SOCIAL NETWORKS AND SANITATION ADOPTION

By: Chima Rickards

March 2023

Supervisors: Dr Bansi Malde & Professor Zaki Wahhaj

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy, in the School of Economics, at the University of Kent.

Contents

List of Figures

List of Tables

Declaration

I hereby declare that this thesis is the result of my original work. I have acknowledged all main sources of help and have cited all sources used in this thesis. This work has not been submitted for any other degree or diploma. All data used in this research have been obtained in accordance with ethical guidelines, and where necessary, appropriate permissions and approvals have been obtained.

Acknowledgements

I am thankful to the School of Economics at the University of Kent for giving me the space to be all versions of myself. I am very grateful to my supervisors: Dr Bansi Malde for her patience and mercy towards me and commitment to this work, and Professor Zaki Wahhaj for his wisdom.

I am also grateful to my PhD cohort colleagues: Aniema, Kemi, Mumba and Lucas for providing me a sense of fellowship, and Bob Chisholm for offering me the required balance of vocational, psychological and spiritual mentorship. Our weekly conversations have been a vibe.

I am very thankful to the Quaker community in Canterbury for being a source of respite and belonging. It goes without saying, but thank you to family and friends for teaching me all I needed to learn along the way, and reminding me of what I behold.

Noted last, but of most significance, I give thanks to the Holy Spirit for constantly holding me in their light, most especially, when all I could witness was darkness.

Introduction

Improvements in sanitation is a key global development concern (SDG 2022). Lack of access to basic sanitation services leads to unfavourable social, economic and health impacts at both the micro and macro level (WHO 2022). This thesis makes a contribution to addressing sustained behaviour change in sanitation, by investigating key social structures perpetuating poor sanitation and exploring leverage points through social networks for altering social norms towards increased latrine usage, latrine ownership, and reduced open defecation (OD).

Beyond economic fundamentals, understanding key drivers of poor sanitation has fed a growing literature in social interactions, social networks and sanitation decision making. A growing amount of the contemporary literature has focused on the importance of sustaining behaviour change in the medium to long term (Gupta et al. 2019, Augsburg et al. 2022, Pakhtigian et al. 2022) and achieving community wide coverage, not just household level uptake of sanitation adoption (Cameron et al. 2021, Cameron et al. 2022). This speaks ever more to the importance of understanding which network structures offer the largest spill-overs in sanitation adoption. With this knowledge one can better understand key groups and reference points for targeting, that induce significant spill-overs which can help shift the norm of sanitation from open defecation to consistent latrine usage. This motivates the use of different reference points for peer effects in this thesis: village, caste[1], and distance.

Understanding the social interactions, social norms and social barriers that contribute to poor sanitation can play a crucial role in promoting sustainable behaviour change. By identifying relevant social factors and dynamics that influence sanitation practices, interventions can be designed to effectively target and address these underlying issues. This thesis contributes to that understanding. It provides insight into some underlying social structures and dynamics within communities and sanita-

[1] Caste is a system of social hierarchy common in South Asia; especially India. It is determined by birth and is characterised by socio-economic boundaries and restrictions shaping the way of life of individuals. This can include occupation, social interactions, and cultural practices

tion decisions. This can improve targeting of interventions that harness these social structures to enable sustained behaviour change.

All chapters utilise the same dataset collected on over 40,000 households, and Gram Panchayats (GP) located in rural Maharashtra, India. The sanitation outcomes analysed in all three chapters are the same: latrine ownership, latrine usage, and open defecation. While previous studies have explored sanitation issues in India, this thesis stands out due to the substantial size and scope of the dataset, contributing evidence to the existing literature on sanitation behaviours in rural India.

The value of focusing on rural environments for understanding social effects comes from the fact that rural environments more often than not, have lower rates of access to basic sanitation services compared to urban environments. This is due to the increased presence of various physical institutions and infrastructure that agglomerate in urban areas compared to rural areas - especially in the global south. Focusing on rural areas expresses areas of greater need, especially due to reduced investment in rural environments. Additionally, focusing on rural environments is partially motivated by the simplicity of identifying reference groups in these areas. Social reference groups in rural environments are more geographically dense, and overlapping compared to urban environments (Kresch et al. 2020).

The value of using the Indian context comes from the stringent norm of open defecation within the country that has rates in paradox to the stage of economic maturity the country has reached (Drèze and Sen 2013). This is tied to social norms around sanitation derived from historic institutions within the caste system (Coffey and Spears 2007). India also provides useful context for understanding behavioural drivers in sanitation given its relatively low rates in sanitation coverage. As of 2020, only 46% of the country had coverage of safely managed sanitation services (WHO-UNICEF JMP 2021).

The country has had several waves of policy campaigns, programmes and instruments to increase sanitation and related WASH activities in the country. These include: Central Rural Sanitation Programme (1986), Total Sanitation Campaign (1999-2012), Nirmal Bharat Abhiyan (Clean India Campaign) (2012), Swachh Bharat Abhiyan (Clean India Mission) (2014-2019), and most recently Open Defecation Free

(ODF) Plus (2020). Policy programmes have included subsidy provision for latrine building, to behaviour change programmes such as Community Led Total Sanitation (CLTS).

India has seen some of the highest drops in OD behaviour since 2015, but the rate of progress nationwide is highly divergent; where some regions are near universal coverage, others seriously lag behind with still very low coverage (WHO-UNICEF JMP 2021).

Even though between 2015-2020, the nation experienced a 15% increase in at least basic sanitation[2] progress is too slow, especially in rural India, to meet the 2030 SDG target 6 of clean water and sanitation for all.

By looking at the effect of social networks on rural sanitation behaviours, one can understand with greater clarity the role different social groups have as a social multiplier on household and collective sanitation outcomes. The methodologies used across chapters differ but align with notable literature in the field. All chapters draw from Brock and Durlauf (2001, 2007) discrete choice framework for social interactions and binary outcomes. Empirical application in chapters 1 and 2 incorporate the work of Gautam 2018. Empirical application in chapter 3 incorporates the work of Lee at al. 2014, which offers a framework for the context of heterogeneous rational expectations; a situation where networks are incomplete. Furthermore, all chapters focus primarily on endogenous social effects. That is, how group behaviour affects individual behaviour. The identification of social effects builds upon the non-linear nature of discrete outcomes, drawing inspiration from the work of Brock and Durlauf. However, in Chapter 1, where the neighborhood network is complete, there are potential concerns regarding identification. On the other hand, in Chapters 2 and 3, where social networks exhibit more heterogeneity, identification becomes less of a concern.

Chapter one uses a complete network defined by the entire village. It looks at neighbourhood peer effects on sanitation outcomes. It follows the work of Brock and Durlauf (2001,2007) discrete choice models for social interactions. Using the entire

[2]Basic sanitation is defined as having access to facilities for the safe disposal of human waste (faeces and urine), as well as having the ability to maintain hygienic conditions, through services such as garbage collection, industrial/hazardous waste management, and wastewater treatment and disposal.

village as the reference group, the chapter estimates endogenous and contextual social effects for three different sanitation outcomes: latrine ownership; latrine usage, and Open Defecation (OD). The chapter finds complementary social effects to exist for all three behaviours. In the range of 8% increased likelihood for a household choosing an outcome, for every 10% increase in neighbourhood households engaging in the same outcome.

Chapter 2 uses networks referenced by caste groups. This includes broad caste categories, and jati (sub-caste) groupings. The motivation for this is the use of a key social identifier of relevance to the Indian context. Networks are complete within caste and within jati, bounded by the village. Results show that positive endogenous social effects exist for these groups at similar rates found in chapter 1. However, effects reduce in size once correlated factors are controlled for by including village and caste (or jati) fixed effects. They show an approximate 5%-7% likely increase in own household sanitation behaviours when own caste group proportion of the same behaviour increases by 10%. The chapter also looks at how these effects differ for specific caste groups, and measures endogenous social effects between caste groups. It finds that Forward Caste and Other Backward Class groups exhibit the highest in-group endogenous social effects. As for between castes, complementary endogenous social effects also exist. Highest other group endogenous effects are experienced by Scheduled Tribe and Denotified Tribe households. For most households, these other group behaviour effects are less than own caste group behaviour effects. The exception are Denotified Tribes who present the highest other group endogenous social effects. These other group effects for Denotified Tribes surpass own group endogenous social effects for latrine ownership and usage.

Chapter 3 compares three spatial networks to assess their relative impact on sanitation behaviours. 30 metres, 100 metres, and 300 metres are the chosen networks in line with realities of the data, and context of social interactions and sanitation outcomes in these environments. Results show that endogenous social effects exist for all three networks. However, the direction and magnitude of effects differ depending on sanitation behaviour and network observed. For open defecation, larger networks feed ever increasing positive endogenous social effects. However, for latrine ownership and usage, endogenous social effects are only positive at the 30-metre network, becoming negative for the 100-metre and 300-metre networks. Results also show complementarity in social effects on the basis of gender. Female household heads tend to feed greater likelihood of adopting complementary sanitation behaviours.

Policy implications of this work emphasise the importance of interventions that focus on behaviour change at the community level, and not just the household level to ensure sustained behaviour change. The work also brings light to the key groups and reference points for effective targeting. The thesis shows that caste, sub-caste (jati), and spatial clusters could be useful target points for policy interventions. Policy aimed at these social networks can facilitate and reinforce positive sanitation behaviour. This thesis also highlights the value of tailoring interventions to different sanitation behaviours. Chapter 3 especially shows the value of targeting latrine ownership, usage and OD differently based on the differing social dynamics associated with each behaviour. Finally, this thesis brings to light the importance of considering gender dynamics. The analysis in chapter 3 reveals that female household heads play a significant role in influencing sanitation behaviours. Policies should consider gender dynamics and empower women to lead and promote complementary sanitation behaviours within their communities.

References

Augsburg, B., Bancalari, A., Durrani, Z., Vaidyanathan, M., White, Z., (2022) When nature calls back: Sustaining behavioral change in rural Pakistan, Journal of Development Economics, Volume 158, 2022, 102933

Brock, W.A. and Durlauf, S.N. (2001). Discrete Choice with Social Interactions. The Review of Economic Studies [Online] 68:235–260.

Brock, W. and Durlauf, S. (2007). Identification of binary choice models with social interactions. Journal of Econometrics [Online] 140:52–75.

Cameron L, Santos P, Thomas M, Albert J. Sanitation, financial incentives and health spillovers: A cluster randomised trial. J Health Econ. 2021 May;77:102456.

Cameron, L., Gertler, P., Shah, M., Alzua, M.L., Martinez, S., Patil, S. (2022) The dirty business of eliminating open defecation: The effect of village sanitation on child height from field experiments in four countries, Journal of Development Economics, Volume 159, 102990, ISSN 0304-3878

Coffey, D. and Spears, D. (2017) Where India Goes: Abandoned Toilets, Stunted Development and the Costs of Caste. New Delhi: HarperCollins Publishers India.

Drèze, J. and Sen, A., (2013). An Uncertain Glory: India and its Contradictions. Princeton University Press, Princeton, NJ, 2013. 434 pp. (Hardback). ISBN 978-0-691-16079-5

Gupta A, Khalid N, Desphande D, Hathi P, Kapur A, Srivastav N, Vyas S, Spears D, Coffey D (2019), Changes in Open Defecation in Rural North India: 2014-2018 ,IZA Discus sion Papers No. 12065, Institute of Labor Economics (IZA), Bonn

Kresch, E., Lipscomb, M., Schechter, L. 2020, Externalities and Spillovers from Sanitation and Waste Management in Urban and Rural Neighborhood, Applied Economic Perspectives and Policy (2020) volume 42, number 3, pp. 395–420

Lee, L., Li, J., Lin, X. (2014). Binary Choice Models with Social Network under Heterogeneous Rational Expectations. The Review of Economics and Statistics, 96(3), 402–417.

Pakhtigian, E., Dickinson, K., Orgill-Meyer, J., Pattanayak, S., 2022. "Sustaining latrine use: Peers, policies, and sanitation behaviors," Journal of Economic Behavior and Organization, Elsevier, vol. 200(C), pages 223-242

The Sustainable Development Goals Report 2022. United Nations Statistics Division 2022.

WHO/UNICEF JMP (2021), Progress on household drinking water, sanitation and hygiene 2000-2020 Five years into the SDGs.

1 Neighbourhood effects and sanitation adoption

Abstract

This chapter measures the impact of neighbourhood endogenous social effects on sanitation behaviour in rural Maharashtra, India. The theoretical framework for identification follows that of Brock and Durlauf's (2001, 2007) social effects in a non-linear binary response framework. Multiple equilibria are found in this setting, and estimation is undertaken using two different methods to select the equilibrium: an equilibrium-in-data rule, and predicted probabilities using a logit functional form. Endogenous social effects are found to exist at the neighbourhood level. To address the existence of neighbourhood level unobservable factors, an instrumental variable approach offered by Shang and Lee (2011) is used.

1.1 Introduction

There are significant disparities in global sanitation coverage and access. As of 2020, only 54% of the global population (4.2 billion people) use safely managed sanitation services,[3] with least developed countries estimated to have only 26% of safely managed sanitation services (WHO/UNICEF JMP 2021). This leaves 3.6 billion people without access to safely managed services; comprising of 494 million people still practising open defecation; 92% of which reside in rural areas.

In areas where sufficiently sophisticated institutions are not present to provide population wide access to sanitation as a public good, sanitation adoption becomes more of a personal choice. The questions that underlie an agent's utility decision around sanitation in areas where access and coverage is relatively low, are very different to areas where sanitation access is widespread. This difference in context can lead to aggregate outcomes of sanitation behaviour that are counter-intuitive to outcomes that would have been chosen by a social planner. The epidemiological importance of sanitation makes it a public health concern, making adequate sanitation a public good (Dickinson and Pattanayak, 2009).

Policy objectives and implementation programmes have been put in place to address this specific public issue, and policy actions taken to nudge private sanitation decision making in the direction of increased coverage and access. This includes international level initiatives such as The United Nations' Sustainable Development Goal 6.2 which aims to achieve access to adequate and equitable sanitation and hygiene for all, and to end open defecation by 2030. National initiatives include India's policy waves and campaigns such as the Swachh Bharat Abhiyan (Clean India Mission) launched in 2014 which aimed to eliminate open defecation and improve sanitation services in the country.[4] Other countries in the developing world have also had their own versions and waves of National Sanitation Campaigns (NSC) over

[3] According to the WHO this is defined as "The population using an improved sanitation facility that is not shared with other households and where excreta are safely disposed of in situ or treated off site. Improved sanitation facilities include flush/pour flush to piped sewer systems, septic tanks or pit latrines; pit latrines with slabs (including ventilated pit latrines), and composting toilets."

[4] Other policies in the country include the Central Rural Sanitation Programme (1986), Total Sanitation Campaign (1999-2012), National Urban Sanitation Policy (NUSP) (2008) and Nirmal Bharat Abhiyan (Clean India Campaign) (2012).

time.[5]

Empirical attempts to test the benefits of public and private policies on sanitation improvement have succeeded to some degree (BenYishay et al., 2017, Garn et al. 2017, Freeman et al. 2017, Caruso et al. 2022, Augsburg et al. 2021, Deutschmanm et al. 2023). Nonetheless, uptake has been limited notwithstanding attempts made to address matters such as financing and knowledge around sanitation (Venkataramanan et al. 2018, Cameron et al. 2021). One possible reason behind this is the entrenchment of certain sanitation practices, which overtime have persisted in social behaviour, even in the presence of factors intended to improve sanitation adoption. One such social practice around sanitation is open defecation (OD).

The persistence of OD in many parts of the global south, especially in rural and urban slums (WHO/UNICEF JMP 2021) suggest that it is a social norm. This is evidenced by work which shows a tendency for households to slip back into OD even after policy assistance that enabled them to own latrines (Coffey and Spears 2017, Orgill-Meyer et al. 2019). This suggests sanitation behaviour change efforts may improve if operated via mechanisms that feed these social norms; outside economic fundamentals such as prices, supply etc (Moffitt, 1998). Social interactions are one such out-of-market mediator of agent behaviour (both economic and otherwise). Literature such as Chuang and Schechter (2014) argue that social interactions are especially relevant in developing countries where missing markets exist. They argue social links are necessary and frequently used by individuals in developing countries for things such as financial insurance, informal loans, information sharing, risk-sharing, learning, adoption of new technology and more. Therefore these social links can be used to nudge social norms towards increased ownership and usage of latrines.

Social effects are noted to play a role around sanitation behaviour in developing country contexts. Coffey and Spears (2017) for example ascertain that open defecation issues in India are not solely due to poverty, education, governance, quality of latrines provided by governments or other economic factors that affect access to

[5] For example Bangladesh's National Sanitation Campaign 2003-2006; Tanzania's 'Nipo Tayari' (I am ready) national sanitation campaign 2017; Peru's National Sanitation Plan (2006-2015).

adequate sanitation. They note that social factors perpetuate behaviours such as OD even when access to a community or private latrine is available. One's peers, or who they socially interact with can influence sanitation practice through social learning or social conformity. Guiteras et al. (2019) and Gautam (2018a) for example were able to empirically show that social interactions act as a means of influencing sanitation adoption. These papers used neighbourhood and village level groupings as the basis for quantifying social effects. This motivates the current chapter. Using a dataset from rural Maharashtra, social effects in sanitation adoption are measured within rural administrative villages (Gram Panchayats). The social interactions of interest are neighbourhood effects. The chapter seeks to test the presence of any neighbourhood effects on individual household sanitation behaviour (latrine ownership, latrine usage, and open defecation) coming from neighbourhood characteristics and behaviour, and measure its magnitude. The novelty of this chapter is the range of sanitation behaviours modelled, and the benefit of a large dataset of almost 40,000 households in assessing impacts.

Social effects can operate through exogenous social factors, such as the socio-economic or demographic context of the social setting; or endogenous social factors which are the behaviours of others within the social setting. These two kinds of social effects are modelled and estimated separately, to show whether, and to what degree they exist at a neighbourhood level when it comes to sanitation behaviour. However, endogenous social effects are of particular focus here. This is because the presence of endogenous social effects reveal social multipliers in collective behaviour adoption. This means that households can be influenced to adopt better sanitation practices even without being directly targeted. Additionally externalities in sanitation behaviour further motivate the reasons for social drivers in shaping sanitation outcomes at neighbourhood levels. Increases in overall coverage of latrine usage is of benefit to all in a neighbourhood (Cameron et al. 2022).

The theoretical framework for identification follows that of Brock and Durlauf's (2000, 2001 and 2007) (B&D) non-linear binary response model of social effects. Endogenous social effects are known to face an identification issue known as the reflection problem (Manski, 1993). Brock and Durlauf address this by using a non-linear framework to identify endogenous social effects. Identification of social effects is feasible through the non-linearity that exists between household sanitation behaviours and individual and group level characteristics. Equilibrium is found on the basis of matching rational expectations in agent's perception of each other's behaviour. Multiple equilibria are said to exist if endogenous social effects are high enough in magnitude. This potentially disrupts the ease of finding one solution to the equilibrium equation. Due to the computational burden of arriving at an equilibrium choice probability for each group before determining social effects, two estimation procedures are followed to estimate the model. The first involves the data in equilibrium assumption (Gautam 2018a) where equilibrium choice probabilities of adoption for each agent household is derived from their actual choices in the data. The second estimation procedure involves a two step process in which the model is estimated without group behaviour in the first stage, and then predicted probabilities are plugged into to the equilibrium equation for the variable representing group behaviour. The computational burden in determining an equilibrium, and uncertainty in defending a specific equilibrium selection rule over another motivates the use of these two estimation options as simplified means of expressing the equilibrium.

The data used to estimate the model consists of household level surveys across 107 villages in rural Maharashtra, India. The suitability of having census level data at village and sub-village level allows one to reference the entire village as a suitable social reference group to capture neighbourhood peer effects. Results show that neighbourhood effects exist for all three sanitation behaviours, with OD exhibiting the largest social effects. Additionally, endogenous effects specifically exhibit higher magnitudes than other social (contextual or correlated) effects. This suggests the possibility of multiple equilibria.

A key assumption in the identification strategy of the model, is the non-existence of group level unobservables. When this assumption is relaxed, identification fails. In estimating the model, not accounting for group level unobservables may bias the endogenous effects estimated. This is because ignoring the effect of group level unobservables means unknown (to the researcher) group based factors that influence group behaviour, but not captured by the data used to test the model are ignored.

Shang and Lee (2011) offer an instrumental variable (IV) estimation strategy that is tested on the data to address the issue of group level unobservables. This involves the use of IVs to capture group level behaviour, which is estimated on a group fixed effects equation. Two types of IVs are created. Results show that most IVs provide similar results to the model that does not separately measure group level unobservables.

The next section will review some literature on social effects and decision-making. It highlights the reasoning behind, and the power of social interactions in decision-making. This specifies on some work using neighbourhood effects to test sanitation adoption. Section 1.3 will look into the theoretical framework of the proposed household decision-making model. This is used as a basis to estimate social effects on sanitation behaviour. Section 1.4 discusses data and estimation. Section 1.5 notes results, and discusses the use of the IV. Section 1.6 concludes the chapter.

1.2 Literature Review

1.2.1 Understanding social effects in decision-making

Manski (2000) defines social interactions as non-market based interactions that affect agent decision-making via constraints, preferences, and expectations. The way of modelling how social interactions impact decision making used in this chapter is via interdependent preferences that agents incorporate into their utility function. An example of this was modelled by Brock and Durlauf (2000, 2001,2007). This models agent decisions in a way that incorporates caring about the same decision made by others. It incorporates the expected choices of others into one's utility framework when making the same choice. This form is analogous to strategic interactions with incomplete information in a game theoretic setting (Jackson and Zenou 2015).

Empirical literature has provided quantitative means of expressing the mechanisms through which social interactions impact decision-making. Peer effects as a means of modelling social interactions for example has been applied to literature on educational outcomes. This was popularised by the controversial Coleman report (Coleman, 1966), which found that educational outcomes for black students, was positively correlated with the proportion of students in their school that were white. The idea of peer effects follows that the outcomes of one's decisions correlate similarly with others they tend to associate with. As an example, a group of friends that smoke might make the activity of smoking a part of the group's way of socialising. As a result, the idea of smoking may seem more appealing to friends that do not smoke, because it offers a point of connection in their social engagements.

Neighbourhood effects as another social interaction measure were popularised through observations of neighbourhood segregation in the U.S. It has also been used to understand outcomes such as neighbourhood poverty and community level provision of publicly beneficial goods. Popularised by the work of (Wilson, 1987), the idea follows that a community facing common environmental factors are similarly influenced by these factors. This can lead to similar decision making on aggregate, leading to less varied average outcomes in those neighbourhoods. Examples include quality of schools, crime rates or employment opportunities.

One means of modelling social effects can be simplified into two main constructs. The first are exogenous social effects. These can be split into those effects mediated by communal circumstances – that is correlated factors, or those effects drawn from average demographic similarities - these are contextual factors. The other are social effects mediated by social behaviour; termed endogenous effects. Manski (1993), made one of the earliest attempts within economics to model these two kinds of social effects. Within that formulation, social effects are derived from social groups. Agents are grouped into a social dimension based on the connections that bind them, and their decision-making is then based not only on their private individual characteristics, but also group characteristics, and group choice probabilities regarding the same decisions to be made.

The linear model has been the baseline functional form of social interaction modelling for some time, but it has also revealed key issues around identification and estimation of the causality of social interactions on decision-making. One of these is the reflection problem which highlights the difficulty in separately identifying exogenous and endogenous social effects because the social group behaviour is derived from the contextual characteristics. As an alternative, a non-linear model, such as a binary response model has been said to also facilitate the theoretical modelling of social interaction effects. For this setting, the Brock and Durlauf (B&D) logistic model is used as the framework for modelling sanitation decisions. The non-linear framework with a few added assumptions enables separate identification of exogenous and endogenous social effects.

1.2.2 Neighbourhood effects and adoption of new technology

Social interaction effects are noted to play an important role in adoption of new technology, especially when applied to rural and agricultural developing country environments. Despite the mechanisms used, the same idea remains – that social interactions have something to offer in inducing individuals to behave in certain ways. BenYishay and Mobarak (2014) argue that social ties are credible sources of information about new technologies, which can induce adoption of said technologies. In some rural environments where open defecation is a norm, adoption of a latrine can be akin to adoption of a new technology. Latrines in whatever forms they come have to be installed. New owners must learn usage, upkeep and replacement. If social learning, desire to conform or other social mechanisms are key for sanitation adoption, appreciating relevant social ties used to spread information is useful for facilitating latrine adoption.

Conley and Udry (2010) for example use survey data to construct information dissemination ties amongst farmers, as the relevant social group of influence. They showed increased likelihood of adopting a new agricultural technology for farming in Ghana. They based their social ties on informational links defined by who agents know and talk to about farming. They found these social interactions to impact the likelihood of adopting the new technology through social learning as the mechanism for decisions made around agricultural input use.

Another work that highlights the importance of social interactions for facilitating technology adoption decisions is Beaman et al. (2021). Through a randomised control trial (RCT) in rural Malawi, the authors used threshold based diffusion models to show that an agent needs to be aware of adoption from multiple sources, before adopting a new agricultural technology themselves. The social networks in their work was based on social network census data in which households reported names of people they consult to make agricultural decisions.

A unifying note is that these literatures utilised detailed surveys to accumulate social network data relevant to the outcomes tested. Given the higher budget and time costs entailed in attaining such detailed network data, other more cost effective proxies have been identified to model social groups, and identify social interaction effects. One such proxy is neighbourhood. Neighbourhoods act as a proximity based social group dimension which is especially salient in rural areas where there is greater overlap (compared to urban areas) between one's physical neighbours, work associates, and friendship ties, making social reference groups more geographically dense in rural environments (Kresch et al. 2020). Literature has shown that neighbourhood based peer effects impact individual agent behaviours (Geruso and Spears, 2017) and this translates into impacts on sanitation behaviour (Guiteras et al. 2019).

The added value of focusing on the neighbourhood as the reference point for social effects is that it is also a relevant and normally utilised domain for policy implementation. Therefore, understanding neighbourhood spillovers in sanitation behaviours can help inform neighbourhood based policy applications. Examples of such policy measures include directing resources towards neighbourhood-focused infrastructure development, implementing sanitation campaigns tailored to specific neighbourhoods, and facilitating community-led sanitation programs that are designed and applied within the neighbourhood context. These policy measures aim to address the unique challenges and opportunities present at the neighbourhood level, thereby maximising their effectiveness in improving sanitation conditions by capitalising on the potential spillover effects of adoption within the neighbourhood.

1.2.3 Social effects and Sanitation

Social effects have been noted to operate either via a substitution or complementary aspect. Complementarity follows the idea that the marginal utility one derives from making a decision such as latrine adoption is positive with the number of others also adopting. An example could come from social learning effects – seeing other households adopt latrines could allow one to better understand how the product works, its use, upkeep etc and thus would induce them to also adopt. Substitution effects relate to the opposite case; the marginal utility of an agent's decision is negatively correlated with the number of others making the same choice. An example could be in the form of implicit free riding. One might decide to share the latrine of other households, instead of installing their own, creating strategic substitutes in latrine ownership. Alternatively, one might choose to continue to open defecate, but benefit from the epidemiological improvements that come from the reduced open defecation behaviour of others generating strategic substitutes in latrine usage. It is thought that whether strategic complements or substitutes exist in sanitation adoption is an empirical question (Guiteras et al. 2019). However, to date most literature has shown strategic complements in sanitation outcomes. For example, Guiteras et al. (2019) using an RCT procedure in rural Bangladesh found that the demand for latrines operate as a strategic complement. They estimated a structural model of utility in which household latrine adoption was interdependent with the share of others within a neighbourhood that also adopted. They found complementary social effects for latrine adoption at a neighbourhood level. However, the paper did not tease out the difference in exogenous or endogenous social effects, but instead created social network ties to test potential mechanisms for the neighbourhood social effects found. They suggested that social learning might have been the mechanism through which the social spillovers found were induced.

Geruso and Spears (2017) using a number of empirical procedures to uncover the Muslim infant mortality advantage in India were able to show that open defecation (OD) rates were socially influenced by the religious composition of others within a neighbourhood. They ascertained that OD was lower in neighbourhoods that had a higher share of Muslim households showing the influence a neighbourhood's religious composition had on socially influencing sanitation practice.

Gautam (2018a, 2018b) empirically show complementary spillovers in sanitation adoption in rural India. Like others, these papers used the village as the relevant reference group of social influence for sanitation adoption. BenYishay et al. (2017) is one of the few exceptions showing negative peer effects in sanitation behaviour. The authors' study on the effect of microcredit on latrine adoption in rural Cambodia revealed negative spillovers from latrine purchases of other households. Through the use of variations in latrine purchases, they showed that increases in the purchase of latrines by neighbours were associated with lower installation rates and reduced latrine coverage. In essence, their findings provide evidence of strategic substitutes in sanitation behaviour.

The sanitation behaviours modelled in this chapter are latrine adoption, latrine usage and any open defecation. The use of these three overlapping categories in representing a form of sanitation behaviour, reduces the risk of misspecification of the model in-case strategic substitutes exist in sanitation. For example, if strategic substitutes were at play with latrine ownership due to free riding off the use of others' latrines, then usage would remain complementary behaviour. Additionally, there is a known discrepancy between ownership and usage; in which some households own latrines but do not use them for their intended purposes (Coffey and Spears 2017). While this chapter does not explicitly differentiate between private and shared latrines, the conceptual framework employed for analysing various sanitation behaviours is expected to capture some distinction between substitution and complementary effects (if they exist).

1.3 Theoretical Framework

As previously noted, the model adopted follows B&D's binary response models in the presence of strategic interactions. It is a framework used to analyse individual decision making. The aim of the model is to identify and estimate the social effects of a reference group. The reference group in this case is a type of administrative village akin neighbourhood[6] in rural India known as Gram Panchayat (GP), on three separate sanitation behaviours: latrine ownership; latrine usage; and any OD. The terms GP and village will be used interchangeably throughout this chapter.

A static random utility model is presented. There are a finite number of decision-

[6] In the data, neighbourhood sizes range from 100+ to almost 500 households.

making households that represent the population N:

$$i = 1, ..., N$$

Where i is each household. The sample is partitioned into sub-populations reflecting the social groups. Each household belongs to a group (g), and these groups are indexed by:

$$g = 1, ..., G$$

with the number of agents per group represented by:

$$I_g$$

The total number of agents per group is such that:

$$\sum_{g=1}^{G} I_g = N$$

i.e. the sum of the number of agents per group, across groups, makes up the total population. The group being the GP.

Each household faces a finite choice set:

$$\omega_i = \{-1, 1\}$$

reflecting a binary choice between one sanitation behaviour and its opposite conduct, such as

$$\omega_i = \begin{pmatrix} 1 & \text{if HH}_i \; \textit{uses (owns) a latrine} \\ -1 & \text{if HH}_i \; \textit{does not use (own) a latrine} \end{pmatrix}$$

where HH_i refers to the individual decision making household.

The binary choice model is based on a random utility value function:

$$V(\omega_i) = u(\omega_i) + S(\omega_i, m_i^e(\omega_i) + \varepsilon(\omega_i) \tag{1}$$

Where $V(\omega_i)$ is the overall utility an individual receives from a making a choice on sanitation.

$u(\omega_i)$ is the individual private deterministic utility from the choice.

$S(\omega_i, m_i^e(\omega_i))$ is the social utility element. It relates the individual household choice to an expectation by the household on the choice of everyone else in their reference group. In other words it represents the conditional probability measure that the decision maker places on the decision of others when making their own choice. Where this may exhibit increasing differences on the decision of others, the social utility component becomes directly dependent on:

$$m_{ig}^e = (I_g - 1)^{-1} \sum_{j \neq i} m_{(i,j)}^e \tag{2}$$

Which is the household's subjective expectation on the outcome of other household's (within their group) choices. This is explained in greater detail later on, as the basis of rational expectations that lead to equilibrium outcomes.

$\varepsilon(\omega_i)$ is the individual error term.

Each household's characteristics are represented by:

$$(x_{ig}, z_g, \varepsilon_i(\omega_i))$$

which is additively segmented into their own individual characteristics (x_{ig}) such as age, characteristics shared by all members of the group (z_g) such as the caste composition in the neighbourhood, or village level availability of certain sanitation related services, and a preference shock on sanitation behaviour ($\varepsilon_i(\omega_i)$).

Household characteristics are grouped into a common knowledge variable (k_{ig}):

$$(x_{ig}, z_g) = k_{ig}$$

with:

$$(k_{ig}, k_{-ig}) = k_g$$

and a private variable denoted by $\varepsilon_i(\omega_i)$

1.3.1 Assumptions

1. Strategic complementarity:

$$\frac{\partial^2 S(\omega_i \overline{\omega}_{-ig})}{\partial \omega_i \overline{\omega}_{-ig}} = J > 0 \tag{3}$$

The above expresses the relationship between household behaviour, and reference group behaviour i.e. interdependence of household decision making. Strategic complementarity makes the directional relationship between these two variables positive. This means that a household is more likely to choose a behaviour if other members of their reference group carry out the same behaviour. This probability grows as the share of group members making the same choice increases.

2. Logistic distribution of difference in error terms for binary choice:

$$Prob(\varepsilon_i(-1) - \varepsilon_i(1) \leq \omega_i) = \frac{1}{1 + exp(-(\beta_0 + \beta_1 x_{ig} + \beta_2 z_g + \beta_3 \omega_g))} \tag{4}$$

3. Error terms are independently and identically distributed (i.i.d) within and across the reference groups.

4. Non-cooperative setting of decision making. This means agents do not coordinate their decision making, but make their decision based on the expectation of the mean value of group decision which is independent of the realisation of group behaviour (ω_g). The value of specifying non-cooperate decision making is that it en-

ables one to model social interaction effects as a strategic behaviour for individual, self-interested agents. However, this may be a strict assumption in settings where frequent social interactions between households allow for collective decisions, such as building shared toilets. Alternative models considering cooperative decision-making processes could provide a more nuanced understanding of social effects[7].

5. Random Assignment which means there is no self-selection into the reference group of interest. The validity of this assumption in a neighbourhood setting, is arguable to the degree that there is a limit to the self-selection choice behind living in a particular neighbourhood. The neighbourhoods people reside in are usually not entirely exogenously determined. However, in this setting it may be a reliable assumption to make when carrying out a cross sectional analysis on the basis that households do not often move residential location in rural areas. Rural-rural or urban-rural migration is less common than urban-urban or rural-urban migration. This allows a static view of random assignment plausible. Additionally 90% of households in the dataset have resided in the same neighbourhood their entire lifetime. This means the initial decision to reside there was not chosen by them, providing some degree of exogeneity in household location. Furthermore, as Pakhtigian et al. 2022 note, it is unlikely that endogeneity in household location may be directly linked to sanitation preferences.

6. No group level unobservables which relate to shared aspects of the group that are not observed to the researcher but may be correlated with the sanitation behaviour decision. This is a strict assumption as it means there are no unobservable village level factors that could shape sanitation outcomes. This could include the presence of village level infrastructure that facilitate sanitation adoption or community level willingness to contribute towards a public good such as sanitation. Relaxation of this assumption will be addressed in more detail later on with the use

[7] Data is not available on the frequency and nature of social interactions in the specific setting to ascertain the degree to which cooperative decisions were made on sanitation adoption between households. What is clear is that given caste based dynamics in the settings, it is likely that any decisions made would have existed solely within caste groups.

of Instrumental Variables (IV).

1.3.2 Identification

In order to ensure the model is identified, the parameters ought to be deducible from the distribution of regressors, and distinguishable from any other observationally equivalent model that could depict a different causal relationship to that sought.

B&D makes assumptions that allow sufficient variation between group characteristics and group behaviour which use the non-linearity that exists between individual behaviour (ω_{ig}), average group behaviour ($\overline{\omega}_{-ig}$), and group characteristics (z_g) enabling separate identification of exogenous and endogenous social effects. This non-linearity is inherent in the outcome variable, which pertains to sanitation behaviour and is characterised as a binary choice between (e.g. owning (1) or not owning (0) a latrine). The binary nature of the outcome, combined with the underlying assumptions, serves to disrupt the linear interdependence that would arise when considering individual characteristics, average group characteristics, and group behaviour.

In other words B&D's assumptions create enough diversity in both group characteristics and behaviour. This diversity allows them to examine how individual behaviour, the average behaviour of the group excluding the individual, and group characteristics interact in a non-linear manner. By focusing on a binary decision, and incorporating their assumptions, they are able to avoid a simple linear relationship between group characteristics, and group behaviour; allowing separate identification of exogenous and endogenous social effects. [8]

Despite the strong assumptions used for identification, the attempt to use alternative methodologies to identify social network effects in this context encounters challenges due to the presence of a complete network, where all households are directly interconnected within the network in this neighbourhood context.

The use of a complete network poses methodological difficulties when attempting to disentangle the effects of social interactions. In a complete network, every household is connected to every other household, creating a dense web of social ties.

[8] The difficulty of utilising some other newer methods to identify social network effects in this setting comes from the use of a complete network in which all households are directly linked to each other within the network.

This high interconnectedness and assumed homogeneity of agents in the neighbourhood network complicates the identification of individual and group-level effects using other modern methods since it becomes challenging to isolate the impact of a specific social connection or distinguish between endogenous and exogenous influences (the issue in linear settings which the B&D method tried to resolve). In such a setting, alternative approaches that rely on specific network structures or assumptions may not be applicable or be heavily computationally burdensome.

There are certain cases in which the presence of multiple equilibria can affect identification. As multiple equilibria can be a feature of social effects models, this queries the validity of identification. As noted by B&D this not an issue because agents know the equilibrium they exist in due to the symmetry of the rational expectations (m^e_{ig}) they have on group behaviour (m_g), which is a moment based on observed variables. Therefore observational equivalence is met for identification.[9] Multiple equilibria can pose a challenge when estimating the model because an equilibrium selection rule must be determined.

1.3.3 Empirical tractability

Determination of sanitation choice depends on a number of factors such that the adoption decision follows the following specification of utility outcomes[10]:

$$u_{ig}(1) - u_{ig}(0) = \beta_0 + \beta_1 x_{ig} + \beta_2 z_g + \beta_3 \overline{y}_{-ig} - \varepsilon_{ig} \tag{5}$$

Utility difference between adoption $u_{ig}(1)$ and non-adoption $u_{ig}(0)$ is based on 4 main factors:

- Individual Characteristics x_i
- Group Characteristics z_g which highlight contextual factors and observed correlated factors

[9] Infact B&D 2007 note that multiple equilibria may facilitate identification in some cases due to pattern reversals in the outcomes that are counter-intuitive to the expected direction of certain regressors, meaning social effects are strong enough to push agents from one equilibrium to another.

[10] Where now decision ω_i (-1,1) is converted to $y_i(0,1)$ for empirical simplicity

- Average Group Behaviour $\overline{y}_{-ig}$
- Individual errors ε_i

Individual choice is then based on the outcome of the above utility framework such that one adopts a latrine (i.e. $y_{ig} = 1$) if:

$$u_{ig}(1) - u_{ig}(0) \geq 0$$

1.3.4 Equilibrium

Each household maximises their expected utility based on their own choice and the expected choice of other households within the GP. They opt for a decision if the utility from that decision is higher than the utility from making another decision:

$$y_i = \begin{Bmatrix} 1 \text{ if } u_{ig}(1) \geq u_{ig}(0) \\ 0 \text{ if } u_{ig}(0) \geq u_{ig}(1) \end{Bmatrix}$$

The decision making rule function of each household is:

$$\psi_i(k_g, \varepsilon_i; \theta) = arg \max_{y \epsilon Y} [\tilde{v}_d(y_i, k_g; \theta) + \varepsilon_i(y_i)] \tag{6}$$

Where $\psi_i(k_g, \varepsilon_i; \theta)$ refers to the determination of the parameter vector (θ) being dependent on common knowledge state variables (k_g), and private information specific to the household (ε_i).

