University of Connecticut

Department of Economics Working Paper Series

Gender Composition of Children and Sanitation Behavior In India

by

Deepak Saraswat
University of Connecticut

Working Paper 2018-12
July 2018

365 Fairfield Way, Unit 1063
Storrs, CT 06269-1063
Phone: (860) 486-3022
Fax: (860) 486-4463
http://www.econ.uconn.edu/

Gender Composition of Children and Sanitation Behavior In India*

Deepak Saraswat

July 29, 2018

Abstract

Open Defecation has been linked to various public health issues and has gained significant policy attention. Investing in adoption of better sanitation has also been advocated on the grounds of providing women with privacy and protection from potential harassment. Nonetheless, previous research has shown that due to son-biased preferences, households in India under-invest in outcomes for their female children. I use the gender of the first-born child as an indicator of the presence of adult female children in households and find that, in certain cases, households reduce open defecation if the first-born child is a girl. The findings in this paper provide a new first stage association between gender composition of children and sanitation behavior and also contribute to the economic literature on decision making in households belonging to developing countries.

Keywords: Sanitation, Open Defecation, Gender Composition of Children, India

JEL Codes: O10, O18, J18

*Contact: deepak.saraswat@uconn.edu. I thank Dean Spears, Michael Geruso, Nishith Prakash, Richard Murphy, Shrijata Chattopadhyay and Gaurav Chiplunkar for helpful comments and discussions. I am responsible for any errors that remain.

1 Introduction

India accounts for 60% of world's open defecation. Open defecation - that is, defecating in open places, behind bushes, near roads, near railway tracks- is considered to be a huge burden on public health and it has been a focus of policy attention for more than a decade and half in India. Open defecation has been associated with various measures of weak health indicators in India. A central theme of many research studies and experiments have been that defecting in open has significant negative externalities (mainly related to health) which show up in form of worse outcomes at a community level. In this paper, I look at other negative externalities of defecating in the open and how individual households are incentivised to switch away from it. Defecating in the open may have significant costs to women and girls, because as compared to men and boys, they need more privacy. These include costs to dignity (due to lack of privacy) and the potential threat of harassment when they defecate, urinate or attend to menstrual hygiene in the open. In this paper, I ask a specific question - Do households adopt better sanitation or stop defecating in the open due to the presence of adult female children?

Looking at the literature on sanitation practices, it has been linked to various public health issues and reduction in OD has been associated with gains in health and well-being. Bleakley (2007), Cutler & Miller (2005), and Watson (2006) have provided evidence about the role of sanitation in achieving better health and human capital in US. Recent studies in Indian context also provide evidence of association between sanitation practices and health and well-being. Duflo *et al.* (2015) suggest clear pathways by which exposure to fecal pathogens introduced by neighbors could lead to acute malnutrition and ultimately death. (Geruso & Spears (2018)) explore the effect of open defecation on childhood mortality which answers a long-standing puzzle of higher mortality among Hindu kids in India. Spears & Lamba (2016) find that exposure to open defecation negatively impacts child cognitive function. On the flip side, there is evidence that reduction in open defection is associated with gains in health. Coffey *et al.* (2017) links reduction in open defecation to reduction in anemia in Nepal.

While there is an extensive literature on negative health externalities of open defecation, there are other negative externalities such as gender based harrasment, lack of privacy and loss of dignity which have not received much attention. There is recent empirical evidence that influenced by social marketing campaigns, entry of new female member may incentivise households to adopt better sanitation practices. Stopnitzky (2017) looks at toilet adoption for new daughter-in-laws in North India and finds significant effects of a "No Toilet, No Bride" campaign[1] on reduction in the open defecation. There is, however, a considerable room for expansion in literature analyzing these incentives to adopt toilets. In this paper, I look at the same outcome (reduction in the open defecation) but for a different household structure where the incentives to adopt better sanitation may come from the presence of female children.

In this paper, I exploit the gender of the first born child as a natural experiment, predicting the presence of female children in a household. Using the National Family Health Survey (NFHS) conducted in 2015-16 in India, I use this natural variation to understand a potential association between presence of female children in a household and their sanitation behavior. Findings suggest that poor households living in urban areas, where the costs of open defecation are higher[2], reduce open defecation by 14-17% (3.1-3.7 percentage points) if the first born child in that household is a girl versus a boy. However, this association does not exist for richer urban households and households in rural areas. Putting the magnitude of results in perspective, Geruso & Spears (2018) find that a 10 percent point reduction in open defecation around the neighborhood is associated with a decline in infant mortality of 6 per 1,000, or about 8 percent of the population mean infant mortality rate.

Reviewing the link between practices and sexual harassment, (Jadhav *et al.* (2016)) provide evidence from Indian context that women who openly defecate are twice as likely to

[1] The campaign known as "No Toilet, No Bride" was initiated by government of state of Haryana in India in 2005. The social marketing campaign encouraged families of marriage-age girls to demand that potential suitor's family construct a toilet prior to the marriage. Stopnitzky (2017) finds a 21% increase in ownership of toilets.

[2] Due to space constraints.

suffer from non-partner sexual violence compared to those who do not. (JAGORI & UN-Women (2010)) provide evidence from Urban slums in New Delhi that 66% women living in these slums report a verbal abuse, 46% report a visual abuse and 10% report a sexual assault. In light of these evidences, having access to a toilet (not defecting in open) has private benefits for girls and women. Having access to toilet not only reduces the costs due to lack of privacy and harassment but also has other private benefits - Adukia (2017) shows that school latrine construction program in India increases school enrollment of pubescent-age girls and much more when there is access to sex-specific toilets.

In Indian context where there is a general preference for male child and private benefits for girls are ignored, the role of toilet as a better sanitary option becomes much more crucial. There is a big strand of literature showing that households in India selectively under-invest in outcomes for girls before and after birth due to presence of son-biased preferences. Sen (2003), Jeffery *et al.* (1989) show that India has a widespread preference for male child. Barcellos *et al.* (2014), Jayachandran & Kuziemko (2011) and Deaton (2003) show that households in India do not invest in nutrition, education, postnatal time and attention for a girl child (e.g. breastfeeding).

While almost all sanitation campaigns advocate adopting toilets due to privacy concerns for women and girls, in the light of evidence shown above, it is likely that incentives to adopt better sanitation for girl child are weak at a household level. Given the under-investment in female children, it is conceivable that households will only provide them private benefits (such as a toilet) when the costs/consequences of not doing so falls on the household as a whole. Women's lack of privacy while defecating may act as a negative consequence for the household as a whole if she faces verbal/visual abuse and/or sexual harassment. In Indian cultural context, sexual harassment and abuse is not only a cost to dignity or physical abuse for a woman, but are considered costs on a household as a whole. Most parts of India follow the structure of a *Patriarchal society*, where to defend a male offender, an incidence of harassment is often portrayed to be incited by a female. Due to this, there are significant costs to a household which may manifest in the form of shame, difficulty in the marital matching of the female child, higher dowry etc. Given these potential costs on the household as a whole, they have incentives to not openly defecate and adopt better sanitation for adult girl child in the household. I analyze the timing of reduction in open defecation and find that urban households with first born girl reduce open defecation when the first born female child is likely to leave the household (most likely getting married). I also provide evidence that the reduction in open defecation due to presence of first born girl is likely driven by households living in Indian states with higher rate of crime against women.

Given that households in India do not invest in private benefits for a girl child, it is an interesting finding that in presence of non-private costs, they are incentivised to provide the private good for the female child. The *first* contribution of this paper is to the literature analyzing decision making at a household level with differential outcomes for male and female children in presence of strong gender biases.

Second, this paper provides a unique new first stage result related to the gender of the first born child and its association with sanitation behavior of a household. To best of my knowledge, this has not been previously looked at in a rigorous empirical setting. This first stage result has a potential to contribute to further economic research like understanding peer-effects of sanitation behavior on neighbors or in social network of a household. Guiteras *et al.* (2015) show that rural households in Bangladesh adopt sanitation when they are in proximity to a household who received subsidy for building toilets, thereby highlighting the importance of social connections in sanitation behavior. The new first stage result from this paper may help invite future research in spillovers of sanitation behavior through kinship and social networks. Identifying the reasons for lack of research in better understanding the externalities of open defecation, Geruso & Spears (2018) mention lack of a strong first-stage in take-up of better sanitation in experimental studies as one of the key reasons. Difficulty in generating a large enough first-stage effect has been demonstrated by three recent field experiments in rural India (Hammer & Spears (2016); Clasen *et al.* (2014), Pl *et al.* (2015)).

The result from this paper provides evidence of previously not known natural incentives for taking up better sanitation, and when coupled with potential spillovers in sanitation practices, it can help future experimental studies in generating larger first-stage effects.

Finally, the study and findings are relevant for policy makers. It contributes to the understanding of who adopts toilet and who does not, which is likely to improve the targeting of resources towards improving sanitation.

Rest of the paper is organized as follows. In Section 2, I start by providing the background of the research problem in this paper. Section III organizes a conceptual framework looking at a household level decision making regarding sanitation practices and deriving testable implications from it. Section IV describes the data and identification. Main results are discussed in Section V along with a falsification test. Section VI conducts tests to verify the statistical strength of results. Section VII provides various robustness tests to check the association between hypothesis in Section III and results of Section V. Section VIII used two alternative specifications (difference in difference & two-stage least squares), and Section IX concludes.

2 Background

In one of his pre-election speeches in 2013, the current Prime Minister of India, Mr. Narendra Modi started a campaign to improve sanitation in India by saying “toilets before temples". For India, a developing and religious country, these were strong words. Whether or not this campaign reaches its successful conclusion is for the time to tell but a stress on sanitation in a pre-election campaign meant serious business. Sanitation in India is a serious concern not just for Indian policy makers but multilateral organizations like UNICEF, WHO and focus of donor agencies worldwide (like Bill and Melinda Gates Foundation).

A billion people worldwide defecate in the open and Indian alone accounts for 60% of them (UNICEF & WHO (2014)). These stark numbers along with well-known health consequences of open defecation makes it a high-priority policy concern. Apart from the effects on public health, defecating in open has negative externalities like the potential for harassment of girls and women who go out to defecate in the open. Women value toilets to a greater extent than men because they suffer disproportionately from male harassment when they defecate, urinate, or attend to menstrual hygiene in open (Stopnitzky (2017)).

Whether or not households practice open defecation also depends on critical factors like the region a household lives in (for example urban vs. rural) and how rich or poor they are. Households living in rural areas have access to large fields, open space, and more privacy while defecating in open which contributes significantly to high open defecation rates in these areas[3]. Defecating in the open in fields far from their home does not pose any immediate cost of pollution and impurity near houses and hence does not result in a higher social cost. These factors may induce low enough social costs that the monetary cost of adopting a toilet is higher, and we can expect even richer households in rural areas to be practicing open defecation. In rural areas, where the community is more integrated and privacy concerns are lower for women, harassments while defecating in the open may also be a lower probability event. On the contrary, in urban areas, households live in constrained spaces which provide less access to open space and less privacy while defecating in the open. Since urban areas have better and modern infrastructure, the social costs (shame) of polluting it and spreading impurity (as perceived by residents) is also higher. These costs are higher in high-income areas and we can expect richer households to be adopting a toilet, irrespective of gender composition of their children. Households in poorer pockets of urban areas, have higher costs of defecating in the open due to significant space constraints, lack of privacy and a higher probability of harassment. In spite of being poor, we may expect the households living in the slum with adult female children to be incentivised to invest in a toilet and not

[3]Qualitative work (Coffey & Spears (2017)) to understand open defecation practices has documented some interesting features of sanitation practices in rural areas, like - a) Households in rural areas may also prefer to defecate in open since they like open environment and not being constrained by walls of a toilet, b) They have been doing so for generations and a behavioral change is harder for them, c) Women in rural areas also prefer to defecate in open since it gives them a chance to go out of house and meet friends.

defecate in the open.

3 Conceptual Framework

3.1 Setup

In this sub-section, I look back at Section I and set up a framework in which a household decided whether or not to adopt a toilet (or stop defecating in the open). This decision depends on the presence of a female child, the income of a household and the region they live in (rural, urban, etc.).

In a one-period framework, a household i is maximizing their utility over a bundle of consumption good X and adopting a toilet t (or a decision to not Openly Defecate), as follows:

$$\max_{\{X_i,t_i\}} V_i = U_i(X_i, \varphi_i, t_i) - \omega C_i(F_i, t_i) \quad s.t. \quad X_i + Pt_i \leq I_i \tag{1}$$

Where, X_i is a bundle of all consumption goods, the price for which is normalized to 1. $t_i \in \{0, 1\}$ is a decision to adopt a toilet, the price for which is P. I_i is the aggregate set of resources a household has.

The first term $U_i(.)$ represents the utility from consuming X, disutility from φ_i which represents the social cost of open defecation (as seen in Section I), and the utility from having a toilet t. The factor $\varphi_i \in [\varphi^L, \varphi^H]$ is the social cost factor related to the region a household i lives in (such as the cost of shame, lack of privacy etc., when defecating in the open). It is positive and is bounded[4]. The second term relates to the cost of open defecation, specific to the presence of female children in household. It is explained below in more detail.

