

UNIVERSITY OF WARWICK

Economics

When the Norm is Wrong: Social Norm Misspecification and Targeting with an Application to Microfinance Adoption

Shantanu Chadha

November 2026
No: 1639

ISSN 2059-4283 (online)
ISSN 0083-7350 (print)

Warwick Economics Research Papers

When the Norm is Wrong: Social Norm Misspecification and Targeting with an Application to Microfinance Adoption

Shantanu Chadha*

November 9, 2026

Abstract

Structural models of peer effects typically capture peer influence through the average, aggregate, or in some cases through extreme peer actions. Such specifications, however, are imposed ex ante and are typically ad hoc. We study the consequences of social norm misspecification by extending the constant elasticity of substitution (CES) norm from continuous to discrete choices. Considering a binary-choice game with incomplete information, we derive results on existence and uniqueness of a rational expectations equilibrium and establish conditions for identification of the structural parameters. We estimate the model by the Nested Pseudo-Likelihood estimator, implemented using the spectral algorithm. Monte Carlo simulations show that the CES estimates of the peer effects approach their true values as the sample grows, whereas imposing a linear norm biases them even in large samples. We apply the framework to microfinance adoption in India and find a strongly max-leaning norm, with significant heterogeneity in curvature and peer effects across villages grouped by caste composition and wealth. In a counterfactual exercise that allocates a fixed subsidy budget to maximise adoption, the estimated and linear norms target very different households, with the former favouring village "leaders", who are more likely to adopt and were the first to be informed about microfinance.

Keywords: Peer Effects, Social Norms, Misspecification, Discrete Choice

*Department of Economics, University of Warwick. E-mail: Shantanu.Chadha@warwick.ac.uk.

1 Introduction

Economic and social interactions are rarely confined to choices of isolated agents. Whether it be firms, households, or individuals, an agent's actions often generate indirect costs or benefits that shape the incentives and behaviours of others. Peer effects are one important class of such externalities, generated when agents interact through a social or economic network that mediates the flow of information, influence or norms that shape their decisions. Substantial empirical evidence supports their importance.[1]

A particularly salient class of problems with peer effects involves discrete choices, such as whether to adopt microfinance (Banerjee et al., 2013), smoke (Nakajima, 2007), buy a new car (Shemesh et al., 2022), take up disaster insurance (Cai et al., 2015) or install solar panels (Bollinger and Gillingham, 2012). At the same time, discrete choice models raise complications that differ from those with continuous outcomes, from existence and uniqueness of an equilibrium to identification and estimation. The policy relevance of these choices nevertheless motivates incorporating peer effects into structural models, enabling counterfactual and welfare analysis. Any model-based prediction, however, depends critically on correctly specifying model primitives. Omitting or misspecifying peer effects where they exist can therefore bias policy recommendations.[2]

Typically, structural models capture peer effects through a benefit from *complementarity* between an agent's own action and the social norm, and a *conformity* cost from deviating from said norm.[3] Two kinds of misspecification can arise in such models. The first concerns the functional form of the utility, usually lacking an axiomatic foundation,[4]

[1] For example, Mas and Moretti (2009) document peer effects in worker productivity and Calvó-Armengol et al. (2009) in educational performance. See Bramoullé et al. (2020) for a recent survey.

[2] For instance, Yakovlev (2018) studies peer influence and habit formation in alcohol consumption, finding that accounting for peer effects makes demand more elastic and more responsive to tax increases.

[3] The most general form, as studied in Liu et al. (2014), Lambotte (2025b) and Boucher et al. (2024), is

$$U_i(y_i, y_{-i}, \mathbf{G}) = \alpha_i y_i + \underbrace{\theta_1 y_i \tilde{y}_{-i}}_{\text{Complementarity}} \underbrace{-\tfrac{1}{2}\theta_2 (y_i - \tilde{y}_{-i})^2}_{\text{Conformity}}.$$

α_i captures the intrinsic utility of choosing action y_i, while $\tilde{y}_{-i}$ represents the local social norm determined by the actions of others (y_{-i}) and the network $\mathbf{G}$. θ_1 and θ_2 capture complementarity and conformity.

[4] Chambers et al. (2023), for instance, provide axioms embedding the linear peer effects model within the Luce framework.

while the second concerns the specification of the social norm itself. This paper focuses on the latter, studying how peer influence is aggregated in discrete choice settings.

Since Manski (1993), most peer effect studies assume that peer influence is captured by the mean or aggregate behaviour of the group (Boucher and Fortin, 2016). Alternative specifications use extreme peer actions—for example, the maximum (Carrell et al., 2010), the minimum (Hahn et al., 2020), or both (Tatsi, 2015),[5] but there is little *a priori* reason to prefer any one. For continuous outcomes, recent work accommodates more flexible norms, with Boucher et al. (2024) adopting a CES aggregator and Tchuente (2026) treating the aggregator itself as the object of identification. The assumption of a linear norm, however, persists and remains ad hoc even in discrete choice models, where the analysis for continuous outcomes does not carry over directly.

The main contribution of this paper is to extend the CES aggregator of Boucher et al. (2024),[6] nesting the mean, maximum, and minimum norms, to discrete choice models with both complementarity and conformity. Consistent with other discrete choice models of peer effects, we define a simultaneous game of incomplete information in which agents form rational expectations over their neighbours' actions and best-respond to those beliefs (Brock and Durlauf, 2001a; Lee et al., 2014). We present results that address the specific theoretical and econometric challenges that discrete choice raises for equilibrium, identification and estimation.

Existence and uniqueness of equilibrium are essential for yielding consistent and empirically meaningful counterfactual predictions without an equilibrium selection rule. Furthermore, multiple equilibria may also hamper identification and estimation. Unlike similar models, our system is neither linear nor, in general, a contraction, and lacks global curvature. Therefore, standard arguments such as those in Boucher et al.

[5] See also Brock and Durlauf (2001b) and Blume et al. (2011) for related peer effect models.

[6] Their social norm is represented by the constant elasticity of substitution (CES) norm given by:

$$\tilde{y}_{-i}(\beta) = \left(\sum_{j=1}^{N} \tilde{g}_{ij} y_j^{\beta} \right)^{\frac{1}{\beta}}. \tag{1}$$

Here, $\tilde{g}_{ij}$ denotes the ij^{th} element of the row-normalised adjacency matrix. When $\beta = 1$, the norm reduces to the average, while extreme values ($\beta \to \pm\infty$) imply that the highest or lowest action in an agent's neighbourhood determines the norm.

(2024); Lambotte (2025b) fail to apply directly. We instead impose supermodularity[7] for existence, and use the order structure of the set of fixed points together with a bound on the elasticity of best responses to the norm for uniqueness. Under this bound, best responses move less than proportionally with the norm, making relative differences in beliefs contract. In this regard, our proof techniques are novel in the peer effects literature, providing an alternative methodology to Boucher et al. (2024).

While models with continuous outcomes face the reflection problem of Manski (1993), since outcomes and the norm are determined simultaneously, the non-linearity of the incomplete-information structure breaks this link. On the other hand, the CES aggregator introduces a countervailing difficulty for global identification. Using the observational equivalence strategy (see Brock and Durlauf, 2007 and Lambotte, 2025b), we show that under rank conditions and variation in neighbours' covariates, discrete or continuous, the CES elasticity, and hence the full parameter vector, is point identified.[8]

Lastly, we jointly estimate the unobserved equilibrium choice probabilities and the parameters as a fixed point of the system using the nested pseudo-likelihood (NPL) estimator of Aguirregabiria and Mira (2007).[9] Despite guarantees of a unique equilibrium for a fixed vector of parameters, the joint fixed point in conditional choice probabilities and parameters need not be unique, nor is convergence to a fixed point guaranteed. Together with the absence of a contraction mapping in our setting, this complicates estimation. To address these issues, we adopt the spectral algorithm of Aguirregabiria and Marcoux (2021), which overcomes challenges with convergence without arbitrary tuning parameters (as in Kasahara and Shimotsu, 2012) and without substantially increasing the computational burden.

We run Monte Carlo simulations to assess the finite-sample properties of the estimator under different CES elasticities and adoption rates. We find that imposing linearity distorts the composition of peer effects and this bias persists as the sample grows. In the largest sample, the restricted model under-states conformity in every design, by up to

[7] A supermodular game is one in which actions are strategic complements; see Topkis (1998).
[8] In the absence of peer effects, the curvature is not identified so the model is only partially identified.
[9] The algorithm has been used in peer effect applications like Lin and Hu (2024) and Lambotte (2025b).

about 50%, while complementarity is underestimated by about 50% when the norm is min-like and overestimated between 27 – 93% when it is max-like. Under the CES norm, the peer effect estimates approach their true values as the sample grows although the elasticity itself is very noisy with nearly flat likelihood in small samples. For extreme elasticities, the estimator cannot pin down the elasticity but still recovers the correct regime and the peer effects for large samples, particularly for the max-like norm.

We apply our framework to the microfinance-adoption data of Banerjee et al. (2013). Leading evaluations of microcredit find only modest average effects (Banerjee et al., 2015a,b), shifting attention towards increasing take-up, and adoption spreads through village networks, so the shape of the norm matters for which households a programme should target. While Banerjee et al. (2013) model the diffusion of information, we, like Lambotte (2025b), structurally model adoption as an equilibrium choice, with peer effects entering households' utility directly. The data reject the linear norm ($p = 0.03$) in favour of a strongly max-leaning one and imposing linearity distorts the peer mechanism. It overstates complementarity more than threefold, with non-overlapping confidence intervals, and understates conformity by about 25%, with overlapping ones. For a household of average degree, the peer effects are about 9% larger under the linear norm.

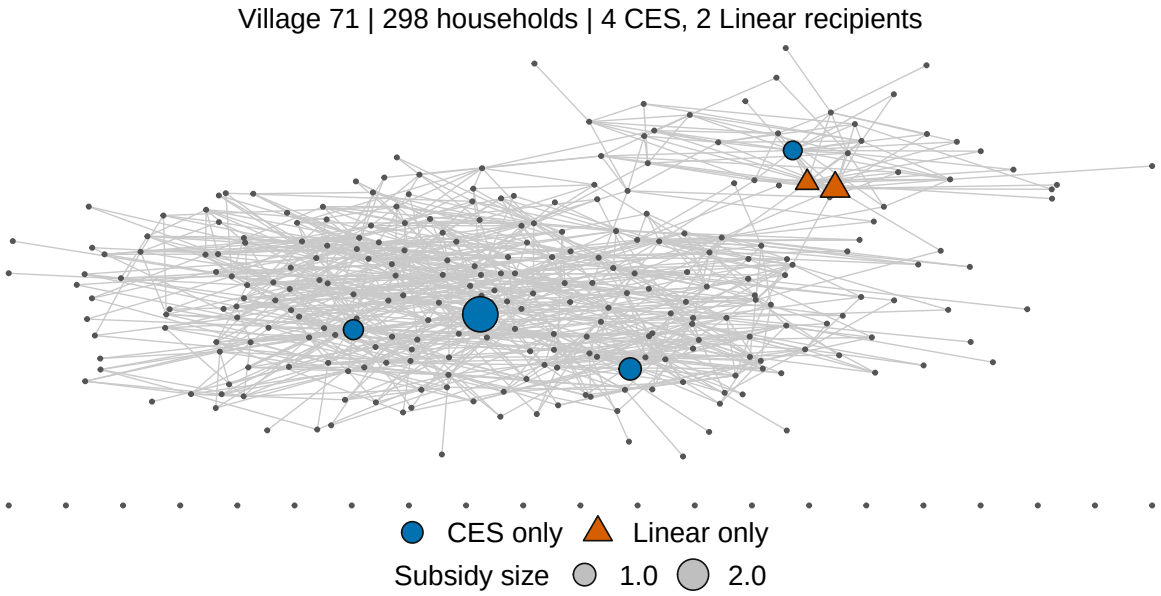
The villages in the dataset are structurally distinct,[10] and imposing a single norm and peer parameters ignores important dimensions of heterogeneity in caste composition and wealth.[11] Partitioning villages by caste (proportion of Scheduled Castes/Tribes) and wealth (bed(s)/latrine ownership), we estimate segment-specific norms and peer effects. Under one partition (bed ownership), the norm lines up with wealth, max-leaning in richer villages and min-leaning in poorer ones. The peer parameters echo this pattern, with poorer villages showing stronger complementarity and weaker conformity. A plausible explanation is that adoption is less risky for richer households, therefore, one informed neighbour suffices, whereas poorer households may require many adopting peers to participate. These patterns, including how conformity varies with caste, are

[10] No cross-village ties exist and the villages differ systematically in caste composition and wealth.

[11] Munshi and Rosenzweig (2006) show that caste networks are relevant for economic interactions in India. The social norm therefore plausibly varies with the caste composition of a neighbourhood, while wealth governs responsiveness to the product.

sensitive to the wealth proxy.

Using the estimated model, we then ask *whom* a planner should target, characterising the allocation of a fixed subsidy budget that maximises aggregate expected adoption. The planner funds the households whose subsidy raises adoption most, directly and through network feedback, and equates these marginal returns across recipients. The marginal return is a generalised Katz–Bonacich centrality in the Jacobian of the norm, similar to Ballester et al. (2006) when the norm is linear but weighted toward each household's most- or least-adopting neighbours as the norm becomes max- or min-leaning. The shape of the norm therefore governs *whom* to subsidise, and by how much.



Notes: Village 71 (298 households), chosen to make the contrast visible. The estimated norm is strongly max-leaning ($\hat{\beta} \approx 491$) and the linear norm imposes $\beta = 1$. Each model uses its own estimates and equilibrium to allocate a common budget of 100 utility-index units. The two norms fund disjoint sets for this village, four households under CES and two under the linear norm, with no overlap, and yield different incremental adoption rates.