Optimal utility is defined by:

$$\tilde{V}^*(y_i, k_g, \varepsilon_i(y_i); \theta) = \max_{d \epsilon 0,1)} [\tilde{V}_1(y_i = 1, k_g \varepsilon_i(1); \theta), \tilde{V}_0(y_i = 0, k_g, \varepsilon_i(0); \theta)] \tag{7}$$

The group probability of adopting a specific behaviour is the expected average level of adoption for the group:

$$\overline{p}_g = (I_g^{-1} - 1)\sum_{j\neq i} p_{i,j}(y_{jg} = 1|k_g;\theta) \tag{8}$$

which is based on the sum of every other households probability of adoption divided by the number of households in the GP (excluding HH_i). The assumed form of the error term (i.i.d and logistically distributed) means the probability of adoption is a logistic function based on the deterministic components of utility. This determines the conditional choice probability for HH_i:

$$p_i^*(y_i = 1|k_g,\theta) = \frac{exp(\beta_0 + \beta_1 x_i + \beta_2 z_g + \beta_3 \overline{p}_g^*)}{1 + exp(\beta_0 + \beta_1 x_i + \beta_2 z_g + \beta_3 \overline{p}_g^*)} \tag{9}$$

The above when solved provides the equilibrium probability of sanitation adoption. Equilibrium is possible because of the assumption that households have rational expectations. The symmetry of rational expectations m_{ig}^e across all households creates a self-consistent equilibrium outcome such that expected outcome of others is the same as their realised outcome (m_g).

If social effects are large enough, there is the risk of multiple equilibria such that a range of self-consistent average choice outcomes could exist. The existence of multiple equilibria outcomes is also influenced by the private deterministic aspect of utility ($u(\omega_i)$). This could capture individual characteristics that strongly favour one behaviour over the other. Characteristics such as gender or age of the household head, may have a threshold below which they lead to one preferred sanitation outcome and above which they lead to a different preferred outcome. On the other hand, households with relatively indifferent characteristics towards the two behaviours may end up on any equilibrium outcome. One example of an individual characteristic that could lead to any equilibrium is household size. Within certain boundaries household size may not strongly lean towards any particular behaviour preference and may not have a clear threshold that would consistently lead to one equilibrium outcome over another. As a result, on the basis of household size, one may end up on any equilibrium outcome.

1.4 Empirical Strategy

1.4.1 Estimation

The empirical model estimated is:

$$p(y_{ig} = 1) = F(\mathbf{X}_{ig}, \mathbf{Z}_g, \overline{y}_{-ig}; \boldsymbol{\beta}) \tag{10}$$

Using a logistic functional form is:

$$p(y_{ig} = 1) = \frac{1}{1 + exp(\beta_0 + \beta_1 \mathbf{X}_{ig} + \beta_2 \mathbf{Z}_g + \beta_3 \overline{y}_{-ig})} \tag{11}$$

where:

y_{ig} is a vector of household sanitation behaviour

$\mathbf{X}_{ig}$ is a matrix of household characteristics

$\mathbf{Z}_g$ is a matrix of group (GP) characteristics

$\overline{y}_{-ig}$ is a vector of group (GP) behaviour

β_1 are individual effects

β_2 are exogenous social effects (contextual effects and correlated effects)

β_3 are endogenous social effects

GP characteristics ($\mathbf{Z}_g$) include correlated characteristics used to capture correlated effects. It also includes contextual characteristics which are the group average of individual characteristics used to capture contextual effects.

The structure of the model creates the possibility of multiple equilibria due to the inclusion of group behaviour $\overline{y}_{-ig}$ as a regressor. Multiple equilibria can arise because different degrees of dependency (β_3) on others' choices, can lead to individual outcomes that are compatible with different aggregate outcomes. That is the strength of social interaction effects can determine whether as individuals make their sanitation decision, the group ends up on a low latrine usage rate equilibrium or a high latrine usage rate equilibria. In order to estimate the parameters of interest, the equilibrium equation should have a single solution for which a maximum likelihood estimator can solve. In the case of multiple equilibria, an equilibrium selection rule is required to determine the equilibrium level of group behaviour used to estimate the model.

It is unclear if ex ante the model in this chapter exhibits social effects sufficiently strong enough for multiple equilibria to exist. To test the hypothesis of multiple equilibria the model was estimated, and predicted probabilities of group behaviour were calculated. This was plotted in a graph against actual group behaviour. This was carried out for all sanitation behaviours, for each GP. The graph of equilibrium choice probabilities for a few GPs is shown in Figures 1-3 below. Where predicted group behaviour intersects with observed group behaviour, can be deemed the equilibrium. Of note is that predicted group behaviour line (blue) crosses the observed group behaviour line (green) multiple times. This supports the idea that for sanitation behaviours neighbourhood peer effects are strong enough to bring about multiple equilibria.

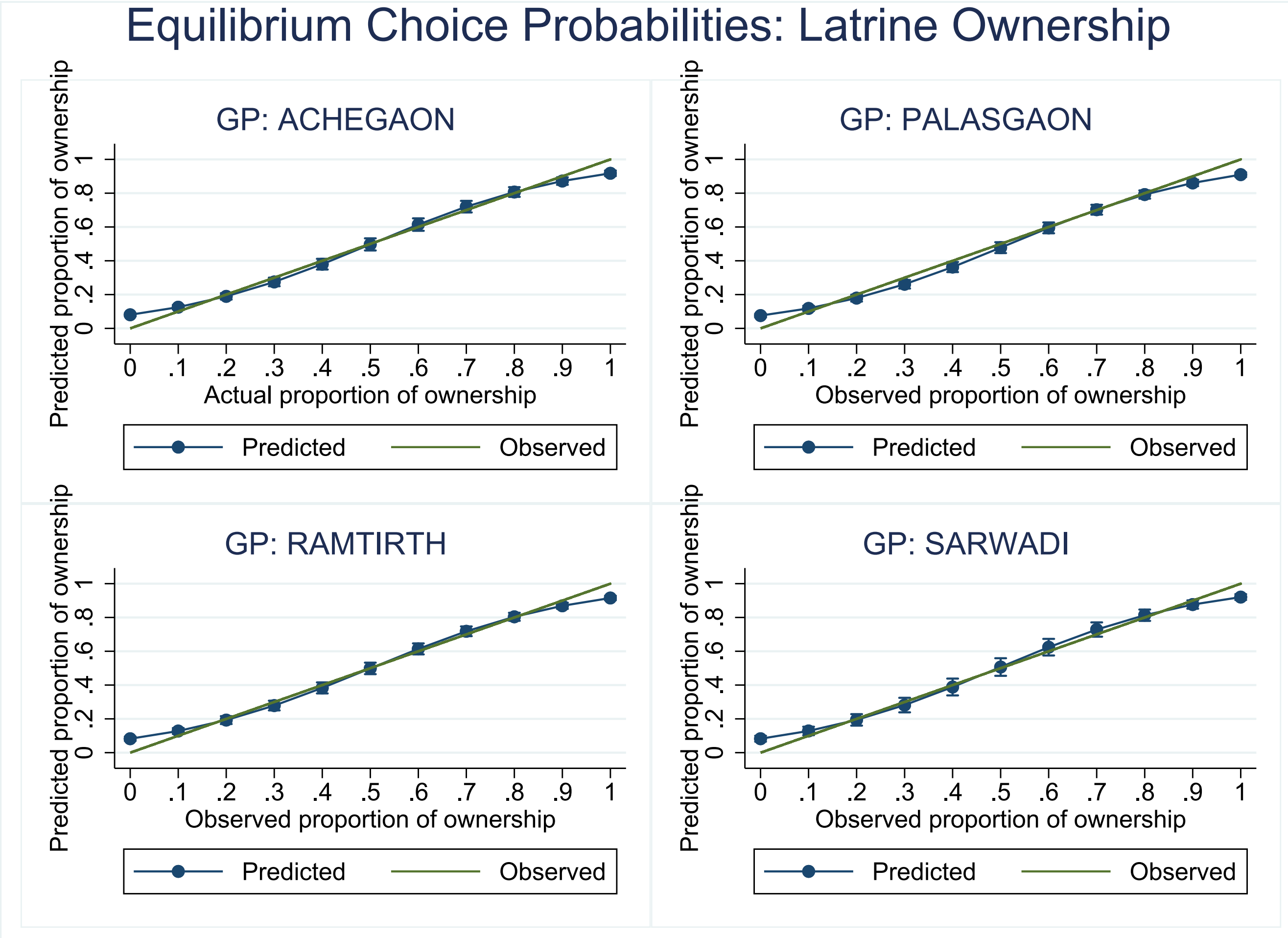

Figure 1: Latrine ownership choice probabilities

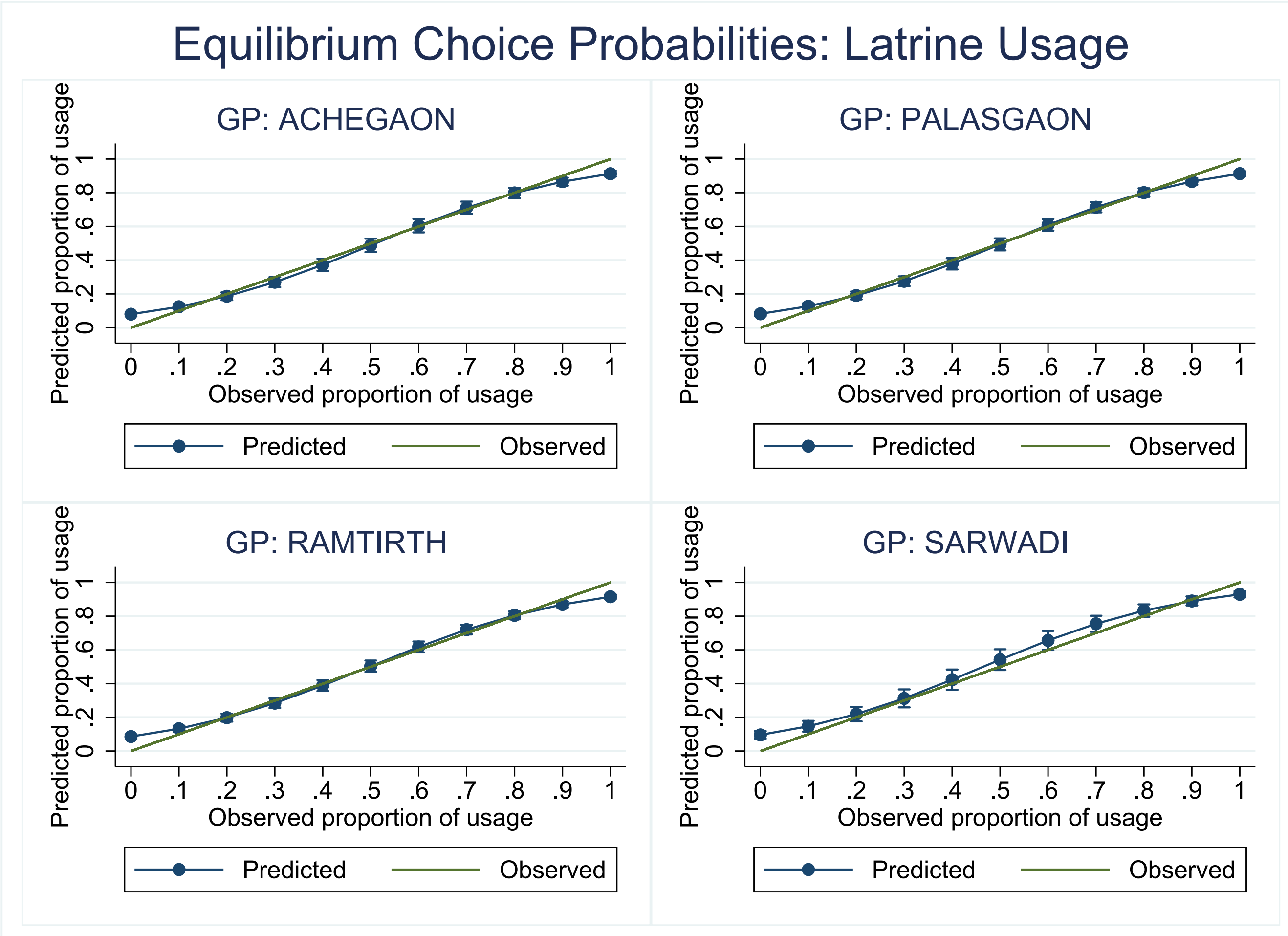

Figure 2: Latrine usage choice probabilities

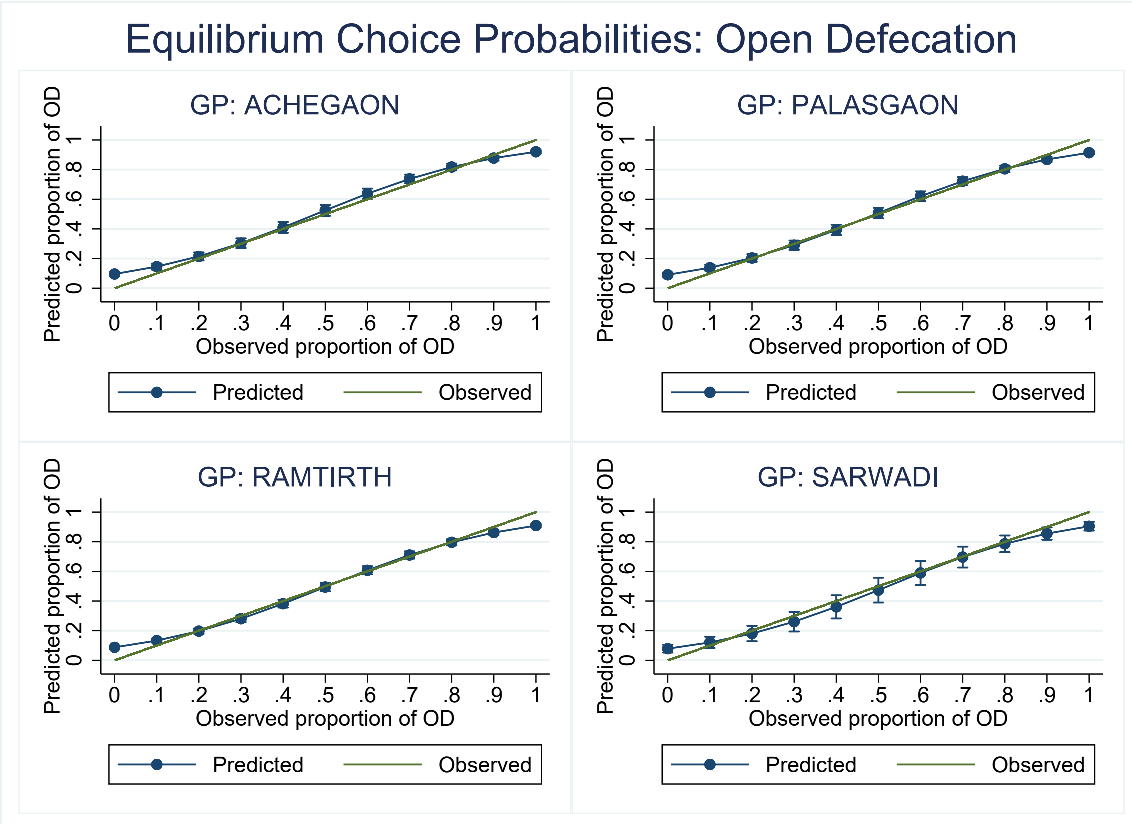

Figure 3: Open defecation choice probabilities

Different options have been proposed in the literature for dealing with estimation in the case of multiple equilibria. The baseline option entails repeated computation of solutions for the equilibrium equation for different sets of explanatory parameters, and then testing which parameter offers the highest likelihood given the data available. This can be computationally burdensome (Gautam, 2018a) and the assumptions made about the function for a chosen fixed-point theorem used would have to be justified. Justifying one equilibrium selection rule over another is difficult.

Alternative estimation methodologies for this kind of model involve a two-step estimation procedure (de Paula (2013); Bajari et al. (2010); Aguirregabiria and Mira (2007)) or the use of an Instrumental Variable (Shang and Lee (2011); Shang (2014)). The two step- estimation procedures involve the use of a non-parametric estimation method such as kernel or sieve estimators to first attain predicted probabilities of sanitation behaviour ($\hat{p}_i$ for p_i^*). The group level average of these predicted probabilities could be plugged in the second stage to maximise the log likelihood of adopting particular sanitation behaviour. The value of an initial non-parametric first stage is that it allows one to compute the choice probabilities without making any assumptions on the model's functional form. As such the estimates can be a useful predictor of choice probabilities in some equilibrium setting, or at least in the region of that setting.

An alternative option (Gautam 2018a) involves the use of the "equilibrium-in-data-rule". In this process a consistent estimate ($\hat{p}_i$) is computed from the actual choices made by households shown in the data. It is assumed this estimate coincides with the actual equilibrium choice probabilities of group decision ($\overline{p}_g$) in which

$$\overline{p_g} = \overline{y}_{-ig} = (I_g^{-1} - 1) \sum_{j \neq i} 1(y_j = 1) \tag{12}$$

For robustness of social effects estimation, two estimation procedures are followed. The first uses the equilibrium in data rule assumption. The second involves a two-step estimation in which in the first step predicted probabilities are estimated following a logit functional form. In the second stage these predicted probabilities are used to calculate group average outcomes to estimate endogenous social effects.

Equilibrium-in-data assumption

This method entails using group choice probabilities from the sample data. This is calculated in the first step as leave out one mean of group outcomes as shown in equation 12..[11]

In the second step this average level of adoption is plugged into the log-likelihood for choice probabilities and estimated using maximum likelihood:

$$L(\theta|\overline{y}_{-ig}, x_{ig}, z_g, \overline{p_g}) = \frac{1}{N}\sum_{i=1}^{N}[y_i log(\hat{p_i}(y_i = 1|x_{ig}, z_g, \overline{p_g}; \theta)]+(1-y_i)(log[1-\hat{p_i}(y_i = 1|x_{ig}, z_g, \overline{p_g}; \theta)])$$

Two-step predicted probabilities

This estimation method entails a two step process. The steps involved are:

Step 1: Estimate p_i^* as $\hat{p}_i$

Step 2: Attain group averages of predicted $\hat{p}_i$. Plug $\hat{p_g}$ into the log likelihood function and find the maximum likelihood.

$$L(\theta|\overline{y}_{-ig}, x_{ig}, z_g, y_i) = \frac{1}{N}\sum_{i=1}^{N}[y_i log(\hat{p_i}(y_i = 1|x_{ig}, z_g, \hat{p_g}, \theta)]+(1-y_i)(log[1-\hat{p_i}(y_i = 1|x_{ig}, z_g, \hat{p_g}; \theta)])$$

In this chapter, a logistic regression is modelled to estimate both stages.

[11] Leave-out-one mean is used such that the decision making household excludes themself from the group behaviour; controlling for increasing differences between individual and group behaviour in small groups.

1.4.2 Data

The data used in this study was collected between 2012-2014 from a sample of households in rural Maharashtra, India. The sample forms part of the baseline and listing data collection for a larger research project focused on sanitation adoption. The data used in this chapter comes from a census collected on Gram Panchayats (GPs), which are administrative units comprising households. GPs with 480 households or fewer were considered non-segmented GPs. In cases where a GP had more than 480 households and spanned multiple villages, one village was randomly chosen for data collection, while the others were excluded. GPs that consisted of a single village with more than 480 households were further divided into neighbourhoods, and a census was conducted in randomly selected neighbourhoods. These GPs were classified as segmented GPs. A majority of GPs in the dataset were not segmented. The collected sample included 120 GPs comprising over 48,000 households. The final dataset used for this chapter is a cross-section of 107 GPs with 38,000+ households.

Household survey data collected includes characteristics of the household head such as their age and gender. Other general household characteristics include whether there were kids under age two in the household, the caste group the household belongs to and the the sanitation behaviours of household members. The focus of this chapter is on three specific kinds of sanitation behaviour; latrine ownership, latrine use, and any open defecation. Ownership, Usage and OD are separately defined dummy variables, due to realities of the data, and context of sanitation. Latrine ownership accounts for whether a household owns a latrine in their abode or not. Latrine usage accounts for whether any members of the household use a toilet; be it a privately or publicly owned latrines. The inclusion of latrine use, as a separate dependent variable from latrine ownership is to capture the possibility that households may not own a toilet, but use one publicly available such as in school, work, or elsewhere within the community environment. Additionally, this accounts for the reality that not all households that own a latrine, use it. Ownership could be unbound to usage due to behavioural norms of OD (Routray et al. 2015), but also due to disrepair or ill-built latrines creating reversion to OD (Orgill-Meyer et al. 2019). Additionally, latrine usage is a subjective measure which could suffer from respondent desirability bias. The reason latrine usage and open defecation are separate variables is because OD captures if any household member engages in OD behaviour. Households can have some members engaging in OD whilst others use toilets, or both depending on the individual agent's circumstances. The aim is to ensure a range of sanitation behaviours are captured. Summary statistics of these household variables are shown in table 1.1.

Table 1.1: Household characteristics

Variable	Observations	Mean	Std. dev.	Min	Max
Age household head	38457	47.766	14.472	18	99
Household Size	38673	4.877	2.384	1	45
Kids under 2 years	38724	1.018	0.148	1	6
Female Head	38724	0.092	0.029	0.037	0.178
Other Backward Class (OBC)	38724	0.292	0.185	0	0.884
Scheduled Castes (SCs)	38724	0.254	0.106	0.022	0.597
Scheduled Tribes (STs)	38724	0.055	0.069	0	0.388
Denotified Tribes (DTs)	38724	0.030	0.063	0	0.306
Nomadic Tribes (NTs)	38724	0.118	0.114	0	0.730
Forward Caste (FC)	38724	0.252	0.156	0	0.694
Latrine Ownership	38724	0.278	0.183	0.071	0.868
Latrine Usage	37608	0.296	0.182	0.076	0.923
Open Defecation (OD)	37608	0.665	0.211	0.064	0.929

Note: Toilet usage and Any OD are reduced to 37,608 observations because there are households surveyed that do not own toilets, but when asked where they defecate answered as "Don't Know". Age of household head, and Household size reduce in sample size due to those who responded "Don't Know".

Households head are on average 47.76% years old, and contain about 5 members.

Only 9.2% of households are headed by women. The rates of latrine ownership, usage and OD rates are 27.8%, 29.6% and 66.5% respectively. Caste group variable show the distribution of each caste group across the dataset. Not all caste groups are present within all GPs, except for Scheduled Castes who have a presence in all GPs.

Village level data collected includes a list of factors pertaining to the presence of certain local institutions and sanitation complementing characteristics. These GP level characteristics used to control for potential correlated characteristics that could shape sanitation outcomes are noted in table 1.2.[12] Intuition suggests that some of these group level characteristics are more relevant for sanitation behaviour than others.

Caste relations are known to play a role in sanitation outcomes (Coffey and Spears 2017, Gupta et al. 2019) from choice of defecation behaviour, to choice in the types of latrines built. Additionally caste composition have been known to affect adoption of new behaviours (de Janvry et al. 2022). To account for these caste related factors that could confound sanitation outcomes, proportions of each caste group in the village are included as controls in the models estimated. This is included alongside group level averages of other household characteristics. Both contextual characteristics and correlated characteristics account for overall group characteristics.[13]

41.4% of GPs have had some sanitation activity carried out in the last 3 years. 86.8% of GPs have programmes that provide some sort of support for the construction of toilets. There are on average about 12 masons to hand in the villages. For 58.6% of villages, their current GP leader is female, whilst for 34.5% of villages the leader was female in the previous year. These variables are included as correlated characteristics because they could inform differences in sanitation outcomes between groups. Women are known to have a higher preference for latrines (Augsburg et al. 2022) but also objectively benefit more from access to latrines (Sahoo et al. 2015).

[12] Contextual characteristics are calculated as the average of individual characteristics excluding the household's characteristics. These are included as controls in estimation to capture exogenous social effects.

[13] Income is not included in the specification as income data was not collected for all households. The issue of controlling for income is addressed in chapter 2. There it is shown that even with income and wealth proxies, estimation results do not change significantly.

Therefore a female village leader could have political preferences towards increasing latrine building and usage in the village. The number of masons account for the market availability of components required to construct a latrine. More masons in a village could mean greater availability of not only the skill set required to construct latrines, but also the tools and materials required. There are about 394 households per village in the sample, ranging from 100+ to 400+ households per GP in the sample data.

Table 1.2: GP characteristics

Variable	Observations	Mean	Std. dev.	Min	Max
Segmented GP	38724	0.472	0.499	0	1
Households per GP	38724	669.816	402.906	107	2250
Sanitation Activity - Last 3 years	38724	0.414	0.493	0	1
Latrine Construction Support	38724	0.868	0.338	0	1
Female GP leader	38724	0.586	0.493	0	1
Previous year female GP leader	38724	0.345	0.475	0	1
Number of Masons	38724	12.181	22.224	0	100

Note: Households per GP accounts for entire GPs which includes villages where the GP was segmented.

1.5 Results

Marginal effects results using the equilibrium in data rule, and the two step predicted probability method are shown in table 1.3. Models (1), (3), and (5) are estimates using the equilibrium in data rule. Models (2), (4), and (6) are estimates using predicted probabilities. Standard errors are bootstrapped and clustered at GP level for all models.

1.5.1 Social Effects

Endogenous social effects are present and positive on all sanitation behaviours modelled. Marginal effects are higher but statistically insignificant for models (2), (4), and (6) which use a two step method with predicted probabilities. Following the equilibrium in data rule, the likelihood of household latrine ownership increases by 8.13% when there is a 10% increase in neighbourhood level ownership. This effect is 8.39% for latrine usage and 8.42% for OD.

As for exogenous social effects, contextual social effects are largely statistically insignificant. However, some exhibit intuitive results. Group average age of household head is positively related to latrine ownership and usage and negatively related to OD. Additionally, a greater presence of female headed households is positively correlated to latrine ownership and usage; 1.64% and 1.18% respective increase from a 10% increase in neighbourhood composition of female headed households. It creates a 0.997% decrease in OD. Correlated social effects are all statistically insignificant and provide mixed results on sanitation outcomes.

Endogenous social effects exhibit the highest magnitude of social effects impact on sanitation behaviour within the models estimated, and they are marginally stronger for OD than other behaviours. Though statistically insignificant for the most part, predicted probabilities based estimated provide the same direction of effects.

Table 1.3: Marginal effects on sanitation behaviours

	Own		Use		OD	
	(1)	(2)	(3)	(4)	(5)	(6)
Group Sanitation Behaviour	0.813***	1.426	0.839***	1.706	0.842***	2.007
	(0.0121)	(1.415)	(0.0124)	(1.626)	(0.0101)	(1.394)
Age household head	0.00801***	0.00965***	0.00619***	0.00777***	-0.00604***	-0.00748***
	(0.00126)	(0.00136)	(0.00120)	(0.00129)	(0.00126)	(0.00142)
Age household head squared	-0.0000544***	-0.0000703***	-0.0000383***	-0.0000535***	0.0000387***	0.0000528***
	(0.0000116)	(0.0000128)	(0.0000110)	(0.0000122)	(0.0000116)	(0.0000136)
Household Size	0.00892***	0.00861***	0.00860***	0.00826***	-0.00865***	-0.00866***
	(0.00118)	(0.00131)	(0.00117)	(0.00133)	(0.00128)	(0.00115)
Female Head	-0.0525***	-0.0528***	-0.0512***	-0.0515***	0.0478***	0.0480***
	(0.00730)	(0.00742)	(0.00637)	(0.00718)	(0.00592)	(0.00790)
Kids under 2 years	-0.00917	-0.00967	-0.00654	-0.00830	-0.000498	0.0106
	(0.0140)	(0.0176)	(0.0157)	(0.0180)	(0.0165)	(0.0247)
Scheduled Castes (SCs)	-0.120***	-0.117***	-0.126***	-0.128***	0.111***	0.116***
	(0.0160)	(0.0172)	(0.0152)	(0.0171)	(0.0198)	(0.0200)
Scheduled Tribes (STs)	-0.0846***	-0.0846***	-0.0843***	-0.0864***	0.0796***	0.0826***
	(0.0143)	(0.0175)	(0.0138)	(0.0190)	(0.0176)	(0.0210)
Denotified Tribes (DTs)	-0.176***	-0.157***	-0.181***	-0.166***	0.163***	0.156***
	(0.0181)	(0.0254)	(0.0232)	(0.0270)	(0.0294)	(0.0281)
Nomadic Tribes (NTs)	-0.102***	-0.0989***	-0.108***	-0.109***	0.0950***	0.106***
	(0.0125)	(0.0148)	(0.0145)	(0.0159)	(0.0168)	(0.0178)
Forward Caste (FC)	0.00657	0.00485	0.0106	0.00536	-0.00527	0.00767
	(0.00988)	(0.0148)	(0.0116)	(0.0132)	(0.0145)	(0.0164)
Avg. age household head	0.0301	-0.00835	0.0377	-0.0216	-0.0397	0.0730
	(0.0236)	(0.150)	(0.0218)	(0.154)	(0.0322)	(0.213)
Avg. age household head squared	-0.000334	0.0000624	-0.000422	0.000208	0.000448	-0.000754
	(0.000246)	(0.00159)	(0.000229)	(0.00164)	(0.000338)	(0.00227)
Avg. Household Size	-0.00258	-0.0445	-0.0101	-0.0392	0.00844	-0.0686
	(0.00561)	(0.113)	(0.00606)	(0.0871)	(0.00690)	(0.116)
Avg. no. Kids under 2 years	-0.228	-0.444	-0.0551	-0.726	-0.0651	4.247
	(0.244)	(2.298)	(0.249)	(2.573)	(0.255)	(5.846)
Proportion Female Head	0.164	0.124	0.118	0.128	-0.0997	0.409
	(0.0874)	(0.796)	(0.102)	(0.775)	(0.124)	(1.104)
Proportion OBC	-0.0541*	-0.219	-0.0468*	-0.318	0.0308	0.312
	(0.0228)	(0.425)	(0.0236)	(0.478)	(0.0349)	(0.332)
Proportion SCs	0.0692*	-0.0632	0.0813*	-0.0943	-0.0482	0.266
	(0.0285)	(0.348)	(0.0356)	(0.390)	(0.0382)	(0.482)
Proportion STs	-0.00753	0.168	-0.0119	0.316	0.0125	-0.975
	(0.0521)	(0.847)	(0.0604)	(0.896)	(0.0779)	(1.536)
Proportion DTs	0.0216	0.359	0.0378	0.515	-0.0253	-0.675
	(0.0900)	(0.867)	(0.0912)	(0.946)	(0.0986)	(0.972)
Proportion NTs	0.0423	0.0729	0.0516	0.0931	-0.0455	-0.403
	(0.0346)	(0.229)	(0.0349)	(0.256)	(0.0385)	(0.553)
Proportion FC	-0.0632*	-0.133	-0.0614*	-0.188	0.0357	-0.00250
	(0.0255)	(0.296)	(0.0253)	(0.286)	(0.0297)	(0.226)
Segmented GP	0.0117	-0.0447	0.0149	-0.0724	-0.00848	0.0832
	(0.00674)	(0.116)	(0.00777)	(0.149)	(0.00811)	(0.125)
Households per GP	-0.0000124	0.0000131	-0.0000107	0.0000335	0.00000257	-0.000105
	(0.00000906)	(0.000103)	(0.0000126)	(0.000123)	(0.0000125)	(0.000185)
Latrine Construction Support	-0.00633	-0.0470	-0.00482	-0.0806	0.000209	0.0787
	(0.00650)	(0.172)	(0.00811)	(0.188)	(0.00860)	(0.128)
Sanitation Activity - Last 3 years	0.000550	0.00916	0.00159	0.0158	0.000694	0.0362
	(0.00509)	(0.0520)	(0.00682)	(0.0525)	(0.00716)	(0.0613)
Number of Masons	-0.000148	0.000233	-0.000143	0.000375	0.000179	-0.000606
	(0.000115)	(0.00119)	(0.000137)	(0.00142)	(0.000143)	(0.00136)
Female GP leader	0.00356	0.00737	-0.000863	0.00903	0.00289	-0.0315
	(0.00493)	(0.0463)	(0.00469)	(0.0462)	(0.00582)	(0.0533)
Previous year female GP leader	-0.00577	0.0312	-0.00652	0.0472	0.00542	-0.0878
	(0.00567)	(0.0895)	(0.00621)	(0.0994)	(0.00661)	(0.126)
N	38406	38406	37295	37295	37295	37295

Notes: Models (1), (3), and (5) are estimates using equilibrium in data assumption. Models (2), (4), and (6) are estimates using predicted probabilities. Standard errors in parentheses. * $p<0.05$, ** $p<0.01$, *** $p<0.001$. Standard errors are bootstrapped and clustered at GP level.

1.5.2 Individual Effects

Individual level effects appear to have intuitively understandable directions regarding the likelihood of adoption, except for the effect of a female household head which appears to be negative and statistically significant. Being a household headed by a female means a 0.53% reduction in latrine ownership, 0.51% decrease in latrine usage, and 0.48% increase in OD. Given the gender differentials in income and wealth that are known to exist in rural developing country environments, these results may be capturing the fact that female household heads may have lower incomes. As such the negative effects may be incorporating the impact of lower incomes affecting sanitation adoption, especially as income is not included as a control variable.

Age of household head has a small positive effect on latrine ownership (0.08%) and usage (0.06%), and negative effect on OD (0.06%). Age of household head squared exhibits the opposite effect. The inclusion of squared age is to account for non-linearities on how age affects the likelihood of adoption. Increases in age may be tied to increased likelihood in sanitation adoption, but this is likely non-monotonic, hence the inclusion of squared age of household head.

1.5.3 Links between sanitation behaviours

As mentioned previously, households exhibit heterogeneity in their sanitation behaviours. This includes owning but not using a latrine, using a latrine they do not own, and different members within the same household exhibiting different sanitation behaviours. To further explore the interconnectedness of these sanitation behaviours, the estimates from table 1.3 are replicated by focusing on latrine usage and OD behaviours conditional on latrine ownership. In other words, households latrine usage and OD behaviours dependent on whether they own a latrine or not are examined. The results, presented in Appendix Table A1, indicate that owning a latrine has a significant positive effect on the likelihood of a household using it, and it significantly reduces the probability of OD. Additionally, households that own latrines are still influenced by the behaviours of their neighbours, although to a lesser extent when compared to the findings in table 1.3.

1.5.4 Group Unobservables

The results in table 1.3 are based on the identification of a model where no group level unobservables that affect sanitation outcomes are present. The inclusion of group level characteristics accounts for some group level observables. Once the assumption of no group level unobservables is relaxed, identification fails. B&D offer solutions to the identification issue created when group level unobservables are present. However, their solutions only lead to partial identification of social effects.

To address this issue of group level unobservables, the work of Shang and Lee (2011) is followed. Their paper involves the use of instrumental variables (IV) for group behaviour. This aims to estimate endogenous social effects whilst creating exogeneity from unobserved group level characteristics. The authors propose a two step estimation procedure[14], in which all group level effects are put together into a group fixed effect (GFE) variable. That is shared group characteristics, group behaviour, and group environmental factors. The GFE variable is then estimated with the use of IVs to replace the endogenous group behaviour variable, whilst reducing the interference group level unobservables would have on endogenous social effects. This estimation method is compatible with single or multiple equilibria scenarios.

Prerequisite assumptions for this method are that group sizes and number of groups are sufficiently large, and that error terms follows a normal distribution. The error term assumption can be generalised to other parametric or semi-parametric social binary choice social interaction models. So the logistic distribution of error terms in this chapter's case remains valid.

The theoretical framework behind the application of this method can be set as a continuation of the model framework used thus far. The previously estimated empirical model equation (in linear form):

$$y_{ig} = \beta_0 + \beta_1 \mathbf{X}_{ig} + \beta_2 \mathbf{Z}_g + \beta_3 \overline{y}_{ig} + \varepsilon_i \tag{13}$$

[14]Though they follow a probit model framework, their processes is noted to be applicative to a logit model without loss of applicability.

can be re-written as:

$$y_{ig} = x_{ig}\delta \ + \alpha_g \tag{14}$$

where x_{ig} represents individual characteristics, and α_g represents group fixed effects encompassing all group variables: group contextual characteristics, correlated characteristics, group behaviour and group unobservables:

$$\alpha_g \ = \gamma_1 \overline{x}_g + \gamma_2 s_g + \gamma_3 \overline{y}_{ig} + u_g \tag{15}$$

Where:
$\overline{x}_g$ are exogenous contextual group characteristics
s_g are exogenous correlated group characteristics[15]
$\overline{y}_{-ig}$ is group level behaviour
u_g are the group level unobservables

The first step is to estimate equation $y_{ig} = x_{ig}\delta \ + \alpha_g$ via a maximum likelihood estimator such as logit. The first step likelihood estimation provides a consistent estimate for each group under the assumption that group sizes and number of groups tends to infinity, but group size going to infinity at a rate either faster or similar to that of the number of groups.

The second step estimates the equation:

$$\hat{\alpha_g} = \gamma_1 \hat{\overline{x}}_g + \gamma_2 s_g + \gamma_3 \hat{\overline{y}}_{ig} + u_g + v_g \tag{16}$$

where the GFE coefficients from the first step are used as predicted values to estimate group variables separately. v_g is the group level error term, correlated group characteristics (s_g) are group level environmental factors, and contextual group characteristics ($\overline{x}_g$) are the GP average of individual characteristics[16]:

[15] What was previously represented by Z_g as group characteristics, is now split between contextual group characteristics ($\overline{x}_g$) which are group average of individual characteristics, and correlated group characteristics (s_g) which are group level environmental factors

[16] Note here that in this version of the social effects estimation equation, group averages of individual charac-

$$\hat{\bar{x}}_g = \frac{1}{N}\sum_{i=1}^{N} x_{ig}$$

Group behaviour is either the group average of individual behaviour as available in the data:

$$\hat{\bar{y}}_g = \frac{1}{N}\sum_{i=1}^{N} y_{ig}$$

or the average of the predicted individual behaviour from estimating the first step equation:

$$\hat{\bar{y}}_g = \frac{1}{N}\sum_{i=1}^{N} \phi(x_{ig}\delta + \hat{\alpha})$$

Improved parameter estimates are shown in table 1.3 from using averages based on the sample data, instead of predicted probabilities for estimation. Therefore in the first step, actual group behaviours are used instead of group behaviour based on predictions.

As noted from the functional form of the second step's equation, the existence of group level unobservables u_g, would influence the formation of the group behaviour(y_g), making the estimation of endogenous effects bias if these unobservables are not accounted for. Shang and Lee (2011) suggests the creation and use of instrumental variables to replace group behaviour in this second step equation that would asymptotically eliminate group level unobservables.

They offer two potential ways to create these IVs in this setting. The first comes

teristics (contextual effects) and group behaviour are averages derived from the data, and not a leave-one-out mean averages as calculated in the previous equilibrium in data estimation procedure. So $\overline{y}_{-ig}$ is now y_g

from using the sample equivalence of group level behaviour. Given that:

$$y_{ig} = x_{ig}\delta \ + \alpha_g \tag{17}$$

$$y_g = \int \phi(x_{ig}\delta + \alpha) dF_{x_{ig}|g} \tag{18}$$

This allows for the relevance necessity of the IV to be met, because it is based on creating the IV from factors that shape group behaviour. Group fixed effects (α_g) are then averaged across groups, becoming a constant that is invariant across groups such that the first variation of the instrumental variable (IV1) becomes:

IV1

$$\frac{1}{N}\sum_{i=1}^{N} \phi(x_{ig}\delta + \hat{\overline{\alpha}})$$

By averaging the GFE across all groups, making it group invariant, this arguably enables the IV to be asymptotically uncorrelated with group level unobservables.

