIJ} = \tilde{v}^{l}_{IJ} + \tau^{l}_{IJ} + \eta^{IJ}_{j} \quad l \in \{T, NT\} \tag{7}$$

Lastly, the payoff for a female j of type J who remains unmatched ($I = 0$) is given by:

$$V_{j0J} = \tilde{v}_{0J} + \eta^{0J}_{j} \tag{8}$$

where $\tilde{v}_{0J}$ denotes the systematic gain for a single type J women.

The payoff specifications above are analytically convenient allow us to incorporate sanitation. The idiosyncratic marital preference for individual men i and women j are assumed to be independently and identically distributed with a type I extreme value distribution. In addition the idiosyncratic marital preference only depends on their spousal type I or J and not their individual identity i or j.

4.4 Marriage Market Equilibrium

The matching market clears when, given equilibrium transfers τ^{l}_{IJ}, the demand by type I males is equal to the supply of females of type J for all (l, I, J)

$$(\mu^{l}_{IJ})^{d} = (\mu^{l}_{IJ})^{s} = \mu^{l}_{IJ}; \; l = (T, NT) \tag{9}$$

substituting (9) into the quasi demand and supply equations above yields the market equilibrium moment condition where the measure of type I men who chose to marry type J women equals the measure of type J women who chose to marry type I men. The equilibrium distribution of marriages is a function of population vectors and exogenous parameters that determine the systematic and idiosyncratic payoffs. Thus, this yields the equilibrium distribution of marriages as a Marriage Matching Function (MMF) in Equation (10) under different living arrangements.

$$\pi^{l}_{IJ} = \ln \frac{{\mu^{l}_{IJ}}^{2}}{\mu_{I0} \cdot \mu_{0J}} \quad for\, l \in \{T, NT\} \tag{10}$$

Equation (10) has an intuitive interpretation where the left hand side reflects the total marital gain for any couple (I, J) matched in living arrangement l relative to the total gain to them from remaining unmarried. The right hand side denotes a log transformation of the marriage matching function characterized by the ratio of number of I, J marriages to a geometric average of the number of singles. To account for how the presence of sanitation in the match affects the gains to marriage we follow CS and measure the numerator of Equation (10) alternatively by the number of marriage living arrangements l with and without sanitation $l \in \{T, NT\}$.

The marriage matching function under each living arrangement $\mu(M, F; l,)$ returns a $\mathcal{I} \times \mathcal{J}$ matrix, where the element μ^l_{IJ} denotes the measure of marriages between type I men and type J women in living arrangement l. The matching function must satisfy the following accounting conditions:

$$\mu_{0J} + \sum_l \sum_{I=1}^{K} \mu^l_{IJ} = f_J \ \forall J \ \ l = (T, NT) \tag{11}$$

$$\mu_{I0} + \sum_l \sum_{J=1}^{K} \mu^l_{IJ} = m_I \ \forall I \ \ l = (T, NT) \tag{12}$$

$$\mu_{0J}, \mu_{I0}, \mu^l_{IJ} \geq 0 \ \forall \ \ l, I, J \tag{13}$$

where the first equation denotes that the total number of men who marry type J women and the number of unmarried type J women must be equal to the number of available type J women, for all J. While the second equation provides a set of accounting conditions that must hold for all male types. The last accounting constraint holds because the number of unmarried persons of any type and gender and the number of marriages between type I men and type J women must be non-negative. These accounting constraints are crucial towards ensuring that the predicted marriage rates are not above one i.e., the matching is feasible.

4.5 Empirical Implementation

Identification

The crucial feature of our identification approach, like others in the literature, is that we observe and thereby are able to exploit multiple observations of the marriage market. While several papers have made use of time variation in male and female characteristics in the population, this requires assumptions on the independence across different age cohorts. In this paper we combine quasi-random variation from the TSC policy with non-overlapping marriage markets along caste and district to generate over-identifying restrictions from cross-sectional variation from the IHDS sample while controlling for time effects across the two survey rounds.

Estimation

To bring the model to data we implement the Type (1) extreme value assumption for the idiosyncratic marital tastes.[28] The model is estimated in two parts using a two-step conditional choice probability (CCP) estimator. First stage estimates the sub-marriage market level demand and supply using male and female quasi-demand equations for each spouse type conditional on their own type with additional identifying restrictions using our caste and region demographic groups. In the second stage we use first stage estimates for marital preference and the distribution of wealth type for both men and women, to estimate the total gains from marriage for each type of match relative to single-hood. Specifically, to identify the surplus matrix Π and generate testable restrictions we include gender specific drifts ν^I and ζ^J that vary by caste and region categories denoted by g, such that the Π^{IJ} terms vary according to:

$$\Pi_g^{IJ} = \Pi^{IJ} + \nu_g^I + \zeta_g^J \tag{14}$$

The drifts absorb possible changes across castes and regions in the surplus generated by marriage capturing taste heterogeneity across different cultural and socio-economic groups. Marital preference estimates are used to solve a system of equalities derived from the key identified moment of the matching model under the assumption of transferable utility.[29]

$$\log\left(\hat{P}(J|I,g)\cdot\hat{P}(I|J,g)\right) - G_{IJ} - E_g^I - F_g^J = 0 \tag{15}$$

Using our sample of multiple non-overlapping groups and within group exposure to the TSC adds to the total number of equations and thereby generates additional identifying variation. The parameters of interest here are $\left(G_{IJ}, E_g^I, F_g^J\right)$, that together capture and provide an estimate of the marital gains. The parameters are estimated using a Minimum Distance Estimator (MDE) as outlined in Chiappori et al. (2017). For the Choo and Siow (2006b) framework, Decker et al. (2013) show that the solution is locally unique.

4.6 Model Estimates

Systematic/Total Gains

To show the difference in systematic gains between marriages with and without sanitation - we show in Figure (7) the systematic total gains in marriages for different living arrangements i.e., with π_{IJ}^T and without π_{IJ}^{NT} sanitation for couples of the same wealth type in 2005 prior to TSC exposure. Recall that the systematic (or total) gain measures the expected marital surplus to a random (I, J) pair marrying relative to them not marrying.

[28] This is a standard assumption and widely employed within the marriage market literature for parametrizing the distribution of the shocks see for e.g. Choo and Siow (2006b), Chiappori et al. (2017)

[29] In practice we exploit variation across a total of 20 groups ($g = 1, ..., 20$) using caste affiliation and region.