Figure 1: Optimal subsidy recipients under the estimated CES and linear norms in an illustrative village.

Comparing the optimal allocation under the estimated versus linear norm, we find that the two fund almost disjoint sets of households, with their rankings negatively correlated (Spearman −0.52). Figure 1 presents this difference clearly in an illustrative village, where the two norms share no common recipients. Moreover, the linear norm spreads the budget thinly, whereas the estimated max-leaning norm concentrates it on

village "Leaders" (62% of its unique recipients, against 12% under the linear norm). These are exactly the households that Banerjee et al. (2013) find central to information diffusion, and that imposing linearity would miss. Misspecifying the norm or ignoring underlying heterogeneity can therefore misdirect the intervention, targeting the wrong households and incorrectly forecasting the changes in adoption it achieves.

The rest of the paper is organised as follows: Section 2 provides a brief review of literature, while Section 3 describes the model and the theoretical results of the underlying game. Section 4 covers the data structure and identification, Section 5 presents the estimation procedure and Monte Carlo evidence for its finite sample performance. Finally, Section 6 presents the empirical results while optimal targeting policy is presented in Section 7 before we conclude in Section 8.

2 Related Literature

The availability of network data and advances in network econometrics have led to a steady increase in both theoretical and empirical models of peer effects (see Bramoullé et al., 2020, for a survey). This work belongs to the strand that models peer effects with discrete choices. Early theoretical models of peer effects include Brock and Durlauf (2001b) and Brock and Durlauf (2001a), who study binary-choice under homogeneous beliefs, extended by Lee et al. (2014) to heterogeneous beliefs that depend on peers' characteristics. More recently, Boucher and Bramoullé (2026) establish coherency and micro-foundations for linear interactions in binary outcomes under incomplete information. Liu et al. (2014), Boucher et al. (2024), and Lambotte (2025b) have explored identification of more general utility functions incorporating conformity and complementarity effects. Volpe (2025) generalises the discrete choice model of Brock and Durlauf (2001a) to let interactions differ across types. Recent work by Houndetoungan (2026) models count outcomes as a game of incomplete information, identified through friends of friends, while Lambotte (2025a) lets conformity in binary outcomes depend on the agent's own action. Our generalisation is orthogonal, wherein we keep a common

interaction structure but generalise the social norm itself for discrete choice models.

The vast majority of the empirical literature relies on a linear social norm (e.g., Calvó-Armengol et al., 2009; Yakovlev, 2018; Guerra and Mohnen, 2022; Lambotte, 2025b), under which an agent's norm is the mean or aggregate behaviour of their peers. Some studies instead use the maximum peer action (Carrell et al., 2010; Tao and Lee, 2014; Díaz et al., 2021; Islam et al., 2021), the minimum (Bietenbeck, 2020; Hahn et al., 2020), or both (Hoxby and Weingarth, 2005; Tatsi, 2015), but these are case-specific choices rather than a general theory (see Brock and Durlauf, 2001b; Blume et al., 2011, for overviews). Houndetoungan (2025) lets payoffs depend on quantiles of peer outcomes, and Herstad et al. (2026) make peer influence rank-dependent; both allow non-monotonic peer influence but concern continuous outcomes. Boucher et al. (2024) propose a non-linear CES social norm, governed by an elasticity parameter, that nests averaging and extreme behaviour for continuous outcomes. We extend the CES social norm to discrete choice peer effect models, combining the two strands of literature. Similar to Boucher et al. (2024) and Lambotte (2025b), we model spillover and conformity effects and give conditions for their separate identification.

Tchuente (2026) treats the peer aggregator, including the CES power mean, as a central primitive. Its identification, estimation and application concern continuous outcomes, although the binary-choice extension (Section 2.6.3 and Appendix A.3) proves uniqueness with Lipschitz bounds that shrink differences in probability levels, which for CES depend on curvature and require probabilities bounded away from zero. We instead control proportional differences through a bound on the best-response elasticity, exploiting the monotonicity and degree-one homogeneity of the CES norm, and develop identification and estimation, which Tchuente (2026) does not. Our contribution therefore is the complete structural and econometric development of the CES norm for binary decisions under incomplete information and application to a relevant setting.

In line with Brock and Durlauf (2001a), our norm is belief-based: agents best-respond to an aggregation of their beliefs about peers' choices, not to the expectation of an aggregate of realised actions. Since utility is belief-dependent, the model belongs to

the literature on psychological games and belief-dependent motivations (surveyed in Battigalli and Dufwenberg, 2022), building on Geanakoplos et al. (1989) and Battigalli and Dufwenberg (2009). The equilibrium is a psychological Nash equilibrium, in which the belief-dependent payoffs are evaluated at correct beliefs, equivalently a self-consistent rational-expectations equilibrium. Imposing this self-consistency while letting beliefs enter utility directly parallels the personal equilibrium of Kőszegi and Rabin (2006) in individual decision theory, and gives the conformity motive a behavioural foundation for which Charness et al. (2019) provide experimental evidence. Consistent with a causal role for perceived norms, Bursztyn et al. (2020) show experimentally that correcting beliefs about others' attitudes shifts economically relevant choices.

Estimation of non-linear models with conditional choice probabilities defined as fixed points of best responses has been studied in (Aguirregabiria and Mira, 2007; Pesendorfer and Schmidt-Dengler, 2008, 2010; Kasahara and Shimotsu, 2012; Aguirregabiria and Marcoux, 2021) among others with previous application to the peer effect models such as (Lin and Hu, 2024; Lambotte, 2025b). We leverage results from (Menzel, 2016; Lin and Hu, 2024) on the asymptotic properties of this estimator in the peer effect setting. However, unlike existing work, we employ the spectral algorithm of Aguirregabiria and Marcoux (2021), which overcomes convergence issues of Aguirregabiria and Mira (2007) without requiring the selection of an arbitrary contraction parameter, as in Kasahara and Shimotsu (2012). To the best of our knowledge, this represents the first application of this algorithm to a peer effects model.

Finally, we contribute to the literature on targeting interventions in networks (Ballester et al., 2006; Galeotti et al., 2020). Whereas existing work studies targeting under local-average or asymmetric conformity (Meng, 2025; Houndetoungan and Lambotte, 2026) or the seeding of information (Banerjee et al., 2019; Beaman et al., 2021), we characterise adoption-maximising subsidies under a non-linear CES norm, showing how its curvature changes whom to target. We find that the CES-based allocations disproportionately target village leaders, the households informed first in Banerjee et al. (2013), while imposing a linear norm changes the recommended recipients.

3 Binary-Choice Peer Effects Model with CES Social Norm

3.1 Setting

Let there be N individuals who interact over a known and fixed network $\mathbf{G}(V, E)$, where V is the set of nodes and E the set of edges. The associated $N \times N$ adjacency matrix is $\mathbf{G} = [g_{ij}]$. We assume the network is undirected ($g_{ij} = g_{ji}$) and unweighted, with $g_{ij} \in \{0, 1\}$ indicating whether agents i and j are connected.[12] For each agent i, let $N_i = \{j \mid g_{ij} = 1\}$ denote the set of neighbours, excluding self-loops ($i \notin N_i$). The degree of node i (number of connections) is given by $d_i = |N_i|$, the cardinality of N_i. We also define the row-normalised adjacency matrix $\tilde{\mathbf{G}} = [\tilde{g}_{ij}]$, where

$$\tilde{g}_{ij} = \begin{cases} \dfrac{g_{ij}}{d_i}, & d_i > 0, \\ 0, & d_i = 0, \end{cases}$$

and the diagonal degree matrix $\mathbf{D}$ with entries $D_{ii} = d_i$.

We consider a simultaneous-move game with incomplete information. Each agent knows their own type but does not observe the types of other players, including their neighbours. This setup follows the standard discrete choice literature on network games (e.g., Brock and Durlauf, 2001a; Lee et al., 2014).

The action space of each agent is binary, i.e. $Y_i := \{0, 1\}$. Here, $y_i = 1$ corresponds to the "high" action (example, adopting a new technology), while $y_i = 0$ corresponds to the "low" action (example, non-adoption). An outcome of the game is a strategy profile $y = (y_1, \ldots, y_N)' \in Y = \times_{i=1}^{N} Y_i$, with y_{-i} denoting the actions of all players except agent i. The utility function, in line with Lambotte (2025b) and Boucher et al. (2024), incorporates two distinct social interaction effects: (i) *complementarity* capturing strategic spillovers, and (ii) *conformity* capturing the pressure to match peers' actions.

[12] Our framework can be extended to directed and weighted networks.

Formally,[13]

$$U_i(y_i, y_{-i}, \mathbf{G}) \;=\; \alpha_i\, y_i \;+\; \underbrace{\theta_1 d_i\, y_i\, \tilde{y}_{-i}}_{\text{Complementarity}} \;-\; \frac{1}{2}\underbrace{\left[\theta_2\left(y_i - \tilde{y}_{-i}\right)^2\right]}_{\text{Conformity}} + \epsilon_i(y_i), \tag{2}$$

where $\epsilon_i(y_i)$ is an idiosyncratic preference shock.

For isolated nodes (no neighbours, i.e. $d_i = 0$), the utility reduces to

$$U_i(y_i, y_{-i}, \mathbf{G}) \;=\; \alpha_i\, y_i + \epsilon_i(y_i). \tag{3}$$

Here α_i captures the intrinsic utility of agent i from selecting the high action ($y_i = 1$).[14] Peer influence is captured by the *social norm function*, which we define as the CES norm of Boucher et al. (2024), allowing for flexible aggregation of peer behaviours:

$$\tilde{y}_{-i}(\beta) \;=\; \left(\sum_{j=1}^{N} \tilde{g}_{ij}\, y_j^{\beta}\right)^{\frac{1}{\beta}}. \tag{4}$$

When $\beta = 1$, it reduces to the usual linear norm and captures the proportion of peers that adopt the high action. For $\beta < 0$, it is defined as:

$$\tilde{y}_{-i}(\beta) = \begin{cases} \left(\sum\limits_{j=1}^{N} \widetilde{g}_{ij} y_j^{\beta}\right)^{\frac{1}{\beta}}, & \text{if } y_j > 0 \;\; \forall j \in N_i, \\ 0, & \text{if } y_j = 0 \;\; \text{for some } j \in N_i. \end{cases} \tag{5}$$

Here $\epsilon_i(y_i)$ represents the private information of agent i, which is observed only by them and not by other players (or by the econometrician). We assume that the distribution of types is common knowledge, i.i.d. across players, and that the network $\mathbf{G}$ is exogenous.[15]

[13] Unlike the continuous case of Boucher et al. (2024), separate identification of the two effects in discrete choice models relies on scaling the complementarity term by d_i. This normalisation follows Lambotte (2025b). In the continuous case, isolated individuals can help identify them (see Boucher et al. (2024)), but this strategy fails in discrete models due to scale and location normalisations for the distribution of shocks.

[14] This intrinsic term depends on observable individual characteristics, group-level factors, and correlated effects; we will specify the exact functional form when turning to the empirical model.

[15] Unlike Lambotte (2025b), we treat the network as exogenous, since our focus is on the consequences

Assumption 1 (A1)**.** We assume that $\epsilon(y_i = 0)$ and $\epsilon(y_i = 1)$ are i.i.d. Gumbel random variables with cumulative distribution function $G(x) = \exp\big(-\exp(-x)\big)$. Moreover, we assume that $\mathbf{G}$ and α are independent of the idiosyncratic shocks $\epsilon(y_i)$ $\forall i$.

The public information of the game therefore, is the vector of intrinsic utilities $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_N)$ together with the network $\mathbf{G}$, i.e. $\overline{\mathcal{I}} = \{\boldsymbol{\alpha}, \mathbf{G}\}$. The private information set of agent i additionally includes their own preference shocks, $\mathcal{I}_i = \{\boldsymbol{\alpha}, \mathbf{G}, \epsilon_i(y_i)\}$. In line with the literature on discrete choice games with incomplete information, we assume that players form rational expectations about the actions of others and maximise utility given these beliefs. This implies that agents base their decisions on their beliefs about others' actions instead of the realised actions themselves. Furthermore, we impose the consistency requirements on these beliefs as in Brock and Durlauf (2001a); Lee et al. (2014); Lambotte (2025b) that ensure that these rational expectations of agents are in fact correct in equilibrium.

Assumption 2 (A2)**.** Each agent forms rational expectations about others' actions conditional on their private information. That is, agent i's belief about the probability that agent j chooses the high action is denoted by $p_j(\mathcal{I}_i) \in [0, 1]$.

Given Assumption 1, the shocks $\epsilon_i(y_i)$ are independent of both α_i and the network $\mathbf{G}$. Hence, agent i's belief about player j's action depends only on the public information set, i.e. $p_j(\mathcal{I}_i) = p_j(\overline{\mathcal{I}})$. It follows that all players hold the same belief about player j's action conditional on public information, and we denote this common expectation by p_j. We further assume that the norm itself is evaluated at these beliefs.

Assumption 3 (A3)**.** Let p_j be the common belief about j's choice, then the *expected social norm* for agent i is

$$\tilde{y}^e_{-i}(\beta) := \left(\sum_{j=1}^{N} \tilde{g}_{ij}\, p_j^{\beta} \right)^{1/\beta},$$

and agent i maximises the belief-dependent utility,

$$U_i^e\big(y_i, \mathbf{p}_{-i}, \mathbf{G}\big) \;=\; \alpha_i\, y_i \;+\; \theta_1\, d_i\, y_i\, \tilde{y}^e_{-i}(\beta) \;-\; \tfrac{1}{2}\theta_2 \big(y_i - \tilde{y}^e_{-i}(\beta)\big)^2 \;+\; \epsilon_i(y_i), \tag{6}$$

of misspecifying the norm on peer effect estimates and counterfactuals.

where $\mathbf{p}_{-i} = (p_j)_{j \neq i}$ is the vector of common beliefs about the other players.