The second type of IV offered is derived from the first, with the inclusion of the difference in average individual characteristics (contextual characteristics) and their averages across GPs. This is based on the idea that the first IV whilst eliminating group level unobservable effects, may lose some important information around group behaviour. As such, the difference between exogenous group contextual characteristics and their averages across all GPs, could capture some of the lost information, whilst retaining asymptotic independence from the influence of group level unobservables.

IV2

$$(\overline{x}_{l,ig} - \overline{\overline{x}}_l)\frac{1}{N}\sum_{i=1}^{N} \phi(x_{ig}\delta + \hat{\overline{\alpha}})$$

where

$$\overline{x}_l = \frac{1}{G}\sum_{g=1}^{G}\overline{x}_{l,ig}$$

The implementation of this IV strategy when applied to the data leads to the creation of five instrumental variables. The first is as IV1, the second to fifth are variations of IV2 using contextual characteristics: Average age of household head; Average household size; Proportion of female household heads; Average number of kids under two years of age. The use of these IVs in isolation and combined allows for 31 different variations of estimation for endogenous group effects. Tables 1.A2-1.A4 in the appendix reports results from using IVs that provided statistically significant outcomes for group level behaviour.[17] Of note, is that IV1 in isolation created statistically insignificant results of unusually high magnitude of endogenous social effects for all three sanitation behaviours.

The IVs for the most part provide estimates of endogenous social effects that are similar to those previously estimated. However the similarity in magnitude and direction of social effects to table 1.3 adds robustness more than it offers new insight into social effect estimation using this method that explicitly accounts for group level unobservables. This means either the previously estimated model sufficiently captures group level unobservables, or even if it does not, does not lead to unreliable estimates of endogenous social effects as hypothesised.

The non-linearity of the logit models lends validity to the to construction of the IVs (Shang and Lee 2011). However, the overall validity of the IVs remains subject to uncertainty due to the assumption that averaging across distinct groups sufficiently eliminates the impact of other specific factors associated with these groups (whether observable or unobservable) from the estimated outcomes, while still preserving relation with group behaviour. This uncertainty could potentially explain the emergence

[17] Variations of the IV which were also excluded from reported results include those that use the variable no. of Kids under 2. This is because this variable had little variation in the dataset, making it unlikely that its inclusion in the construction of the IVs adds sufficient strength or power to explain variations in group behaviour.

of misleading outcomes when solely relying on IV1 in isolation. Alternatively, the inclusion of IV2 (comprising IV1 along with the difference between the average group characteristics and the overall average of those characteristics across all groups) could enhance the validity of the IV. This expanded instrument captures additional household characteristics that may play a role in determining group behaviour, thereby contributing to a more valid IV estimation.

1.6 Conclusion

This chapter estimates the impact of neighbourhood based peer effects on sanitation behaviour: latrine ownership, latrine usage, and open defecation (OD). It follows the framework of Brock and Durlauf's (2001, 2007) non-linear discrete choice models. Multiple equilibria exist in the model. To address the risk of this affecting estimation, two separate estimation procedures were followed: equilibrium in data rule, and predicted probabilities.

Following the equilibrium in data rule, results show that social multipliers in sanitation behaviours exist at the neighbourhood level, with 10% increase in group level behaviour leading to an approximate marginal effect magnitude of 8% in likelihood of households carrying out the same behaviour. These effects are minutely larger for OD, and consistently larger than exogenous social effects. With the predicted probabilities method, results are of the same direction as the equilibrium in data method. However, the results were also statistically insignificant, and over-inflated when compared to estimation using the equilibrium in data rule.

Identification of the model was based on the assumption of no-group level unobservable factors that could influence sanitation outcomes. Given identification failure and risk of estimate bias when this assumption is relaxed, an IV estimation procedure used by Shang and Lee (2011) is followed to estimate social effects in the presence of group level unobservables. The IVs were created to asymptotically remove the influence of these group level unobservables. Results using variations of the IV show similar magnitude and direction of social effects on sanitation behaviour.

Marginal effects result suggest that neighbourhood level social effects on individual household sanitation behaviour plays a key role in persisting certain sanitation outcomes such as high open defecation rates and low latrine usage. This supports the use of neighbourhood focused policies to improve sanitation outcomes as they can benefit from the social multiplier in behaviours that come from social effects.

The strength of endogenous effects in this chapter may be exaggerated or underplayed due to the model specification that allows every household to be equally and symmetrically affected by every other household in the GP. An improved specification, open to further research would be to create social groups in a manner that allows heterogeneity in social ties.[18] For example neighbourhood effects based on spatial distance between households within the neighbourhood, or peer effects through well known social structures in the Indian context, such as caste.

[18]This would also improve identification of the different social effects.

1.7 References

Aguirregabiria, V. and Mira, P. (2007). Sequential Estimation of Dynamic Discrete Games. Econometrica [Online] 75:1–53.

Augsburg, B., Baquero, J. Gautam, S., RodriguezLesmes, P. (2021) : Sanitation and marriage markets in India: Evidence from the total sanitation campaign, IFS Working Paper, No. W21/44, Institute for Fiscal Studies (IFS), London.

Augsburg, B., Malde, B., Olorenshaw, H., Wahhaj, Z. (2022). To invest or not to invest in sanitation: the role of intra-household gender differences in perceptions and bargaining power. 23/04. London: Institute for Fiscal Studies.

Bajari, P., Hong, H., Krainer, J., and Nekipelov, D. (2010). Estimating Static Models of Strategic Interactions. Journal of Business and Economic Statistics, 28(4), 469–482.

Beaman, L., BenYishay, A., Magruder, J., and Mobarak, A.M. 2021. "Can Network Theory-Based Targeting Increase Technology Adoption?" American Economic Review, 111 (6): 1918-43.

BenYishay, A. and Mobarak, A.M. (2014). Social Learning and Communication. National Bureau of Economic Research Working Paper Series [Online] No. 20139.

BenYishay, A.,Fraker, A.,Guiteras, R., Palloni, G., Shah, N.,Shirrell, S., Wang, P.(2017). Microcredit and willingness to pay for environmental quality: Evidence from a randomized-controlled trial of finance for sanitation in rural Cambodia, Journal of Environmental Economics and Management, Volume 86, Pages 121-140, ISSN 0095-0696

Brock, W. and Durlauf, S.N. (2000). Interactions-Based Models. National Bureau of Economic Research Technical Working Paper Series [Online] No. 258.

Brock, W.A. and Durlauf, S.N. (2001). Discrete Choice with Social Interactions. The Review of Economic Studies [Online] 68:235–260. .

Brock, W. and Durlauf, S. (2007). Identification of binary choice models with social interactions. Journal of Econometrics [Online] 140:52–75.

Cameron, L., Gertler, P., Shah, M., Alzua, M.L., Martinez, S., Patil, S. (2022) The dirty business of eliminating open defecation: The effect of village sanitation on child height from field experiments in four countries, Journal of Development Economics, Volume 159, 102990, ISSN 0304-3878

Caruso, B., Sclar G.,, Routray P., Nagel C., Majorin F., Sola S., Koehne W., Clasen T., Effect of a low-cost, behaviour-change intervention on latrine use and safe disposal of child faeces in rural Odisha, India: a cluster-randomised controlled trial, The Lancet Planetary Health, Volume 6, Issue 2, 2022, Pages e110-e121, ISSN 2542-5196.

Chuang, Y. and Schechter, L.(2014) Social Networks in Developing Countries. (2014). Social Networks in Developing Countries. Annual Review of Resource Economics.

Coffey, D. and Spears, D. (2017) Where India Goes: Abandoned Toilets, Stunted Development and the Costs of Caste. New Delhi: HarperCollins Publishers India.

Coleman, J. et al. (1966). Equality of educational opportunity. Washington D. C.: U. S. Government Printing Office.

Conley, T.G. and Udry, C.R. (2010). Learning about a New Technology: Pineapple in Ghana. The American Economic Review [Online] 100:35–69.

de Janvry, A., Rao, M., and Sadoulet, E. “Seeding the Seeds: Role of Social Structure in Agricultural Technology Diffusion.” Working Paper, August 2022.

de Paula, Á. (2013). Econometric Analysis of Games with Multiple Equilibria. Annual Review of Economics [Online] 5:107–131.

Deutschmann, J.W., Gars, J., Houde, J., Lipscomb, M., and Schechter, L., Privatization of public goods: Evidence from the sanitation sector in Senegal (2023), Journal of Development Economics, Volume 160, 102971, ISSN 0304-3878

Dickinson, K., and Pattanayak, S. (2009). Open sky latrines: Social reinforcing in the case of a (very) impure public good.

Freeman MC, Garn JV, Sclar GD, Boisson S, Medlicott K, Alexander KT, Penakalapati G, Anderson D, Mahtani AG, Grimes JET, Rehfuess EA, Clasen TF. (2017) The impact of sanitation on infectious disease and nutritional status: A systematic review and meta-analysis. Int J Hyg Environ Health. Aug;220(6):928-949.

Gautam, S. (2018). Household (Under) Adoption of Sanitation: Externalities and Borrowing Constraints.

Gautam, S. (2018). Quantifying Welfare Effects in the presence of Externalities: An Ex-ante Evaluation of a Sanitation Intervention.

Garn JV, Sclar GD, Freeman MC, Penakalapati G, Alexander KT, Brooks P, Rehfuess EA, Boisson S, Medlicott KO, Clasen TF. The impact of sanitation interventions on latrine coverage and latrine use: A systematic review and meta-analysis. Int J Hyg Environ Health. 2017 Apr;220(2 Pt B):329-340.

Geruso, Michael, and Dean Spears. 2018. ”Neighborhood Sanitation and Infant

Mortality.” American Economic Journal: Applied Economics, 10 (2): 125-62.

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

Gupta A, Khalid N, Desphande D, Hathi P, Kapur A, Srivastav N, Vyas S, Spears D, Coffey D (2019), Changes in Open Defecation in Rural North India: 2014-2018 ,IZA Discussion Papers No. 12065, Institute of Labor Economics (IZA), Bonn.

Jackson, M. and Zenou, Y., 2015. ”Games on Networks,” Handbook of Game Theory with Economic Applications,, Elsevier.

Kennedy-Walker, R., Amezaga, J.M., Paterson, C.A. (2015) The impact of community social dynamics on achieving improved sanitation access for the urban poor: The case of Lusaka, Zambia. Habitat International, Volume 50, Pages 326-334, ISSN 0197-3975.

Kresch, E., Lipscomb, M., and Schechter, L., Externalities and Spillovers from Sanitation and Waste Management in Urban and Rural Neighborhoods (July 26, 2019).

Manski, C.F. (1993). Identification of Endogenous Social Effects: The Reflection Problem. The Review of Economic Studies [Online] 60:531–542.

Manski, C.F. (2000). Economic Analysis of Social Interactions. Journal of Economic Perspectives [Online] 14:115–136.

Moffitt, R. (2000). Policy Interventions, Low-Level Equilibria And Social Interactions. Social Dynamics:45–82.

Orgill-Meyer J, Pattanayak SK, Chindarkar N, Dickinson KL, Panda U, Rai S, Sahoo B, Singha A, Jeuland M. Long-term impact of a community-led sanitation campaign in India, 2005-2016. Bull World Health Organ. 2019 Aug 1;97(8):523-533A.

Pakhtigian, E., Dickinson, K., Orgill-Meyer, J., Pattanayak, S., 2022. "Sustaining latrine use: Peers, policies, and sanitation behaviors," Journal of Economic Behavior and Organization, Elsevier, vol. 200(C), pages 223-242

Routray P, Schmidt WP, Boisson S, Clasen T, Jenkins MW. Socio-cultural and behavioural factors constraining latrine adoption in rural coastal Odisha: an exploratory qualitative study. BMC Public Health. 2015 Sep 10;15:880.

Sahoo KC, Hulland KR, Caruso BA, Swain R, Freeman MC, Panigrahi P, Dreibelbis R.(2015) Sanitation-related psychosocial stress: A grounded theory study of women across the life-course in Odisha, India. Soc Sci Med. 2015 Aug;139:80-9.

Shang, Q. (2014). Endogenous neighborhood effects on welfare participation. Empirical Economics [Online] 47:639–667.

Shang, Q. and Lee, L. (2011). Two-Step Estimation of Endogenous and Exogenous Group Effects. Econometric Reviews [Online] 30:173–207. .

Venkataramanan, V., Crocker, J., Karon, A., and Bartram, J. (2018) Community-Led Total Sanitation: A Mixed-Methods Systematic Review of Evidence and Its Quality, Environmental Health Perspectives 126:2 CID: 026001

WHO/UNICEF JMP (2021), Progress on household drinking water, sanitation and hygiene 2000-2020 Five years into the SDGs.

1.8 Appendix A: Results

Table A1: Marginal effects on sanitation behaviours conditional on latrine ownership

	Use	OD
	(1)	(2)
Group Sanitation Behaviour	0.0342***	0.257***
	(0.00996)	(0.0546)
Household owns a latrine	0.153***	-0.315***
	(0.00914)	(0.0256)
Age household head	-0.00122***	0.000663
	(0.000320)	(0.000613)
Age household head squared	0.0000113***	-0.00000520
	(0.00000293)	(0.00000592)
Household Size	-0.000219	0.000634
	(0.000290)	(0.000468)
Female Head	-0.0000914	0.00289
	(0.00272)	(0.00284)
Kids under 2 years	0.000705	-0.0106
	(0.00368)	(0.00981)
Scheduled Castes (SCs)	-0.0132***	0.0125
	(0.00294)	(0.0114)
Scheduled Tribes (STs)	-0.00369	0.00633
	(0.00401)	(0.0127)
Denotified Tribes (DTs)	-0.0165**	0.0151
	(0.00571)	(0.0175)
Nomadic Tribes (NTs)	-0.00606	0.00318
	(0.00334)	(0.0116)
Forward Caste (FC)	-0.00143	0.0155
	(0.00305)	(0.0131)
Avg. age household head	0.00731	-0.0547
	(0.0134)	(0.0805)
Avg. age household head squared	-0.0000895	0.000589
	(0.000144)	(0.000864)
Avg. Household Size	-0.00412	0.0357**
	(0.00298)	(0.0137)
Avg. no. Kids under 2 years	-0.108	-0.795
	(0.134)	(0.736)
Proportion Female Head	-0.0628	-0.236
	(0.0475)	(0.271)
Proportion OBC	0.00500	0.0244
	(0.0135)	(0.0691)
Proportion SCs	-0.0175	-0.00965
	(0.0160)	(0.103)
Proportion STs	0.0196	0.0849
	(0.0334)	(0.116)
Proportion DTs	-0.0158	0.00528
	(0.0286)	(0.0974)
Proportion NTs	-0.0235	0.187*
	(0.0156)	(0.0874)
Proportion FC	0.000508	0.0675
	(0.0150)	(0.0663)
Segmented GP	0.00819	-0.000996
	(0.00439)	(0.0276)
Households per GP	-0.0000114*	0.0000150
	(0.00000582)	(0.0000289)
Latrine Construction Support	0.000415	0.0261
	(0.00383)	(0.0272)
Sanitation Activity - Last 3 years	0.00460	-0.0293
	(0.00310)	(0.0180)
Number of Masons	0.0000355	-0.0000282
	(0.0000726)	(0.000280)
Female GP leader	0.00189	-0.000205
	(0.00296)	(0.0162)
Previous year female GP leader	-0.000119	0.0154
	(0.00404)	(0.0147)
N	37295	37295

Notes: Models (1) and (2) are estimates conditional on whether household own a latrine or not. Standard errors in parentheses. *p<0.05 , ** p<0.01 , *** p<0.001. Standard errors are bootstrapped and clustered at GP level.

Table 1.A2: Marginal effects parameter estimates using Instrumental Variables (IV) - Latrine Ownership

	IV2: Gender	IV1 and IV2: Gender	IV2: Age, Gender	IV1 and IV2: Age, Gender	IV2: Household size	IV1 and IV2: Household size	IV2: Age, Household size	IV2: Gender, Household size	IV1 and IV2: Age, Household size	IV1 and IV2: Age, Gender, Household size	IV2: Age, Gender, Household size	IV1 and IV2: Age, Gender, Household size
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Constant	-0.0268	-0.0103	0.0427	0.0834	-0.131	-0.143	-0.0817	-0.104	-0.0802	-0.107	-0.0619	-0.0508
	(0.337)	(0.344)	(0.358)	(0.384)	(0.306)	(0.306)	(0.311)	(0.304)	(0.311)	(0.304)	(0.312)	(0.315)
Group Ownership behaviour	0.895***	0.906***	0.939***	0.965***	0.830***	0.823***	0.861***	0.847***	0.862***	0.845***	0.873***	0.880***
	(0.0896)	(0.0916)	(0.0944)	(0.109)	(0.0604)	(0.0615)	(0.0520)	(0.0545)	(0.0520)	(0.0548)	(0.0494)	(0.0496)
Avg. age household head	-0.0000304	-0.000231	-0.000876	-0.00137	0.00123	0.00138	0.000637	0.000909	0.000619	0.000945	0.000396	0.000261
	(0.00450)	(0.00445)	(0.00437)	(0.00442)	(0.00451)	(0.00458)	(0.00428)	(0.00442)	(0.00427)	(0.00444)	(0.00424)	(0.00420)
Avg. age household head squared	-0.0000108	-0.00000914	-0.00000375	0.000000375	-0.0000214	-0.0000226	-0.0000164	-0.0000187	-0.0000162	-0.0000190	-0.0000144	-0.0000133
	(0.0000417)	(0.0000411)	(0.0000390)	(0.0000385)	(0.0000436)	(0.0000442)	(0.0000410)	(0.0000425)	(0.0000409)	(0.0000426)	(0.0000403)	(0.0000398)
Avg. Household Size	-0.00484	-0.00548	-0.00753	-0.00911	-0.000811	-0.000348	-0.00271	-0.00184	-0.00277	-0.00173	-0.00348	-0.00391
	(0.00943)	(0.00976)	(0.0104)	(0.0119)	(0.00738)	(0.00744)	(0.00699)	(0.00729)	(0.00700)	(0.00729)	(0.00720)	(0.00733)
Avg. no. Kids under 2 years	-0.554	-0.563	-0.591	-0.613	-0.498	-0.492	-0.524	-0.512	-0.525	-0.511	-0.535	-0.541
	(0.299)	(0.308)	(0.334)	(0.364)	(0.276)	(0.276)	(0.283)	(0.275)	(0.283)	(0.275)	(0.286)	(0.290)
Proportion Female Head	0.129	0.131	0.137	0.141	0.118	0.116	0.123	0.121	0.123	0.120	0.125	0.126
	(0.0858)	(0.0869)	(0.0933)	(0.101)	(0.0875)	(0.0892)	(0.0839)	(0.0853)	(0.0838)	(0.0855)	(0.0838)	(0.0838)
Proportion OBC	-0.0772*	-0.0802*	-0.0896*	-0.0968*	-0.0588*	-0.0567*	-0.0675**	-0.0635**	-0.0678**	-0.0630**	-0.0710**	-0.0730**
	(0.0318)	(0.0325)	(0.0349)	(0.0389)	(0.0246)	(0.0247)	(0.0230)	(0.0236)	(0.0231)	(0.0236)	(0.0230)	(0.0233)
Proportion SC	0.0640	0.0609	0.0508	0.0431	0.0837**	0.0860**	0.0744*	0.0787*	0.0741*	0.0792*	0.0707*	0.0686*
	(0.0438)	(0.0447)	(0.0476)	(0.0525)	(0.0324)	(0.0328)	(0.0317)	(0.0327)	(0.0317)	(0.0327)	(0.0328)	(0.0331)
Proportion ST	0.0211	0.0266	0.0441	0.0576	-0.0133	-0.0172	0.00293	-0.00449	0.00343	-0.00545	0.00950	0.0132
	(0.0617)	(0.0620)	(0.0615)	(0.0673)	(0.0672)	(0.0684)	(0.0619)	(0.0614)	(0.0618)	(0.0617)	(0.0577)	(0.0571)
Proportion DT	0.0274	0.0329	0.0507	0.0643	-0.00743	-0.0114	0.00895	0.00144	0.00946	0.000469	0.0156	0.0193
	(0.0743)	(0.0750)	(0.0770)	(0.0833)	(0.0578)	(0.0588)	(0.0545)	(0.0577)	(0.0545)	(0.0579)	(0.0559)	(0.0558)
Proportion NT	0.0472	0.0478	0.0496	0.0509	0.0437	0.0433	0.0454	0.0446	0.0454	0.0445	0.0460	0.0464
	(0.0281)	(0.0287)	(0.0315)	(0.0341)	(0.0266)	(0.0268)	(0.0267)	(0.0264)	(0.0268)	(0.0264)	(0.0270)	(0.0273)
Proportion FC	-0.0822***	-0.0833***	-0.0869**	-0.0896**	-0.0752**	-0.0744**	-0.0785***	-0.0770***	-0.0786***	-0.0768***	-0.0798***	-0.0806***
	(0.0235)	(0.0240)	(0.0281)	(0.0311)	(0.0235)	(0.0237)	(0.0231)	(0.0227)	(0.0232)	(0.0227)	(0.0229)	(0.0232)
Segmented GP	0.0105	0.00954	0.00659	0.00432	0.0162	0.0169	0.0135	0.0148	0.0134	0.0149	0.0124	0.0118
	(0.0120)	(0.0121)	(0.0123)	(0.0132)	(0.00889)	(0.00900)	(0.00853)	(0.00896)	(0.00853)	(0.00898)	(0.00879)	(0.00881)
Households per GP	-0.0000128	-0.0000123	-0.0000105	-0.00000921	-0.0000162	-0.0000166	-0.0000146	-0.0000153	-0.0000146	-0.0000154	-0.0000140	-0.0000136
	(0.0000103)	(0.0000104)	(0.0000106)	(0.0000114)	(0.00000930)	(0.00000944)	(0.00000885)	(0.00000921)	(0.00000884)	(0.00000924)	(0.00000894)	(0.00000892)
Latrine Construction Support	-0.0140	-0.0152	-0.0189	-0.0217	-0.00674	-0.00590	-0.0102	-0.00859	-0.0103	-0.00839	-0.0116	-0.0123
	(0.0107)	(0.0110)	(0.0114)	(0.0135)	(0.00967)	(0.00985)	(0.00826)	(0.00864)	(0.00825)	(0.00868)	(0.00766)	(0.00769)
Sanitation Activity - Last 3 years	0.00573	0.00590	0.00643	0.00684	0.00469	0.00457	0.00518	0.00496	0.00520	0.00493	0.00538	0.00549
	(0.00596)	(0.00607)	(0.00654)	(0.00704)	(0.00592)	(0.00596)	(0.00588)	(0.00582)	(0.00588)	(0.00582)	(0.00587)	(0.00591)
Number of Masons	-0.000103	-0.0000981	-0.0000822	-0.0000700	-0.000134	-0.000138	-0.000120	-0.000126	-0.000119	-0.000127	-0.000114	-0.000110
	(0.0000901)	(0.0000930)	(0.000103)	(0.000116)	(0.0000942)	(0.0000957)	(0.0000918)	(0.0000889)	(0.0000918)	(0.0000890)	(0.0000893)	(0.0000901)
Female GP leader	0.00711	0.00724	0.00766	0.00798	0.00628	0.00619	0.00667	0.00649	0.00668	0.00647	0.00683	0.00692
	(0.00570)	(0.00582)	(0.00613)	(0.00657)	(0.00544)	(0.00547)	(0.00536)	(0.00539)	(0.00537)	(0.00539)	(0.00541)	(0.00545)
Previous year female GP leader	-0.00759	-0.00683	-0.00436	-0.00247	-0.0124	-0.0130	-0.0101	-0.0112	-0.0101	-0.0113	-0.00922	-0.00871
	(0.00997)	(0.0100)	(0.00949)	(0.0102)	(0.00817)	(0.00834)	(0.00757)	(0.00805)	(0.00755)	(0.00808)	(0.00761)	(0.00756)
N	38724	38724	38724	38724	38724	38724	38724	38724	38724	38724	38724	38724

Note: Standard errors in parentheses. *p<0.05, **p<0.01, ***p<0.001. Standard errors are clustered at GP level. IV1 refers to the IV where group fixed effects are calculated across all villages to create a constant. IV2 refers to the IV calculated by using IV multiplied by the difference between group average and sample average of contextual characteristics. As some IVs are combined, each column notes which combinations (if any) of household characteristics are used in the IVs utilised.

Table 1.A3: Marginal effects parameter estimates using Instrumental Variables (IV) - Latrine Usage

	IV2: Age	IV1 and IV2: Age	IV2: Gender	IV1 and IV2: Gender	IV2: Age, Gender	IV1 and IV2: Age, Gender	IV2: Household size	IV1 and IV2: Household size	IV2: Age, Household size	IV2: Gender, Household size	IV1 and IV2: Age, Household size	IV1 and IV2: Age, Gender, Household size	IV2: Age, Gender, Household size	IV1 and IV2: Age, Gender, Household size
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Constant	-0.0660	0.196	-0.181	-0.0810	-0.155	-0.0683	-0.273	-0.178	-0.247	-0.250	-0.159	-0.168	-0.229	-0.152
	(0.535)	(0.888)	(0.377)	(0.428)	(0.371)	(0.409)	(0.309)	(0.316)	(0.312)	(0.314)	(0.322)	(0.324)	(0.318)	(0.329)
Group Usage behaviour	0.989**	1.167*	0.910***	0.978***	0.928***	0.987***	0.848***	0.912***	0.865***	0.863***	0.925***	0.919***	0.878***	0.930***
	(0.313)	(0.509)	(0.130)	(0.135)	(0.117)	(0.123)	(0.0785)	(0.0650)	(0.0646)	(0.0742)	(0.0564)	(0.0611)	(0.0652)	(0.0553)
Avg. age household head	-0.00251	-0.00635	-0.000816	-0.00229	-0.00121	-0.00248	0.000528	-0.000864	0.000154	0.000199	-0.00114	-0.00101	-0.000116	-0.00125
	(0.00764)	(0.0121)	(0.00492)	(0.00492)	(0.00470)	(0.00472)	(0.00424)	(0.00379)	(0.00410)	(0.00423)	(0.00375)	(0.00386)	(0.00414)	(0.00383)
Avg. age household head squared	0.00000258	0.0000356	-0.0000120	0.000000696	-0.00000858	0.00000229	-0.0000235	-0.0000115	-0.0000203	-0.0000207	-0.00000918	-0.0000103	-0.0000180	-0.00000823
	(0.0000634)	(0.0000964)	(0.0000445)	(0.0000420)	(0.0000416)	(0.0000393)	(0.0000406)	(0.0000348)	(0.0000388)	(0.0000402)	(0.0000338)	(0.0000352)	(0.0000388)	(0.0000343)
Avg. Household Size	-0.0151	-0.0199	-0.0130	-0.0148	-0.0135	-0.0151	-0.0113	-0.0130	-0.0118	-0.0117	-0.0134	-0.0132	-0.0121	-0.0135
	(0.0123)	(0.0200)	(0.00808)	(0.00978)	(0.00815)	(0.00979)	(0.00704)	(0.00770)	(0.00684)	(0.00701)	(0.00779)	(0.00777)	(0.00694)	(0.00788)
Avg. no. Kids under 2 years	-0.363	-0.518	-0.294	-0.354	-0.310	-0.361	-0.239	-0.296	-0.255	-0.253	-0.307	-0.302	-0.266	-0.311
	(0.414)	(0.720)	(0.320)	(0.383)	(0.324)	(0.378)	(0.281)	(0.296)	(0.283)	(0.283)	(0.305)	(0.302)	(0.287)	(0.310)
Proportion Female Head	0.103	0.136	0.0890	0.101	0.0923	0.103	0.0777	0.0894	0.0808	0.0805	0.0918	0.0906	0.0831	0.0927
	(0.112)	(0.188)	(0.100)	(0.109)	(0.0987)	(0.108)	(0.0992)	(0.0911)	(0.0956)	(0.0977)	(0.0917)	(0.0928)	(0.0953)	(0.0934)
Proportion OBC	-0.0814	-0.133	-0.0587	-0.0784	-0.0640	-0.0809	-0.0408	-0.0594*	-0.0458	-0.0452	-0.0630*	-0.0613*	-0.0494	-0.0645*
	(0.0970)	(0.160)	(0.0438)	(0.0469)	(0.0417)	(0.0456)	(0.0273)	(0.0274)	(0.0255)	(0.0275)	(0.0269)	(0.0274)	(0.0266)	(0.0275)
Proportion SC	0.0569	0.00724	0.0787	0.0597	0.0737	0.0573	0.0961**	0.0781*	0.0913**	0.0919**	0.0746*	0.0763*	0.0878**	0.0731*
	(0.0938)	(0.158)	(0.0505)	(0.0576)	(0.0482)	(0.0548)	(0.0314)	(0.0339)	(0.0306)	(0.0330)	(0.0339)	(0.0354)	(0.0327)	(0.0356)
Proportion ST	0.0528	0.152	0.00912	0.0471	0.0192	0.0519	-0.0255	0.0104	-0.0159	-0.0171	0.0174	0.0140	-0.00892	0.0203
	(0.176)	(0.290)	(0.0847)	(0.0868)	(0.0784)	(0.0797)	(0.0771)	(0.0663)	(0.0714)	(0.0716)	(0.0626)	(0.0627)	(0.0677)	(0.0600)
Proportion DT	0.0814	0.176	0.0397	0.0760	0.0494	0.0805	0.00670	0.0409	0.0159	0.0148	0.0477	0.0444	0.0225	0.0504
	(0.176)	(0.282)	(0.0897)	(0.0925)	(0.0842)	(0.0878)	(0.0627)	(0.0580)	(0.0586)	(0.0633)	(0.0563)	(0.0590)	(0.0607)	(0.0581)
Proportion NT	0.0663	0.0810	0.0598*	0.0655*	0.0613*	0.0662*	0.0547	0.0600*	0.0561*	0.0560*	0.0611*	0.0606*	0.0572*	0.0615*
	(0.0414)	(0.0680)	(0.0289)	(0.0332)	(0.0295)	(0.0337)	(0.0287)	(0.0282)	(0.0281)	(0.0279)	(0.0286)	(0.0282)	(0.0277)	(0.0287)
Proportion FC	-0.0850	-0.103	-0.0770**	-0.0839*	-0.0788**	-0.0848*	-0.0706**	-0.0772**	-0.0724**	-0.0722**	-0.0785**	-0.0779**	-0.0737**	-0.0790**
	(0.0520)	(0.0900)	(0.0279)	(0.0327)	(0.0294)	(0.0353)	(0.0247)	(0.0260)	(0.0242)	(0.0245)	(0.0270)	(0.0261)	(0.0244)	(0.0272)
Segmented GP	0.00757	-0.00959	0.0151	0.00855	0.0134	0.00772	0.0211*	0.0149	0.0195*	0.0197*	0.0137	0.0143	0.0183	0.0132
	(0.0314)	(0.0495)	(0.0153)	(0.0156)	(0.0143)	(0.0146)	(0.00935)	(0.00873)	(0.00882)	(0.00969)	(0.00852)	(0.00901)	(0.00941)	(0.00891)
Households per GP	-0.00000302	0.00000939	-0.00000849	-0.00000373	-0.00000723	-0.00000313	-0.0000128	-0.00000834	-0.0000116	-0.0000118	-0.00000745	-0.00000788	-0.0000108	-0.00000709
	(0.0000241)	(0.0000382)	(0.0000120)	(0.0000129)	(0.0000116)	(0.0000126)	(0.0000102)	(0.00000966)	(0.00000953)	(0.00000979)	(0.00000947)	(0.00000945)	(0.00000941)	(0.00000938)
Latrine Construction Support	-0.0201	-0.0406	-0.0111	-0.0190	-0.0132	-0.0199	-0.00393	-0.0114	-0.00592	-0.00568	-0.0128	-0.0121	-0.00737	-0.0134
	(0.0360)	(0.0604)	(0.0147)	(0.0162)	(0.0131)	(0.0148)	(0.0110)	(0.0100)	(0.00911)	(0.00993)	(0.00892)	(0.00914)	(0.00857)	(0.00833)
Sanitation Activity - Last 3 years	0.00657	0.00903	0.00549	0.00643	0.00574	0.00655	0.00463	0.00552	0.00487	0.00484	0.00570	0.00561	0.00504	0.00577
	(0.00794)	(0.0133)	(0.00592)	(0.00687)	(0.00605)	(0.00696)	(0.00627)	(0.00628)	(0.00613)	(0.00606)	(0.00632)	(0.00621)	(0.00599)	(0.00626)
Number of Masons	-0.0000775	0.00000462	-0.000114	-0.0000822	-0.000105	-0.0000782	-0.000142	-0.000113	-0.000134	-0.000135	-0.000107	-0.000110	-0.000129	-0.000104
	(0.000172)	(0.000298)	(0.000106)	(0.000129)	(0.000106)	(0.000126)	(0.000105)	(0.000103)	(0.000101)	(0.0000996)	(0.000104)	(0.000102)	(0.0000979)	(0.000103)
Female GP leader	0.00257	0.00425	0.00183	0.00247	0.00200	0.00255	0.00124	0.00185	0.00140	0.00138	0.00197	0.00191	0.00152	0.00202
	(0.00616)	(0.0101)	(0.00565)	(0.00642)	(0.00561)	(0.00632)	(0.00545)	(0.00547)	(0.00532)	(0.00542)	(0.00549)	(0.00553)	(0.00534)	(0.00556)
Previous year female GP leader	-0.00388	0.00956	-0.00980	-0.00465	-0.00843	-0.00399	-0.0145	-0.00963	-0.0132	-0.0134	-0.00867	-0.00913	-0.0122	-0.00828
	(0.0228)	(0.0349)	(0.0135)	(0.0138)	(0.0122)	(0.0121)	(0.00854)	(0.00787)	(0.00814)	(0.00905)	(0.00755)	(0.00839)	(0.00877)	(0.00810)
N	37608	37608	37608	37608	37608	37608	37608	37608	37608	37608	37608	37608	37608	37608

Note: Standard errors in parentheses. *p<0.05, **p<0.01, ***p<0.001. Standard errors are clustered at GP level. IV1 refers to the IV where group fixed effects are calculated across all villages to create a constant. IV2 refers to the IV calculated by using IV multiplied by the difference between group average and sample average of contextual characteristics. As some IVs are combined, each column notes which combinations (if any) of household characteristics are used in the IVs utilised.

Table 1.A4: Marginal effects parameter estimates using Instrumental Variables (IV) - Open Defecation (OD)

	IV2: Age	IV1 and IV2: Age	IV2: Gender	IV1 and IV2: Gender	IV2: Age, Gender	IV1 and IV2: Age, Gender	IV2: Household size	IV1 and IV2: Household size	IV2: Age, Household size	IV2: Gender, Household size	IV1 and IV2: Age, Household size	IV1 and IV2: Age, Gender, Household size	IV2: Age, Gender, Household size	IV1 and IV2: Age, Gender, Household size
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Constant	-0.249	-0.392	-0.765	-0.831	-0.582	-0.658	-5.305	-6.794	-0.460	-0.758	-0.692	-0.841	-0.609	-0.717
	(0.710)	(0.608)	(0.572)	(0.589)	(0.490)	(0.480)	(23.12)	(24.03)	(0.593)	(0.570)	(0.548)	(0.592)	(0.487)	(0.481)
Group Usage behaviour	0.819***	0.852***	0.937***	0.953***	0.896***	0.913***	1.976	2.316	0.868***	0.936***	0.921***	0.955***	0.902***	0.926***
	(0.121)	(0.103)	(0.0967)	(0.0992)	(0.0726)	(0.0711)	(5.286)	(5.416)	(0.102)	(0.0961)	(0.0997)	(0.100)	(0.0727)	(0.0720)
Avg. age household head	-0.000646	-0.000381	0.000313	0.000436	-0.0000270	0.000115	0.00875	0.0115	-0.000253	0.000299	0.000176	0.000455	0.0000225	0.000223
	(0.00545)	(0.00519)	(0.00492)	(0.00493)	(0.00496)	(0.00492)	(0.0446)	(0.0486)	(0.00510)	(0.00492)	(0.00491)	(0.00494)	(0.00494)	(0.00489)
Avg. age household head squared	0.0000283	0.0000263	0.0000211	0.0000201	0.0000236	0.0000226	-0.0000431	-0.0000641	0.0000254	0.0000212	0.0000221	0.0000200	0.0000233	0.0000217
	(0.0000519)	(0.0000494)	(0.0000454)	(0.0000450)	(0.0000469)	(0.0000460)	(0.000338)	(0.000374)	(0.0000485)	(0.0000454)	(0.0000458)	(0.0000449)	(0.0000466)	(0.0000455)
Avg. Household Size	0.0126	0.0100	0.00328	0.00209	0.00658	0.00521	-0.0784	-0.105	0.00877	0.00342	0.00461	0.00191	0.00610	0.00416
	(0.0130)	(0.0118)	(0.0118)	(0.0124)	(0.00988)	(0.0102)	(0.417)	(0.429)	(0.0109)	(0.0118)	(0.0112)	(0.0125)	(0.00987)	(0.0104)
Avg. no. Kids under 2 years	-0.0538	0.0682	0.388	0.444	0.231	0.296	4.271	5.545	0.127	0.381	0.325	0.453	0.254	0.346
	(0.609)	(0.520)	(0.504)	(0.521)	(0.425)	(0.419)	(19.86)	(20.68)	(0.509)	(0.502)	(0.475)	(0.525)	(0.424)	(0.423)
Proportion Female Head	-0.0959	-0.0824	-0.0471	-0.0409	-0.0644	-0.0572	0.382	0.522	-0.0759	-0.0478	-0.0541	-0.0399	-0.0619	-0.0517
	(0.132)	(0.123)	(0.105)	(0.108)	(0.105)	(0.105)	(2.394)	(2.550)	(0.117)	(0.104)	(0.118)	(0.109)	(0.104)	(0.106)
Proportion OBC	0.0110	0.0180	0.0363	0.0396	0.0273	0.0311	0.260	0.333	0.0214	0.0360	0.0327	0.0401	0.0287	0.0340
	(0.0412)	(0.0376)	(0.0298)	(0.0301)	(0.0295)	(0.0290)	(1.173)	(1.214)	(0.0351)	(0.0298)	(0.0343)	(0.0302)	(0.0292)	(0.0286)
Proportion SC	-0.0839	-0.0730	-0.0444	-0.0393	-0.0584	-0.0526	0.303	0.417	-0.0678	-0.0450	-0.0501	-0.0386	-0.0564	-0.0481
	(0.0558)	(0.0493)	(0.0454)	(0.0474)	(0.0393)	(0.0395)	(1.781)	(1.860)	(0.0450)	(0.0453)	(0.0440)	(0.0476)	(0.0388)	(0.0395)
Proportion ST	0.0752	0.0427	-0.0425	-0.0576	-0.000760	-0.0181	-1.078	-1.417	0.0270	-0.0408	-0.0257	-0.0599	-0.00683	-0.0314
	(0.154)	(0.133)	(0.112)	(0.114)	(0.0971)	(0.0940)	(5.303)	(5.442)	(0.130)	(0.111)	(0.120)	(0.115)	(0.0967)	(0.0940)
Proportion DT	0.0310	0.0125	-0.0360	-0.0446	-0.0122	-0.0221	-0.625	-0.819	0.00360	-0.0350	-0.0264	-0.0459	-0.0157	-0.0297
	(0.0927)	(0.0825)	(0.0772)	(0.0779)	(0.0695)	(0.0679)	(3.062)	(3.167)	(0.0798)	(0.0770)	(0.0764)	(0.0782)	(0.0690)	(0.0674)
Proportion NT	-0.0353	-0.0467	-0.0766	-0.0820	-0.0620	-0.0681	-0.441	-0.560	-0.0522	-0.0760	-0.0707	-0.0828	-0.0641	-0.0728
	(0.0565)	(0.0488)	(0.0505)	(0.0520)	(0.0420)	(0.0420)	(1.898)	(1.945)	(0.0486)	(0.0502)	(0.0470)	(0.0524)	(0.0422)	(0.0428)
Proportion FC	0.0464	0.0437	0.0366	0.0354	0.0401	0.0386	-0.0497	-0.0781	0.0424	0.0368	0.0380	0.0352	0.0396	0.0375
	(0.0329)	(0.0307)	(0.0323)	(0.0334)	(0.0298)	(0.0302)	(0.485)	(0.541)	(0.0303)	(0.0322)	(0.0303)	(0.0336)	(0.0299)	(0.0307)
Segmented GP	-0.0178	-0.0151	-0.00806	-0.00681	-0.0115	-0.0101	0.0776	0.106	-0.0138	-0.00820	-0.00945	-0.00662	-0.0110	-0.00897
	(0.0145)	(0.0131)	(0.0119)	(0.0123)	(0.0107)	(0.0107)	(0.444)	(0.454)	(0.0121)	(0.0119)	(0.0118)	(0.0124)	(0.0106)	(0.0106)
Households per GP	0.00000714	0.00000335	-0.00000661	-0.00000838	-0.00000173	-0.00000376	-0.000128	-0.000167	0.00000151	-0.00000641	-0.00000465	-0.00000864	-0.00000244	-0.00000532
	(0.0000176)	(0.0000157)	(0.0000151)	(0.0000156)	(0.0000127)	(0.0000128)	(0.000626)	(0.000637)	(0.0000151)	(0.0000150)	(0.0000151)	(0.0000157)	(0.0000127)	(0.0000129)
Latrine Construction Support	-0.00387	-0.00120	0.00580	0.00704	0.00237	0.00380	0.0909	0.119	0.0000899	0.00566	0.00442	0.00723	0.00287	0.00489
	(0.0130)	(0.0117)	(0.0105)	(0.0108)	(0.00928)	(0.00925)	(0.438)	(0.454)	(0.0112)	(0.0105)	(0.0111)	(0.0109)	(0.00925)	(0.00929)
Sanitation Activity - Last 3 years	-0.00487	-0.00371	-0.000652	-0.000110	-0.00215	-0.00153	0.0365	0.0487	-0.00315	-0.000713	-0.00125	-0.0000286	-0.00193	-0.00105
	(0.00830)	(0.00762)	(0.00896)	(0.00914)	(0.00797)	(0.00800)	(0.197)	(0.207)	(0.00777)	(0.00893)	(0.00793)	(0.00919)	(0.00802)	(0.00815)
Number of Masons	0.000216	0.000197	0.000147	0.000138	0.000172	0.000161	-0.000460	-0.000660	0.000188	0.000148	0.000157	0.000137	0.000168	0.000154
	(0.000149)	(0.000138)	(0.000137)	(0.000143)	(0.000126)	(0.000128)	(0.00310)	(0.00334)	(0.000138)	(0.000137)	(0.000141)	(0.000144)	(0.000127)	(0.000132)
Female GP leader	0.00219	0.00126	-0.00117	-0.00160	0.0000221	-0.000473	-0.0307	-0.0403	0.000813	-0.00112	-0.000689	-0.00166	-0.000151	-0.000852
	(0.00677)	(0.00648)	(0.00708)	(0.00721)	(0.00648)	(0.00649)	(0.157)	(0.167)	(0.00662)	(0.00705)	(0.00679)	(0.00725)	(0.00652)	(0.00662)
Previous year female GP leader	0.0170	0.0140	0.00596	0.00453	0.00989	0.00825	-0.0915	-0.123	0.0125	0.00612	0.00754	0.00432	0.00931	0.00700
	(0.0151)	(0.0133)	(0.0123)	(0.0125)	(0.0109)	(0.0106)	(0.494)	(0.515)	(0.0128)	(0.0122)	(0.0121)	(0.0126)	(0.0108)	(0.0106)
N	37608	37608	37608	37608	37608	37608	37608	37608	37608	37608	37608	37608	37608	37608

Note: Standard errors in parentheses. *p<0.05, **p<0.01, ***p<0.001. Standard errors are clustered at GP level. IV1 refers to the IV where group fixed effects are calculated across all villages to create a constant. IV2 refers to the IV calculated by using IV multiplied by the difference between group average and sample average of contextual characteristics. As some IVs are combined, each column notes which combinations (if any) of household characteristics are used in the IVs utilised.

