Journal of Development Economics 164 (2023) 103083

Contents lists available at ScienceDirect

Journal of Development Economics

journal homepage: www.elsevier.com/locate/devec

Regular article

# Quantifying welfare effects in the presence of externalities: An ex-ante evaluation of sanitation interventions

Sanghmitra Gautam *

*Department of Economics, Washington University in St. Louis, United States of America*

## ARTICLE INFO

*JEL classification:*
C13
D12
D62
H23
I38

*Keywords:*
Welfare
Demand
Externalities
Policy
Subsidy
Sanitation

## ABSTRACT

This paper analyzes the impact of externalities on household demand for sanitation and the subsequent welfare effects generated from policy interventions. A critical feature of household sanitation (e.g., toilets) is that the take-up generates externalities where the privately chosen level is less than the socially optimal. To analyze the impact of policy interventions, I explicitly model household choice, taking into account the interdependence of household decision-making within the village. I identify and estimate the model using micro-survey data from India. Using the estimated model, I show how *untargeted price subsidies*, although cost effective at increasing sanitation coverage, have a regressive effect. I contrast this policy response with a *targeted cash transfer* to households with children, which ameliorates the regressive impact at the expense of a lower take-up.

## 1. Introduction

Many preventive healthcare goods, such as vaccines, water filters, bed nets, and sanitation, have consumption externalities whose benefit depends on the adoption decisions of other agents. While the adoption of such goods can generate large welfare gains, the allocation is likely to be inefficient. In the presence of an externality, households do not internalize all the benefits their adoption produces, and the privately chosen level is less than what is socially optimal. This paper focuses on quantifying the divergence between the public and private value of one such household good: a sanitation facility at home.

While the public and private health gains from sanitation coverage are well established (Coffey et al., 2018; Geruso and Spears, 2018), the distributional impact of policies to increase sanitation coverage is less understood. As a result, how to design policies that better capture the spillover benefits associated with sanitation take-up remains an open question. Moreover, the detrimental implications on human capital (Spears and Lamba, 2016) make this a key policy issue for developing country governments with limited resources wanting to implement effective policies. This paper analyzes the impact of externalities on household demand for sanitation and the subsequent welfare effects generated from policy interventions.

To analyze the importance of spillovers and the subsequent welfare gains of different policies on an individual household, I make explicit use of a theoretical framework. First, I model household behavior in the presence of externalities by specifying a household utility maximization problem that is embedded within a static incomplete information game to incorporate household interdependence in behavior. The model is identified and structurally estimated using both household survey and village-level data from India. I demonstrate the identification of the model primitives and implement a simple two-step method to obtain consistent parameter estimates, taking into account the possibility of multiple equilibria.

Second, I use the estimated model to compute price and income elasticity estimates to analyze household sanitation demand. Household demand for sanitation is significantly more responsive to changes in price relative to income. A median wealth household has a price elasticity eight times larger than the corresponding income elasticity estimates. The interdependence in household behavior generates spillover

*E-mail address:* sanghmitra.gautam@wustl.edu.

* I thank the editor and two referees for their valuable comments. I also thank Orazio Attanasio, Andrew Chesher, Aureo De Paula, Valerio Dotti, Bruce Petersen, Dean Spears, Duncan Thomas, Michela Tincani, Petra Todd, Giorgio Topa, and seminar participants at University College London, University of Mannheim, ESEM Geneva, ADBI-IFS Sanitation and Development conference. I also thank EDePo Institute for Fiscal Studies (IFS) and Britta Augsburg for access to the FINISH household dataset, and the Municipal Corporation of Gwalior (India) for sub-district and village census linked codes and their advice and assistance in data collection of cost variables. This paper comprises content from the first and second chapters of my PhD thesis. All remaining errors are my own.

https://doi.org/10.1016/j.jdeveco.2023.103083
Received 28 December 2021; Received in revised form 13 February 2023; Accepted 8 March 2023
Available online 11 March 2023

effects. Using the estimated model, I show that a change in price (or income) has both a direct and indirect impact, where the latter depends on the externalities in the village. Decomposition of the elasticity shows that the indirect effect comprises close to $2/3$ of the total demand response. In addition, I quantify the expected private and social benefits from sanitation adoption using a money measure. Specifically, a one dollar decrease in price generates a private household value equivalent to 9 dollars and a social value of 15 dollars from sanitation. The wedge between the two corresponds to the size of the public benefits from externalities.

Third, I use the empirical framework to conduct an ex-ante policy evaluation of two policy instruments: targeted cash transfers and price subsidies. The policy performance is evaluated on two outcomes: (i) average sanitation take-up rate and (ii) welfare effects. To analyze take-up, I quantify household sanitation adoption response to a 25% price subsidy and a cash transfer under budget neutrality. A subsidy increases sanitation take-up by 2.1 percentage points in villages with low adoption. In contrast, for the same total cost, a cash transfer policy increases sanitation take-up by only 0.2 percentage points. The results of this analysis are striking: under budget neutrality, a price subsidy generates an increase in take-up that is ten times the increase under a cash transfer. Consequently, a price subsidy is more cost effective than a targeted cash transfer in increasing the average sanitation adoption rate. This result shows that when considering the demand for goods that generate spillovers, that is, consumption externalities, an ex-ante evaluation before introducing a policy is critically important for the effective design of policy interventions.

Last, I compute an ex-ante money measure to quantify the expected welfare gains from each policy intervention. The welfare analysis provides two policy-relevant takeaways. First, an untargeted price subsidy has strong distributive effects. Specifically, a large share of the welfare gain is concentrated among relatively wealthy households. Second, a price subsidy has larger positive welfare effects on households that include children and female members. This analysis is of practical value to governments implementing large-scale policies with a limited budget. Field experiments measuring the impact of policies or provision of goods that generate externalities may benefit from an understanding of spillover effects ex-ante (Todd and Wolpin, 2008). By incorporating distributive welfare analysis, I show how untargeted price subsidies for sanitation adoption can have a regressive impact in terms of welfare.

Focusing on the welfare impacts and importance of externalities, this paper contributes to the existing literature on sanitation and impact evaluation on two fronts. First, the empirical analysis in this paper emphasizes the importance of an ex-ante evaluation to help avoid the high cost of implementing programs and is complementary to the existing literature on the impact evaluation of sanitation interventions. Recent policy evaluations of sanitation interventions in the Indian context include (Clasen et al., 2014; Hammer and Spears, 2016; Patil et al., 2014; Barnard et al., 2013). Community-led total sanitation (CLTS) interventions have been evaluated in several other country contexts, including Tanzania (Briceno et al., 2017), Mali (Pickering et al., 2015), Nigeria (Abramovsky et al., 2023), Ghana (Crocker et al., 2016), and Ethiopia (Crocker, 2016). In Bangladesh, Guiteras et al. (2015) demonstrate the effectiveness of CLTS with subsidy provision. In recent work, Guiteras et al. (2019) incorporate a BLP framework to analyze the ex-post impact of price subsidies delivered in a field experiment setting. I add to the existing sanitation policy evaluation literature, which thus far has been exclusively ex-post, by highlighting the complementary role of an ex-ante policy analysis first outlined by Todd and Wolpin (2008). I show how such cost-efficiency concerns can be incorporated into program design for an expensive good like a household sanitation facility and are important for governments and policymakers.

Second, in a related strand, Attanasio et al. (2013) emphasize the importance of structural policy analysis in quantifying the distributive impacts of *untargeted* price changes by quantifying the welfare gains from the Progresa conditional cash transfers (CCT). In this paper, I show how existing CCT programs offer an attractive policy tool for sanitation adoption in the form of a *targeted* cash transfer. Moreover, I demonstrate how efficiency and equity gains are often at odds in the case of sanitation adoption with externalities. Nevertheless, a policy-maker's objectives can be incorporated and evaluated rigorously using an ex-ante evaluation approach. The key policy implication emphasizes the attractiveness of existing CCT programs that can be incorporated within the arsenal of policy tools for sanitation adoption available to resource-constrained governments.

The rest of the paper proceeds as follows. Section 2 outlines the theoretical framework specifying a household utility maximization problem that is embedded within a static incomplete information game to incorporate household interdependence in behavior. Section 3 describes the data and discusses the identification and estimation of the model. Section 4 presents the model estimates and discusses the main results and model fit. Section 5 implements an ex-ante policy analysis to compare the cost efficiency and welfare gains derived under two policy instruments: a price subsidy and a targeted cash transfer. Section 6 concludes.

## 2. Model

This section introduces a static model of household choice of sanitation adoption. In this context, a private sanitation facility at home is a good that generates externalities, that is, costs and benefits derived from sanitation not reflected in the good's market price. A decision-making household derives utility that depends on its own and other households' sanitation adoption. Such spillover effects imply that the sanitation adoption choices of different households within a village are potentially interdependent. To capture this feature, I model household sanitation adoption decisions as the outcome of a simultaneous-move Bayesian–Nash game among households within the village. The use of simple static games has been popular in the recent empirical literature.[1] In this paper, I extend the general structure of a Bayesian–Nash game by embedding within it an explicit household choice structure derived from a standard consumer problem. The context of sanitation adoption is fairly general, that is, the household preference model captures several key incentives and trade-offs common to the choice of preventative health goods, which generate externalities.

### 2.1. Setup

A household is assumed to be a single decision-making unit, making a discrete decision on whether to adopt sanitation. The population of households is partitioned into groups indexed by $g = 1, \ldots, G$, each corresponding to a village community. There are a finite number of decision-making households $i = 1, \ldots, I_g$ in each village $g$. Each household $i$ in village $g$ simultaneously chooses an action $d_{ig} \in \{0, 1\}$ where:

$$d_{ig} = \begin{cases} 1 & Household\ adopts\ sanitation \\ 0 & no\ adoption \end{cases}$$

The vector of the actions chosen by all households in village $g$ is $d_g = \left(d_{1g}, \ldots, d_{I_g g}\right)$.

[1] More generally, static games of incomplete information have been used extensively to quantify strategic interactions among firms looking to enter a market (Jia, 2008), teacher–student interactions (Todd and Wolpin, 2018), and technology adoption within networks (Björkegren, 2019). A parallel set of papers have focused on the identification and estimation of such games, for example, Aguirregabiria and Mira (2007, 2019), Bajari et al. (2007), Pakes et al. (2007) and Pesendorfer and Schmidt-Dengler (2008).

### 2.2. Household preferences

Household preference represented by a direct utility function for household $i$ in village $g$ is denoted by:

$$u_{ig}\left(c_{ig}, d_{ig}, \bar{d}_{-ig}\right) \tag{1}$$

where $c_{ig}$ represents the consumption of a composite private good, $d_{ig} \in \{0, 1\}$ is own adoption of sanitation, and $\bar{d}_{-ig}$ is the average level of adoption in the village excluding household $i$, which is written as:

$$\bar{d}_{-ig} = \frac{1}{I_g - 1} \sum_{j \neq i} \mathbf{1}\left\{d_{jg} = 1\right\} \tag{2}$$

The formulation of preferences in Eq. (1) captures the possibility that $i$'s preferences over consumption and adoption of sanitation are affected by the aggregate adoption choices of other villagers. Household $i$ allocates their income $y_{ig}$ between the consumption of a composite private good $c_{ig}$ and sanitation $d_{ig}$, whose price is $p_g$. Thus, the household's budget constraint is written as follows:

$$c_{ig} + p_g d_{ig} - y_{ig} \leq 0 \tag{3}$$

The above setup corresponds to a standard single-agent consumer problem except for the presence of other agents' choices $\bar{d}_{-ig}$ in $i$'s utility function. This feature illustrates how households' optimal choices are interdependent. It generates an externality where each household incorporates the actions of others within its utility gains but does not account for the impact of their own choice on the well-being of others. In aggregate, this lack of accountability results in a divergence in the private and social optimal adoption levels. While this setup is relatively general, I impose parametric restrictions on the functional form of the preferences and the nature of strategic interactions to make the model empirically tractable. Specifically, the utility function has the functional form:

$$u_{ig} = \underbrace{\gamma d_{ig}\bar{d}_{-ig} + A_{ig}\left(\bar{d}_{-ig}\right)}_{S_{ig}} + B_{ig}\left(d_{ig}\right) + C_{ig}\left(c_{ig}, d_{ig}\right) \tag{4}$$

where $C_{ig}\left(c_{ig}, d_{ig}\right)$ is the utility from consumption, which may depend on own adoption, and $B_{ig}\left(d_{ig}\right)$ denotes the pure private utility from own sanitation consumption that is allowed to vary at the household and village level. The inclusion of $\bar{d}_{-ig}$ in the direct utility implies that households derive direct and indirect public gains $S_{ig}$ from other households' consumption of sanitation. An example of direct gains is free-riding on the sanitation facilities of a neighbor, while indirect gains can include public health externalities. Public gains $S_{ig}$ are modeled as two components of the utility function: $\gamma$ and $A_{ig}(\cdot)$.

The parameter $\gamma$ captures how the utility of own sanitation varies with the average level of adoption in the village. This may include health and information externalities affecting the private utility from sanitation, and peer effects. For instance, $\gamma$ may capture the effect of social pressure to adopt sanitation exerted by other households that already possess a sanitation facility at home.

The function $A_{ig}(\cdot)$ represents a pure public gain from the widespread adoption of sanitation in the village and affects the household utility regardless of its adoption choice. It captures, for instance, general health benefits a household enjoys solely from residing in a village with a high adoption rate. For example, a high $\bar{d}_{-ig}$ may imply a lower probability of contracting infectious diseases for the household members even if the household does not possess its own sanitation facility. However, the additive functional form of $A_{ig}(\cdot)$ does not allow for certain types of direct public gains from sanitation adoption. For example, the structure does not explicitly incorporate the possibility that in villages with a relatively high adoption level, household members may access and use their neighbor's sanitation facility. Nevertheless, if such free-riding behavior occurs in a village, it would *implicitly* be captured as a downward bias distortionary effect on the parameter $\gamma$; that is, the size of the social interaction parameter is shaped by the magnitude of the incentives to free ride.

The inclusion of $\bar{d}_{-ig}$ in the direct utility in Eq. (4) allows for interdependent adoption choices, but it also puts restrictions on how neighboring households within the village affect individual household gains. This assumption implies that the aggregate behavior and not the identity of an individual household matters. Nevertheless, the aggregate function captures a host of direct peer-to-peer interactions such as social norms, peer pressure, and indirect public health externalities, where the level of sanitation coverage in a village captures the cleanliness of the village environment and overall disease burden. In related work, Hammer and Spears (2016) find village-level sanitation coverage to be an important determinant of children's human capital. A general formulation that incorporates household identity when $i > 2$, as in this context, substantially complicates the model solution and raises additional identification issues.[2] For example, in recent work, Todd and Wolpin (2018) estimate a coordination game with player identity to capture strategic interactions among teachers and students in the production of student knowledge.

To summarize, notice that under the two choice realizations, the utility function in Eq. (4) takes the following forms[3]:

$$u^1_{ig} = C_{ig}\left(c^1_{ig}, d_{ig}\right) + \gamma\bar{d}_{-ig} + A_{ig}\left(\bar{d}_{-ig}\right) + B_{ig}(1)$$

$$u^0_{ig} = C_{ig}\left(c^0_{ig}, d_{ig}\right) + A_{ig}\left(\bar{d}_{-ig}\right) + B_{ig}(0)$$

These formulas illustrate two key features of the model. First, a household's utility is allowed to vary with the average adoption level in the village $\bar{d}_{-ig}$ both in the case of own adoption and non-adoption via the function $A_{ig}(\cdot)$. This implies that each household may be affected by the externality even if it decides not to adopt. Second, both the pure public gains from a high adoption rate in the village $A_{ig}(\cdot)$ and the private gains $B_{ig}(1) - B_{ig}(0)$ are allowed to vary across individual households and villages.

### 2.3. Information and indirect utility

***States and information.*** Each household $i$ is endowed with a set of state variables $\left(y_{ig}, x_{ig}, p_g, z_g, \epsilon_{ig}\right)$ that include household income $y_{ig}$, price of sanitation $p_g$, and other household- ($x_{ig}$) and village- ($z_g$) specific characteristics. In addition, each household is also endowed with a set of taste/preference shocks $\epsilon_{ig}$. These taste shocks are private information, known only to household $i$, and are unobservable to all other households within the group, including the econometrician. Let $\epsilon_{ig} = \left(\epsilon_{ig}(1), \epsilon_{ig}(0)\right)$ denote the $1 \times 2$ vector of the individual taste shocks $\epsilon_{ig}(d_{ig})$ and $y_g = (y_{1g}, \ldots, y_{ig}, \ldots, y_{I_g g})$ and $x_g = (x_{1g}, \ldots, x_{ig}, \ldots, x_{I_g g})$. The density of $\epsilon_{ig}$ is denoted by $f(\epsilon_{ig})$. It is assumed that $w_g = \left(y_g, x_g, p_g, z_g\right)$ is a vector of common knowledge variables that are observable to all households in village $g$ and the econometrician. Thus, the information available to each household $i$ is incomplete and restricted to the set of publicly observable variables $w_g$ and the own taste shock $\epsilon_{ig}$. This description is formalized through the following two assumptions:

**[A1]**: *The vector of observables $w_g$ is common knowledge possessed by all households $i$ within a village/group $g$.*

**[A2]**: *The unobserved taste shocks $\epsilon_{ig}$ are private information possessed only by a household $i$ and are distributed i.i.d across households and choice alternatives.*

[2] Such a specification, while more general and thus able to capture more complex patterns of social interactions, must be motivated by the overall objective. I leave the development and estimation of such a model of household behavior in the context of sanitation for future work.