Remark. Evaluating the norm at beliefs is distinct from taking the expectation of the norm of realised actions. For $\beta \neq 1$, $\tilde{y}^e_{-i}(\beta) \neq \mathbb{E}[\,\tilde{y}_{-i}(\beta)]$ (Jensen's inequality), and the two coincide only in the linear case $\beta = 1$. Utility therefore depends directly on beliefs about others' actions.[16] Appendix D presents a more formal discussion on this assumption.

The parameter β therefore governs how much weight the norm places on the peers most and least likely to adopt. The linear and the extreme norms then become special cases depending on the value of β:

$$\underbrace{\min_{j \in N_i} p_j}_{\beta \to -\infty} \;\leq\; \underbrace{\prod_j p_j^{\tilde{g}_{ij}}}_{\beta \to 0} \;\leq\; \underbrace{\sum_j \tilde{g}_{ij}\, p_j}_{\beta = 1} \;\leq\; \underbrace{\max_{j \in N_i} p_j}_{\beta \to +\infty}.$$

At $\beta = 1$ the norm is the believed share of adopting peers. As $\beta \to +\infty$ the peer most likely to adopt sets the norm, and a single likely adopter can sway the neighbourhood. As $\beta \to -\infty$ the norm is high only when all peers are likely to adopt, as when collective agreement is required. Estimating β therefore avoids choosing an aggregator ex ante.[17]

Furthermore, to separate the two peer effect mechanisms, we scale the complementarity term by each agent's degree. At $\beta = 1$, $d_i \tilde{y}^e_{-i} = \sum_j g_{ij} p_j$ is the expected number of i's neighbours who adopt, and in general $d_i \tilde{y}^e_{-i}(\beta)$ is the effective number of peers expected to adopt. Complementarity θ_1 is therefore the payoff from each expected adopting peer, whatever i's degree (Lambotte, 2025b).

Finally, we assume that the maximum degree remains bounded, which controls the degree distribution as the network size grows.

[16] Charness et al. (2019) provide experimental evidence of agents responding to their beliefs about peer behaviour with conformity effects. This places the game among models with belief-dependent payoffs, namely psychological games (Geanakoplos et al., 1989; Battigalli and Dufwenberg, 2009) (see Battigalli and Dufwenberg, 2022, for a survey) and reference-dependent preferences (Kőszegi and Rabin, 2006), whose equilibrium requires those beliefs to be correct.

[17] Evaluating the norm at beliefs is what keeps β meaningful. For realised binary actions, $y_j^\beta = y_j$, so the CES of realised actions is just a power of the adoption share, and for $\beta \leq 0$, it only records whether every peer adopts or not. Because beliefs lie strictly in $(0, 1)$, p_j^β varies with β, and different curvatures generate distinguishable norms.

Assumption 4 (A4)**.** The maximum degree of the network is bounded, i.e. there exists a $\bar{d} < \infty$, independent of N, such that $d_{\max} \leq \bar{d}$, where $d_{\max} = \max_{1 \leq i \leq N} d_i$.

Because the utility function includes the complementarity term which is scaled by the degree, bounding the degree prevents the spillover term from scaling with N. Similar restrictions are common in the literature with Lin and Hu (2024); Lambotte (2025b) imposing bounded degree for existence and uniqueness, while de Paula et al. (2018) use it for identification of preferences.

3.2 Best Responses

For the remainder of this section we assume that there are no isolated individuals in the network. We relax this in the empirical analysis. Under Assumptions 2 and 3, agents maximise their belief-dependent utility in (6) by choosing the action that yields higher utility. In particular, agent i chooses $y_i = 1$ if and only if

$$\Delta U_i^e \equiv U_i^e(1, \mathbf{p}_{-i}, \mathbf{G}) - U_i^e(0, \mathbf{p}_{-i}, \mathbf{G}) \geq 0.$$

Solving further gives us,

$$\begin{aligned}
\Delta U_i^e &= \alpha_i + \theta_1 d_i \tilde{y}_{-i}^e(\beta) - \frac{1}{2}[\theta_2(1 + \tilde{y}_{-i}^e(\beta)^2 - 2\tilde{y}_{-i}^e(\beta))] + \epsilon_i(1) \\
&\quad - \left[-\frac{1}{2}\left[\theta_2 \tilde{y}_{-i}^e(\beta)^2\right] + \epsilon_i(0)\right] \geq 0, \\
&= \alpha_i + \theta_1 d_i \tilde{y}_{-i}^e(\beta) + \theta_2(\tilde{y}_{-i}^e(\beta) - \frac{1}{2}) - (\epsilon_i(0) - \epsilon_i(1)) \geq 0, \\
&\iff z_i(\mathbf{p}, \boldsymbol{\theta}) \geq \eta_i. \qquad (7)
\end{aligned}$$

We define $\eta_i \equiv \epsilon_i(0) - \epsilon_i(1)$, which by Assumption 1 follows a logistic distribution with mean 0, scale 1, and cumulative distribution function $F_\eta(x)$. Since η_i is independent of the public information by Assumption 1, i's conditional choice probability is given by:

$$p_i \equiv \Pr(y_i = 1 \mid \boldsymbol{\theta}, \overline{\mathcal{I}}) = \Pr(\Delta U_i^e \geq 0) = \Pr\left(z_i(\mathbf{p}, \boldsymbol{\theta}) \geq \eta_i\right) = F_\eta\big(z_i(\mathbf{p}, \boldsymbol{\theta})\big). \qquad (8)$$

3.3 Existence and Uniqueness

Let $\Gamma_i(\mathbf{p};\boldsymbol{\theta}) = F_\eta\big(z_i(\mathbf{p},\boldsymbol{\theta})\big)$, with $\Gamma_i : [0,1]^N \to [0,1]$, denote the best-response mapping. A rational expectations equilibrium of the game is defined as

Definition 3.1 (Rational Expectations Equilibrium)**.** A Rational Expectations Equilibrium is a vector of beliefs $\mathbf{p}^* \in [0,1]^N$ such that, given these beliefs, each agent i chooses $y_i = 1$ if and only if $\Delta U_i^e \geq 0$, and the beliefs are correct: $p_j^* = \Pr(y_j = 1 \mid \boldsymbol{\theta}, \overline{\mathcal{I}})$ for all j.

Stacking across agents, an equilibrium is equivalently a fixed point of the system

$$\mathbf{p}^* = \Gamma(\mathbf{p}^*;\boldsymbol{\theta}) = \mathbf{F}_\eta\big(Z(\mathbf{p}^*)\big), \tag{9}$$

where $Z(\mathbf{p}^*) = \boldsymbol{\alpha} + \theta_1 \mathbf{D}\,\tilde{\mathbf{y}}^e(\beta) + \theta_2\big(\tilde{\mathbf{y}}^e(\beta) - \frac{1}{2}\iota\big)$ and $\tilde{\mathbf{y}}^e(\beta) = \big(\tilde{y}^e_{-1}(\beta), \ldots, \tilde{y}^e_{-N}(\beta)\big)'$ is the vector of stacked norms.

There is no guarantee that a rational expectations equilibrium exists or if it is unique. Both properties are essential for coherent empirical analysis. Moreover, multiple equilibria complicate identification and estimation, and make counterfactual predictions unreliable without an equilibrium selection mechanism.

Compared to standard problems, the mapping of conditional choice probabilities given by equation (9) is neither linear nor a contraction. Depending on β, the derivative of the social-norm aggregator $\tilde{y}^e_{-i}(\beta)$ with respect to p_j can be arbitrarily large, ruling out contraction mapping arguments commonly employed in such models. Instead, under Assumption 5 below, the best-response mapping, $\Gamma(\mathbf{p};\boldsymbol{\theta})$, is monotone in $\mathbf{p}$. We then invoke Tarski's fixed-point theorem, standard in supermodular games, to establish the existence of a rational expectations equilibrium.[18]

Assumption 5 (A5)**.** Let $d_{\min} = \min_{i:d_i>0} d_i$ and $d_{\max} = \max_{i:d_i>0} d_i$ denote the minimum and maximum degrees of non-isolated individuals in the network. We assume that

$$\min_{d\in[d_{\min},d_{\max}]} \{\theta_1 d + \theta_2\} > 0.$$

[18]We can get existence using Brouwer's fixed-point theorem which only requires continuity. However, Tarski's fixed-point theorem provides additional lattice structure that we exploit for uniqueness.

Remark. Assumption 5 requires peer effects to be *strategic complements*, but it allows either mechanism to be negative. This is needed for the best-response mapping to be monotone, which places the model within the class of supermodular games, a structure commonly assumed in the literature (see, for example, Boucher et al. (2024)). The assumption is natural in many socio-economic environments where individuals benefit from coordination, imitation, or conformity. Examples include adopting a new technology, consuming the same product brand, or choosing high effort in a task. At the same time, Assumption 5 rules out settings with *strategic substitutability*, such as competition for scarce resources or congestion effects.[19]

Theorem 1 below presents results on the existence and on the structure of fixed points of the system of conditional choice probabilities.

Theorem 1 (Existence of Rational Expectations Equilibrium)**.** *Let* $\mathbf{p} = (p_1, \ldots, p_N) \in [0,1]^N$ *be the vector of individual choice probabilities and*

$$\Gamma_i(\mathbf{p}; \boldsymbol{\theta}) = F_{\eta_i}\left(\alpha_i + (\theta_1 d_i + \theta_2)\tilde{y}^e_{-i}(\beta) - \frac{\theta_2}{2}\right).$$

Under Assumptions 1-5, there exists a Rational Expectations equilibrium $\mathbf{p}^*$ *of the generalised peer effects model with the CES social norm such that:*

$$\mathbf{p}^* = \Gamma(\mathbf{p}^*; \boldsymbol{\theta}) = \mathbf{F}_\eta(Z(\mathbf{p}^*)).$$

Furthermore, the set of fixed points is a complete lattice, with a least element $\underline{\mathbf{p}}^*$ *and a greatest element* $\overline{\mathbf{p}}^*$*, such that* $\underline{\mathbf{p}}^* \leq \mathbf{p}^* \leq \overline{\mathbf{p}}^*$ *for every fixed point* $\mathbf{p}^*$*.*

Proof. Proof in Appendix Section A.1. □

Bajari et al. (2010, 2015), among others, handle multiple equilibria in discrete choice games by computing all of them and estimating an equilibrium selection mechanism. However, this becomes infeasible as the number of players and actions grows, since

[19] Volpe (2025) is a recent paper that studies heterogeneous peer effects where they allow for strategic substitutability and negative interactions showing that it can yield a unique but unstable equilibrium, or multiple equilibria.

the parameter count rises faster than the available independent conditional choice probabilities. Identification and estimation then demand stronger restrictions.

The fixed point mapping in our model is neither a contraction,[20] nor globally convex or concave.[21] Instead, we use the lattice structure of the fixed points to establish uniqueness and provide an alternative proof to Boucher et al. (2024). A common sufficient condition for uniqueness in peer-effect models, both continuous (Liu et al., 2014; Boucher et al., 2024) and discrete choice (Brock and Durlauf, 2001a; Lee et al., 2014; Lambotte, 2025b), is that peer effects be sufficiently small. We adopt a related condition, where instead of bounding the peer-effect parameters directly, we bound the elasticity of the best response with respect to the social norm over the range of norm values, which guarantees that the set of fixed points is a singleton.

We define the social norm mapping $\tilde{\mathbf{y}}^e(\cdot;\beta) : [0,1]^N \to [0,1]^N$, that takes $\mathbf{p}$ and returns the social norm for each agent at fixed β. Writing the best response mapping directly as a function of the norm,

$$\Gamma_i(\mathbf{p};\boldsymbol{\theta}) = \Gamma_i(\tilde{y}^e_{-i};\boldsymbol{\theta}) = F_\eta\Big(\alpha_i + (\theta_1 d_i + \theta_2)\,\tilde{y}^e_{-i} - \tfrac{\theta_2}{2}\Big),$$

where $\tilde{y}^e_{-i} \equiv \tilde{y}^e_{-i}(\mathbf{p};\beta)$, we define its elasticity with respect to i's norm as,

$$\varepsilon_i^{BR} := \frac{\partial \log \Gamma_i(\tilde{y}^e_{-i};\boldsymbol{\theta})}{\partial \log \tilde{y}^e_{-i}}. \tag{10}$$

For the logistic CDF, this elasticity is

$$\varepsilon_i^{BR} = (\theta_1 d_i + \theta_2)\,\tilde{y}^e_{-i}\Big[1 - F_\eta\big(\alpha_i - \tfrac{\theta_2}{2} + (\theta_1 d_i + \theta_2)\,\tilde{y}^e_{-i}\big)\Big]. \tag{11}$$

[20] It fails to be a contraction because its Jacobian,

$$\frac{\partial \Gamma_i}{\partial p_j} = f_\eta(z_i)\,(\theta_1 d_i + \theta_2)\,\frac{\partial \tilde{y}^e_{-i}(\beta)}{\partial p_j}, \qquad \frac{\partial \tilde{y}^e_{-i}(\beta)}{\partial p_j} = \tilde{g}_{ij}\left(\frac{\tilde{y}^e_{-i}(\beta)}{p_j}\right)^{1-\beta},$$

need not have norm below one: its magnitude scales with the strength of peer effects $(\theta_1 d_i + \theta_2)$, which may be large, and the gradient of the CES aggregator diverges as a peer's belief $p_j \to 0$ whenever $0 \le \beta < 1$. Hence no uniform Lipschitz constant below one exists.

[21] Our mapping is contingent on the cumulative distribution function F_η which is neither globally concave nor globally convex. This is true for most empirically relevant CDFs.