2 Caste and sanitation decision making in India

Abstract

This chapter assesses the impact of caste and jati based endogenous social effects on individual household sanitation behaviours across 120 villages in rural Maharashtra, India. Latrine ownership, latrine usage, and open defecation are the sanitation behaviours of interest. It is found that a household's own caste and jati group's sanitation behaviour, positively affect the likelihood of a household exhibiting the same behaviour. Endogenous social effects from the behaviour of other caste groups on own household behaviour are also estimated at caste level. This chapter finds that endogenous social effects from the choice of other caste groups differ depending on one's caste relative to the comparison caste group.

2.1 Introduction

Access to sanitation and safe water is goal 6 of the UN's Sustainable Development Goals. As of 2020, 1,7 billion people live without basic sanitation.[19] Of those, 494 million persons practise open defecation (WHO - UNICEF 2021). Open defecation (OD) is defined as excreting outdoors in open fields, forests, rivers, outside the home, gutters et cetera. The consequences of persistent and wide spread OD include increased child mortality, malnutrition, gender inequity, limited access to education, negative externalities on the environment, and disproportionate impacts on vulnerable groups such as the elderly and handicapped.

As of 2012 India captured 59% of global OD (UNICEF - WHO 2012). However, from 2015-2020 OD reduced in India by 14% making it amongst countries making the greatest progress in reducing OD (WHO - UNICEF 2021). However, progress has been significantly varied regionally with some states reaching only OD rates of 1% whilst others are still above 50%. The country has enacted a range of sanitation related policy over the decades. The most recent government attempt being the Swachh Bharat Abhiyan (Clean India Mission); a nation wide policy campaign from 2014-2019 which aimed to achieve nationwide sanitation coverage. This included the building of household-owned and community-owned latrines and making the nation universally OD free. The campaign claims to have met its goal of 100% sanitation coverage and thus claims to be OD free.[20] Notwithstanding, if 100% sanitation coverage has been achieved, the discrepancy between latrine ownership and usage may still persist as the normative behaviour of OD continues. Subsequently, an understanding of social factors that perpetuate these norms would be useful for policy targeting that ensures sustainable behaviour change in OD. This will help ensure latrines offered are used consistently, and that future sanitation policy can be more effective at sustaining change. This includes Phase 2 of Swachh Bharat Abhiyan known as ODF plus which aims to reinforce latrine use behaviours.

[19]Where basic sanitation is defined as "having access to facilities for the safe disposal of human waste (faeces and urine), as well as having the ability to maintain hygienic conditions, through services such as garbage collection, industrial/hazardous waste management, and wastewater treatment and disposal."

[20]The factuality of this has been challenged (Exum et al. 2020, Gupta et al. 2020).

This chapter aims to investigate the impact of caste social interactions on sanitation behaviour, by shedding light on the caste dynamics that influence the likelihood of household choice in latrine ownership, usage and OD behaviour. The research questions of interest are: Does the sanitation behaviour of households in your caste influence your own sanitation behaviour? Does the sanitation behaviour of other caste groups influence your own sanitation behaviour? To address these questions, neighbourhood level caste groups are used to identify the endogenous social effects of sanitation behaviour.

The social effects of interest in this study are endogenous social effects; how behaviour of the group affects the likelihood of the same behaviour by the decision making agent - in this case the household. Most literature so far has supposed that social effects in sanitation adoption operate as strategic complements (Guiteras et al. 2019, Geruso and Spears 2017, Gautam (2018a, 2018b), Jenkins and Cairncross 2010). That is to say, households' utility and decision making follow the decision of the group. The strategic complements may arise through social learning, desire to conform, or an expression of signalling status - a form of 'Keeping up with the Joneses'. When the social group is one's caste, these strategic complementarities can come from preferences for in-group association or greater willingness to be guided by in-group social norms. Mechanisms to facilitate these strategic complements can include information sharing, herd behaviour, or visual observation of peers' choices. One can see how in the case of caste due to greater in-group interactions, information sharing or observation of other's conduct is more accessible to agents thereby increasing the chances of these mechanisms working within caste to shape outcomes.

When measuring social effects, it is key to use a relevant metric for social interactions and outcomes as the basis for social ties. In this chapter caste and jati (sub-caste) are the metric through which social networks are defined. The value of caste grouping comes from the importance this social identity holds in Indian society and its embeddedness in Indian social institutions. One's caste group shapes a range of socio-economic outcomes such as labour market participation (Chandrasekhar and Mitra 2019), marriage choice (Munshi 2019), social class and economic access

(Munshi 2016).[21] Both caste and jati can hold important information about social networks. Broad caste reflects the general hierarchy of class based groupings of people, while the jati represent a more socially cohesive sub-grouping reflecting clan or tribe. The use of these two distinct groupings can inform different impacts of social interaction effects on sanitation behaviour. Jati as a more intimate social tie may exhibit greater intensity of norms around trust and reciprocity leading to tight-nit social networks that could have a more powerful social effect on sanitation behaviour. Caste however, which is more reflective of social hierarchy groupings may have broader networks that are associated through wealth, social status and education through which sanitation behaviour may be influenced.

The relevance of these groupings in relation to sanitation comes from norms in Indian social dynamics around the role of different groups when it comes to activities such as sewage maintenance (and relatedly toilet maintenance). For example, manual scavenging was a role primarily attributed to lower caste groups known as Dalits. The caste system is also tied to traditional norms around an idea of ritual purity which is noted to include prescriptions around the ownership and usage of a latrine within one's household (Coffey and Spears 2017).

Though previous works (Guiteras et al. 2019) have shown the impact of social effects on sanitation, and others have expressed the importance caste plays in sanitation behaviour (Gatade, 2015), there remains a gap in quantitatively showing the effects of caste on sanitation adoption. Therefore, this study contributes to the literature on social interactions, caste and sanitation behaviour. It estimates correlation between caste group and sanitation outcomes through the metric of social effects. It focuses on caste and jati as the reference groups for which social interactions in sanitation behaviour take place.

This chapter follows Brock and Durlauf (2000, 2001, 2007) binary response model of strategic interactions to identify caste based social effects on sanitation behaviour. The non-linearity of that framework ensures social effects are separately identified. Models of social effects are known to exhibit multiple equilibria which can be an issue

[21] Though affirmative action programmes have been set out in Indian legislation, aspects of the institution still hold strong in daily society and caste based discrimination still prevails.

for estimation. The model is estimated using a data in equilibrium assumption as used by Gautam (2018a). This assumption simply holds that the outcomes present in the data reflect and can thus proxy for the equilibrium.[22] Correlated shocks are addressed with the separate and joint inclusion of village level fixed effects and caste fixed effects for different model specifications.

It is found that the sanitation behaviour of one's caste positively influences the likelihood of the same behaviour by an individual household in the same caste. This persists within caste and jati (sub-caste) levels. However, the social effects are of slightly larger magnitude for caste than jati, across all behaviours. For example, increasing own caste (jati) group latrine uptake by 10% increases the likelihood of own uptake by 8% (7%) in model 1 (no fixed effects). This falls to 6% (4%) when both caste (jati) group and village level fixed effects (FE) are included (model 4). The inclusion of group and village level fixed effects allows one to control for unobserved factors at both the village and caste level that may influence sanitation outcomes. That is, it controls for group level unobservables.[23] Similar magnitudes are present with latrine usage and OD behaviour.

When comparing the behaviour of other caste groups to household behaviour, complementary social effects are also found between caste groups even when taking account of the social hierarchy that exists between caste groups. However, these complementary effects are larger for lower castes than they are for higher castes. That is, social effects of higher caste behaviour are generally more impactful on lower castes, than social effects of lower caste behaviour on higher caste groups. Additionally, in most cases the magnitude of effects of other caste behaviour is lower than that of own caste behaviour.

For policy, this implies that increasing access to sanitation can be achieved cost-effectively by aiming behaviour change policy in latrine usage towards few key actors within various castes, enabling spillovers to occur both within and between caste groups. The idea being that when some within the caste group adopt, the strategic

[22] Additionally as shown in chapter 1, this methodology is simple and estimates social effects better than others used.

[23] This does not include controlling for within-caste (jati) in GP unobservables that may exist

complementarity can spread across the group. The importance still lies on targeting policy toward increased latrine usage and reduced OD as opposed to latrine ownership alone.

Section 2.2 provides context on the Indian caste system and its connection to social interactions, caste norms and sanitation. Section 2.3 discusses the theoretical framework through which endogenous social effects are identified. Section 2.4 covers the empirical strategy, describes the data, and the estimation procedure. Section 2.5 discusses the results and Section 2.6 concludes.

2.2 Context

2.2.1 Sanitation and social networks

Sanitation infrastructure in the form of sewage systems is often too resource intensive for governments in developing countries to willingly implement in rural areas. As an alternative sanitation access has been improved in rural areas by provision of cost-friendly basic latrines of varying designs that can be supplied on a household basis (Brikké and Bredero 2003). These latrines are simpler and of a different nature to the flush toilet that is prevalent in most developed countries and urban cities.

A variety of interventions such as subsidies, education and awareness raising campaigns and programs have been used to facilitate sanitation improvements. This has gone some way in increasing sanitation coverage, and reducing the incidence of OD (Garn et al. 2017). Despite increase in coverage, the reduction of OD also requires behaviour change beyond information and education (Coffey et al. 2014, Shakya et al. 2015, Vyas and Spears 2018). The act of OD in many communities has persisted for generations, developing it into a social norm. Therefore access to latrines does not necessarily drop the incidence of OD within a community (Coffey et al. 2015, O'Reilly et al. 2017a, Augsburg and Rodriguez-Lesmes 2018).

Social interactions have been identified as affecting likelihood of adopting new sanitation behaviour such as constructing a latrine (Stopnitzky 2017). Of key importance in social interaction analysis is attaining accurate data on agents' social ties, and realm of influence. Failing detailed network data on every kind of relationship an agent has with other agents, using social identities as a reference point helps provide social identifiers by which people are affected. This could be through peer effects, information sharing, herd behaviour amongst others. In this sense, it is important to choose a reference for defining social networks that is both relevant to the agents and the outcome of interest. In the Indian context caste is intricately linked to perceptions around sanitation (Coffey and Spears 2017). These include perceptions around purity, pollution and having latrines in one's household (O'Reilly et al. 2017b, Vyas and Spears 2018, Coffey and Spears 2017), as well as sanitation maintenance like sewage cleaning being historically ascribed to specific caste groups (Gatade 2015). Alongside being relevant to the outcome of interest, caste is also a strong social identifier in other aspects of Indian society and life (Munshi, 2019).

2.2.2 Indian Caste System

Caste in the Indian society is a form of social stratification. Key elements of it include hierarchy, endogamy, and social exclusion. Hierarchy here refers to the linear ordering of people into social classes, also referred to as varnas. There are four main varnas within the Indian caste system with a fifth group excluded from the class structure altogether referred to as Dalits or Untouchables. Endogamy refers to marrying within one's own social group. Within the Indian caste system, this is reflected by socially restricting marriage to those of one's own jati. Jati is a subgrouping of varna that refers to clan-like associations. It is said there are about 4000 different jatis operating in India (Munshi, 2019). Social exclusion relates to rules around interactions that can exist between varnas and between jatis. For example a practice known as "untouchability" considers those of the excluded dalit class as polluting to physically interact with and thus would segregate them from mainstream society (Bathran, 2011). These three elements can make the Indian caste system akin to a mixture of religion (given its Hindu origins), ethnicity (given its endogamous aspect), and class (given its hierarchal structure).

The operation of the caste system has a long standing history in Indian society, though linked to the Hindu religion, its contemporary influence persists beyond religion; with caste ideologies such as endogamy also operating within Christian and Muslim Indian communities (Munshi, 2019). The prevalence of the caste system endures in significant aspects of the average life of a person in India. This includes economic, social, and naturally religious life. Caste doctrine shapes the kind of access to education and labour market opportunities an individual gets, who they marry, the religious practices they follow, who they interact with and how they can interact with others socially. For example the persistence of endogamy has meant that over 95% of Indians still marry within their jati (Munshi, 2019; Stopnitzky, 2017).

The hierarchal aspect of the caste system, mixed with the ethnic boundaries that comes with endogamy, characterises a strong within caste interaction, and between caste exclusion in Indian culture. Caste interactions permeate "otherness", resulting in the marginalisation of groups of people in ways that is similarly depicted in other social segregation and stratifications of social identity – sex, gender, race, ethnicity, class, religion et cetera. Its ability to support in-group social mobility is paralleled by its ability to be exclusionary or be expressed at the explicit cost of others. Arora and Sanditov (2015) describe this as a "sharedness" shaped by a "sense of belonging" to those one shares social commitments and expectations with, breeding a cultural identity that can feed exclusion of those belonging to an "inferior identity".

The between castes exclusion is marked by limitations in inter-caste relations. This extends beyond marriage and enters other dimensions of day-to-day life. Neighbourhood structures have also evolved so Dalits on occasion technically reside outside the main village community (Bathran, 2011; Coffey and Spears, 2017), and different castes may share different sources of water (Dutta, Behera and Bharti 2015).

Indian social identity being strongly defined across caste lines makes it a useful reference point for social interaction analysis in this context. This does not deny the fact that multi-caste based networks exist but simply to say that caste is a significant determinant of social network formation for rural environments in India (Arora and Sanditov 2015).

2.2.3 Caste social interactions and Sanitation

OD behaviour exists across all caste groups in rural India (Spears and Thorat 2019). Coffey and Spears (2017) use qualitative methods to show that sanitation behaviour in India is marked by social norms around open defecation which feeds persistently poor sanitation practice despite government actions to improve sanitation outcomes.

If an individual household derives utility from following the collective behaviour of the group, and the group actualises its identity through certain practices, it becomes clear that these strategic social interactions are what feeds the perpetuation of certain norms - such as OD. One said social norm specific to OD and caste is that of ritual purification, pollution and untouchability. The concept of purity and pollution when fixated upon, and applied to a contemporary context means that having a latrine in one's household premises is perceived unclean, and thus undesirable (Vyas and Spears 2018). As such the caste norm of untouchability for instance, when actualised in collective group behaviour, can create spillovers on agents by feeding the unwillingness to construct or use a latrine and instead chose to OD. Spears and Thorat (2019) show that the commonality of the practice of untouchability was linked to higher prevalence of OD. In their empirical analysis, the practice of untouchability was measured at the individual household and village level (excluding self) and regressed against the likelihood of OD. Though they did not estimate social effects on OD, the work showed the link between certain caste based practices and OD.

Another norm around caste and OD is noted to derive from the occupational guilds the caste system and British colonialism in the region historically built (Munshi 2016). Removal of human excreta, sewage cleaning, waste disposal, rearing of pigs, removal of human and cattle corpse and such tasks that involved ritually polluting substances, were roles ascribed to certain lower castes. This modernised into a dichotomy whereby personal involvement with faeces in any manner that includes some sanitation practices and upkeep (such as emptying latrines) is something left to the lower castes that were known to have these roles, specifically Dalits (Coffey and Spears, 2017; Vyas and Spears, 2018). This can further limit sanitation adoption as upper caste members may avoid affordable latrines to maintain ritual purity.

Lower caste members may avoid latrine adoption, maintenance and replacement to avoid discriminatory affiliation of "impurity" or further re-subjection to these jobs and as their hereditary lot in life. The avoidance of the social stigma thus exists for both lower and upper caste households (Vyas and Spears 2018), reinforcing OD behaviour at the collective and individual level. This in part motivates calculating the magnitude to which the collective feeds the individual in sanitation behaviour.

Perceptions around sanitation in India are not limited to those based on caste alone. Agents are known to have positive ideas around the benefits of OD (Coffey et al. 2014) such as perceiving it as a wholesome activity. Some farmers perceive it as efficient practice during their work day. Some women see it positively as an opportunity to socialise with other women in the community (Routray et al. 2015). Nonetheless, the relationship between caste and sanitation is argued to be so strong that attempts to delink this relationship whilst trying to improve sanitation access is unrealistic (Gatade 2015, O'Reilly et al. 2017b, Pais 2021).

A number of qualitative studies have argued that caste relations around sanitation plays a role in which households are more likely to OD in environments where the practice of untouchability is more common (Spears and Thorat 2019), and the ideas of pollution and purity are more strongly enforced (Vyas and Spears 2018, O'Reilly et al. 2017b).

Gautam (2018a), Dickinson and Pattanayak (2009) and Shakya et al. (2015) empirically show social interactions influence sanitation adoption in rural parts of India. These papers use neighbourhood and village level groupings as the basis for quantifying such social effects. Any reference to caste largely relied on using it as a control, as opposed to the reference by which networks were defined and used to measure the likelihood of adoption through social effects.

Despite known intricacies between caste and sanitation in India, little to no empirical work has analysed how caste relations (at both a caste and jati level) affect likelihood of specific sanitation behaviours. As noted caste level groupings hold strong class based hierarchies, and jatis reflect more potentially cohesive social groups around clan, tribe, and language. How the behaviour of one's group or other groups' impact the likelihood of latrine ownership, latrine usage, and OD could operate in different directions or magnitudes dependent on caste or jati (sub-caste) as the social reference point. A clearer understanding on how different caste based ties shape decision making for sanitation would be pragmatic for sanitation policy formation and targeting. This can help policy design regarding the socio-demographic structures that need to be accounted for in order to nudge sanitation decisions in the right direction. It could also aid cost effectiveness on whom to target policy actions that creates a sufficiently large social multiplier to ensure universal access and usage.

2.3 Methodology

2.3.1 Theoretical Framework

The theoretical model used to estimate endogenous social effects is Brock and Durlauf (2001, 2007) binary response model with strategic interactions. It follows in part Manski's (1993) social effects framework which structures social effects in three forms – contextual effects, correlated effects and endogenous effects. Contextual effects are determined by average characteristics shared by the group. Examples in this setting include group average age of household head or proportion of households headed by women within the group. Correlated effects are shaped by similar factors or personal characteristics that members of the same group face. This can be environmental factors such as the presence of masons in the vicinity or group level access to water facilities. Endogenous effects are the response of the individual household from the group's collective behaviour. Of specific interest in this setting are endogenous effects based on sanitation behaviour. The model applied is structured such that these different social effects can be separately identified and estimated. The focus is to determine the endogenous effects of caste and jati affiliation on latrine ownership, usage and OD behaviour.

The linear in means model for social effects makes identification of endogenous effects tasking given the reflection problem (Manski 1993). Brock and Durlauf (B&D) offer a non-linear model that circumvents the reflection problem allowing separate identification of endogenous, contextual and correlated social effects. Of importance is identifying a relevant reference group to shaping social interactions and their resultant effects on outcomes. The social reference group must sufficiently capture social interactions, but also be tied to the outcomes of interest. The use of caste and sub-caste as the relevant reference group in this chapter is drawn from the relevance these two groups have in shaping socio-economic dynamics, as well as the role they play in prescriptions around latrine ownership in the household, latrine maintenance and thus usage, and open defecation.

2.3.2 Setup

The B&D binary response model is well suited to this setting where household sanitation decision is represented in binary form: to own or not own a latrine, to use or not use a latrine, to OD or not OD. The model has a random utility structure with a finite number of decision-making households that represent the population N:

$$i = 1, ..., N$$

The sample is partitioned into sub-populations reflecting the social groups. Each household belongs to a group (g). The model is applied to two different contexts of representative social groups: caste within the village and jati within the village. These groups are indexed by:

$$g = 1, ..., G$$

which total G number of groups, and each group represented by:

$$I_g$$

The total number of agents is expressed as the sum of the number of agents per group, across groups, making up the total population:

$$\sum_{g=1}^{G} I_g = N$$

So for caste, the population is the sum of the number of households per caste group in the village. The same applies for jati.

Household utility is represented as:

$$V(\omega_i) = u(\omega_i) + S(\omega_i, \mu_i^e(\omega_{-i})) + \varepsilon(\omega_i) \tag{1}$$

where ω_i represents the decision, for example latrine ownership. $u(\omega_i)$ is the private deterministic aspect of utility from that decision. In other words, individual agent's characteristics that determine the utility of a specific sanitation behaviour. $S(\omega_i, \mu_i^e(\omega_{-i})$ is the social component of utility, that is utility from the decision based on social expectations. This makes $\mu_i^e(\omega_{-i})$ the agent's expectation of the decision made by others. $\varepsilon(\omega_i)$ is the individual error term. In this chapter's context, the individual agent is represented by an individual household.

The social utility component captures household's subjective expectation of the decision of others within the reference group of interest (own group or other group). This expectation is represented by the proportion of those within the group (excluding self in own group case) that make a specific choice:

$$\mu_i^e(\omega_{-i}) = \overline{m}_i^e = (I_g - 1)^{-1} \sum_{j \neq i} m_{(i,j)}^e \tag{2}$$

2.3.3 Assumptions

1. Strategic complementarity:

$$\frac{\partial^2 S(\omega_{ig}\overline{m}_{-ig})}{\partial \omega_{ig}\overline{m}_{-ig}} = J > 0 \tag{3}$$

The above expresses the relationship between household behaviour, reference group behaviour, and social utility. It presents a positive relationship between household sanitation decision and group sanitation decision. The more households in one's reference group that make a certain choice on sanitation, the more attractive that choice is to the household's own choice set on sanitation. The applicability of this assumption in the current context is validated on the belief that same caste groups share peer relations and so are more likely to associate with each other in similar conduct. However, in a situation where members of another group influence one's own behaviour, it is arguable that this assumption may fail, that is, when comparing the behaviour of other caste groups to own household behaviour. A hypothetical of this might be a situation in which higher ranked caste groups are less likely to pursue a behaviour that lower caste groups pursue creating social distance or disassociation with the ways of those lower caste groups. In such a setting strategic substitutes would be the case; which is a negative relationship between household sanitation decision and group sanitation decision. The hypothesis of this chapter is that strategic complementarities exist for both own and other group behaviour in sanitation. The intuition behind this is motivated by the extra costs involved in latrine building, and associated status that may come with owning a latrine. Building a private latrine in one's household costs money, and so to be able to achieve this can be a signal of wealth; a form of conspicuous consumption. Additionally building a latrine enables others to become more familiar with the process, the costs and other opportunities for learning that may limit households' decision making on whether to construct and use latrines or not.

2. Non-cooperative decision making: Households do not work together in their determination of sanitation behaviour. This allows social effects to be modelled as a strategic response from non-coordinating agents based solely on their own self-interest.

3. No group level unobservables: This requires that no group level factors that impact likelihood of individual adoption are left unaccounted for. If this is not met, identification of endogenous social effects is not feasible because shared group aspects that correlate with household sanitation may confound the behavioural effect one is trying to uncover. Though a strong assumption in other settings, here the inclusion of caste (or jati), and village level fixed effects in the model can capture potential correlated effects that may exist at the group levels. This includes common factors within caste groups across villages, and common factors across caste groups within

villages. However the inclusion of these fixed effects do not control for any potential correlated effects that may exist at the caste by village level.[24]

4. Random assignment into reference group: This ensures exogeneity in reference group formation. This is met because one does not choose the caste or jati they are born into. As for neighbourhood, though there is choice in where one lives, it can be presumed that in a rural setting, family residences exist generationally, allowing rigidity in neighbourhood assignment for current households.[25]

5. Difference between error terms are logistically distributed:

$$Prob(\varepsilon(0) - \varepsilon(1) \leq x) = \frac{1}{1 + exp(-\beta x)}$$

This imposes a non-linearity in the structure between one choice and another.

6. Error terms are independently and identically distributed (i.i.d). However, in the more realistic case where error terms are likely correlated within social groups, the assumption can be generalised to median independence of error terms. This addresses the potential issue of heteroskedasticity.

The difference in the utility a household gains from one sanitation choice $u_i(1)$ compared the other $u_i(0)$ is based on 5 main factors:

- Individual household characteristics (X_{ig}) such as age or gender of household head.
- Group characteristics (Z_g) which highlight contextual effects. This is represented by the group average of the individual characteristics.

[24] Interacting caste fixed effects and village effects may be one way to address this, but it may introduce too many fixed effects into the model risking the incidental parameter problem. Another way to address this could be to account for caste level dynamics within villages such as proportion of group sizes within the village.

[25] Additionally in the dataset for which this model is applied, approximately 90% of households are noted to have resided in the same village their entire lives.

- Average group behaviour (m^e_{-ig}) which expresses endogenous effects represented by coefficient J.

- Correlated group characteristics (α_g) which are those environmental or normative factors shared by the group. These are captured with the inclusion of fixed effects at group and village level.

- Individual household errors (ε_i) which are unknown factors specific to each household.

Household choice is based on the difference in outcome between the utilities:

$$u_i(1) - u_i(0) \geq 0$$

where determination of sanitation behaviour depending on the above factors follows the specification:

$$u_i(1) - u_i(0) = k + cX_i + dZ_g + Jm^e_{i,g} + \alpha_g - \varepsilon_i \tag{4}$$

Making the model to be estimated:

$$P(y_{ig} = 1) = F(k + cX_i + dZ_g + Jy_g + \alpha_g + \varepsilon_i)$$

where y_{ig} is the household's discrete choice on sanitation behaviour - to own or not own a latrine; to use or not use a latrine; and to OD or not OD, and household expectation of group behaviour($m^e_{i,g}$) is the actual group behaviour (y_g). The feasibility of this comes from symmetric rational expectations in group behaviour.

2.4 Empirical Strategy

2.4.1 Identification

B&D prove the model is identified under specific assumptions which prescribe sufficient variation in individual characteristics and group characteristics. The endogeneity (or reflection) problem is largely based on having a linear model structure.

A linear model without further assumptions makes it difficult to separately identify group characteristics from group behaviour. The identification assumptions in the B&D model reveals a non-linearity that addresses this through two broad intuitions. Firstly, there is a difference in support of the variables between the group behaviour (y_g), and the group level factors (Z_g) and (α_g), that determine behaviour. Secondly, there is a non-linearity between individual behaviour (which is limited strictly as 0 or 1), and group behaviour (which is a proportion range between 0 and 1). These are further supported by the necessity for attributes to vary within and across groups such that there is only a partial dependency between variables. These factors allow for the separate identification of endogenous social effects from exogenous social effects.

The issues of identification noted in chapter 1 may also apply in this chapter. However, the use of caste as the reference group provides greater support for the random assignment assumption. On the other hand, the assumption of non-cooperative decision making may be strict if intra-caste coordination takes place for latrine adoption and usage.

2.4.2 Equilibrium and empirical tractability

A self-consistent equilibrium exists in this model because of rational expectations and symmetric individual household expectations. By virtue of the symmetry, expectations on what the mean decision by others on sanitation behaviour would be, becomes the actual mean outcome. Hence how ($m^e_{i,g}$) translates to (y_g).

There is also the possibility of multiple equilibria depending on the strength of endogenous social effects compared to the private deterministic aspects. According to B&D, there exists a threshold for endogenous effects J in the strategic complementarity equation (3). Below the threshold there is a unique equilibrium, and above it there are multiple equilibria. When the group's behaviour has a strong influence on the household due to the endogenous social effect J, there can be multiple expected group behaviours that meet self-consistent complementary expectations $m^e_{i,g}$. In other words, when endogenous social effects are sufficiently high such that households are more swayed in their behaviour based on the group behaviour, then the equilibrium outcome can realise itself on a number of dimensions that are compatible with individual household utility outcomes which is what renders $(m^e_{i,g})$ as (y_g). Whilst in other cases multiple equilibria can be a problem for identification, in this context, the existence of multiple equilibria can further facilitate identification of endogenous social effects (B&D 2007). When complementary endogenous effects are present and sufficiently high to create multiple equilibria, it indicates that endogenous social effects are present. This is because a key property that models with endogenous social effects exhibit is the presence of multiple equilibria.

Each household maximises their expected utility based on their own choice and the expected choice of others within the group. The household opts for a decision if the utility from that decision is higher than the utility from making another decision:

$$y_i = \begin{Bmatrix} 1 \text{ if } \tilde{U}_1 \geq \tilde{U}_0 \\ 0 \text{ if } \tilde{U}_0 \geq \tilde{U}_1 \end{Bmatrix}$$

The decision making rule function of each household is:

$$\psi_i(\omega_g, \varepsilon_i; \theta) = arg \max_{y \epsilon Y} [\tilde{U}_d(y_i, \omega_g; \theta) + \varepsilon_i(y_i)] \tag{5}$$

Optimal utility is defined as:

$$\tilde{U}^*(y_i, \omega_g, \varepsilon_i(y_i); \theta) = \max_{d \epsilon 0,1)} [\tilde{U}_1(y_i = 1, \omega_g \varepsilon_i(1); \theta), \tilde{U}_0(y_i = 0, \omega_g, \varepsilon_i(0); \theta)] \tag{6}$$

The probability of every other households' adoption decision (from the perspective of the decision making household) is represented by:

$$p_{-i}(y_{-ig} | \omega_g, \theta)$$

The group probability of adopting a specific behaviour is the expected average level of adoption for the group:

$$\overline{p}_g = (I_g - 1)^{-1} \sum_{j \neq g} p_{i,j}(y_{ig} = 1 | \omega_g; \theta) \tag{7}$$

which is based on the sum of every other household's probability of adoption divided by the number of households in the group (excluding the decision making household).
The assumed form of the error term makes the probability of adoption a logistic function based on the deterministic components of utility. This determines the conditional choice probability for the decision making household:

$$p_i^*(y_i = 1 | \omega_g, \theta) = \frac{exp(\beta_0 + \beta_1 X_i + \beta_2 Z_g + \beta_3 p_g^* + \alpha_g)}{1 + exp(\beta_0 + \beta_1 x_i + \beta_2 Z_g + \beta_3 p_g^* + \alpha_g)} \tag{8}$$

The empirical implementation of the model tests household likelihood of a specific sanitation behaviour (e.g. latrine ownership) based on the estimation equation:

$$y_{ig} = \beta_0 + \beta_1 x_{ig} + \beta_2 Z_g + \beta_3 p_g^* + \alpha_g + \varepsilon_i \tag{9}$$

where y_{ig} represents household likelihood of a particular sanitation outcome. Three separate outcomes are modelled: latrine ownership; latrine usage; OD. The inference of interest is that from $\beta_3 p_g^*$ which represents group behaviour p_g^* and the endogenous social effect β_3. The application of this framework to the data allows one to ascertain the effect of own (or other) caste/jati group sanitation behaviour within the neighbourhood on own individual likelihood of the same sanitation behaviour.

Private deterministic components are captured by individual household characteristics variables shown in table 2.1. The group variable capturing sanitation behaviour is captured at either a caste or jati level within the Gram Panchayat (GP).[26] The group level characteristics that capture contextual effects are represented by group

[26] The GP represents administrative unit similar to a village and shall be used interchangeably from with village henceforth.

averages (within GP) of the individual characteristics. Group sanitation behaviour is measured as average of same caste/jati behaviour within the village. Group level unobservable factors such as caste norms, or village level specific characteristics are captured by the inclusion of fixed effects. This allows one to capture group level unobservables operating at the village or caste/jati level that could affect sanitation behaviour. This could be neighbourhood specific traits, or caste/jati specific norms that may inform sanitation behaviour. The inclusion of fixed effects allows one to reveal the true underlying behavioural spillover from others engaging in a particular sanitation behaviour within the group. This enables understanding of the extent to which caste/jati social ties may shape individual household behaviour. This could be through behaviour of the group, or the practices affiliated with belonging to the group identity.

The inclusion of fixed effects also ensures the assumption of no group level unobservables is met. An exhaustive list of group level correlated variables that may affect sanitation behaviour is not feasible. Therefore fixed effects are intended to capture these potential correlated group level characteristics and their resulting correlated social effects.

2.4.3 Data

The data used to test social effects through caste based social networks comes from rural Maharashtra, India. It is a cross section of 120 distinct Gram Panchayats (GPs) which are local council villages, and represents over 40,000 households. Census surveys were taken for GPs with less than 480 households. For villages with more than 480 households, the village was segmented. The segmented villages were divided into neighbourhoods, and a census was conducted of chosen neighbourhoods. This facilitated the reduction of measurement errors because even though census data was not possible for those villages with above 480 households, the collection of census data for the sub-neighbourhoods within the village provides a geographic boundary within which it can be assumed caste based dynamics are strongest. That is to say, caste networks are likely tighter within neighbours than between neighbourhoods in villages. Therefore by capturing at least a census of a subsection of the entire village, one is still able to more accurately capture caste based social networks that exist within the village. In the data set 66 different Jatis and 6 general caste classifications are represented.

Table 2.1: Descriptive Statistics

	Mean	Standard Deviation	Min	Max
Caste groups (proportions)				
Other Backward Class (OBC)	28.93%	18.81	0%	88.44%
Scheduled Castes	25.26%	10.56	2%	59.69%
Scheduled Tribe	5.50%	6.96	0%	38.79%
Denotified Tribe	3.49%	7.90	0%	48.83%
Nomadic Tribe	11.65%	11.44	0%	73.01%
Forward Caste	25.15%	15.76	0%	69.36%
Household Characteristics				
Age household head	47.76	14.46	18	99
Household Size	4.89	2.41	1	45
Female Head	0.09	0.03	0.04	0.18
Kids under 2 years	1.02	0.15	1	6
Latrine Ownership	28.27%	18.50	2.34%	86.75%
Latrine Usage	30.04%	18.45	2.34%	92.27%
Any OD	65.30%	21.23	6.44%	95.91%
GP Characteristics				
Households per GP	393.93	100.08	98	509
No. of Jati groups per GP	17.72	8.93	5	35
No. of Caste groups per GP	4.85	1.87	3	6
Jati size per GP	72.72	63.72	1	348
Caste size per GP	125.22	70.29	1	352

Notes: Sample averages and standard deviation of village, caste/jati and household characteristics. 40,000+ observations consisting of 120 GPs, 6 caste groups and 66 jati groups. Percentages represent total sample proportion of variable.

Table 2.1 highlights summary statistics on individual household characteristics, and group compositions. The average number of households per GP is 393. Average caste and jati group size per GP are 125 and 72 groups respectively. This ranges from one household in the village belonging to a specific caste or jati, to 300+ households in GP belonging to the same caste or jati group. Not all caste or jati groups are present within all villages. There are on average about 5 caste groups per village, and about 18 jati groups per village. This highlights the wide range distribution of caste and jati presence across the data. Some caste have a dominant presence within their village, whilst having no presence in other villages. The average age of household head is 47.76 years, and 9% of all households are headed by a female.

Table 2.2 presents descriptives for sanitation behaviours across caste groups. Latrine ownership and usage rates across all caste groups are under 40% and OD rates across all caste groups exceed 55%. Anova tests were conducted to assess the differences in mean sanitation behaviours amongst the various caste groups. The results indicate statistically significant variations in sanitation behaviours across most pairwise comparisons of caste groups. For latrine ownership and OD, the differences are not statistically significant only between Scheduled Castes (SC) and Scheduled Tribes (ST), as well as between these two groups and Nomadic Tribes (NT). The same pattern of differences in caste averages exists for latrine usage. However, for latrine usage the caste means are also insignificant between Other Backward Class (OBC) and Forward Castes (FC).