[3] Where $u^1_{ig} = u\left(c^1_{ig}, 1, \bar{d}_{-ig}\right)$ and $u^0_{ig} = u\left(c^0_{ig}, 1, \bar{d}_{-ig}\right)$ denotes choice-specific utility levels, and $c^1_{ig}$, $c^0_{ig}$ denotes consumption levels that are choice-specific.

In addition to assumption [A2], $F(\epsilon)$ is assumed to have a parametric distribution from a known family. Specifically, the private taste shock terms $\epsilon_{ig}$ are distributed according to the type 1 extreme value distribution, implying that the difference $\left(\epsilon_{ig}(0)-\epsilon_{ig}(1)\right)$ possesses a logistic distribution.

***Decision rule.*** Given this information structure, I define a decision rule $\tilde{d}_{ig}$ for household $i$ as a function that maps observables $w_g$ and $i$'s realization of the taste shock $\epsilon_{ig}$ into a choice $d_{ig} \in \{0,1\}$. For an arbitrary decision rule $\tilde{d}_{ig}\left(w_g,\epsilon_{ig}\right)$, it is possible to construct the conditional choice probability of sanitation adoption for each household in the village. Let $\mathbb{P}_{ig}\left(d_{ig}=1 \mid w_g,\tilde{d}_{ig},\theta\right)$ denote the probability that household $i$ chooses to adopt sanitation ($d_{ig}=1$) conditional on $w_g$, preferences $\theta$ and given a decision rule $\tilde{d}_{ig}$. The conditional probability $\mathbb{P}_{ig}\left(d_{ig}=1 \mid w_g,\tilde{d}_{ig},\theta\right)$ is obtained by integrating the decision rule over the distribution of the taste shocks:

$$\mathbb{P}_{ig}\left(d_{ig}=1 \mid w_g,\tilde{d}_{ig},\theta\right)=\int \mathbf{1}\left\{\tilde{d}_{ig}\left(w_g,\epsilon,\theta\right)=1\right\} f(\epsilon)\, d\epsilon \tag{5}$$

where $\mathbf{1}\left\{\tilde{d}_{ig}\left(w_g,\epsilon,\theta\right)=1\right\}$ is an indicator function that household $i$'s choice is $d_{ig}=1$ (adoption) given the vector of state variables $\left(w_g,\epsilon_{ig}\right)$ and preferences $\theta$. Because the taste shock $\epsilon_{ig}$ is unobservable to all households $j \neq i$, $\mathbb{P}_{ig}\left(d_{ig}=1 \mid w_g,\tilde{d}_{ig},\theta\right)$ also equals the choice probability of adoption for household $i$ conditional on the information set $\left(w_g,\epsilon_{jg}\right)$ of any household $j \neq i$. In order words, it denotes the belief household $j$ has about $i$'s adoption choice. More generally, $\mathbb{P}_{ig}\left(d_{ig} \mid w_g,\tilde{d}_{ig},\theta\right)$ denotes the choice probability for any given choice $d_{ig}$ made by household $i$.

***Choice-specific indirect utility.*** Returning to the direct utility specification, I assume that the private benefits from consumption of sanitation in Eq. (4) are a function of household-level characteristics $x_{ig}$ with the functional form:

$$B_{ig}\left(d_{ig}\right)=\left(\delta_g+\beta x_{ig}\right) d_{ig}+\epsilon_{ig}\left(d_{ig}\right) \tag{6}$$

where $\delta_g$ represents a village-level fixed effect, and $\epsilon_{ig}\left(d_{ig}\right)$ is an i.i.d. taste shock. In principle, $\delta_g$ can be identified with appropriate variation in the data. However, because of data limitations — specifically, lack of repeated observations of the same village — I restrict such group effects to be a smooth function of village-level observable characteristics $z_g$, with the functional form $\delta_g=\kappa+\delta z_g$.[4] Moreover, I assume that the utility from the consumption of private goods has the functional form:

$$C_{ig}\left(c_{ig},d_{ig}\right)=\left[\alpha-\left(\zeta+\lambda x_{ig}\right)\left(1-d_{ig}\right)\right] c_{ig} \tag{7}$$

where the term in square brackets is the marginal utility of consumption, which is allowed to vary with own sanitation adoption $d_{ig}$ and with household-level observables $x_{ig}$. The specification of the function $C_{ig}(\cdot)$ in Eq. (7) captures the possibility that the extent to which own sanitation affects the marginal utility of private consumption may depend on some specific household characteristics, such as the number of children.

Substituting the formulas for $B_{ig}\left(d_{ig}\right)$ and $C_{ig}\left(c_{ig},d_{ig}\right)$ into Eq. (4) and using the budget constraint in Eq. (3), I obtain the choice-specific utility enjoyed by $i$ given adoption level $\bar{d}_{-ig}$, which is written as:

$$\begin{aligned} &V_{ig}^{d}\left(y_{ig},p_g;x_{ig},z_g,\bar{d}_{-ig},\epsilon_{ig},\theta\right)= \\ &\underbrace{a_{ig}+\kappa d_{ig}+\beta x_{ig}d_{ig}+\delta z_g d_{ig}+\zeta y_{ig}d_{ig}+\lambda x_{ig}y_{ig}d_{ig}+\xi p_g d_{ig}+\gamma d_{ig}\bar{d}_{-ig}}_{v_{ig}^{d}\left(y_{ig},p_g;x_{ig},z_g,\bar{d}_{-ig},\theta\right)} \\ &+\epsilon_{ig}\left(d_{ig}\right) \end{aligned} \tag{8}$$

where the superscript $d \in \{0,1\}$, $\theta$ is a vector of preference parameters $\theta=[\kappa,\beta,\delta,\lambda,\zeta,\xi,\gamma]$ with $\xi=-\alpha$, and $a_{ig}=A_{ig}\left(\bar{d}_{-ig}\right)+\alpha y_{ig}$ is a component of the household's indirect utility that is independent of the household's own sanitation adoption. Thus, each household's utility depends on the state variables $\left(y_{ig},x_{ig},p_g,z_g,\epsilon_{ig}\right)$, own choice $d_{ig}$, and a function of the choice of other households $\bar{d}_{-ig}$. Under incomplete information, a household forms beliefs about the adoption decision made by its neighbors. Because household $i$ does not have information on the taste preference of other households $\epsilon_{jg}$, it constructs beliefs about the expected choice of other households using all relevant observable information. That is, $i$ integrates its choice-specific indirect utility in Eq. (8) over the conditional probability measure $\mathbb{P}_{-ig}\left(d_{-ig} \mid w_g,\tilde{d}_{-ig},\theta\right)=\prod_{j\neq i}\mathbb{P}_{jg}\left(d_{jg} \mid w_g,\tilde{d}_{jg},\theta\right)$ where $\tilde{d}_{-ig}=\left(\tilde{d}_{1g},\ldots,\tilde{d}_{(i-1)g},\tilde{d}_{(i+1)g},\tilde{d}_{I_g g}\right)$ denotes the vector of decision rules of all households other than $i$. As a result, the expected utility $\sum_{d_{-ig}} V_{ig}^{d}\left(y_{ig},p_g;x_{ig},z_g,\bar{d}_{-ig},\epsilon_{ig},\theta\right)\mathbb{P}_{-ig}\left(d_{-ig} \mid w_g,\tilde{d}_{-ig},\theta\right)$ received by household $i$ from choosing $d_{ig}$ conditional on $\left(w_g,\epsilon_{ig}\right)$, denoted by $\tilde{V}_{ig}^{d}$, is written as:

$$\begin{aligned} &\tilde{V}_{ig}^{d}\left(y_{ig},p_g;w_g,\tilde{d}_{-ig},\epsilon_{ig},\theta\right) \\ &=\sum_{d_{-ig}}\left[a_{ig}+\kappa d_{ig}+\beta x_{ig}d_{ig}+\delta z_g d_{ig}+\zeta y_{ig}d_{ig}+\lambda x_{ig}y_{ig}d_{ig}\right. \\ &\quad\left.+\xi p_g d_{ig}+\gamma d_{ig}\bar{d}_{-ig}+\epsilon_{ig}\left(d_{ig}\right)\right]\cdot\mathbb{P}_{-ig}\left(d_{-ig} \mid w_g,\tilde{d}_{-ig},\theta\right) \end{aligned} \tag{9}$$

By applying the expectation over the conditional distribution of $d_{-ig}$, Eq. (9) simplifies to:

$$\begin{aligned} &\tilde{V}_{ig}^{d}\left(y_{ig},p_g;w_g,\tilde{d}_{-ig},\epsilon_{ig},\theta\right)= \\ &\underbrace{\bar{a}_{ig}+\kappa d_{ig}+\beta x_{ig}d_{ig}+\delta z_g d_{ig}+\zeta y_{ig}d_{ig}+\lambda x_{ig}y_{ig}d_{ig}+\xi p_g d_{ig}+\gamma d_{ig}\overline{\mathbb{P}}_{-ig}\left(w_g,\tilde{d}_{-ig},\theta\right)}_{\tilde{v}_{ig}^{d}\left(y_{ig},p_g;w_g,\tilde{d}_{-ig},\theta\right)} \\ &+\epsilon_{ig}\left(d_{ig}\right) \end{aligned} \tag{10}$$

where $\overline{\mathbb{P}}_{-ig}\left(w_g,\tilde{d}_{-ig},\theta\right)=\sum_{d_{-ig}}\bar{d}_{-ig}\mathbb{P}_{-ig}\left(d_{-ig} \mid w_g,\tilde{d}_{-ig},\theta\right)$ denotes the expected value of $\bar{d}_{-ig}$, and $\bar{a}_{ig}=\sum_{d_{-ig}}a_{ig}\mathbb{P}_{-ig}\left(d_{-ig} \mid w_g,\tilde{d}_{-ig},\theta\right)$ is a component that includes the expected value of the pure public gains for a household $i$ from the adoption of sanitation by other households in the village. Under the two choice realizations, we get:

$$\begin{aligned} \tilde{V}_{ig}^{1}\left(y_{ig},p_g;w_g,\tilde{d}_{-ig},\epsilon_{ig},\theta\right)= &\quad \bar{a}_{ig}+\kappa+\beta x_{ig}+\delta z_g+\zeta y_{ig}+\lambda x_{ig}y_{ig} \\ &+\xi p_g+\gamma\overline{\mathbb{P}}_{-ig}\left(w_g,\tilde{d}_{-ig},\theta\right)+\epsilon_{ig}(1) \\ \tilde{V}_{ig}^{0}\left(y_{ig},p_g;w_g,\tilde{d}_{-ig},\epsilon_{ig},\theta\right)= &\quad \bar{a}_{ig}+\epsilon_{ig}(0) \end{aligned} \tag{11}$$

Note that $\tilde{V}_{ig}^{d}$ is no longer a function of $\bar{d}_{-ig}$. Instead, it is solely a function of the information set of household $i$ and the vector of decision rules $\tilde{d}_{-ig}$. The formulas for $\tilde{V}_{ig}^{1}$ and $\tilde{V}_{ig}^{0}$ in Eq. (11) illustrate how $\tilde{V}_{ig}^{1}-\tilde{V}_{ig}^{0}$ is independent of $\bar{a}_{ig}$. Thus, it can be written as a function of observables and the expected average adoption level $\overline{\mathbb{P}}_{-ig}\left(w_g,\tilde{d}_{-ig},\theta\right)$.

***Optimal choice.*** A household makes a choice so as to maximize its expected utility, which depends on its taste shock $\epsilon_{ig}$ and on the expectation of the choices made by other households. Using the choice-specific expected indirect utilities defined in Eq. (11) household $i$'s indirect utility is:

$$\begin{aligned} &V_{ig}^{*}\left(y_{ig},p_g;w_g,\tilde{d}_{-ig},\epsilon_{ig},\theta\right) \\ &=\max\left\{\tilde{V}_{ig}^{1}\left(y_{ig},p_g;w_g,\tilde{d}_{-ig},\epsilon_{ig},\theta\right),\tilde{V}_{ig}^{0}\left(y_{ig},p_g;w_g,\tilde{d}_{-ig},\epsilon_{ig},\theta\right)\right\} \end{aligned} \tag{12}$$

Lastly, the household's optimal choice given information set $\left(w_g,\epsilon_{ig}\right)$ and decision rules $\tilde{d}_{-ig}$ — that is, the *best response function* of household $i$ — can be expressed in the familiar form:

$$d_{ig}^{BR}\left(w_g,\tilde{d}_{-ig},\epsilon_{ig},\theta\right)=\begin{cases}1 & if \quad \tilde{V}_{ig}^{1}\geq\tilde{V}_{ig}^{0} \\ 0 & otherwise\end{cases} \tag{13}$$

which resembles that of a standard discrete-choice random utility model, except for the inclusion of the decision rules of all other households $\tilde{d}_{-ig}$.

[4] The implicit assumption is that all relevant village-level effects are captured by variation in village-level observables included in the vector $z_g$, and there are no residual unobserved group-level effects.

### 2.4. Equilibrium

A Bayesian–Nash equilibrium (BNE) in this game is a vector of decision rules $\tilde{d}_g^* = \left(\tilde{d}_{1g}^*, \tilde{d}_{2g}^*, \ldots, \tilde{d}_{I_g g}^*\right)$ in each village $g$ that satisfies:

$$\tilde{d}_{ig}^*\left(w_g, \epsilon_{ig}, \theta\right) = d_{ig}^{BR}\left(w_g, \tilde{d}_{-ig}^*, \epsilon_{ig}, \theta\right) \ \forall i = 1, 2, \ldots, I_g \tag{14}$$

That is, the adoption choices prescribed by the decision rule of each household are *best responses* to other households' decision rules. Following Aguirregabiria and Mira (2019), we can represent a BNE in the space of choice probabilities. This representation is convenient for the econometric analysis of this model. Substituting the equilibrium condition of Eq. (14) into the definition of choice probabilities in Eq. (5), taking into account the form of the expected payoffs, I can characterize a BNE as a collection of conditional choice probabilities $\left\{\mathbb{P}_{ig}^*\left(d_{ig} = 1 \mid w_g, \theta\right)\right\}_{i=1}^{I_g}$, where each element $\mathbb{P}_{ig}^*\left(d_{ig} = 1 \mid w_g, \theta\right) \equiv \mathbb{P}_{ig}\left(d_{ig} = 1 \mid w_g, \tilde{d}_{ig}^*, \theta\right)$, that solves the following system of equations:

$$\mathbb{P}_{ig}^*\left(d_{ig} = 1 \mid w_g, \theta\right) = \frac{\exp\left(\kappa + \beta x_{ig} + \delta z_g + \zeta y_{ig} + \lambda x_{ig} y_{ig} + \xi p_g + \gamma \overline{\mathbb{P}}_{-ig}^*\left(w_g, \theta\right)\right)}{1 + \exp\left(\kappa + \beta x_{ig} + \delta z_g + \zeta y_{ig} + \lambda x_{ig} y_{ig} + \xi p_g + \gamma \overline{\mathbb{P}}_{-ig}^*\left(w_g, \theta\right)\right)} \quad \forall i = 1, 2, \ldots, I_g \tag{15}$$

where $\overline{\mathbb{P}}_{-ig}^*\left(w_g, \theta\right) = \frac{1}{I_g - 1} \sum_{j \neq i} \mathbb{P}_{jg}^*\left(d_{jg} = 1 \mid w_g, \theta\right)$ is the equilibrium expected average adoption level of households other than $i$. Thus, $\overline{\mathbb{P}}_{-ig}^*\left(w_g, \theta\right)$ also represents the equilibrium belief of household $i$ regarding the expected average level of adoption by other households in the village, which marginalizes household $i$'s uncertainty about the choices of other households to compute the expected return from choosing $d_{ig} = 1$ given available information $w_g$.

For a fixed $w_g$, the definition of equilibrium given above implies $I_g$ equilibrium probabilities $\mathbb{P}_{ig}^*\left(d_{ig} = 1 \mid w_g, \theta\right)$ per village. That is, Formula (15) represents a system of $I_g$ equations for each village $g = 1, \ldots, G$ that can be solved to determine equilibrium probabilities of adoption. Note that the system of equations in Formula (15) can be written as fixed point mapping in vector form. In a closely related model, Aguirregabiria and Mira (2019) use such a mapping to prove the existence of a Bayesian Nash equilibrium for discrete choice under noncooperative decision-making. As the empirical content of the model is closely related to their baseline formulation, I do not include a formal proof of existence. Instead, I take existence as given and provide a heuristic argument on the source of multiplicity in the model.

***Multiplicity of equilibria.*** Multiple equilibria are possible in this model. For a given set of parameters, there could be more than one solution for the system of Equations in (15). Multiplicity in the structure arises as a result of interdependence in household choice. Intuitively, the stronger the interdependence, as captured by the social interaction parameter $\gamma$ relative to the private utility component, the more likely it is to have multiplicity. These multiple solutions imply the existence of distinct expected average choice levels that are each compatible with individual optimal decisions. While the possibility of multiple equilibria may be theoretically an attractive feature in capturing observed household behavior, it generates profound challenges in structurally identifying and estimating the model, which is discussed in the next section.