Since Theorem 1 delivers a least and a greatest fixed point, every equilibrium lies between them, and equilibrium norms lie in a compact sub-interval of $(0,1)$.[22] Because Γ is increasing under Assumption 5 and continuous, iterating the best responses from $\mathbf{0}$ and from $\mathbf{1}$ bounds every equilibrium, $\Gamma^t(\mathbf{0}) \le \mathbf{p}^* \le \Gamma^t(\mathbf{1})$ for every $t \ge 1$, and the two sequences converge to the least and greatest fixed points as $t \to \infty$ (Topkis, 1998). A single iteration already gives an interval $[a,b]$ that contains every equilibrium norm and does not depend on β, with $a = \min_{j:\, d_j>0} F_\eta(\alpha_j - \frac{\theta_2}{2})$ and $b = \max_{j:\, d_j>0} F_\eta(\alpha_j + \theta_1 d_j + \frac{\theta_2}{2})$. Assumption 6 bounds the best-response elasticity (11) uniformly over such an interval.

Assumption 6 (A6)**.** There exists a closed interval $[a,b] \subset (0,1)$ such that

$$\left\{ \tilde{y}^e_{-i}(\mathbf{p}^*;\beta) \;:\; 1 \le i \le N,\;\; \mathbf{p}^* = \Gamma(\mathbf{p}^*;\boldsymbol{\theta}) \right\} \;\subseteq\; [a,b],$$

and the best-response elasticity is uniformly below one on it,

$$\bar{\varepsilon}^{BR} := \max_{1\le i\le N} \sup_{\tilde{y}^e_{-i}\in[a,b]} \varepsilon_i^{BR} < 1.$$

Remark. Assumption 6 requires each agent's best response to move less than proportionally to a proportional change in the expected social norm. It makes the map a contraction in log coordinates, equivalently the Thompson metric, which together with homogeneity of degree one of the CES norm gives uniqueness. Peer effects may still be large in levels but cannot generate strong enough proportional feedback to sustain two distinct fixed points. See Appendix A.2 for a discussion.

By (11), ε_i^{BR} is single-peaked in the norm. It rises until $(\theta_1 d_i + \theta_2)\,\tilde{y}^e_{-i} F_\eta(z_i) = 1$ and falls thereafter with its peak value at e^{-z_i}, with z_i evaluated at the peak.[23] This gives sufficient conditions for uniqueness, two of which recover restrictions used in the literature.

- *Floor-probability bound,* $\min_i(\alpha_i - \frac{\theta_2}{2}) > -2$. The peak value is below one if and

[22] The CES norm is a power mean of the beliefs, so it lies between their smallest and largest values.

[23] Since $F'_\eta = F_\eta(1-F_\eta)$, the derivative of (11) in the norm is $(\theta_1 d_i+\theta_2)[1-F_\eta(z_i)][1-(\theta_1 d_i+\theta_2)\tilde{y}^e_{-i}F_\eta(z_i)]$, whose last factor falls from one and so changes sign at most once. At the peak, $(\theta_1 d_i+\theta_2)\tilde{y}^e_{-i} = 1/F_\eta(z_i) = 1+e^{-z_i}$, so $\varepsilon_i^{BR} = e^{-z_i}$ and $z_i - e^{-z_i} = \alpha_i - \frac{\theta_2}{2} + 1$.

only if $\alpha_i - \frac{\theta_2}{2} > -2$, so under this bound the elasticity is below one at every norm and Assumption 6 holds for every β and every strength of peer effects. It requires each agent to adopt with probability above $1/(1+e^2) \approx 0.12$ when no peer is expected to adopt.

- *Below-one bound,* $\max_i(\theta_1 d_i + \theta_2)\, b < 1$. Since $\varepsilon_i^{BR} < (\theta_1 d_i + \theta_2)\, \tilde{y}^e_{-i} \leq (\theta_1 d_i + \theta_2)\, b$ on $[a, b]$, it holds whenever $\max_i(\theta_1 d_i + \theta_2) < 1$, the discrete analogue of the restrictions imposed by Boucher et al. (2024) for continuous outcomes.

- *Linear-norm bound,* $\max_i(\theta_1 d_i + \theta_2) < 4$. For $\beta \geq 1$, including the linear norm, the CES norm is non-expansive in levels. Therefore, since $\sup_z f_\eta(z) = \frac{1}{4}$, this bound makes Γ a contraction in levels, giving uniqueness. For any β, if also $\Gamma_i(\tilde{y}^e_{-i}; \boldsymbol{\theta}) \geq \tilde{y}^e_{-i}$ on $[a, b]$, the bound gives $\varepsilon_i^{BR} \leq (\theta_1 d_i + \theta_2)/4 < 1$, so Assumption 6 holds. It is exactly the coefficient restriction Lambotte (2025b) use for the linear social norm.

Assumption 6 is weaker than the first two bounds, and than the third when $\Gamma_i(\tilde{y}^e_{-i}; \boldsymbol{\theta}) \geq \tilde{y}^e_{-i}$ on $[a, b]$. We now state our uniqueness result.

Theorem 2 (Uniqueness of Fixed Point)**.** *Under Assumptions 1–6, the Rational Expectations Equilibrium of the generalised peer effects model with the CES social norm given by the system in equation* (9)*,* $\mathbf{p}^* \in [0, 1]^N$ *of* Γ*, is unique.*

Proof. See Appendix Section A.2. □

Remark. Assumptions 5 and 6 are sufficient but not necessary for existence and uniqueness. Uniqueness does not strictly require changes in best responses to be small. Under Assumption 5, uniqueness can also be checked directly by iterating the mapping from $\mathbf{0}$ and from $\mathbf{1}$.[24]

[24]For our application for the pooled sample, the elasticity at the equilibrium norm exceeds one for some very high degree households (also fails for linear norm and condition of Lambotte (2025b)). However, the two iterations nevertheless agree to within 10^{-10}, ensuring equilibrium is unique. It does however fail under the estimated parameters for some segments, once we partition the villages.

4 Identification

We have provided conditions under which there exists a unique equilibrium of the discrete choice model with the non-linear CES social norm. In this section, we provide results that help establish identification of the structural parameters under reasonable restrictions.

4.1 Data Structure

We assume that the intrinsic utility that agents derive from choosing the high action takes an additively separable form:

$$\alpha_i = M_i\xi + X_i\gamma,$$

where X_i collects agent i's observable characteristics. In our application these are household covariates,[25] and they proxy for income and social standing of households. We assume that we observe an exogenous network with row-normalised adjacency matrix $\tilde{\mathbf{G}}$. In general, $\tilde{\mathbf{G}}$ may instead be weighted, letting the strength of interactions vary across pairs, or directed, capturing asymmetric relationships, although we require it to be exogenous.

M_i is the i^{th} row of the selection matrix $\mathbf{M}$ $(N \times m)$, which indicates the subnetwork (out of m disjoint subnetworks) to which an individual belongs. These could be villages, schools or firms, so $M_i\xi$ is a subnetwork fixed effect, with ξ_s capturing correlated effects common to all agents in subnetwork s, such as local infrastructure or institutions.

Thus, the structural utility for agent i can be written as:

$$z_i(\mathbf{p}) = M_i\xi + X_i\gamma + \theta_1 d_i \tilde{y}^e_{-i}(\beta) + \theta_2\left(\tilde{y}^e_{-i}(\beta) - \tfrac{1}{2}\right). \tag{12}$$

[25] We omit contextual effects, the average characteristics $\tilde{G}_i\mathbf{X}$ of i's peers. We assume peers' characteristics affect i's adoption only indirectly, through their influence on peers' own adoption, which already enters i's payoff via the social norm $\tilde{y}^e_{-i}$; they have no separate direct effect. This assumption may not hold in other situations and then identification variation needs to come from higher order neighbourhoods.

We now allow for the possibility of isolated individuals in the network:

$$z_i(\mathbf{p}) = M_i\xi + 1\{d_i = 0\}\delta + X_i\gamma + \theta_1 d_i \tilde{y}^e_{-i}(\beta) + 1\{d_i > 0\}\left(\theta_2\left(\tilde{y}^e_{-i}(\beta) - \frac{1}{2}\right)\right), \tag{13}$$

with $1\{d_i > 0\}$ the indicator equal to one if agent i has positive degree and zero otherwise.

There is one additional fixed effect that we add for isolated agents, δ. For a connected agent, $\theta_2\left(\tilde{y}^e_{-i} - \frac{1}{2}\right) = \theta_2\tilde{y}^e_{-i} - \frac{\theta_2}{2}$. Therefore, the constant $-\frac{\theta_2}{2}$ is present for agents with peers but absent for isolated ones. Without the additional dummy for isolated individuals, the model would assign baseline differences between connected and isolated individuals to θ_2 which is behaviourally incoherent and conflates conformity with being connected. The dummy absorbs the level shift and identifies θ_2 from variation in the social norm among connected agents instead of the connected–isolated gap. We present both specifications, with and without isolated-dummy, in the empirical section.

Let $\mathbf{1}_{d=0}$ be the indicator vector for isolated nodes. In matrix form, the terms in (13) that do not involve the social norm are $\mathbf{M}\xi + \mathbf{1}_{d=0}\delta + \mathbf{X}\gamma$, with regressor matrix $[\,\mathbf{M} \;\; \mathbf{1}_{d=0} \;\; \mathbf{X}\,]$. Here, what is observable is the network, the individual characteristics X_i, the selection matrix $\mathbf{M}$, and the realised choices $\mathbf{y}$. While actual realised actions are observed, we do not observe the conditional choice probabilities $\mathbf{p}$ directly. They are equilibrium objects determined by the parameters, the network and the covariates through the fixed point of (9). Moreover, this implies that the expected social norms $\tilde{\mathbf{y}}^e(\beta)$ are unobserved as well. Similarly, the preference shocks η_i are unobservable to the econometrician for all i and we only know the distribution of the shocks. Our aim is therefore to guarantee that we can uniquely identify $\boldsymbol{\theta} = (\xi, \gamma, \delta, \theta_1, \theta_2, \beta)' \in \boldsymbol{\Theta}$.

4.2 Identification Strategy

In contrast to the continuous outcomes case, which suffers from the reflection problem of Manski (1993) because of simultaneous determination of social and individual outcomes, discrete choice models avoid this issue due to the non-linearity introduced by the cumulative distribution function. Nevertheless, once a CES social norm is incorporated, the

resulting non-linearities make identification non-trivial. We utilise variation in degree to separate the two peer effects and variation in neighbours' covariates to identify β.

While local identification in non-linear models can often be guaranteed by an appropriate rank condition on the Jacobian, global identification is usually more complex (see Rothenberg (1971) and Bowden (1973)). We instead establish global identification using the observational equivalence strategy (Brock and Durlauf, 2007; Lambotte, 2025b; Lee et al., 2014). Under Assumptions 1 and 2, the distribution of $\mathbf{y}$ given $(\mathbf{G}, \mathbf{M}, \mathbf{X})$ is summarised by the conditional choice probabilities $p_i^* = \Pr(y_i = 1 \mid \mathbf{G}, \mathbf{M}, \mathbf{X})$, which are therefore what the data identify. Two parameter vectors are then observationally equivalent if they generate the same conditional choice probabilities at every draw of $(\mathbf{G}, \mathbf{M}, \mathbf{X})$ in the support of the joint distribution. We assume that the admissible parameter space contains $(\theta_1, \theta_2) = (0, 0)$.

We prove two main results. When at least one of the peer effect parameters is non-zero, the model is point identified under either of two alternative sets of conditions on the observed covariates. One relies on a continuous covariate, while the other requires only a covariate taking at least three values. The second result is when both peer effect parameters are zero, β remains unidentified, and we achieve only partial identification.

We first make the following weak assumption on the observables and parameters.

Assumption 7 (A7)**.** We assume that:

1. **Finite utility.** The intrinsic utility $\alpha_i = M_i\xi + 1\{d_i = 0\}\delta + X_i\gamma$ is finite for every individual.

2. **Payoff relevance.** Some component x of X_i has a common coefficient $\gamma_x \neq 0$.

3. **Full rank.** $[\,\mathbf{M} \;\; \mathbf{1}_{d=0} \;\; \mathbf{X}\,]$ has full column rank.

Finite utility ensures that equilibrium probabilities lie strictly in $(0, 1)$, and that the norm is well defined for every β. Unlike typical discrete choice peer effect models, which assume full rank of a matrix that includes regressors built from the equilibrium choice probabilities (e.g., Lambotte, 2025b), we impose no such condition. Instead, payoff relevance ensures that variation in a neighbour's x moves that neighbour's choice

probability and, through it, the norm. The full-rank condition involves only the observed regressors and is standard. For simplicity, we assume that the population contains isolated individuals to keep δ meaningful. This is without loss as otherwise δ is dropped from $\boldsymbol{\theta}$, and $\mathbf{1}_{d=0}$ from the full-rank condition.

We also assume that Assumptions 1-4 hold for all supported configurations and Assumptions 5 and 6 hold at every supported configuration over a common interval, when at least one of θ_1 and θ_2 is non-zero. We now state our main identification result.

Theorem 3 (Identification)**.** *Let x be the payoff-relevant covariate, $\mathbf{X}^{-x}$ the other covariates and, for a set S of individuals, $\mathbf{x}_{-S} = (x_k)_{k \notin S}$. Under Assumptions 1-7:*

(i) ***Point identification with a continuous variable.*** *If $(\theta_1, \theta_2) \neq (0, 0)$ and:*

(a) some i with $d_i \geq 2$ is such that, for fixed $(\mathbf{G}, \mathbf{M}, \mathbf{X}^{-x}, \mathbf{x}_{-N_i})$, the support of $(x_j)_{j \in N_i}$ contains a non-empty open set;

(b) some k with $d_k \neq d_i$ has a neighbour j such that, for fixed $(\mathbf{G}, \mathbf{M}, \mathbf{X}^{-x}, \mathbf{x}_{-j})$, the support of x_j contains at least two points.