Figure 7: Systematic Gains under different living arrangements in 2005

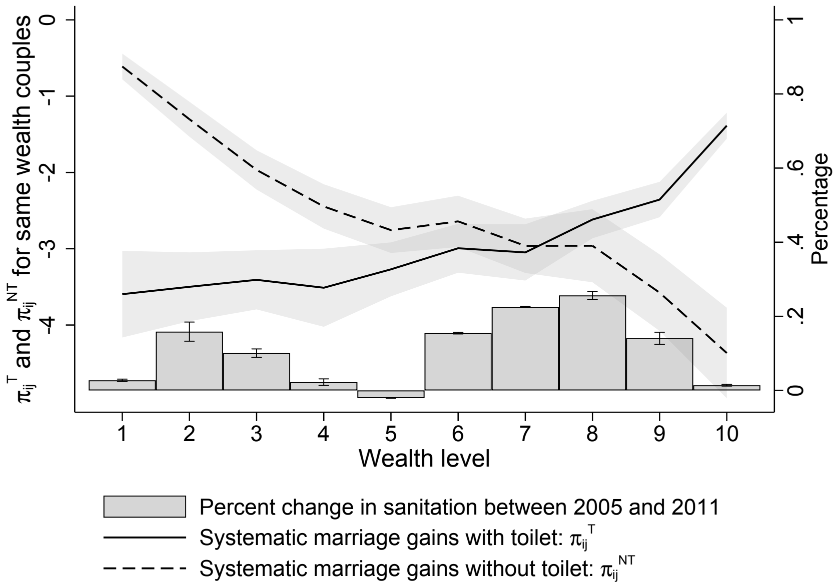


Figure (7) highlights two important points. First, the systematic gains under both living arrangements lie below zero indicating the large fraction of younger women and men in our sample. An average annual population growth rate of 1.44% between 2005 and 2011 (World Bank) would necessarily imply a further decrease in the systematic gain from marriage. Moreover, these composition changes due to population shifts are distinct from any changes in marital preferences but are observationally equivalent.

Second, the systematic gains to marriage with sanitation are not unilaterally larger but instead vary with the wealth distribution across the sub-marriage markets. Highlighting the scope for large heterogeneous policy effects. Specifically the systematic gains to marriage with sanitation are higher than that of marriages without only among the wealthiest of the same wealth matches. In our framework the equilibrium effects operate on both the type of living arrangement and type of spouse. In the first case the increase in living arrangements which come with sanitation, with TSC exposure, will imply a necessary decrease in the systematic gains π_{IJ}^{T} vs π_{IJ}^{NT}. In the second case, heterogeneous prevalence of living arrangements across the wealth distribution may generate composition changes in the choice of spouses. Therefore to quantify the total effect from TSC exposure would necessitate fixing the level of systematic gains.

Sorting Outcomes

In this section we use the estimated equilibrium model to analyse how the total sanitation campaign affected outcomes in the marriage markets. Table 3 presents how the probability of marriage between types is affected by the program intensity. The estimates in Table 3 captures how exposure to TSC changes the overall gains from entering the marriage market and is captured in the choice of *whether* to marry. We find that both poor men and poor women are more likely to marry in equilibrium with TSC exposure. This increase in probability of marriage with TSC exposure suggests an increase in gains to entering the marriage market for men and women. Moreover, we would expect this difference-in-difference in the gains to entering the marriage market to be concentrated (largest) among individuals at the lower end of the wealth distribution. [30]

Table 3: Probability of Marriage

	Men				Women			
	High TSC	Low TSC	Change	p-val	High TSC	Low TSC	Change	p-val
Wealth Type L	0.473	0.434	0.038	**0.042**	0.680	0.615	0.065	**0.006**
Wealth Type M	0.383	0.374	0.009	0.503	0.490	0.503	-0.012	0.478
Wealth Type H	0.361	0.360	0.001	0.635	0.401	0.398	0.003	0.838

Notes: Probability of marriage with caste, region and time group effects. Wealth Types L,M,H refer to low, medium and high wealth respectively Household level controls include age, assets, Education of Household Head. Standard errors clustered.

The partner choice for men and women in Table (4) and Table (5) rationalize patterns that we observe previously. By definition, the sum of the shares across different types of spouses for a given type of male or female is one. Table (4) shows a clear pattern for men where with TSC exposure men were more likely to marry "down" and less likely to marry "up".[31] However, the composition changes induced by TSC for women are less clear. Table (5) shows that while non-poor women were less-likely to marry "down" this change was not accompanied by a corresponding shift to marry "up" among poor and non-poor women alike.

[30] Appendix tables 14 and 13 presents the same results for high and low sex ratio districts separately. Results are similar to those presented here.

[31] Sorting at the top of the wealth distribution showed no discernible changes.

Table 4: Partners of Men

Wife →	Wealth Type L	Wealth Type M	Wealth Type H
Husband ↓		High TSC	
Wealth Type L	0.703	0.256	0.041
Wealth Type M	0.324	0.369	0.307
Wealth Type H	0.073	0.247	0.679
Husband ↓		Low TSC	
Wealth Type L	0.640	0.299	0.061
Wealth Type M	0.245	0.453	0.301
Wealth Type H	0.046	0.254	0.700
Husband ↓		Change	
Wealth Type L	0.063**	-0.043*	-0.020
Wealth Type M	0.078***	-0.084***	0.006
Wealth Type H	0.027	-0.006	-0.021

Notes: Conditional Probability of matching with Woman by wealth type with caste, region and time group effects. Wealth Types L,M,H refer to low, medium and high wealth respectively. Household level controls include age, assets, Education of Household Head. Standard errors clustered.

As such the fall in the marriage shares for women where the husband was as wealthy or less would be the equilibrium response to the increase in men who are now marrying "down". It is important to note that while Table (4) and Table (5) indicate changes in spousal choice i.e., whom to marry they do not incorporate the decision of whether or not to marry. In contrast the marriage matching function in equation (10) encapsulates both these effects into a single summary statistic. In the subsequent section, we use this summary statistic to analyse the total effect of the TSC looking at changes to both the marital surplus and the expected surplus share. Overall, the changes in marital choices in Table (4) and Table (5) induced by the Total Sanitation Campaign illustrate the importance of heterogeneous changes in marital surplus.

Table 5: Partners of Women

Husband →	Wealth Type L	Wealth Type M	Wealth Type H
Wife ↓		High TSC	
Wealth Type L	0.647	0.273	0.079
Wealth Type M	0.196	0.391	0.413
Wealth Type H	0.023	0.208	0.769
Wife ↓		Low TSC	
Wealth Type L	0.625	0.303	0.072
Wealth Type M	0.252	0.422	0.326
Wealth Type H	0.043	0.228	0.729
Wife ↓		Change	
Wealth Type L	0.023	-0.030	0.007
Wealth Type M	-0.056***	-0.031**	0.087
Wealth Type H	-0.020*	-0.020**	0.040

Notes: Conditional Probability of matching with Woman by wealth type with caste, region and time group effects. Wealth Types L,M,H refer to low, medium and high wealth respectively. Household level controls include age, assets, Education of Household Head. Standard errors clustered.

5 How does the TSC affect Marriage Markets

The analysis below is specific to the structural approach and decomposes the impact of the policy on marital surplus and expected female surplus share using simulations from the estimated model.

5.1 Effect of TSC on Marital Surplus

In order to empirically quantify the impact of the TSC policy on the gains to marriage, we consider variation to a type of match characterised by wealth and whether the matched couple belongs to a high or low TSC district in each year. The change in the total match surplus for an (I,J) match type in the High TSC exposed districts can be estimated by a difference in difference type estimator [32]

Table (6) shows the change in the gains to marriage across district exposure for spouses of different wealth types. We find that households with marriageable men and women exposed to the TSC policy

[32]This is similar to the approach implemented by Choo and Siow (2006b) to analyse the impact of the partial legalization of abortion in the US on the total gains to marriage.

experience a heterogeneous impact in match surplus for a given match. The total surplus from a match of a given type for each sub marriage market is given in Table (6) where a small negative number denotes a larger match specific surplus. Overall the match surplus increases in 5 out of total of 9 match types concern couples formed where the man is at least as wealthy as their spouse if not more i.e., lower triangle of the match surplus matrix. Table (6) highlights the importance of the general equilibrium effects generated by the Total Sanitation Campaign on the redistribution of gains across different sub-markets or matched types.