[4] φ_i can range from very small values in sparsely populated remote rural areas to slightly higher values in somewhat dense areas to the high value in poor space constrained urban region to very high value in a posh urban residential society. It can also be negative for some households living in poor regions who have been defecating in the open for generations and have strong preferences for it. For simplicity in mathematical proofs, I assume it to be continuous between the defined bounds and non-negative.

The factor ω represents the probability of harassment an adult female child might face while defecating in open[5]. $C_i(F_i, t_i)$ is the cost factor representing a number of female children a household has and if or not they have a toilet. The factor $F_i = \sum_j \alpha_{ij} f_{ij}$ represents the total number of female children above a certain age cutoff, where, f_{ij} is the jth female child of household i and $\alpha_{ij} = 1$ if $f_{ij} \geq \bar{f}$[6] and 0 otherwise. The interaction of ω & $C(.)$ determines the cost a household faces when their adult female child defecates in the open.

Some key assumptions related to the setup are as follows:

1. $C_F(F_i, t_i) > 0$ and $C(F_i, 1) = C(0, t_i) = 0$ i.e. the cost factor is increasing in number of female kids above a certain age cutoff[7] and is 0 if there are no female kids above a certain age cutoff or the household has a toilet.

2. The utility term $U(.)$ is increasing and concave in the level of consumption X and decreasing in social cost factor φ.

I assign the term $U_i(.)$ in equation (1) a specific form: $U_i(X_i, \varphi_i, t_i) = u(X_i) - \varphi_i(1 - t_i)$, and accordingly, the value function becomes, $V_i = u(X_i) - \varphi_i(1 - t_i) - \omega C_i(F_i, t_i)$.

Given the binary nature of decision to adopt a toilet and the assumptions mentioned above, a household chooses an optimal V_i^* amongst following two options:

$$V_i^* = \begin{cases} V_i(0) = & u(I_i) - \varphi_i - \omega C_i(F_i, 0) \quad if \quad t_i = 0, \\ V_i(1) = & u(I_i - P) \quad\quad\quad\quad\quad\quad if \quad t_i = 1 \end{cases} \tag{2}$$

Given the optimal choice of value based on toilet adoption decision, I have following proposition and subsequent cases (Appendix A provides relevant proofs):

[5]The probability of harassment links more closely to the crime rate in the region a household lives in. For mathematical simplicity, I assume it is constant. Even if we let it vary by region (low for rural areas, higher for urban areas), the direction of results would not change but the mathematical arguments become cumbersome

[6]$\bar{f}$ is a specific age cutoff, such as puberty, beyond which a female child needs privacy and other harassment related risks kick in.

[7]This will be tested in empirical analysis in Section VI.

Proposition: $\exists$ a level of social cost $\bar{\varphi}$, such that,

1. **Case 1**: $\forall\ \ \varphi_i < \bar{\varphi}$,

$$u(I_i) - \varphi_i - \omega C_i(F_i, 0) > u(I_i - P) \implies V_i^*(0) > V_i^*(1),$$

2. **Case 2**: and, $\forall\ \ \varphi_i \geq \bar{\varphi}$, $\exists$ an $\bar{I}$, such that, $\forall\ I_i \geq \bar{I}$,

$$u(I_i - P) \geq u(I_i) - \varphi_i - \omega C_i(F_i, 0) \implies V_i^*(1) \geq V_i^*(0)$$

3. **Case 3**: and, $\forall\ \ \varphi_i \geq \bar{\varphi},\ \ \&\ \ \forall\ I_i < \bar{I}$,

$$u(I_i) - \varphi_i - \omega C_i(F_i, 0) > u(I_i - P) \implies V_i^*(0) > V_i^*(1) \quad \text{if } F_i = 0, \text{ i.e. } C_i(.) = 0$$
$$u(I_i - P) \geq u(I_i) - \varphi_i - \omega C_i(F_i, 0) \implies V_i^*(1) \geq V_i^*(0) \quad \text{if } F_i > 0, \text{ i.e. } C_i(.) > 0$$

Where, $V_i^*(1) \geq V_i^*(0)$ means that, not openly defecating (adopting better sanitation) gives a household higher utility. Conversely, $V_i^*(0) > V_i^*(1)$ means that continuing to defecate openly gives a household higher utility.

3.2 Testable Predictions

The proposition and the cases put forth in the previous section provides us cases where a representative household may or may not adopt a toilet depending on aggregate resources, the region they live in and the presence of elder female children they have. These theoretical cases give us empirically testable hypothesis, that, given the *treatment* (presence of or some female children) status of a household, under what conditions they are likely to be incentivised to adopt better.

Taking into account the Indian context (as discussed in Section I), we can relate the social costs to sanitation choices of households. A large proportion of households in rural areas are likely to be living in areas where there are low regional costs to open defecation (low φ) and going by Case 1 in the previous sub-section, irrespective of income and gender composition of kids; they are likely to not adopt a toilet, i.e., *Never Takers*. Going by Case 2, richer households in urban areas face high enough social costs of open defecation (high φ) that, irrespective of the gender composition of children, they are likely to adopt a toilet, i.e., *Always Takers*. Households living in poor urban regions have a higher cost of defecating in the open but, at the same time, they are poorer. Their marginal utility gain from not spending in adopting a toilet is high enough, such that, a) In the absence of an adult female child, the utility gain outweighs the social cost, but b) In the presence of an adult female child, the total costs (social + potential cost of female harassment) outweighs the utility gain. Going by Case 3, these households will only adopt a toilet if they have adult female children in a household, i.e., *Compliers*. However, some of these poor households will be soo poor that the budget constraint remains tight for them even when they have female children; they will remain *never takers*. Table 1 summarizes the cases in an experimental framework.

4 Data and Identification

4.1 Data

The main dataset used in the analysis is National Family Health Survey (NFHS) of India, conducted in 2015-16. The NFHS (India's version of Demographic and Health Survey) is a large, nationally representative survey and is regarded as a very high-quality demographic survey. The respondents are women aged 15-49 and report birth histories and other information for their children. This survey also includes information on household assets, infrastructure and other health related reports. The main variables I use in the analysis are the birth records from NFHS. These include birth order, gender, date of birth, whether the child is alive, and whether or not s/he continues to live in the household, for each of the child ever born to the surveyed women. Apart from these, I use characteristics of the women surveyed, of the head of the household, the residence (rural or urban) and indicators of wealth in a household (data on categories of assets). As the main outcome of interest, I use the survey question where a household reports: *"What kind of toilet facility do members of your household usually use?"*. I create an indicator OD equals to 1 if a household reports: "*Having no toilet facility, going to field/bush to relieve themselves*" and 0 otherwise.

As discussed in previous sections, household wealth could be a key factor related to adoption of better sanitation facilities. Although, NFHS does not record income or consumption of surveyed households, it records the assets a household owns. Recent empirical studies use these asset ownerships as a proxy of household wealth (Geruso & Spears (2018)). I create an 'Asset Index' as a measure of the wealth of a household by summing over the dummy variables recording presence of various assets in household and creating a standard normal index of it[8].

Another key factor which could relate to the adoption of sanitation facilities is the age of oldest girl child. As discussed before, households might be incentivized to adopt a toilet when the eldest girl child is entering (or near) puberty. A general age range of attaining Puberty in Girls is about 10-14 years in India (Khadgawat *et al.* (2016)). Other similar studies also document early puberty starting as early as eight years of age. I take the minimum age cutoff for my sample to be eight years for a first born child. It includes the usual starting range of 10 years for the onset of puberty and also two years before that to account for the earlier onset of puberty and/or an earlier recognition of a need for privacy by parents. When married, the girl child leaves their parent's house to live with husband's family. This event not only changes the gender composition of children living in the household, but may also affect the wealth of household. One one hand, households may become poorer after paying dowry for the girl child's marriage, on the other hand, it is conceivable that after

[8]This Index is created at the level of Residence x Survey Round; total six levels. Standardizing the Index: $Index = [(Sum\ of\ Indicators)_i - (Mean\ within\ a\ level)]/[Standard\ Deviation\ within\ the\ level]$

the marriage of girl child, a household has more per capita income. In either case a change in wealth of a household may change their sanitation behavior. Therefore, I restrict the maximum age of first born child to be 16 years in the data used for analysis[9].

For main analysis, I use households with, 1) first born child alive, 2) first born child is 8-16 years of age, and 3) either the responding women or her husband are head of the household. This is referred as 'Main Analysis Sample' in rest of the paper.

4.2 Identification

The main independent variable in my research question is the presence of and/or a number of female children between age 8-16 years in the household. An ideal (but hypothetical) comparison would be between households with and without an elder female child in which the presence of that child is randomly assigned. To get close to this ideal comparison, various studies in Indian context use gender of first child as a plausible random assignment [(Barcellos *et al.* (2014)), (Kishore & Spears (2014))]. I use the same identification strategy in the main analysis of this paper. Gender of first born child is considered random in many economic studies [(Rosenblaum (2013)), (Bhalotra & Cochrane.C (2010))]. A problem central to these studies is that, due to son-biased preferences, households in India practice fertility stopping rules. As a result, a correlation develops over time between household level outcomes and the gender of the first child. To get around this problem, these studies restrict the samples to households with a very young first born child.

In this paper, the sample of interest are the households in which the first born child is at least eight years old, hence the solution to restrict sample does not apply here. Households in which the first-born child is a girl are more likely to have a higher fertility rate to achieve the desired number of sons, and hence larger family on an average. Larger families have lower per capita resources which make them poorer on average, and hence more likely to practice open defecation (or not invest in a toilet). The bias induced due to the gender of

[9] 16 years mark a point in NFHS data where first born girls are systematically and significantly more likely to be "not living with their parents" than boys. See Figure 5 (Panel A).

the first child if at all, will only induce a downward push on any association I find between gender of the first child and the sanitation practices. In the empirical analysis below, I first check if household characteristics and other relevant outcomes are statistically similar across households with young first born children. Next, I use the Main Analysis Sample (where at least eight years have passed since the birth of first child), to check if relevant outcomes change in a direction which could potentially induce upward bias on the results.

Finally, it remains to see if the gender of the first born child affects the presence of female children in my data. Figure 1 uses the Main Analysis Sample and plots the average number of female kids vs. the average total number of kids for all the households in the sample. The association shown in the plot is separated by the gender of the first born child. As observed, households with the first born child as a girl have, on average, a higher number of female kids compared to households with the first born child as a boy.

5 Results

5.1 How do households with first born girl child compare to first born boy child

Before the main analysis I run mean difference checks on various household characteristics and relevant outcomes. Outcome variable included in these checks are sanitation behavior of a household and other factors that may affect it. I use following reduced form regression for these checks.

$$Y_{ir} = \alpha_1 FB_{ir}^{Girl} + \epsilon_{ir} \tag{3}$$

Y_{ir} is the outcome for mean difference check for household i living in region r: urban or rural. FB_{ir}^{Girl} is the indicator for first born child being a girl in household i living in region r.

Standard errors are clustered at the level of primary sampling unit (PSU)[10]. The coefficient of interest is α_1. Table 3 & 4 report the coefficient of interest, its p-value and the mean of the outcome.

The first set of mean difference check looks at the difference in relevant outcomes of households with first born child of age 5 years or less[11]. Table 3 reports the results, separately by Rural and Urban regions. Reassuringly, the main outcome of interest, Open Defecation is not statistically different in both regions across the comparison groups. Households with the first born girl are not significantly richer or poorer than households with the first born boy. Households with better infrastructure such as piped water, cement/concrete floor, walls, etc. may find it easier to adopt a toilet. As observed, households with first child as girl do not differ significantly from those with first child a boy in any of these categories. Households with the first born girl are also not observed to be different in parent's education and level of social awareness (reading newspapers and watching TV). It appears that fewer mothers in rural areas have primary education but it cannot be ruled out that this difference is significant just by chance[12].

The second set of mean difference checks uses the main analysis sample where at least eight years have passed since the birth of first child in the household. Table 4 reports these results. Households show evidence of gender-biased fertility stopping rules, as the mothers of the first-born girl are more likely to be pregnant at the time of survey as compared to mothers of first-born boys. Looking at asset index, households with first born girl child are significantly poorer than households with the first born boy. Households with the first born girl go for a larger fertility and hence are poorer on average due to lower per capita resources. This, however, does not induce any upward bias on my results. As expected, first born girls live in larger households and with more siblings but again they do not put

[10] The NFHS is a two-stage random sample, first sampling Primary Sampling Unit (PSU) and then households within sampled PSUs. PSU is that sense is a sampling cluster. It is usually city blocks in urban areas and villages in rural areas.

[11] Similar to the main analysis sample, households included in this set of check are those where first bron child is alive and respondent is either head of household or head's wife.