Then $\boldsymbol{\theta} = (\xi, \gamma, \delta, \theta_1, \theta_2, \beta)'$ is point identified.

(ii) ***Point identification with a discrete variable.*** *If $(\theta_1, \theta_2) \neq (0, 0)$ and:*

(a) some i with $d_i = 1$ has a neighbour j such that, for fixed $(\mathbf{G}, \mathbf{M}, \mathbf{X}^{-x}, \mathbf{x}_{-j})$, the support of x_j contains at least two points;

(b) $[\mathbf{M} \;\; \mathbf{X}]$, restricted to individuals with $d_i = 1$, has full column rank;

(c) some i with $d_i = 2$ has a neighbour j such that, for fixed $(\mathbf{G}, \mathbf{M}, \mathbf{X}^{-x}, \mathbf{x}_{-j})$, the support of x_j contains at least three points.

Then $\boldsymbol{\theta} = (\xi, \gamma, \delta, \theta_1, \theta_2, \beta)'$ is point identified.

(iii) ***Partial identification with no peer effects.*** *If $\theta_1 = \theta_2 = 0$ and some i has a neighbour j such that, for fixed $(\mathbf{G}, \mathbf{M}, \mathbf{X}^{-x}, \mathbf{x}_{-j})$, the support of x_j contains at least two points, then $(\xi, \gamma, \delta, \theta_1, \theta_2)'$ is point identified, but β is not.*

Proof. Proof in Appendix Section A.3. □

Variation in a neighbour's covariate moves their choice probability, and hence i's norm, while holding i's own characteristics fixed, so the resulting change in i's choice isolates the peer effect. In part (i), a payoff-relevant covariate that varies continuously across the neighbours of an individual with at least two neighbours moves their equilibrium choice probabilities over a non-empty open set, so i's choices trace out the shape of the norm, which identifies β. Neighbour variation at a second degree d then pins down $\theta_1 d + \theta_2$.

Without a continuous covariate, part (ii) relies on individuals with one and two neighbours. At degree one, the norm is independent of β, so variation in the neighbour's covariate identifies $\theta_1 + \theta_2$, and full rank among these individuals identifies γ, and ξ up to the constant $\theta_2/2$. At degree two, β affects the norm only through the spread between the two neighbours' choice probabilities. A neighbour's x taking at least three values changes the spread, which identifies β and both θ_1 and θ_2. In parts (i) and (ii), the full-rank condition in Assumption 7 identifies the remaining own coefficients once the peer parameters are known. As in Lambotte (2025b) and Liu et al. (2014), two different degrees separate θ_1 from θ_2.

Finally, in part (iii), when peer effects are absent, variation in a neighbour's covariate moves their choice probability while holding i's own characteristics fixed, and i's choice does not respond, so no admissible alternative can have peer effects either. The full-rank condition in Assumption 7 then identifies the remaining parameters, while β never affects the choices.

Our application satisfies the sample counterparts of the conditions in part (ii), with the number of rooms as the payoff-relevant discrete covariate. Every village has isolated households (628 in total), households with one neighbour (727) and households with two neighbours (687), and both full-rank conditions hold. The number of rooms takes 11 values at the neighbours of degree-one households, and 10 at the neighbours of degree-two households.

5 Estimation and Simulations

In this section we present an appropriate estimator and present Monte Carlo simulations to illustrate its finite sample properties. We also compare it to the misspecified linear model with $\beta = 1$.

5.1 Nested Pseudo-Likelihood Estimator

Similar to other discrete choice models, we estimate our model parameters using a maximum likelihood–based approach. However, the conditional choice probabilities are not directly observed and are instead generated endogenously within the model, with equilibrium probabilities determined as the unique fixed point of the system in equation (9). This feature makes the *Nested Pseudo-Likelihood Estimation* (NPL) strategy of Aguirregabiria and Mira (2007) particularly well-suited, and in-fact been previously applied to estimate peer effects under rational expectations.[26] The likelihood is pseudo exactly because the conditional choice probabilities are not directly observed but are instead derived.

By Assumptions 1–2, agents' rational expectations depend only on the public information set $\overline{\mathcal{I}}$, so peers' realised choices do not affect i's beliefs. Since the private shocks are also independent across agents, peers' realised choices do not affect y_i, and the individual contributions to the likelihood are independent conditional on $\overline{\mathcal{I}}$. Thus, the conditional log-likelihood is expressed as:

$$\mathcal{L}_N(\boldsymbol{\theta}; \mathbf{p}^*) := \mathcal{L}_N(\boldsymbol{\theta}; \mathbf{y}|\overline{\mathcal{I}}) = \frac{1}{N}\sum_{i=1}^{N}\{y_i \ln p_i^* + (1 - y_i)\ln(1 - p_i^*)\} \tag{14}$$

where $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ is the parameter vector and $p_i^* = \Gamma_i(\mathbf{p}^*; \boldsymbol{\theta})$.

If instead the conditional choice probabilities were observable, we could use the

[26] See, for example, Lambotte (2025b), Lin and Hu (2024), and others.

standard Maximum Likelihood Estimator (MLE):

$$\hat{\boldsymbol{\theta}}_{MLE} = \arg\max_{\boldsymbol{\theta}} \mathcal{L}_N(\boldsymbol{\theta}; \mathbf{p}^*) \quad \text{s.t. } \mathbf{p}^* = \Gamma(\mathbf{p}^*; \boldsymbol{\theta}). \tag{15}$$

where $\Gamma(\mathbf{p}; \boldsymbol{\theta}) = F_\eta(Z(\mathbf{p}))$. Since the conditional choice probabilities are not directly observed, we require an estimator that consistently recovers both the probabilities and the structural parameters. Given the high dimensionality of the public information set $\overline{\mathcal{I}}$, non-parametric estimation of these probabilities is infeasible. The NPL estimator of Aguirregabiria and Mira (2007) can instead help iteratively recover the two objects of interest. The pseudo log-likelihood is thus defined as:

$$Q_N(\boldsymbol{\theta}; \mathbf{p}) = \frac{1}{N} \sum_{i=1}^{N} \Big\{ y_i \ln(\Gamma_i(\mathbf{p}; \boldsymbol{\theta})) + (1 - y_i) \ln(1 - \Gamma_i(\mathbf{p}; \boldsymbol{\theta})) \Big\}, \tag{16}$$

where $\mathbf{p}$ now need not be the equilibrium probabilities. Let $q_i(\boldsymbol{\theta}, \mathbf{p}) = y_i \ln(\Gamma_i(\mathbf{p}; \boldsymbol{\theta})) + (1 - y_i) \ln(1 - \Gamma_i(\mathbf{p}; \boldsymbol{\theta}))$ be agent i's contribution to the pseudo log-likelihood function and let $Q_0(\boldsymbol{\theta}, \mathbf{p}) = \mathbb{E}[q_i(\boldsymbol{\theta}, \mathbf{p})]$. Then define:

$$\tilde{\boldsymbol{\theta}}_0(\mathbf{p}) = \arg\max_{\boldsymbol{\theta}} Q_0(\boldsymbol{\theta}, \mathbf{p}) \text{ and } \Gamma_0(\mathbf{p}) := \Gamma(\mathbf{p}; \tilde{\boldsymbol{\theta}}_0(\mathbf{p}))$$

$$\tilde{\boldsymbol{\theta}}_N(\mathbf{p}) = \arg\max_{\boldsymbol{\theta}} Q_N(\boldsymbol{\theta}, \mathbf{p}) \text{ and } \Gamma_N(\mathbf{p}) := \Gamma(\mathbf{p}; \tilde{\boldsymbol{\theta}}_N(\mathbf{p}))$$

with θ_0 and $\mathbf{p}^*$ the true values of the parameters and the conditional choice probabilities. Furthermore, let

$$\Lambda_{0N} \equiv \{(\boldsymbol{\theta}, \mathbf{p}) : \boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}_0(\mathbf{p}), \mathbf{p} = \Gamma_0(\mathbf{p})\}$$

$$\Lambda_N \equiv \{(\boldsymbol{\theta}, \mathbf{p}) : \boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}_N(\mathbf{p}), \mathbf{p} = \Gamma_N(\mathbf{p})\}$$

For any $\mathbf{p}$, an NPL fixed point is a pair $(\boldsymbol{\theta}, \mathbf{p})$. Λ_{0N} is the set of fixed points in the population and Λ_N is the set of fixed points in the sample. Although the fixed point of the system is unique under our assumptions for a given parameter vector, this does not imply that Λ_N is a singleton. Multiple parameter–probability pairs may satisfy the fixed point conditions, especially since the mapping is not a contraction. Hence, the NPL

estimator is defined as the maximiser of the pseudo log-likelihood over all such fixed points:

$$(\hat{\boldsymbol{\theta}}_{NPL}, \hat{\mathbf{p}}_{NPL}) = \arg \max_{(\boldsymbol{\theta},\mathbf{p})\in\Lambda_N} Q_N(\boldsymbol{\theta}, \mathbf{p}). \tag{17}$$

We define two matrices $V_1(\theta_0)$ and $V_2(\theta_0)$:

$$\mathbb{E}\left[\frac{\partial^2 Q_N(\theta_0,\mathbf{p}^*)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} + \frac{\partial^2 Q_N(\theta_0,\mathbf{p}^*)}{\partial\boldsymbol{\theta}\partial\mathbf{p}'}\cdot\left[I-\left(\frac{\partial\Gamma(\mathbf{p}^*;\theta_0)}{\partial\mathbf{p}}\right)\right]^{-1}\cdot\frac{\partial\Gamma(\mathbf{p}^*;\theta_0)}{\partial\boldsymbol{\theta}'}\middle|\overline{\mathcal{I}}\right] \to_p V_1(\theta_0) \tag{18}$$

$$\mathbb{E}\left[\frac{\partial q_i(\theta_0,\mathbf{p}^*)}{\partial\boldsymbol{\theta}}\frac{\partial q_i(\theta_0,\mathbf{p}^*)}{\partial\boldsymbol{\theta}'}\middle|\overline{\mathcal{I}}\right] \to_p V_2(\theta_0) \tag{19}$$

Under regularity conditions (discussed in Appendix Section B), Lin and Hu (2024), relying on Menzel (2016) show that the NPL estimator as described in equation (17), is consistent and asymptotically normal.

Theorem 1 of Lin and Hu (2024). Under Assumptions 1-8,

$$\hat{\boldsymbol{\theta}}_{NPL} \to_p \theta_0, \qquad \sqrt{N}\left(\hat{\boldsymbol{\theta}}_{NPL} - \theta_0\right) \to_d \mathcal{N}(0, V_{NPL}),$$

and where $V_{NPL} = V_1^{-1}(\theta_0)V_2(\theta_0)V_1^{-1}(\theta_0)^\top$ is the variance of the NPL estimator where $V_1(\theta_0)$ and $V_2(\theta_0)$ are defined in equation (18).

Proof. See Lin and Hu (2024) (Appendix A. Proofs, Proof of Theorem 1, p. 14-15). □

For the actual estimation procedure, we use the spectral algorithm of Aguirregabiria and Marcoux (2021). This algorithm overcomes the non-convergence of the usual iterative estimation of the NPL algorithm of Aguirregabiria and Mira (2007). In addition, it does not require selection of an arbitrary parameter such as in the algorithm of Kasahara and Shimotsu (2012). For details on the spectral algorithm, see Appendix Section B.

In practice, two issues complicate estimating the CES elasticity jointly with the other parameters. First, the likelihood can be nearly flat in β, since the norm converges to its max or min very quickly as $|\beta|$ grows. Second, the gradient with respect to β can be on a different scale from those of the other parameters, such as θ_1 and θ_2, which makes

joint non-linear optimisation ill-conditioned. We address both by exploiting that β is the only parameter entering the utility index non-linearly. We fix β on a grid and, at each value, estimate the remaining parameters using the spectral NPL, recording the profiled pseudo log-likelihood. The grid is dense near zero, with a geometric tail on each side and the exact $\pm\infty$ (max- and min-aggregator) endpoints.

We pick the best grid point for $\beta < 0$ and for $\beta \geq 0$, refine each by a one-dimensional search over β between its neighbouring grid points, re-estimating the other parameters by spectral NPL at every candidate, and keep the point with the higher refined likelihood. Although Theorem 3 identifies β exactly, its numerical recovery becomes fragile as $\beta \to \pm\infty$. When the profiled likelihood is flat, up to a tolerance, out to a $\pm\infty$ limit, the candidate values of β are numerically indistinguishable, without affecting the other estimates. We classify that side as a flat tail. In those cases, we report both the point estimate and whether it exhibits tail behaviour, giving the corresponding max- or min-aggregator fit in the latter case.

5.2 Monte Carlo Simulations

We assess the finite-sample behaviour of the NPL estimator and profile-based algorithm by Monte Carlo simulations, comparing the CES norm to the restricted linear norm ($\beta = 1$). We calibrate the network and structural parameters to match the setting of the empirical application. Networks are generated from the Chung–Lu model (Chung and Lu, 2002) to match the mean degree (5.75) and share of isolated households (9.3%) (Table E.12). We set $\theta_1 = 0.15$ and $\theta_2 = 1$, vary $\beta \in \{-8.3,\ 3.9,\ 22.3\}$, consider a low- and a high-adoption design, and run 500 replications for each $N \in \{500, 1000, 2000, 4000\}$. Appendix C gives the set-up details, and all simulation tables are in Appendix E.