Table 2.2: Caste Sanitation behaviour

	Mean (%)	Standard Deviation	Min (%)	Max (%)
Latrine Ownership				
Overall	28.27	45.03	2.27	86.75
Other Backward Class	33.15	47.09	0	100
Scheduled Castes	21.99	41.42	0	86.87
Scheduled Tribe	19.8	39.86	0	100
Denotified Tribe	14.53	35.25	0	100
Nomadic Tribe	22.11	41.5	0	100
Forward Caste	35.56	47.87	0	100
Latrine Usage				
Overall	30.06	45.85	2.27	92.27
Other Backward Class	36.02	48	0	100
Scheduled Castes	22.84	41.98	0	89.72
Scheduled Tribe	21.99	41.43	0	100
Denotified Tribe	15.34	36.05	0	100
Nomadic Tribe	23.25	42.25	0	100
Forward Caste	37.51	48.42	0	100
Any OD				
Overall	65.97	47.38	6.44	95.75
Other Backward Class	58.35	49.3	0	100
Scheduled Castes	71.9	44.95	5.32	100
Scheduled Tribe	74.19	43.77	0	100
Denotified Tribe	80.81	39.39	0	100
Nomadic Tribe	73.52	44.12	0	100
Forward Caste	61.26	48.72	0	100

Notes: Averages, standard deviation, minimum and maximum of sanitation behaviours across caste groups. Latrine ownership is defined by whether household owns a latrine. Latrine usage is defined as whether any household member uses a latrine (own or publically provided elsewhere such as school or work). Any OD is defined as whether any household member practices open defecation.

Table 2.3: Jati sanitation behaviour

Jati	Latrine Ownership rate (%)	Latrine Usage rate (%)	Any OD rate (%)
Banjara	15.12	16.06	78.87
Beldar	25.68	29.51	70.49
Berad	9.09	9.09	90.91
Bhil	23.33	23.33	73.33
Bhoi	18.55	17.45	78.91
Bramhan	69.14	69.14	29.63
Buddhist	26.83	26.83	73.17
Burud	13.64	13.64	59.09
Chambhar (Mochi)	24.64	24.88	69.86
Dhangar	20.57	21.96	75.45
Dhobi	17.22	18.91	72.21
Dhor	38.38	37.37	62.63
Gavandi	39.02	43.24	56.76
Gawade/Gawde/Gavade	11.11	11.11	88.89
Golla/Gollewar/Golka/Goler	14.01	15.38	72.8
Gond	54.84	54.84	45.16
Gondhali	37.04	33.33	66.67
Gosai/Gusai	29.17	29.17	70.83
Gosavi	22.95	27.5	75
Gowaari	16.98	18.37	61.22
Gurav	33.55	32.89	66.45
Holer	0	0	100
Jangam	32.8	33.42	57.95
Kaikadi	26.25	25.32	75.95
Kalal	31.86	31.53	64.86
Kasar (sub-castes -Kachar, Kachari	11.36	13.64	40.91
Katkari	11.76	17.39	78.26
Khatik	21.99	26.95	73.76
Koli	16.7	19	76.54
Komati/Komti	56.63	59.04	40.96
Koshti/Koshta	55.81	53.49	46.51
Kulekadgi/Kullekadgi/Kulakadgi/Kull	54.02	55.17	45.98
Kumbhar or Kumhar	33.33	36.11	62.63
Kunbi ,(Maratha Kunbi)	32.16	36.06	59.13
Kureshi	44.44	46.15	42.31
Lingayat	37.27	40.66	54.1
Lohar	27.81	28.4	64.5
Maana	15.38	16.67	50
Mahar	25.21	25.73	68.38
Mali	36.19	37.14	63.81
Mang (Matang)	18.92	19.92	75.2
Mannerwarlu	27.27	28.65	72.18
Maratha	36.65	37.65	63.18
Marwardi	32.43	32.42	67.12
Momin (Ansari)	38.19	39.82	41.12
Muslim	30.87	33.67	65.83
Nhavi (Salmani, Hajam)	23.99	26.2	62.36
Nirali/Nirhali	48.72	48.72	51.28
Padmashali	38.46	37.18	55.13
Panchal	41.18	35.29	64.71
Pardhi	33.8	33.8	60.56
Pathan	33.33	36.36	63.64
Perakewad/Peraki/Perike/Peraka	16.49	17.02	65.96
Pinjari	17.86	16.07	83.93
Rangari	36.84	31.58	63.16
Shimpi	39.73	41.1	56.16
Sonar	42.45	41.9	53.33
Tamboli	62.5	75	25
Teli	39.39	40.4	56.57
Vanjari	34.26	33.87	65.66
Wadar/Vaddar	8.71	9.76	88.85
Wadhai (Sutar)	16.27	17.66	78.26
Wani/Vani/Wain	34.54	34.77	63.91
Warik	14.29	14.29	85.71
Warthi	25	31.58	63.16
Yelam	48.28	49.93	48.13

Notes: Sanitation behaviour rates across jatis. 66 jatis are represented. Muslim and Buddhist, though not official jati categorisations, were self-reported by a significant number of households as their jati group.

Sanitation behaviours across jatis are displayed in table 2.3. The lowest and highest latrine ownership rates are 0 and 69% respectively. The lowest and highest latrine usage rates are 0 and 75% respectively. Lowest and highest open defecation rates are 25% and 100% across jati groups. Muslim and Buddhist are included in the jati list even though they are not officially jati groups. This is because a significant number of survey respondents (across various GPs) self reported these as their jati groups.[27] Though overall ownership and usage rates are low, and overall OD rates relatively high within the sample, these rates vary by caste group. Table 2.2 shows higher ownership and usage rates for caste groups higher up the caste hierarchy such as Forward and Other Backward Class (OBC). Whilst OD rates are generally lower for these caste groups. Conversely, lower ranked caste groups (Scheduled Castes (SC), Scheduled Tribes (ST), Denotified Tribes (DT) and Nomadic Tribe (NT)) have higher OD rates, and lower average ownership and usage rates.

2.4.4 Estimation

Estimating the effect of own group behaviour

The model estimated is:

$$y_{icg} = \beta_0 + \beta_1 X_{icg} + \beta_2 Z_{cg} + \beta_3 \Theta_{cg} + \beta_4 V_{cg} + \beta_5 K_{cg} + \varepsilon_{icg} \tag{10}$$

where:

index i refers to the individual household, c to the caste/jati and g to the village/GP.

y_{icg} is the sanitation outcome; latrine ownership, latrine usage and Any OD estimated separately.

X_{icg} are the individual household characteristics.

Z_{cg} are the group characteristics for contextual effects – (calculated as the in-group average of individual household characteristics).

[27] The inclusion of these two groups does not notably alter estimation results.

Θ_{cg} is the group behaviour measured as the group average of each sanitation activity.

V_g are the village specific correlated characteristics captured by GP fixed effects.

K_c are the caste/jati specific correlated characteristics captured by caste/jati fixed effects.

ε_{icg} are idiosyncratic error term for each household.

Given the possibility of multiple equilibria in the model, a specific chosen equilibrium is required for estimation of the model. Given the complexity in determining a suitable equilibrium choice selection method, the equilibrium in data assumption as used by Gautam (2018a) to estimate a similar B&D model is adopted. The equilibrium in data is used to fit in the log likelihood function that is maximised. This means that the variable representing group behaviour (Θ_{cg}) intended to reveal endogenous social effects β_3, is proxied by the in-group sanitation behaviour rate in the data. This is assumption is more realistic to the model context. Even with the existence of multiple equilibria, households are assumed to know the actual average expected choice of the group, which is presumed to be observed in the data. To ensure a household's decision does not interfere with representations of group behaviour, a leave-out-one mean is used to measure group behaviour.

Different specifications of equation (10) are estimated to include and exclude fixed effects. The first model specification includes no fixed effects. The second includes GP fixed effects only. This supposes that there may be GP level factors unaccounted for that affect sanitation outcomes at the individual household and GP level. Thus to address the potential of bias in endogenous social effects from these unaccounted for GP level unobservables, GP fixed effects are included instead. The third specification includes caste/jati fixed effects only. This addresses a similar problem that may arise from caste/jati group level unobservables that may confound caste/jati level endogenous social effects. The final specification includes both GP and caste/jati level fixed effects.

Estimating the effect of other group behaviour

Given hierarchies which persist in caste groups, it is of value to test the presence of endogenous social effects derived from the behaviour of other caste groups. That is to see if higher (lower) ranked caste groups' behaviour impact lower (higher) ranked household caste sanitation behaviour. This could be a scenario in which the assumption of strategic complementarity fails in endogenous social effects. It could be the case that an individual household has a utility preference for distancing itself from the average behaviour of other caste groups. Estimation for this is carried out on caste group specific samples of the data.

Additionally, different caste groups may exhibit different social structures and dynamics. Therefore one needs to account for heterogeneity amongst the different caste groups in endogenous social effects. The results using all caste groups only provide an average amongst all castes. This may underestimate how this varies between caste groups; the degree of relevance of social effects for the different caste groups. Therefore, estimation is carried out to:

a. Note the difference in magnitude of endogenous social effects within caste groups.

b. Note the presence of any endogenous social effects from the behaviour of other caste groups.

c. Note any difference in own caste endogenous social effects when other caste endogenous social effects are accounted for.[28]

The model estimated is:

$$y_{icg} = \beta_0 + \beta_1 X_{icg} + \beta_2 Z_{cg} + \beta_3 \Theta_g + \beta_6 \Psi_{j1} + \beta_7 \Psi_{j2} + \varepsilon_{ig} \tag{11}$$

where $\Psi_{j()}$ represents the behaviour of other caste groups.

[28] The following specifications are for general caste groupings only with estimation based on each caste subsample. It is not done for jati. There is no known hierarchy in jati groupings beyond those expressed at the caste level. Additionally, given there are 66 different jatis categorised, with overall group sizes in the data ranging from 8 to 6000+ it is infeasible to reasonably compare the effect of other jati group behaviour on own household behaviour.

This model specification is similar to the previous specification on the effect of own group behaviour, except fixed effects are not included here (given only subsamples are estimated).[29] The model is estimated in two forms. The first includes the behaviour of own group within the subsample. The second with own group behaviour alongside the inclusion of the behaviour of other groups. This allows one to see what changes in own group endogenous social effects may exist when other group behaviour is and is not accounted for. As well as note any influence (and direction) of other caste groups' behaviours on own behaviour.

2.5 Results

2.5.1 Own jati group effects

Tables 2.4-2.6 present marginal effect results on sanitation behaviours; ownership, usage and OD respectively, with jati as the social reference group by which social effects are estimated. Results show that for all three sanitation behaviours, endogenous social effects exist and are positive. In other words, positively correlated behavioural spillovers exist in sanitation conduct at the jati level.

In table 2.4, without any fixed effects (model 1), one sees that increasing own jati group ownership levels by 10% increases an average household's likelihood of owning a latrine by 7%. When GP only fixed effects are included (model 2), the effect remains positive but the magnitude drops to 5.33%. When jati only fixed effects are included the magnitude drops as well when compared to no fixed effects, but remains positive (6.88%) and is slightly higher in magnitude than the specification which only incorporates GP fixed effects. When both GP and jati fixed effects are included in the estimation, endogenous social effects remain positive but the magnitude is at its lowest (4.39%). This pattern remains for both latrine usage and OD behaviours. However, usage and OD have slightly larger marginal endogenous effects on a household than latrine ownership. Endogenous social effects on usage (OD) being 4.75% (5.41%) in the models with jati and GP fixed effects included.

[29] The inclusion of caste fixed effects is redundant given each estimation is on a sample of one caste group. The model does not converge when GP fixed effects are included.

The drop in endogenous effects when fixed effects are included suggests there are neighbourhood and jati level unobservables that impact average household sanitation behaviour. These unobservables, when not captured by fixed effects, bias upwards the endogenous social effect of jati group behaviour. GP fixed effects implies that there are village specific characteristics previously unaccounted for, and jati fixed effects implies that there are jati specific traits, or codes of conduct. Both of these correlate with sanitation behaviour, highlighting the importance of accounting for correlated factors when estimating endogenous social effects.

There is a notable difference in the fall in magnitude of endogenous social effects when village fixed effects are included compared to the inclusion of jati fixed effects. GP fixed effects appear to reduce the magnitude more than jati fixed effects. This suggests that GP level unobservables more markedly impact household sanitation behaviour than jati level unobservables. This could be due to a range of sanitation complementing infrastructure that are present in some GPs and absent in others. Although, this does not exclude the possibility that this may be due to jati specific factors at GP level, (differing across GPs) that impact sanitation behaviour.[30]

Contextual social effects are also present. Most average characteristic estimates are statistically insignificant. However, group level proportion of female headed households remains statistically significant for the most part with sanitation conduct. For usage and ownership, the effect is positive. For OD, the effect is negative. This means the more female headed households within the group, the higher the likelihood for a household belonging to that same jati, to exercise the same sanitation behaviour. This applying to both positive (ownership and usage), and negative (OD) sanitation conduct implies the useful role collectivised female headed households can play in inducing sanitation behaviour change within the villages.

On the other hand, the characteristic of female headed household has opposite effects when observed at an individual level. Having a female headed household suggests one is less likely to own or use a latrine, and more likely to OD, consistently across model specifications. The individual effect of a female headed household

[30] One such factor could be jati fractionalization within the village. Other literature (de Janvry et al. 2022) have shown that jati heterogeneity within villages can impact variation in adoption of new behaviours.

may seem counter intuitive given the intuition that women may desire and seek sanitation access more due to their greater need for it, both in the biological sense (e.g. during pregnancy, menstruation), but also in the social sense (e.g. safety, privacy). This would assume that a female household head would have a higher demand for sanitation access. Therefore one would expect this to show through a positive individual effect with latrine ownership and usage, and negative individual effect with OD. One reason for this result may be due to the confounding effect of income. Female headed households may desire sanitation more, but they may also have less income on average than their male counterpart headed households. This may show as a reduced likelihood of adopting sanitation because it is not high in their household expenditure list.

Other individual factors have an expected relationship with sanitation behaviour. Age of household head is positively correlated with ownership and usage and negatively with OD. Age squared has the opposite effect, though with a very small magnitude. This suggests that up to a point, older household heads tend to care more about sanitation access. It could also be the case that household heads at certain bandwidths (such as marriage age) may be more incentivised to own latrines.

Having more household members increases likelihood of ownership and usage whilst reducing likelihood of OD. The intuition behind this is unclear, especially as household size has a negative impact when looked at on average. A better understanding of this relationship between household size and sanitation adoption, may be better understood if composition of the household is known. Larger households may have less willingness to own a latrine due to pre-allocation of resources elsewhere to account for a larger household. On the other hand, a larger household could be more disposed to own a latrine because it has more income earners.

Table 2.4: Determinants of latrine ownership within own jati

	(1)	(2)	(3)	(4)
Own Jati Ownership Behaviour	0.707***	0.533***	0.688***	0.439***
	(0.0145)	(0.0189)	(0.0171)	(0.0238)
Age household head	0.00851***	0.00852***	0.00842***	0.00821***
	(0.00114)	(0.00113)	(0.00114)	(0.00113)
Age household head squared	-0.0000615***	-0.0000613***	-0.0000611***	-0.0000594***
	(0.0000104)	(0.0000104)	(0.0000105)	(0.0000103)
Household Size	0.00862***	0.00850***	0.00866***	0.00844***
	(0.00101)	(0.00104)	(0.000998)	(0.00101)
Female Head	-0.0547***	-0.0550***	-0.0537***	-0.0524***
	(0.00666)	(0.00667)	(0.00667)	(0.00678)
Kids under 2 years	-0.00533	-0.00845	-0.00480	-0.00868
	(0.0151)	(0.0152)	(0.0151)	(0.0152)
Avg. age household head	0.00452	0.00231	0.00102	-0.00301
	(0.00460)	(0.00427)	(0.00472)	(0.00414)
Avg. age household head squared	-0.0000512	-0.0000178	-0.0000220	0.0000175
	(0.0000462)	(0.0000420)	(0.0000479)	(0.0000410)
Avg. Household Size	-0.00359	-0.00343	-0.00209	-0.00328
	(0.00333)	(0.00366)	(0.00381)	(0.00355)
Avg. no. Kids under 2 years	0.0150	-0.0547	0.0220	-0.0652
	(0.0484)	(0.0507)	(0.0542)	(0.0553)
Proportion Female Head	0.0625*	0.0303	0.0854***	0.0736**
	(0.0245)	(0.0281)	(0.0257)	(0.0282)
N	39626	39626	39626	39626
Psuedo R Squared	0.185	-	-	-

Notes: Logit model estimates.Model (1) includes no fixed effects, (2) village level fixed effects only, (3) Jati fixed effects only and (4) GP and Jati fixed effects included. Standard errors in parentheses, and clustered at group level. Significance based on: * p<0.05, ** p<0.01, *** p<0.001

Table 2.5: Determinants of latrine usage within own jati

	(1)	(2)	(3)	(4)
Own Jati Usage Behaviour	0.728***	0.566***	0.706***	0.475***
	(0.0147)	(0.0186)	(0.0174)	(0.0225)
Age household head	0.00639***	0.00641***	0.00634***	0.00616***
	(0.00104)	(0.00104)	(0.00105)	(0.00104)
Age household head squared	-0.0000431***	-0.0000430***	-0.0000430***	-0.0000415***
	(0.00000971)	(0.00000969)	(0.00000976)	(0.00000964)
Household Size	0.00806***	0.00804***	0.00808***	0.00797***
	(0.00109)	(0.00112)	(0.00107)	(0.00109)
Female Head	-0.0550***	-0.0555***	-0.0539***	-0.0528***
	(0.00658)	(0.00656)	(0.00660)	(0.00666)
Kids under 2 years	-0.00179	-0.00517	-0.00149	-0.00596
	(0.0151)	(0.0152)	(0.0152)	(0.0152)
Avg. age household head	0.00623	0.00405	0.00298	-0.00142
	(0.00414)	(0.00412)	(0.00429)	(0.00415)
Avg. age household head squared	-0.0000686	-0.0000366	-0.0000417	0.000000971
	(0.0000420)	(0.0000410)	(0.0000440)	(0.0000413)
Avg. Household Size	-0.00552	-0.00299	-0.00453	-0.00341
	(0.00334)	(0.00361)	(0.00374)	(0.00349)
Avg. no. Kids under 2 years	0.0212	-0.0663	0.0250	-0.0863
	(0.0477)	(0.0520)	(0.0533)	(0.0565)
Proportion Female Head	0.0634**	0.0332	0.0870***	0.0807**
	(0.0239)	(0.0281)	(0.0255)	(0.0287)
N	38469	38469	38469	38469
Psuedo R Squared	0.184	-	-	-

Notes: Logit model estimates.Model (1) includes no fixed effects, (2) village level fixed effects only, (3) Jati fixed effects only and (4) GP and Jati fixed effects included. Standard errors in parentheses, and clustered at group level. Significance based on: * p<0.05, ** p<0.01, *** p<0.001

Table 2.6: Determinants of OD within own jati

	(1)	(2)	(3)	(4)
Own Jati OD Behaviour	0.756***	0.597***	0.743***	0.541***
	(0.0101)	(0.0218)	(0.0116)	(0.0280)
Age household head	-0.00647***	-0.00652***	-0.00642***	-0.00630***
	(0.00111)	(0.00112)	(0.00111)	(0.00111)
Age household head squared	0.0000447***	0.0000449***	0.0000446***	0.0000436***
	(0.0000103)	(0.0000104)	(0.0000103)	(0.0000103)
Household Size	-0.00769***	-0.00790***	-0.00769***	-0.00786***
	(0.00115)	(0.00118)	(0.00113)	(0.00115)
Female Head	0.0529***	0.0538***	0.0519***	0.0514***
	(0.00633)	(0.00636)	(0.00634)	(0.00637)
Kids under 2 years	-0.00624	-0.00213	-0.00681	-0.00180
	(0.0150)	(0.0150)	(0.0150)	(0.0151)
Avg. age household head	-0.00239	-0.00179	0.000663	0.00207
	(0.00431)	(0.00409)	(0.00444)	(0.00405)
Avg. age household head squared	0.0000329	0.0000168	0.00000663	-0.00000696
	(0.0000442)	(0.0000410)	(0.0000460)	(0.0000412)
Avg. Household Size	0.00886**	0.00182	0.00860*	0.00241
	(0.00311)	(0.00344)	(0.00347)	(0.00350)
Avg. no. Kids under 2 years	-0.0147	0.0817	-0.0253	0.0924
	(0.0489)	(0.0565)	(0.0538)	(0.0637)
Proportion Female Head	-0.0732**	-0.0369	-0.0942***	-0.0797**
	(0.0238)	(0.0287)	(0.0253)	(0.0289)
N	38475	38475	38475	38475
Psuedo R Squared	0.217	-	-	-

Notes: Logit model estimates.Model (1) includes no fixed effects, (2) village level fixed effects only, (3) Jati fixed effects only and (4) GP and Jati fixed effects included. Standard errors in parentheses, and clustered at group level. Significance based on: * p<0.05, ** p<0.01, *** p<0.001

2.5.2 Own caste group effects

Tables 2.7-2.9 present marginal effects results when caste is the social reference group by which household social networks are defined. Effects are slighter higher in magnitude but similar to the effects using jati as the social reference group.

Table 2.7 shows that complementary social effects exist in latrine adoption for caste groups. In model 1, it shows that a 10% increase in the proportion of one's caste group owning latrines, results in an approximately 8% increase in the average household's likelihood of latrine ownership. Model 2 includes village fixed effects only, model 3 caste fixed effects only, and model 4 includes both caste and village fixed effects. Similarly to that seen with jati groups, endogenous social effects remain positive, but fall in magnitude when fixed effects are included. Similarly with the jati estimations, the magnitude of endogenous effects is highest without any FE, and lowest when both caste and village FE are included. This implies that there are unobserved village specific factors and caste specific characteristics which influence household latrine adoption. When these factors are not accounted for, there is a positive bias in endogenous social effects. The same pattern of results is observed in regards to social effects on latrine usage and any OD captured in tables 2.8 and 2.9 respectively.

The key difference between the caste based and jati based results is that caste groupings show endogenous social effects of slightly higher magnitude across all three sanitation behaviours than jati based groupings. This implies that endogenous social effects are stronger at the general caste level, than the jati level. This is unexpected as jati is known to be a tighter social identifier than caste. However, the results indicate that when it comes to sanitation behaviour peer effects in caste are stronger than peer effects in jati. One explanation could be the cause of not controlling for variance in village level jati groups. It has been shown that the structure of jati social networks could impact social effects in technology adoption. de Janvry et al. (2022) using an randomised control trial (RCT) in Odisha, India to test adoption of flood resistant rice varieties, show that technology diffuses quicker in villages with less distinct jati groups. This means the level of jati fractionalization within villages can influence how social dynamics enable adoption of new behaviours such as latrine usage and ownership. As shown in table 2.1, there are 5 to 35 distinct jati groups across villages. In more heterogeneous villages, there could be more social tension which can affect social effects in sanitation behaviour. Thereby not controlling for these within jati-village correlated factors could be creating notable variance in the social effects of jati groups. One cannot not be sure without further investigation into a variety of possibilities behind this difference in social effects.

Regarding contextual social effects, a higher proportion of female household heads belonging to the same caste, has a similar relationship as seen with jati. However, this is not consistent across all specifications, and mostly not statistically significant. Nonetheless, the same inconsistency with female household head at an individual level exists in the caste specific models. Individual factors such as age, and household size remain positive relative to latrine ownership, and usage, whilst negative with OD. Endogenous social effects are highest for OD and lowest for ownership.

Given that endogenous social effects exist at both the jati and caste level for a household's own jati and caste group, it would be useful to know if these endogenous effects also exist across groups. As well as the variety in effects across distinct caste groups. Is a household also impacted by the sanitation behaviour of households belonging to other groups? If so, is this effect weaker or stronger? What does accounting for this mean for own group behavioural spillover effects? To address these questions, own caste group specific samples are tested to note own group endogenous social effects, compared to the behaviour of other caste groups. This is done only for caste groups, as there are too many individual jati groupings to make this kind of estimation feasible.

Table 2.7: Determinants of latrine ownership within own caste

	(1)	(2)	(3)	(4)
Own Caste Ownership Behaviour	0.799***	0.679***	0.788***	0.573***
	(0.00912)	(0.0219)	(0.00944)	(0.0288)
Age household head	0.00831***	0.00828***	0.00823***	0.00811***
	(0.00116)	(0.00117)	(0.00116)	(0.00117)
Age household head squared	-0.0000576***	-0.0000572***	-0.0000570***	-0.0000561***
	(0.0000107)	(0.0000107)	(0.0000107)	(0.0000108)
Household Size	0.00886***	0.00887***	0.00889***	0.00881***
	(0.00102)	(0.00105)	(0.00102)	(0.00104)
Female Head	-0.0549***	-0.0559***	-0.0544***	-0.0541***
	(0.00655)	(0.00655)	(0.00654)	(0.00659)
Kids under 2 years	-0.00642	-0.00769	-0.00563	-0.00759
	(0.0151)	(0.0151)	(0.0151)	(0.0151)
Avg. age household head	0.0144	0.0135	0.0128	0.00640
	(0.00823)	(0.00758)	(0.00786)	(0.00713)
Avg. age household head squared	-0.000159	-0.000133	-0.000153	-0.0000942
	(0.0000844)	(0.0000771)	(0.0000809)	(0.0000732)
Avg. Household Size	-0.00410	-0.00311	-0.00213	-0.00698
	(0.00296)	(0.00608)	(0.00336)	(0.00673)
Avg. no. Kids under 2 years	-0.0144	-0.0927	0.0408	-0.0932
	(0.0518)	(0.0568)	(0.0652)	(0.0661)
Proportion Female Head	0.0436	-0.0390	0.0801*	0.0828
	(0.0346)	(0.0489)	(0.0346)	(0.0518)
N	42180	42180	42180	42180
Psuedo R Squared	0.1772	-	-	-

Notes: Logit model estimates.Model (1) includes no fixed effects, (2) village level fixed effects only, (3) Caste fixed effects only and (4) GP and Caste fixed effects included. Standard errors in parentheses, and clustered at group level. Significance based on: * p<0.05, ** p<0.01, *** p<0.001

Table 2.8: Determinants of latrine usage within own caste

	(1)	(2)	(3)	(4)
Own Caste Usage Behaviour	0.822***	0.692***	0.810***	0.587***
	(0.0107)	(0.0217)	(0.0114)	(0.0277)
Age household head	0.00629***	0.00625***	0.00621***	0.00608***
	(0.00106)	(0.00107)	(0.00106)	(0.00107)
Age household head squared	-0.0000400***	-0.0000395***	-0.0000395***	-0.0000383***
	(0.00000974)	(0.00000985)	(0.00000976)	(0.00000991)
Household Size	0.00839***	0.00842***	0.00841***	0.00837***
	(0.00110)	(0.00113)	(0.00109)	(0.00112)
Female Head	-0.0535***	-0.0548***	-0.0530***	-0.0530***
	(0.00668)	(0.00670)	(0.00669)	(0.00674)
Kids under 2 years	-0.00413	-0.00558	-0.00342	-0.00572
	(0.0153)	(0.0152)	(0.0153)	(0.0152)
Avg. age household head	0.0183**	0.0156*	0.0164**	0.00813
	(0.00592)	(0.00625)	(0.00578)	(0.00669)
Avg. age household head squared	-0.000201**	-0.000153*	-0.000192**	-0.000110
	(0.0000614)	(0.0000639)	(0.0000604)	(0.0000686)
Avg. Household Size	-0.00620*	-0.00258	-0.00467	-0.00587
	(0.00316)	(0.00586)	(0.00347)	(0.00639)
Avg. no. Kids under 2 years	0.0109	-0.0921	0.0625	-0.109
	(0.0453)	(0.0580)	(0.0563)	(0.0616)
Proportion Female Head	0.0406	-0.0405	0.0747*	0.0828
	(0.0350)	(0.0493)	(0.0354)	(0.0536)
N	40946	40946	40946	40946
Psuedo R Squared	0.1711	-	-	-

Notes: Logit model estimates.Model (1) includes no fixed effects, (2) village level fixed effects only, (3) Caste fixed effects only and (4) GP and Caste fixed effects included. Standard errors in parentheses, and clustered at group level. Significance based on: * p<0.05, ** p<0.01, *** p<0.001

Table 2.9: Determinants of OD within own caste

	(1)	(2)	(3)	(4)
Own Caste Any OD behaviour	0.836***	0.722***	0.829***	0.664***
	(0.00822)	(0.0235)	(0.00841)	(0.0304)
Age household head	-0.00639***	-0.00632***	-0.00632***	-0.00619***
	(0.00117)	(0.00118)	(0.00117)	(0.00119)
Age household head squared	0.0000421***	0.0000413***	0.0000417***	0.0000406***
	(0.0000108)	(0.0000109)	(0.0000108)	(0.0000109)
Household Size	-0.00817***	-0.00827***	-0.00818***	-0.00825***
	(0.00117)	(0.00120)	(0.00116)	(0.00118)
Female Head	0.0507***	0.0513***	0.0503***	0.0498***
	(0.00630)	(0.00637)	(0.00630)	(0.00635)
Kids under 2 years	-0.00446	-0.00197	-0.00505	-0.00198
	(0.0161)	(0.0161)	(0.0161)	(0.0161)
Avg. age household head	-0.0141*	-0.0119	-0.0121	-0.00620
	(0.00711)	(0.00646)	(0.00676)	(0.00617)
Avg. age household head squared	0.000161*	0.000120	0.000148*	0.0000897
	(0.0000734)	(0.0000664)	(0.0000702)	(0.0000646)
Avg. Household Size	0.00715*	0.0000528	0.00680*	0.00264
	(0.00306)	(0.00563)	(0.00332)	(0.00620)
Avg. no. Kids under 2 years	0.0232	0.176**	-0.0194	0.185*
	(0.0528)	(0.0670)	(0.0582)	(0.0731)
Proportion Female Head	-0.0624	-0.0221	-0.0919**	-0.117*
	(0.0340)	(0.0458)	(0.0348)	(0.0497)
N	40946	40946	40946	40946
Psuedo R Squared	0.1992	-	-	-

Notes: Logit model estimates.Model (1) includes no fixed effects, (2) village level fixed effects only, (3) Caste fixed effects only and (4) GP and Caste fixed effects included. Standard errors in parentheses, and clustered at group level. Significance based on: * p<0.05, ** p<0.01, *** p<0.001

2.5.3 Effect between groups

Estimation on the effect of other group behaviour on household behaviour is determined by taking the population sample for each caste group and estimating them separately. The classification of caste groups found in the data was developed by the government to determine affirmative action policies. Forward Castes (FC) are considered highest in the hierarchy and so receive no government affirmative action support. This is followed by Other Backward Classes (OBC) who have government assigned quotas in areas such as education or political leadership. The OBC are considered educationally and social disadvantaged, but not to the extent of the Scheduled Castes (SC/Dalits), Scheduled Tribes (ST), Denotified Tribes (DT) and Nomadic Tribes (NT) who are afforded higher reservations than OBC. These groups fall below OBC in the historic hierarchy.

It has not as clear what contemporary hierarchy may exist between Scheduled Castes, Scheduled Tribes, Nomadic Tribes, and Denotified Tribes. Due to this, when looking at any hierarchy related social effects on behaviour, comparison is made only with castes distinctly in another hierarchy. For example, when estimating social effects for the OBC sample, estimates are made for their own group endogenous social effect. Comparisons are then made with the group behaviour of FC, and all other caste groups (comprising SC, ST, DT, and NT). On the other hand whilst estimating endogenous social effects for ST for example, comparison is made only with OBC and FC group behaviours. This formulation also reduces attrition of data when caste sample estimates are made, as not all caste groups are present in every single village, and estimation is only made for those villages where the comparable groups also reside. Additionally, given the significantly smaller sample size of ST, DT and NT, grouping them together with SC into an All Others category helps to increase power when estimating endogenous social effects.

Tables 2.10-2.15 present these results. In table 2.10 (OBC sample), model 1 shows that increasing latrine ownership by 10% in own group, increases likelihood of household adoption by 8.71%. Including other groups' sanitation behaviour reduces these endogenous social effects slightly, though they remain positive, and statistically significant (7.91%). Both other groups (FC and All others) show positive endogenous social effects i.e. complementarity in sanitation behaviour. However, their magnitudes are significantly lower than own group's impact. FC group behaviour has a 0.27% (statistically insignificant) effect of an OBC household owning a latrine. Whilst all other caste groups collectively have a slightly larger 0.8% impact on likelihood of an OBC household owning a latrine. These patterns remains for usage and OD except these two behaviours have slightly larger endogenous social effects.

In table 2.11 for Scheduled Castes, own group endogenous social effects are 6.66% for ownership. There is a minor drop in these effects when FC and OBC behaviours are included. Although the social effect of FC and OBC group behaviours are positive, they are statistically insignificant and of significantly smaller magnitude compared to own group endogenous social effects. For FC this is 0.286% and for OBC this is 0.173%. This pattern is consistent for all sanitation outcomes estimated, and again effects are slightly larger for usage and OD.

For Scheduled Tribes (table 2.12), own group endogenous effects are 6.4% for ownership. These drop notably when FC and OBC group behaviours are accounted for (4.77%). For latrine ownership and usage specifically, FC and OBC behaviour exhibit similar effects. This is in the range of 0.1%-0.2% effect on own behaviour. However for OD, OBCs have a larger (1.24%) effect on household OD behaviour compared to FC (0.5%).

Denotified Tribes (table 2.13) follow a similar pattern of social effect responses to Scheduled Tribes. When other group behaviours are included, the effect of own group behaviour drops substantially. For latrine ownership and usage, the effect of OBC group behaviour are higher than the effect of own group behaviour. OBC social effects are 3.71% (4.72%) for ownership (usage) whilst own group effects are 2.77% (1.97%). This implies that the behaviour of OBC exhibits larger social spillovers on the behaviour of households belonging to DT than own group behaviour. For OD, own group behaviour is positive and statistically significant. It does not drop as much in magnitude when other groups' behaviour is included. The effect of FC behaviour are much smaller, negative and statistically insignificant. The effect of OBC OD behaviour is positive, statistically significant, but lower than own group behaviour.

Table 2.14 highlights results for Nomadic tribes. Endogenous social effects on ownership drops from 6.77% to 5.3% when other groups behaviours are included. Though the effect of other groups' behaviours is positive, only the behaviours' of OBC remain statistically significant. The endogenous social effects for FC are not statistically significant in most cases and notably smaller than the effects of OBC behaviour.

Table 2.15 reveals results for Forward Castes. The effects of own group and other groups' behaviours are positive. Own group behaviour for ownership falls from 8.88% to 7.76% when other groups' behaviour are included. This patter of higher social effects from OBC than all other caste groups persists for usage and OD. The effect of OBC behaviour is notably higher in magnitude and statistically significant than the effect of all other caste groups' behaviour. For ownership, OBC behaviour has a 1.04% effect, whilst the behaviours of all other caste groups, has a 0.0358% effect.

Overall, it appears that households in all caste groups are positively affected by the sanitation behaviour of households in other caste groups. This supports the existence of strategic complementarity in sanitation adoption both within and between caste groups. For the most part, own group endogenous social effects are stronger than the endogenous social effects from other groups' behaviours. Highest other group endogenous effects are experienced by Scheduled Tribes and Denotified Tribes. Of all caste groups own group social effects are highest for Forward Castes and lowest for Denotified Tribes. Additionally, OBC appears to be the group that exhibits the highest endogenous social effects on other groups. This implies that apart from own group complementarities in sanitation adoption, OBC can be a useful target group for social multipliers in sanitation adoption for other caste groups.