[12] Nevertheless, I control for all these factors while estimating the regressions for main results.

any upward bias on my results. Households with the first born girl are also not observed to be having better construction (material for floor, walls, etc.) or more educated and/or aware parents (education, reading newspaper, watching TV, etc.). Household head in urban households with the first born girl are slightly older. I include parents age as one of the controls in main analysis.

Overall, these mean difference checks help support the identification strategy that there are no significant differences in the variables related to Open Defecation which may put an upward bias on main outcome of interest.

5.2 Do households reduce open defecation due to presence of female children?

In Table 5, I analyze the hypothesis that households might reduce open defecation (adopt a better sanitation facility) if they have a girl child; the presence of whom is identified by a the gender of first born child. I estimate the regression of following form on the entire NFHS dataset[13]:

$$OD_i = \beta_1 FB_i^{Girl} + X_i + \epsilon_i, \tag{4}$$

where, as before i indicated household, and FB_i^{Girl} is indicator for first born gender being female. Outcome is the indicator variable OD_i. X indicates key factors (as discussed in Section 3) which may predict and relate with sanitation behavior of a household. Table 5 reports the results, starting with no controls X in Column 1 and adding each key control in Columns 2-5. Standard errors are clustered at PSU.

Starting with Column 1, households with first born girl are shown be to be practicing less OD, but the coefficient is very small and statistically insignificant. It is important to account for economic status of a household (which is related to the gender of first born

[13]Similar to main analysis sample, the households included in this analysis are the ones where the first born child is alive and responding women or her husband are household head.

child). In Column 2, I include the asset index and the coefficient on OD becomes larger in magnitude and statistically significant. Also as expected, the coefficient on asset index shows that richer households practice less open defecation. The place of residence (rural or urban) is an important factor and including region in Column 3 shows that. Households in rural areas practice more open defecation, but that does not take away much from the coefficient on first born girl child. First born being a girl has consequences not only on the number of girl child in household but also on the total fertility (Figure 1), each of these consequences are expected to have a differential impact on sanitation facilities adopted by a household. Column 4 includes two controls, one for total number of children in the household and other for the total number of girls among all children (excluding the first born girl, if present). As expected, a higher number of female children, controlling for total number of children, relates to lower open defecation. This column suggests that while more female children in a household may relate to lesser open defecation, the first born being a female alone relates to a strong persistent reduction in OD as shown by the stable coefficient and significance. Lastly, as discussed in Section 3 before, the incentive to reduce OD may only arise when the first born girl child has attained puberty or near it; Column 5 includes an indicator variable equal to 1 if first born is 8 years of age or older and 0 if younger. As expected, households with older first born child are shown to be reducing OD. This however does not clearly indicate if it is just the older girl child or also the older boy child relating to reduced OD (Section 5.4 will shed more light on this).

Analysis in Table 5 provides support for gender of first born affecting the sanitation behavior and other key factors discussed in Section 2 & 3 which could be related to reduction in OD.

5.3 Who drives the reduction in OD in presence of female children?

As observed in previous section, the gender of first born child relates to the sanitation behavior of a household. It gives empirical support to the research question and testable predictions from Section 3 but it is not yet clear if this association is supported by all households belonging to different income groups and or is driven by a few specific sections of the society. It is conceivable that households belonging to different economic and social groups may not change sanitation practices due to presence of a girl child. As suggested by the conceptual framework and Table 1 & 2 the households living in poor urban regions are more likely to respond to the need of toilet for female children in household.

To investigate the validity of conceptual framework, I start by restricting the age range of first born child to 8-16 years in the analysis pool of 220,203 households and dividing the remaining 91,246 households by Rural and Urban regions, and within each region into groups of asset index (ranging from poorest to richest). Finally, I estimate the regression of form using OLS:

$$OD_{irw} = \gamma_1 FB^{Girl}_{irw} + \chi_{irw} + \epsilon_{irw}, \tag{5}$$

where i is a household, r is region (urban or rural), and w denotes the wealth group. In addition to the variables described before, χ denotes household level controls. These controls mainly comprise of household infrastructure, awareness and characteristics of parents, all of which are conceivably related to a household's choice of sanitation[14]. Regressions as per equation 5 are estimated with and without including controls for both rural and urban regions. Standard errors are clustered at PSU level.

[14] Full list of control are - Standardized value of asset index, number of household members, number of women and girls in the household, age of household head and the mother of first botn, education of household head and the mother, frequency of watching TV and reading newspaper in the household, main floor material, main roof material, main wall material, religion followed by the household and if the responding mother currently pregnant

Table 6 & 7 report the coefficient of interest γ_1 for urban and rural regions, respectively. Results in both table are reported in two panels, Panel A with no household controls and Panel B including the full range of controls. Finally, the regression are run on separate groups made using asset index. The main analysis sample for each region is divided into *ten* groups, one for each decile of asset index[15]. For example, the 1st group is the poorest 10 percentiles (according to the distribution of asset index) of households in main analysis sample for each region, and similarily, 10th group is the richest 10 percentiles.

Looking first at the results for Urban areas in Table 6, the row showing mean OD rate suggests that open defecation rate is falling rapidly as households get richer. Open defecation rate is high and ranges from almost 40% to 22% for the two poorest groups of asset index (Column 1 & 2). It starts to fall as households get richer and almost vanishes after the 5th group. Poorer households also have higher OD rates are most likely to be the ones to reduce OD if they have female children at home[16]. As expected in Panel A, households falling in 2nd decile of asset index show a reduction in OD if the first born child is girl. However, this reduction does not show up for the poorest decile, which is consistent with the conceptual framework that there might be some households who are poor enough for the budget constraint to be very tight for them. Panel B includes a large range of household controls and the results retain their patterns after including these controls. Overall, this amount to a 14-17% reduction in OD in urban households belonging to the 2nd poorest decile of asset index. Moving along the richer groups, this association becomes weaker and statistically insignificant.

Moving to results in rural areas in Table 7 it shows that open defecation rates are not only much higher on average than urban areas, but they stay higher for even the richer groups of households. For example, while the group of households belonging to median asset index in

[15]The choice of dividing the sample into deciles of asset index comes from the objective of identifying the section of households driving the reduction in OD (which needs division by wealth) but not ending up with small number of observations in each group and loosing statistical power.

[16]Poorer households with higher OD rates signifies they are living in worser conditions of the urban area, might be facing the higher crime rate and anti-social elements. Also given a higher OD rate, they have much larger room for improvement

urban areas had OD rate of lower than 10%, its counterpart in rural areas has almost a 50% OD rate. In contrast to the case in urban areas, this provides empirical evidence for lower costs of open defecation in rural households where space is not constrained. As discussed in the conceptual framework, rural households may have very low costs of defecating in the open and may not have incentives to switch from OD even if they have elder female kids in the household. Results in Panel A show that point estimates of the difference in open defecation between households with first born girl versus boy are very low for all decile groups of asset index. Given large confidence intervals, a conclusion of 'no statistically significant differences' cannot be ruled out. Panel B includes household controls and most of the patterns shown in Panel A are unchanged here. 7th decile however, shows a decline of 2.3% points which is significant at 90% confidence level. As will be analyzed in next section, this is not a strong enough statistical significance for claiming the causal association. Looking at the 5th, 6th and 7th groups together, they show small reductions in OD if first born is girl but the association is weak in magnitude and significance. One possible explanation, consistent with the conceptual framework could be that, since defecating openly is easier/less costly in rural areas (weaker space constraints), only the richer households would want to build a toilet and switch away from OD if they have a first born girl child. However, the smaller magnitude and weaker statistical significance does not allow claiming a causal association in this case.

5.4 Falsification

As discussed in Section 3 and demonstrated in the conceptual framework, a central notion is that households will be incentivised to invest in better sanitation when the expected costs of female children practicing OD is higher. If the reduction in open defecation also shows up when the female kids in the household are below the age of puberty, it may indicate that the results are potentially driven by causal factors other than what I have considered so far. To test this, I run a falsification test using regression in equation 5. While the data used for falsification analysis is same as the Main Analysis Sample for two criteria, a) where the first born child is alive, b) where the respondent or her husband is the head of household, it differs in the age cutoff. Only the households with the first born child less than eight years of age are included. Same as the analysis in Section 5.3 this falsification test is run on households separated into rural and urban regions, and within each region, into deciles of asset index.

Results are reported in Table 8. As observed in Panel A, unlike Table 6 the poorest wealth categories do not show reduction open defecation here if the first born child is girl and young. Same holds for rural region in Panel B. This test confirms that the reduced form association between first born female child and reduction in open defecation is valid only for elder female children and provides support for testable predictions from conceptual framework.

6 Inference Checks

All results reported in the previous section are generated by using comparisons amongst multiple groups of asset index. Thus, it cannot be ruled out that the reduction in OD, where it shows up is appearing 'by chance'. To mollify this concern, I put the analysis through stress test using three different inference checks.

6.1 Bonferroni Correction

Under multiple hypothesis testing, the chances of rejecting the null (making a Type-I error) increases by the factor of number of hypothesis being tested. The analysis then requires correction for multiple hypothesis. In Section 5.3, within each region (rural and urban), the regression is estimated for 10 groups of asset index and thus, p-value of 2nd decile in urban analysis being less than 0.05 alone is not enough for claiming its statistical significance. A Bonferroni multiple comparison correction for n independent comparisons requires a significance threshold of $\alpha = \frac{0.05}{n}$ for each comparison to recover a desired $\alpha = 0.05$.

Using this criterion, the p-value reported in Column 2 of Table 6 (Panel B) is just enough to recover a $\alpha = 0.05$ for within urban area comparisons and just enough to recover a $\alpha = 0.1$ when an additional division of data among rural and urban areas is taken into consideration. Looking at results for rural areas in Table 7, coefficient in Column 7 of Panel B does not provide enough statistical significance to recover a satisfactory α.

6.2 Power of sample

This test deals with a statistical question - Is the sample size in 2nd decile of urban analysis enough to detect the observed effect size? Focusing on reduced form results for 2nd decile in urban areas, I conduct a retrospective power calculation. This exercise takes households in main analysis sample where the first born child was boy as a control group, and the ones in which the first born child was girl as treatment group. Given the distribution of data and the outcome (Open Defecation), I generate the minimum detectable effect sizes and the sample size required to be powered to detect those effect sizes. The effect size is measured here as a percentage change in open defecation with respect to the mean open defecation rate of households with born child as boy. Standard deviations of OD in both comparison groups are taken into account. Calculations were made with a preset power level = 0.8 and an $\alpha = 0.05$.

Figure 2 reports the result of this exercise. The available sample size of 3609 households in Column 2 of Table 6 is enough to detect an effect size of 14-15% of control mean while the regression detects a reduction of 17%. It suggests that the available sample size is just a little more than the required to detect the observed effect size. Figure 3 reports the result of the same exercise for rural households in Column 7 of Table 7. It shows that the available sample size of 5013 households is not enough to detect the observed effect size.

6.3 Placebo Test

Since, the sample size in Column 2 of Table 6 is just about enough to detect the observed effect, a potential concern arises that the associated confidence intervals may be large enough to allow for the result to just appear 'by chance'. I put an additional stress test on the data to alleviate this concern. I do a placebo test where I reassign the treatment status (first being child being girl) randomly across all the sample households in urban areas belonging to 2nd decile of asset index and run the regression in equation (4) on it. I do this random assignment 3000 times, resulting in 3000 counterparts of γ_1 from equation (4). I then plot a distribution of these randomly generated coefficients and put the original effect size along with it. Figure 4 shows the distribution and the original coefficient (a vertical line in red) from Column 2 of Table 6 . As observed, the set of randomly generated coefficients are centered around a mean of 0 and the original coefficients lie far left on the tail. This test strengthens the statistical relevance of the result in Table 6.

7 Robustness

Previous sections have established the causal link between reduction in open defecation and gender of the first born child, and also the section of population which is most likely to support this link. This section looks specifically at the households demonstrating this causal link, i.e. households belonging to 2nd decile of asset index in urban areas and attempt to test its relevance with respect to hypothesis in Section XXX and robustness to potentially conflicting channels.

7.1 When is the toilet *needed* and what drives the need?

As suggested by Table 6 (Column 2) and falsification analysis in Table 8 (Column 2), the incentive to adopt a toilet (reduce open defecation) does not appear until the first born girl child is older and/or near puberty. It is only after the first born girls crosses 8 years or age that the reduction in OD starts to show up. It is however unclear if the incentive to reduce OD comes with the increased demand of menstrual hygiene and privacy (as first born girls reach menarche) or when protecting them from potential harassment becomes important (when they are about to be married). If households start adopting toilets when their first born child reaches menarche, it would suggest that households are responding to the private demands of their girl child, which however is contradictory to literature about girl's status in India. On the other hand, if households start adopting a toilet when the eldest girl is about to be married, it would support my hypothesis that households are reducing the costs (like delay in marriage, bad match or higher dowry) due to potential harassment.