The estimator performs well when β is not extreme, particularly with high adoption. Under the CES norm, $\hat{\theta}_1$ is within 0.02 of its true value at $N = 4000$ in every design, and $\hat{\theta}_2$ is within 0.07 in four of the six (Tables E.1 and E.2). The exceptions are in the low-adoption designs. The restricted linear model, by contrast, stays biased as N grows. At $N = 4000$, it under-estimates complementarity by about 50% when $\beta = -8.3$, and

when $\beta = 22.3$ it over-estimates complementarity by 60 to 95% and under-estimates conformity by 45 to 50%. In most designs its bias and MSE persist, whereas under the CES norm both shrink towards zero (Tables E.3–E.4).

For draws with an interior $\hat{\beta}$, the coverage of $\hat{\theta}_2$ is at or above the nominal level in every design, at 90 to 100% (Tables E.5 and E.6). That of $\hat{\theta}_1$ is lower (70 to 95%). For tail draws, where the profile likelihood is nearly flat, the local Wald standard error understates the uncertainty and coverage collapses. As N grows, the median $\hat{\beta}$ moves towards the truth and the share of tail solutions falls, though slowly for $\beta < 0$, while the tail draws distort the mean (Tables E.1 and E.2). The LR rejection rate rises with N; at $N = 4000$ it is essentially 100% at $\beta = 22.3$ but only 15 to 32% at $\beta = -8.3$. For the extreme values, the estimator recovers the sign of β but not its magnitude, both models under-estimate complementarity when the norm is strongly min-like, and the LR test has power only against the max-like norm (Tables E.7–E.11). See Appendix C for a more detailed discussion of the results.

6 Empirical Results

To analyse the effect of norm misspecification on the distortion of optimal targeting policy, we take the framework developed above to the microfinance-adoption dataset of Banerjee et al. (2013). Their original focus was the diffusion of information about the programme through a small set of key agents in each village and its effect on take-up. In line with our framework, and similar to Lambotte (2025b), we instead take awareness of the programme as given and model the adoption choice itself as a simultaneous game of incomplete information.

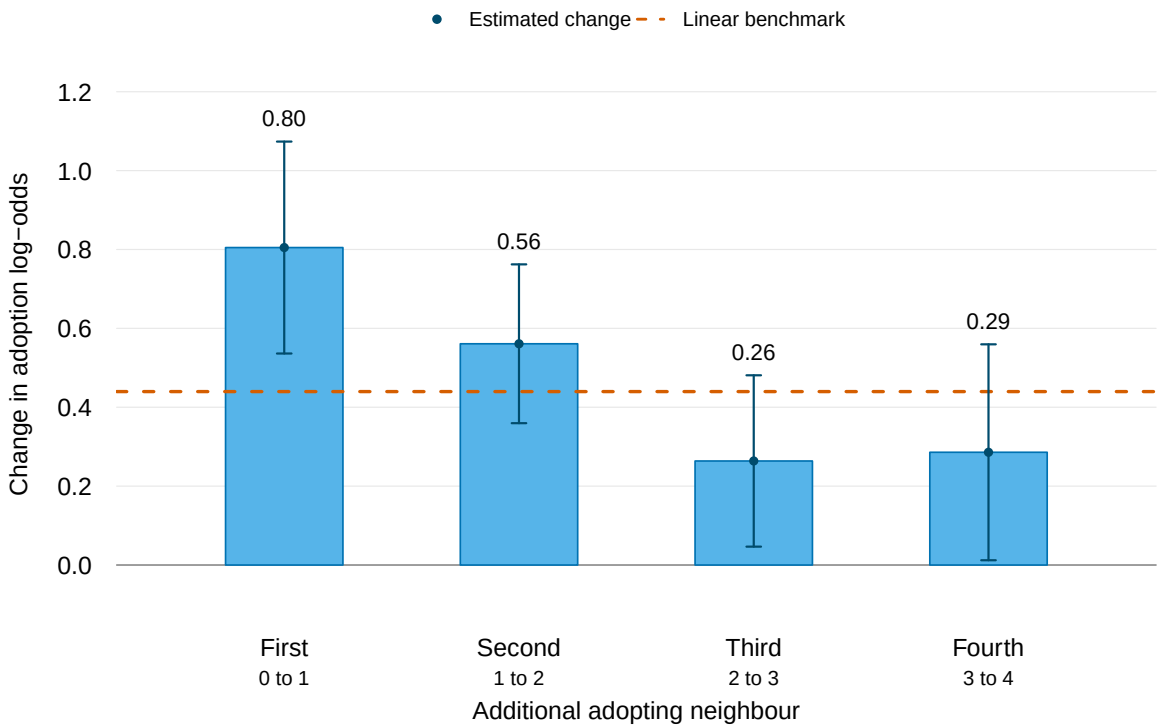
The programme was operated by Bharatha Swamukti Samsthe (BSS), a Bangalore-based microfinance institution that made joint-liability loans, principally to women organised in small borrowing groups. The 75 sample villages span five districts of rural Karnataka, in southern India, and BSS began lending in 43 of them from 2007 onward, entering villages one at a time. We retain the 29 of these 43 villages for which the house-

hold census records each household's caste, which leaves 6,733 households. The action y_i equals one if the household *participates*, as recorded in the lender's administrative records. Take-up was gradual and far from universal, with 17.4% of households in our estimation sample participating.

Crucially, Banerjee et al. (2013) fielded a baseline survey in 2006, *before* BSS entered any village, so the network is measured prior to the adoption decisions it later shapes, therefore is pre-determined. The survey paired a household census with a social-network module covering twelve dimensions of interaction, such as visits, borrowing and lending, and advice. In line with Lambotte (2025b), we use the visiting network, which links two households if either reports visiting the other or being visited by it. It is sparse, with a mean degree of 5.75 and 9.3% of households isolated. The individual characteristics X_i comprise the number of rooms and beds in the home; indicators for a non-Hindu head of household and for scheduled-caste or scheduled-tribe status; and indicators for owning a latrine, for being a designated village "leader" household, and for renting rather than owning the home. Summary statistics are reported in Appendix Table E.12.

Two features of microfinance diffusion make the linear assumption ex-ante implausible. Banerjee et al. (2013) find that, once informed, households barely respond to more neighbours adopting, pointing to a maximum-like norm. The group-liability structure of microfinance, by contrast, requires a minimum group size, suggesting a minimum-like norm. The opposite pull of these two mechanisms implies that the curvature of the social norm is an empirical question instead of an *a priori* assumption. Figure 2 illustrates that the data favour a non-linear norm. The effect of each additional adopting neighbour on adoption declines, whereas a linear norm implies a constant effect.

We first estimate the model on the pooled sample of all 29 villages. Since villages are not inter-connected and differ in their average household characteristics, we then partition them into four segments along two dimensions, caste homogeneity and income, the latter proxied by asset ownership (beds and latrines). Our main specification includes village fixed effects and the isolated-household dummy. Appendix Table E.13 reports estimates without the dummy and relates them to Lambotte (2025b), whose



Notes: 4,123 households with at least four neighbours. Each bar is the estimated change in the log-odds of adoption from one more adopting neighbour, from a logit of adoption on dummies for the number of adopting neighbours and on their share, with village and degree fixed effects and household controls. The dashed line is the constant change implied when the number and share enter linearly, as under a linear norm. The changes decline over the first four adopting neighbours, and a village-clustered Wald test rejects their equality ($p = 0.011$). Whiskers are village-clustered 90% confidence intervals.

Figure 2: Microfinance adoption and additional adopting neighbours.

control function for network endogeneity appears to absorb the same isolated-household heterogeneity.

6.1 Pooled sample

Table 1 reports the pooled estimates with the full estimates presented in Table E.13. The data reject the linear norm in favour of a strongly max-leaning one, with an estimated β of 491.5 and the likelihood-ratio test rejects imposing linearity (LR = 4.63, $p = 0.03$). However, as Figure 3 illustrates, the profiled likelihood rises monotonically towards the maximum aggregator but plateaus (within set tolerance), indicating tail like behaviour where β is numerically indistinguishable and we report the optimal value in that regime. This is in line with the pattern in Figure 2 where the first adopting neighbour does most

of the work.

	CES (β free)			Linear ($\beta = 1$)		LR: $\beta = 1$	
N	β	θ_1	θ_2	θ_1	θ_2	stat.	p
6733	$491.5^{\dagger}$	0.053^{**}	2.064^{***}	0.180^{***}	1.538^{***}	4.63	0.031^{**}
		(0.023)	(0.445)	(0.032)	(0.501)		

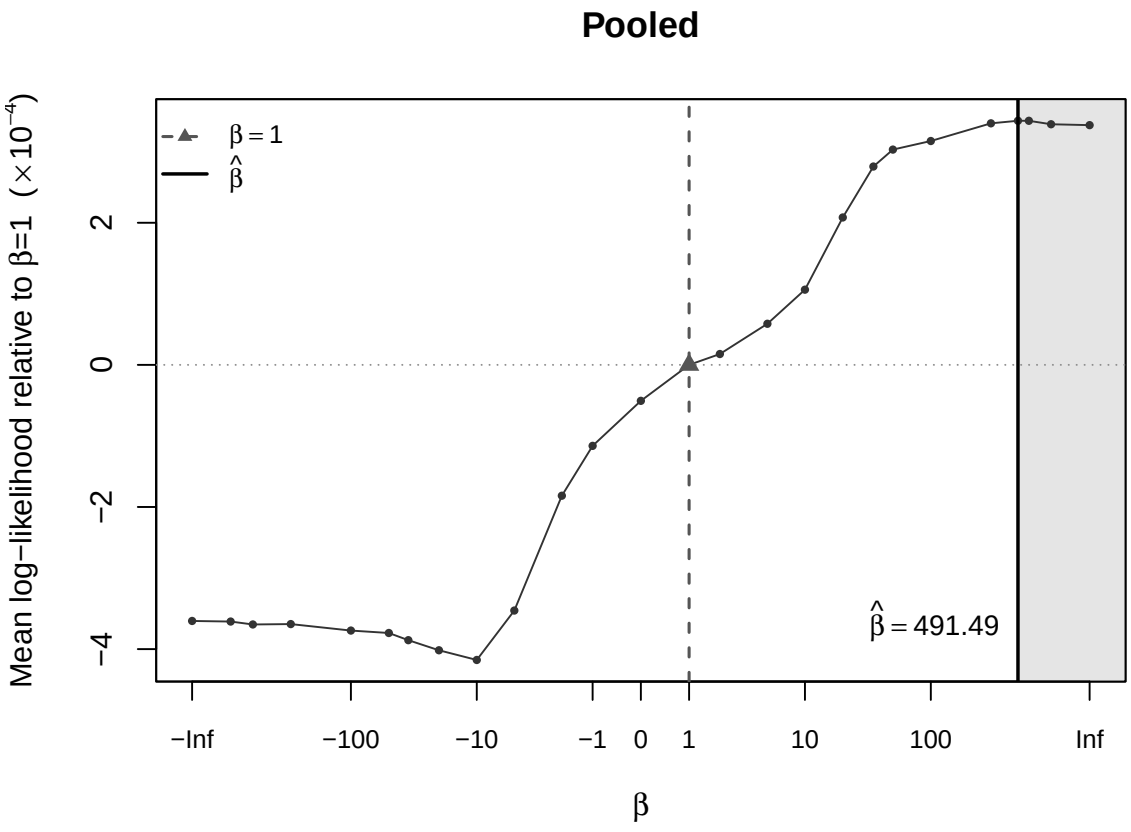
Notes: All 29 villages pooled with village fixed effects and an isolated-household dummy. † marks a flat-tail fit, where β is numerically indistinguishable and the likelihood is flat. The reported value is the refined estimate in the plateau. Standard errors in parentheses, stars are Wald tests against zero. The LR statistic tests $\beta = 1$ against a χ^2_1 reference. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 1: Microfinance Adoption for pooled sample: Point Estimates for CES Elasticity and Peer Effects (CES vs. Linear Norm) and the LR Test of $\beta = 1$.

Furthermore, imposing linearity distorts the two peer channels as anticipated by our simulations. The restricted fit overstates strategic complementarity more than threefold (0.18 vs. 0.05) and understates conformity (1.54 vs. 2.06). A researcher assuming the linear norm would thus attribute adoption far more to spillovers, and less to conformity, than the data support.

Moreover, the dummy on isolated households is significantly different from zero and negative under both the linear and the non-linear norm (−0.894 and −1.033, respectively). Appendix Table E.13 re-estimates the pooled model with and without it, and the exclusion changes the results drastically with conformity turning negative ($\hat{\theta}_2 = -0.725$ under the CES norm and −0.182 under the linear norm) and the profiled β runs to the minimum aggregator. Without a separate intercept, isolated households' lower baseline adoption is mechanically loaded onto the conformity term, which in turn distorts the estimated norm and its curvature β. This mirrors the pattern in Lambotte (2025b),[27] whose homophily-based network-endogeneity control plays the same role as our isolated-household dummy. Comparing their Models 3 and 6 in Table 2, removing the control turns the conformity coefficient from positive to negative while complementarity does not change much, similar to our no-dummy results, for both the linear and the non-linear specifications. Their control function is therefore likely to capture, at least in

[27] They include exogenous peer effects **GX**, which we do not as these effects enter only through feedback and provide identifying information, similar to Boucher et al. (2024). Their inclusion also reduces the information available to identify β precisely.



Notes: The y-axis is the Log-Likelihood of the unrestricted CES model relative to the linear case. The profiled likelihood rises towards the maximum aggregator and then plateaus, with refined estimate $\hat{\beta} = 491.5$, while the linear norm ($\beta = 1$) is dominated. Shaded region denotes the flat tail-region where β is numerically indistinguishable.