Table 2.10: Sanitation behaviours for Other Backward Class

	Ownership		Usage		Any OD	
	(1)	(2)	(3)	(4)	(5)	(6)
Own Caste Behaviour	0.871***	0.791***	0.898***	0.803***	0.905***	0.862***
	(0.0172)	(0.0350)	(0.0171)	(0.0322)	(0.0144)	(0.0284)
FC group behaviour	-	0.0269	-	0.0249	-	0.0217
		(0.0227)		(0.0232)		(0.0232)
All other groups' behaviour	-	0.0840*	-	0.128**	-	0.0322
		(0.0373)		(0.0392)		(0.0258)
Age household head	0.00510**	0.00509**	0.00355	0.00348	-0.00364*	-0.00366*
	(0.00178)	(0.00178)	(0.00184)	(0.00184)	(0.00184)	(0.00184)
Age household head squared	-0.0000259	-0.0000258	-0.0000128	-0.0000122	0.0000169	0.0000171
	(0.0000172)	(0.0000172)	(0.0000178)	(0.0000178)	(0.0000179)	(0.0000179)
Household Size	0.0108***	0.0108***	0.00962***	0.00961***	-0.00911***	-0.00918***
	(0.00163)	(0.00163)	(0.00171)	(0.00171)	(0.00177)	(0.00177)
Female Head	-0.0785***	-0.0788***	-0.0707***	-0.0714***	0.0606***	0.0608***
	(0.0132)	(0.0131)	(0.0140)	(0.0140)	(0.0145)	(0.0145)
Kids under 2 years	-0.0101	-0.0115	-0.0123	-0.0140	-0.0104	-0.00966
	(0.0283)	(0.0283)	(0.0297)	(0.0298)	(0.0308)	(0.0308)
Avg. age household head	0.0173	0.0153	0.0148	0.0112	-0.0144	-0.0145
	(0.0220)	(0.0219)	(0.0230)	(0.0228)	(0.0232)	(0.0235)
Avg. age household head squared	-0.000199	-0.000181	-0.000177	-0.000145	0.000173	0.000173
	(0.000228)	(0.000227)	(0.000239)	(0.000237)	(0.000241)	(0.000244)
Avg. Household Size	-0.0130	-0.0133	-0.0140	-0.0164*	0.0114	0.00508
	(0.00715)	(0.00718)	(0.00758)	(0.00763)	(0.00772)	(0.00855)
Avg. no. Kids under 2 years	-0.0927	-0.241	-0.0522	-0.227	-0.00919	0.0584
	(0.254)	(0.262)	(0.265)	(0.271)	(0.271)	(0.278)
Proportion Female Head	0.0926	0.0621	0.0710	0.0199	-0.0146	0.00481
	(0.0990)	(0.0997)	(0.103)	(0.104)	(0.105)	(0.106)
N	12130	12130	11674	11674	11674	11674
Psuedo R Squared	0.1507	0.1511	0.1474	0.1483	0.1677	0.1679

Notes: Logit model estimate of OBC subsample.Models (1),(3) and (5) include own group behaviour only. Models (2), (4) and (6) include other groups' behaviour. Standard errors in parentheses. Significance based on: * p<0.05, ** p<0.01, *** p<0.001

Table 2.11: Sanitation behaviours for Scheduled Castes

	Ownership		Usage		Any OD	
	(1)	(2)	(3)	(4)	(5)	(6)
Own Caste Behaviour	0.666***	0.633***	0.689***	0.651***	0.740***	0.694***
	(0.0127)	(0.0219)	(0.0137)	(0.0223)	(0.00957)	(0.0208)
FC group behaviour	-	0.0286	-	0.0312	-	0.0149
		(0.0222)		(0.0228)		(0.0224)
OBC group behaviour	-	0.0173	-	0.0255	-	0.0445
		(0.0275)		(0.0282)		(0.0259)
Age household head	0.0109***	0.0109***	0.00914***	0.00914***	-0.00885***	-0.00892***
	(0.00170)	(0.00170)	(0.00176)	(0.00175)	(0.00176)	(0.00176)
Age household head squared	-0.0000874***	-0.0000876***	-0.0000729***	-0.0000729***	0.0000693***	0.0000700***
	(0.0000169)	(0.0000169)	(0.0000174)	(0.0000174)	(0.0000175)	(0.0000175)
Household Size	0.00418**	0.00414**	0.00388*	0.00378*	-0.00295	-0.00292
	(0.00158)	(0.00158)	(0.00164)	(0.00163)	(0.00170)	(0.00170)
Female Head	-0.0272*	-0.0269*	-0.0296*	-0.0293*	0.0315**	0.0311**
	(0.0111)	(0.0111)	(0.0115)	(0.0115)	(0.0119)	(0.0119)
Kids under 2 years	-0.00813	-0.00792	-0.00938	-0.00877	0.0107	0.0102
	(0.0262)	(0.0262)	(0.0275)	(0.0274)	(0.0291)	(0.0290)
Avg. age household head	-0.0151	-0.0185	-0.00294	-0.00663	0.00868	0.0100
	(0.0139)	(0.0139)	(0.0147)	(0.0147)	(0.0152)	(0.0152)
Avg. age household head squared	0.000165	0.000205	0.0000349	0.0000766	-0.0000917	-0.000105
	(0.000154)	(0.000155)	(0.000163)	(0.000163)	(0.000169)	(0.000169)
Avg. Household Size	0.0115	0.0104	0.00708	0.00385	0.00613	0.00271
	(0.00731)	(0.00743)	(0.00806)	(0.00831)	(0.00736)	(0.00770)
Avg. no. Kids under 2 years	0.171	0.181	0.193	0.217	-0.178	-0.195
	(0.209)	(0.209)	(0.215)	(0.216)	(0.224)	(0.224)
Proportion Female Head	0.00687	0.0372	0.0252	0.0699	-0.0517	-0.0984
	(0.0732)	(0.0756)	(0.0750)	(0.0782)	(0.0772)	(0.0794)
N	10591	10591	10233	10233	10233	10233
Psuedo R Squared	0.2198	0.2202	0.2062	0.2067	0.2512	0.2517

Notes: Logit model estimate of SCs subsample. Models (1),(3) and (5) include own group behaviour only. Models (2), (4) and (6) include other groups' behaviour. Standard errors in parentheses. Significance based on: * p<0.05, ** p<0.01, *** p<0.001

Table 2.12: Sanitation behaviours for Scheduled Tribes

	Ownership		Usage		Any OD	
	(1)	(2)	(3)	(4)	(5)	(6)
Own Caste Behaviour	0.641***	0.477***	0.684***	0.488***	0.726***	0.623***
	(0.0381)	(0.0592)	(0.0408)	(0.0636)	(0.0318)	(0.0467)
FC group behaviour	-	0.135*	-	0.165*	-	0.0547
		(0.0593)		(0.0649)		(0.0523)
OBC group behaviour	-	0.139*	-	0.190**	-	0.124*
		(0.0624)		(0.0683)		(0.0589)
Age household head	0.00274	0.00316	-0.000350	-0.0000743	-0.000258	-0.000504
	(0.00374)	(0.00374)	(0.00392)	(0.00392)	(0.00403)	(0.00403)
Age household head squared	-0.00000280	-0.00000719	0.0000233	0.0000200	-0.0000197	-0.0000172
	(0.0000370)	(0.0000370)	(0.0000390)	(0.0000390)	(0.0000403)	(0.0000402)
Household Size	0.0101**	0.00983**	0.00916**	0.00875*	-0.00778*	-0.00796*
	(0.00329)	(0.00329)	(0.00354)	(0.00354)	(0.00372)	(0.00370)
Female Head	-0.0363	-0.0367	-0.0411	-0.0416	0.0542*	0.0543*
	(0.0234)	(0.0232)	(0.0250)	(0.0248)	(0.0256)	(0.0255)
Kids under 2 years	0.0230	0.0227	0.0538	0.0544	-0.0590	-0.0596
	(0.0451)	(0.0451)	(0.0473)	(0.0474)	(0.0500)	(0.0500)
Avg. age household head	0.0123	0.0225	0.0176	0.0300	-0.0113	-0.0128
	(0.0263)	(0.0249)	(0.0280)	(0.0272)	(0.0291)	(0.0288)
Avg. age household head squared	-0.000136	-0.000235	-0.000189	-0.000315	0.000121	0.000140
	(0.000276)	(0.000260)	(0.000294)	(0.000284)	(0.000305)	(0.000302)
Avg. Household Size	-0.00841	-0.0191	-0.00893	-0.0210	0.00868	0.000126
	(0.0129)	(0.0132)	(0.0140)	(0.0143)	(0.0145)	(0.0147)
Avg. no. Kids under 2 years	0.0411	0.0343	0.0640	0.0522	0.102	0.103
	(0.236)	(0.238)	(0.250)	(0.254)	(0.254)	(0.253)
Proportion Female Head	0.0851	0.0174	0.0839	0.00741	-0.136	-0.103
	(0.114)	(0.115)	(0.123)	(0.123)	(0.129)	(0.130)
N	2296	2296	2231	2231	2231	2231
Psuedo R Squared	0.1443	0.1505	0.136	0.1433	0.1793	0.1829

Notes: Logit model estimate of ST subsample. Models (1),(3) and (5) include own group behaviour only. Models (2), (4) and (6) include other groups' behaviour. Standard errors in parentheses. Significance based on: * $p<0.05$, ** $p<0.01$, *** $p<0.001$

Table 2.13: Sanitation behaviours for Denotified Tribes

	Ownership		Usage		Any OD	
	(1)	(2)	(3)	(4)	(5)	(6)
Own Caste Behaviour	0.490***	0.277***	0.503***	0.197**	0.598***	0.410***
	(0.0383)	(0.0579)	(0.0405)	(0.0688)	(0.0269)	(0.0636)
FC group behaviour	-	-0.108	-	-0.0701	-	-0.0305
		(0.0942)		(0.0916)		(0.0863)
OBC group behaviour	-	0.371**	-	0.472***	-	0.234*
		(0.114)		(0.126)		(0.103)
Age household head	0.00793*	0.00729	0.00745	0.00670	-0.00434	-0.00355
	(0.00403)	(0.00399)	(0.00421)	(0.00416)	(0.00417)	(0.00415)
Age household head squared	-0.0000715	-0.0000634	-0.0000690	-0.0000590	0.0000344	0.0000249
	(0.0000411)	(0.0000406)	(0.0000431)	(0.0000426)	(0.0000430)	(0.0000428)
Household Size	-0.000880	-0.00163	-0.000371	-0.00151	-0.000698	-0.0000568
	(0.00334)	(0.00332)	(0.00363)	(0.00360)	(0.00381)	(0.00379)
Female Head	0.0847*	0.0783*	0.105*	0.0902*	-0.0639	-0.0596
	(0.0380)	(0.0358)	(0.0414)	(0.0383)	(0.0389)	(0.0377)
Kids under 2 years	0.0153	0.00741	0.0192	0.00645	-0.0179	-0.0117
	(0.0266)	(0.0262)	(0.0280)	(0.0284)	(0.0303)	(0.0304)
Avg. age household head	0.00444	-0.00423	0.00556	-0.00247	-0.00159	0.00286
	(0.0143)	(0.00947)	(0.0109)	(0.00940)	(0.0136)	(0.0109)
Avg. age household head squared	-0.0000487	0.0000770	-0.0000616	0.0000623	0.0000274	-0.0000420
	(0.000156)	(0.000108)	(0.000124)	(0.000108)	(0.000153)	(0.000124)
Avg. Household Size	0.0201*	-0.000577	0.0182	-0.00719	-0.00342	0.0106
	(0.00971)	(0.0109)	(0.0106)	(0.0114)	(0.0101)	(0.0104)
Avg. no. Kids under 2 years	0.180	-0.0959	0.182	-0.267	-0.0574	0.177
	(0.134)	(0.163)	(0.135)	(0.181)	(0.139)	(0.159)
Proportion Female Head	0.0157	0.0372	0.00174	-0.00999	-0.111	-0.0853
	(0.0975)	(0.0970)	(0.103)	(0.103)	(0.114)	(0.116)
N	1446	1446	1414	1414	1414	1414
Psuedo R Squared	0.2595	0.2768	0.2257	0.2486	0.2897	0.297

Notes: Logit model estimate of DT subsample.Models (1),(3) and (5) include own group behaviour only. Models (2), (4) and (6) include other groups' behaviour. Standard errors in parentheses. Significance based on: * p<0.05, ** p<0.01, *** p<0.001

Table 2.14: Sanitation behaviours for Nomadic Tribes

	Ownership		Usage		Any OD	
	(1)	(2)	(3)	(4)	(5)	(6)
Own Caste Behaviour	0.677***	0.530***	0.700***	0.560***	0.712***	0.626***
	(0.0246)	(0.0444)	(0.0249)	(0.0433)	(0.0178)	(0.0300)
FC group behaviour	-	0.0611*	-	0.0565	-	0.0526
		(0.0304)		(0.0315)		(0.0294)
OBC group behaviour	-	0.130**	-	0.133**	-	0.0817*
		(0.0402)		(0.0405)		(0.0336)
Age household head	0.00297	0.00291	0.000991	0.000806	-0.00201	-0.00193
	(0.00256)	(0.00255)	(0.00258)	(0.00257)	(0.00257)	(0.00256)
Age household head squared	-0.0000150	-0.0000148	0.00000361	0.00000496	0.00000832	0.00000791
	(0.0000247)	(0.0000246)	(0.0000249)	(0.0000248)	(0.0000249)	(0.0000249)
Household Size	0.0124***	0.0123***	0.0120***	0.0119***	-0.0118***	-0.0121***
	(0.00235)	(0.00234)	(0.00242)	(0.00241)	(0.00244)	(0.00244)
Female Head	-0.0350	-0.0334	-0.0355	-0.0338	0.0304	0.0296
	(0.0192)	(0.0192)	(0.0197)	(0.0197)	(0.0200)	(0.0200)
Kids under 2 years	-0.0329	-0.0330	-0.0345	-0.0350	0.0192	0.0202
	(0.0339)	(0.0341)	(0.0347)	(0.0350)	(0.0339)	(0.0341)
Avg. age household head	0.0200	0.0234	0.0197	0.0149	-0.0218	-0.0215
	(0.0199)	(0.0196)	(0.0177)	(0.0171)	(0.0189)	(0.0186)
Avg. age household head squared	-0.000220	-0.000265	-0.000224	-0.000183	0.000254	0.000260
	(0.000211)	(0.000209)	(0.000191)	(0.000185)	(0.000204)	(0.000200)
Avg. Household Size	0.00344	0.000739	-0.000196	-0.00235	-0.00148	-0.0116
	(0.00966)	(0.00961)	(0.00948)	(0.00941)	(0.00932)	(0.00950)
Avg. no. Kids under 2 years	0.106	0.0685	0.137	0.0830	-0.0859	-0.0407
	(0.235)	(0.237)	(0.240)	(0.242)	(0.241)	(0.242)
Proportion Female Head	0.119	0.161	0.107	0.154	-0.175	-0.187
	(0.110)	(0.109)	(0.114)	(0.112)	(0.116)	(0.114)
N	4831	4831	4803	4803	4803	4803
Psuedo R Squared	0.1503	0.1535	0.146	0.149	0.2014	0.2037

Notes: Logit model estimate of NT subsample. Models (1),(3) and (5) include own group behaviour only. Models (2), (4) and (6) include other groups' behaviour. Standard errors in parentheses. Significance based on:* p<0.05, ** p<0.01, *** p<0.001

Table 2.15: Sanitation behaviours for Forward Caste

	Ownership		Usage		Any OD	
	(1)	(2)	(3)	(4)	(5)	(6)
Own Caste Behaviour	0.888***	0.776***	0.901***	0.778***	0.901***	0.810***
	(0.0211)	(0.0389)	(0.0210)	(0.0387)	(0.0191)	(0.0357)
OBC group behaviour	-	0.104**	-	0.116**	-	0.0827*
		(0.0389)		(0.0390)		(0.0340)
All other groups' behaviour	-	0.0358	-	0.0420	-	0.0181
		(0.0352)		(0.0369)		(0.0316)
Age household head	0.0128***	0.0126***	0.0101***	0.00983***	-0.0101***	-0.00997***
	(0.00201)	(0.00201)	(0.00205)	(0.00205)	(0.00204)	(0.00204)
Age household head squared	-0.0000930***	-0.0000908***	-0.0000689***	-0.0000663***	0.0000707***	0.0000691***
	(0.0000193)	(0.0000193)	(0.0000197)	(0.0000197)	(0.0000197)	(0.0000197)
Household Size	0.0110***	0.0110***	0.0111***	0.0110***	-0.0115***	-0.0116***
	(0.00179)	(0.00179)	(0.00185)	(0.00185)	(0.00184)	(0.00184)
Female Head	-0.0864***	-0.0857***	-0.0869***	-0.0863***	0.0808***	0.0802***
	(0.0150)	(0.0150)	(0.0156)	(0.0156)	(0.0157)	(0.0157)
Kids under 2 years	-0.00178	-0.000885	0.000220	0.00148	-0.00409	-0.00350
	(0.0252)	(0.0251)	(0.0258)	(0.0258)	(0.0257)	(0.0256)
Avg. age household head	-0.00130	-0.00225	-0.00540	-0.00343	0.00600	0.00846
	(0.0258)	(0.0258)	(0.0267)	(0.0269)	(0.0265)	(0.0264)
Avg. age household head squared	-0.0000160	-0.0000142	0.0000260	-0.00000391	-0.0000323	-0.0000487
	(0.000270)	(0.000271)	(0.000280)	(0.000282)	(0.000278)	(0.000277)
Avg. Household Size	-0.00482	-0.00940	-0.0101	-0.0208	0.0164	0.0142
	(0.0101)	(0.0102)	(0.0110)	(0.0114)	(0.0110)	(0.0111)
Avg. no. Kids under 2 years	-0.178	-0.0872	-0.133	-0.0117	0.0975	0.130
	(0.245)	(0.245)	(0.252)	(0.253)	(0.251)	(0.250)
Proportion Female Head	0.135	0.183	0.106	0.137	-0.116	-0.166
	(0.109)	(0.110)	(0.114)	(0.115)	(0.114)	(0.117)
N	10593	10593	10298	10298	10298	10298
R Squared	0.1332	0.1341	0.1241	0.1252	0.1366	0.1373

Notes: Logit model estimate of FC subsample. Models (1),(3) and (5) include own group behaviour only. Models (2), (4) and (6) include other groups' behaviour. Standard errors in parentheses. Significance based on: * p<0.05, ** p<0.01, *** p<0.001

2.5.4 Potential Mechanisms

Although this chapter has provided empirical evidence of the existence of social effects in sanitation behaviour within jati and caste groups, as well as between different caste groups, it does not offer a definitive analysis of the underlying mechanisms driving these endogenous social effects. Within caste and jati groups, it is unlikely that only one mechanism is responsible for these effects. It is plausible that multiple mechanisms are at play, including information sharing within groups, such as the dissemination of knowledge regarding the actual costs of building latrines. Additionally, social learning may occur regarding the benefits of latrine usage and the disadvantages of OD. Furthermore, the effects may also be influenced by the emergence of within-group norms, potentially driven by observations of fellow caste or jati members constructing and using latrines, thereby fostering a desire to conform to similar behaviours. This desire to conform may also be driven by peer pressure. Observing others in the same caste and jati adopting improved sanitation practices may create social pressure to do the same in order to maintain group cohesion and avoid being socially ostracised.

On the other hand, the mechanisms behind the positive social effects in sanitation behaviour between different caste groups are likely to differ from those within caste groups. For instance, status signalling could be a significant mechanism originating from upper caste groups, as they may engage in behaviours such as latrine construction to subtly signal their social standing. Similarly, the desire to imitate or avoid being surpassed by other caste groups could also influence sanitation behaviour. It is important to note that these mechanisms may also operate within caste groups, although the emphasis may differ. Another potential mechanism feeding positive social effects between caste groups may be coming from intergroup interactions. Agents may observe improved sanitation practices from another group and may be inspired to adopt similar practices because they seem beneficial.

Further research is needed to explore these mechanisms in greater detail and to identify other potential factors driving social effects in sanitation behaviour. By understanding the specific mechanisms at work, policymakers and practitioners can design targeted interventions to promote improved sanitation practices within and between caste groups.

2.5.5 Accounting for income

Notwithstanding the influence of social effects, income is still limiting factor in sanitation adoption. The full data sample did not include income factors and so estimates are limited in interpretation when income is not accounted for.

As part of the same project which collected the listing data used for this chapter, a detailed household survey was carried on a subset of households which includes asset and income related variables. Therefore for robustness, the caste-average income made available through this subsample was used to control for wealth status of households within their caste-village groupings. The variables chosen were average caste-village income, and an index of average benefits received. The index accounts for whether a household receives one or multiple government allocated benefits such as a rural employment card, or low income health insurance card. A furniture index is also included which captures the number of basic furnishings: chairs, tables and beds owned by a household. Table 2.16 provides summary statistics of these asset variables. Caste-village averages of these variables were calculated to be used on the full data sample.

Results are provided in the appendix tables A2.1-A2.3 for OBC, All Other Castes, and FC respectively. Own group endogenous social effects remain positive and of similar magnitude when these wealth controls are included. Further, the social effects of other groups' behaviours remain positive but not as high as own group endogenous effects.

The effects of these wealth factors on sanitation adoption are generally of the expected direction, but largely statistically insignificant and of very minimal magnitude. This is not much different when other wealth related controls such as transport, assets, or household properties are used instead. The small effects of income and other household assets is unexpected. However this may be due to measurement error based on the calculation of income variables. The income and asset variables are based on a subsample of 2000-3000+ households which were used to create the caste-village averages for the full sample. This may not fully capture range of household specific incomes, which may be wide ranging within caste.

Table 2.16: Income and assets summary statistics

Variable	Observations	Mean	Standard Deviation	Min	Max
Yearly Income	3395	68560.11	226091.90	4	7500000
Benefits cards	3589	0.71	0.83	0	3
Furniture	2083	1.90	0.84	1	3

Notes: Summary statistics of asset variables from subsample household surveys. Yearly income is measure in Indian Rupees. Benefits card is the number of different government allocated benefits card households have. Furniture is the number of different types of furniture households have: chair, tables, and beds.

2.6 Conclusion

This chapter estimates endogenous social effects using caste and jati based social groups. For both caste and jati groups, there is a positive relationship between group mean behaviour and individual behaviour in sanitation adoption. This is of a higher magnitude for general caste groups than jati groups. Endogenous social effects remain positive, but fall in magnitude when fixed effects are included at the village and caste/jati level. This highlights the importance of accounting for group level unobservables in determining social effects on sanitation behaviour.

Caste appears to have marginally higher social effects than jati. This may be because social effects in caste are in fact higher, or the results cover heterogeneity on how jati social effects impact behaviour.

Comparison of endogenous social effects within and between caste groups is also estimated. Across all subsampled groups, and for all three sanitation outcomes, own group endogenous effects are positive and mostly statistically significant. These effects drop in magnitude when other groups' behaviours are included. This implies that part of the positive endogenous social effect observed in how own group behaviour affects the household is confounded by how other group behaviour affects the household's sanitation decision. For all subsamples, complementary endogenous social effects exist for other groups' sanitation behaviours but is not always statistically significant, and magnitudes vary group to group.

Hierarchy appears to play the highest role when samples are compared to OBC, which reveals higher magnitude of endogenous social effects on other groups, than FC or All Others. This may come from the middle tier OBC holds between caste groups, which may make them perhaps socially more visible to both FC and All Others, than FC and All Others are to each other.

It may be the case that other castes are positively affected by each other due to reputation concerns. Sanitation adoption is a conspicuous activity and caste groups may care enough about what other caste groups think, and how they are represented, sufficiently enough to create social effects between caste groups in sanitation outcomes.

The policy implications of the results show that caste can be used to facilitate sanitation adoption. To take most advantage of the social multiplier from endogenous social effects, sanitation adoption should be targeted directly towards caste groups. These spillovers which are strongest on the most part from own group behaviour, suggest that spreading sanitation action and support through more cohesive groupings such as caste, could best inspire uptake by all others within the same groups.

The findings of this chapter demonstrate that government subsidies targeting lower-ranked castes for latrine construction have positive spillover effects on other caste groups. This study also reveals that endogenous social effects are more substantial within higher-ranked caste groups, particularly Forward Castes (FC) and Other Backward Classes (OBC), and stronger between caste groups when OBC is the other group. Therefore, complementary sanitation policies aimed at these caste groups would ensure that the spillover effect of sanitation adoption is more extensive. It is suggested that policies should target each caste group separately, particularly in areas with strong caste-based segregation, with complementary interventions tailored to the circumstances and norms of each caste group. Possible measures could include subsidies for lower-ranked castes and information programs for higher-ranked castes that educate them about latrine maintenance and dispel the need to practice ritual impurity in ways that impede their ability to access suitable sanitation.

2.7 References

Arora, S., and Sanditov, B. (2015). Cultures of Caste and Rural Development in the Social Network of a South Indian Village. SAGE Open, 5(3).

Augsburg, B., and Rodríguez-Lesmes, P., (2018), Sanitation and child health in India, World Development, 107, issue C, p. 22-39.

Bathran, R (2011): "Scavenging Caste: Between Classand Caste?," Economic and Political Weekly, Vol 26,No 13, 26 March

Brikké, F. and Bredero, M., (2003) Linking technology choice with operation and maintenance in the context of community water supply and sanitation: a reference document for planners and project staff. Water Supply and Sanitation Collaborative Council, Operation and Maintenance Network, World Health Organization.

Brock, W. and Durlauf, S.N. (2000). Interactions-Based Models. National Bureau of Economic Research Technical Working Paper Series [Online] No. 258.

Brock, W.A. and Durlauf, S.N. (2001). Discrete Choice with Social Interactions. The Review of Economic Studies [Online] 68:235–260.

Brock, W. and Durlauf, S. (2007). Identification of binary choice models with social interactions. Journal of Econometrics [Online] 140:52–75.

Chandrasekhar, S. and Mitra, A. (2019) Migration, caste and livelihood: evidence from Indian city-slums, Urban Research and Practice, 12:2, 156-172

Coffey, D., Gupta, A. , Hathi, P. , Khurana, N. Spears, D. Srivastav, N. and Vyas, S.. (2014). Revealed preference for open defecation: Evidence from a new survey in rural north India. Economic and Political Weekly. 49. 43-55.

Coffey, D. and Gupta, A. and Hathi, Payal and Spears, D. and Srivastav, N and Vyas, S.. (2017). Understanding open defecation in rural India untouchability, pollution, and latrine pits. Economic and Political Weekly. 52. 59-66.

Coffey, D. and Spears, D. - Where India Goes: Abandoned Toilets, Stunted Development and the Costs of Caste. New Delhi: HarperCollins Publishers India, 2017

de Janvry, A., Rao, M., and Sadoulet, E., "Seeding the Seeds: Role of Social Structure in Agricultural Technology Diffusion." Working Paper, August 2022.

Dickinson, K., and Pattanayak, S. (2009). Open sky latrines: Social reinforcing in the case of a (very) impure public good.

Dutta, S., Behera, S.,and Bharti, A. (2015). Access to Drinking Water by Scheduled Castes in Rural India: Some Key Issues and Challenges. Indian Journal of Human Development, 9(1), 115–132.

Exum NG, Gorin EM, Sadhu G, Khanna A, Schwab KJ. Evaluating the declarations of open defecation free status under the Swachh Bharat ('Clean India') Mission: repeated cross-sectional surveys in Rajasthan, India. BMJ Glob Health. 2020 Mar 25;5(3):e002277.

Garn JV, Sclar GD, Freeman MC, Penakalapati G, Alexander KT, Brooks P, Rehfuess EA, Boisson S, Medlicott KO, Clasen TF. The impact of sanitation interventions on latrine coverage and latrine use: A systematic review and meta-analysis. Int J Hyg Environ Health. 2017 Apr;220(2 Pt B):329-340. doi: 10.1016/j.ijheh.2016.10.001. Epub 2016 Oct 11. PMID: 27825597; PMCID: PMC5414716.

Gatade, S.. (2015). Silencing caste, sanitising oppression: Understanding Swachh Bharat Abhiyan. Economic and Political Weekly. 50. 29-35.

Gautam, S. (2018). Household (Under) Adoption of Sanitation: Externalities and Borrowing Constraints.

Gautam, S. (2018). Quantifying Welfare Effects in the presence of Externalities: An Ex-ante Evaluation of a Sanitation Intervention.

Geruso, M., and Spears, D. 2018. "Neighborhood Sanitation and Infant Mortality." American Economic Journal: Applied Economics, 10 (2): 125-62.

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

Gupta, A. , Khalid, N. , Desphande, D. , Hathi, P., Kapur, A., Srivastav, N., Vyas, S., Spears, D., and Coffey, D.. 2019. Changes in Open Defecation in Rural North India: 2014-2018. IZA Discussion Paper.

Jenkins, M. and Cairncross, S.. (2010). Modelling latrine diffusion in Benin: Towards a community typology of demand for improved sanitation in developing countries. Journal of water and health. 8. 166-83. 10.2166/wh.2009.111.

Manski, C.F. (1993). Identification of Endogenous Social Effects: The Reflection Problem. The Review of Economic Studies [Online] 60:531–542.

Munshi, K. (2016), Caste Networks in the Modern Indian Economy, ch. Chapter 2, p. 13-37, Springer

Munshi, K. (2019), "Caste and the Indian Economy." Journal of Economic Literature, 57 (4): 781-834.

O'Reilly, K., Dhanju, R., Goel, A., (2017a), Exploring “The Remote” and “The Rural”: Open Defecation and Latrine Use in Uttarakhand, India, World Development, Volume 93, 2017, Pages 193-205, ISSN 0305-750X.

O'Reilly, K., Dhanju, R. and Louis, E., (2017b), Subjected to Sanitation: Caste Relations and Sanitation Adoption in Rural Tamil Nadu, Journal of Development Studies, 53, issue 11, p. 1915-1928.

Pais, R., (2021) Sanitation and Unchanging Role of Caste. Indian Anthropologist, Vol. 51, No. 2, Special Issue on: LANDSCAPE OF SANITATION IN INDIA: REFLECTIONS ON SWACHHTA (July – December 2021), pp. 57-69 (13 pages)

Routray P, Schmidt WP, Boisson S, Clasen T, Jenkins MW. Socio-cultural and behavioural factors constraining latrine adoption in rural coastal Odisha: an exploratory qualitative study. BMC Public Health. 2015 Sep 10;15:880.

Shakya HB, Christakis NA, Fowler JH. Association between social network communities and health behavior: an observational sociocentric network study of latrine ownership in rural India. Am J Public Health. 2014 May;104(5):930-7. doi: 10.2105/AJPH.2013.301811. Epub 2014 Mar 13. PMID: 24625175; PMCID: PMC3987610.

Shakya HB, Christakis NA, Fowler JH. Social network predictors of latrine ownership. Soc Sci Med. 2015 Jan;125:129-38.

Spears, D. and Thorat, A. (2019). The Puzzle of Open Defecation in Rural India: Evidence from a Novel Measure of Caste Attitudes in a Nationally-Representative Survey. Economic Development and Cultural Change. 67. 10.1086/698852.

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

Vyas, S. and Spears, D., (2018). Sanitation and Religion in South Asia: What Accounts for Differences across Countries?. The Journal of Development Studies. 54. 1-17.

WHO/UNICEF JMP (2012), Progress on drinking water and sanitation, 2012 update.

WHO/UNICEF JMP (2021), Progress on household drinking water, sanitation and hygiene 2000-2020 Five years into the SDGs.

2.8 Appendix B: Results with income controls

Table A2.1: OBC sanitation behaviours with wealth controls

	Ownership		Usage		OD	
	(1)	(2)	(3)	(4)	(5)	(6)
Own Caste Behaviour	0.872***	0.796***	0.908***	0.809***	0.898***	0.869***
	(0.0208)	(0.0399)	(0.0214)	(0.0381)	(0.0168)	(0.0297)
All Others Group Behaviour	-	0.0943*	-	0.139**	-	0.0266
		(0.0437)		(0.0460)		(0.0272)
FC Group Behaviour	-	0.0130	-	0.0165	-	0.0108
		(0.0232)		(0.0237)		(0.0237)
Age household head	0.00526**	0.00529**	0.00413*	0.00415*	-0.00360	-0.00362
	(0.00186)	(0.00186)	(0.00192)	(0.00192)	(0.00192)	(0.00192)
Age household head squared	-0.0000284	-0.0000287	-0.0000183	-0.0000186	0.0000178	0.0000180
	(0.0000179)	(0.0000179)	(0.0000186)	(0.0000185)	(0.0000187)	(0.0000187)
Household Size	0.0104***	0.0104***	0.00950***	0.00947***	-0.00834***	-0.00839***
	(0.00175)	(0.00175)	(0.00183)	(0.00183)	(0.00191)	(0.00191)
Female Head	-0.0756***	-0.0758***	-0.0662***	-0.0665***	0.0598***	0.0599***
	(0.0137)	(0.0137)	(0.0147)	(0.0146)	(0.0152)	(0.0152)
Kids under 2 years	-0.0276	-0.0293	-0.0260	-0.0281	0.00665	0.00735
	(0.0313)	(0.0313)	(0.0329)	(0.0330)	(0.0339)	(0.0339)
Avg. age household head	0.0253	0.0210	0.0251	0.0193	-0.0328	-0.0308
	(0.0250)	(0.0249)	(0.0262)	(0.0259)	(0.0267)	(0.0270)
Avg. age household head squared	-0.000284	-0.000240	-0.000292	-0.000232	0.000369	0.000347
	(0.000260)	(0.000259)	(0.000273)	(0.000270)	(0.000278)	(0.000282)
Avg. Household Size	-0.0120	-0.0138	-0.0121	-0.0164	0.0115	0.00685
	(0.00790)	(0.00799)	(0.00847)	(0.00860)	(0.00868)	(0.00961)
Avg. no. Kids under 2 years	-0.120	-0.301	-0.0141	-0.238	-0.0381	0.0332
	(0.299)	(0.311)	(0.316)	(0.324)	(0.324)	(0.332)
Proportion Female Head	0.0972	0.0753	0.0574	0.0350	-0.0117	0.000534
	(0.110)	(0.111)	(0.116)	(0.116)	(0.118)	(0.119)
Avg Log Caste Income	-0.0000345	-0.000279	0.000152	-0.000154	-0.000114	-0.0000488
	(0.000592)	(0.000602)	(0.000631)	(0.000641)	(0.000656)	(0.000661)
Benefits Index	0.00207	0.00149	0.00347	0.00320	-0.000864	-0.000838
	(0.00428)	(0.00430)	(0.00456)	(0.00458)	(0.00455)	(0.00455)
Furniture Index	0.00131	0.00228	0.000767	0.00177	-0.00287	-0.00298
	(0.00309)	(0.00313)	(0.00327)	(0.00330)	(0.00333)	(0.00333)
N	10889	10889	10455	10455	10455	10455
R squared	0.1563	0.1566	0.1515	0.1522	0.1733	0.1734

Notes: Logit model estimate of Other Backward Caste. Models (1),(3) and (5) include own group behaviour only. Models (2), (4) and (6) include other groups' behaviour. Standard errors in parentheses. Significance based on: * p<0.05, ** p<0.01, *** p<0.001

Table A2.2: All Other Castes sanitation behaviours with wealth controls

	Ownership		Usage		OD	
	(1)	(2)	(3)	(4)	(5)	(6)
Own Caste Behaviour	0.668***	0.612***	0.703***	0.646***	0.748***	0.692***
	(0.0131)	(0.0267)	(0.0142)	(0.0276)	(0.0101)	(0.0225)
OBC Group Behaviour	-	0.0373	-	0.0440	-	0.0474*
		(0.0241)		(0.0254)		(0.0222)
FC Group Behaviour	-	0.0249	-	0.0224	-	0.0161
		(0.0178)		(0.0186)		(0.0179)
Age household head	0.00702***	0.00706***	0.00537***	0.00539***	-0.00576***	-0.00579***
	(0.00135)	(0.00135)	(0.00139)	(0.00139)	(0.00142)	(0.00142)
Age household head squared	-0.0000488***	-0.0000492***	-0.0000349*	-0.0000351*	0.0000388**	0.0000392**
	(0.0000133)	(0.0000133)	(0.0000137)	(0.0000137)	(0.0000140)	(0.0000140)
Household Size	0.00649***	0.00645***	0.00631***	0.00627***	-0.00588***	-0.00589***
	(0.00123)	(0.00123)	(0.00129)	(0.00129)	(0.00136)	(0.00136)
Female Head	-0.0234*	-0.0232*	-0.0242*	-0.0241*	0.0254*	0.0253*
	(0.00934)	(0.00934)	(0.00977)	(0.00977)	(0.0101)	(0.0101)
Kids under 2 years	-0.00187	-0.00162	0.00314	0.00352	-0.00977	-0.0101
	(0.0179)	(0.0179)	(0.0187)	(0.0188)	(0.0196)	(0.0197)
Avg. age household head	-0.0101	-0.0175	0.0107	0.00128	-0.00101	0.00595
	(0.0191)	(0.0195)	(0.0199)	(0.0204)	(0.0212)	(0.0215)
Avg. age household head squared	0.000124	0.000206	-0.000104	-0.00000189	-0.00000107	-0.0000754
	(0.000207)	(0.000212)	(0.000218)	(0.000223)	(0.000231)	(0.000234)
Avg. Household Size	0.0107	0.00970	0.00184	-0.000607	-0.000614	-0.00433
	(0.00637)	(0.00640)	(0.00690)	(0.00698)	(0.00646)	(0.00665)
Avg. no. Kids under 2 years	-0.0868	-0.0708	-0.0471	-0.0266	0.0457	0.0476
	(0.217)	(0.217)	(0.226)	(0.227)	(0.234)	(0.234)
Proportion Female Head	0.128	0.150	0.111	0.137	-0.205*	-0.238**
	(0.0807)	(0.0813)	(0.0842)	(0.0851)	(0.0852)	(0.0863)
Avg Log Caste Income	-0.000893*	-0.000946*	-0.000413	-0.000465	0.000345	0.000392
	(0.000417)	(0.000419)	(0.000431)	(0.000431)	(0.000445)	(0.000445)
Benefits Index	0.00664*	0.00795*	0.00531	0.00675*	-0.00508	-0.00613
	(0.00304)	(0.00309)	(0.00312)	(0.00318)	(0.00319)	(0.00321)
Furniture Index	0.00506*	0.00515*	0.00486*	0.00482*	0.00127	0.00187
	(0.00233)	(0.00233)	(0.00238)	(0.00237)	(0.00242)	(0.00243)
N	16409	16409	15982	15982	15982	15982
R squared	0.1698	0.1702	0.1557	0.1561	0.1987	0.1991

Notes: Logit model estimate of all Scheduled Caste, Scheduled Tribes, Denotifed Tribes, Nomadic Tribes collectively. Models (1),(3) and (5) include own group behaviour only. Models (2), (4) and (6) include other groups' behaviour. Standard errors in parentheses. Significance based on: * $p<0.05$, ** $p<0.01$, *** $p<0.001$

Table A2.3: Forward Caste sanitation behaviours with wealth controls

	Ownership		Usage		OD	
	(1)	(2)	(3)	(4)	(5)	(6)
Own Caste Behaviour	0.883***	0.730***	0.898***	0.744***	0.895***	0.799***
	(0.0225)	(0.0472)	(0.0225)	(0.0455)	(0.0208)	(0.0401)
OBC Group Behaviour	-	0.126**	-	0.136**	-	0.0888*
		(0.0457)		(0.0454)		(0.0389)
All Others Group Behaviour	-	0.0580	-	0.0527	-	0.0150
		(0.0426)		(0.0430)		(0.0365)
Age household head	0.0133***	0.0131***	0.0113***	0.0110***	-0.0102***	-0.0101***
	(0.00215)	(0.00215)	(0.00220)	(0.00220)	(0.00218)	(0.00219)
Age household head squared	-0.0000985***	-0.0000963***	-0.0000800***	-0.0000776***	0.0000721***	0.0000707***
	(0.0000207)	(0.0000206)	(0.0000211)	(0.0000211)	(0.0000210)	(0.0000210)
Household Size	0.0114***	0.0114***	0.0115***	0.0115***	-0.0127***	-0.0128***
	(0.00191)	(0.00190)	(0.00196)	(0.00196)	(0.00196)	(0.00196)
Female Head	-0.0762***	-0.0761***	-0.0768***	-0.0769***	0.0727***	0.0724***
	(0.0162)	(0.0162)	(0.0169)	(0.0169)	(0.0170)	(0.0170)
Kids under 2 years	0.000955	0.00144	0.00239	0.00325	-0.00340	-0.00247
	(0.0264)	(0.0264)	(0.0271)	(0.0271)	(0.0269)	(0.0268)
Avg. age household head	-0.00948	-0.0155	-0.00958	-0.0139	0.0111	0.0162
	(0.0289)	(0.0295)	(0.0298)	(0.0306)	(0.0298)	(0.0300)
Avg. age household head squared	0.0000656	0.000117	0.0000649	0.0000964	-0.0000823	-0.000124
	(0.000303)	(0.000309)	(0.000313)	(0.000321)	(0.000312)	(0.000314)
Avg. Household Size	-0.00832	-0.0134	-0.0126	-0.0253*	0.0183	0.0139
	(0.0110)	(0.0111)	(0.0120)	(0.0123)	(0.0121)	(0.0122)
Avg. no. Kids under 2 years	-0.251	-0.186	-0.193	-0.0864	0.137	0.201
	(0.299)	(0.301)	(0.309)	(0.312)	(0.305)	(0.308)
Proportion Female Head	0.112	0.137	0.109	0.0959	-0.119	-0.164
	(0.144)	(0.145)	(0.149)	(0.150)	(0.150)	(0.152)
Avg Log Caste Income	0.00000942	-0.000172	-0.000130	-0.000329	0.0000570	0.000198
	(0.000806)	(0.000812)	(0.000830)	(0.000834)	(0.000826)	(0.000832)
Benefits Index	0.00102	0.00463	0.00233	0.00580	-0.00244	-0.00352
	(0.00577)	(0.00584)	(0.00589)	(0.00596)	(0.00590)	(0.00592)
Furniture Index	0.00259	0.00607	0.00253	0.00594	-0.00190	-0.00429
	(0.00412)	(0.00423)	(0.00423)	(0.00433)	(0.00423)	(0.00434)
N	9152	9152	8899	8899	8899	8899
R squared	0.1369	0.138	0.1301	0.1314	0.1432	0.1439

Notes: Logit model estimate of Forward Caste. Models (1),(3) and (5) include own group behaviour only. Models (2), (4) and (6) include other groups' behaviour. Standard errors in parentheses. Significance based on: * p<0.05, ** p<0.01, *** p<0.001

3 Neighbourhood spatial networks and sanitation adoption. Does proximity matter?

Abstract

This chapter investigates the impact of endogenous and contextual social effects on three distinct sanitation behaviours, namely: latrine ownership, latrine usage, and open defecation (OD), by employing spatial networks as the benchmark group to define social networks. The study assesses how the magnitude of social effects changes with household network size in the village. The spatial networks are specified using geographic distances of 30 metres, 100 metres, and 300 metres diameters. The findings reveal that the magnitude of social effects varies with both network size and sanitation behaviour, and do not necessarily increase with network size as presumed. The endogenous social effects of OD intensify with network size, whereas the social effects of latrine ownership and usage diminish with network size. The results have implications for understanding how social effects in negative sanitation behaviours contrast with social effects in positive sanitation behaviours.

3.1 Introduction

The provision of adequate sanitation services is a key element of Goal 6 of the Sustainable Development Goals (SDGs), which aims to ensure access to clean water and sanitation for all by 2030. Insufficient access to safe sanitation services and widespread open defecation pose significant global concerns, as they have adverse impacts on social and economic indicators, including increased childhood morbidity, heightened prevalence of waterborne illnesses like cholera, elevated healthcare costs, greater pollution, gender imbalances, and limited human capital development. As of 2020, global rates of sanitation have declined, with the percentage of people with access to safely managed sanitation services increasing from 47% to 54% between 2015 and 2020. However, at this rate of progress, it is unlikely that the goal of universal access to basic sanitation services will be achieved by 2030, unless there is a fourfold increase in the rate of progress (WHO - UN 2021).

It has become a stylised fact within the sanitation context that market-based factors such as price, access or information asymmetry, are not the only barriers to improved sanitation adoption in many developing contexts (Coffey et al. 2014, Routray et al. 2015). Behavioural factors, especially in the form of social norms and preferences for open defecation (OD) is a common practice which also hinders progress in improved sanitation outcomes.