Figure 5 using the NFHS data helps in understanding the timing of toilet adoption. Panel A at the top is a local polynomial smoothed value plot of the age of first born child (on x-axis) and if they are living with their parents or out of house (on y-axis). First born girl and boy are separated by green and blue lines along with their associated confidence intervals. Panel B plots the coefficients from the regression below:

$$OD_{ij} = \gamma_1 FB_{ij}^{Girl} + \chi_{ij} + \epsilon_{ij}, \tag{6}$$

Where i is the household and j indicates the age of first born child being less than or equal to j years (for $j = 1, 2, 3.....25$ years). χ as before represents the household level controls used in the main analysis. This regression is estimated only on the part of main analysis sample belonging to 2nd decile of asset index in urban areas. SE's are clustered at the level of PSU. Panel B plots the coefficient of interest γ_1 along with associated confidence intervals.

As observed in the Figure 5, gender of first born is not associated with reduction in OD until they are 13 years of age and only starts to show significant difference when they hit 16 years. It remains constant thereafter. In Panel A, age 16 is associated with girls leaving their parents house to live outside (most likely getting married). It suggests that households are responding more in line of the hypothesis of "reducing harassment costs" and not so much to the private demand of the girl child.

7.2 Are results driven by difference in household characteristics?

As observed in previous sub-section, the age at which first born girls are leaving the household is associated closely with household's sanitation behavior. A question arises if household characteristics change around the time of first born girl's marriage in the direction which can explain the reduction in OD. Marriage being a big event in a household in India, it is conceivable that people would renovate their houses, purchase new assets etc and as a part of those, they may construct toilets as well. However, a large portion of marriage expenses are financed by girl's family and in addition to that, they pay dowry. Given such high expenses, it is less likely that family will will engage in own house renovations or asset accumulations. Nevertheless, I estimate the regression as in equation (3) on the sample of 3609 households belonging to urban areas and in 2nd decile of asset index. This regression is run on households level outcomes as mean difference checks and on the assets owned by household.

Table 9 shows the mean difference checks for household level outcomes (same as Table 3 and Table 4) while Table 10 shows the mean difference checks for various household assets reported in NFHS survey. As seen in Table 9, none of the household characteristics change in the direction which could suggest household level improvements[17]. Table 10 shows the mean differences in various assets. Just a few households with first born girl report having no electricity and bicycle (the differences which are significant), and that does not suggest any potential biases. Households with first born girl have slightly higher number of mattresses, which may be due to larger family size or for giving their female children separate beds to sleep.

[17]Households with first born girl report having more roofs made of 'Asbestos Sheet'. Asbestos sheet roofs are usually an inferior quality roofing material (as opposed to Cement, stone or Concrete) and are less likely to be an improvement. in addition, this is the only category showing difference and its appearance 'by chance' cannot be ruled out.

7.3 Are results driven by female members other than female children?

If there are additional female members in the household who are not the daughters of the household head, it cannot be ruled out that the results shown in Table 6 are driven by the additional female member and not by the female child. Table 4 shows that in households with first born girl, there are significantly higher number of women who are 15-49 years of age. Possible explanation of this could be the presence of first born girl herself, additional female kids the household had after first child (in order to achieve desired number of sons) or systematic presence/entry of additional female members in these households. The latter one, if present is of a deep concern for validity of the results. There are some potential channels of presence/entry of additional female member. After the marriage of first born girl child, per capita resources in a household increases and household heads of nuclear families (which are more prevalent in Urban areas) may decide to bring in their parents (or parents in-law) to live them them. If parents are old and/or unwell thereby restricting their mobility (either gender), households might need to start using a toilet for them. Another channel could be entry of new daughter in law, in regard to which Stopnitzky (2017) provides causal empirical evidence of reduction in OD. Households with first born girl have higher fertility and smaller age gap between children. It is conceivable that after marrying first born girl child, household would seek to balance out the dowry payment by marrying the son and receiving dowry, hence bringing in a daughter in law.

A straightforward way to examine this possibility is to analyze the gender composition (female members to be specific) of households with first born child as girl versus boy and look for additional female member. NFHS records details of all the household members, in which they record the relation of the member with the head of household. I use the relationship information for households in urban areas in 2nd decile of asset index (restricting the maximum age to 16 years). Table 11 reports the result of estimating equation (3) with four different outcomes (Column 1-4). Households with first born girl have just about 1 additional female member (Column 1), and if looked at only the number of female children as outcome (Column 2), the coefficient seems to explain the additional female member in Column 1. Column 3 estimates the regression with all female members other than daughter of household head as outcome and there are no differences (both statistically and by magnitude). Looking specifically at the presence of daughter in law (Column 3), households with first born girl has lower number of daughters in law (potentially due to delay in marriage of son or son being young). These results strengthen the association of results in Section 5.3 to the testable predictions in Section 3.

7.4 Crime Against Women and Sanitation Behavior

Hypothesis in Section 3 and results in Section 5.3 establish that households respond to private demand of sanitation when the costs of female children defecating in open are likely to be higher. It is conceivable that households living in areas where there are higher crime against women would have higher incentives to reduce OD as their girl child reaches marriageable age. I look at this possibility using the state level crime data reports by National Crime Record Bureau (NCRB)[18] for the year 2015 in India and the urban households in 2nd decile of asset index in NFHS (maximum age of first born child being restricted to 16 years).

Table 12 ranks 36 states/UTs in India in increasing order of Urban OD rate (Column 2) and of rate of Crime against Women (Column 3). It is likely that poorer states will have both high OD rate and higher crime rates in general, hence a great deal of overap between these two respective rankings. However, there are considerable mismatches in both ranking. It could be due to difference in composition of urban population by state or state specific factors (which is beyond the scope of this paper).

I estimate the regression same as equation (4) on the urban households (as explained above), separated by crime rankings. Table 13 reports the results separated in two categories of states; one with crime rates in lower half of rankings (Column 1,2 & 3) and second with

[18]Retrieved from: http://ncrb.gov.in/StatPublications/CII/CII2015/FILES/Table%205.1.pdf

crime rates in upper half of rankings (Column 4,5 & 6). Results show that households belonging to states with higher crime against women reduce open defecation if their first born child is girl while households belonging to states with lower crime against women do not. Results are consistent with inclusion of state fixed effects (Column 2 & 5), and household level controls (Column 3 & 6). While this result seem to support the hypothesis discussed at the start, as observed the open defecation rate of urban households in states with lower crime is also low. A possibility cannot be ruled out that households in lower crime ranking states are already better off in terms of sanitation practices and hence the room for improvement in that area is lower as compared to the states with higher crime.

Table 14 attempts at investigating this possibility in more detail. The sample of households is same as the one used in Table 13, except that only the 18 states with highest urban OD rate are included (ranking 1-18 on urban OD rate). As before, the states are then divided into 2 equal categories, one with 9 states having highest crime against women and other with 9 states having lower crime. As opposed to the categories in Table 13, the category with lower crime against women has higher urban OD rate and possibly more room to improve on sanitation practices as compared to category with higher crime against women. Results show that households in category with higher crime against women reduce OD if they have first born girl child. While the category with lower crime shows negative association between first born girl and OD, the coefficient in smaller in size and statistically insignificant. Since the category with lower crime still comprises of states with higher OD rate, they might have higher room for improvement and shows some evidence in that direction, but the association is possibly weakened by lower crime against women (lower costs due to lower possibility of harassment).

8 Alternative Specifications

Given the gender preferences in India, it cannot be ruled out that after the birth of first child, unobservable characteristics of a household may change which may drive the reduction in open defecation and cast doubt on the hypothesis of "shame costs"" due to potential harassment driving the results. In addition, while the observables seem balanced across households with first born girl versus boy, it is possible that their fertility decision (which are observed in data) are based on their characteristics before they had their first child (the characteristics we do not observe). These pre-fertility characteristics may not necessarily be balanced and while affecting fertility decisions they may also be correlated with the sanitation behavior. I use two alternative specifications to test the validity of results.

8.1 Difference in Differences

Taking the gender of first born child as a quasi-experimental shock and their age, we can classify households in four categories as in Table 15 along with their observed and unobserved characteristics. Assuming there are some fixed unobservable factors linked to the gender of first born child; these factors may drive the results observed so far. Table 15 gives one such example where separate fixed factors associated with the gender of first born are coupled with a potential of harassment costs (or "shame" associated with it), which exists (or is higher) only when you have the first born as a girl and she is above the age cut off of 8 years.

Using this framework, I use a difference-in differences specification as follows:

$$Y_{iw} = \delta_1 FB_{iw}^{Girl}\ X\ AC_{iw} + FB_{iw}^{Girl} + AC_{iw} + \chi_{iw} + \epsilon_{iw} \tag{7}$$

Where, i is the household and w is the asset decile the household belongs to. AC is the dummy = 1 if the child is between 8-16 years of age and 0 otherwise. χ is a vector of household level controls. Coefficient of interest is δ_1

Table 16 reports the result of estimating this regression on the main analysis sample for urban region. A reduction in OD is observed in the 2nd decile of asset index (Column 2). Few subsequent groups show reductions which are small in magnitude and statistically insignificant. The coefficient is similar in size with that of Column 2 in Table 6 and overall these results strengthen the validity of analysis in Section 5.3.

8.2 Two stage least squares

As argued in the conceptual framework in Section 3, a household might be incentivised to adopt a toilet more if they have higher number of female kids. A simple OLS of the sanitation behavior on number of female kids would be biased as child bearing decisions are endogenous to various observed and unobserved household characteristics. For example, after the first-born child is female, a household can be more successful in their decision to have more kids (and eventually achieve desired number of sons) if they are wealthy and can afford a large number of kids. If the households observed in survey have more female kids because they could afford to, they were probably wealthier to start with (the before fertility period I dont observe in data) and are likely to have a toilet (or not openly defecate) as well. In a simple OLS of open defecation rate on a number of female children, household's wealth would act on both the independent and dependent variable and hence, an OLS is biased.

Figure 6 plots the asset index against a total number of kids in all urban households[19]. First, households with first born girl are on average poorer than households with the first-born boy (represented by Green dots). Second, as shown by connected lines, households with first born girl are richer in general than households with first born boy after second birth order (although, households in both these categories are much poorer compared to households with one or two children).

As argued in the Section 4.2 before, gender of first born child is random, and a higher number of female children in a household is a consequence of (randomly) having the first child as female. The gender of first child seems to be a potential candidate for being an

[19] This figure represents all Urban households irrespective of their position in asset index.

instrument for a number of female children a household has, and attenuate the upward bias from an OLS regression.

The requirement for an exclusion restriction dictates that instrument should not be correlated with the outcome of interest through any other channel but only through the instrumented independent variable. However, the first born child being itself could contribute directly to reduction in open defecation. This weakens the exclusion restriction.

Before moving forward with the analysis, I would like to re-clarify the purpose of this analysis. This set of analysis has branched out of a causal reduced form result shown in Section 5.3 and is not intended to explore causality in itself. In this analysis, I aim to see if the direction of results change when I take a step towards attenuating the potential bias discussed before. Due to a weaker exclusion restriction, and more so because the purpose here is not to check causality, I refrain from making a causal claim from this analysis. These results should be interpreted carefully and in light of the argument above.

Going forward with analysis, I estimate the first stage as follows on the sample used in previous sub-section. This analysis is conducted only for the urban households in 2nd decile of asset index of the main analysis sample (Column 2 of Table 6).

$$NF_i = \beta_1 FB_i^{Girl} + \mu_i \tag{8}$$

Where, NF is the number of female children (between 8 and 16 years of age) and are instrumented by FB_i^{Girl}, the indicator equals to 1 if first born child of household i is female.

Using the predicted values of $\widehat{NF}_i$ from the first stage above, I run the second stage as follows:

$$OD_i = \gamma_1 \widehat{NF}_i + \chi_i + \epsilon_i \tag{9}$$

Coefficient of interest is γ_1. Table 17 reports the results from this two-stage least squares regression. Column 1 shows the first stage regression which is of expected sign, statistically significant and with high value of F-stat. Column 2 and 3 show the second stage results i.e. γ_1 with and without controls, respectively. As seen from these results, the number of female kids above the age cutoff used in analysis are associated with a reduction of about 4 percentage points in open defecation in poorer urban regions. These results are very close to the reduced form results shown in the previous section.

9 Conclusion

This paper focuses on negative externalities of open defecation on female children/adult girls and its potential association with sanitation behavior of a household. Findings suggest that incentivised by the presence of female child, poor households living in urban areas (where cost of open defecation is higher) reduce open defecation. This association does not exist for richer households in urban areas and households in rural areas. This association seems to be driven by "shame" costs on the household if the female child gets harassed while defecating in the open. In addition, this relation seems to be stronger in Indian states which have higher rate of crime against women. These results provide a new first stage association between gender composition of children and household sanitation behavior and also inform policy about who adopts toilets for girls and who does not.

References

Adukia, A. 2017. Sanitation and Education. *American Economic Journal: Applied Economics*, **9(2)**, 23–59.

Barcellos, Silvia H., Carvalho, L., & Lleras-Muney, A. 2014. Child Gender and Parental Investments in India: Are Boys and Girls Treated Differently? *American Economic Journal: Applied Economics*, **6**, 157–189.