Figure 3: Profiled log-likelihood in β for pooled sample.

part, the difference between connected and isolated households rather than network endogeneity.[28]

6.2 Heterogeneity across segments

Two dimensions of heterogeneity that are especially salient in Indian villages can shape peer influence and adoption– social stratification, captured by caste, and wealth. Caste in our data is a household's Scheduled Caste or Tribe (SC/ST) status, and caste-based networks act as channels along which information, norms and risk-sharing flow in

[28]Their control enters as a quartic polynomial in a node position μ estimated from a homophily-based link-formation model with node fixed effects (`homophily.fe` in CDatanet). A household with no links has an unidentified fixed effect that the estimator drives to the lower boundary ($\mu \to -\infty$, numerically a floor), so after the normalisation to $[-1, 1]$ isolated households are pinned at the bottom of the μ range.

rural India. This makes caste homogeneity the natural axis for conformity (Munshi and Rosenzweig, 2006; Banerjee et al., 2013). Wealth, in contrast, shapes both the demand for and the returns to microcredit, and hence take-up (Banerjee et al., 2015a,b), while also conditioning the strategic, complementarity side of adoption through credit constraints and the social-interaction returns to borrowing (Feigenberg et al., 2013). Both dimensions also shape the norm and whose behaviour matters for adoption decisions.

We therefore split villages along these two axes at their village-level medians, giving four segments (heterogeneous/homogeneous × poor/rich). Since there are multiple wealth proxies in the data, we pick two of them, bed and latrine ownership, based on how well they segment the villages and report results for both.[29] The segment composition and the proxies are described in Appendix Tables E.14 and E.15. Here we discuss results for the case with the isolated-dummy included while Appendix Tables E.18 and E.19 present the other specification.

	Caste × Beds				**Caste × Latrine**			
Segment	N	$\hat{\beta}$	LR	p	N	$\hat{\beta}$	LR	p
Het/P	1,616	-83.1†	1.81	0.179	1,691	-0.97	0.67	0.412
Het/R	1,387	491.5†	3.39	0.066*	1,312	-282.8†	3.16	0.075*
Hom/P	1,728	-296.7†	1.56	0.212	1,718	-4.05	5.23	0.022**
Hom/R	2,002	787.7†	0.75	0.387	2,012	2.00	0.35	0.556

Notes: Segments are Het/Hom (heterogeneous vs homogeneous on the caste-minority share) × P/R (poor vs rich), split at village-level medians of each partition's wealth proxy (bed or latrine ownership). † marks a flat-tail fit (β numerically indistinguishable; the value is the refined estimate in the plateau). LR tests $\beta = 1$ against a χ^2_1 reference, with $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 2: Norm elasticity and test of $\beta = 1$ by segment for the Caste × Beds and Caste × Latrine partitions.

Table 2 reports the β estimates by segment for both partitions, with the LR statistic and p-value for $H_0 : \beta = 1$. Only 3 out of 8 segments reject the $\beta = 1$ restriction, likely due to the smaller sample size. The norm is markedly heterogeneous, where for the Caste × Beds case every segment sits in the flat-tail regime, whereas under Caste × Latrine three of the four segments yield interior $\hat{\beta}$ (Figures E.1 and E.2 show the likelihood

[29] An alternative would be to construct a composite wealth index from all the income-related covariates.

profiles).[30] The norm regime lines up with wealth with poorer segments leaning towards the minimum ($\hat{\beta} < 1$) in both partitions and the richer ones towards the maximum, the one exception being the heterogeneous–rich latrine segment, which is min-like.

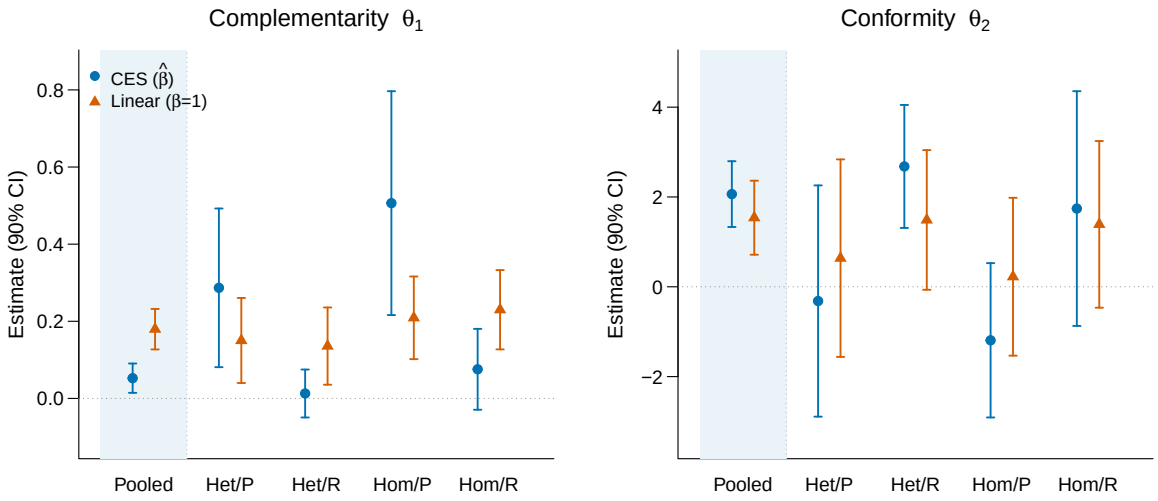


Figure 4: Peer parameters by segment, Caste × Beds: CES vs. linear, with 90% CIs and the pooled estimates for reference.

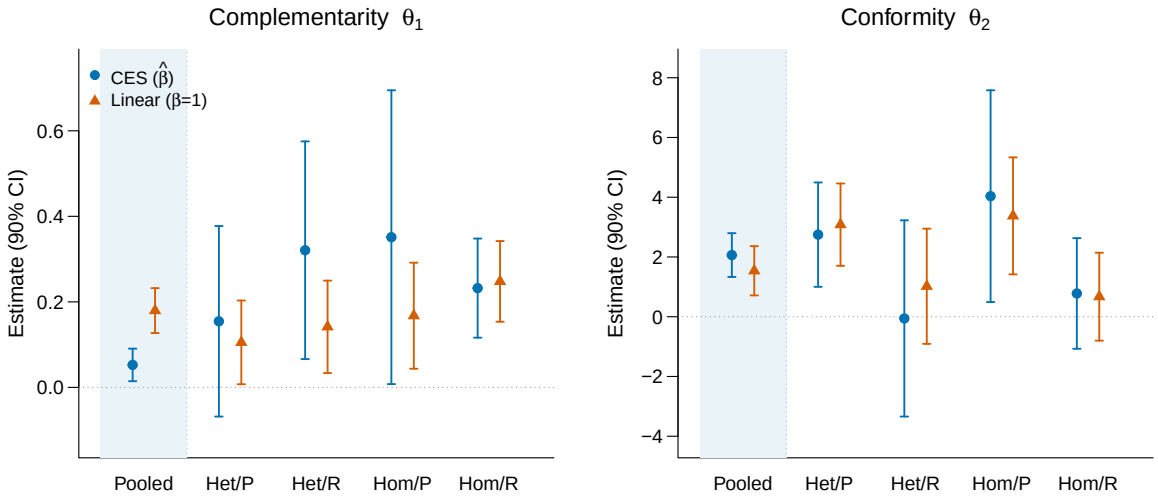


Figure 5: Peer parameters by segment, Caste × Latrine: CES vs. linear, with 90% CIs and the pooled estimates for reference.

Figures 4–5 plot the peer parameters against the pooled estimates and the linear restriction. As in the pooled sample, imposing linearity distorts the peer-effect estimates

[30] In several segments the estimated peer parameters are large enough that our sufficient condition for equilibrium uniqueness (the best-response elasticity bound) is likely violated, under both the CES and the linear ($\beta = 1$) specifications, thus multiplicity of equilibria likely exists.

in several segments, even flipping sign in some segments, although the segment estimates are noisier given the smaller sample size. Relative to the pooled fit, the segments are highly heterogeneous with complementarity generally stronger, and conformity is far more dispersed and even negative. Under the beds split, poorer villages show stronger complementarity and weaker conformity while for the latrine split the conformity ordering reverses, with the poorer segments showing the strongest conformity. Among poorer villages, caste-homogeneous villages exhibit stronger conformity than heterogeneous ones holding wealth fixed, although this pattern does not survive the bed-ownership split. The results illustrate that there is indeed heterogeneity in how peer influence shapes adoption and how we define its dimensions matters. These are patterns that the pooled sample would mask both in terms of the norm as well as the peer effect parameters. Appendix Tables E.16 and E.17 report the full estimates.

7 Optimal Targeting

The previous section illustrated the impact of norm misspecification on the peer effect estimates. An important next issue is how this misspecification biases counterfactual analysis and the resulting policy recommendations. A natural policy question in the case of microfinance adoption is whom a planner should target in the network, and by how much, with financial inclusion an explicit government objective in many developing countries, including India (for example, Pradhan Mantri Jan Dhan Yojana).[31] The answer depends on which households are central, and hence on the shape of the norm. The optimal allocation is a network-centrality problem, and misspecifying the norm as linear distorts which households count as central for propagation of peer influence, so the

[31]A separate but important question is whether to target welfare improvements instead of adoption in discrete choice models, for example Bhattacharya et al. (2024); Volpe (2025) and others. We focus on adoption rather than welfare for two reasons. First, raising take-up is the explicit objective of microfinance and financial-inclusion programmes (Banerjee et al., 2013, 2015b,a), and more broadly of the network-targeting literature on product and technology adoption that we build on (Beaman et al., 2021; Banerjee et al., 2019; Bursztyn et al., 2014). Second, welfare analysis in these models is especially difficult, requiring a stance on the private value of microfinance, as well as on whether a planner should consider conformity as welfare relevant. Moreover, Bhattacharya et al. (2024) show that choice data, while sufficient to predict counterfactual demand, may not identify welfare, since distinct mechanisms behind the same interaction coefficient imply different welfare effects.

planner targets on the wrong centrality measure.

We set the problem of the planner as follows. The planner maximises expected adoption by allocating subsidies, such as conditional cash transfers or take-up incentives conditional on adoption, to individuals in the network subject to a fixed budget. A subsidy $\tau_i \geq 0$ raises agent i's incentive from α_i to $\alpha_i + \tau_i$ and we assume that these subsidies are public knowledge. For an agent with at least one peer, the post-policy equilibrium satisfies

$$p_i = \Gamma_i(\mathbf{p}, \boldsymbol{\tau}; \beta) = F_\eta\left(\alpha_i + \tau_i + (\theta_1 d_i + \theta_2)\widetilde{y}^e_{-i}(\mathbf{p}; \beta) - \frac{\theta_2}{2}\right). \tag{20}$$

Denote the new conditional choice probabilities for a vector of subsidies by $\mathbf{p}(\boldsymbol{\tau}; \beta)$. For a fixed and finite budget B, the planner's problem is to,

$$\max_{\boldsymbol{\tau} \geq \mathbf{0}} \frac{1}{N}\sum_{i=1}^{N} p_i(\boldsymbol{\tau}; \beta) \qquad \text{subject to} \qquad \sum_{i=1}^{N} \tau_i = B \qquad \mathbf{p}(\boldsymbol{\tau}, \beta) = \Gamma(\mathbf{p}, \boldsymbol{\tau}; \beta). \tag{21}$$

The Jacobian of the post-policy best-response mapping $\Gamma(\cdot; \beta)$ in (20), evaluated at the equilibrium, has typical element

$$J_{ij} = \frac{\partial \Gamma_i}{\partial p_j} = p_i(1 - p_i)\,(\theta_1 d_i + \theta_2)\,\frac{\partial \widetilde{y}^e_{-i}(\mathbf{p}; \beta)}{\partial p_j}, \qquad J_{ij} = 0 \text{ for isolated } i.$$

Here $\partial \widetilde{y}^e_{-i}/\partial p_j$ is the marginal effect of j's adoption on i's norm. This term defines an effective influence network through which peer effects propagate (similarly defined in Tchuente (2026)). Its form however, is governed by β and in the linear case this factor is simply $\tilde{g}_{ij}$. While a closed-form solution for optimal adoption does not exist, Proposition 1 below presents the solution for a generalised norm with elasticity β.

Proposition 1 (Optimal Targeting)**.** *Suppose Assumptions 1–7 hold with α_i replaced by the post-policy incentives $\alpha_i + \tau_i^*$. At any optimal allocation $\boldsymbol{\tau}^*$ where the norm is continuously*

differentiable, there exists λ *such that*

$$\frac{p_i(1-p_i)}{N}\left[(I-J)^{-\top}\iota\right]_i \quad \begin{cases} = \lambda, & \tau_i^* > 0, \\ \leq \lambda, & \tau_i^* = 0, \end{cases}$$

where all quantities are evaluated at the post-policy equilibrium.