Community led total sanitation (CLTS) campaigns have been a way in which behaviour change in sanitation has tried to been attempted. CLTS has been present in 59 countries (Zuin et al. 2019). It involves the use of community based self-enforcement to end OD and initiate latrine building. Part of the focus of CLTS often involves evoking shame and disgust as incentives to change sanitation behaviours in communities. However, heterogeneity in success of CLTS (Venkataramanan et al. 2018, Orgill-Meyer et al. 2019, Kresch et al. 2020, Abramovsky et al. 2023) suggests gaps on when, where and how CLTS can be most effective. This includes achieving intended outcomes in sanitation behaviour without socially problematic externalities (Bartram et al. 2012), but also ensuring long-term behaviour change.

The role of sanitation social norms is especially stark in the Indian context where data on sanitation status does not align with other measures of socio-economic development as it does in other parts of the world (Drèze and Sen 2013). Part of this is due to cultural beliefs around sanitation (Coffey and Spears 2017). These beliefs are intertwined with the complexities of India's caste system (Coffey and Spears 2017).[31] Consequently, this unique context presents an opportunity to examine the social interaction effects on sanitation behaviour.

India has implemented various policy packages to enhance its sanitation status, with the most recent being the Swachh Bharat Mission (SBM), which aimed to achieve Open Defecation Free (ODF) status by October 2019. The government of India has declared that this goal has been achieved. However, several academic writings have challenged the validity of this assertion (Exum et al. 2019, Gupta et al. 2019, Behera et al. 2021, Caruso et al. 2022). Thus, while sanitation access in India may have improved, OD has not been entirely eradicated, making it a relevant area for investigation in India's socio-economic development.

This chapter employs spatial distance to define social networks and investigates the impact of endogenous and contextual social effects on household sanitation behaviours within rural villages of Maharashtra, India. This study benefits from the use of a large dataset of 29,000+ household GPS locations to understand the impact proximity plays in social effects. Three distinct spatial networks are utilised, each with bandwidths of 30, 100, and 300 meters.

The next section explains the significance of examining spatial networks and sanitation. Section 3.3 explains the methodology employed to gauge social effects impacts, comprising the theoretical framework, identification strategy, equilibrium determination and empirical tractability. Section 3.4 describes the data and provides justification for the selected spatial network boundaries. Section 3.5 discusses estimation results and limitations. Section 3.6 concludes.

[31] Ideas around ritual purity, and the role of lower caste groups feed hesitancy to access and maintain latrine ownership for usage.

3.2 Spatial Networks and Sanitation

In the economics discipline, social interaction effects are often evaluated through Manski's (1993) work, which distinguishes between various types of social effects and the challenges associated with their identification. Manski identifies three types of social effects; endogenous, contextual, and correlated social effects. Endogenous social effects are the behaviour spill-overs resulting from the behaviour of others within an individual's social group. Contextual social effects arise from the impact of the average characteristics of one's social group. Correlated effects encompass common shocks and circumstances that affect the social group as a whole. They can shape outcomes for all members of the group in a specific way. Distinguishing between distinct types of social effects is important as the different social effects offer different policy implications. Identifying the dominant social effects underpinning specific outcomes, can be informative in shaping policy.

Using the framework of social effects, social interactions mediated through neighbourhoods have been measured to understand social effects on different outcomes including education, labour market participation, poverty, adoption of new technology and more (Helmers and Patnam 2014, Hellerstein et al. 2014). This has included sanitation on a few occasions (Shakya et al. 2014, Lunn 2020). However, the impact of a neighbourhood network defined by spatial distance between households on endogenous and contextual social effects in sanitation outcomes has hardly been analysed for the Indian context (Pakhtigian et al. 2022).

The broad neighbourhood (e.g. at the village or sub-village level as a whole) is usually the boundary through which neighbourhood effects are measured for sanitation outcomes (Gautam 2018a, Gautam 2018b, Guiteras et al. 2019). Other attempts using more refined spatial configuration to assess neighbourhood effects measure the spatial networks in other ways. For example, Tukahirwa et al. 2011 show that social and spatial proximity are relevant for sanitation outcomes in informal urban settlements in Uganda. However spatial distance in that paper was relative to offices and other locations in which sanitation services were on offer. The spatial distance between agents in assessing spill-overs of sanitation outcomes was not used. The estimation sample in the paper was also limited to under 350 households. Kennedy-Walker et al. 2015 inputs spatial proximity in relation to sanitation outcomes. However, their paper was focused on Zambia, and spatial proximity there was defined in terms of distance between households and the local institutions such as administrative bodies for amenities such as water provision, health, and community services. The paper was also limited to a sample size of under 200 respondents. Lunn 2020 has shown that familial networks that exist from rural households with those in urban environments can positively influence the likelihood of latrine ownership. However, the paper was limited in its focus on social ties that exist outside the rural environment. These studies all offer uniquely different spatial configurations on assessing social effects in sanitation outcomes. This chapter contributes to the literature on the relationship between space and sanitation outcomes, by focusing on within village social interactions. It builds the spatial networks through pairwise household distances, using over 29,000 household respondents. Pakhtigian et al. 2022 has shown the existence of peer effects using spatial networks to define reference groups based on hamlets. Their work covered 1000 households in 39 villages in Orissa, India over a 14 year period to test the impact of sanitation adoption policy vis a vis peer effects. They find that social effects persist over time and are thus a useful means of catalysing long-term sanitation adoption.

Chakraborty (2020) argues that neglecting the spatial heterogeneity of sanitation drivers may result in incomplete information for practitioners. To address this gap, this chapter examines how peer effects in sanitation behaviour manifest as the socio-spatial network changes. It is unclear whether social effects increase with expansion of the network. Or if they decrease as the network bandwidth increases because proximity has a strong influence on peer effects. This chapter aims to contribute to the literature in this area by analysing the GPS locations of households in rural villages of Maharashtra, India. It examines the effects of neighbourhood networks on sanitation behaviours: latrine ownership, usage, and open defecation. By defining neighbourhood groups at a more granular spatial level than the entire village neighbourhood, the aim is to enhance our understanding of the role of more granular spatial boundaries on sanitation outcomes. Three spatial boundaries (30 metres, 100 metres, and 300 metres) are used to limit the spatial network of households, and the rationale for these boundaries will be discussed further in this chapter. The study focuses solely on contextual and endogenous social effects and does not examine correlated effects.[32]

The research question of interest is whether larger spatial networks exhibit larger endogenous and contextual social effects. It is expected that larger networks would have more significant endogenous social effects due to the increased number of people one interacts with, and greater influence that comes from being part of a larger social reference group. Though this work follows a local average model framework, it is reasonable to anticipate that larger networks, characterised by a greater mass of individuals generates more pressure to conform with the behaviour and outcome of others.

The accuracy of social effects measured relies on the identification of a relevant reference group which influences individuals' behaviour through social interactions. Del Bello et al. (2014) highlight the significance of this reference group in determining the accuracy of the networks involved. However, determination and measurement of this reference group remain challenging, particularly in terms of cost-effectiveness. Geography can be a straightforward way to proxy for social networks. It has also been demonstrated as a relevant factor for social referencing (Topa and Zenou 2015), including the context of sanitation outcomes (Chakraborty et al. 2022).

The structure by which social networks have been defined in the literature vary widely. Proxies such as classroom, or neighbourhood catchment are used when more granular information of the exact networks are unattainable. Despite acquisition of other detailed information, it frequently remains constrained to a singular dimension such as friendship groups, colleagues, or administrative bandwidth of neighbours. This makes it difficult to structure social networks on all possible dimensions that agents use to interact with the rest of the world. This does not negate the value a specific dimension plays on behavioural outcomes. It simply highlights the need for

[32] Correlated effects were attempted through the use of fixed effects and random effects in variations of the model. However, the computation complexity of this led to either spurious results, or non-convergence of the model. For simplicity's sake, a model without correlated effects is presented. However, extension of this work with the inclusion of correlated effects modelled in alternative ways is underway.

caution in relating defined social groups to specific outcomes, and assuming those social references are the only references by which social effects in a particular outcome is mediated. It also speaks to the importance of ensuring a specific social dimension measured is relevant to the outcome studied e.g. educational outcomes and classroom groupings.

This chapter centres on sanitation behaviour of individuals residing in rural neighbourhoods in a global south context. Peer influence can be a significant determinant of sanitation behaviours. For instance, OD is a conspicuous activity that attracts attention within community settings. Perpetuated over long periods of time it develops into a social norm making it susceptible to peer influence. Therefore, change in this behaviour could spill-over to peers in the same neighbourhood. Another behaviour analysed in the context is latrine ownership. If latrine ownership becomes more prevalent in a neighbourhood, it's visibility could be perceived as a signal of status, or offer opportunities for others to learn about the value of this technology. This could incentivise other households in the neighbourhood to adopt the behaviour as well. This highlights how these outcomes naturally have a spatial component in the social intereactions that can influence them; providing justification for distance as a useful means for defining networks of peer effects.

The uniqueness of sanitation in the social and spatial space is also marked by the degree of negative externalities from OD. OD affects the health outcomes of the wider community when it leaks into water sources, agricultural land, and other spaces agents encounter in daily life. These spatial spill-overs in disease transmission, and pollution can be reduced significantly through collective community wide increase in latrine usage, and reduction in OD. This is because of the spatial correlations that exist in disease transmission (Emch 2012); creating spatial externalities of poor or good sanitation behaviour on the health outcomes of others. Concurrently, positive externalities can be attained at the spatial level from increased latrine usage (Cameron et al. 2021, Cameron et al. 2022).

Social structure defined by spatial distance plays an important role for a number of other reasons. The cost of social interaction increases with physical distance. Literature has noted decay in social interactions as distance increases (Hipp and Perrin

2009, Patacchini and Zenou 2012(a), Preciado et al. 2012). Proximity increases the frequency of social interactions by increasing the likelihood of chance encounters and relationship formation. It expresses part of the physical limits of one's day-to-day social interactions. This can especially be true in a rural environment within a developing country due to constraints in communication technologies and transportation. This can enhance the significance of neighbourhoods in social dynamics within rural areas, especially compared to urban environments. Limited access to external resources, smaller communities, and restricted interactions make rural settings typically characterised by a limited number of individuals with whom one can interact, leading to a stronger sense of community. As a result, spatial referencing is useful for measuring the links and partial boundaries of social interactions, making geographic proximity serve as a valuable proxy for estimating social networks in rural environments.

Furthermore, a number of other social ties such as caste, religion, education, income, and social status are social dimensions that can affect choice of location. This enables spatial location to proxy for multiple social dynamics operating at the same time. Proximity can serve as a catalyst for social interactions, and also serve as an indicator of pre-existing social ties. As Neal (2020) highlights: "...everything is related to everything else, but topographically and topologically near things are more related than topographically and topologically distant things".

3.3 Methodology

The theoretical framework used to define and estimate social effects follows the work of Brock and Durlauf (2000, 2007), and Lee et al. 2014. It is akin to an SAR framework (Lin 2010) in which social relationships are defined through a spatial weights matrix (or adjacency matrix). The weights matrix determines the connections between households as neighbours within the village. The equation which defines the social effects to be estimated is:

$$p(y_{ig}) = F(\beta_0 + \beta_1 \mathbf{W_{ig}} y_{ig} + \beta_2 \mathbf{X_{ig}} + \beta_3 \mathbf{W_{ig}} \mathbf{X_{ig}} + \varepsilon_{ig}) \tag{1}$$

Where $p(y_{ig})$ represents the probability of choosing a sanitation outcome, such as to own or not own a latrine. $F(.)$ is a logistic distribution function.

Where:

y_{ig} is a vector of individual household sanitation behaviour of everyone in the village. In which subscript g marks the village.

$\mathbf{W_{ig}}$ is the village spatial weights matrix. This determines the neighbours of households based on the distance cut-off used to define the network.

$\mathbf{X_{ig}}$ is an n by k matrix of individual household characteristics. Each column represents the individual level characteristics that determines households' independent likelihood of a specific sanitation outcome. With inclusion of the adjacency matrix $\mathbf{W_{ig}}$, this manifests as the average characteristics of households' network members, creating contextual characteristics.

β_1, β_2, and β_3 respectively represent endogenous social effects, individual household effects and contextual social effects.

ε_{ig} is a column vector that represents idiosyncratic shocks to each household's decision.

This specification implies that individual and social determinants of choices act as strategic complements. Mechanisms that could generate strategic complements in sanitation outcomes include conformity, social learning, social norms, status signalling, or a mix of these operating together. However, the specific mechanisms at play are not the focus of this chapter. Of note is the acknowledgement that social effects in neighbourhoods tend to be strategic complements. Case and Katz 1991 for example note that the probability of social ills in one neighbourhood is increasing in the prevalence of the same ills in adjacent neighbourhoods. The same would apply to positive or neutral outcomes. This means that there is a tendency for households to behave similarly to their neighbourhood peers than in opposition to them when it comes to certain decisions.

The social networks defined in this chapter are bounded within the Gram Panchayat (GP) of households. A GP represents a village type administrative boundary of a neighbourhood, and so the terms village and GP will be used interchangeably through this chapter.

To facilitate comparative analysis, three distinct distance thresholds are used to define the spatial networks under consideration: 30 metres, 100 metres, and 300 metres. Social connections are defined as follows. Two households i and j are considered to be connected if they reside within a specific radius of each other. If not, they are not directly connected.

3.3.1 Model Assumptions

The model has a non-linear structure with a binary dependent variable (e.g. to own or not own a toilet), and group behaviour variable as the average proportion of other network members engaging in the same behaviour. This format overcomes the reflection problem (Manski 1993) of separately identifying group characteristics from group behaviour. Error terms are heteroskedastic coming from the fact that household rational expectations are heterogeneous. This is due to the presence of different but potentially overlapping reference groups for each household. It is assumed that there are homogenous effects of all neighbours within a household's network. This is due to the use of a local average model structure, in which the adjacency matrices are row-normalised, such that all others in a household's network are given equal weighting.

From data on surveyed households, over 90% of the respondents have resided in the village and occupied the same dwelling for their entire lives. This supports the assumption of exogenous neighbourhood formation due to generational constraints on household and village assignment over time. Specifically, the fact that individuals are born and continue to reside in their current location suggests that their decision to remain in that specific area is firstly determined by circumstantial factors, rather than conscious choice. Therefore, it would be incorrect to assume that all residents of the village made an active decision to live in their current location, as this is not the case for those households whose dwelling is inherited.[33] Additionally, any endogeneity in location is unlikely to associate with sanitation preferences (Pakhtigian et al. 2022).

The assumption of exogenous network formation could fail on the basis of caste. It is known that certain caste groups may agglomerate creating neighbourhood segregations within the same village. To address this issue, caste groups are included as controls in the model. Risk of location choice based on other outcomes that could affect sanitation behaviour are also controlled for by caste which is a strong socio-economic indicator that correlates with education, income, wealth status and other factors.

Finally, the model presented does not include correlated effects in household sanitation decision. This is a strong assumption which asserts that beyond the variables modelled, there are no other correlated factors that could impact sanitation adoption.[34] Not including correlated effects in model estimates that control for shared environmental factors faced by agents, could result in an over-estimation of social effects.

3.3.2 Identification

Manski 1993 popularised difficulty in identification of social effects. Referred to as the reflection problem, it posits difficulty in separately identifying group behaviour (and its resulting effect) from group characteristics (and their effects). However, as shown by Brock and Durlauf (2001, 2007) (B&D), this issue is less prevalent in the case of binary outcomes. Binary outcomes create a non-linear relationship between group characteristics that determine outcomes, and the endogenous group behaviour. Intrinsic non-linearities in the binary choice framework produce identification due to variations in group characteristics $W_{ig}X_{ig}$ and group behaviour $W_{ig}y_{ig}$. Given that group behaviour is bounded between 0 and 1, whilst group characteristics varies with large support (A4 B&D 2007), this reveals nonlinearities that allow linear independence between the two variables. This enables separate identification of endogenous

[33] Some exceptions to this include the movement of female household members upon marriage, urban migration for economic reasons, or displacement.

[34] This assumption can be relaxed with inclusion of fixed effects and random effects which will be added to an extension of this work.

social effects and contextual social effects.

Identification is further supported within this model context as shown by Lee et al. 2014. Unlike the B&D setting of a complete network, the general network structure of this model allows for variation in households' reference groups within the village. In other words, each household is impacted by a heterogeneous set of overlapping neighbours within the network (De Giorgi et al. 2009, Lin 2010). The incomplete nature of the network breaks any linear dependency between group characteristics and group average behaviour amongst individual households. Identification requirements are not as rigid in this case of a non-linear network compared to a linear network. However, the requirements of in-transitivity in the network, differing group sizes, and overlapping networks (Lee 2007, Bramouille et al. 2009) still apply in this network structure if one chose a linear formulation for modelling and identifying social effects.[35]

3.3.3 Equilibrium

When modelling social interaction effects, the risk of multiple equilibria arises if the endogenous social effect is strong enough to create multiple equilibria points that align with individual characteristics. In this case of heterogenous rational expectations, a unique rational expectations equilibrium is generated from the model. This is achieved on the basis that the system of equilibrium equations for agents of a network[36] is a contraction mapping which can reach a rational expectation equilibrium using the Brouwer fixed point theorem. Lee et al. (2014) detail out conditions for the contraction mapping proprieties needed to reach this unique equilibrium.

3.3.4 Empirical Strategy

The model to be estimated (in linear individual household form) is:

[35] A further assumption for identification in the linear case would be that individual characteristics (X) are independent from correlated unobservables. This strict exogeneity assumption allows for identification whilst accounting for correlated unobservables.

[36] As each agent has a different set of neighbours, they each have different equilibrium determining equation parameters characterised by a vector the same length as the network size.

$$y_{ing} = \beta_0 + \beta_1 W_{ing} y_g + \beta_2 x_{ing} + \beta_3 W_{ing} x_g + \varepsilon_{ing} \tag{2}$$

Where:

y_{ing} is individual household sanitation behaviour. In which subscript i is the household, subscript n marks their neighbourhood network and subscript g marks their village.

W_{ing} is the village spatial weight matrix for the decision making household (HH_i). This determines one's neighbourhood network. Where $W_{ij} = 1$ if HH_i and HH_j are neighbours, and $W_{ij}= 0$ if they are not. Households do not include themselves in the weighting matrix so W_{ii}=0. Furthermore, the weighting matrix is row normalised such that if HH_i and HH_j are neighbours, $W_{ij} = \frac{1}{N_{ng}}$, where N_{ng} is the number of households belonging to HH_i's neighbourhood network within the village.[37]

y_g are the outcomes for households within the village. With inclusion of the adjacency matrix W_{ing}, this manifests as the average behaviour of those who are in HH_i's neighbourhood network.

x_{ing} represents the individual characteristics of households within the village.

x_g are the characteristics of households within the village. With inclusion of the adjacency matrix W_{ing}, it becomes the average characteristics of those who are in HH_i's neighbourhood network.

β_1, β_2, and β_3 respectively represent endogenous social effects, contextual social effects and individual household effects.

ε_{ing} represent idiosyncratic shocks to households' sanitation decision

[37] This results in the local average model in which social effects are estimated by the average of neighbours characteristics and behaviours. Unlike the local aggregate model which is based on the total number of neighbours in the network.

Given the binary outcome, a logistic parametrisation is used to structure the estimation of the model such that the probability of engaging in a specific sanitation behaviour (owning a latrine, using a latrine, or open defecating) is equal to the joint distribution of the estimation function:

$$P(y_{ing} = 1) = F(\beta_0 + \beta_1 W_{ing} y_g + \beta_2 x_{ing} + \beta_3 W_{ing} x_g) \tag{3}$$

The logistic expression of this choice probability is:

$$\frac{\exp(\beta_0 + \beta_1 W_{ing} y_g + \beta_2 x_{ing} + \beta_3 W_{ing} x_g)}{1 + \exp(\beta_0 + \beta_1 W_{ing} y_g + \beta_2 x_{ing} + \beta_3 W_{ing} x_g)}$$

Where the village level equilibrium outcome y_g^* is the solution to:

$$\mathbf{y_g^*} = (y_{1g} ..., y_{ng})' = \tanh(\beta_0 + \beta_1 W_{ing} y_g + \beta_2 x_{ing} + \beta_3 W_{ing} x_g) \tag{4}$$

With the log likelihood function:

$$\ln L(\boldsymbol{\beta}|X, W) = \sum_{i=1}^{N_g} [y_{ing} \ln(P(y_{ing} = 1)) + 1 - y_{ing} \ln(P(y_{ing} = 0))] \tag{5}$$

Rational expectation equilibrium $\mathbf{y_g^*}$ is solved iteratively, as a subroutine to the maximum likelihood estimation of the above log likelihood function. One starts with initial parameter guesses of $\boldsymbol{\beta}$, solves for $\mathbf{y_g^*}$, plugs this into the log likelihood function, and carries out the maximum likelihood estimation. Updated values of $\boldsymbol{\beta}$ are used to update and solve for $\mathbf{y_g^*}$ and run the maximum likelihood estimation again. This process continues until convergence is reached. This is done separately for each GP.

The model takes for granted exogenous network formation, and no correlated effects. The limits of these assumptions have already been previously discussed. As most households have resided with the same household members, in the same location all or a significant proportion of their lives, people are quite spatially stable. Furthermore, considering the significant role played by caste groups in influencing various socio-economic attributes of households, incorporating it into the model should help account for certain network or GP level unobservables. The inclusion of caste adds value on two other fronts. It helps account for any non-random assignment that may exist in neighbourhood location determination on the basis of caste. Additionally, it serves as a mechanism to mitigate the impact of caste-based social norms on sanitation practices, thereby maintaining the emphasis on spatially-mediated social interactions.

3.4 Data

3.4.1 Determination of spatial networks

Spatial networks are defined by geodetic distances which calculate the length of the shortest distance between two points on the surface on a mathematical model of the earth.[38] Therefore, edges in the networks are defined by geographic distance between households. The networks are undirected meaning ever household tie is equally mirrored. That is, if HH_i is a neighbourhood to HH_j, the reverse is also true. Metres are the distance metric used in this chapter.

To ensure comparison of useful spatial networks to compare social effects, the networks are defined by distances that would make economic sense to have as a marker for social interactions and sanitation behaviours. Additionally, these spatial networks are necessarily defined in a way that accurately aligns with the context of the dataset.

Figure 1 shows the distribution of pairwise distances between households within their village. Within the dataset households are on average 10 metres away from their nearest neighbours. Mean pairwise distance between households across all villages is 268 metres. The minimal distance between neighbours is 0, because 28% of households have multiple distinct family units living within the same structure, and therefore have the same GPS location. This ranges from 2-14 households per structure. The maximum minimal distance in the entire dataset required to make two households neighbours (i.e. a situation in which there are no isolated nodes) is 350 metres. This ranges from 30 metres to 350 metres, meaning for some GPs there

[38] Used by most geographic information systems (GIS), including Google.

are no isolated nodes by 30 metres. For most GPs, households have a neighbour at least within 50 metres away from them. In 12 GPs households have their nearest neighbour more than 50 metres away, and most of these are within 100 metres.

At 3.1 kilometres all nodes are connected to each other within all GPs. In other words, networks are complete. The 75th percentile of pairwise distances is 330 metres. The smallest maximum distance to make any neighbourhood networks complete is 201 metres. 6 GPs have complete networks below a cut-off of 300 metres.

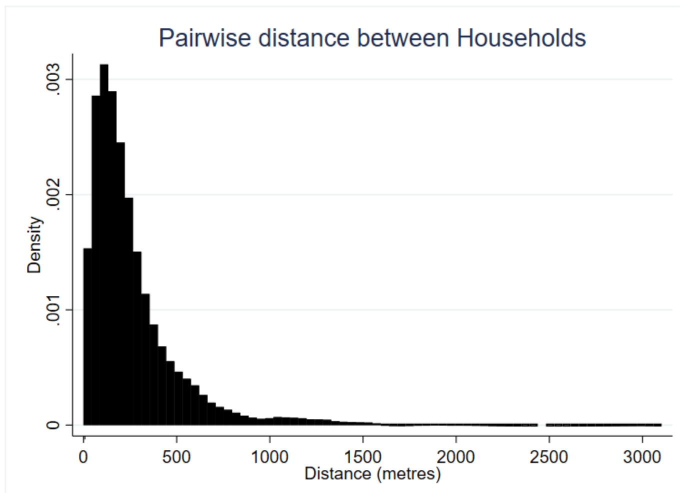

Figure 4: Pairwise distances

Based on the spatial distribution of households across the sample data, the limit for any defined spatial network cut-off must be below 3.1 kilometres. This ensures complete networks are not created across all villages.[39] However, for most households this complete network threshold starts from 300 metres upwards.

Three spatial network cut-offs are chosen based on the above context: 30 metres, 100 metres, and 300 metres.[40] 30 metres reflects close neighbours. This sufficiently captures a net of immediate neighbours within the dataset, and is broad enough

[39]This is necessary because a key part of the model is the use of incomplete networks which define heterogeneous rational expectations, and facilitate identification..

[40]10 metres to use as a cut off is too small. There was a 10-metre error in GPS data collection. Anything from 10 metres or less would be spurious due to this margin of error.

to ensure most households within the dataset have at least one neighbour in their network. 300 metres is a workable upper limit of spatial network to study within the data's context, as well as within economic reasoning. It sufficiently encompasses a sizeable proportion of villages into network clusters without making complete networks. Within villages, about 70% of household pairs lie within 300 metres of each other. This distance also embodies a reasonable extent of spatial mobility within a village; encompassing routine journeys between various amenities such as markets, places of worship, and access to transport. This makes 300 metres provide sufficient scope for varied social interactions to occur.

100 metres acts an intermediary distance between 30 metres and 300 metres. It is broad enough to capture a net of possible social interactions that can create sanitation spill-over effects. This could include for instance, how far parents may allow their children to travel in order to play with other children in the neighbourhood.

All these distance cut-offs reasonably reflect sufficient distance to be affected by a neighbours' negative health externality from OD or positive health externality from toilet use. 300 metres being about 5 mins walk, it also encompasses how far one could be reasonably expected to walk (at any hour of the day) to perhaps access a community toilet.[41]

3.4.2 Household and village characteristics

The dependent variables are latrine ownership, latrine usage and open defecation.[42] Regressors include individual household characteristics, network averages of these characteristics, and average network behaviour. Individual characteristics include: Age of the household head, their gender, household size, and caste group. Caste is categorised into three broad groups: General Castes, Other Backward Class (OBC), and All other caste groups. All other caste groups encompasses Scheduled Castes (SC), Scheduled Tribe (ST), Denotified Tribes (DT), and Nomadic Tribes (NT).

[41] There are no specific guidelines identified in literature which defines reasonable distance cut-offs between neighbours in this context.

[42] OD represents any OD behaviour within a household. It does not refer to OD by all members of a household. Therefore, it only captures if at least one household members is noted to OD irrespective of if the household has a latrine that some of its members use.

Notwithstanding their individual distinctions, these groups have been aggregated to ease estimation. Their grouping as one is supported by the case that unlike General Castes, and OBCs, these groups experience the most socio-economic prejudice in the Indian context.

There are a limited number of regressors in the data, in part restricted to facilitate estimation given the computation burden of equilibrium determination. The trade-off however is the use of census-based data within villages.[43] The data comprises a sample collected over 2012-2014 as part of the baseline and listing data collection for a larger project focused on sanitation adoption in rural Maharashtra, India. A census was collected of GPs with about 480 households or smaller (deemed non-segmented GPs). Where GPs consisted of more than 480 households, if spread over more than one village, one village was selected at random for complete listing and the other(s) were excluded from listing. Additionally, GPs consisting of only one village with more than 480 households were partitioned into segments of 120 households and 4 segments were subsequently randomly selected for data collection. These are considered segmented GPs. A majority of GPs in the dataset were not segmented. The final dataset available for estimation in this case are a total of 29,469 households covering of 95 GPs. Summary statistics of variables are shown in table 3.1.

[43]For some larger villages, data collection was based on segmentation of these villages, and census data was collected in these sub-neighbourhoods. 46% of the sample data is from segmented villages.

Table 3.1: Village level Mean, Standard Deviation, Minimum and Maximum of Individual Characteristics. (N=29,469)

Variable	Mean	Std. Dev.	Min	Max
Age	47.711	14.448	18	99
Household Size	4.946	2.447	1	45
Female Head	0.093	0.291	0	1
OBC	0.275	0.446	0	1
All Others	0.458	0.498	0	1
General Caste	0.268	0.443	0	1
Latrine Own	0.306	0.461	0	1
Latrine Usage	0.315	0.464	0	1
Any OD	0.641	0.480	0	1

Village sizes range from 95 to 484 households with a mean of 310 households per village. Average age of household head is about 48 years, average household size is approximately 5 members and 9% of households are headed by women. Most castes are represented in every GP. Ownership and usage rates are approximately 31% and 32%, whilst OD rate is 64%. Latrine usage and OD are defined by whether at least one household member engages in the behaviour. Therefore, OD, latrine usage and latrine ownership are all present in some households. This motivates the separate estimation of these outcomes as they are not mutually exclusive in the data. Amongst caste groups, latrine ownership and usage rates are generally higher for OBC and General caste households than All Others. Correspondingly, OD rates are lower for these households than for All Others. This is shown in table 3.2 which highlights sanitation rates across the various caste groups within the dataset.

Table 3.2: Sanitation behaviours by caste groups. (N=29,469)

Caste	Own	Use	Any OD
Other Backward Class (OBC) (All Others)	0.358	0.373	0.570
Scheduled Castes	0.240	0.241	0.698
Scheduled Tribe	0.222	0.232	0.717
Denotified Tribe	0.217	0.221	0.715
Nomadic Tribe	0.208	0.218	0.737
General Caste	0.381	0.391	0.597
Total	0.306	0.315	0.641

Note: Table highlights proportion of each caste group that engages in each sanitation behaviour shown in columns.

3.4.3 Network summary statistics

Table 3.3 shows summary statistics for the three spatial networks analysed. Networks are bounded by villages. The data shows summary statistics at both individual and network level. Households on average have 12 neighbours in their network at 30 metres, 79 neighbours at 100 metres and 245 neighbours at 300 metres. As expected, the larger the spatial network, the larger the average number of neighbours a household has within their network. Correspondingly, the larger the defined spatial network, the fewer the number of isolated households.

Degree centralisation measures the concentration of connections among a limited number of nodes. This captures the extent of inequality in the distribution of links throughout the network. A degree centralisation of 0 implies that a network is complete, meaning all households are directly connected to each other. On the other end, a degree centralisation of 1 means a single household is the sole source of direct connections in the network.[44] Table 3.3 shows that degree centralisation is very low across all spatial networks. Maximum centralisation is 0.177 at 30 metres, 0.321 at 100 metres and 0.331 at 300 metres. This removes the risk of centralisation biasing results in social effects due to this particular network characteristic. This could be a

[44] A star network would be the typical example of a network of degree centralisation of 1.

concern in estimation as there are no network fixed effects used to capture unobserved correlated factors at network level. If degree centralisation was high, it would imply the network structure and its attributes such as the location of important nodes (degree central nodes) could significantly impact the assessment of social interaction effects.

Network density is the total number of connected households divided by the possible number of connections if the network was complete. As shown and expected in the data, this density increases with the network size; averages of 0.04, 0.25 and 0.74 for the 30 metre, 100 metre and 300 metre networks respectively.

Table 3.4 displays mean and standard deviation (SD) of all variables for each spatial network. That is the mean and SD of contextual characteristics. The means are very similar. However, the SD become smaller the larger the network. Specifically, a wider network breadth has less fluctuation in mean characteristics and outcomes. This may suggest greater disparities between smaller neighbourhood networks. It is also plausible that this reduction in SD is a result of overlap in networks at larger distance cut-offs compared to smaller ones. Alternatively, it could be an expression of the central limit theorem whereby a larger sample size yields a more precise estimate of the SD.

Table 3.3: Mean, Standard Deviation, Minimum and Maximum of network characteristics.

	Mean	SD	Min	Max
30m				
Neighbours/Degree	12.442	7.861	0	55
Isolates	0.007	0.084	0	1
Degree centrality	0.000	0.084	0	0.002
Degree centralisation	0.064	0.030	0.025	0.177
Density	0.041	0.024	0.014	0.193
100m				
Neighbours/Degree	79.456	39.144	0	253
Isolates	0.0001	0.010	0	1
Degree centrality	0.0027	0.001	0	0.009
Degree centralisation	0.1862	0.061	0.066	0.321
Density	0.2537	0.119	0.100	0.737
300m				
Neighbours/Degree	245.307	91.123	0	483
Isolates	0.000	0.006	0	1
Degree centrality	0.008	0.003	0	0.016
Degree centralisation	0.128	0.096	0	0.331
Density	0.738	0.204	0.301	1.000

Note: Degree and centrality are calculated at an individual node level (N=29,469), whilst centralisation and density are calculated at the network level (N=95).

Table 3.4: Network mean and standard deviation of individual characteristics. (N=29,469)

	Mean			Standard Deviation		
Variable	30m	100m	300m	30m	100m	300m
Age	47.659	47.700	47.723	6.227	3.698	3.062
Household Size	4.945	4.944	4.949	1.102	0.680	0.581
Female Head	0.093	0.093	0.093	0.115	0.052	0.034
OBC	0.274	0.274	0.277	0.326	0.247	0.207
All Others	0.459	0.457	0.456	0.371	0.266	0.180
General Caste	0.267	0.269	0.267	0.321	0.236	0.180
Latrine Ownership	0.306	0.307	0.307	0.279	0.221	0.204
Latrine Usage	0.314	0.316	0.315	0.280	0.221	0.202
Any OD	0.641	0.638	0.640	0.312	0.256	0.230

Table 3.5 highlights correlations between own variables and contextual variables. All are positive, however caste group and OD behaviour exhibit the highest correlation between own and contextual variables. There is a positive relationship between a household's own sanitation behaviour and the sanitation behaviour of their network, with a stronger correlation observed in smaller networks.

Table 3.5: Correlations between own and network level of individual characteristics. (N=29,469)

Variable	30m	100m	300m
Age	0.142	0.195	0.194
Household Size	0.183	0.219	0.215
Female Head	0.065	0.082	0.073
OBC	0.679	0.569	0.459
All Others	0.703	0.565	0.366
General Caste	0.670	0.554	0.407
Latrine Ownership	0.467	0.462	0.436
Latrine Usage	0.464	0.457	0.429
Any OD	0.535	0.524	0.479

The correlation between caste groups of households and their network is notably higher the smaller the network. In fact caste groups exhibit the highest amount of individual and network correlation; ranging 0.4-0.5 at 300 metres, and 0.7 at 30 metres. This suggests the presence of caste homophily. Following the work of Currarini et al. 2009 to test homophily in these spatial networks, two measures of homophily are used to test for spatial clustering of households by caste group.

Table 3.6 shows relative inbreeding homophily for all outcome and independent variables.[45] This is measured by comparing H_i with w_i where $H_i = \frac{s_i}{d_i+s_i}$ representing the proportion of same caste (s_i) neighbourhood links as a proportion of same and different caste links in the network ($d_i + s_i$).

$w_i = \frac{N_i}{N}$ is the relative proportion of a caste group's size to all caste groups in the village. Therefore, a situation in which $H_i = w_i$ implies baseline homophily. That is households cluster in the same proportion of their group size. Relative inbreeding homophily is when $H_i > w_i$. This would mean that households tend to reside nearer their own caste groups disproportionately to relative caste proportions in the village. The opposite reflects relative heterophily $H_i < w_i$.

There is no presence of baseline homophily within the data set at any network level. Table 3.6 shows the proportion of relative inbreeding. At 30 metres, there is 99.95% relative inbreeding homophily for all caste groups. This is the highest amongst the networks. The table also notes that relative inbreeding is higher at 300 metres than 100 metres. This may be indicative of multiple, separated caste neighbourhoods within the same village.

As Currarini et al. 2009 note, the measure of relative homophily may obscure the influence of relative group sizes within a village, resulting in potential bias. Though relative homophily shows a degree of homophily, it fails to capture the extent to which caste groups exhibit locational bias towards others of the same caste, relative to the extent that they could be biased. Two groups could have the same rate of H_i, fulfilling relative homophily with $H_i > w_i$ and yet the extent of this bias is not fully captured because say one group is only 3% of the population, whilst the other

[45] Age and household size homophily are measured using bands. Distribution of bandwidths used are displayed in appendix: table 3.A1 and table 3.A2

group is 90% of the population. To correct for this, they provide an Inbreeding Homophily index (IHI) where $\text{IHI}_i = \frac{H_i - w_i}{1 - w_i}$. This measures the amount of bias in baseline homophily relative to the maximum possible bias (denominator) that could exist in caste group spatial allocation. If IHI_i=0, then we have baseline homophily. Greater than zero means inbreeding homophily, less than 0 means heterophily. At extremes, IHI_i=1 would mean complete inbreeding of caste groups spatially, and IHI_i=-1 means complete heterophily; that is complete mixing between caste groups in household location.

Table 3.6: Proportion of relative caste inbreeding across networks. (N=29,469)

Variable	30m	100m	300m
Caste	0.9995	0.3144	0.5336
Age	0.4191	0.2959	0.0693
Gender	0.4376	0.4351	0.2417
Household Size	0.4225	0.3462	0.1785
Latrine Ownership	0.5260	0.3661	0.1848
Latrine Usage	0.5147	0.3874	0.1524
Any OD	0.4953	0.3810	0.1057

Note: Data shows proportion of relative Inbreeding, as no baseline homophily exists for most variables in data. Only household size exhibits baseline homophily at 30m, and 300m networks. These are respectively 0.06%, and 0.01% of the data.

Table 3.7 shows summary statistics for the IHI for the spatial networks. The highest amount of mean caste inbreeding homophily is in the 30 metre network (0.417). In certain villages this reaches a maximum of 1, meaning for some GPs at 30 metres, caste groups only reside near others of the same caste. At 100 metres, the IHI is on average -0.32 indicating some relative mixing amongst castes. Yet, at 300 metres, this becomes marginally positive at 0.03 indicating a greater tendency towards social grouping in household location compared to a distance of 100 meters. This observation supports the likelihood of multiple separated caste groups residencies in villages.

Caste based neighbourhood segregation is known to be a norm in rural Indian villages. Literature has noted ethnic segregation to be a norm in neighbourhood composition (Schelling 1971).[46] The presence of caste homophily indicates that caste-based networks may also be at play within neighbourhoods.

The use of caste of groups as a control variable in the model estimation should mitigate the bias in the results that may arise from unobserved characteristics correlated with caste, such as income, access to education, and other factors not captured in the data but relevant to sanitation decision making.

[46] Patacchini and Zenou 2012 show importance of geographic distance mediated by ethnic homophily plays a stronger role on employment outcomes. Hellerstein et al. 2014 shows the same around spatial networks and labour outcomes which are stronger across ethnic lines.

Table 3.7: Inbreeding homophily index (IHI) at network levels. (N=29,469)

30m				
Variable	Mean	Standard Deviation	Min	Max
Caste	0.417	0.172	-0.027	1.000
Age	-0.003	0.019	-0.070	0.058
Gender	-0.004	0.055	-0.146	0.146
Household Size	-0.004	0.028	-0.089	0.085
Latrine Ownership	0.001	0.032	-0.097	0.103
Latrine Usage	0.001	0.034	-0.087	0.102
Any OD	0.000	0.032	-0.088	0.102
100m				
Variable	Mean	Standard Deviation	Min	Max
Caste	-0.315	0.493	-2.411	0.816
Age	-0.004	0.006	-0.026	0.015
Gender	-0.003	0.018	-0.082	0.041
Household Size	-0.003	0.010	-0.031	0.050
Latrine Ownership	-0.003	0.013	-0.044	0.042
Latrine Usage	-0.003	0.012	-0.046	0.029
Any OD	-0.003	0.013	-0.045	0.030
300m				
Variable	Mean	Standard Deviation	Min	Max
Caste	0.027	0.102	-0.153	0.775
Age	-0.003	0.003	-0.029	0.007
Gender	-0.002	0.006	-0.020	0.031
Household Size	-0.002	0.004	-0.021	0.014
Latrine Ownership	-0.003	0.004	-0.022	0.019
Latrine Usage	-0.003	0.004	-0.027	0.012
Any OD	-0.003	0.004	-0.028	0.015

Note: IHI > 0 is inbreeding, and IHI < 0 is heterophily. At extremes 0 = Pure baseline homophily , 1 = Complete Inbreeding, -1 = Complete heterophily.