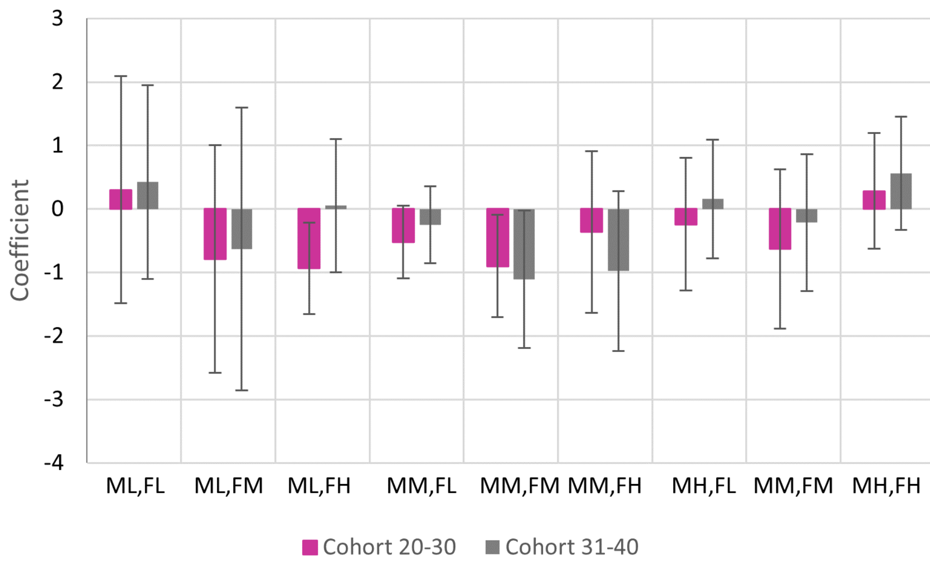
Table 6: Match Surplus

Wife →	Wealth Type L	Wealth Type M	Wealth Type H
Husband ↓		High TSC	
Wealth Type L	-0.788	-2.991	-6.969
Wealth Type M	-2.425	-1.936	-2.751
Wealth Type H	-5.146	-2.281	-0.649
Husband ↓		Low TSC	
Wealth Type L	-0.918	-2.586	-5.931
Wealth Type M	-2.598	-1.653	-2.678
Wealth Type H	-5.712	-2.493	-0.673
Husband ↓		Change	
Wealth Type L	**0.130**	-0.405	-1.038
Wealth Type M	**0.173**	-0.282	-0.074
Wealth Type H	**0.565**	**0.211**	**0.024**

Notes: Men in rows and women in columns. Wealth Types L,M,H refer to low, medium and high wealth respectively. Bold indicates increase in total match surplus.

As outlined in Section 3.3, we do not have data to assess if the identified differences could be obtained with information prior to the TSC implementation. However, to provide robustness, we can do a comparison against older to cohorts to assess if the parallel trends assumption is plausible. Figure (8) presents the result of such exercise, showing the coefficients for the nine combinations of male and female wealth categories (ML/M/H: male low/medium/high wealth; FL/M/H: female low/medium/high wealth). We find that expected total gains are stable across older cohorts, specifically, 31-40 vs 41-60 (further details are available in appendix (B)).

Figure 8: DiD coefficients for the impact of TSC on expected total gains through comparison of cohorts



Note. These figures correspond to the coefficients, and their 90% confidence intervals, after a DiD regression that compares the difference of expected total gains of cohorts 20-30 (treated) and 31-40 (non-treated) with cohort 41-60 (control), between low and high TSC exposure districts. The coefficients and its standard errors are availabel in Table (17) in appendix B.

5.2 Effect of TSC on Surplus Share

While our context does not allow for any exogenous variation in the population supplies we are able to use model simulations to analyse the impact of the TSC program on different underlying marriage market conditions.[33] Using the estimated parameters for the systematic gains from marriage under different living arrangements Table (7) illustrates the overall impact of the TSC on the wife's expected surplus share. Even though the relative scarcity of women is thought to imply a larger sharing weight the net equilibrium impact of the TSC seems to have eroded away these female gains. In other words, while for some type of couples there are benefits for getting married as seen in Table (6), wives will have less control of the resources of the household. The counterfactual estimates from the model in Table (8) show us that some women, in particular the poorest, would have experienced an increase in their expected surplus share if the gender differences were removed. TThe result is particularly striking in light of the policy implications but straightforward once equilibrium effects are taken into account.

[33]Note that there is very little regional variation in sex ratio with over 80% of high sex ratio districts located in the South and East of the country, according to 2011 census data. As such implied "shortage" of women and unobserved differences in marital preferences between the South and the rest of the country are observationally equivalent.

Table 7: Wife's Expected Surplus Share (Female Shortage)

Wife →	Wealth Type L	Wealth Type M	Wealth Type H
Husband ↓		High TSC	
Wealth Type L	0.623	0.537	0.455
Wealth Type M	0.668	0.586	0.504
Wealth Type H	0.684	0.603	0.523
Husband ↓		Low TSC	
Wealth Type L	0.635	0.563	0.480
Wealth Type M	0.673	0.603	0.521
Wealth Type H	0.681	0.612	0.530
Husband ↓		Change	
Wealth Type L	-0.012	-0.026	-0.025
Wealth Type M	-0.004	-0.018	-0.017
Wealth Type H	0.004	-0.009	-0.007

Notes: Men in rows and women in columns. Wealth Types L,M,H refer to low, medium and high wealth respectively.

Table (7) and (8) also illustrate the importance of equilibrium effects where the underlying sharing weight is not a fixed parameter solely dependent on the scarcity of women relative to men. More generally the patterns of intra-household distribution of resources and welfare are not invariant primitive parameters but endogenous entities reflecting marriage market conditions (Chiappori et al., 2018). For example, along with matching attributes the surplus share of husbands and wives may be also determined by their willingness to enter the marriage market. From the tables we see that the average surplus share for women in markets with female shortage is 0.582 while the average surplus share in markets with an excess of women is 0.581. A corresponding two-tailed p-value of 0.927 of the mean difference in female share further illustrates that in fact, in our context, there are no fixed differences in the surplus share associated with the supposed scarcity of women.