### 2.5. Identification

In this section, I briefly discuss the identification of the model. With the possibility of multiple equilibria, the model is empirically incomplete without the specification of an equilibrium selection mechanism. This "incompleteness" makes it difficult to construct a proper likelihood and objective function, which has implications for the estimation of the model. One way of dealing with the multiplicity would be to develop a theory of the underlying equilibrium selection mechanism and thus complete the model for the purpose of identification.[5] An alternative approach abstracts away from imposing an equilibrium selection mechanism and instead makes an assumption about the observed data. This approach is an empirically attractive alternative when the implied set of equilibrium selection mechanisms is large, and observed data do not rule out certain behavior. To obtain consistent parameter estimates, I assume that in each village, the data observed are generated from only one of the possible equilibria. The restriction is formalized in assumption [A3] below:

**[A3]** : *Given a value of primitives of the model $Y = (w_g, \theta)$ households in a group $g$ select only one equilibrium from the set of possible equilibria, and they do not switch to other equilibria as long as $Y$ does not change.*

This assumption, referred to as the *single-equilibrium-in-data*, has been frequently used in the context of estimating incomplete information games and is independent of the choice of estimation method.[6] The assumption is also less restrictive than explicitly assigning ex-ante an equilibrium selection mechanism the village might be at. Under assumption [A3], I can take the equilibrium choice probabilities $\mathbb{P}_{ig}^*\left(d_{ig} \mid w_g, \theta\right)$ as given. Then, the question of identification is whether it is possible to reverse engineer the structural parameters $\theta = [\kappa, \beta, \delta, \lambda, \zeta, \xi, \gamma]$. Using Eq. (15) to construct an odds ratio $\frac{\mathbb{P}_{ig}^*}{(1-\mathbb{P}_{ig}^*)}$ and performing the Hotz and Miller (1993) inversion by taking logs of the equation yields the following familiar result:

$$\tilde{v}_{ig}^{1*} - \tilde{v}_{ig}^{0*} = \ln\left[\mathbb{P}_{ig}^*\left(d_{ig} = 1 \mid w_g, \theta\right)\right] - \ln\left[1 - \mathbb{P}_{ig}^*\left(d_{ig} = 1 \mid w_g, \theta\right)\right] \tag{16}$$

where $\tilde{v}_{ig}^{d*} \equiv \tilde{v}_{ig}^{d}\left(y_{ig}, p_g; w_g, \tilde{d}_{-ig}^*, \theta\right)$. This equation demonstrates that it is possible to obtain the expected choice-specific value function for any $w_g$ from the empirical counterparts of the equilibrium choice probabilities observed in the data. However, Eq. (16) also illustrates

**Table 1**
Descriptives: Household characteristics.

| | Mean | SD |
|---|---|---|
| Household has sanitation | 0.378 | (0.48) |
| *Household head* | | |
| Age (in years) | 42.802 | (12.881) |
| Upper caste | 0.134 | (0.341) |
| Education (in years) | 4.607 | (4.729) |
| *Household demographics* | | |
| Nr. of adult women | 1.641 | (0.805) |
| Nr. of children (<14 years) | 1.987 | (1.471) |
| *Cash-on-hand* | | |
| Household income (x Rs. 1000) | 76.539 | (96.421) |
| Liquid assets (x Rs. 1000) | 4.504 | (5.073) |
| Observations (Households) | 1467 | |

*Note.* This table presents descriptive statistics for household characteristics and demographic composition for the FINISH household survey. Standard deviation is included in parentheses. Household head characteristics include age, upper caste (including general and forward caste membership), and education in years. The household has a sanitation facility at home (==1). Household demographics include the number of adult women, the number of children under the age of 14. Cash-on-hand measure of total household income (per Rs. 1000) and value of liquid assets and savings (per Rs. 1000). Indian Rs. 1000 ≈ GBP 10 or USD 13.

[5] For example, one can make an assumption about the equilibrium played or, more formally, model (i.e., parametrically) an equilibrium selection mechanism such as implemented by Bajari et al. (2010). The use of an appropriate equilibrium selection rule assures the existence of a well-defined likelihood function over the entire space of observable outcomes. However, a key limitation of this approach is that the consistency of the estimation depends critically on the validity of the assumed selection rule, which is not always testable.

[6] See, for example, Aguirregabiria and Mira (2007), Bajari et al. (2007), Pakes et al. (2007) and Pesendorfer and Schmidt-Dengler (2008).

a potential issue with identification, that, a single equation with two unknown objects ($\tilde{v}_{ig}^{1*}$ and $\tilde{v}_{ig}^{0*}$) on the LHS leaving only the difference $\tilde{v}_{ig}^{1*} - \tilde{v}_{ig}^{0*}$ identified. To proceed, the expected value from non-adoption is normalized to zero, such that $\tilde{v}_{ig}^{1*} - \tilde{v}_{ig}^{0*} = \tilde{v}_{ig}^{1*}$. Note that the normalization $\tilde{v}_{ig}^{0*} = 0$ does not imply that the deterministic part of the indirect utility of non-adoption $v_{ig}^{0} \equiv v_{ig}^{0}\left(y_{ig}, p_g; x_{ig}, z_g, \bar{d}_{-ig}, \theta\right)$ in Eq. (8) must also be equal to zero for each household. In fact, $v_{ig}^{0}$ is allowed to vary with $\bar{d}_{-ig}$, but its expectation is normalized to zero. Moreover, the normalization does not prevent the pure public benefits from the average level of adoption in the village, captured by the function $A_{ig}\left(\bar{d}_{-ig}\right)$, from varying with $\bar{d}_{-ig}$, as well as being heterogeneous across households and villages. For instance, conditional on both households not adopting sanitation, a household with several children may enjoy larger marginal public benefits from an increase in the average level of adoption in the village than one featuring no children, as indicated by Hammer and Spears (2016).

Eq. (16) establishes a link between the equilibrium choice probabilities and the expected choice-specific value functions. The latter can be connected with the model primitives using Eq. (9) to derive a system of equations:

$$\underbrace{\tilde{v}_{ig}^{1*} - \tilde{v}_{ig}^{0*}}_{known} = \sum_{d_{-ig}} \left\{ \left[ v_{ig}^{1}\left(y_{ig}, p_g; x_{ig}, z_g, \bar{d}_{-ig}, \theta\right) - v_{ig}^{0}\left(y_{ig}, p_g; x_{ig}, z_g, \bar{d}_{-ig}, \theta\right) \right] \cdot \underbrace{\mathbb{P}_{-ig}^{*}\left(d_{-ig} \mid w_g, \theta\right)}_{D} \right\} \; for \; i = 1, \ldots, I_g \tag{17}$$

For each equation in (17) the difference $(\tilde{v}_{ig}^{1*} - \tilde{v}_{ig}^{0*})$ is a known object, and the equilibrium choice probabilities $\mathbb{P}_{-ig}^{*}$ denoted by term D have an empirical counterpart observed in the data. Identifying the model requires finding a unique set of primitives $v_{ig}^{1}$ and $v_{ig}^{0}$ that solves this system of equations. For a fixed $w_g$, there are $I_g \times I_g \times 2$ unknowns but only $I_g$ equations. This implies that without additional restrictions, the structural parameters of the model are not identified. However, by restricting our focus to the difference $v_{ig}^{1} - v_{ig}^{0}$ the number of unknowns reduces to $I_g \times I_g$.[7] To proceed further with the model identification, I make use of exclusion restrictions that are discussed in Section 3.

## 3. Empirical implementation

### 3.1. Data

I use data from India for the empirical analysis. In particular, I use micro-data from the FINISH household survey to analyze sanitation adoption behavior. The survey was conducted between 2009–2010 in the district of Gwalior, located within the state of Madhya Pradesh in India.[8] For village and community level characteristics, I use supplementary data from the Indian Census, District Census Handbook 2011 Village Amenities, and Town release. Lastly, I collected data on the purchase cost of sanitation shortly after the first round of the household survey was completed.

***Household characteristics.*** For analyzing household behavior, I use data from Round 1 of the FINISH household survey. The survey was conducted among households in neighboring villages and slum communities around the city of Gwalior. As the FINISH household sample comprises data from rural and semi-urban communities, I refer to them as village communities in the data description. The estimation sample consists of 1467 households across 44 village communities. Table 1 provides descriptive statistics for household characteristics and determinants of household sanitation choice. As seen from Table 1, less than 40% of the households have a sanitation facility at home, most often the *twin-pit pour flush* (TPPF) system. Over 80% of households with a sanitation facility indicated having a pit pour-flush system. TPPF is the standard and most popular sanitation design unit implemented by the Indian Government under the Total Sanitation Campaign (TSC). The twin-pit technology is reasonably versatile with a small land footprint, access to a piped sewage system is not required, and the amount of space needed to build the facility is minimal.[9]

Table 1 also includes household head characteristics and demographic composition. The mean age of the household head is approximately 43 years, with just below primary school education. Approximately 44% of household heads have no education. In addition, 13% of household heads identify themselves as upper caste, which includes Brahmins and other forward caste groups. Household composition variables include the number of women and children within the household. On average, households have at least one adult woman (over the age of 18) and two children under 14. In households with one adult woman, the woman's marital status is almost always married or widowed, indicating the presence of a spouse or elderly parent of the household head. To include women, I construct a categorical variable that denotes the presence of one, two, three, or four (or more) adult women within the household. Approximately 23% of households in the sample do not have a child under 14. To incorporate children, I include three categories for no (zero), one, or two (or more) children under the age of 14, respectively. Lastly, the mean annual household income is approximately Rs. 76,500 (approx. USD 995). To capture a household's effective wealth, I construct a cash-on-hand measure that includes total household income and the value of liquid assets, including household savings.

***Village and community controls.*** Information on the village and community-level characteristics were collated from the 2011 Indian Census. The district census handbook for the district of Gwalior, including the town release, provides rural and urban census denominations for each village community. I use sub-district and village code data provided by the Municipal Corporation of Gwalior to merge the sample of village communities with the census data. Based on the census classification, the sample of 44 villages is 47% rural and 53% semi-urban or urban. The census village amenities and town information provides supplemental data on the FINISH survey's village community information.

Table 2 presents descriptive statistics for the village-level characteristics of the 44 communities observed in the data. Approximately 43% of the village communities have a drainage infrastructure within the village. In contrast, only 29% of the communities have a public sanitation facility. A public facility's presence may substitute for a private sanitation facility at home. Close to 36% of the communities have access to a bank or post office within 1 km. Government post offices in many parts of rural India offer basic banking services like savings accounts, cashier checks, and fixed-rate deposits. Proximity to a bank may ease a household's constraints to save, consequently increasing the likelihood of purchasing sanitation. Like household characteristics, the presence of public sanitation, drainage infrastructure, or a bank within the community affects a household's net utility from adoption. Unlike household variables, village characteristics drive the part of private benefits from sanitation that is common to all households in that village. By including village observables, the model can account for

[7] Note that in a model where the identity of the household matters for the adoption decision the degree of under identification would be greater $I_g \times 2^{I_g - 1}$.

[8] Further details on the FINISH household survey are described in Augsburg and Rodriguez-Lesmes (2018).

[9] *WHO Sanitation System Fact Sheet.* The minimum recommended distance between the sanitation facility and any other structural foundation is 1 m. In regions with high groundwater pollution risk or where pits are located in areas with a high or variable water table and/or fissures or cracks in the bedrock, the recommended minimum horizontal distance is 30 m between the sanitation facility and the water source to limit exposure to microbial contamination.

**Table 2**
Descriptives: Village characteristics.

| | Mean | SD |
|---|---|---|
| Drainage infrastructure in village | 0.428 | (0.490) |
| Public sanitation facility | 0.293 | (0.421) |
| Bank or post office (within 1 km) | 0.356 | (0.463) |
| Market price of sanitation (x Rs. 1000) | | |
| Labor costs | 0.368 | (0.308) |
| Material costs | 8.259 | (1.561) |
| Sanitation prevalence in village | 0.378 | (0.304) |
| Observations (village communities) | 44 | |

*Note.* This table presents descriptive statistics for village community-level characteristics. Standard deviation is included in parentheses. The statistics are computed using calculated using the District Census Handbook (DCHB) Village Amenities for the district of Gwalior, DCHB Town release, and village information variables include in the FINISH household survey. The market purchase cost of sanitation is shown in Indian rupees (per Rs. 1000). Indian Rs. 1000 ≈ GBP 10 or USD 13.

differences in the net utility from adoption across villages and for group effects that drive household adoption other than social interactions.

***Market price of sanitation.*** Lastly, the village-level purchase cost comprises labor and material costs to build a sanitation facility. Data on price estimates were collected separately and shortly after round 1 of the household survey was completed.

$$p_g = F(r, qty) + L(w, days) \tag{18}$$

The labor costs include the daily wage rate ($w$) and approximate time to construct a TPPF sanitation unit, which is, on average, between 3 and 4 days. The daily wage rate also varies across villages, with a mean of $Rs.$ 168 (USD 2.2) and standard deviation $Rs.$ 125 (USD 1.6). Material costs, denoted by $F(.)$ comprise between 80%–95% of the total price of sanitation. Material cost measures include quantity ($qty$) and unit cost ($r$) for the five main raw materials used to construct the sanitation unit — bricks, cement mortar, tiles, ceramic fixtures, and tin sheets. Table 2 decomposes the total cost into labor and raw material components. The price estimate has a mean of $Rs.$ 8628 (USD 112) and a standard deviation of $Rs.$ 1549 (USD 20) with substantial variation across village groups. On average, sanitation price is almost twice the average value of household assets/savings and approximately 12% of the annual household income. These numbers provide descriptive evidence of the sizable expenditure households encounter when making their adoption decision. Raw materials used in the construction of sanitation are widely produced and demanded in the region for other industrial and domestic construction. The overall demand for these raw materials to build a sanitation facility constitutes a small proportion of the overall demand in the region. The primary source of variation in the material costs across villages is inversely related to distance to the center of Gwalior city, reflecting transportation costs.

### 3.2. Identification of preference parameters

Following the identification discussion in Section 2, exclusion restrictions are imposed to identify the model primitives. These restrictions are embedded in the partition of the state space of a household $\left(y_{ig}, x_{ig}, p_g, z_g\right)$ relative to its information set $\left(w_g, \epsilon_{ig}\right)$ where $w_g = \left(y_g, x_g, p_g, z_g\right)$. This can be seen in Eq. (17) where the vector $x_{-ig} = \left(x_{1g}, \ldots, x_{(i-1)g}, x_{(i+1)g}, \ldots, x_{I_g g}\right)$ enters the beliefs $\mathbb{P}_{-ig}\left(d_{-ig} \mid w_g, \theta\right)$ but is excluded from the difference $v^1_{ig} - v^0_{ig}$. By holding $x_{ig}$ fixed, and varying $x_{-ig}$ it is possible to increase the number of equations that $v^1_{ig} - v^0_{ig}$ must satisfy. If there are at least $I_g$ points in the support of the conditional distribution of $x_{-ig}$ given $x_{ig}$, it is possible to increase the number of equations relative to the model parameters. In principle, one could use any of the elements of $x_{-ig}$ to generate additional equations, such as the number of children or adult females. However, some of these variables may not have enough points in their support to satisfy the required conditions. This is not a concern in the application considered as $x_{-ig}$ includes variables with rich support. To provide intuition, consider the example of a specific household characteristic, such as household head education. Though the education of all households excluding $i$, determines household $i$'s beliefs about the village adoption, it does not affect household $i$'s directly utility from sanitation adoption. The restrictions embedded within the model help identify parameters that capture public gains from sanitation.

Lastly, I discuss what variation observed in the data allows for the identification of the specific parametric form of the direct utility function. The mean sanitation adoption conditional on household and village-level observables can be used to construct data moments, for example, the odds of adoption. These data moments are an empirical counterpart to the model conditional choice probabilities of sanitation adoption $\mathbb{P}_{ig}$. Intuitively, within-village variations in household demographics and income generate data moments that identify differences in the net utilities from sanitation adoption across households within a village. While across-village variations in village observables and prices generate data moments that capture across-village differences in the net utilities from sanitation adoption for a household with the same household-level characteristics living in a different village.

### 3.3. Estimation

The estimation proceeds in two stages using a Hotz and Miller (1993) conditional choice probability (CCP) type estimator in the context of incomplete information games. There are several methods that implement the two-step approach; the underlying intuition behind the estimation steps is the same and follows from the identification discussion. In the first stage, the conditional choice probability (CCP) estimates for each household defined in Eq. (15) are recovered directly from the observed data. This step is followed by a second stage, which recovers the structural parameters of interest using either a maximum likelihood or a method of moments estimator. I implement a two-step pseudo maximum likelihood estimator in the spirit of Bajari et al. (2010), and Aguirregabiria and Mira (2007) to estimate the model.

***First stage.*** In general first stage estimation requires obtaining consistent estimates of the CCPs on the left-hand side of Eq. (15). In practice, however, without a large amount of data, nonparametric estimation methods can be subject to a severe curse of dimensionality, especially when the dimension of state variables is large. The simple structure of the model, specifically how strategic interactions are incorporated through the expected beliefs, simplifies the first stage substantially. Instead of implementing a semi-parametric estimator to compute the CCP estimates $\mathbb{P}_{ig}$, I proceed by obtaining a consistent estimate of the beliefs $\overline{\mathbb{P}}_{-ig}$ within each village $g$ using a simple frequency estimator. For a household choosing whether or not to adopt, the belief will be close to the equilibrium level of sanitation in the village. A key limitation of this approach arises when the observed sanitation adoption rate is measured with error. Measurement error causes the distribution generating data to differ from the true distribution. To address this issue, I allow for measurement error in the estimate of latent beliefs as observed by the researcher.[10] I implement a measurement error correction approach from Chesher (1991) and Chesher et al. (1985) to quantify the effect of measurement error on the parameter estimates. The next paragraph describes the small error variance approximation approach from Chesher (1991) and summarizes the steps needed to implement the correction method in the model.