3.5 Results

Parameter coefficients are shown in tables 3.8, 3.10 and 3.12 for latrine ownership, latrine usage, and OD respectively. To ease readability and interpretation of results, marginal effects are reported in tables 3.9, 3.11, and 3.13 for ownership, usage and OD correspondingly. Following the strategy of Lee et al. 2014, marginal effects are measured in two different ways. The first referred to as *naïve* marginal effects shows the standard marginal effects shown in econometric models. It captures a change in outcomes, from a change in a regressor with other variables kept constant, usually at their mean. However, when looking at social effects structured in this way, one ignores the change in network level equilibrium outcomes due to a change in one's characteristics. This can occur in two ways. First it alters the contextual characteristics of other households who have them as a neighbour. Secondly it changes the network average behaviour. The equilibrium changes from this affects the outcome for other households in the changer's network. Therefore, the authors offer what is referred to as *sophisticated* marginal effects. In this version, marginal effects are separately calculated for the Initiator household who experiences an exogenous change in their characteristics, and the Affected households who through social effects are indirectly impacted in their own individual likelihood of an outcome.

To capture this, instead of calculating marginal effects for a single household, a village whose mean household characteristic most closely reflect that of the entire dataset, is chosen.[47] Marginal effects are repeatedly calculated for all as Initiator and Affected households in the village. The average of these calculations is taken to get sophisticated marginal effects.

Tables 3.9, 3.11, and 3.13 show the two types of marginal effects. Endogenous and contextual social effects have no output for the sophisticated marginal effects calculation. The reason for this is that marginal effects typically refer to a change in an independent or exogenous variable. As a result, in the sophisticated approach, marginal effects cannot be calculated for contextual characteristics and endogenous characteristics which are intrinsically determined endogenously.

[47] Descriptives of the representative village are shown in table 3.A3 of the appendix.

3.5.1 Social Effects

For latrine ownership, one sees in table 3.9 (naïve calculation) that endogenous social effects are positively related with latrine ownership. For the 30-metre network, a household's likelihood of owning a latrine increases by 15% when their group's average behaviour towards latrine ownership increases by 50%. In other words, moving from a group in which no one owns a latrine to a group where half the members own a latrine. This becomes an approximate 8% negative impact at both the 100-metre and 300-metre networks. All statistically significant, this implies that latrine ownership is most positively influenced in the smaller network.

Table 3.8: Coefficient parameter estimates for latrine ownership. N= 29,469

	30m	100m	300m
Constant	1.7848***	5.3126***	6.2059***
	(0.0734)	(0.4373)	(0.4707)
Network Behaviour	0.6262***	-0.6388*	-0.6224*
	(0.0593)	(0.331)	(0.3627)
Age	-0.0127***	-0.0129***	-0.0125***
	(0.0004)	(0.0005)	(0.0005)
Female Head	0.3002***	0.2891***	0.2846***
	(0.0224)	(0.023)	(0.023)
Household Size	-0.0429***	-0.0424***	-0.0428***
	(0.0025)	(0.0026)	(0.0026)
OBC	-0.3631***	-0.4343***	-0.4623***
	(0.0206)	(0.0183)	(0.0167)
General Caste	-0.5318***	-0.571***	-0.5883***
	(0.0204)	(0.0176)	(0.0159)
Avg. Age	-0.0044***	-0.0397***	-0.0524***
	(0.0008)	(0.0033)	(0.0037)
Proportion Female Head	-0.238***	0.0214	0.9664***
	(0.0456)	(0.1499)	(0.2286)
Avg. Household Size	-0.0047	-0.1535***	-0.1823***
	(0.0049)	(0.0148)	(0.0172)
Proportion OBC	-0.1716***	-0.7855***	-1.2729***
	(0.029)	(0.1035)	(0.1148)
Proportion General caste	0.0283	-0.4293***	-0.8276***
	(0.0315)	(0.1251)	(0.1424)

t statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.01

Table 3.9: Marginal effects parameter estimates for latrine ownership.

	30m			100m			300m		
	Naïve	Initiator	Affected	Naive	Initiator	Affected	Naive	Initiator	Affected
Group Behaviour	15.315	-	-	-8.592	-	-	-8.819	-	-
Age	-0.310	-0.634	-0.001	-0.174	-0.346	-0.002	-0.177	-0.354	-0.003
Female Head	-7.279	44.623	39.669	-3.594	44.079	39.165	-3.732	42.720	37.836
Household Size	-1.048	-2.129	-0.003	-0.570	-1.133	-0.007	-0.607	-1.211	-0.009
OBC	8.845	35.350	39.666	6.253	34.082	39.156	7.042	32.296	37.793
General Caste	12.990	35.310	39.675	8.203	34.431	39.164	8.920	32.794	37.805
Avg. Age	-0.107	-	-	-0.534	-	-	-0.743	-	-
Proportion Female Head	-5.820	-	-	0.288	-	-	13.693	-	-
Avg. Household Size	-0.115	-	-	-2.065	-	-	-2.584	-	-
Proportion OBC	-4.196	-	-	-10.566	-	-	-18.037	-	-
Proportion General caste	0.692	-	-	-5.774	-	-	-11.727	-	-

Notes: Naïve marginal effects refer to a change in outcomes from a change in a variable, leaving other regressors fixed at their means. Sophisticated marginal effects account for the system change in contextual characteristics and behaviour from a change in a variable, and so reflects a separate change on the Changer from those in the changer's network. Therefore the Initiator column represents the direct effect from the change on the "Initiator" of change. the Affected column reflects an indirect effect on the "Affected" households in the same network as the Initiator household.

Contextual effects show mixed results depending on the contextual characteristic observed. Focusing on significant results, one can see that the average household head age of a group has a small negative impact (0.1%) on the likelihood of a household owning a latrine. As the network grows, this negative effect increases slightly but remains below 1%.

Though only statistically significant at the 30-metre and 100-metre network, higher proportions of female headed households goes from negatively impacting household own likelihood of adoption at 5% (30m) to positively impacting it at 13% (300m).

Statistically significant at 100 metres, and 300 metres, the impact of average household size is consistently negative at around 2% with the likelihood of owning a latrine.

Table 3.10: Coefficient parameter estimates for latrine usage. N= 29,469

	30m	100m	300m
Constant	1.675***	4.7934***	4.6928***
	(0.0734)	(0.444)	(0.4825)
Group Behaviour	0.6089***	-0.6105*	-0.1763
	(0.0612)	(0.3459)	(0.3803)
Age	-0.0112***	-0.0115***	-0.0112***
	(0.0004)	(0.0005)	(0.0005)
Female Head	0.291***	0.2791***	0.2812***
	(0.0224)	(0.0229)	(0.0236)
Household Size	-0.0402***	-0.04***	-0.0403***
	(0.0025)	(0.0026)	(0.0026)
OBC	-0.3769***	-0.4422***	-0.4816***
	(0.0205)	(0.0182)	(0.0171)
General Caste	-0.5363***	-0.5752***	-0.609***
	(0.0204)	(0.0175)	(0.0164)
Avg Age	-0.0042***	-0.0326***	-0.0339***
	(0.0008)	(0.0032)	(0.0036)
Proportion Female Head	-0.2465***	0.0789	0
	(0.0461)	(0.15)	(0.2156)
Avg. Household Size	0.0004	-0.1321***	-0.1296***
	(0.0049)	(0.0147)	(0.0168)
Proportion OBC	-0.2035***	-0.9635***	-1.1518***
	(0.0308)	(0.1218)	(0.1353)
Proportion General caste	0.0062	-0.5412***	-0.6991***
	(0.033)	(0.1391)	(0.157)

t statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.01

Table 3.11: Marginal effects parameter estimates for latrine usage.

	30m			100m			300m		
	Naïve	Initiator	Affected	Naive	Initiator	Affected	Naive	Initiator	Affected
Group Behaviour	14.908	-	-	-8.780	-	-	-3.403	-	-
Age	-0.274	-0.562	-0.001	-0.166	-0.330	-0.002	-0.215	-0.431	-0.003
Female Head	-7.061	42.864	37.983	-3.735	42.245	37.416	-5.140	41.719	36.803
Household Size	-0.985	-1.991	-0.002	-0.575	-1.143	-0.007	-0.778	-1.555	-0.010
OBC	9.196	33.459	37.979	6.781	32.225	37.402	9.743	31.000	36.782
General Caste	13.116	33.523	37.989	8.812	32.614	37.412	12.360	31.550	36.796
Avg. Age	-0.104	-	-	-0.469	-	-	-0.655	-	-
Proportion Female Head	-6.036	-	-	1.135	-	-	0.000	-	-
Avg. Household Size	0.011	-	-	-1.901	-	-	-2.501	-	-
Proportion OBC	-4.982	-	-	-13.856	-	-	-22.230	-	-
Proportion General caste	0.153	-	-	-7.784	-	-	-13.493	-	-

Notes: Naïve marginal effects refer to a change in outcomes from a change in a variable, leaving other regressors fixed at their means. Sophisticated marginal effects account for the system change in contextual characteristics and behaviour from a change in a variable, and so reflects a separate change on the Changer from those in the changer's network. Therefore the Initiator column represents the direct effect from the change on the "Initiator" of change. the Affected column reflects an indirect effect on the "Affected" households in the same network as the Initiator household.

Table 3.11 highlights marginal effects for latrine usage. At 30 metres there is a positive marginal effect of 15% increase in likelihood of using a latrine when one's group increases average latrine usage by 50%. Similarly, to latrine ownership, this effects changes direction for the larger spatial networks. Latrine usage is negative at 9% for 100 metres, and 3% for 300 metres.

Contextual effects in latrine usage show mixed results. Increases in average household head age of one's group is negatively tied to latrine usage. This negative effect grows in magnitude as the network grows, but remains under 1%.

For the 30-metre network, an increase in the proportion of female headed households reduces likelihood of latrine usage by 6%. This changes direction for the other networks, but those results are statistically insignificant.

Only statistically significant at 100 metres and 300 metres, increases in average household size is negatively tied to latrine usage.

Table 3.12: Coefficient parameter estimates for OD. N= 29,469

	30m	100m	300m
Constant	-0.989***	-0.4252	-0.4001
	(0.0736)	(0.4043)	(0.8269)
Group Behaviour	0.4813***	0.8235**	0.9309
	(0.0738)	(0.3695)	(0.7542)
Age	0.0091***	0.01***	0.0099***
	(0.0004)	(0.0004)	(0.0004)
Female Head	-0.2085***	-0.2233***	-0.2186***
	(0.0218)	(0.0207)	(0.0203)
Household Size	0.0316***	0.0357***	0.0381***
	(0.0025)	(0.0024)	(0.0023)
OBC	0.2784***	0.3129***	0.3419***
	(0.0204)	(0.0168)	(0.015)
General Caste	0.4307***	0.4819***	0.4941***
	(0.0204)	(0.0166)	(0.0146)
Avg Age	-0.0001	-0.0086***	-0.0089*
	(0.0008)	(0.0027)	(0.0051)
Proportion Female Head	0.423***	0.9558***	1.2891***
	(0.0478)	(0.09)	(0.1383)
Avg. Household Size	-0.0477***	-0.0592***	-0.0619***
	(0.0051)	(0.0093)	(0.0161)
Proportion OBC	0.3106***	0.1911	0.1574
	(0.0382)	(0.158)	(0.3211)
Proportion General caste	-0.0844**	-0.2148*	-0.1938
	(0.0333)	(0.1184)	(0.2387)

t statistics in parentheses

* p<0.10, ** p<0.05, *** p<0.01

Table 3.13: Marginal effects parameter estimates for OD behaviour.

	30m			100m			300m		
	Naïve	Initiator	Affected	Naive	Initiator	Affected	Naive	Initiator	Affected
Group Behaviour	11.805	-	-	20.081	-	-	22.439	-	-
Age	0.224	0.451	0.000	0.243	0.484	-0.001	0.239	0.477	-0.001
Female Head	5.072	-33.559	-29.910	5.480	-36.027	-31.985	5.320	-38.060	-34.034
Household Size	0.774	1.495	-0.006	0.870	1.724	-0.008	0.918	1.831	-0.009
OBC	-6.844	-26.449	-29.917	-7.515	-28.184	-32.041	-8.090	-29.916	-34.125
General Caste	-10.570	-26.040	-29.930	-11.624	-27.815	-32.058	-11.776	-29.797	-34.142
Avg. Age	-0.002	-	-	-0.210	-	-	-0.214	-	-
Proportion Female Head	10.375	-	-	23.308	-	-	31.073	-	-
Avg. Household Size	-1.169	-	-	-1.443	-	-	-1.493	-	-
Proportion OBC	7.619	-	-	4.660	-	-	3.793	-	-
Proportion General caste	-2.070	-	-	-5.237	-	-	-4.672	-	-

Notes: Naïve marginal effects refer to a change in outcomes from a change in a variable, leaving other regressors fixed at their means. Sophisticated marginal effects account for the system change in contextual characteristics and behaviour from a change in a variable, and so reflects a separate change on the Changer from those in the changer's network. Therefore the Initiator column represents the direct effect from the change on the "Initiator" of change. the Affected column reflects an indirect effect on the "Affected" households in the same network as the Initiator household.

The marginal effects on OD behaviour in table 3.13 show that households are positively influenced by the share of one's neighbourhood that also OD. At 30 metres a household is 12% more likely to OD when half their neighbours also OD. This increases to 20% impact at 100 metres and 22% at 300 metres (though statistically insignificant).

Having a 50% increase in female headed households increases the likelihood of OD by 10%, 23%, and 31% for the corresponding 30 metre, 100 metre, and 300 metre networks. The effect increases as the network size grows, similar to latrine ownership (but with a negative impact in that case). This would suggest that similarly to latrine ownership, female headed households influence each other's behaviour as their neighbourhood network of fellow female headed household grows, irrespective of whether the sanitation behaviour is a socially desirable outcome or not.

Overall endogenous social effects show different effects based on network size and dependent on sanitation outcomes. Positive behaviours such as latrine ownership and usage are more positively influenced by group behaviour at small networks (30m), whilst the negative sanitation outcome of OD persistently increases with network size.

Contextual effects vary depending on contextual characteristics of reference. However, gender and caste distribution have notably higher magnitudes of contextual social effects on sanitation outcomes. Contextual effects of female household head grow with network size, and is most prevalent for OD. Greater presence of female headed households tends to be positively tied to increased likelihood of sanitation behaviours as the network cut off increases, irrespective of whether the sanitation outcome is desirable (latrine ownership) or not (OD). This suggests gender may be another strong source of peer effects.

3.5.2 Individual Effects

In the naïve marginal effects, age of the household head has a marginally negative effect for latrine ownership, and usage both at 0.3% for the 30-metre network. The impact of age on OD is weakly positive, with a marginal effect of 0.2%. The impact of age exhibits relatively stable magnitude on sanitation outcomes as network size increases.

The empirical findings reveal that the negative impact of age is pronounced on Initiator households with negligible effects on Affected households. This is consistent across network sizes. Holding constant all other relevant variables, the Initiator marginal effect reveals the influence of ageing on outcomes. While this change in age works through the equilibrium, it comes as no surprise that it exerts little to no influence on the outcomes of affected households in the network.

Female household heads in the naïve marginal effects appear to be less likely to own, and use a latrine and more likely to OD. However, when the social effects character of the system is accounted for it is found that female headed households are significantly more likely own and use a latrine, and less likely to OD compared to male headed households. For instance, in latrine usage (table 3.11), it is shown that initiator female headed households are 43% more likely to use latrines, whilst affected female headed households are 38% more likely to use a latrine. The magnitude of these effects remains similar across network sizes. This observation that female household heads are more likely to engage in improved sanitation practices is consistent with the economic intuition that women generally derive greater benefits from latrine usage. These benefits can be objectively measured in terms of the increased safety, health, and privacy that latrine access provides women. Additionally, women tend to have a higher subjective valuation of latrine ownership compared to men across households (Augsburg et al. 2022).

Household size is negatively tied to latrine ownership (0.5%-1% across networks) and usage (0.6%-1% across networks), and positively tied to OD (0.8%-0.9% across networks). When comparing initiator and affected households, though the direction of effects remains the same, as in the naïve form, the impact of one's household size is larger for initiator (1%-2%) than affected (0%) households. Consistent with the intuition regarding the age of household head, if a household increases in size, it is expected that their sanitation outcomes will be more directly affected than the sanitation outcomes of their neighbours.

Compared to All Other castes General castes and OBC are less likely to OD, more likely to own and use latrines. This is as expected because these groups are generally socio-economically advantaged than All Other castes. This enables them to be able to access and afford to build latrines more readily relative to other caste groups. As the summary statistics showed these two groups had higher rates of latrine ownership and usage, and lower rates of OD than other caste groups.

Naïve marginal effects generally range from 6%-13% for all sanitation outcomes for OBC and General castes. However, with sophisticated marginal effects, these effects increase significantly for all sanitation outcomes. At 30 metres, for latrine ownership, Initiator OBC and General caste households are both 35% more likely to own latrines than All Other castes. These effects drop slightly in magnitude as the network grows, dropping to 34% at 100 metres and 33% at 300 metres. The effects are positive and similar in magnitude for latrine usage; displaying a similar patter of slightly reduced effects as the network size grows.

OD for Initiator OBC and General caste household at 30 metres is negative. OBC and General caste Initiator households are both 26% less likely to OD than All Others. This negative effect increases to 28% and 30% for the 100-metre and 300-metre networks respectively.

Of additional note is the difference in effect between initiator and affected households. For both OBC and General castes, the affected households are just as, if not more likely, to engage in the same sanitation outcome as the initiator household. As the sophisticated marginal effects accounts for the environmental spill-overs of a marginal change in a household's characteristic, this implies that caste belonging has a significant impact on the equilibrium effect, and resultant indirect impact on how other households determine their sanitation behaviour.

The marginal effects of individual characteristics are similar across networks, in the range of 3% difference within variables. This is the case whether looking at naïve or sophisticated marginal effects. This aligns with the expectation that individual characteristics are less variant to the influence of social networks. Additionally, for Initiator households, the sophisticated marginal effects are consistently higher than the naïve ones.

3.5.3 Discussion

The difference in social effects from the spatial networks on positive and negative sanitation behaviours implies different mechanisms through which these behaviours diffuse within villages. OD is socially prevalent across all, and is a norm for all, only getting more embedded as a collective practice the more people engage in the behaviour. However, with latrine usage and latrine ownership, tighter spatial networks are more relevant for positive spill-overs from these behaviours. This in part could reflect the fact that smaller networks induce more tight-knit interactions, leading to more socially cohesive groups. This social cohesion may come from multiple attributes of which caste is amongst. As shown through the homophily measure, the presence of social-spatial network sorting along caste lines is strongest at 30 metres compared to the other two networks. These findings suggest that the greater endogenous social effects observed at a 30-metre radius for latrine ownership and usage may indicate households' greater propensity to conform to the behaviour of their perceived social peers, particularly within their caste group. Alternatively, it is possible that the strategic complementarity of positive sanitation outcomes is more discernible and, therefore, easier to conform to at 30 meters compared to larger distances, whilst OD is more visible across the entire GP.

There is a possibility that a combination of mechanisms, including social learning, pressure, coordination, and physical proximity interdependencies, may contribute to the observed results. This is because spatial networks can capture the influence of spatial proximities on infection, facilitate the formation of new social connections, and serve as a proxy for pre-existing social ties. As discussed earlier in this chapter, sanitation intersects with the social and spatial dimensions. The negative externalities associated with OD, particularly in terms of disease transmission and pollution, are emphasised in close geographic proximity. Previous literature has also highlighted the decay of social interactions with physical distance (Hipp and Perrin 2009, Preciado et al. 2012). This indicates that proximity plays a crucial role in the frequency of social interactions within rural environments. Moreover, spatial location can proxy for other social dimensions such as caste, religion, education, income, and social class, further supporting the notion that geographical proximity interdependencies contribute to the intensity of social contact and the observed correlations between distance and sanitation behaviour.

The findings imply that female-headed households have a strong influence on other households in their network, including male-headed households. The impact on male-headed households is evident from their inclusion in the network of affected households in the female-headed household group. This gender-based asymmetry in peer effects is consistent with other studies (Pakhtigian et al. 2022). Female-headed households exhibit significant effects on their own sanitation adoption behaviour and that of other households. These outcomes imply that gender can serve as a critical reference point for amplifying spillover effects in sanitation adoption.

The policy implications from the results suggest a difference in approach to community level sanitation programmes. Much work done on sanitation policy at neighbourhood levels involve community led total sanitation (CLTS) programmes. Crocker et al. (2016) have shown that CLTS works best in more socially cohesive villages.[48] Therefore focusing sanitation programmes on these subgroups which can be identified through spatial referencing, may improve success in such behaviour change programmes.

Other community based sanitation programmes involve interdependent decision making through public commitments (Bakhtiar et al. 2023). These can be useful for catalysing interdependency in sanitation decisions. Given this chapter shows that social spill-overs are strongest at smaller distanced communities for ownership and usage, these could be key levels for targeting such strategic commitment devices.

While specific to the Indian context, the results support the importance of spatial networks for shaping social norms and other socially relevant outcomes. This holds true for other contexts where space is a key boundary for social interactions, as well as a proxy for other social components such as neighbourhood socio-economic status.

3.5.4 Limitations

One of the initial limitations of this chapter is the assumption that geographic proximity sufficiently captures the social structures of relevance. However, the reality is that social interactions exist beyond geography, so some tighter and more influential social relations may be geographically further away than relations that are spatially close. Therefore, the influence of different neighbours on a household's decision-making is likely to vary. However, this study primarily focuses on the role of spatial proximity in shaping social interactions. Specifically, it examines how social effects manifest through spatial distance. That said, a means of separately measuring household reported social ties, as well as spatial locations to help differentiate spatial network and social network effects distinctly would be of interest for future

[48] Their work was applied to Ghana.

research.

Secondly, there is lack of information on spatial structure such as where key hubs are based, or how close households are to complementary infrastructure that affect sanitation such as water sources (Bharat et al. 2020). Blind location may distort results when not accounting for other spatial dynamics at play. Therefore, there might be effects linked to geographic structure; distance between key hubs, households in relation to these hubs, and means of access. This could create bias in results which look solely at distance between households, without controlling for these other spatial components.

The presence of limited control variables such as household wealth or household risk preferences which may influence adoption of health technologies such as latrines might bias social effects.[49]

Lastly, social effects from correlated effects are not included in this chapter. The inclusion of group fixed effects and random effects would have helped address the presence of any correlated effects that exist at the network or village level. The inclusion of caste group controls should partially have addressed this, given correlated effects in sanitation that exist at the caste group level. However, other correlated effects which exist means the results for endogenous social effects are potentially slightly biased upwards.

Finally, it is worth considering the possibility of an incorrect equilibrium selection rule. Nevertheless, Lee et al. 2014 observe that the potential bias arising from this is small.

3.6 Conclusion

In this chapter, spatially defined neighbourhood networks are to used to assess endogenous and contextual social effects on household sanitation decision: latrine ownership, latrine usage and OD. Three different networks are defined to asses how these social effects may differ depending on the bandwidth of the neighbourhood network: 30 metres, 100 metres, and 300 metres. Results show that endogenous social effects

[49] In this study, limited control variables were used to ease computational burden of estimation. Inclusion of too many variables restricted the ability of the maximum likelihood estimation to converge.

in latrine ownership and usage are positive at 30 metres, but negative at 100 metres and 300 metres. On the other hand, endogenous social effects in OD are consistently positive, only getting larger in magnitude as the network bandwidth increases. Results show that contextual social effects vary depending on contextual characteristic observed. Average age of household head and average household size are negatively correlated with latrine ownership and usage, whilst positively correlated with OD. The findings suggest that higher proportions of female-headed households are positively correlated with both open defecation (OD) and latrine ownership at larger network sizes, with the magnitude of these effects increasing as the network bandwidth expands.

Neighbourhood specific implementation of sanitation programmes can take advantage of the role of social effects in improving outcomes. However, results show the importance of ensuring that sanitation programmes such as CLTS are tied to the most relevant subsets of communities in which they aim to serve, in order to take full advantage of these social multiplier effects. For the Indian context, doing this in a caste targeted way takes advantage of the homophily in neighbour network structure. Also, it means targeting can be more nuanced to account for varying perceptions around sanitation across different caste groups. The larger contextual effects of female headed households suggests gender based targeting may be influential for sanitation adoption.

The mixed effects from positive versus negative sanitation behaviours imply different things about how behaviour change can be facilitated. Endogenous social effects in OD are consistently positive, and grow with network size. However, positive behaviours such as ownership and usage are only positively linked at the closer network proximity of 30 metres. This means that for increasing latrine ownership and usage, the value of endogenous social effects is most effective if behaviour change is targeted at this lower spatial level. Whilst concurrently policies to reduce OD can be carried out across whole villages given that this behaviour is strongly and increasingly spatially correlated the wider the spatial boundary.

This chapter intentionally overlaps social effects with spatial effects by using the spatial networks to proxy for social networks. However, spatial networks are not the only boundary through which social interactions relevant for sanitation outcomes operate in rural environments. It may be interesting for future research to measure household reported social ties as well as spatial locations to help differentiate spatial network and social network effects distinctly, as well as identify any degree of overlap in these effects.

In order to gain better understanding of variation in social effects it would be useful for future research to establish the capacity to differentiate between heterogeneous social effects across households within a network. Such an approach would enable one to assess the susceptibility to social effects amongst different individuals and groups.

3.7 References

Abramovsky, L., Augsburg, B., Lührmann, M., Oteiza, F., Rud, J., 2023. "Community Matters: Heterogeneous Impacts of a Sanitation Intervention," Working Papers 210, Red Nacional de Investigadores en Economía (RedNIE).

Augsburg, B., Malde, B., Olorenshaw, H., Wahhaj, Z. (2022). To invest or not to invest in sanitation: the role of intra-household gender differences in perceptions and bargaining power. 23/04. London: Institute for Fiscal Studies.

Bakhtiar, M., Guiteras, R., Levinsohn, J., Mobarak, A. (2023) Social and financial incentives for overcoming a collective action problem, Journal of Development Economics, Volume 162, 103072, ISSN 0304-3878

Bartram J, Charles K, Evans B, O'Hanlon L, Pedley S. Commentary on community-led total sanitation and human rights: should the right to community-wide health be won at the cost of individual rights? J Water Health. 2012 Dec;10(4):499-503.

Behera MR, Pradhan HS, Behera D, Jena D, Satpathy SK. Achievements and challenges of India's sanitation campaign under clean India mission: A commentary.

J Edu Health Promot 2021;10:350

Bharat, G., Dkhar, N., and Abraham, M. 2020. Aligning India's Sanitation Policies with the Sustainable Development Goals (SDGs), TERI Discussion Paper.

Bramoullé Y., Djebbari H., Fortin B. (2009) Identification of peer effects through social networks, Journal of Econometrics, Volume 150, Issue 1, Pages 41-55

Cameron L, Santos P, Thomas M, Albert J. Sanitation, financial incentives and health spillovers: A cluster randomised trial. J Health Econ. 2021 May;77:102456.

Cameron, L., Gertler, P., Shah, M., Alzua, M.L., Martinez, S., Patil, S. (2022) The dirty business of eliminating open defecation: The effect of village sanitation on child height from field experiments in four countries, Journal of Development Economics, Volume 159, 102990, ISSN 0304-3878

Caruso, B., Sclar G.,, Routray P., Nagel C., Majorin F., Sola S., Koehne W., Clasen T., Effect of a low-cost, behaviour-change intervention on latrine use and safe disposal of child faeces in rural Odisha, India: a cluster-randomised controlled trial, The Lancet Planetary Health, Volume 6, Issue 2, 2022, Pages e110-e121, ISSN 2542-5196.

Chakraborty, S., Novotný, J., Das, J., Bardhan, A., Roy, S., Mondal, S., Patel, P.P., Santra, S., Maity, I., Biswas, R., Maji, A. and Pramanik, S. (2022), Geography matters for sanitation! Spatial heterogeneity of the district-level correlates of open defecation in India. Singapore Journal of Tropical Geography, 43: 62-84.

Coffey, D., Gupta, A., Hathi, P., Khurana, N., Spears, D., Srivastav, N., Vyas, S. (2014). Revealed preference for open defecation: Evidence from a new survey in rural north India. Economic and Political Weekly. 49. 43-55.

Coffey, D. and Spears, D. (2017) Where India Goes: Abandoned Toilets, Stunted Development and the Costs of Caste. New Delhi: HarperCollins Publishers India.

Crocker J, Abodoo E, Asamani D, Domapielle W, Gyapong B, Bartram J. Impact Evaluation of Training Natural Leaders during a Community-Led Total Sanitation Intervention: A Cluster-Randomized Field Trial in Ghana. Environ Sci Technol. 2016 Aug 16;50(16):8867-75.

Currarini, S., Jackson, M. O., Pin, P. (2009). An Economic Model of Friendship: Homophily, Minorities, and Segregation. Econometrica, 77(4), 1003–1045.

De Giorgi, G., Pellizzari, M., and Redaelli, S., Be as Careful of the Company You Keep as of the Books You Read: Peer Effects in Education and on the Labor Market (May 2009). NBER Working Paper No. w14948.

Del Bello, C.L., Patacchini, E., Zenou, Y., 2014. Peer effects: social distance or geographical distance? Unpublished manuscript, Stockholm University.

Drèze, J. and Sen, A., (2013). An Uncertain Glory: India and its Contradictions. Princeton University Press, Princeton, NJ, 2013. 434 pp. (Hardback). ISBN 978-0-691-16079-5

Emch M. , Root E.D., Giebultowicz S. , Ali M. , Perez-Heydrich C., Yunus M. (2012) Integration of Spatial and Social Network Analysis in Disease Transmission Studies, Annals of the Association of American Geographers, 102:5, 1004-1015.

Exum NG, Gorin EM, Sadhu G, Khanna, A., Schwab, K. Evaluating the declarations of open defecation free status under the Swachh Bharat ('Clean India') Mission: repeated cross sectional surveys in Rajasthan, India. BMJ Global Health 2020;5:e002277.

Gautam, S. (2018). Household (Under) Adoption of Sanitation: Externalities and Borrowing Constraints.

Gautam, S. (2018). Quantifying Welfare Effects in the presence of Externalities: An Ex-ante Evaluation of a Sanitation Intervention.

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

Gupta A, Khalid N, Desphande D, Hathi P, Kapur A, Srivastav N, Vyas S, Spears D, Coffey D (2019), Changes in Open Defecation in Rural North India: 2014-2018 ,IZA Discussion Papers No. 12065, Institute of Labor Economics (IZA), Bonn.

Hellerstein J. Kutzbach M., Neumark D., Do labor market networks have an important spatial dimension? Journal of Urban Economics, Volume 79, 2014, Pages 39-58, ISSN 0094-1190.

Helmers, C. and Patnam, M. (2014), Does the rotten child spoil his companion? Spatial peer effects among children in rural India. Quantitative Economics, 5: 67-121.

Hipp, J. R., Perrin, A. J. (2009). The Simultaneous Effect of Social Distance and Physical Distance on the Formation of Neighborhood Ties. City and Community, 8(1), 5–25.

Kennedy-Walker, R., Amezaga, J.M., Paterson, C.A. (2015) The impact of community social dynamics on achieving improved sanitation access for the urban poor: The case of Lusaka, Zambia. Habitat International, Volume 50, Pages 326-334, ISSN 0197-3975.

Kresch, E., Lipscomb, M., Schechter, L. 2020, Externalities and Spillovers from

Sanitation and Waste Management in Urban and Rural Neighborhood, Applied Economic Perspectives and Policy (2020) volume 42, number 3, pp. 395–420

Lee L., Identification and estimation of econometric models with group interactions, contextual factors and fixed effects (2007) Journal of Econometrics, Volume 140, Issue 2, Pages 333-374.

Lee, L., Li, J., Lin, X. (2014). Binary Choice Models with Social Network under Heterogeneous Rational Expectations. The Review of Economics and Statistics, 96(3), 402–417.

Lin, X. (2010). Identifying Peer Effects in Student Academic Achievement by Spatial Autoregressive Models with Group Unobservables. Journal of Labor Economics, 28(4), 825–860.

Lunn A (2020) Urban family ties and household latrines in rural India: A cross-sectional analysis of national data. PLoS ONE 15(7): e0235677.

Manski, C.F. (1993). Identification of Endogenous Social Effects: The Reflection Problem. The Review of Economic Studies [Online] 60:531–542.

Orgill-Meyer J, Pattanayak SK, Chindarkar N, Dickinson KL, Panda U, Rai S, Sahoo B, Singha A, Jeuland M. Long-term impact of a community-led sanitation campaign in India, 2005-2016. Bull World Health Organ. 2019 Aug 1;97(8):523-533A.

Pakhtigian, E., Dickinson, K., Orgill-Meyer, J., Pattanayak, S., 2022. "Sustaining latrine use: Peers, policies, and sanitation behaviors," Journal of Economic Behavior and Organization, Elsevier, vol. 200(C), pages 223-242.

Patacchini, E., and Zenou, Y., (2012), Ethnic networks and employment outcomes, Regional Science and Urban Economics, 42, issue 6, p. 938-949.

Preciado, P., Snijders, T., Burk, W., Stattin, H., Kerr, M. (2012). Does Proximity Matter? Distance Dependence of Adolescent Friendships. Lancet. 34. 10.1016/j.socnet.2011.01.002.

Rogers, E.M. (2003) Diffusion of Innovations. Free Press, New York.

Routray P, Schmidt WP, Boisson S, Clasen T, Jenkins MW. Socio-cultural and behavioural factors constraining latrine adoption in rural coastal Odisha: an exploratory qualitative study. BMC Public Health. 2015 Sep 10;15:880.

Shakya HB, Christakis NA, Fowler JH. Association between social network communities and health behavior: an observational sociocentric network study of latrine ownership in rural India. Am J Public Health. 2014 May;104(5):930-7.

Schelling, T. (1971) Dynamic models of segregation, The Journal of Mathematical Sociology, 1:2, 143-186.

The Sustainable Development Goals Report 2022. United Nations Statistics Division 2022.

Topa, G., and Zenou Y., 2015, Neighborhood and Network Effects, Handbook of Regional and Urban Economics, Volume 5A. ISSN 1574-0080.

Tukahirwa J.T. , Mol A.P.J. , Oosterveer P. , Access of urban poor to NGO/CBO-supplied sanitation and solid waste services in Uganda: The role of social proximity. Habitat International, Volume 35, Issue 4, 2011, Pages 582-591, ISSN 0197-3975.

Venkataramanan, V., Crocker, J., Karon, A., and Bartram, J. (2018) Community-Led Total Sanitation: A Mixed-Methods Systematic Review of Evidence and Its Quality, Environmental Health Perspectives 126:2 CID: 026001

WHO/UNICEF JMP (2021), Progress on household drinking water, sanitation and hygiene 2000-2020 Five years into the SDGs.

Zuin, V., Delaire, C., Peletz, R., Cock-Esteb, A., Khush, R., Albert, J., 2019. "Policy Diffusion in the Rural Sanitation Sector: Lessons from Community-Led Total Sanitation (CLTS)," World Development, Elsevier, vol. 124(C), pages 1-1.

3.8 Appendix C: Descriptive Statistics

Table 3.A1: Household size distribution

Household Size	Frequency	Percent
3 and under	6,896	23.4
9 and under	21,224	72.02
10 plus	1,349	4.58
Total	29,469	100

Table 3.A2: Age of household head distribution

Age of household head	Frequency	Percent
29 under	2,497	8.47
39 under	6,384	21.66
49 under	7,010	24
59 under	5,151	17
69 under	5,639	19
79 under	2,214	7.51
80 plus	574	1.95
Total	29,469	100

Table 3.A3: Summary statistics of GP used to calculate sophisticated marginal effects

	Network characteristics		GP characteristics	
30m	Avg. No. neighbours	11.083	GP size	460
	Density	0.024	Avg. Age	47.750
	Centrality	0.000	Avg. household size	4.974
	Centralisation	0.033	Female household head	0.074
	Avg. No. of Isolates	0.002	OBC	0.230
100m	Avg. No. neighbours	87.461	All Other Castes	0.465
	Density	0.191	General Caste	0.304
	Centrality	0.003	Latrine ownership	0.163
	Centralisation	0.093	Latrine usage	0.167
	Avg. No. of Isolates	0.000	Any OD	0.839
300m	Avg. No. neighbours	305.313		
	Density	0.665		
	Centrality	0.010		
	Centralisation	0.231		
	Avg. No. of Isolates	0.000		

4 Conclusion

In this thesis, the role of social networks in shaping sanitation behaviours in rural India are explored. Three different network structures were examined: the complete network of the entire village, networks based on caste and jati groupings, and networks defined by varying spatial distances. The findings suggest that social networks play a significant role in shaping sanitation behaviours and norms, and that their impact varies depending on the behaviour and network structure being considered.

In Chapter 1, the complete network of the village was used to examine the impact of neighbourhood peer effects on sanitation outcomes. Results show that complementary social effects exist for all three sanitation behaviours studied: latrine ownership, latrine usage, and open defecation. The effects were largest for open defecation.

In Chapter 2, networks based on caste and jati groupings were examined. The findings show that endogenous social effects exist for these groups at similar rates to those found in Chapter 1. However, the magnitude of these effects reduces once correlated factors are controlled for by village and caste (or jati) fixed effects. It was also found that endogenous social effects between caste groups are less than within-group effects, with Forward Caste and Other Backward Class groups exhibiting the highest in-group endogenous social effects, and Denotified Tribes exhibiting the highest other group endogenous social effects.

Chapter 3 compared the impact of three different spatial networks on sanitation behaviours. The findings suggest that the direction and magnitude of endogenous social effects differ depending on the sanitation behaviour and network structure being considered. Specifically, it found that larger networks are associated with increasing positive endogenous social effects for open defecation. However, for latrine ownership and usage, endogenous social effects are only positive at the 30-metre network, becoming negative for the 100-metre and 300-metre networks.

Taken together, the findings provide insights into some of the nuanced ways in which social networks shape sanitation behaviours. The results suggest that the impact of social networks is both positive and negative, depending on the behaviour and network structure being considered. The thesis shows that behavioural spillovers exist in sanitation adoption with proximity, caste, and gender appearing to be especially significant drivers of these social multipliers. These findings have relevant policy implications for interventions aimed at improving sanitation outcomes in rural India, and other rural global south contexts. By understanding the role of social networks in shaping sanitation behaviours, policymakers can develop more effective interventions that are tailored to the social context of the communities they serve.

Community level collective change, targeted along reference lines of socially cohesive groups such as caste and women can be useful leverage points for improving sanitation outcomes. This means that these easier to observe social identifiers (caste, gender, proximity) can also be more cost effect means of galvanising behaviour change than needing to attain all social relations directly from agents.

As important as these social spill-overs are for sanitation behaviour change, they are best leveraged when initiated alongside complementary soft and hard infrastructure in support of sanitation improvements. This means that sustaining sanitation improvements requires both improvements in tangible factors such as access to water sources, but also intangible factors such as sufficient market support for the construction of quality latrines. This will ensure sanitation maintenance is a core part of sanitation adoption.