Table 8: Wife's Expected Surplus Share (No Female Shortage)

Wife →	Wealth Type L	Wealth Type M	Wealth Type H
Husband ↓		High TSC	
Wealth Type L	0.720	0.434	0.388
Wealth Type M	0.797	0.540	0.492
Wealth Type H	0.826	0.586	0.539
Husband ↓		Low TSC	
Wealth Type L	0.597	0.490	0.435
Wealth Type M	0.675	0.573	0.519
Wealth Type H	0.698	0.600	0.546
Husband ↓		Change	
Wealth Type L	0.122	-0.055	-0.048
Wealth Type M	0.122	-0.033	-0.027
Wealth Type H	0.127	-0.013	-0.007

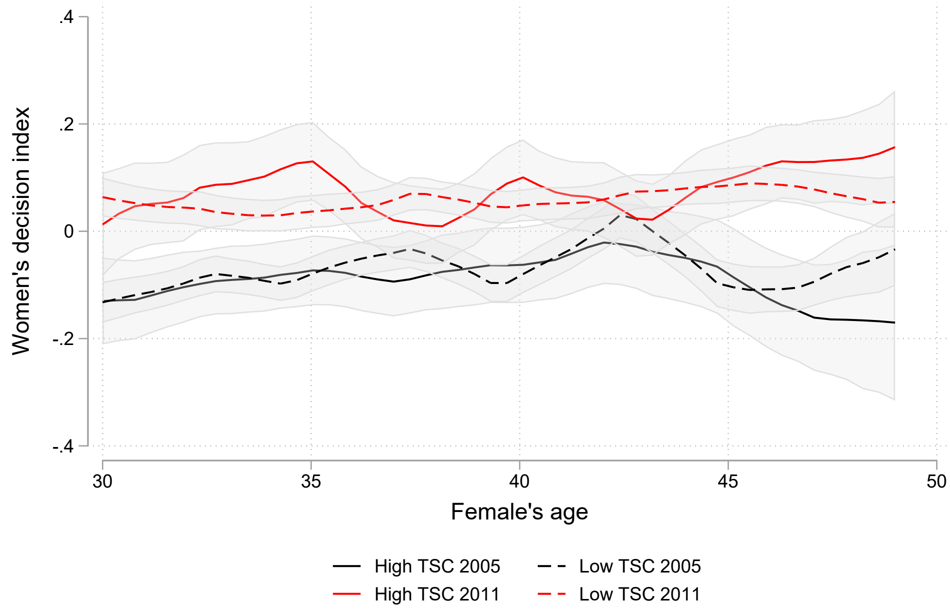
Notes: Men in rows and women in columns. Wealth Types L,M,H refer to low, medium and high wealth respectively.

To provide corroborative evidence of our model estimates on surplus division, we construct a measure of female financial decision making within the household. We construct a proxy of intra-household allocation using an independent measure of female financial decision making from the IHDS. Women are asked who has the say in buying an extensive item such as a TV or a fridge; they could answer that she, her husband, other woman, other man, or someone else; and if it is more than one person, who has the more say. In our index, 0 corresponds to no women having a say, 1 to the respondent or a senior women having a say, and 2 if any of them have the most say. [34] Figure (9) shows that for women above 30, there is no difference of this variable between high and low TSC districts, and is also stable across cohorts. The only difference is between waves of the IHDS.[35] This pattern indicates that parallel trends over control of resources is a plausible assumption. We use a wave-standardised version of the index with mean 0 and standard deviation 1.

[34]The questionnaire includes other decisions such as how to cook in a daily basis or how many children to have. As our central concern is about control of resources, we focus only on the reported question which is related to financial resources and is present in both IHDS waves. The mean of index before being standardised in 2004 is of 0.847 and SD of 0.596. In 2011, the mean is 0.94 and SD of 0.55.

[35]This could be either because of a generalised cultural change in the recent years, but also because the 2011 questionnaire includes a longer gender relations questionnaire.

Figure 9: Women decision index for individuals not in the marriage market, by TSC implementation group and IHDS wave



Note. Local polynomial smoother using an Epanechnikov kernel. Includes 95% confidence intervals.

Table (9) presents the regression estimates of TSC exposure on the female financial decision making measure, which are positive and insignificant. It is important to note that these estimates, which are not exact model moments, are qualitatively comparable if not in magnitude. Moreover, as an independent observation of intra-household behaviour they serve as a robustness check on our model estimates on surplus division. Appendix Table (16) provides an alternate specification with finer cuts of the sample used above in Table (9) to include additional subgroups on the wealth measure. The results are qualitatively similar and in line with the main findings from the model.

Table 9: TSC impact on women decision index by wealth and sex ratio

	Low Sex Ratio		High Sex Ratio	
	(1)	(2)	(3)	(4)
Wealth Type:	Low wealth	High Wealth	Low Wealth	High Wealth
dT	-0.0233	-0.0538	0.118	-0.123
	(0.0575)	(0.0597)	(0.124)	(0.121)
TSC	-0.0582	-0.134	-0.0597	-0.131
	(0.108)	(0.111)	(0.143)	(0.162)
$dT \times TSC$	0.0673	0.0753	0.192	0.179
	(0.124)	(0.117)	(0.205)	(0.192)
Constant	-0.505**	-0.0404	-0.0676	-0.126
	(0.240)	(0.231)	(0.346)	(0.272)
Observations	8,790	6,437	2,107	1,252
R^2	0.018	0.029	0.107	0.075

Notes: Own calculations using data from the IHDS waves 2004 ($dT = 0$) and 2011 ($dT = 1$). The sample consists of households of all single and married females (males) aged 15 to 34 at the survey time. High Sex Ratio corresponds to districts with at least 999 women per 1000 men in the age range 15-34 (Columns 3 and 4), while Low Sex Ratio is districts with less than 999 (Columns 1 and 2). District level sex ratio information was computed using data from the population census 2001 and 2011. High wealth correspond to individuals whose asset index is above the 50 percentile of the entire country distribution per wave (Columns 2 and 4). Those households below such cutoff are classified as low wealth (columns 1 and 3). $TSC = 1$ correspond to a grand score for implementation of 61 or above, and $TSC = 0$ to those with a score of 61 or above. The grand score per district is taken from WSP (2011). Apart from the coefficients presented in the table, as controls, we consider the age and marital status of the individual for whom the household is in the sample; the wealth index of the household; household size; education level of the household head (no education [base], primary, incomplete secondary, secondary, and above secondary); caste (Brahmin [base], High caste, Other backward caste, Dalit, Adivasi, Muslim, Sikh - Jain); an indicator variable for the rural zone; and fixed effects at the state level. Clustered at the district level standard errors in parentheses. Significance: *** p<0.01, ** p<0.05, * p<0.1

6 Conclusion

While the case for investment in sanitation has generally been convincingly made, there remains an incomplete understanding of impacts of sanitation interventions (Augsburg and Sainati (2020); Gautam (2017)). This maybe due to the complexity and heterogeneity of sanitation, which implies direct monetary and non-monetary costs and benefits, many of which are hard to measure, as well as positive and negative externalities. Gautam (2020) highlights the importance of inter-household interactions and quantifies the welfare effects in the presence of externalities from sanitation through an ex-ante approach. Nevertheless, there may be indirect impacts on individuals via alternative markets, such as marriage markets, which to date, have been largely ignored in the literature. In this paper we show that sanitation matters in marriage markets: the Indian Total Sanitation Campaign changed marriage market outcomes. We exploit in this paper quasi-random variation from the TSC, that shifted the distribution of households with sanitation, and the incentives of men and women to sort into marriage, or not. We show that exposure to a high-performing TSC district had a significant impact on both the composition of marriage types, and the sorting into marriages by men and women. The analysis relies on reduced-form techniques and a structural approach. Both approaches exploit information on marriages - including the chosen living arrangement of couples - across geographical areas and time. Identification is achieved through a difference-in-differences approach in a multi-market framework. For the structural analysis we develop and estimate a simple friction-less matching model à la Choo and Siow (2006b), allowing for different types of living arrangements through a couple-specific random component affecting marriage surplus and other socio-economic spousal characteristics, such as wealth.