[10] In a different context, Todd and Wolpin (2018) incorporate measurement error in the latent variables that determine effort decisions to estimate a complete information game of effort coordination between teacher and student.

$$\mathbb{P}_{ig}^{corr}\left(d_{ig}=1 \mid w_g,\theta\right)=\frac{\exp\left(k+\beta x_{ig}+\delta z_g+\zeta y_{ig}+\lambda x_{ig}y_{ig}+\xi p_g+\gamma\overline{d}_{-ig}+\overbrace{\gamma(\rho\sigma_g^2)G_{\overline{d}}^{(1)}(\overline{d}_g)}^{C1}+\overbrace{\frac{1}{2}\gamma^2(\rho\sigma_g^2)H(w_{ig}\hat{\theta}_1)}^{C2}\right)}{1+\exp\left(k+\beta x_{ig}+\delta z_g+\zeta y_{ig}+\lambda x_{ig}y_{ig}+\xi p_g+\gamma\overline{d}_{-ig}+\underbrace{\gamma(\rho\sigma_g^2)G_{\overline{d}}^{(1)}(\overline{d}_g)}_{C1}+\underbrace{\frac{1}{2}\gamma^2(\rho\sigma_g^2)H(w_{ig}\hat{\theta}_1)}_{C2}\right)} \tag{20}$$

**Box I.**

***Measurement error correction.*** The observed level of adoption in a village is an equilibrium outcome of the underlying strategic interactions among households. Thus, each household's (probabilistic) best response adoption function is fully characterized by its beliefs about the mean adoption in the village, given household and village-level observable characteristics. In what follows, I treat the belief $\overline{\mathbb{P}}_{-ig}$ as a latent variable measured with error.

$$\overline{d}_g=\overline{\mathbb{P}}_{-ig}+\sigma_g u_{ig} \tag{19}$$

where the researcher does not observe $\overline{\mathbb{P}}_{-ig}$ but instead observes realizations $\overline{d}_g$ contaminated with measurement error. The variable $u_{ig}$ is assumed to be continuously distributed and independent of both the decision to adopt and the beliefs $\overline{\mathbb{P}}_{-ig}$. It possesses mean zero and variance one, while $\sigma_g$ captures across-village variations, and correlations $corr(u_g, u_{g'})=\rho$. Eq. (19) illustrates how the presence of measurement error $\sigma_g u_{ig}$ causes the true data generating distribution to differ from the observed data distribution. The small error variance approximation method treats the observed density as a distorted version of the true density of interest. Intuitively, the method constructs the distorted density by taking a Taylor series approximation of the conditional density around the point of no measurement error. Chesher (1991) shows how the mean regression function of interest can be derived using the distorted density, which explicitly accounts for measurement error through specific correction parameters. Under the type 1 extreme value assumption and the direct application of the mean regression function in Chesher (1991), the approximation under the logit specification can be used to define a corrected CCP for each household $i$, whose formula is shown in Eq. (20) which is given in Box I. where $H(w_{ig}\hat{\theta}_1)=1-2\frac{e^{w_{ig}\hat{\theta}_1}}{\left(1+e^{w_{ig}\hat{\theta}_1}\right)}$. The correction terms $C1$ and $C2$ capture the first- and second-order effects of measurement error. Together they illustrate the effect of measurement error on the CCPs and the structural parameter estimates.

The first term $C1$ generates the overall attenuation effect of measurement error. It includes the strategic interaction parameter $\gamma$ and $\rho$, which accounts for differences in measurement error across villages $\sigma_g^2$. The behavior of $C1$ also depends on the first derivative of the logarithm of the density of the error-free covariate $\overline{\mathbb{P}}_{-ig}$, which is approximated by $G_{\overline{d}}^{(1)}(\overline{d}_g)$. Appendix A.2 provides a detailed description of the construction of the variable $G_{\overline{d}}^{(1)}(\overline{d}_g)$ and conditioning on $\overline{d}$. Consequently, the first-order effect of measurement error is to raise the density of the error-free covariate where it is convex and to depress it where it is concave. The second term $C2$ corrects the curvature of the nonlinear regression as conditioning moves from the error-free $\overline{\mathbb{P}}_{-ig}$ to error contaminated $\overline{d}_g$. The term accounts for the additional nonlinear effect in nonlinear regression models, raising the error contaminated regression function where the error free regression function is convex and lowering it where it is concave. The behavior of $C2$ depends on the probability of sanitation adoption for a given household defined by $H(w_{ig}\hat{\theta}_1)$ constructed using $\hat{\theta}_1$ which are a vector of parameter estimates from a logit regression on the error contaminated data, that is, with no measurement error correction. Note that in the case of a linear model, $C2$ vanishes and, unlike $C1$, does not depend upon the distribution of the error-free covariate $\overline{\mathbb{P}}_{-ig}$. The approximation method establishes a direct link between the regression functions of interest and the underlying latent beliefs, characterizing the distortion induced by measurement error on the true data generating process. Correction terms $C1$ and $C2$ in Eq. (20) can be used to quantify and correct for the distortive effect of measurement error on the second-stage parameter estimates.

***Second stage.*** The CCPs $\mathbb{P}_{ig}^{corr}$ in Eq. (20) depend on the structural parameters, and can be used to construct a log-likelihood function.

$$\mathcal{L}(\theta|w_g,d_{ig})=\frac{1}{G\cdot I_g}\sum_{g=1}^{G}\sum_{i=1}^{I_g}\Big(d_{ig}\ln\mathbb{P}_{ig}^{corr}\left(d_{ig}=1\mid w_g,\theta\right) \\ +(1-d_{ig})\ln\left(1-\mathbb{P}_{ig}^{corr}\left(d_{ig}=1\mid w_g,\theta\right)\right)\Big) \tag{21}$$

The second stage estimates – denoted by $\hat{\theta}_2$ – are obtained by using a maximum likelihood estimator with the modified log-likelihood function defined in Eq. (21).[11] The vector of parameter estimates obtained from the second stage include $\hat{\theta}_2=\left[\hat{k},\hat{\beta},\hat{\delta},\hat{\zeta},\hat{\lambda},\hat{\xi},\hat{\gamma},\widehat{\gamma\rho},\widehat{\gamma^2\rho}\right]$. The first seven parameter estimates are the parameters of interest followed by the correction terms that control the attenuation and the degree of curvature of the probabilistic best response function. A clustered bootstrap procedure is employed to construct standard errors.

Before discussing results, I return to the discussion in Section 2 on unobserved heterogeneity. Suppose a large panel with a time dimension for each village was available. In that case, the estimation procedure described above could be implemented village-by-village to allow for a substantial amount of unobserved heterogeneity. This approach would relax the assumption that village-specific unobservables are a smooth function of observed state variables. Despite widespread use in many empirical applications, the assumption does impose strong restrictions on the choice-specific utilities. Moreover, such restrictions are unlikely to hold in many other important applications. In those cases, a more general approach to coping with unobserved heterogeneity may be needed, as developed in Aguirregabiria and Mira (2007).

## 4. Estimation results

### 4.1. Model fit

I begin by discussing the fit of the model. To assess if the estimated model captures the observed features in the data, I compare the observed sanitation choice distribution at the village level with those predicted by the model. Fig. 1 shows the model fit of sanitation

[11] Chesher et al. (1985) describe such a procedure in the context of correcting for the effects of random parameter variation.

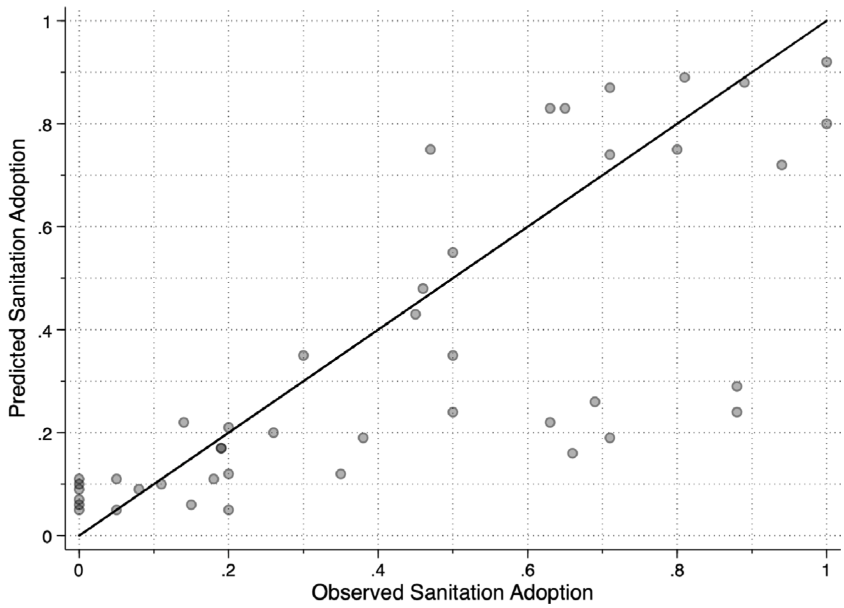



**Fig. 1.** Model fit across villages. *Note.* Each dot represents a particular village. Both model predicted and observed sanitation adoption is represented as a fraction. The solid line corresponds to the 45-degree line. Adjusted $R^2 = 0.819$ computed using all 44 villages in the sample. Correlation coefficient $\rho = 0.768$.

adoption patterns across villages. I use the model estimates for each village to compute a predicted sanitation adoption level and compare it with the prevalence of sanitation observed in the data. The data moments are plotted on the horizontal axis, while the model-predicted sanitation adoption level for each village is plotted on the vertical axis. The solid diagonal line corresponds to the 45° line. The correlation between the observed and predicted sanitation adoption is equal to 0.768. With a few exceptions, the model fit is close to the observed data.

While it is reassuring that the model fits the target data moments in the estimation across villages, Table A.1 in Appendix A presents results from an in-sample validation exercise. For each village, the model predicted equilibrium sanitation level is compared to a regression of observed sanitation adoption level in the data for each village. The first column of Table A.1 shows the regression coefficient for the sanitation level using observed data. The column based on actual data also reports the 95% confidence interval and the $p$-value. The next two columns show the model-predicted sanitation adoption level and the percent difference between the model-predicted equilibrium from the observed data. Overall, the model estimates are precisely estimated for most villages and lie within the 95% confidence interval of the observed data.

Furthermore, the quasi-nested structure of the estimation procedure allows me to incorporate a likelihood ratio (LR) test where the restricted version of the model with no measurement error is tested against the unrestricted version accounting for measurement error. The null hypothesis here essentially tests that the "smaller" is the true model against an unrestricted model that includes additional correction terms. A larger test statistic would indicate that the null hypothesis is false. The LR test statistic is 12.58 distribution chi-squared with two degrees of freedom with a $p$-value of 0.0019.[12] The test statistic indicates that the unrestricted model fits significantly better than the model without measurement error correction. Sweeting (2006) incorporates a similar LR test to assess the strength of strategic complementarities. While my model and log-likelihood function are different, the underlying intuition behind the LR test is supported.

***Impact of measurement error.*** To assess the impact of measurement error on parameter estimates, Table 3 presents results from two estimation procedures. For ease of reference, only village-level parameters directly affected by measurement error are shown. This includes the parameter on the price variable ($\xi$), spillover effect ($\gamma$), and village-level controls ($\delta s$). Column (A) presents estimates from an estimation procedure that does not account for measurement error. While column (B) presents estimates from the measurement error correction estimation method described in Section 3.3.

Overall, accounting for measurement error in the estimation increases the magnitude of the village-level parameter estimates. Excluding the estimate for drainage infrastructure, these differences are not found to be significant. Similarly, the absolute value of the price estimate increase in column (B) but is not significantly different. In contrast, the parameter capturing the externality effect $\gamma$ is significantly larger with measurement error correction. Not accounting for measurement error would have resulted in a significant downward bias on the parameter estimate capturing the externality. In other words, the size of the spillovers would be underestimated. The impact of measurement error is captured by the estimates from the two correction terms $C1$ and $C2$, which account for the attenuation effect and degree of curvature of the probability response curve, respectively.

### 4.2. Structural parameter estimates

The parameter estimates shown in Table 4 correspond to estimates of the household demand for sanitation or "take-up" response in equilibrium and can be used to compute demand elasticities at the household level, which is discussed in the next section. However, the parameter estimates in front of income and price, denoted by $\zeta$ and $\xi$ along with the interaction terms $\lambda^{1Ch}$ and $\lambda^{2Ch}$ relate directly to the primitives of the utility function discussed in Section 2. In particular, the estimate of $\xi$ reflects the negative of the marginal utility of consumption, while the extent to which this marginal utility varies with a household's own adoption of sanitation is captured by $\zeta + \lambda^{\#Ch}$, where # denotes the number of children in the household and $\lambda^{0Ch} = 0$. The estimates for the parameters $\zeta$, $\lambda^{1Ch}$ and $\lambda^{2Ch}$ in Table 4 suggest the existence of a positive interaction between the private benefits from sanitation and other types of consumption in the household's utility

[12] The chi-square test statistic with two degrees of freedom is 5.99 at 0.05 critical value.

**Table 3**
Impact of measurement error.

| | Parameter | (A) No correction | | (B) With correction | | |
|---|---|---|---|---|---|---|
| | | (1) Value | (2) Std Err | (3) Value | (4) Std Err | (1)=(3) $p$-value |
| *Village parameters* | | | | | | |
| Drainage infrastructure | $\delta^{dr}$ | 0.256 | (0.113)** | 0.437 | (0.169)** | 0.058 |
| Public sanitation | $\delta^{pu}$ | 0.033 | (0.111) | −0.064 | (0.134) | 0.180 |
| Bank or Post office | $\delta^{ba}$ | 0.031 | (0.016)* | 0.035 | (0.021)** | 0.322 |
| Price of sanitation (per Rs. 1000) | $\xi$ | −0.068 | (0.038)* | −0.084 | (0.045)* | 0.279 |
| Average sanitation in village (excl. $i$) | $\gamma$ | 4.626 | (0.273)*** | 4.899 | (0.313)*** | 0.007 |
| *Measurement error correction* | | | | | | |
| Correction term 1 | $C1$ | | | −1.794 | (0.905)** | |
| Correction term 2 | $C2$ | | | 15.243 | (4.093)*** | |

*Note.* The table presents village level parameter estimates. Standard errors are included in parentheses. Column panel (A) presents parameter estimates from an estimation procedure that does not account for measurement error. Column panel (B) presents estimates of the measurement error correction estimation procedure. Both specifications include household characteristics and income. Household head characteristics include age (in years), upper caste (including general and forward caste membership), and education (in years). Household composition characteristics include the number of adult married women, the number of children under the age of 14, and a cash in hand measure of total income. Price denotes village-level purchase cost of sanitation in Indian rupees. Price and household income variables are per Rs. 1000. Average village sanitation denotes a measure of mean sanitation excluding household $i$. Village and community level characteristics include the presence of drainage infrastructure (==1), public sanitation facility (==1), and bank or post office within 1 km (==1) of the village. A clustered bootstrapped procedure is used to construct standard errors in the second stage 5000 reps. Indian Rs. 1000 ≈ GBP 10 or USD 13.

*Denotes signf at 0.10.

**Denotes signf at 0.05.

***Denotes signf at 0.01.

**Table 4**
Structural parameter estimates.

| | Parameter | Value | Std Err |
|---|---|---|---|
| *Panel A. Income, price and externality* | | | |
| Household income | $\zeta$ | 0.024 | (0.011)** |
| | $\lambda^{1Ch}$ | −0.007 | (0.004)* |
| | $\lambda^{2Ch}$ | −0.021 | (0.011)* |
| Price of sanitation | $\xi$ | −0.084 | (0.045)* |
| Average sanitation (excl $i$) | $\gamma$ | 4.899 | (0.313)*** |
| *Panel B. Household characteristics* | | | |
| Age of head | $\beta^{Age}$ | 0.006 | (0.003)* |
| Upper caste | $\beta^{UC}$ | 0.863 | (0.259)*** |
| Education of head | $\beta^{Edu}$ | 0.115 | (0.017)*** |
| Nr. of adult females | | | |
| | $\beta^{2Wo}$ | 0.365 | (0.157)** |
| | $\beta^{3Wo}$ | 0.746 | (0.223)*** |
| | $\beta^{4Wo}$ | 1.653 | (0.411)*** |
| Nr. of children < 14 | | | |
| | $\beta^{1Ch}$ | 0.227 | (0.138)* |
| | $\beta^{2Ch}$ | 0.521 | (0.179)*** |

*Note.* This table presents structural parameter estimates for price, income, strategic interaction and household characteristics with village level controls. Standard errors are in the parentheses. Household head characteristics include age, upper caste membership (includes general and forward caste), and education (in years). Household composition characteristics include the number of adult women (base 1 woman), the number of children under the age of 14 (base 0 children), and cash-in-hand household income. Sanitation prevalence in the village in measured with a leave one out mean. Price and household income variables are per Rs. 1000. Village and community level controls include the presence of drainage infrastructure (==1), public sanitation facility (==1), bank or post office within 1 km (==1). A clustered bootstrapped procedure is used to construct standard errors 5000 reps. Indian Rs. 1000 ≈ GBP 10 or USD 13.

*Denotes signf at 0.10.