Bhalotra, S., & Cochrane.C, T. 2010. Where Have All the Young Girls Gone? Identification of Sex Selection in India. *IZA Discussion Paper no. 5381, Institute for the Study of Labor, Bonn.*

Bleakley, H. 2007. Disease and Development: Evidence from Hookworm Eradication in the American South. *Quarterly Journal of Economics*, **122 (1)**, 73–117.

Clasen, T, Boisson, Sophie, Routray, Parimita, Torondel, Belen, Bell, Melissa, Cumming, Oliver, Ensink, Jeroen, & 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. *Lancet Global Health*, **2(11)**, e645–53.

Coffey, D., & Spears, D. 2017. *Where India Goes: Abandoned Toilets, Stunted Development and Costs of Caste*. Harper Collins.

Coffey, D., Geruso, M., & Spears, D. 2017. Sanitation, Disease Externalities, and Anemia: Evidence From Nepal. *The Economic Journal*, **Accepted**.

Cutler, D., & Miller, G. 2005. The role of public health improvements in health advances: The twentieth-century United States. *Demography*, **42 (1)**, 1–22.

Deaton, A. 2003. Health, Inequality, and Economic Development. *Journal of Economic Literature*, **41, No.1**, 113–58.

Duflo, E., Greenstonea, M., Guiteras, R., & Clasen, T. 2015. Toilets Can Work: Short and Medium Run Health Impacts of Addressing Complementarities and Externalities in Water and Sanitation. *NBER WP 21521*.

Geruso, M., & Spears, D. 2018. Neighborhood Sanitation and Infant Mortality. *American Economic Journal: Applied Economics*, **10**, 125–62.

Guiteras, R., Levinsohn, J., & Mobarak, AM. 2015. Sanitation subsidies. Encouraging sanitation investment in the developing world: a cluster-randomized trial. *Science*, **348(6237):903-6**.

Hammer, J., & Spears, D. 2016. Village sanitation and child health: Effects and external validity in a randomized field experiment in rural India. *Journal of Health Economics*, **48**, 135–48.

Jadhav, A., Weitzman, A., & Smith-Greenway, E. 2016. Household sanitation facilities and women's risk of non-partner sexual violence in India. *BMC Public Health*.

JAGORI, & UN-Women. 2010. *Safe Cities Free of Violence Against Women and Girls Initiative: Baseline Survey in Delhi*. Report. Jagori and UN Women.

Jayachandran, S., & Kuziemko, I. 2011. Why Do Mothers Breastfeed Girls Less than Boys? Evidence and Implications for Child Health in India. *Quarterly Journal of Economics*, **126**, 1485–1538.

Jeffery, P., Jeffery, R., & Lyon, A. 1989. Labor Pains and Labour Power. *London:Zed*.

Khadgawat, R., Marwaha, RK., Mehan, N., Surana, V., Dabas, A., Sreenivas, V., Gaine, MA., & Gupta, N. 2016. Age of Onset of Puberty in Apparently Healthy School Girls from Northern India. *Indian Pediatrics*, **53 Issue 5**, 383–387.

Kishore, A., & Spears, D. 2014. Having a son promotes clean cooking fuel use in urban India: Women's status and son preference. *Economic Development and Cultural Change*, **62 Issue 4**, 673–699.

Pl, Gertler., Shah, Manisha, Alzua, Maria Laura, Cameron, Lisa, Martinez, Sebastian, & Patil., Sumeet. 2015. How Does Health Promotion Work? Evidence from the Dirty Business of Eliminating Open Defecation. *NBER WP 20997*.

Rosenblaum, D. 2013. The Effect of Fertility Decisions on Excess Female Mortality in India. *Journal of Population Economics*, **26 No. 1**, 147–80.

Sen, A. 2003. Missing Women Revisited: Reduction in Female Mortality Has Been Counter balanced by Sex Selective Abortions. *British Medical Journal*, **327**, 1297–98.

Spears, D., & Lamba, S. 2016. "Effects of Early-Life Exposure to Rural Sanitation on Childhood Cognitive Skills: Evidence from India's Total Sanitation Campaign. *Journal of Human Resources*, **51 (2)**, 298–327.

Stopnitzky, Y. 2017. No Toilet No Bride? Intrahousehold Bargaining in Male-Skewed Marriage Markets in India. *Journal of Development Economics*, **127**, 269–282.

UNICEF, & WHO. 2014. *Progress on Drinking Water and Sanitation 2014 Update*. Report. UNICEF.

Watson, T. 2006. Public health investments and the infant mortality gap: Evidence from federal sanitation interventions on U.S. Indian reservations. *Journal of Public Economics*, **90 (8–9)**, 1537–60.

Mathematical Appendix

In addition to assumptions in Section III-A, I assume, $C(1,0)$ is given; the aggregate resources an household i, $I \in [I_L, I_H]$ where $I_L > 0$ and $I_H < \infty$; household can only have a finite number of girl children above the cut-off age, $F_i = 0, 1, 2, ..., N$; the minimum φ, $\varphi_L < u(I_H) - u(I_H - P) - \omega C(N, 0)$.

Proposition: ***Case 1***: $\forall\ I_i\ \exists\ \bar{\varphi}$, such that, for $\varphi_i < \bar{\varphi}$,

$$V_i(0) > V_i(1) \qquad \forall i \tag{10}$$

Proof: Consider equation (10),

$$\begin{aligned} V_i(0) &> V_i(1) \\ \implies U(I_i, \varphi_i, 0) - \omega C(F_i, 0) &> U(I_i - P, \varphi_i, 1) \\ \implies u(I_i) - \varphi_i - \omega C(F_i, 0) &> u(I_i - P) \\ \implies \hat{\varphi}(I_i, F_i) \equiv u(I_i) - u(I_i - P) - \omega C(F_i, 0) &> \varphi_i \end{aligned} \tag{11}$$

In equation (11), we obtain the cutoff as a function of income and the number of children. In order to make the cutoff independent of them, we observe that, $\hat{\varphi}(I_i, F_i)$ is decreasing in both, I_i[20] and F_i[21]. Hence, the lowest value of $\hat{\varphi}(I_i, F_i)$ is the cutoff level below which all households do not adopt a toilet. Now, we obtain,

$$\bar{\varphi} \equiv \hat{\varphi}(I_H, N) = u(I_H) - u(I_H - P) - \omega C(N, 0) \tag{12}$$

■

[20] The derivative of $\hat{\varphi}(I_i, F_i)$ wrt I_i is $u'(I_i) - u'(I_i - P)$. Since $u(.)$ is concave in X, $I_i > I_i - P$ implies that $u'(I_i) < u'(I_i - P)$.

[21] The derivative of $\hat{\varphi}(I_i, F_i)$ wrt F_i is $-\omega C_F(F_i, 0)$. Here, $C_F(F_i, 0) > 0$ implies that $-\omega C_F(F_i, 0) < 0$.

Case 2: For all $\varphi_i > \bar{\varphi}$, $\exists\ \bar{I}$ such that for $I_i > \bar{I}$,

$$V_i(1) > V_i(0) \quad \forall i, \forall F_i \geq 0, \tag{13}$$

Proof: Equation (13) suggests that the household has a higher net utility from adopting a toilet:

$$U(I_i - P, \varphi_i, 1) > U(I_i, \varphi_i, 0) - \omega C(F_i, 0) \tag{14}$$

In this case, equation (14) should hold for all possible values of F_i. Since, the RHS is decreasing in F_i, if the inequality holds for $F_i = 0$, it also holds for $F_i > 0$. Therefore, the condition in equation (14) reduces to:

$$U(I_i - P, \varphi_i, 1) > U(I_i, \varphi_i, 0) \tag{15}$$

Note, that equation (15) does not depend on ω. From equation (15), we get,

$$u(I_i - P) > u(I_i) - \varphi_i \tag{16}$$

$$\implies \varphi_i > u(I_i) - u(I_i - P) \tag{17}$$

Note, that the LHS is constant for all I_i[22] and the RHS is decreasing in I_i[23]. Hence, corresponding to each level of $\varphi_i > \bar{\varphi}$, there exists an $\bar{I}(\varphi_i)$ such that for $I_i > \bar{I}(\varphi_i)$ households will always adopt a toilet. $\bar{I}(\varphi_i)$ is given by:

$$\varphi_i = u(\bar{I}(\varphi_i)) - u(\bar{I}(\varphi_i) - P) \tag{18}$$

■

[22]

[23] Derivative of the RHS wrt I_i is $u'(I_i) - u'(I_i - P)$. Since $u(.)$ is concave in X, $I_i > I_i - P$ implies that $u'(I_i) < u'(I_i - P)$

Case 3: For given cost $C(1,0) = C_1$, $\varphi_i > \bar{\varphi}$, and $I_i < \bar{I}(\varphi_i)$ $\exists \tilde{I} < \bar{I}(\varphi_i)$ such that $\forall i$ with $I_i \in [\tilde{I}, \bar{I}(\varphi_i)]$,

$$U(I_i, \varphi_i, 0) > U(I_i - P, \varphi_i, 1), \qquad \text{if } F_i = 0 \tag{19}$$

$$U(I_i - P, \varphi_i, 1) > U(I_i, \varphi_i, 0) - \omega C(F_i, 0) \qquad \text{if } F_i > 0 \tag{20}$$

Conceptual Explanation: I will provide a conceptual explanation supporting the validity of Case 3 above. Consider equation (19),

$$U(I_i, \varphi_i, 0) > U(I_i - P, \varphi_i, 1) \tag{21}$$

$$\implies u(I_i) - \varphi_i(1-0) - \omega C(F_i, 0) > u(I_i - P) \tag{22}$$

$$\implies u(I_i) - u(I_i - P) > \varphi_i \tag{23}$$

In this case, there are no adult girl child in household, $I_i \in [\tilde{I}, \bar{I}(\varphi_i)]$ and I_i is sufficiently low, such that marginal utility out of income is higher and a gain in utility because of not adopting a toilet is higher than the social cost of open defecation. Hence, in this case a household will not adopt a toilet.

Similarly, consider equation (20),

$$U(I_i - P, \varphi_i, 1) > U(I_i, \varphi_i, 0) - \beta C(F_i, 0) \tag{24}$$

$$\implies u(I_i) - u(I_i - P) < \varphi_i + \omega C(F_i, 0) \tag{25}$$

In this case, there are adult girl child in household (hence, $C(.) > 0$), all other factors are same as equation (19). The gain in utility by not adopting a toilet in this case is outweighed by the social cost + cost associated with an adult female child in the household. Hence, in this case a household will adopt toilet only if $C(.) > 0$

■

Tables

Table 1: Testable Predictions

		No Female Children (T = 0) 0	1
Female Children (T = 1)	0	Never a Toilet **(Case 1)**	Defiers
	1	Compliers **(Case 3)**	Always a Toilet **(Case 2)**

Notes: This Table represents the Testable Predictions from Section III. Decision to adopt a toilet or not is represented in Second Row and Second Column by 0 and 1. It is 1 if households adopts a toilet and 0 otherwise. Treatment is the presence of female children in a household and is represented by $T = 1$ for treated and $T = 0$ for control.

Table 2: Revisiting Testable Predictions

		No Female Children (T = 0)	
		0	1
Female Children (T = 1)	0	Never a Toilet **(Rural HHs &) Poorest urban HHs)**	Defiers
	1	Compliers **(Poorer Urban HHs)**	Always a Toilet **(Richer Urban HHs)**

Notes: This Table represents the version of Table 1 above with tested empirical evidence. Decision to adopt a toilet or not is represented in Second Row and Second Column by 0 and 1. It is 1 if households adopts a toilet and 0 otherwise. Treatment is the presence of female children in a household and is represented by $T = 1$ for treated and $T = 0$ for control.