To analyse the importance of marriage market conditions and to illustrate the overall sorting effect we use the estimated equilibrium model to conduct two model simulations. First, we show that cohorts within TSC exposed markets experienced a shift in marital gains both across matches but also within a given match. Specifically, the resultant sorting patterns display a marked gender asymmetry with an increase in marital surplus among matches, where men are wealthier than their spouse, and a decrease in surplus where the wife is wealthier. Second, we consider the role of relative scarcity of women within the marriage market and find that the increase in female gains were negligible. Instead, the TSC had an unintended countervailing impact on female empowerment as measured by the effective control over household resources and reflected in the division of surplus. Our results indicate that this increased sorting of men and women into households with sanitation on the marriage market came at the expense of women losing control of household resources and a decline in female empowerment within marriage. While evaluations of the TSC generally highlighted the policy's positive impacts on sanitation uptake, our paper draws attention to an important indirect effect, that might well be present in other such policies, including the 'No toilet, no bride' campaign. We argue that – to obtain a comprehensive picture of such female-focused policies including costs and benefits – their interaction with marriage markets should not be ignored.

The insights derived from our empirical analysis further highlight interesting avenues for future research that are beyond the scope of this paper. In particular, the importance of sanitation in the marriage market determined by the magnitude of unobserved heterogeneity raises interesting questions on the presence and nature of frictions specifically on the marriage market. Even though the transferable utility assumption limits our ability to explore such questions, our results emphasize a promising avenue to be explored in future research. In addition, while the marriage matching function in our analysis is static, it may be reasonable to expect additional aspects of marital behaviour, in response to policy exposure, to be dynamic e.g. marriage delay. Moreover, this response may differ across men and women. In such a case, we would need a dynamic marriage matching function. While a dynamic matching model lies beyond the primary focus of this paper, an important contribution of this paper is to provide a rigorous structural foundation that can be extended to explore resulting dynamics from the TSC. In summary, our analysis emphasises the importance of accounting for general equilibrium effects which necessitates going beyond reduced form methods and yet has been largely missing from the policy discourse within the literature at the intersection of sanitation and marriage markets.

References

Abramitzky, R., Delavande, A., and Vasconcelos, L. (2011). Marrying up: the role of sex ratio in assortative matching. *American Economic Journal: Applied Economics*, 3(3):124–57.

Abramovsky, L., Augsburg, B., Luuhrmann, M., Oteiza, F., and Rud, J. P. (2019). Community matters: heterogenous impacts of a sanitation intervention.

Adams-Prassl, A. and Andrew, A. (2019). Preferences and beliefs in the marriage market for young brides.

André, P. and Dupraz, Y. (2019). Education and polygamy: Evidence from Cameroon. Technical report, Warwick Economics Research Papers No 1219.

Andres, L. A., Deb, S., Joseph, G., Larenas, M. I., and Grabinsky Zabludovsky, J. (2020). A multiple-arm, cluster-randomized impact evaluation of the Clean India (Swachh Bharat) Mission program in rural Punjab, India. Working Paper 9249, World Bank".

Angrist, J. (2002). How do sex ratios affect marriage and labor markets? Evidence from America's second generation. *The Quarterly Journal of Economics*, 117(3):997–1038.

Arnold, B. F., Khush, R. S., Ramaswamy, P., London, A. G., Rajkumar, P., Ramaprabha, P., Durairaj, N., Hubbard, A. E., Balakrishnan, K., and Colford, J. M. (2010). Causal inference methods to study non-randomized, preexisting development interventions. *Proceedings of the National Academy of Sciences*, 107(52):22605–22610.

Attanasio, O. and Kaufmann, K. (2017). Education choices and returns on the labor and marriage markets: Evidence from data on subjective expectations. *Journal of Economic Behavior Organization*, 140.

Augsburg, B. and Sainati, T. (2020). Editorial: WASH economics and financing: towards a better understanding of costs and benefits. *Journal of WASH for Development*, 10:615—-617.

Banerjee, A., Duflo, E., Ghatak, M., and Lafortune, J. (2013). Marry for what? Caste and mate selection in modern India. *American Economic Journal: Microeconomics*, 5(2):33–72.

Barnard, S., Routray, P., Majorin, F., Peletz, R., Boisson, S., Sinha, A., and Clasen, T. (2013a). Impact of Indian Total Sanitation Campaign on latrine coverage and use: A cross-sectional study in Orissa three years following programme implementation. *PLOS One*, 8.

Barnard, S., Routray, P., Majorin, F., Peletz, R., Boisson, S., Sinha, A., and Clasen, T. (2013b). Impact of indian total sanitation campaign on latrine coverage and use: a cross-sectional study in orissa three years following programme implementation. *PloS one*, 8(8):e71438.

Basu, A. (1999). Fertility decline and increasing gender imbalance in India, including a possible South Indian turnaround. *Development and Change*, 30:237–263.

Beauchamp, A., Calvi, R., and Fulford, S. (2017). Terms of engagement: Marriage and migration in India. In *Structural Models in Development: Migration, Marriage and the Family*. Econometric Society.

Becker, G. S. and Becker, G. S. (2009). *A Treatise on the Family*. Harvard university press.

Borker, G., Eeckout, J., Luke, N., Minz, S., Munshi, K., and Swaminathan, S. (2018). Wealth, marriage, and sex selection. PAA 2018 Annual Meeting.

Borker, G., Eeckout, J., Luke, N., Minz, S., Munshi, K., and Swaminathan, S. (2019). Wealth, marriage, and sex selection. Working paper.

Botticini, M. and Siow, A. (2003). Why dowries? *American Economic Review*, 93(4):1385–1398.

Briceno, B., Coville, A., Gertler, P., and Martinez, S. (2017). Are there synergies from combining hygiene and sanitation promotion campaigns: Evidence from a large-scale cluster-randomized trial in rural Tanzania. *PLoS One*, 12(11):e0186228.

Cavill, S., Mott, J., Tyndale-Biscoe, P., Bond, M., Edström, J., Huggett, C., and Wamera, E. (2018). Men and boys in sanitation and hygiene: A desk-based review. CLTS Knowledge Hub and Learning Paper, Institute of Development Studies.

CBGA, U. (2011). Total sanitation campaign (TSC). budgeting for change series, 2011. Centre for Budget and Governance Accountability; Social Policy, Planning, Monitoring and evaluation (SPPME), UNICEF.

Charles, K. K. and Luoh, M. C. (2010). Male incarceration, the marriage market, and female outcomes. *The Review of Economics and Statistics*, 92(3):614–627.

Chiappori, P.-A. (2020). The theory and empirics of the marriage market. *Annual Review of Economics*, 12:547–578.

Chiappori, P.-A., Dias, M. C., and Meghir, C. (2018). The marriage market, labor supply, and education choice. *Journal of Political Economy*, 126(S1):S26–S72.

Chiappori, P.-A., Oreffice, S., and Quintana-Domeque, C. (2012). Fatter attraction: anthropometric and socioeconomic matching on the marriage market. *Journal of Political Economy*, 120(4):659–695.

Chiappori, P.-A. and Salanié, B. (2016). The econometrics of matching models. *Journal of Economic Literature*, 54(3):832–61.

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

Choo, E. and Siow, A. (2006a). Estimating a marriage matching model with spillover effects. *Demography*, 43(3):463–490.