**Denotes signf at 0.05.

***Denotes signf at 0.01.

function. A test of $\zeta = 0$ has a $p$-value of 0.011, implying that the interaction is statistically significant.

The model also incorporates heterogeneity in the utility from adoption due to both household- and village-level observable characteristics, which corresponds to the component $B_{ig}(d_{ig})$ of the direct utility function in Eq. (4). The parameter vector $\beta$ captures the household-level heterogeneity. The estimates for all the elements of $\beta$ are statistically significant. The parameters capturing the effect of the number of adult females ($\beta^{2Wo}$, $\beta^{3Wo}$, $\beta^{4Wo}$) and the number of children ($\beta^{1Ch}$, $\beta^{2Ch}$) affect the net utility from sanitation adoption in a positive and significant way. Parameters capturing the effect of the age ($\beta^{Age}$), upper-caste status ($\beta^{UC}$) and education ($\beta^{Edu}$) of the household have a similar impact. Overall, the results suggest that households featuring many women and children and people who are more educated and from upper castes, on average, display a stronger taste for sanitation consumption that translates into higher take-up rates.

The impact of village-level heterogeneity, captured by the parameter vector $\delta$, can be seen in Table 3 column (B). Although negative, the effect of a public sanitation facility on the net utility from sanitation adoption is not significantly different from zero. A public alternative within the village does not seem to significantly reduce the net utility gain from owning a private sanitation facility at home. Lack of significance provides indirect evidence on the extent of free-riding behavior within the village. However, the evidence is limited because the quality and accessibility of public sanitation facilities may be very different from private facilities. In contrast, the presence of drainage infrastructure and a bank/post office has a positive and significant effect on the net utility from adoption for all households in the village. These observable village characteristics account for group effects that drive household correlation in adoption independent from the externality effects. Without these controls, the strategic behavior within the village would be overestimated. Lastly, the social interaction parameter $\gamma$ estimate is positive and statistically significant, providing compelling support for the hypothesis that sanitation adoption generates positive externalities at the village level. Moreover, the magnitude of the social interaction effect on the net utility from sanitation adoption is considerable: a percentage point increase in the adoption rate of the village has the same effect on a household's take-up probability as a 584 rupee price discount, equal to 6.8% of the average price of a sanitation facility in the sample.

While the structural parameter estimates of household preferences are interesting objects per se and provide useful insights into the key drivers of household adoption, a clearer picture of the magnitudes implied by the economic model can be obtained from the income and price demand elasticities. For instance, one must calculate the price elasticity to quantify the effect of a uniform price subsidy on sanitation take-up. However, this is not a straightforward calculation because the effect of a price change is not solely captured by its direct impact on a household's adoption probability. The additional feedback effect of a price change through the externality channel must also be considered. Thus, the overall impact of a price change consists of two components:

1. **Direct Effect:** The primary effect on household demand is characterized by an individual household's isolated response to a price change.
2. **Indirect Effect:** A secondary effect generated by the dependence of a household's adoption choice on the adoption behavior of other households that, in equilibrium, also respond to the price change.

If externalities affect individual household decisions, then the computation of the overall price elasticity of demand must differentiate between these two components. Without this separation, the direct price effect on household demand would be overestimated, with potential consequences for policy design. For instance, in designing an optimal price subsidy, such overestimation may result in an inefficient use of government resources.

I use the equilibrium condition in Eq. (15) to calculate the price elasticity of demand and separate out the direct and indirect effects. For ease of interpretation, some arguments of $\mathbb{P}^*_{ig}\left(d_{ig}=1 \mid w_g, \theta\right)$ are suppressed in this section. Specifically, the equilibrium conditional choice probability of adoption is denoted by $\mathrm{P}_{ig}\left(y_{ig}, p_g, \overline{\mathbb{P}}^*_{-ig}\left(w_g, \theta\right)\right)$, and its village level average is given by $\overline{\mathrm{P}}_g\left(y_g, p_g\right) \equiv \frac{1}{I_g}\sum_{j=1}^{I_g}\mathrm{P}_{jg}$ $\left(y_{jg}, p_g, \overline{\mathbb{P}}^*_{-jg}\left(w_g, \theta\right)\right)$. For a sufficiently large village, the equilibrium expected average probability of adoption for households other than $i$ is approximately equal to $\overline{\mathrm{P}}_g\left(y_g, p_g\right)$:

$$\overline{\mathbb{P}}^*_{-ig}\left(w_g, \theta\right) = \frac{1}{I_g - 1}\sum_{j\neq i}\mathbb{P}^*_{jg}\left(d_{jg}=1 \mid w_g, \theta\right) \simeq \overline{\mathrm{P}}_g\left(y_g, p_g\right) \tag{22}$$

The equilibrium conditional choice probability of adoption $\overline{\mathbb{P}}^*_{-ig}\left(w_g, \theta\right)$ is approximated by $\overline{\mathrm{P}}_g\left(y_g, p_g\right)$ in the elasticity formula. Using this newly defined notation, I can derive the price derivative of $\mathrm{P}_{ig}\left(y_{ig}, p_g, \overline{\mathrm{P}}_g\left(y_g, p_g\right)\right)$, which can be written as:

$$\frac{d\mathrm{P}_{ig}\left(y_{ig}, p_g, \overline{\mathrm{P}}_g\left(y_g, p_g\right)\right)}{dp_g} = \underbrace{\frac{\partial\mathrm{P}_{ig}\left(y_{ig}, p_g, \overline{\mathrm{P}}_g\left(y_g, p_g\right)\right)}{\partial p_g}}_{A} + \underbrace{\frac{\partial\mathrm{P}_{ig}\left(y_{ig}, p_g, \overline{\mathrm{P}}_g\left(y_g, p_g\right)\right)}{\partial\overline{\mathrm{P}}_g}\frac{\partial\overline{\mathrm{P}}_g\left(y_g, p_g\right)}{\partial p_g}}_{B} \tag{23}$$

where $A$ is the direct effect of the price change on $i$'s probability of adoption, and $B$ is the indirect effect due to the increase in the adoption rate in a village. The component $B$ is derived using the equilibrium condition in Eq. (15) under the approximation in Eq. (22). Differentiating the equilibrium condition with respect to $p_g$, I get:

$$\frac{\partial\overline{\mathrm{P}}_g\left(y_g, p_g\right)}{\partial p_g} = \frac{\frac{1}{I_g}\sum_{i=1}^{I_g}\Lambda'\left(\kappa + \beta x_{ig} + \delta z_g + \zeta y_{ig} + \lambda x_{ig}y_{ig} + \xi p_g + \gamma\overline{\mathrm{P}}_g\left(y_g, p_g\right)\right)\xi}{1 - \frac{1}{I_g}\sum_{i=1}^{I_g}\Lambda'\left(\kappa + \beta x_{ig} + \delta z_g + \zeta y_{ig} + \lambda x_{ig}y_{ig} + \xi p_g + \gamma\overline{\mathrm{P}}_g\left(y_g, p_g\right)\right)\gamma} \tag{24}$$

where $\Lambda$ is the cumulative distribution function (CDF) of the logistic distribution. Similarly, I can write the effect of a uniform change in income that affects only some households (e.g., a targeted transfer). To do so, let $t_g$ be the amount of the transfer and $s_g$ be a $1\times I_g$ vector whose elements $s_{ig}$ equals 1 if household $i$ receive the transfer, and zero if it does not. The equilibrium effect of this type of income change on the average level of adoption in the village can be derived by differentiating the equilibrium condition in Eq. (15) and evaluating the derivative of interest at $t_g = 0$, resulting in the formula:

$$\left.\frac{\partial\overline{\mathrm{P}}_g\left(y_g + t_g s_g, p_g\right)}{\partial t_g}\right|_{t_g=0} = \frac{\frac{1}{I_g}\sum_{i=1}^{I_g} s_{ig}\Lambda'\left(\kappa + \beta x_{ig} + \delta z_g + \zeta y_{ig} + \lambda x_{ig}y_{ig} + \xi p_g + \gamma\overline{\mathrm{P}}_g\left(y_g, p_g\right)\right)\left(\zeta + \lambda x_{ig}\right)}{1 - \frac{1}{I_g}\sum_{i=1}^{I_g}\Lambda'\left(\kappa + \beta x_{ig} + \delta z_g + \zeta y_{ig} + \lambda x_{ig}y_{ig} + \xi p_g + \gamma\overline{\mathrm{P}}_g\left(y_g, p_g\right)\right)\gamma} \tag{25}$$

which can be used to derive the direct and indirect effect of income changes in a manner similar to that described in Eq. (23). Using the above formulas and parameter estimates, I can measure the responsiveness of sanitation demand to changes in price and income. These elasticity estimates are discussed next.

**Table 5**
Price and income elasticity for a household.

| Wealth | El. | Price | | |
|---|---|---|---|---|
| | | *10th %tile* | *50th %tile* | *90th %tile* |
| *10th %tile* | **M** | −0.339 | −0.518 | −0.642 |
| | | (0.194) | (0.310) | (0.412) |
| | **I** | 0.010 | 0.011 | 0.011 |
| | | (0.003) | (0.003) | (0.004) |
| *25th %tile* | **M** | −0.337 | −0.516 | −0.639 |
| | | (0.193) | (0.310) | (0.409) |
| | **I** | 0.022 | 0.024 | 0.025 |
| | | (0.006) | (0.007) | (0.008) |
| *50th %tile* | **M** | −0.331 | −0.508 | −0.630 |
| | | (0.188) | (0.308) | (0.402) |
| | **I** | 0.054 | 0.058 | 0.061 |
| | | (0.017) | (0.018) | (0.019) |
| *75th %tile* | **M** | −0.312 | −0.480 | −0.599 |
| | | (0.174) | (0.285) | (0.384) |
| | **I** | 0.152 | 0.165 | 0.173 |
| | | (0.044) | (0.049) | (0.053) |
| *90th %tile* | **M** | −0.257 | −0.407 | −0.514 |
| | | (0.140) | (0.239) | (0.333) |
| | **I** | 0.336 | 0.373 | 0.396 |
| | | (0.073) | (0.087) | (0.098) |

*Note.* This table presents the price and income elasticity estimates for a representative household at different points of the price and income distribution observed in the data. Standard errors included in the parentheses are computed using the delta method. Elasticity estimates are computed under a marginal change in prices and income. **M** denotes the Marshallian (or uncompensated) price elasticity. **I** denotes the income elasticity.

### 4.3. *Elasticity of sanitation adoption*

#### 4.3.1. *Price and income elasticity*

The estimates of the household demand function shown in Table 4, for the most part, are not easy to interpret. For this reason, I use the model estimates to compute sanitation take-up elasticity measures for changes in price and income. Given the heterogeneity embedded within the demand system, the elasticity estimates do vary across households. Table 5 shows the price and income elasticity estimates, computed using term $A$ in Eq. (23), at different points of the household wealth and price distribution observed in the sample.

I first discuss the Marshallian (or uncompensated) price elasticity, which computes changes in the demand for sanitation for a change in the price of sanitation. The price elasticity estimates, denoted by **M**, are negative, indicating a decrease in the probability of sanitation adoption for a small increase in price. At first glance, we see that the price elasticity *increases* with the price level (across the columns) and *decreases* with

**Table 6**
Direct and indirect impact of a price change.

| Wealth | El. | Price | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | $10th\,\%tile$ | | | $50th\,\%tile$ | | | $90th\,\%tile$ | | |
| | | Dir. | Indir. | Tot. | Dir. | Indir. | Tot. | Dir. | Indir. | Tot. |
| $25th\,\%tile$ | **M** | −0.337<br>(0.193) | −0.723<br>(0.058) | −1.060<br>(0.201) | −0.516<br>(0.310) | −1.177<br>(0.052) | −1.693<br>(0.324) | −0.639<br>(0.409) | −1.458<br>(0.106) | −2.097<br>(0.4.24) |
| $75th\,\%tile$ | **M** | −0.312<br>(0.174) | −0.486<br>(0.041) | −0.798<br>(0.179) | −0.480<br>(0.285) | −0.710<br>(0.056) | −1.190<br>(0.293) | −0.599<br>(0.384) | −1.143<br>(0.096) | −1.742<br>(0.341) |

*Note.* This table presents the price elasticity for a representative household at different points of the price and wealth distribution. Standard errors included in the parentheses are computed using the delta method. Elasticity estimates are computed under a marginal change in prices and income. The total demand response in elasticity estimates is decomposed into the direct and indirect effect. **Dir.** denotes the direct effect, **Indir.** denotes the indirect effect generated by the spillover. **Tot.** denotes the total household response. **M** denotes the Marshallian (or uncompensated) price elasticity.

the wealth level (down a column). More specifically, Table 5 shows that the demand response to a change in the price of sanitation *declines* with household wealth; that is, poorer households are more sensitive to price changes. For example, a household at the 10th percentile of wealth distribution is 1.27 times ($-0.519/-0.403$) more price sensitive than its wealthier counterpart at the 90th percentile. This difference in price sensitivity for poorer and wealthier households is mostly stable across the price distribution: 1.32 times at the 10th percentile, 1.27 at the 50th percentile, and 1.25 at the 90th percentile, respectively. In addition, the household price elasticity response increases with the price of sanitation for a given level of wealth. For a household in the 50th percentile of the wealth distribution, price sensitivity increases by 1.90 ($-0.630/-0.331$) as the market price increases from the 10th to the 90th percentile. Standard errors are included in parentheses. Excluding the 90th price percentile, all price elasticity estimates are significant at the 10% level.

In addition to price, Table 5 also includes income elasticity estimates, denoted by **I**, at different points of the wealth and price distribution. The income elasticity estimates are positive and significant at the 1% level, indicating an *increase* in the probability of adoption for a small *increase* in income. The results are striking: price elasticity estimates are, on average, eight times larger than the income elasticity measures. The difference in magnitude follows directly from the magnitudes of the demand parameter estimates in Table 4. Similar to the price elasticity estimates, the income elasticity of sanitation adoption *increases* with the price of sanitation for a given level of wealth. Perhaps unsurprisingly, I also find that poorer households are income *inelastic* in their sanitation take-up response relative to their wealthier counterparts. However, the response to changes in income for households in the 75th percentile is approximately seven times that of households in the 25th percentile. The difference in magnitude highlights the considerable distributional impacts of policy interventions and the importance of an in-depth welfare analysis. Overall, the large degree of heterogeneity in both the price and income elasticity response displayed in Table 5 points to significant wealth effects. While the inelastic income response at the lower end of the wealth distribution may point to the presence of liquidity (or borrowing) constraints, the inelastic response at the top end of the wealth distribution is somewhat unexpected. The income elasticity response for households in the 90th percentile of wealth is 0.37, well below one. This finding points to sanitation being a *necessity* for the household and not a *luxury* good. I further expand on the implications from the income elasticity estimates in the household composition analysis below.

### 4.3.2. Direct and indirect effects

Before turning to the household composition effects, I take a closer look at the price elasticity measures in Table 5. Given the underlying interdependence in household sanitation adoption, there will be an additional spillover effect defined as term $B$ in Eq. (23). Table 6 displays the decomposition of the price elasticity estimates into the direct and indirect response to a price change as a function of the total elasticity.

There are two key takeaways from Table 6. First, the household-level demand response captured by the total price elasticity is significantly larger once the indirect effect is taken into account. The indirect demand response to a price change accounts, on average, for 65% of the total response. The sharp increase in demand sensitivity after the indirect effect is included is a direct outcome of incorporating the externalities that arise from sanitation take-up and the interdependence in household choice. Second, the total uncompensated demand response is primarily determined by the magnitude of the indirect effect. While there is some variation across the distribution, as a proportion, the indirect response ranges between 59% at the lower limit and close to 69% at the upper limit. The size of this impact is directly related to the magnitude of the parameter $\gamma$ in the utility function, which captures the strength of the underlying strategic interactions.

The decomposition in Table 6 allows us to quantify the proportion of policy impact attributable to the direct household response relative to the indirect effect generated from the presence of externalities. These findings provide useful information for policymakers and governments that want to make effective *ex-ante* program placement decisions with a limited budget (Todd and Wolpin, 2008). To summarize, the fact that policies that target price (and income) generate spillover effects due to interdependence is not surprising. The difference in the magnitude of the spillover effects, which are proportionally related to the relative magnitude of the price and income elasticity estimates in Table 5, is also largely expected. Nevertheless, the substantial heterogeneity in the price and income elasticities highlights substantial scope to explore policy design questions. I return to this discussion in Section 5, where I compare the effectiveness of price subsidy and cash transfers in increasing sanitation take-up and welfare.

#### *Household composition*

The estimates of household characteristics in Table 4 point to the importance of household composition as a determinant of sanitation demand. Tables 7 and 8 display the price and income elasticities for different household composition groups. Table 7 summarizes households' price and income sensitivity along two dimensions of policy interest: the gender of adults and the number of children in the household. For conciseness, the elasticity estimates are computed at the 50th percentile of the price and wealth distribution. The estimates document substantial heterogeneity in demand responsiveness to price and income as the composition of women and children within the household changes. Demand responsiveness *decreases* with the number of adult females in the household. In other words, a household with more women has a more "rigid" demand for sanitation relative to one with only one adult female member. Similarly, both price and income sensitivity of sanitation demand *decrease* with the number of children under the age of 14. Further details on the household composition groups can be found in the table notes.