Table 3: Mean Differences Check - *First born ≤ 5 years*

	Rural			Urban		
Outcome	Mean	Difference Girl - Boy	P-Value	Mean	Difference Girl - Boy	P-Value
Wealth: Asset Index	0	-0.016	0.54	0	-0.014	0.414
Sanitation: Open Defecation Rate	0.105	0.003	0.699	0.477	0.005	0.582
Source of Water						
Piped in Dwelling	0.304	-0.018	0.122	0.094	-0.004	0.372
Piped to yard/plot	0.186	0.001	0.915	0.099	0.002	0.672
Public tap/standpipe	0.156	0.01	0.297	0.139	0	0.964
Cooking Fuel						
LPG/Natural Gas	0.767	-0.003	0.786	0.218	-0.001	0.91
Kerosene	0.028	-0.007	0.077	0.008	0.001	0.691
Coal/Lignite	0.018	0.002	0.623	0.009	0	0.945
Wood	0.129	0.006	0.511	0.631	-0.007	0.426
Animal Dung	0.014	0.002	0.458	0.059	0.002	0.561
Household Construction						
Floor Type: Mud/Clay/Earth	0.093	-0.011	0.132	0.43	0	0.974
Floor Type: Brick	0.007	0.002	0.42	0.007	-0.001	0.445
Floor Type: Stone	0.042	-0.005	0.366	0.026	-0.004	0.179
Floor Type: Cement	0.556	0.004	0.748	0.326	0.008	0.349
Roof Type: Metal/Gi	0.19	-0.006	0.582	0.271	0.006	0.45
Roof Type: Calamine/Cement Fibre	0.036	0.007	0.127	0.022	0.001	0.686
Roof Type: Asbestos Sheet	0.067	0.006	0.34	0.069	-0.002	0.696
Roof Type: RCC/RBC/Cement/Concrete	0.508	-0.009	0.459	0.212	-0.002	0.755
Wall Type: Mud	0.045	-0.006	0.224	0.218	0.008	0.293
Wall Type: Bamboo with Mud	0.027	0.002	0.697	0.095	0.001	0.814
Wall Type: Cement/Concrete	0.481	-0.002	0.854	0.217	0.004	0.533
Wall Type: Burnt Bricks	0.222	-0.013	0.237	0.21	0	0.999
Household/Child Characteristics						
Religion: Hindu	0.717	-0.006	0.58	0.731	-0.007	0.328
Religion: Muslim	0.153	-0.004	0.698	0.123	0.005	0.393
Mother's Age	26.077	-0.073	0.548	24.452	0.056	0.467
Father's Age	31.072	-0.136	0.359	29.469	0.026	0.808
Is the mother currently pregnant?	0.131	0.03	0.001	0.166	0.03	0
Number of HH members	3.638	0.008	0.803	3.864	0.009	0.747
Number of Children (under 5 years)	0.998	0.006	0.334	0.999	0.005	0.3
Number of Women (15-49 yrs)	1.137	0.013	0.253	1.167	-0.001	0.951
First Born received prenatal care?	0.769	-0.002	0.831	0.588	0.002	0.795
First Born alive?	0.981	0.002	0.61	0.967	0.005	0.082
Education/Awareness						
Mother's Education: No Education	0.095	-0.001	0.847	0.23	0.011	0.137
Mother's Education: Primary	0.084	-0.009	0.218	0.14	-0.012	0.046
Mother's Education: Secondary	0.547	-0.001	0.963	0.545	-0.002	0.79
Mother's Education: Higher	0.273	0.011	0.355	0.085	0.003	0.474
Father's Education: No Education	0.07	0.005	0.477	0.141	0	0.959
Father's Education: Primary	0.102	-0.003	0.665	0.169	0.002	0.797
Father's Education: Secondary	0.538	0.002	0.86	0.577	-0.002	0.825
Father's Education: Higher	0.289	-0.004	0.752	0.113	0	0.988
Reading Paper atleast once a week	0.382	0.002	0.848	0.146	0.01	0.117
Watching TV once a week	0.866	-0.006	0.511	0.589	0.002	0.808

Notes: This table reports the result on estimating equation (1) on the part of NFHS data in which a). Responding mother is the head of household or if wife of household head, b). First born child is alive and of less than 5 years of age. Standard errors are clustered at PSU level.

Table 4: Mean Differences Check - *Main Analysis Sample*

Outcome	Rural			Urban		
	Mean	Difference Girl - Boy	P-Value	Mean	Difference Girl - Boy	P-Value
Wealth: Asset Index	-0.04	-0.044	0	-0.09	-0.026	0
Source of Water						
Piped in Dwelling	0.337	-0.017	0.002	0.101	-0.007	0.005
Piped to yard/plot	0.173	0.005	0.251	0.115	0.005	0.041
Public tap/standpipe	0.143	0.006	0.141	0.134	0.002	0.579
Cooking Fuel						
LPG/Natural Gas	0.755	-0.017	0.001	0.208	-0.005	0.093
Kerosene	0.015	0.002	0.28	0.005	0.001	0.212
Coal/Lignite	0.019	0	0.962	0.008	0	0.923
Wood	0.153	0.01	0.027	0.605	0.004	0.36
Animal Dung	0.021	0	0.896	0.1	0.001	0.675
Household Construction						
Floor Type: Mud/Clay/Earth	0.078	0.007	0.052	0.386	0.001	0.725
Floor Type: Brick	0.007	0.001	0.207	0.008	0.001	0.253
Floor Type: Stone	0.057	0	0.989	0.034	-0.002	0.095
Floor Type: Cement	0.514	0.001	0.833	0.342	-0.004	0.313
Roof Type: Metal/Gi	0.146	0.005	0.28	0.2	0.003	0.416
Roof Type: Calamine/Cement Fibre	0.032	0.002	0.435	0.022	0	0.714
Roof Type: Asbestos Sheet	0.057	0.008	0.008	0.055	-0.002	0.23
Roof Type: RCC/RBC/Cement/Concrete	0.552	-0.009	0.135	0.274	0.002	0.615
Wall Type: Mud	0.04	0.004	0.151	0.184	0.004	0.254
Wall Type: Bamboo with Mud	0.021	-0.001	0.668	0.063	0.001	0.626
Wall Type: Cement/Concrete	0.481	-0.007	0.223	0.241	-0.004	0.271
Wall Type: Burnt Bricks	0.238	-0.001	0.8	0.255	0.004	0.273
Household/Child Characteristics						
Religion: Hindu	0.717	-0.004	0.428	0.765	0.001	0.702
Religion: Muslim	0.171	-0.001	0.828	0.111	0.003	0.286
Mother's Age	33.411	0.035	0.537	32.354	0.045	0.207
Father's Age	45.292	0.15	0.028	44.065	0.051	0.266
Is the mother currently pregnant?	0.019	0.007	0	0.029	0.016	0
Number of HH members	5.669	0.286	0	6.008	0.391	0
Number of Children (under 5 years)	0.511	0.143	0	0.682	0.217	0
Number of Women (15-49 yrs)	1.341	0.171	0	1.314	0.157	0
Education/Awareness						
Mother's Education: No Education	0.172	0.003	0.501	0.399	0.004	0.284
Mother's Education: Primary	0.115	0.003	0.529	0.172	-0.004	0.221
Mother's Education: Secondary	0.529	-0.002	0.731	0.39	0	0.924
Mother's Education: Higher	0.184	-0.004	0.402	0.039	0	0.938
Father's Education: No Education	0.157	-0.001	0.785	0.307	0.004	0.251
Father's Education: Primary	0.145	0.003	0.502	0.197	-0.001	0.744
Father's Education: Secondary	0.508	0.003	0.574	0.437	-0.005	0.176
Father's Education: Higher	0.191	-0.005	0.318	0.059	0.003	0.207
Reading Paper atleast once a week	0.406	-0.017	0.004	0.138	0.001	0.81
Watching TV once a week	0.881	-0.004	0.352	0.578	-0.001	0.867

Notes: This table reports the result on estimating equation (1) on the part of NFHS data in which a). Responding mother is the head of household or if wife of household head, b). First born child is alive and is between 8-16 years of age. Standard errors are clustered at PSU level.

Table 5: Gender of first born child and reduction in open defecation

	(1) OD	(2) OD	(3) OD	(4) OD	(5) OD
First Born (Female = 1)	-0.000	-0.017***	-0.013***	-0.016***	-0.017***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Asset Index		-0.051***	-0.043***	-0.042***	-0.042***
		(0.000)	(0.000)	(0.000)	(0.000)
Region (Rural = 1)			0.280***	0.275***	0.275***
			(0.004)	(0.004)	(0.004)
Total number of kids				0.026***	0.027***
				(0.001)	(0.001)
Number of female kids				-0.011***	-0.011***
				(0.001)	(0.001)
First Born Age >= 8 (= 1)					-0.0108***
					(0.00260)
Observations	220203	220203	220203	220203	220203

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.0001$

Notes: This table reports the estimates from equation (4) on the entire NFHS data where a). Responding mother is the head of household or if wife of household head and, b). First born child is alive. Outcome varibale in all columns if the indicator variable OD = 1 if household defecates in open and OD = 0 if not. Standard errors in parenthesis. Standard errors are clustered at PSU.

Table 6: Reduced Form Results-*Urban Areas*

	Panel A: No Controls									
	(1st) OD	(2nd) OD	(3rd) OD	(4th) OD	(5th) OD	(6th) OD	(7th) OD	(8th) OD	(9th) OD	(10th) OD
First Born (Female)	0.014	-0.031*	0.002	-0.012	-0.011	-0.004	-0.005	0.003	0.003	0.001
	(0.017)	(0.014)	(0.014)	(0.012)	(0.011)	(0.008)	(0.007)	(0.004)	(0.004)	(0.003)
Mean OD Rate	0.391	0.219	0.144	0.11	0.083	0.046	0.026	0.013	0.004	0.004
P-Value		0.028								
Observations	3351	3609	2528	2727	2829	2635	2423	3740	1364	2358

	Panel B: Including Controls									
	(1st) OD	(2nd) OD	(3rd) OD	(4th) OD	(5th) OD	(6th) OD	(7th) OD	(8th) OD	(9th) OD	(10th) OD
First Born (Female)	0.008	-0.038**	0.002	-0.013	-0.012	-0.011	-0.004	-0.001	0.003	0.004
	(0.016)	(0.013)	(0.014)	(0.011)	(0.01)	(0.008)	(0.006)	(0.004)	(0.004)	(0.003)
Mean OD Rate	0.391	0.219	0.144	0.11	0.083	0.046	0.026	0.013	0.004	0.004
P-Value		0.005								
	3351	3609	2528	2727	2829	2635	2423	3740	1364	2358

Standard errors in parentheses

$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.0001$

Notes: This Table reports the estimates from equation (4) on the Main Analysis Sample in Urban Areas. Outcome varibale in all columns of both Panel A and B is the indicator variable OD = 1 if household defecates in open and OD = 0 if not. Analysis is for households belonging to deciles of asset index; from poorest to richest decile). In both Panel A & B, the sample as described in the text, is the set of households, a) where the husband of respondent (surveyed woman) is head of the household, b) with the first child at least 8 years old and at most 16 years old, and c). where the first born child is alive. Panel A reports the results without any controls while Panel B reports the results including household level controls. Controls include: Standardized value of asset index, number of household members, number of women and girls in the household, age of household head and the mother of first botn, education of household head and the mother, frequency of watching TV and reading newspaper in the household, main floor material, main roof material, main wall material, religion followed by the household and if the responding mother currently pregnant Standard Errors are in parenthesis. Standard errors clustered at Primary Sampling Unit.

Table 7: Reduced Form Results-*Rural Areas*

	(1st) OD	(2nd) OD	(3rd) OD	(4th) OD	(5th) OD	(6th) OD	(7th) OD	(8th) OD	(9th) OD	(10th) OD
Panel A: No Controls										
First Born (Female)	-0.008	-0.004	-0.008	-0.007	0.004	-0.012	-0.018	-0.012	-0.004	-0.000
	(0.008)	(0.013)	(0.009)	(0.013)	(0.013)	(0.013)	(0.014)	(0.014)	(0.011)	(0.009)
Mean OD Rate	0.792	0.728	0.648	0.563	0.493	0.421	0.361	0.3	0.229	0.121
P-Value							0.187			
Observations	10019	4778	11385	6044	5945	5565	5013	4238	5613	5082
Panel B: Including Controls										
	(1st) OD	(2nd) OD	(3rd) OD	(4th) OD	(5th) OD	(6th) OD	(7th) OD	(8th) OD	(9th) OD	(10th) OD
First Born (Female)	-0.008	0.002	-0.006	-0.002	-0.015	-0.017	-0.023^{+}	-0.005	-0.008	-0.007
	(0.007)	(0.01)	(0.008)	(0.011)	(0.012)	(0.012)	(0.012)	(0.013)	(0.01)	(0.009)
Mean OD Rate	0.792	0.728	0.648	0.563	0.493	0.421	0.361	0.3	0.229	0.121
P-Value							0.064			
	10019	4778	11385	6044	5945	5565	5013	4238	5613	5082

Standard errors in parentheses

$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.0001$

Notes: This Table reports the estimates from equation (4) on the Main Analysis Sample in Rural Areas. Outcome varibale in all columns of both Panel A and B is the indicator variable OD = 1 if household defecates in open and OD = 0 if not. Analysis is for households belonging to deciles of asset index; from poorest to richest decile). In both Panel A & B, the sample as described in the text, is the set of households, a) where the husband of respondent (surveyed woman) is head of the household, b) with the first child at least 8 years old and at most 16 years old, and c). where the first born child is alive. Panel A reports the results without any controls while Panel B reports the results including household level controls. Controls include: Standardized value of asset index, number of household members, number of women and girls in the household, age of household head and the mother of first botn, education of household head and the mother, frequency of watching TV and reading newspaper in the household, main floor material, main roof material, main wall material, religion followed by the household and if the responding mother currently pregnant. Standard Errors are in parenthesis. Standard errors clustered at Primary Sampling Unit.