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

Clasen, T., Boisson, S., Routray, P., Torondel, B., Bell, M., Cumming, O., Ensink, J., Freeman, M., Jenkins, M., Odagiri, M., et al. (2014). Effectiveness of a rural sanitation programme on diarrhoea, soil-transmitted helminth infection, and child malnutrition in Odisha, India: A cluster-randomised trial. *The Lancet Global Health*, 2(11):e645–e653.

Crocker, J. (2016). Teachers and sanitation promotion: An assessment of community-led total sanitation in Ethiopia. *Environmental Science and Technology*, 50(12).

Crocker, J., Abodoo, E., Asamani, D., Domapielle, W., Gyapong, B., and Bartram, J. (2016). Impact evaluation of training Natural Leaders during a Community-Led Total sanitation intervention: A cluster-randomized field trial in Ghana. *Environ. Sci. Technol.*, 50(16):8867–8875.

Das Gupta, M., Zhenghua, J., Bohua, L., Zhenming, X., Chung, W., and Hwa-Ok, B. (2003). Why is son preference so persistent in East and South Asia? A cross-country study of China, India and the Republic of Korea. *The Journal of Development Studies*, 40:153–187.

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

Desai, S. and Andrist, L. (2010). Erratum: Fatter attraction: Anthropometric and socioeconomic matching on the marriage market. *Demography*, 47:667–687.

Desai, S., Vanneman, R., and National Council of Applied Economic Research (New Delhi) (2010). India Human Development Survey (IHDS), 2005. ICPSR22626-v8. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor].

Desai, S., Vanneman, R., and National Council of Applied Economic Research (New Delhi) (2018). India Human Development Survey-II (IHDS-II), 2011-12. Inter-university Consortium for Political and Social Research [distributor].

Fernández, R. and Wong, J. C. (2017). Free to leave? A welfare analysis of divorce regimes. *American Economic Journal: Macroeconomics*, 9(3):72–115.

Galichon, A. and Salanie, B. (2021). Cupid' s invisible hand: Social surplus and identification in matching models.

Garn, J. V., Sclar, G. D., Freeman, M. C., Penakalapati, G., Alexander, K. T., Brooks, P., Rehfuess, E. A., Boisson, S., Medlicott, K. O., and Clasen, T. F. (2017). The impact of sanitation interventions on latrine coverage and latrine use: A systematic review and meta-analysis. *International journal of hygiene and environmental health*, 220(2):329–340.

Gautam, S. (2017). *Household Demand in the Presence of Externalities: Model and Applications*. PhD thesis, University College London, United Kingdom.

Gautam, S. (2020). Quantifying welfare effects in the presence of externalities: An ex-ante evaluation of sanitation interventions. Working Paper, Washington University in St. Louis.

Greenwood, J., Guner, N., Kocharkov, G., and Santos, C. (2016). Technology and the changing family: A unified model of marriage, divorce, educational attainment, and married female labor-force participation. *American Economic Journal: Macroeconomics*, 8(1):1–41.

Grossbard, S. (1993). On the economics of marriage-a theory of marriage, labor and divorce. *MPRA Paper*, 1832.

Guiteras, R., Levinsohn, J., and Mobarak, A. M. (2015). Sanitation subsidies. encouraging sanitation investment in the developing world: A cluster-randomized trial. *Science*, 348(6237):903–906.

Gupta, B. (2014). Where have all the brides gone? Son preference and marriage in India over the twentieth century. *The Economic History Review*, 67(1):1–24.

Hammer, J. S. and Spears, D. (2016). Village sanitation and children's human capital: evidence from a randomized experiment by the Maharashtra government. World Bank Policy Research Working Paper No. 6580.

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

Hitsch, G. J., Hortaçsu, A., and Ariely, D. (2010). Matching and sorting in online dating. *American Economic Review*, 100(1):130–63.

Jaggi, T. (2001). The economics of dowry: Causes and effects of an Indian tradition. *University Avenue Undergraduate Journal of Economics*, 5(1):2.

Kone, Z. L., Liu, M. Y., Mattoo, A., Ozden, C., and Sharma, S. (2018). Internal borders and migration in India. *Journal of Economic Geography*, 18(4):729–759.

Lafortune, J. (2013). Making yourself attractive: Pre-marital investments and the returns to education in the marriage market. *American Economic Journal: Applied Economics*, 5(2):151–78.

Lundberg, S. and Pollak, R. A. (1996). Bargaining and distribution in marriage. *Journal of economic perspectives*, 10(4):139–158.

Miller, G. and Mobarak, A. M. (2013). Gender differences in preferences, intra-household externalities, and low demand for improved cookstoves. National Bureau of Economic Research Working Paper (18964).

Moorjani, P., Thangaraj, K., Patterson, N., Lipson, M., Loh, P.-R., Govindaraj, P., Berger, B., Reich, D., and Singh, L. (2013). Genetic evidence for recent population mixture in India. *The American Journal of Human Genetics*, 93(3):422–438.

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

Patil, S. R., Arnold, B. F., Salvatore, A. L., Briceno, B., Ganguly, S., Colford Jr, J. M., and Gertler, P. J. (2014). The effect of India's Total Sanitation Campaign on defecation behaviors and child health in rural Madhya Pradesh: A cluster randomized controlled trial. *PLoS Med*, 11(8):e1001709.

Pattanayak, S. K., Yang, J.-C., Dickinson, K. L., Poulos, C., Patil, S. R., Mallick, R. K., Blitstein, J. L., and Praharaj, P. (2009). Shame or subsidy revisited: social mobilization for sanitation in orissa, india. *Bulletin of the World Health Organization*, 87:580–587.

Pencavel, J. (1998). Assortative mating by schooling and the work behavior of wives and husbands. *The American Economic Review*, 88(2):326–329.

Pickering, A. J., Djebbari, H., Lopez, C., Coulibaly, M., and Alzua, M. L. (2015). Effect of a Community-Led Sanitation intervention on child diarrhoea and child growth in rural Mali: A cluster-randomised controlled trial. *The Lancet Global Health*, 3(11):e701–e711.

Radtke, I. (2018). WASH guidelines for field practitioners. Part 2: Sanitation. Malteser Inernational.

Rao, V. (1993). The rising price of husbands: A hedonic analysis of dowry increases in rural India. *Journal of political Economy*, 101(4):666–677.

Rasul, I. (2006). Marriage markets and divorce laws. *Journal of Law, Economics, and organization*, 22(1):30–69.

Reynoso, A. (2018). The impact of divorce laws on the equilibrium in the marriage market. *University of Michigan, Job Market Paper*.

Spears, D. and Lamba, S. (2016). Effects of early-life exposure to sanitation on childhood cognitive skills: Evidence from India's Total Sanitation Campaign. *Journal of Human Resources*, 51(2):298–327.

Stopnitzky, Y. (2012). The bargaining power of missing women: Evidence from a sanitation campaign in india. MPRA Paper 37841, University Library of Munich, Germany.

Stopnitzky, Y. (2017). No toilet no bride? Intrahousehold bargaining in male-skewed marriage markets in India. *Journal of Development Economics*, 127:269–282.

Water and Sanitation Program (WSP) (2011). A decade of the Total Sanitation Campaign: Rapid assessment of processes and outcomes. Volume 1: Main Report.

Zha, D. (2019). Schooling expansion and the female marriage age: Evidence from Indonesia. Columbia University, Job Market Paper.