The rigidity of the demand response in Table 7 is directly related to the importance of sanitation for those household groups. The *increase* in the rigidity of the household demand, particularly with respect to income, can be viewed as the *shadow value* of sanitation. Intuitively,

**Table 7**
Household composition of women and children: Price and income elasticity.

| Children | El. | Women | | | |
|---|---|---|---|---|---|
| | | HH w/1Wo | HH w/2Wo | HH w/3Wo | HH w/4Wo |
| HH w/o Ch | **M** | −0.493 | −0.445 | −0.388 | −0.243 |
| | | (0.297) | (0.270) | (0.248) | (0.163) |
| | **I** | 0.092 | 0.083 | 0.072 | 0.045 |
| | | (0.040) | (0.037) | (0.034) | (0.022) |
| HH w/1Ch | **M** | −0.469 | −0.418 | −0.358 | −0.215 |
| | | (0.275) | (0.245) | (0.220) | (0.136) |
| | **I** | 0.061 | 0.054 | 0.046 | 0.028 |
| | | (0.045) | (0.040) | (0.035) | (0.024) |
| HH w/2Ch | **M** | −0.438 | −0.383 | −0.321 | −0.184 |
| | | (0.264) | (0.232) | (0.206) | (0.124) |
| | **I** | 0.012 | 0.010 | 0.009 | 0.005 |
| | | (0.017) | (0.015) | (0.013) | (0.007) |

*Note.* This table presents the direct price and income elasticities estimates for different household composition groups computed at the $50th$ percentile of price and income distribution. Elasticity estimates are computed under a marginal change in prices and income. Standard errors in the parentheses are computed using the delta method. **M** denotes the Marshallian price elasticity. **I** denotes the pure income elasticity. Household composition groups are defined according to the number women and children within the household as follows: Row [1] **HH w/o Ch** households without a child under the age of 14, Row [2] **HH w/1Ch** households with 1 child under the age of 14, Row [3] **HH w/2Ch** households with 2 or more children under the age of 14, Col [1] **HH w/1Wo** households with 1 adult woman over the age of 18, Col [2] **HH w/1Wo** households with 2 adult women over the age of 18, Col [3] **HH w/1Wo** households with 3 adult women over the age of 18, Col [4] **HH w/1Wo** households with 4 adult women over the age of 18.

**Table 8**
Household composition of children and education: Price and income elasticity.

| Children | El. | Household head | | | |
|---|---|---|---|---|---|
| | | No Edu | Primary | Secondary | >Secondary |
| HH w/o Ch | **M** | −0.524 | −0.455 | −0.369 | −0.332 |
| | | (0.321) | (0.274) | (0.220) | (0.196) |
| | **I** | 0.098 | 0.085 | 0.069 | 0.062 |
| | | (0.043) | (0.037) | (0.030) | (0.027) |
| HH w/1Ch | **M** | −0.503 | −0.428 | −0.338 | −0.301 |
| | | (0.301) | (0.250) | (0.191) | (0.168) |
| | **I** | 0.065 | 0.055 | 0.044 | 0.039 |
| | | (0.049) | (0.041) | (0.0032) | (0.029) |
| HH w/2Ch | **M** | −0.476 | −0.393 | −0.301 | −0.264 |
| | | (0.291) | (0.238) | (0.178) | (0.155) |
| | **I** | 0.013 | 0.011 | 0.008 | 0.007 |
| | | (0.019) | (0.015) | (0.011) | (0.010) |

*Note.* This table presents the direct price and income elasticity estimates for different household composition groups computed at the $50th$ percentile of price and income distribution. Elasticity estimates are computed under a marginal change in prices and income. Standard errors in the parentheses are computed using the delta method. **M** denotes the Marshallian price elasticity. **I** denotes the pure income elasticity. Household composition groups are defined according to education of household head and the number of children within the household as follows: Row [1] **HH w/o Ch** households without a child under the age of 14, Row [2] **HH w/1Ch** households with 1 child under the age of 14, Row [3] **HH w/2Ch** households with 2 or more children under the age of 14, Col [1] **No Edu** household head has no education, Col [2] **Primary** household head has primary education (5 years), Col [3] **Secondary** household head has secondary education (10 years), Col [4] **>Secondary** household head has more than secondary education (11–15 years)

holding wealth fixed a lower income elasticity for households with women and children reflects the *necessity* of sanitation for those households. Lastly, the overall heterogeneity in the income sensitivity of take-up response summarized in Table 7 points to important avenues of policy design, such as targeting interventions to households with women and children. In the next section, I explore this aspect further, in which I compare two specific policy instruments: an *untargeted* price subsidy and a *targeted* income transfer.

Similarly, Table 8 summarizes households' price and income sensitivity, featuring different education levels of the household head

**Table 9**
Household composition of children: Change in sanitation take-up.

| ΔPr(Take-up) | Children | | |
|---|---|---|---|
| Women | HH w/o Ch | HH w/1Ch | HH w/2Ch |
| HH w/1Wo | – | +0.035 | +0.082 |
| HH w/2Wo | – | +0.041 | +0.094 |
| HH w/3Wo | – | +0.046 | +0.102 |
| HH w/4Wo | – | +0.044 | +0.094 |

*Note.* This table presents the change in the probability of sanitation take-up for changes in the household composition of children relative to the base category in columns. The number of adult women is held fixed in each row. The estimates are computed at the $50th$ percentile of price and income distribution. HH w/o Ch denotes the base category. Household composition groups are defined according to the number of women and children within the household as follows: Col [1] **HH w/o Ch** households without a child under the age of 14, Col [2] **HH w/1Ch** households with 1 child under the age of 14, Col [3] **HH w/2Ch** households with 2 or more children under the age of 14, Row [1] **HH w/1Wo** households with 1 adult woman over the age of 18, Row [2] **HH w/2Wo** households with 2 adult women over the age of 18, Row [3] **HH w/3Wo** households with 3 adult women over the age of 18, Row [4] **HH w/4Wo** households with 4 adult women over the age of 18.

**Table 10**
Household composition of women: Change in sanitation take-up.

| Δ Pr(Take-up) | Women | | | |
|---|---|---|---|---|
| Children | HH w/1Wo | HH w/2Wo | HH w/3Wo | HH w/4Wo |
| HH w/o Ch | – | +0.071 | +0.156 | +0.379 |
| HH w/1Ch | – | +0.077 | +0.167 | +0.388 |
| HH w/2Ch | – | +0.083 | +0.177 | +0.391 |

*Note.* This table presents the change in the probability of sanitation take-up for changes in household composition of women relative to the base category in columns. The number of children is held fixed in each row. The estimates are computed at the $50th$ percentile of price and income distribution. **HH w/1Wo** denotes the base category. Household composition groups are defined according to the number women and children within the household as follows: Col [1] **HH w/1Wo** households with 1 adult woman over the age of 18, Col [2] **HH w/2Wo** households with 2 adult women over the age of 18, Col [3] **HH w/3Wo** households with 3 adult women over the age of 18, Col [4] **HH w/4Wo** households with 4 adult women over the age of 18, Row [1] **HH w/o Ch** households without a child under the age of 14, Row [2] **HH w/1Ch** households with 1 child under the age of 14, Row [3] **HH w/2Ch** households with 2 or more children under the age of 14.

and child composition. Households with an educated head with two children are much *less responsive* to price and income changes than households with an uneducated head with no children. Specifically, a household with one child where the head has completed secondary education is just over half (0.301/0.503) as responsive to price changes than a head with no education. The magnitude and direction of income *insensitivity* follow a similar pattern to price. The rigidity of the demand response on the income of households with an educated head with more children suggests that sanitation is a stronger *necessity* for these households. In terms of levels, the magnitude of both price and income elasticity *decreases* with the number of children in the household regardless of the education level of the household head. This result can be viewed as further evidence of a monotonic relationship between the number of children in the household and the decline in sensitivity to price and income changes seen in Table 7.

While demand price and income elasticity estimates represent a key dimension to evaluate the effectiveness of a policy aiming to induce higher adoption rates within a village, an assessment of the broader consequences must also account for the heterogeneity in sanitation adoption rates across different household demographic groups. Tables 9 and 10 provide additional evidence on heterogeneity dimension by computing the sanitation take-up response to changes in the household composition relative to the base category. The estimates in the two tables can be viewed as a sanitation take-up elasticity measure for changes in household composition holding price and income fixed.

Table 9 presents the change in sanitation take-up probability for changes in the number of children relative to a household with no children holding fixed the number of adult women. The results show that the probability of sanitation adoption strongly depends on the presence of children in the household. Adding an extra child increases the probability of take-up by 3.5 to 5 percentage points, depending on the number of women in the household. This effect nearly doubles with the addition of two children. Table 10 shows that the sanitation take-up probability similarly *increases* with the number of adult female household members holding fixed the number of children. Relative to a household with only one woman, adding an adult female member implies an increase in the probability of sanitation adoption by approximately eight percentage points. The magnitude of this effect increases with the number of women in the household. Collectively, the results presented in Tables 9 and 10 illustrate a second critical point for policy design. The analysis suggests that as take-up rates differ substantially across household composition groups, the welfare effects of a policy that incentivizes sanitation adoption may also be severely skewed towards specific household types. Therefore, the policy may have substantial distributive impacts that may be of substantive importance for the policymaker. In the next section, I assess the size and heterogeneity of the welfare effects of different types of policy intervention.

## 5. Ex-ante policy evaluation

I conduct an ex-ante policy evaluation in this section. The evaluation is performed on two criteria: (1) the cost effectiveness of the policy in increasing the sanitation take-up rate at the village level, and (2) the distributional effects of the policy on the households within each village. The former analysis is carried out using the price and income elasticities computed in Section (4.3). For the latter analysis, I build on Attanasio et al. (2013), who derive a measure of the welfare effect induced by food price changes using a QUAIDS demand model. I adapt their approach to a discrete choice setup to compute welfare gains.

### 5.1. Ex-ante compensating variation

To quantify the welfare effects under different policy interventions, I compute an ex-ante money measure of the welfare change generated by the policy. This computation determines the amount of money given to (or taken away from) a household to make them indifferent *in expectation* between two scenarios: with and without the policy. The expectation is taken over the realizations of the household's taste shocks. Note that this welfare measure closely resembles the familiar concept of compensating variation, except that it is calculated ex-ante from the household's perspective. That is, it is computed with respect to the expected indirect utility rather than its realized value. This object is useful and attractive for two reasons. First, it allows for a natural interpretation as an ex-ante utility measure **before** introducing the policy intervention. Second, the ex-ante utility function has a closed-form expression that is invertible under the type 1 extreme value distribution assumption. This property is used to back out the expected expenditure function that can be used to compute the ex-ante compensating variation ($ECV$) welfare measure.

The expected expenditure function is computed by inverting the expression for the household's expected indirect utility in Eq. (12) for each household in the village sample.[13] Following McFadden (1978) and Arcidiacono and Ellickson (2011) the expected indirect utility function of household $i$ can be expressed as

$$\int \widetilde{V}^*_{ig}\left(y_{ig}, p_g; w_g, \epsilon, \tilde{d}^*_{-ig}, \theta\right) F(\epsilon)\, d\epsilon = -\ln\left[1 - \mathrm{P}_{ig}\left(y_{ig}, p_g, \overline{\mathrm{P}}_g\left(y_g, p_g\right)\right)\right] + \phi \tag{26}$$

where $\phi$ is the Euler's constant. Consider a policy that changes the income vector and the price of sanitation in village $g$ from $(y_g, p_g)$ to $(y'_g, p'_g)$. For example, a *uniform price subsidy* $S_g$ corresponds to a policy that changes the price from $p_g$ to $p'_g = p_g + S_g$, leaving the vector $y_g$ unchanged, while a *targeted cash transfer* changes the income vector from $y_g$ to $y'_g$ with no effect on the price $p_g$. Using the formula for $\mathrm{P}_{ig}\left(y_{ig}, p_g, \overline{\mathrm{P}}_g\left(y_g, p_g\right)\right) = \mathbb{P}^*_{ig}\left(d_{ig} = 1 \mid w_g, \theta\right)$ in Eq. (15), the corresponding adjustment in income that makes household $i$ indifferent ex-ante between the two scenarios $(y_g, p_g)$ and $(y'_g, p'_g)$, denoted by $ECV_{ig}\left(\left(y_g, p_g\right), \left(y'_g, p'_g\right)\right)$, has the formula:

$$ECV_{ig}\left(\left(y_g, p_g\right), \left(y'_g, p'_g\right)\right) = -\left(y'_{ig} - y_{ig}\right) - \frac{\xi}{\zeta + \lambda x_{ig}}\left(p'_g - p_g\right) - \frac{\gamma}{\zeta + \lambda x_{ig}}\left[\overline{\mathrm{P}}_g\left(y'_g, p'_g\right) - \overline{\mathrm{P}}_g\left(y_g, p_g\right)\right] \tag{27}$$

The value of $ECV_{ig}$ can be calculated using the fixed point mapping in Eq. (15) under the approximation of the expected average equilibrium probability stated in Eq. (22). Similar to a standard compensating variation, $ECV_{ig}$ takes negative values when the household experiences a welfare gain (*money that has to be taken away*), and its magnitude represents the monetary value of such gain. Eq. (27) illustrates how the value of $ECV_{ig}$ depends upon the change in household income induced by the policy $(y'_{ig} - y_{ig})$, in the price of sanitation $(p'_g - p_g)$, and in the average adoption rate in the village $\overline{\mathrm{P}}_g\left(y'_g, p'_g\right) - \overline{\mathrm{P}}_g\left(y_g, p_g\right)$. Note that Eq. (27) implies that a uniform cash transfer (or subsidy) generates a value for household $i$ that exceeds the monetary value of the transfer received. This follows from the fact that in the presence of externalities, such a policy produces a feedback effect on the average adoption level in the village and, in turn, on household $i$'s welfare. The $ECV_{ig}$ measure can also be used to quantify the private and social value of sanitation adoption. As an illustrative example, a one dollar change in price generates an average private value of 8.8 dollars for the household and a social value of 14.77 dollars. The wedge between the two corresponds to the size of the public benefits from externalities. In what follows, Eq. (27) is used to assess the welfare effects under the two counterfactual policy scenarios and compute the average welfare and distributional effects of the policy impacts within the population.

### 5.2. Equilibrium policy effects and social multiplier

Before presenting the results from the counterfactual exercise, I illustrate the mechanism through which different types of policy interventions affect aggregate sanitation adoption behavior at the village level. Fig. 2 provides a graphical representation of the simulated aggregate household choice behavior within a specific village. It compares the effects of two policy interventions – an untargeted cash transfer and a price subsidy of an equal amount – relative to the baseline scenario in which no policy is implemented. The average prevalence of sanitation in the village is plotted on the horizontal axis, while the vertical axis plots the probability of sanitation adoption for an individual household as a function of average sanitation adoption in the village. The diagonal plots the 45° line. The solid black line plots the response curve for the baseline scenario under no policy. Its shape is determined from the simulation of the village-level fixed point described in Eq. (15) for a fine grid of $\bar{d}_g$ that denotes the prevalence of sanitation in the village. The dashed red and green lines plot the same response for a price subsidy and a cash transfer, respectively. An equilibrium is a point at which the response curve crosses the 45° line, where the average probability

[13] Eq. (12) illustrates how such inversion is typically not a straightforward analytical calculation in discrete-choice models because the object of interest $V^*_{ig}$ is the maximum of the two choice-specific indirect utility functions. However, under a specific choice of distribution for the idiosyncratic taste shocks, the formula for $V^*_{ig}$ in Eq. (12) admits a closed-form expression that dramatically simplifies the analysis.

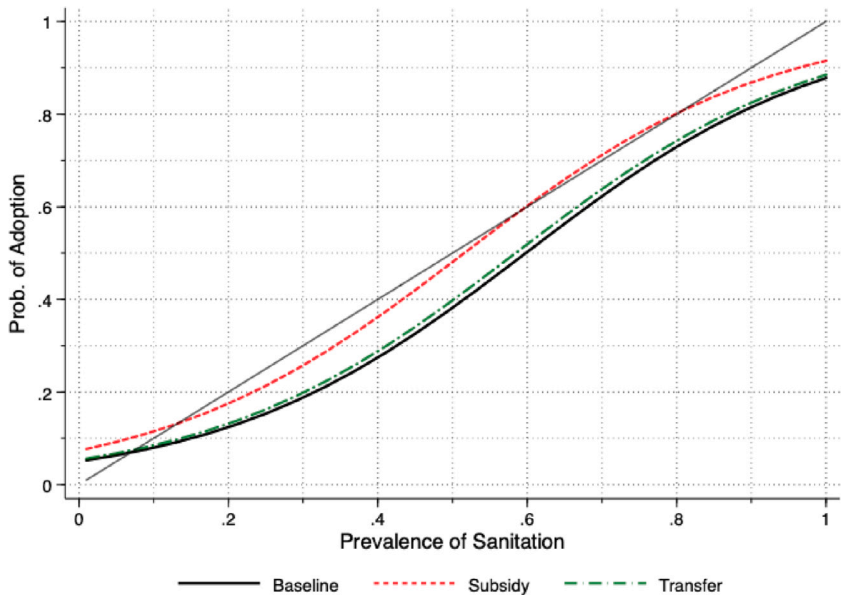



**Fig. 2.** Probability of sanitation adoption. *Note.* This graph plots the probability of sanitation adoption as a function of sanitation prevalence in the village. The model simulation is performed using a representative village from the sample. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of adoption equals the expected prevalence of sanitation in the village. Fig. 2 illustrates the mechanisms which govern the interdependence in household behavior.