Table 8: Falsification Results

Panel A: Urban Areas

	(1st) OD	(2nd) OD	(3rd) OD	(4th) OD	(5th) OD	(6th) OD	(7th) OD	(8th) OD	(9th) OD	(10th) OD
First Born (Female)	-0.000	-0.004	-0.003	-0.003	-0.003	0.008	-0.004	0.001	0.003	0.003
	(0.017)	(0.014)	(0.017)	(0.014)	(0.013)	(0.011)	(0.01)	(0.005)	(0.005)	(0.003)
Mean OD Rate	0.341	0.19	0.136	0.095	0.075	0.042	0.034	0.012	0.004	0.002
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2618	2490	1612	1668	1585	1454	1178	1635	539	870

Panel B: Rural Areas

	(1st) OD	(2nd) OD	(3rd) OD	(4th) OD	(5th) OD	(6th) OD	(7th) OD	(8th) OD	(9th) OD	(10th) OD
First Born (Female)	0.002	0.001	-0.002	0.015	0.003	0.012	0.002	0.005	0.001	0.001
	(0.008)	(0.014)	(0.009)	(0.014)	(0.015)	(0.015)	(0.016)	(0.017)	(0.015)	(0.014)
Mean OD Rate	0.773	0.69	0.615	0.537	0.481	0.416	0.365	0.317	0.264	0.148
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7220	2965	6737	3623	3422	3344	2756	2318	3099	2559

Standard errors in parentheses

$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.0001$

Notes: This Table reports the estimates from equation (4) on the Falsification sample in Urban and Rural Areas (Panel A and B respectively). Outcome varibale in all columns of both Panel A and B is the indicator variable OD = 1 if household defecates in open and OD = 0 if not. Analysis is for households belonging to deciles of asset index; from poorest to richest decile). The falsification sample as described in the text, is the set of households, a) where the husband of respondent (surveyed woman) is head of the household, b) with the first child at most 8 years old, and c). where the first born child is alive. Both Panel A & report the results including household level controls. Controls include: Standardized value of asset index, number of household members, number of women and girls in the household, age of household head and the mother of first botn, education of household head and the mother, frequency of watching TV and reading newspaper in the household, main floor material, main roof material, main wall material, religion followed by the household and if the responding mother currently pregnant. Standard Errors are in parenthesis. Standard errors clustered at Primary Sampling Unit.

Table 9: Mean Differences Check - *Urban households in 2nd decile of Asset Index*

Outcome	Mean	Difference Girl - Boy	P-Value
Wealth: Asset Index	-0.951	-0.001	0.818
Source of Water			
Piped in Dwelling	0.213	-0.007	0.588
Piped to yard/plot	0.182	0.012	0.377
Public tap/standpipe	0.229	0.001	0.948
Cooking Fuel			
LPG/Natural Gas	0.572	-0.004	0.828
Kerosene	0.032	0.001	0.926
Coal/Lignite	0.037	0.003	0.678
Wood	0.27	0.007	0.66
Animal Dung	0.034	-0.004	0.486
Household Construction			
Floor Type: Mud/Clay/Earth	0.149	-0.016	0.19
Floor Type: Brick	0.149	-0.016	0.19
Floor Type: Stone	0.011	0.004	0.223
Floor Type: Cement	0.057	0	0.99
Roof Type: Metal/Gi	0.219	-0.002	0.882
Roof Type: Calamine/Cement Fibre	0.042	0.001	0.916
Roof Type: Asbestos Sheet	0.101	0.018	0.07
Roof Type: RCC/RBC/Cement/Concrete	0.333	-0.024	0.126
Wall Type: Mud	0.08	0.008	0.389
Wall Type: Bamboo with Mud	0.033	-0.004	0.469
Wall Type: Cement/Concrete	0.377	-0.011	0.499
Wall Type: Burnt Bricks	0.253	0.008	0.6
Household/Child Characteristics			
Religion: Hindu	0.69	0.004	0.792
Religion: Muslim	0.195	0.009	0.491
Mother's Age	32.415	0.144	0.349
Father's Age	37.684	0.064	0.742
Is the mother currently pregnant?	0.023	0.013	0.01
Number of HH members	4.949	0.263	0
Number of Children (under 5 years)	0.512	0.139	0
Number of Women (15-49 yrs)	1.121	0.162	0

Notes: This table reports the estimates of equation (1) on the urban areas belonging to the main analysis sample and in 2nd decile of asset index. The main analysis sample as described in the text, is the set of households, a) where the husband of respondent (surveyed woman) is head of the household, b) with the first child is 8-16 years of old, and c). where the first born child is alive. Standard errors are clustered at PSU.

Table 10: Mean Differences Check - *Urban households in 2nd decile of Asset Index*

Assets	Mean	Difference Girl - Boy	P-Value
Mattress	0.677	0.031	0.044
Pressure Cooker	0.721	0.007	0.659
Chair	0.725	0.014	0.362
Cot or Bed	0.846	-0.009	0.439
Table	0.479	0.009	0.581
Electric Fan	0.82	-0.009	0.482
Color Television	0.763	-0.014	0.34
Sewing Machine	0.152	0.011	0.373
Internet	0.013	-0.005	0.176
Computer	0.002	0.001	0.663
Air Conditioner/Cooler	0.085	0.006	0.495
Washing Machine	0.022	0.002	0.761
Water Pump	0.038	0.004	0.556
Thresher	0.002	-0.002	0.29
Electricity	0.976	-0.01	0.06
Radio	0.044	0	0.968
Refrigerator	0.095	0.005	0.613
Bicycle	0.409	-0.039	0.017
Motorcycle/Scooter	0.139	-0.007	0.539
Car/Truck	0.007	-0.003	0.214

Notes: This table reports the estimates of equation (1) on the urban areas belonging to the main analysis sample and in 2nd decile of asset index. The outcome variables are various assets reported in NFHS data. The main analysis sample as described in the text, is the set of households, a) where the husband of respondent (surveyed woman) is head of the household, b) with the first child is 8-16 years of old, and c). where the first born child is alive. Standard errors are clustered at PSU.

Table 11: Gender composition (number of females) of the households

	(1) All Female	(2) Daughters Only	(3) Other than Daughters	(4) Daughters in Law only
First Born (Female = 1)	1.051***	1.057***	-0.00579	-0.00249*
	(0.0251)	(0.0206)	(0.0139)	(0.00110)
Observations	6099	6099	6099	6099

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.0001$

Notes: This table reports the estimates of equation (1) on the urban areas belonging to the main analysis sample and in 2nd decile of asset index, but including the households with the first born child less than 8 years as well. The main analysis sample as described in the text, is the set of households, a) where the husband of respondent (surveyed woman) is head of the household, b) with the first child is 8-16 years of old, and c). where the first born child is alive. The outcome variable here are total number of female members of household (Column 1), number of daughters of household head (Column 2), number of female members other than daughters of household head (Column 3) and, number of daughters in law of household head (Column 4). Standard erros in parenthesis. Standard errors are clustered at PSU.

Table 12: Ranking States/UTs (OD Rate & Crime Against Women)

Rank	By OD Rate	By Crime Against Women
1	Odisha	Uttar Pradesh
2	Jharkhand	West Bengal
3	Bihar	Maharashtra
4	Chattisgarh	Rajasthan
5	Tamil Nadu	Madhya Pradesh
6	Karnataka	Assam
7	Madhya Pradesh	Odisha
8	Maharashtra	Andhra Pradesh
9	Andhra Pradesh	Telangana
10	Rajasthan	Delhi
11	Pondicherry	Karnataka
12	West bengal	Bihar
13	Uttar Pradesh	Kerala
14	Goa	Haryana
15	Gujarat	Gujarat
16	Telangana	Chhatisgarh
17	Daman & Diu	Jharkhand
18	Dadra & Nagar Haveli	Punjab
19	Jammu & Kashmir	Tamil Nadu
20	Uttarakhand	Jammu & Kashmir
21	Andaman & Nicobar	Uttarakhand
22	Himachal Pradesh	Himachal Pradesh
23	Haryana	Tripura
24	Punjab	Chandigarh
25	Delhi	Meghalaya
26	Chandigarh	Goa
27	Assam	Arunanchal Pradesh
28	Meghalaya	Manipur
29	Arunanchal Pradesh	Sikkim
30	Tripura	Mizoram
31	Manipur	Andaman & Nicobar
32	Nagaland	Nagaland
33	Sikkim	Pondicherry
34	Kerala	Daman & Diu
35	Mizoram	Dadra & Nagar Haveli
36	Lakshwadeep	Lakshwadeep

Notes: This table reports the ranking of states in India by a) the open defecation (OD) rate (calculated using NFHS data) and b) by the ranking of crime against women (Using http://ncrb.gov.in/StatPublications/CII/CII2015/FILES/Table%205.1.pdf)

Table 13: Reduction in OD and Crime Against Women - All States

	Crime Rank <= 18			Crime Rank > 18		
	(1)	(2)	(3)	(4)	(5)	(6)
	OD	OD	OD	OD	OD	OD
First Born (Female = 1)	-0.0307*	-0.0323*	-0.0307*	0.000987	-0.00444	-0.000276
	(0.0138)	(0.0133)	(0.0123)	(0.0133)	(0.0122)	(0.0122)
State Fixed Effects		Yes	Yes		Yes	Yes
Household Controls			Yes			Yes
OD Rate		0.25			0.12	
Observations	4066	4066	4066	2033	2033	2033

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.0001$

Notes: This table reports the estimates of equation (4) on the households in urban areas belonging to the main analysis sample and in 2nd decile of asset index. The main analysis sample as described in the text, is the set of households, a) where the husband of respondent (surveyed woman) is head of the household, b) with the first child is 8-16 years of old, and c). where the first born child is alive. The outcome variable is the indicator for open defecation (OD) in all Columns. Sample in Column 1, 2 & 3 are the households in states which rank 18 or lower in ranking of crime against women. Sample in Column 4, 5 & 6 are the households in states which rank more than 18 in ranking of crime against women. Household controls include: Standardized value of asset index, number of household members, number of women and girls in the household, age of household head and the mother of first born, education of household head and the mother, frequency of watching TV and reading newspaper in the household, main floor material, main roof material, main wall material, religion followed by the household and if the responding mother is currently pregnant. Standard errors are in parenthesis. Standard are clustered at PSU.

Table 14: Reduction in OD and Crime Against Women - 18 top OD States

	Crime Rank <= 9			Crime Rank > 9		
	(1)	(2)	(3)	(4)	(5)	(6)
	OD	OD	OD	OD	OD	OD
First Born (Female = 1)	-0.0365*	-0.0337*	-0.0306*	-0.0120	-0.0205	-0.0239
	(0.0166)	(0.0162)	(0.0150)	(0.0235)	(0.0232)	(0.0217)
State Fixed Effects		Yes	Yes		Yes	Yes
Household Controls			Yes			Yes
OD Rate		0.25			0.325	
Observations	2720	2720	2720	1620	1620	1620

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.0001$

Notes: This table reports the estimates of equation (4) on the households in urban areas belonging to the main analysis sample and in 2nd decile of asset index. Only the households ranking 18 or lower in OD rankings in Table 12 are included. The main analysis sample as described in the text, is the set of households, a) where the husband of respondent (surveyed woman) is head of the household, b) with the first child is 8-16 years of old, and c). where the first born child is alive. The outcome variable is the indicator for open defecation (OD) in all Columns. Sample in Column 1, 2 & 3 are the households in states which rank in top half of ranking of crime against women in this sample. Sample in Column 4, 5 & 6 are the households in states which rank in bottom half in ranking of crime against women. Household controls include: Standardized value of asset index, number of household members, number of women and girls in the household, age of household head and the mother of first born, education of household head and the mother, frequency of watching TV and reading newspaper in the household, main floor material, main roof material, main wall material, religion followed by the household and if the responding mother is currently pregnant. Standard errors are in parenthesis. Standard are clustered at PSU.

Table 15: Framework for Diff-in Diff

	First Born (Girl)	First Born (Boy)
FB Child between 8-16 years	Shame Concerns + FF for Girl Birth	No/Less Shame Concerns + FF for Boy Birth
FB Child less than 8 years	No/Less Shame Concern + FF for Girl Birth	No/Less Shame Concern + FF for Boy Birth

Notes: This Table represents the Difference in Differences framework. It shows Shame Concerns and Fixed Factors (FF) for households divided into categories of the gender of first born child and their age.

Table 16: Difference in Differences

	(1) (OD)	(2) (OD)	(3) (OD)	(4) (OD)	(5) (OD)	(6) (OD)	(7) (OD)	(8) (OD)	(9) (OD)	(10) (OD)
FirstBornGirl*AgeCutoff	0.008	-0.038*	0.003	-0.013	-0.014	-0.017	-0.002	-0.003	0.003	0.002
	(0.023)	(0.019)	(0.021)	(0.018)	(0.016)	(0.013)	(0.012)	(0.00696)	(0.007)	(0.004)
FirstBornGirl (=1)	0.005	0.0005	-0.003	-0.004	-0.000	0.009	-0.006	0.002	-0.002	0.001
	(0.017)	(0.014)	(0.017)	(0.013)	(0.013)	(0.011)	(0.01)	(0.005)	(0.006)	(0.003)
AgeCutoff	0.031*	0.023^{+}	-0.002	0.013	0.016	0.01	-0.004	0.001	-0.003	0.003*
	(0.016)	(0.013)	(0.015)	(0.012)	(0.011)	(0.009)	(0.008)	(0.004)	(0.005)	(0.002)
Household Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5969	6099	4140	4395	4414	4089	3601	5375	1903	3228

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.0001$

Notes: This table reports the estimates of equation (7) on the households in urban areas belonging to the main analysis sample. The main analysis sample as described in the text, is the set of households, a) where the husband of respondent (surveyed woman) is head of the household, b) with the first child is 8-16 years of old, and c). where the first born child is alive. The outcome variable is the indicator for open defecation (OD) in all Columns. Household controls include: Standardized value of asset index, number of household members, number of women and girls in the household, age of household head and the mother of first born, education of household head and the mother, frequency of watching TV and reading newspaper in the household, main floor material, main roof material, main wall material, religion followed by the household and if the responding mother is currently pregnant. Standard errors are in parenthesis. Standard are clustered at PSU.