A Descriptives and Additional Exercises

Figure 10: Marriage patterns of women over time

Note. Own calculations using data from the IHDS 2005 and 2011.

Table 10: Descriptive statistics IHDS sample

	2004	2011
Demographics:		
Age at marriage (women):		
p25	15.0	16.0
p50	17.0	18.0
p75	19.0	20.0
mean	17.1	18.0
Age gap (male-female):		
p25	0.0	0.0
p50	1.0	1.0
p75	4.0	3.0
mean	2.1	1.8
Marriage market characteristics:		
Bride participated in matching decision (%)	0.374	0.351
Newly-weds living with groom's parents (%)	0.924	0.974
Wife from same village as groom (%)	0.119	0.105
Hours to natal home (wife)	3.287	3.065
Inter-caste marriage (%)	0.049	0.060
Marriage associated costs:		
Gifts occurrence (0: never 1: usually)	0.448	0.364
Share of household income spent by groom	1.7	1.9
Share of household income spent by bride	2.6	2.9

Notes: Own calculations using data from the IHDS waves 2004 and 2011. Sample includes men and women aged 15 to 34 years and married at the time of the survey.

Table 11: TSC impact on sanitation for households with marriageable male and females by household type

Variables	Only son marriageable households			Only daughter marriageable households		
	All	High sex ratio	Low sex ratio	All	High sex ratio	Low sex ratio
dT	-0.0450***	-0.0378	-0.0483***	-0.0192	-0.00332	-0.0225
	(0.0124)	(0.0340)	(0.0134)	(0.0172)	(0.0450)	(0.0179)
TSC	0.0435	0.132*	0.00612	0.110***	0.145**	0.0847*
	(0.0414)	(0.0675)	(0.0478)	(0.0386)	(0.0695)	(0.0460)
$dT \times TSC$	0.0779**	-0.0653	0.148***	0.0364	-0.0359	0.0738*
	(0.0333)	(0.0492)	(0.0374)	(0.0352)	(0.0605)	(0.0446)
Age oldest marriageable individual	0.00654***	0.00358	0.00678***	0.00846***	0.00943***	0.00769***
	(0.00109)	(0.00251)	(0.00121)	(0.00228)	(0.00295)	(0.00288)
Marital status oldest marriageable individual	-0.0439***	-0.00254	-0.0472***	-0.0825***	-0.0740**	-0.0849**
	(0.0132)	(0.0227)	(0.0150)	(0.0295)	(0.0336)	(0.0389)
Constant	0.267	0.339***	0.227	0.132	0.176	0.133
	(0.179)	(0.0865)	(0.178)	(0.163)	(0.166)	(0.169)
Observations	21,245	4,437	16,808	6,594	1,747	4,847
R^2	0.424	0.491	0.414	0.401	0.462	0.393

Notes: Own calculations using data from the IHDS waves 2004 and 2011. The sample consists of all single and married females (males) aged 15 to 34 at the survey time. The high sex ratio corresponds to districts with at least 999 women per 1000 men in the age range 15-34, while the low sex ratio is districts with less than 999 sex-ratio. District level sex ratio information was computed using information from the population census 2001 and 2011. High TSC districts correspond to a grand score for implementation of 61 or above, and high TSC to those with a score of 61 or above. The grand score per district is taken from WSP (2011). Oldest marriageable female (male) is an indicator variable that takes the value of 1 if the respondent is the oldest marriageable (age 15-34) daughter (son) in the household.Apart from the coefficients presented in the table, as controls, we consider an index of the female's (male's) wealth; age of the female (male); household size; education level of the household head (no education [base], primary, incomplete secondary, secondary, and above secondary); caste (Brahmin [base], High caste, Other backward caste, Dalit, Adivasi, Muslim, Sikh - Jain); and indicator variable for the rural zone; and fixed effects at the state level. Clustered at district level standard errors in parentheses. Significance: *** p<0.01, ** p<0.05, * p<0.1

Table 12: Sensitivity to the cut-offs definitions

		High Sex Ratio cut-off *				
		990	995	999	1005	1015
Oldest marriageable son	High Sex Ratio	-0.065^{*}	-0.093^{**}	-0.083^{*}	-0.096^{**}	-0.106^{**}
	Low Sex ratio	0.144^{***}	0.144^{***}	0.136^{***}	0.138^{***}	0.131^{***}
Oldest marriageable daughter	High Sex Ratio	-0.007	-0.023	-0.038	-0.057	-0.063
	Low Sex ratio	0.074^{*}	0.074^{*}	0.077^{**}	0.086^{**}	0.085^{**}
		TSC expousure cut-off**				
		50	60	61	65	67
Oldest marriageable son	High Sex Ratio	-0.045	-0.083^{*}	-0.083^{*}	-0.090^{*}	-0.098^{**}
	Low Sex ratio	0.088^{***}	0.132^{***}	0.136^{***}	0.147^{***}	0.151^{***}
Oldest marriageable daughter	High Sex Ratio	-0.011	-0.038	-0.038	-0.054	-0.077
	Low Sex ratio	0.061^{*}	0.077^{**}	0.077^{**}	0.082^{**}	0.078^{*}

* TSC=61 ** Sex ratio = 999

Table 13: Probability of Marriage (LSR)

	Men				Women			
	High TSC	Low TSC	Change	p-val	High TSC	Low TSC	Change	p-val
Wealth Type L	0.451	0.438	0.014	0.504	0.629	0.633	-0.004	0.881
Wealth Type M	0.389	0.386	0.003	0.865	0.502	0.524	-0.023	0.256
Wealth Type H	0.367	0.376	-0.008	0.569	0.394	0.412	-0.018	0.302

Notes: Probability of marriage with caste, region and time group effects. Wealth Types L,M,H refer to low, medium and high wealth respectively. Household level controls age, assets, Education of Household Head. Standard errors clustered.

Table 14: Probability of Marriage (HSR)

	Men				Women			
	High TSC	Low TSC	Change	p-val	High TSC	Low TSC	Change	p-val
Wealth Type L	0.493	0.423	0.070	0.169	0.826	0.558	0.268	**0.000**
Wealth Type M	0.359	0.325	0.034	0.141	0.407	0.410	-0.003	0.924
Wealth Type H	0.308	0.297	0.011	0.600	0.350	0.345	0.004	0.881

Notes: Probability of marriage with caste, region and time group effects. Wealth Types L,M,H refer to low, medium and high wealth respectively. Household level controls include age, assets, Education of Household Head. Standard errors clustered.