First, the figure shows how a price subsidy shifts the response curve upwards to a much larger extent than a cash transfer of an equal amount. This implies that a household's response is substantially larger to a subsidy than to a cash transfer. Although the implication is consistent with the elasticity measures in Table 5, households do not act in isolation. A change in price (or income) has both a direct and indirect effect, where the latter depends on the mean adoption in the village. The indirect spillover effect exacerbates the initial disparity in the direct demand response to the two policies. This is because the magnitude of the spillover effect depends on both the social interaction parameter $\gamma$ and the average adoption in the village, which is endogenous. A median wealth household is eight times more responsive to price than income. As a result, many more households adopt sanitation under the direct effect of a subsidy compared to an income transfer. Simultaneously, each household's decision to adopt responds optimally to the policy based on their belief about the change in aggregate village behavior. The larger the direct increase, the larger the spillover effect on households that would otherwise not have adopted but are now motivated to do so. Though the transfer and subsidy both generate spillover effects, in the case of a price subsidy, the same $\gamma$ *multiplies* a bigger village response, thereby shifting the subsidy response curve upward to a larger extent.

Second, I find the emergence of multiple equilibria in the counterfactual policy scenario of a price subsidy. As seen in Fig. 2, the dashed red line crosses the 45° line at three points, implying three different village-level adoption levels that are consistent with households' utility maximizing behavior. Multiple equilibria are a direct consequence of the spillovers in sanitation adoption driven by the interdependence in household behavior. In turn, the degree of inter-household dependence is related to the magnitude of $\gamma$ that captures the importance of the underlying social interactions. The present analysis does not extend to determining which one of the three equilibria the society would actually move to under the counterfactual scenario.

While Fig. 2 provides insight into why a price subsidy tends to generate a stronger response in terms of sanitation take-up than a cash transfer of equal amount, it also raises several questions of policy interest. First, as policymakers operate under limited budgets, it is crucial to evaluate the cost effectiveness of each policy; in other words, which of the feasible policy alternatives can induce the highest village-level adoption rate? Second, as cash transfers are often targeted to specific household types, it is important, from a policy design perspective, to assess whether the cost effectiveness of such policy instruments can be improved through targeting. Third, cash transfers and price subsidies may produce very different distributive effects. Therefore, the assessment of various policy alternatives should consider the impact on the aggregate take-up behavior and how the total welfare benefits from each intervention are allocated across different households within a village.

### 5.3. Welfare analysis

#### 5.3.1. Sample selection

Because the aggregate demand equation is based on household micro-data, it is possible to estimate both the average welfare effect in the sample and the distributional effects of each policy within the population. To motivate the policy evaluation analysis, I first define the sample selection and the scenario under which counterfactual simulations are conducted. For policy simulations, the sample is restricted to those villages where the level of sanitation adoption is below 50%. With this selection, the sample size for the counterfactual exercises is close to 900 households (across 25 villages), which is approximately 60% of the original sample. The mean adoption level in the selected sample is 16.8%, which is necessarily lower than the mean in the full sample (37.9%).

For each village, I simulate the take-up response to an untargeted price subsidy based on a 25% discount on the price of sanitation in the village. The subsidy is disbursed uniformly across the village. For each household that has yet to adopt sanitation, the probability of adoption under this scenario (price change) is computed, taking into account both the direct and indirect effects. The indirect or spillover effect will be generated from inter-household dependence on choice behavior that magnifies the impact of the market-wide subsidy. To maintain budget neutrality across the counterfactual policies, I take the unit amount disbursed under the price subsidy and multiply it by the number of

**Table 11**
Subsidy versus targeted cash transfer.

| | Subsidy | | | | Targeted cash transfer | | | |
|---|---|---|---|---|---|---|---|---|
| | Take-up | ECV (Rs.) | | | Take-up | ECV (Rs.) | | |
| | p.p. chg | Dir. | Indir. | Tot. | p.p. chg | Dir. | Indir. | Tot. |
| *Panel A. Overall* | | | | | | | | |
| Average impact | 2.11 | −20.57 | −16.37 | −36.94 | 0.20 | −0.45 | −2.28 | −2.74 |
| *Panel B. Household income* | | | | | | | | |
| $10th\,\%tile$ | 2.56 | −14.87 | −10.95 | −25.82 | 0.40 | −0.39 | −1.59 | −1.99 |
| $25th\,\%tile$ | 3.03 | −18.76 | −14.26 | −33.03 | 0.42 | −0.43 | −1.97 | −2.39 |
| $50th\,\%tile$ | 3.21 | −20.63 | −16.59 | −37.22 | 0.45 | −0.46 | −2.31 | −2.77 |
| $75th\,\%tile$ | 4.83 | −23.01 | −22.11 | −45.12 | 0.55 | −0.55 | −3.17 | −3.72 |
| $90th\,\%tile$ | 4.42 | −23.64 | −33.02 | −56.67 | 0.85 | −0.57 | −3.45 | −4.02 |

*Note.* This table presents the simulation results for a 25% subsidy and a targeted transfer to households with children. The simulation is performed under budget neutrality across the two policy instruments. Policy performance is measured on take-up and expected compensating variation (ECV) measure of welfare. Take-up measures the percentage point change (p.p. chg) in sanitation prevalence. ECV is measured in Indian rupees scaled per Rs. 1000. Total (Tot.) welfare gain is decomposed into direct (Dir.) and indirect (Indir.) gains generated via spillover effects. The negative value on the ECV measure denotes the amount of money that needs to be "taken away" from a household under each policy. Household income percentile shown in rows in Panel B.

households that are new adopters under the subsidy intervention. This gives the total cost of the untargeted subsidy policy. To simulate the response to the cash transfer policy, an amount equal to such total cost is divided among the targeted households in the village. The total budget is distributed uniformly as a cash transfer among village households with at least one child under the age of 14. I recompute the probability of adoption for each household under the new scenario (change in income), considering both the direct and indirect change in response.

### 5.3.2. Distributive impact of a subsidy and cash transfer

Table 11 summarizes the average effect of the subsidy and transfer policy interventions on a selected sample of villages. For each policy, performance is measured on take-up and expected compensating variation (ECV) measure of welfare. The welfare measure is computed using Eq. (27). Panel A shows the overall impact, while panel B shows the policy impact across different household composition groups. As seen from Panel A, the price subsidy dominates the targeted cash transfer in terms of cost effectiveness. It raises the average level of adoption by 2.1 percentage points, corresponding to an increase of 12.5% in the average adoption rate from a base of 16.8%. In contrast, for the same total cost, the targeted cash transfer generates a 0.2 percentage points increase in take-up, corresponding to a 1.2% increase. Achieving a similar increase in take-up rate under the cash transfer would require a policy budget six times larger than under the subsidy.

In terms of welfare, the average ECV is significantly larger under the price subsidy than the targeted cash transfer. A negative value on the ECV denotes the amount of money that needs to be "taken away" from each household under each policy intervention. The larger the amount that needs to be taken away, the larger the welfare gains realized under the policy intervention. A 25% subsidy, amounting to 2.33 rupees, produces an average per-household welfare benefit of 37 rupees. For a household that adopts sanitation, the private gain is 21 rupees, while the indirect public gain is worth 16 rupees. On average, a sizable $2/5^{ths}$ of the welfare gain is public. Compared with the subsidy, the total welfare gains under the transfer policy are substantially smaller. However, this does not necessarily imply that the cash transfer is unambiguously undesirable as a policy intervention. First, note that the targeted cash transfer is welfare-enhancing: it produces an average per-household welfare benefit equal to 2.74 rupees for a given cost of 0.90 rupees. On average, one rupee spent on a transfer policy generates a value equal to 3.04 rupees. The private gain for a household that receives the transfer is only 0.45 rupees, and the indirect public gain is significantly larger, amounting to 2.3 rupees.

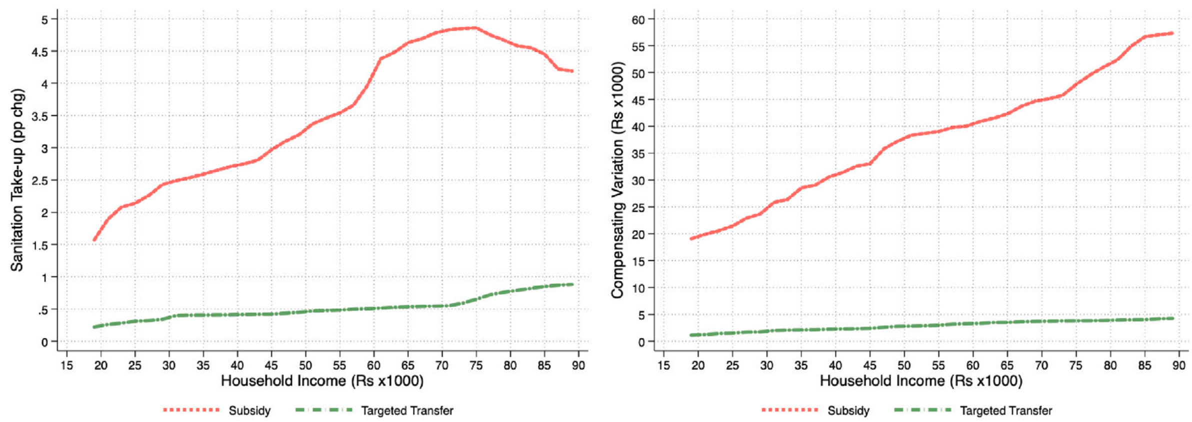
However, panel A does not paint a complete picture of the distributive impacts of the two policies at different levels of income distribution or for different demographic composition groups. A key question is whether the distributional effects of these two policies are regressive and, if so, by how much. Panel B of Table 11 illustrates how both policies affect take-up behavior in a rather heterogeneous way with respect to total equivalized household income.[14] The equivalized income measure for each household is calculated by dividing the household income by the number of equivalent members. This number is constructed using the OECD-modified equivalence scale, which assigns a value of 1 to the household head, 0.5 to each additional adult member, and 0.3 to each child.[15] The increase in sanitation adoption is more considerable among wealthier households under both policies. Similar to the average effect, the welfare gains for all income groups are substantially larger under the subsidy than the transfer. However, comparison across income groups for a given policy highlights interesting distributive effects that can be seen in Fig. 3.

Fig. 3 shows how sanitation take-up and the welfare gains vary by total equivalized household income under the two policies. The distribution of the effects of the price subsidy on sanitation take-up, as seen in (Fig. 3.a), is strongly skewed towards relatively wealthy households. The increase in the probability of sanitation adoption for a household in the 75th percentile of the distribution of equivalized income is more than 1.6 times the corresponding increase for a household in the 25th percentile. (Fig. 3.b) measures the distributive impact on welfare using the expected compensating variation (ECV). The results regarding the distributive welfare impact are even more striking. The sharp increase in ECV with household income is striking. (Fig. 3.b) shows a sharp increase in ECV with household income, suggesting a strong regressive impact of the subsidy policy. The wealthiest households derive a welfare gain almost 2.5 times that of the poorest households. Similar to the subsidy, the effect of the targeted cash transfer on sanitation take-up is skewed towards relatively wealthy households. However, unlike the subsidy, the targeted cash transfer translates to a more equal distribution of welfare gains, even though the gains are lower.

The findings from Table 11 and Fig. 3 are crucial for policy design. Together they show how the distribution of welfare gains differs across income groups under the two policies. This is important if policymakers evaluate the desirability of a policy intervention in terms of targets beyond the mere aggregate effect on sanitation adoption and may be concerned about the distribution of welfare gains. Moreover, the results suggest that while the targeted cash transfer is less cost effective

[14] The importance of analyzing welfare with a more comprehensive measure of wealth such as cash in hand was first put forth by Deaton (1989), and Deaton (1991).

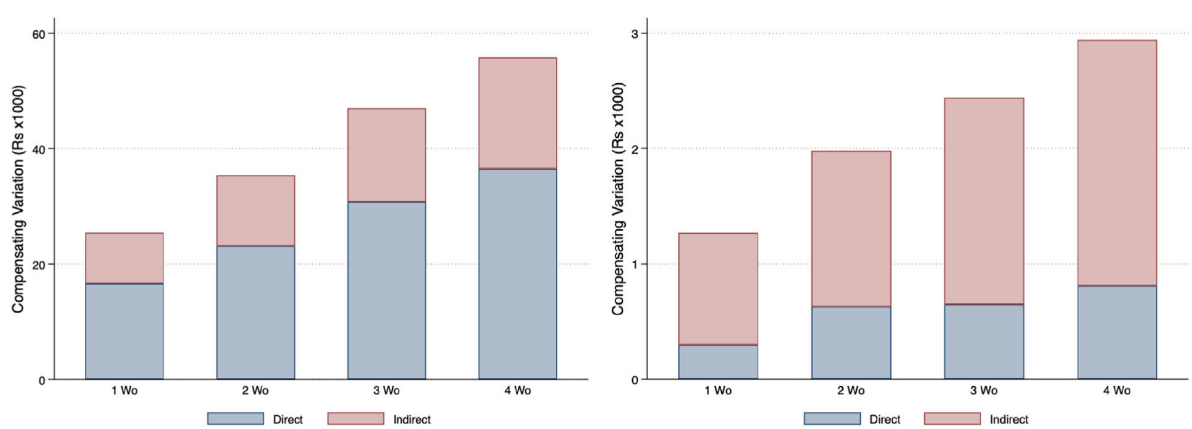
[15] This scale, first proposed by Asghar Zaidi et al. (1995), has been widely used since the late 1990s by several offices of national statistics, such as the Statistical Office of the European Union (EUROSTAT).




(a) Sanitation Take-up

(b) Welfare gains

**Fig. 3.** Subsidy versus targeted transfer. *Note.* This figure shows the distribution of sanitation take-up and welfare gains under a 25% price subsidy and targeted cash transfer to households with children. The simulation is performed under budget neutrality across the two policy instruments. Take-up measures the percentage point change (p.p. chg) in sanitation prevalence. The expected compensating variation (ECV) denotes a welfare measure in Indian rupees scaled per Rs. 1000. Figure (a) shows the change in sanitation take-up at different points of the household income. Figure (b) measures the welfare effect of the two policies at different points of the household income.




(a) Subsidy

(b) Targeted Transfer

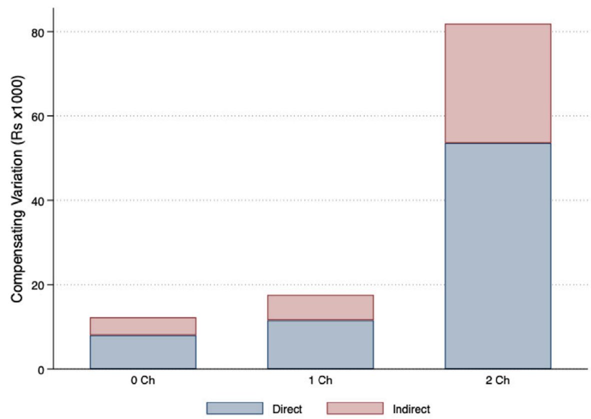
**Fig. 4.** Distribution of Welfare Gains (Women). *Note.* This figure shows the distribution of welfare gains for households with adult women, under a price subsidy and targeted cash transfer. The simulation is performed under budget neutrality across the two policy instruments. Figure (a) shows the change in welfare under a subsidy. Figure (b) shows the change in welfare under a cash transfer. The expected compensating variation (ECV) denotes a welfare measure in Indian rupees scaled per Rs. 1000. The *x*-axis labels denote the number of adult women in the household. [1Wo] households with 1 adult woman over the age of 18, [2Wo] households with 2 adult women over the age of 18, [3Wo] households with 3 adult women over the age of 18, and [4Wo] households with 4 adult women over the age of 18.

than the price subsidy in inducing sanitation adoption, it has fewer regressive distributional consequences. Thus, a policymaker must trade off cost effectiveness and distributive justice when choosing between the two types of intervention. Lastly, this analysis suggests that cash transfers may be used in *conjunction* with price subsidies as they may help ameliorate the stark regressive effect of price subsidies. To explore the feasibility of doing so, I analyze the welfare outcomes of each policy intervention across different household demographic types.

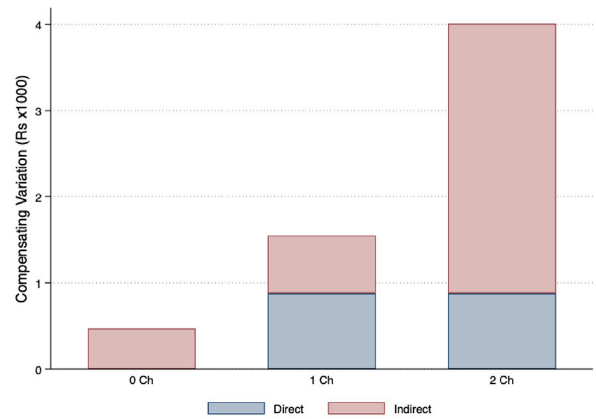
### 5.3.3. Targeting households with women and children

To assess how the welfare benefits are allocated across different households in a village, I analyze the heterogeneity in welfare effects produced by the two policies based on the number of women and children within the household. Fig. 4 shows the average welfare gains generated by the subsidy and the targeted cash transfer for households with different numbers of adult female members. Each measure of welfare gain is divided into a direct effect of the policy on household welfare and an indirect effect due to the change in the average adoption level in the village induced by the policy. The figure shows that the welfare gains from both policy interventions increase with the number of women in the household. Overall, the distributive effects of the two policies in terms of the number of women in the household are similar.