Table 17: 2-Stage Least Squares

	(1) Number of Female Kids ($\geq$ 8 yrs)	(2) OD	(3) OD
First Born (Female = 1)	1.066**		
	(0.029)		
Number of Female Children ($\geq$ 8 yrs)		-0.028*	-0.041**
		(0.013)	(0.014)
Asset Index			-0.121*
			(0.052)
Household Controls	NO	NO	YES
F-Stat (1st Stage)	1318.07		
Observations	3609	3609	3609

$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Notes: This table reports the estimates from equation (8) in Column (1) followed by estimates from equation (9) in Column (2 & 3) on the households in urban areas belonging to the 2nd decile of asset index in the main analysis sample. The main analysis sample as described in the text, is the set of households, a) where the husband of respondent (surveyed woman) is head of the household, b) with the first child is 8-16 years of old, and c). where the first born child is alive. The outcome variable is number of female kids ($\geq$ 8 years of age) in Column 1 and the indicator for open defecation (OD) in Column 2 & 3. Household controls include: Standardized value of asset index, number of household members, number of women and girls in the household, age of household head and the mother of first born, education of household head and the mother, frequency of watching TV and reading newspaper in the household, main floor material, main roof material, main wall material, religion followed by the household and if the responding mother is currently pregnant. Standard errors are in parenthesis. Standard are clustered at PSU.

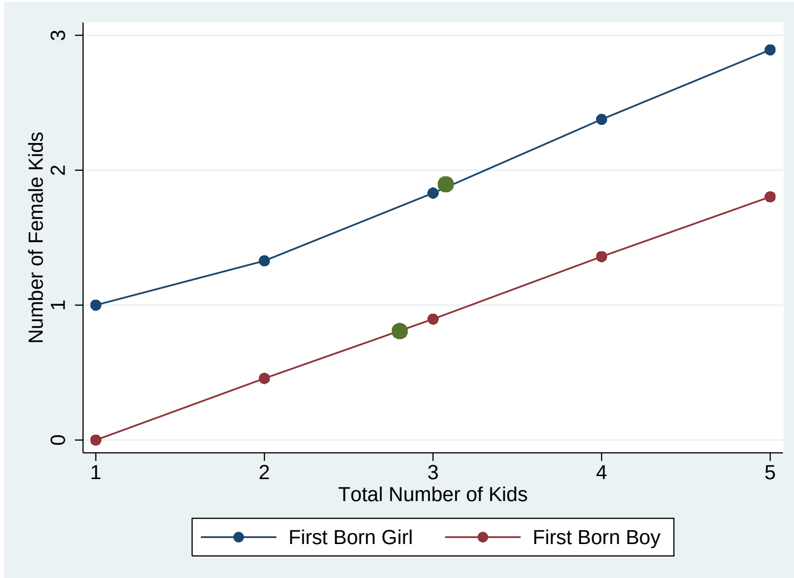

Figure 1: Sibling size and Gender composition by Gender of first born child

This Figure plots the association between Total Number of Kids and Number of Female Kids in a household, separated by the gender of first born child. The sample here is the Main Analysis Sample for Urban areas. The Green dots represent the interaction point of means from both axis.

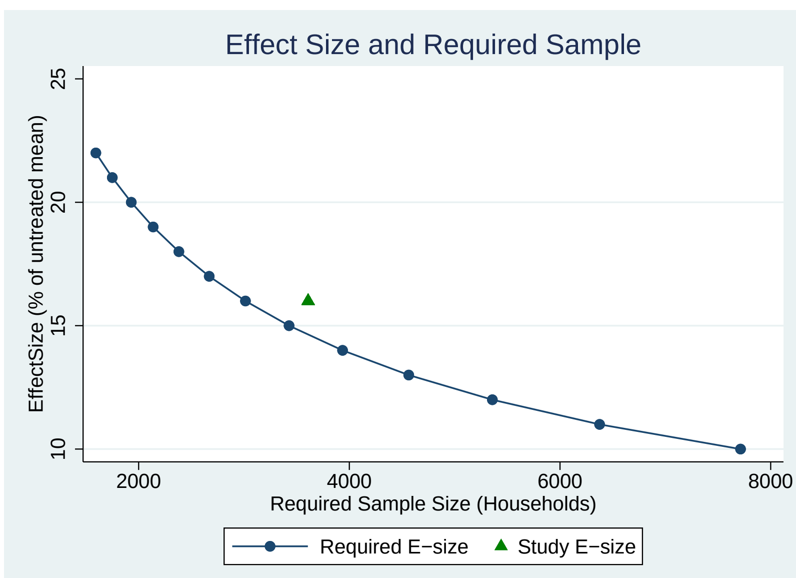

Figure 2: Retrospective Power Calculation (Urban Area)

This Figure shows the result of retrospective power calculations for analysis of urban areas in 2nd decile of asset index. The combination of study effect size and sample (green triangle) being on right side of blue curve of various other hypothetical combinations represents sufficient power to detect the observed effect size.

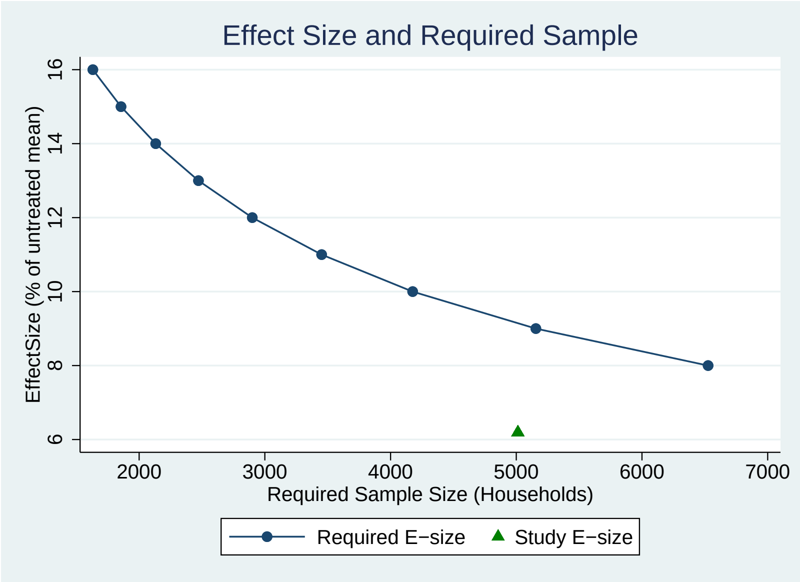

Figure 3: Retrospective Power Calculation (Rural Area)

This Figure shows the result of retrospective power calculations for analysis of rural areas in 7th decile of asset index. The combination of study effect size and sample (green triangle) being below the blue curve of various other hypothetical combinations represents sufficient power to detect the observed effect size.

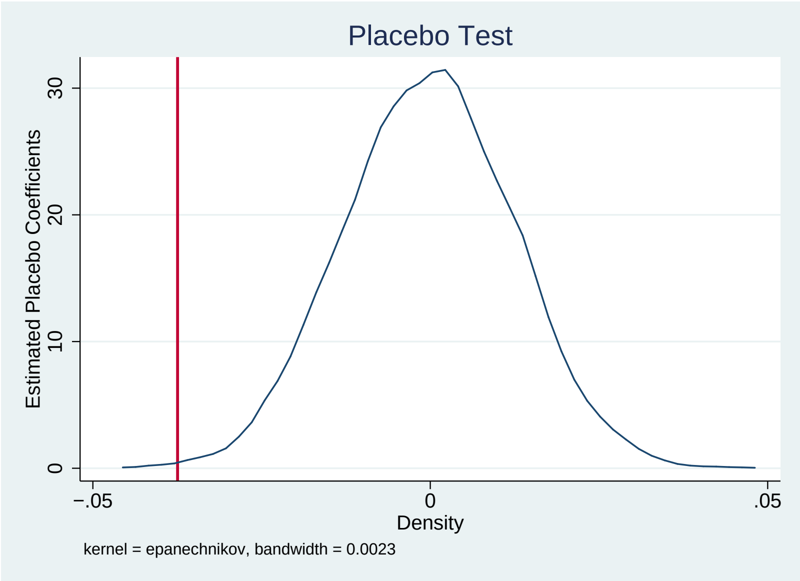

Figure 4: Placebo Test

This Figure plots the results of placebo exercise. The coefficients resulting from 3000 regressions similar to equation (4) with random treatment assignment are plotted along with the original coefficient (in red). Sample here is the Main Analysis Sample for urban households in 2nd decile of asset index.

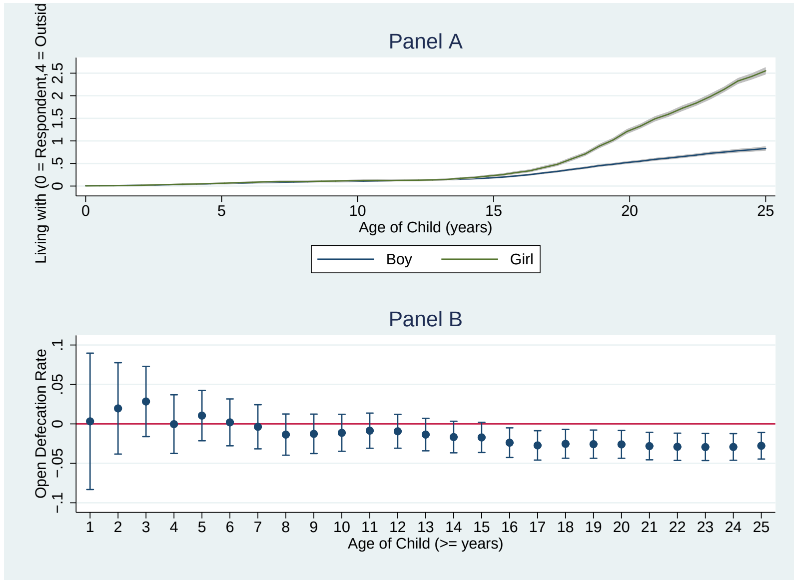

Figure 5

This Figure shows the analysis described in Section 7.1 of the paper. Panel A shows the polynomial fit of age of child on x-axis and whether or not they live with their parents on y-axis. This plot os separated by the gender of First born child. Plot for each gender has associated confidence intervals around them. Panel B shows the coefficient β_1 from estimating the equation (6) on urban households in 3nd decile of asset index of main analysis sample. Coefficients have associated confidence with them.

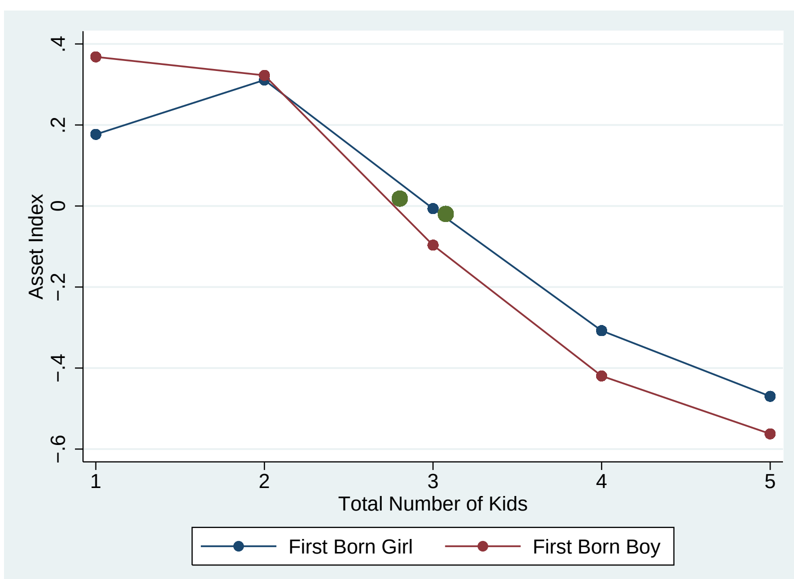

Figure 6: Wealth and Number of Kids by Gender of first born child

This Figure plots the association between Total Number of Kids and the Asset Index of a household, separated by the gender of first born child. The sample here is the Main Analysis Sample for Urban areas. The Green dots represent the interaction point of means from both axis.