Table 15: TSC impact on alternative marriage market outcomes by wealth and sex ratio

	Low Sex Ratio		High Sex Ratio	
	(1)	(2)	(3)	(4)
Wealth Type:	Low wealth	High Wealth	Low Wealth	High Wealth
Panel A: Age of the husband				
$dT \times TSC$	0.358	0.210	1.183	-0.0207
	(0.282)	(0.312)	(1.104)	(0.405)
Observations	11,898	8,797	2,701	1,735
R^2	0.048	0.083	0.086	0.181
Panel B: Age of the wife				
$dT \times TSC$	0.190	0.210	0.896	0.376
	(0.369)	(0.281)	(1.053)	(0.490)
Observations	11,898	8,797	2,701	1,735
R^2	0.044	0.060	0.047	0.096
Panel C: Age gap				
$dT \times TSC$	0.168	0.000460	0.287	-0.396
	(0.189)	(0.144)	(0.330)	(0.367)
Observations	11,898	8,797	2,701	1,735
R^2	0.188	0.161	0.204	0.191

Notes: Own calculations using data from the IHDS waves 2004 ($dT = 0$) and 2011 ($dT = 1$). The sample consists of households of all single and married females (males) aged 15 to 34 at the survey time. High Sex Ratio corresponds to districts with at least 999 women per 1000 men in the age range 15-34 (Columns 3 and 4), while Low Sex Ratio is districts with less than 999 (Columns 1 and 2). District level sex ratio information was computed using data from the population census 2001 and 2011. High wealth correspond to individuals whose asset index is above the 50 percentile of the entire country distribution per wave (Columns 2 and 4). Those households below such cutoff are classified as low wealth (columns 1 and 3). $TSC = 1$ correspond to a grand score for implementation of 61 or above, and $TSC = 0$ to those with a score of 61 or above. The grand score per district is taken from WSP (2011). Apart from the coefficients presented in the table, as controls, we consider the age and marital status of the individual for whom the household is in the sample; the wealth index of the household; household size; education level of the household head (no education [base], primary, incomplete secondary, secondary, and above secondary); caste (Brahmin [base], High caste, Other backward caste, Dalit, Adivasi, Muslim, Sikh - Jain); an indicator variable for the rural zone; and fixed effects at the state level. Clustered at the district level standard errors in parentheses. Significance: *** p<0.01, ** p<0.05, * p<0.1

Table 16: TSC impact on women decision by wealth and sex ratio

	Low Sex Ratio			High Sex Ratio		
	(1)	(2)	(3)	(4)	(5)	(6)
Wealth Type:	L Wealth	M Wealth	H Wealth	L Wealth	M Wealth	H Wealth
dT	-0.0165	-0.0527	-0.0654	0.0784	0.177	-0.186
	(0.0680)	(0.0578)	(0.0711)	(0.132)	(0.134)	(0.138)
TSC	-0.0831	-0.108	-0.0860	-0.217	0.261	-0.310
	(0.129)	(0.101)	(0.132)	(0.201)	(0.159)	(0.194)
$dT \times TSC$	0.119	-0.0581	0.174	0.414	-0.152	0.248
	(0.133)	(0.127)	(0.148)	(0.251)	(0.200)	(0.218)
Constant	-0.403*	-0.307	-0.0449	-0.886***	-0.329	0.0422
	(0.212)	(0.305)	(0.173)	(0.259)	(0.286)	(0.354)
Observations	6,034	5,096	4,097	1,577	957	825
R^2	0.018	0.025	0.036	0.114	0.128	0.085

Notes: Own calculations using data from the IHDS waves 2004 ($dT = 0$) and 2011 ($dT = 1$). The sample consists of households of all single and married females (males) aged 15 to 34 at the survey time. High Sex Ratio corresponds to districts with at least 999 women per 1000 men in the age range 15-34 (Columns 4-6), while Low Sex Ratio is districts with less than 999 (Columns 1-3). District level sex ratio information was computed using data from the population census 2001 and 2011. Wealth levels correspond to tertiles of the asset index of the entire country distribution per wave. Apart from the coefficients presented in the table, as controls, we consider the age and marital status of the individual for whom the household is in the sample; the wealth index of the household; household size; education level of the household head (no education [base], primary, incomplete secondary, secondary, and above secondary); caste (Brahmin [base], High caste, Other backward caste, Dalit, Adivasi, Muslim, Sikh - Jain); an indicator variable for the rural zone; and fixed effects at the state level. Clustered at the district level standard errors in parentheses. Significance: *** $p<0.01$, ** $p<0.05$, * $p<0.1$

B Cohort-based analysis and parallel trends

The identification strategy proposed in this article considers the variation on the exposure to the TSC program of individuals who take part of the marriage market. Yet, there is a valid question on whether there was an ongoing trend of divergence of marriage market gains prior to the program, between high and low exposure to TSC districts. To study this possibility, we use a cohort based analysis using the 2011 IHDS data. While we do not have earlier versions of the IHDS that allows to compute outcomes for the 15-35 cohorts in previous years, some outcomes from older cohorts provide an opportunity to assess differences in the past for those cohorts when they faced the marriage market.

First, couples are organised in cohorts according to the age of both individuals: 20 to 30, 31 to 40, and 41 to 60 years of age. We do not consider couples outside those combinations in order to preserve the simplicity of the analysis. Second, we compute the expected total gains π_{IJ} (equation 9) for all nine combinations of individual types. We do not consider the living arrangement for this exercise as the concern is about the marriage market characteristics, but also as toilet ownership at the time of marriage is not observed.[36] Third, we compare the difference on the expected total gains between the youngest cohorts (20-30 are the treated , 31 to 40 are a not treated group) between high and low exposure of the district to the TSC program, against the difference of the oldest cohort (41 to 60, another non treated group). This is implement using the following linear fixed panel model:

$$\pi^{I,J}_{j,c} = \beta^{I,J}_0 + \beta^{I,J}_1 \cdot TSC_j \cdot \mathbf{1}(c = 20 - 30) + \beta_2 * TSC_j \cdot \mathbf{1}(c = 31 - 40) + \eta_c + \gamma_j, \quad (16)$$
$$\text{for} \quad I \in \{Low, Medium, High\}, J \in \{Low, Medium, High\}$$

where $\pi^{I,J}_{j,c}$ are the expected total gains in district j in cohort $c \in \{20 - 30, 31 - 40, 41 - 60\}$, conditional on considering only data from couples of types $I \times J$. Variable $TSC_j = 1$ corresponds to a district with high exposure to the program (as in section 3.2).

Column 1 of table 17 shows that there is evidence a reduction in the expected total gains for the low-high and medium-medium wealth groups. Those results are qualitative similar to the results in Table 6. However, we should be careful in the comparison as the two exercises are using alternative comparison groups.

Column 2 of table 17 shows a difference for the medium-medium group, significant at the 90% level; apart from this, as expected, there are no differences on the expected gains between cohorts 31-40. This suggest that there is no particular diverging trends on marriage market characteristics along the dimension of the TSC exposure.

[36]Toilet ownership in 2011 for older couples might respond to shocks that occurred after marriage

Table 17: Impact of TSC on expected total gains using cohorts variation

		(1)	(2)
		$TSC_j \times 1(c = 20 - 30)$	$TSC_j \times 1(c = 31 - 40)$
Men I	Women J	β_1	β_2
Low	Low	0.301	0.424
		(1.059)	(0.901)
Low	Medium	-0.79	-0.628
		(1.062)	(1.319)
Low	High	-0.933**	0.05
		(0.426)	(0.619)
Medium	Low	-0.519	-0.253
		(0.338)	(0.358)
Medium	Medium	-0.901*	-1.108*
		(0.477)	(0.639)
Medium	High	-0.363	-0.979
		(0.754)	(0.747)
High	Low	-0.241	0.159
		(0.617)	(0.552)
High	Medium	-0.632	-0.213
		(0.742)	(0.637)
High	High	0.283	0.563
		(0.537)	(0.526)

Clustered at district level errors in parentheses. Significance: * 10%, ** 5%, *** 1%.