Similarly, Fig. 5 shows the allocation of welfare gains for households with different numbers of children. Here, the price subsidy has much stronger distributive effects than the targeted cash transfer. The welfare gains generated by the subsidy policy are concentrated within a particular demographic group: households with two or more children. Those households enjoy an average welfare gain nearly five times that of a household with only one child. Conversely, the welfare gains from the targeted transfer are less concentrated: households with two or more children enjoy 2.6 times the welfare gains experienced by households




(a) Subsidy




(b) Targeted Transfer

**Fig. 5.** Distribution of Welfare Gains (Children). *Note.* This figure shows the distribution of welfare gains for households with children, under a price subsidy and targeted cash transfer to households with children. The simulation is performed under budget neutrality across the two policy instruments. Figure (a) shows the change in welfare under a subsidy. Figure (b) shows the change in welfare under a cash transfer. The expected compensating variation (ECV) denotes a welfare measure in Indian rupees scaled per Rs. 1000. The $x$-axis labels denote the number of children in the household. [0Ch] households without a child under the age of 14, [1Ch] households with 1 child under the age of 14, and [2Ch] households with 2 or more children under the age of 14.

with only one child. To summarize, although the welfare effects of the subsidy and transfer policies were analyzed separately, they should be considered as potentially complementary interventions from a policy design perspective. A policymaker aiming to achieve a satisfactory average sanitation adoption level and correct the regressive effect that certain policy interventions may generate may consider combining both policies under a fixed budget.

## 6. Conclusion

While the public and private health gains from sanitation coverage are well established (Coffey et al., 2018; Geruso and Spears, 2018), the distributional impact of policies to increase sanitation coverage is less understood. A comprehensive understanding of the cost effectiveness and welfare impacts of sanitation interventions at the household level is far from complete. This paper makes an important contribution by conducting an ex-ante policy evaluation of existing government interventions. To this end, I develop a household choice model for goods with externalities and embed it within a simple game to characterize the interdependence in household behavior. I identify and structurally estimate the primitive parameters of the model using market and household survey data from India. The evaluation framework accounts for externalities in the consumption of sanitation facilities and allows for the consequent interdependence in household choice of sanitation adoption. I use the estimated model to quantify the household sanitation adoption response and the subsequent welfare effects under different sanitation policy interventions.

The insights derived from this paper have two key policy implications for the design of sanitation interventions. First, the *price subsidy* is unambiguously superior to the *targeted cash transfer* with respect to the cost effectiveness in increasing sanitation adoption in aggregate. Thus, the analysis suggests that if the policymaker's sole aim is to increase the sanitation adoption rate at the village level, and limited resources can be allocated towards this target, the former policy is strongly preferable to the latter. Second, the *price subsidy* produces distributional effects that are strongly regressive and skewed toward specific demographic groups. Therefore, as the amount of resources available for the policy intervention increase and the relative importance of distributive justice in the policymaker's objective function grows, the *targeted cash transfer* becomes increasingly desirable and preferable to the *price subsidy*.

Incorporating welfare evaluation in the empirical study of externalities is an important addition to the sanitation impact evaluation literature and makes the model applicable to several contexts for studying spillover effects in developing countries. In particular, there is substantial room to explore the theoretical structure developed in this paper for ex-post policy evaluation of outcomes that are beyond the scope of this paper. In a companion paper, I extend this household model to capture inter-temporal trade-offs, incorporating both externalities and borrowing constraints to deliver a richer empirical framework capable of answering optimal policy design questions related to the targeting of loans and subsidies and the associated Pareto efficiency gains from each policy (Gautam, 2019). In addition to household dynamics, there is also scope to explore market-level dynamics, such as the timing of adoption. From a policy perspective, there may be additional cost efficiency gains from targeting specific households that differ in their marginal utility from adoption, for example, early adopters with first-mover advantage. Analyzing the impact of such policies would require extending the present framework to allow for household identity in social interactions and time to capture sequential moves. The development of such a framework is left for future research. Lastly, the policy analysis in this paper highlights the critical role informal and formal institutions jointly play in allocating resources in the absence of markets. There is considerable scope to explore policy responses to externalities discussed in this paper to alternative institutional structures.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

Market price data and additional code available upon request.

## Appendix A

### *A.1. Additional tables*

See Table A.1.

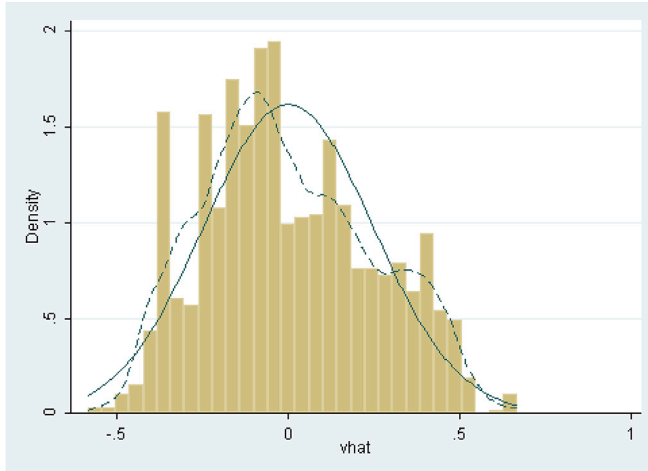



**Fig. A.1.** Density of $\hat{\upsilon}_{ig}$. *Note.* Kernel = Gaussian, bwidth = 0.0262.

**Table A.1**
Model fit by village: Average sanitation adoption at baseline.

| Village ID | Data[a] | | | Model[b] | |
|---|---|---|---|---|---|
| | (1) coef. | (2) 95% CI | (3) $p$-value | (4) est. | (5) diff. |
| 1 | 0.81 | [0.701, 0.919] | <0.001 | 0.89 | −0.08* |
| 3 | 0.14 | [0.076, 0.204] | <0.001 | 0.22 | −0.08** |
| 4 | 0.26 | [0.170, 0.350] | <0.001 | 0.20 | 0.06* |
| 5 | 0.19 | [0.039, 0.341] | 0.018 | 0.17 | 0.02* |
| 6 | 0.35 | [0.137, 0.563] | 0.004 | 0.12 | 0.23** |
| 7 | 0.18 | [−0.052, 0.412] | 0.148 | 0.11 | 0.07 |
| 8 | 0.20 | [−0.007, 0.407] | 0.072 | 0.21 | −0.01 |
| 11 | 0.20 | [−0.158, 0.558] | 0.314 | 0.05 | 0.15 |
| 13 | 0.71 | [0.554, 0.866] | <0.001 | 0.19 | 0.52** |
| 14 | 0.45 | [0.277, 0.623] | <0.001 | 0.43 | 0.02* |
| 16 | 0.94 | [0.821, 1.009] | <0.001 | 0.72 | 0.22** |
| 17 | 0.50 | [0.323, 0.677] | <0.001 | 0.24 | 0.26** |
| 19 | 0.05 | [−0.045, 0.145] | 0.305 | 0.05 | 0.00 |
| 20 | 0.63 | [0.289, 0.971] | 0.006 | 0.22 | 0.41** |
| 21 | 0.19 | [0.069, 0.311] | 0.003 | 0.17 | 0.02* |
| 23 | 0.69 | [0.459, 0.921] | <0.001 | 0.26 | 0.43** |
| 24 | 0.11 | [0.055, 0.165] | <0.001 | 0.10 | 0.01* |
| 25 | 0.20 | [−0.158, 0.558] | 0.314 | 0.12 | 0.080 |
| 26 | 0.63 | [0.507, 0.753] | <0.001 | 0.83 | −0.20** |
| 27 | 0.50 | [0.211, 0.789] | 0.005 | 0.55 | −0.05* |
| 28 | 0.38 | [0.037, 0.723] | 0.058 | 0.19 | 0.19 |
| 29 | 0.66 | [0.545, 0.775] | <0.001 | 0.16 | 0.50** |
| 30 | 0.65 | [0.521, 0.779] | <0.001 | 0.83 | −0.18** |
| 31 | 0.50 | [0.211, 0.789] | 0.005 | 0.35 | 0.15* |
| 32 | 0.46 | [0.272, 0.648] | <0.001 | 0.48 | −0.02* |
| 33 | 0.89 | [0.799, 0.981] | <0.001 | 0.88 | 0.01* |
| 34 | 0.88 | [0.765, 0.995] | <0.001 | 0.24 | 0.64** |
| 35 | 0.05 | [−0.021, 0.121] | 0.165 | 0.11 | −0.06 |
| 36 | 0.71 | [0.467, 0.953] | <0.001 | 0.74 | −0.03* |
| 37 | 0.15 | [0.045, 0.255] | 0.007 | 0.06 | 0.09* |
| 38 | 0.47 | [0.294, 0.646] | <0.001 | 0.75 | −0.28** |
| 39 | 0.88 | [0.718, 1.002] | <0.001 | 0.29 | 0.59** |
| 40 | 0.08 | [−0.010, 0.170] | 0.085 | 0.09 | −0.01 |
| 42 | 0.80 | [0.675, 0.925] | <0.001 | 0.75 | 0.05* |
| 43 | 0.71 | [0.367, 1.053] | 0.004 | 0.87 | −0.16* |
| 44 | 0.30 | [0.166, 0.434] | <0.001 | 0.35 | −0.05* |

*Note.* Eight villages out of 44 are excluded from the table where sanitation adoption was close to zero or one. All 44 villages are included in the model estimation in column (4). In this exercise the 95 percent confidence interval (data) may lie outside of [0,1] as a consequence of using a linear probability model.

*The parameter estimate in column (4) falls within the 95 percent confidence interval and the $p$-value of coef. is <0.05.

**The parameter estimate in column (4) falls outside the 95 percent confidence interval and the $p$-value of coef. is <0.05.

[a]Regression coefficients from a linear probability model (LPM), 95% confidence interval, and $p$-value.

[b]Model predicted sanitation level and pct difference.

### *A.2. Estimation*

This note describes the steps to construct $G^{(1)}_{\overline{d}}(\overline{d}_g)$.

1. Regress $\overline{d}_g = \varphi W_{ig} + \upsilon_{ig}$, where $W_{ig}$ denotes for each household $i$ all of the variables included in its information set except $\overline{d}_g$, and $\upsilon_{ig}$ is the residual/error.
2. Recover an estimate $\hat{\varphi}$ from Step 1 to construct fitted residuals $\hat{\upsilon}_{ig} = \overline{d}_g - \hat{\varphi}' W_{ig}$. Alternatively, plot $\hat{\upsilon}_{ig}$ and calculate for every point $a$ and a small distance $\triangle$ $-\left(\frac{f(a+\triangle)-f(a)}{\triangle}\right)\left[\frac{1}{\left(\frac{f(a+\triangle)+f(a)}{2}\right)}\right]$.

   Note that this method may be problematic at the tails of the distribution. Fig. A.1 plots the density of $\hat{\upsilon}_{ig}$.
3. Use the $\hat{\upsilon}_{ig}$ from Step 2 to construct a kernel density for $\hat{f}(\hat{\upsilon}_{ig}) = \frac{1}{Nh}\sum_{i=1}^{N} K\left(\frac{\hat{\upsilon}_{ig}-\hat{\upsilon}_{0g}}{h}\right)$, where $f(\hat{\upsilon}_{ig}) \approx f(\overline{d}_g|W_{ig})$ and where $N$ is the total sample size, and $h$ is the smoothing parameter or bandwidth.
4. Lastly, use the kernel density $\hat{f}(\hat{\upsilon}_{ig})$ to construct $G^{(1)}_{\overline{d}}(\overline{d}_g)$[16] where

$$G^{(1)}_{\overline{d}}(\overline{d}_g) = \frac{\partial \log f(\overline{d}_g|W_{ig})}{\partial \overline{d}_g} = \frac{1}{f(\overline{d}_g|W_{ig})} \cdot \frac{\partial f(\overline{d}_g|W_{ig})}{\partial \overline{d}_g}$$
$$= \left[\frac{1}{\hat{f}(\hat{\upsilon}_{ig})}\right]\left(\frac{\partial \hat{f}(\hat{\upsilon}_{ig})}{\partial \hat{\upsilon}_{ig}}\right) \tag{28}$$

## References

Abramovsky, L., Augsburg, B., Lührmann, M., Oteiza, F., Rud, J.P., 2023. Community matters: heterogeneous impacts of a sanitation intervention. World Development 165, 106197.

Aguirregabiria, V., Mira, P., 2007. Sequential estimation of dynamic discrete games. Econometrica 75 (1), 1–53.

Aguirregabiria, V., Mira, P., 2019. Identification of games of incomplete information with multiple equilibria and unobserved heterogeneity. Quant. Econ. 10, 1659–1701.

Arcidiacono, P., Ellickson, P.B., 2011. Practical methods for estimation of dynamic discrete choice models. Annu. Rev. Econ. 3 (1), 363–394.

[16] A potential disadvantage of this method that $G^{(1)}_{\overline{d}}(\overline{d}_g)$ comprises a ratio of kernel estimators, and problems may arise when the denominator kernel estimator is close to zero. This is more likely to occur in the tails of the distribution.

Asghar Zaidi, M., Hagenaars, A., de Vos, K., 1995. Poverty Statistics in the Late 1980s: Research Based on Micro-Data. Publications Office, European Commission and Eurostat.

Attanasio, O., Di Maro, V., Lechene, V., Phillips, D., 2013. Welfare consequences of food prices increases: Evidence from rural Mexico. J. Dev. Econ. 104, 136–151.

Augsburg, B., Rodriguez-Lesmes, P.A., 2018. Sanitation and child health in India. World Dev. 107, 22–39.

Bajari, P., Benkard, C.L., Levin, J., 2007. Estimating dynamic models of imperfect competition. Econometrica 75 (5), 1331–1370.

Bajari, P., Hong, H., Krainer, J., Nekipelov, D., 2010. Estimating static models of strategic interactions. J. Bus. Econom. Statist. 28 (4), 469–482.

Barnard, S., Routray, P., Majorin, F., Peletz, R., Boisson, S., Sinha, A., Clasen, T., 2013. Impact of Indian Total Sanitation Campaign on latrine coverage and use: a cross-sectional study in Orissa three years following programme implementation. PLoS One 8 (8), e71438.

Björkegren, D., 2019. The adoption of network goods: Evidence from the spread of mobile phones in Rwanda. Rev. Econom. Stud. 86 (3), 1033–1060.

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

Chesher, A., 1991. The effect of measurement error. Biometrika 78 (3), 451–462.

Chesher, A., Lancaster, T., Irish, M., 1985. On detecting the failure of distributional assumptions. In: Annales de L'Insee. JSTOR, pp. 7–45.

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

Coffey, D., Geruso, M., Spears, D., 2018. Sanitation, disease externalities and anaemia: Evidence from Nepal. Econ. J. 128 (611), 1395–1432.

Crocker, J., 2016. Teachers and sanitation promotion: An assessment of community-led total sanitation in ethiopia. Environ. Sci. Technol. 50 (12).

Crocker, J., Abodoo, E., Asamani, D., Domapielle, W., Gyapong, B., Bartram, J., 2016. Impact evaluation of training natural leaders during a community-led total sanitation intervention: A cluster-randomized field trial in ghana. Environ. Sci. Technol. 50 (16), 8867–8875.

Deaton, A., 1989. Rice prices and income distribution in Thailand: A non-parametric analysis. Econ. J. 99 (395), 1–37.

Deaton, A., 1991. Saving and liquidity constraints. Econometrica 59 (5), 1221–1248.

Gautam, S., 2019. Household (Under) Adoption of Sanitation: Externalities and Borrowing Constraints. Working Paper, Washington University in St. Louis.

Geruso, M., Spears, D., 2018. Neighborhood sanitation and infant mortality. Am. Econ. J.: Appl. Econ. 10 (2), 125–162.

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

Guiteras, R., Levinsohn, J., Mobarak, A.M., 2019. Demand Estimation with Strategic Complementarities: Sanitation in Bangladesh. CEPR Discussion Paper No. DP13498.

Hammer, J., Spears, D., 2016. Village sanitation and child health: effects and external validity in a randomized field experiment in rural India. J. Health Econ. 48, 135–148.

Hotz, V.J., Miller, R.A., 1993. Conditional choice probabilities and the estimation of dynamic models. Rev. Econom. Stud. 60 (3), 497–529.

Jia, P., 2008. What happens when Wal-Mart comes to town: An empirical analysis of the discount retailing industry. Econometrica 76 (6), 1263–1316.

McFadden, D., 1978. Modelling the choice of residential location. In: Snickars, F., Weibull, J. (Eds.), Spatial Interaction and Planning Models. North-Holland, Amsterdam, pp. 75–96.

Pakes, A., Ostrovsky, M., Berry, S., 2007. Simple estimators for the parameters of discrete dynamic games (with entry / exit examples). Rand J. Econ. 38 (2), 373–399.

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

Pesendorfer, M., Schmidt-Dengler, P., 2008. Asymptotic least squares estimators for dynamic games. Rev. Econom. Stud. 75 (3), 901–928.

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

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

Sweeting, A., 2006. Coordination, differentiation, and the timing of radio commercials. J. Econ. Manag. Strategy 15 (4), 909–942.

Todd, P.E., Wolpin, K.I., 2008. Ex ante evaluation of social programs. In: Annales D'Economie Et de Statistique. JSTOR, pp. 263–291.

Todd, P., Wolpin, K.I., 2018. Accounting for mathematics performance of high school students in Mexico: Estimating a coordination game in the classroom. J. Polit. Econ. 126 (6), 2608–2650.