Proof. Proof in Appendix Section A.4. □

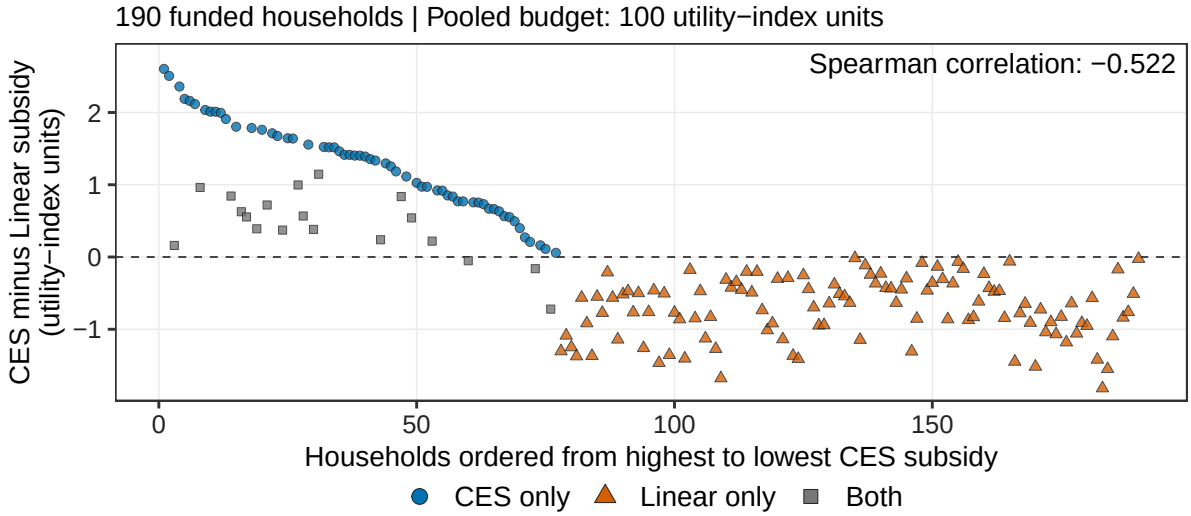
Remark. Here, λ is the Lagrange multiplier on the budget constraint and the left-hand side is the marginal return to subsidising agent i. The planner funds households until these marginal returns are equalised with any household left unsubsidised having a weakly lower return. $[(I-J)^{-\top}\iota]_i$, with ι the vector of ones, is the total increase in adoption, across all households, from a unit increase in i's adoption probability after network feedback is accounted for. It is i's Katz–Bonacich centrality in the transposed effective network $J^\top$. Multiplying it by $p_i(1-p_i)/N$, the response of i's adoption probability to a unit subsidy, gives the marginal return. For $\beta = 1$, this is similar to the standard centrality-based targeting of Ballester et al. (2006); Galeotti et al. (2020) but for a non-linear norm the same centrality is computed on the influence network defined by $\partial \widetilde{y}^e_{-i}/\partial p_j$. Note that the objective is not globally concave in $\boldsymbol{\tau}$ since the logit is not globally concave. Therefore, optimal allocation need not be unique, and the condition above is necessary but not sufficient. In practice we compute the optimal policy $\boldsymbol{\tau}^*$ numerically.[32]

We take the estimates from Section 6 for the pooled sample and compare the optimal policy for the estimated β to that of the restricted linear model under their respective peer effect estimates with a common budget of $B = 100$ utility-index units.[33] Figure 6 presents the comparison. Of the 190 households funded under at least one norm (77

[32] We solve the planner's problem by spectral projected gradient (Birgin et al., 2000), projecting onto the budget set $\{\boldsymbol{\tau} \geq \mathbf{0} : \sum_i \tau_i = B\}$, and stop when the duality gap falls below 2×10^{-6}. Because the objective is not concave, we start from three allocations (uniform, the budget spread over the 1,000 households with the highest marginal returns at $\boldsymbol{\tau} = \mathbf{0}$, and a rescaled larger-budget solution) and report the stationary allocation with the highest adoption. Uniqueness is checked by iterating from the $\mathbf{p} = \mathbf{0}$ and $\mathbf{p} = \mathbf{1}$ end points.

[33] The budget is small and holding peers' adoption probabilities at their baseline values, the optimal allocation raises average adoption by only about a third of a percentage point under either norm.

under CES and 132 in linear), only 19 households overlapping. Figure 1 presents the difference for an illustrative village. Who is targeted and by how much is fundamentally different under the two norms, with a Spearman rank correlation of −0.52. Because the pooled norm is strongly max-leaning, the CES planner concentrates the budget on a small set of high-centrality hubs in the effective network, whereas the linear planner spreads it far more evenly. For the same budget this gap in targeting translates into an increase in predicted adoption of 10.9 percentage points for the CES case while only 1.9 under the linear norm, reflecting the gap due to both the norm and the peer effect estimates.



Notes: Each point is one of the 190 households funded under at least one norm, ordered by CES subsidy and the vertical axis is the CES minus linear subsidy. Each model uses its own estimates and equilibrium to allocate the same 100 utility-index-unit budget. The allocations are strongly negatively rank-correlated (Spearman −0.52), and across all 6,733 households the budget raises predicted adoption by 1.9 points under the linear norm (17.4% → 19.3%) and 10.9 under CES (17.4% → 28.2%).

Figure 6: Optimal subsidy allocation under CES and linear norms in the pooled sample

Table 3 shows that the two norms disagree far less about *how connected* their targets are than about *who* those targets are. Both policies fund well-connected households, with mean degree near 15 against 5.5 in the population that is not targeted, and the households the two norms agree on are the most connected and wealthiest of all, although the size of the subsidy still differs. The decisive difference is whether the household is a leader or not. 62% of the CES-only recipients are village leaders, against only 12% of the linear-only recipients which is close to their proportion in the population. Leaders have a similar positive own coefficient under both norms (0.56 and 0.57), with higher

baseline adoption (0.239 versus 0.165) and degree (7.5 versus 5.5). Under a max-like norm a household's reference point is its highest-adopting neighbour, so a leader, who adopts more and has more links, generates especially strong spillovers when it is the maximal peer for others.

This parallels Banerjee et al. (2013), where information about the programme was seeded through leaders (teachers, shopkeepers, and self-help-group heads) and they found that the centrality of these first contacts predicted participation. Although their mechanism is one of information diffusion, our estimated norm in the simultaneous belief-based game recovers the same pattern by identifying those leaders from the data. In the linear norm, leadership carries no special value as a neighbour's influence is independent of its likelihood of adoption and the budget instead flows to a broader, somewhat wealthier set of high-degree households. Misspecifying the norm as linear therefore does not just rescale a common ranking, it ignores the very households the max-leaning norm makes pivotal.

Recipient group	N	Degree	Rooms	Beds	Non-Hindu	SC/ST	Latrine	Leader
CES only	58	15.9	2.22	0.69	12.1%	56.9%	17.2%	62.1%
Linear only	113	14.7	2.74	0.78	3.5%	38.9%	33.6%	12.4%
Both	19	21.2	3.21	1.05	10.5%	52.6%	36.8%	36.8%
Neither	6,543	5.5	2.44	0.76	4.9%	32.6%	26.1%	11.5%

Notes: Entries are group averages. A household is a recipient if its subsidy exceeds 10^{-6}. Degree is the number of network links, and Rooms and Beds are the household's respective numbers of rooms and beds. SC/ST is the Scheduled-Caste or Scheduled-Tribe flag and Non-Hindu the non-Hindu household-head flag, while Latrine and Leader indicate whether the household has a latrine and is a village leader.

Table 3: Household characteristics by subsidy-recipient group for pooled sample.

A similar exercise for illustrative segments (heterogeneous and rich) under the two wealth proxies is reported in Appendix E in Figure E.3 and Table E.20 with a budget of $B = 20$ for each segment. For the Caste-Beds segment the norm is again max-leaning ($\hat{\beta} \approx 491$) and optimal subsidy is concentrated on 15 high-degree households with 40% of them being village leaders mirroring the pooled sample optimal policy. For the Caste-Latrine segment the estimated norm is instead minimum-like ($\hat{\beta} \approx -283$), and the targeting is very different. The CES planner now spreads the budget across 93 households

and raises adoption by only 0.9 points, similar to the linear norm. Its recipients are low-degree (6.1 versus 13.6 for the linear) but on average richer (3.4 rooms and 1.2 beds, versus 2.3 rooms and 0.6 beds), and, in contrast to the maximum-like regime, they are not village leaders (3.3% versus 26.1% under the linear norm). Under the min-like norm, it is more important to target the ones with the lowest adoption probability. This illustrates that a different estimated norm delivers very different policy conclusions and misspecification would induce bias.

8 Concluding Remarks

With the growing availability of network data and the policy relevance of economic questions that require modelling peer effects, it is crucial to understand the implications of the underlying model primitives and assumptions. Any form of misspecification can lead to biased counterfactual predictions and, consequently, sub-optimal policy recommendations. This paper studies one particular form of misspecification in discrete choice models of peer effects, specifically the social norm or how peer influence is aggregated. Ex-ante assumptions on the social norm are typically ad hoc, and while recent work has relaxed them for continuous outcomes (Boucher et al., 2024; Tchuente, 2026), many empirically relevant settings, for example microfinance adoption, have discrete outcomes. We study norm misspecification with both conformity and complementarity effects extending the CES norm of Boucher et al. (2024) to a setting that raises challenges absent in the continuous case.

We model the game under incomplete information, where agents form beliefs about others' actions, and give a full structural treatment of the model. We establish existence and uniqueness of equilibrium, and identification of the structural parameters. Because the CES norm and the discrete choice setting makes the best-response system highly non-linear, none of these follows directly from the linear or continuous case. Building on Aguirregabiria and Mira (2007) and Aguirregabiria and Marcoux (2021), we propose a consistent and asymptotically normal estimator of the structural parameters and

equilibrium conditional choice probabilities. In Monte Carlo simulations, its peer-effect estimates approach their true values as the sample grows, whereas imposing a linear norm leaves them biased.

In the microfinance data of Banerjee et al. (2013), the pooled estimates reject the linear norm in favour of a max-like one, and imposing linearity overstates complementarity more than threefold while understating conformity. The norm also varies across caste and wealth segments, heterogeneity that pooling hides. The misspecification carries over to policy. The optimal subsidies under the estimated and the linear norm fund almost disjoint sets of households, and under the max-like norm the planner favours village leaders, echoing their role in Banerjee et al. (2013). More broadly, wherever peer effects shape take-up, from financial inclusion to other adoption programmes that governments and NGOs pursue, getting the norm wrong would not only bias the estimated peer effects but also misdirect the subsidy, so getting the shape of the norm right is important.

References

Aguirregabiria, V. and Marcoux, M. (2021). Imposing equilibrium restrictions in the estimation of dynamic discrete games. *Quantitative Economics*, 12(4):1223–1271.

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

Bajari, P., Chernozhukov, V., Hong, H., and Nekipelov, D. (2015). Identification and Efficient Semiparametric Estimation of a Dynamic Discrete Game. NBER Working Papers 21125, National Bureau of Economic Research, Inc.

Bajari, P., Hong, H., and Ryan, S. P. (2010). Identification and estimation of a discrete game of complete information. *Econometrica*, 78(5):1529–1568.

Ballester, C., Calvó-Armengol, A., and Zenou, Y. (2006). Who's who in networks. wanted: The key player. *Econometrica*, 74(5):1403–1417.

Banerjee, A., Chandrasekhar, A. G., Duflo, E., and Jackson, M. O. (2013). The diffusion of microfinance. *Science*, 341(6144):1236498.

Banerjee, A., Chandrasekhar, A. G., Duflo, E., and Jackson, M. O. (2019). Using gossips to spread information: Theory and evidence from two randomized controlled trials. *Review of Economic Studies*, 86(6):2453–2490.

Banerjee, A., Duflo, E., Glennerster, R., and Kinnan, C. (2015a). The miracle of microfinance? evidence from a randomized evaluation. *American Economic Journal: Applied Economics*, 7(1):22–53.

Banerjee, A., Karlan, D., and Zinman, J. (2015b). Six randomized evaluations of microcredit: Introduction and further steps. *American Economic Journal: Applied Economics*, 7(1):1–21.

Battigalli, P. and Dufwenberg, M. (2009). Dynamic psychological games. *Journal of Economic Theory*, 144(1):1–35.

Battigalli, P. and Dufwenberg, M. (2022). Belief-dependent motivations and psychological game theory. *Journal of Economic Literature*, 60(3):833–882.

Beaman, L., BenYishay, A., Magruder, J., and Mobarak, A. M. (2021). Can network theory-based targeting increase technology adoption? *American Economic Review*, 111(6):1918–1943.

Bhattacharya, D., Dupas, P., and Kanaya, S. (2024). Demand and welfare analysis in discrete choice models with social interactions. *Review of Economic Studies*, 91(2):748–784.

Bietenbeck, J. (2020). The long-term impacts of low-achieving childhood peers: Evidence from Project STAR. *Journal of the European Economic Association*, 18(1):392–426.

Birgin, E. G., Martínez, J. M., and Raydan, M. (2000). Nonmonotone spectral projected gradient methods on convex sets. *SIAM Journal on Optimization*, 10(4):1196–1211.

Blume, L. E., Brock, W. A., Durlauf, S. N., and Ioannides, Y. M. (2011). Identification of social interactions. In Benhabib, J., Bisin, A., and Jackson, M. O., editors, *Handbook of Social Economics*, volume 1, pages 853–964. North-Holland.

Bollinger, B. and Gillingham, K. (2012). Peer effects in the diffusion of solar photovoltaic panels. *Marketing Science*, 31(6):900–912.

Boucher, V. and Bramoullé, Y. (2026). Binary outcomes and linear interactions. *Journal of Business & Economic Statistics*, 44(3):865–875.

Boucher, V. and Fortin, B. (2016). Some challenges in the empirics of the effects of networks. In *The Oxford Handbook of the Economics of Networks*. Oxford University Press.

Boucher, V., Rendall, M., Ushchev, P., and Zenou, Y. (2024). Toward a general theory of peer effects. *Econometrica*, 92(2):543–565.

Bowden, R. (1973). The theory of parametric identification. *Econometrica*, 41(6):1069–1074.

Bramoullé, Y., Djebbari, H., and Fortin, B. (2020). Peer effects in networks: A survey. *Annual Review of Economics*, 12(1):603–629.

Brock, W. A. and Durlauf, S. N. (2001a). Discrete choice with social interactions. *The Review of Economic Studies*, 68(2):235–260.

Brock, W. A. and Durlauf, S. N. (2001b). Interactions-based models. In *Handbook of Econometrics*, volume 5, pages 3297–3380. Elsevier.

Brock, W. A. and Durlauf, S. N. (2007). Identification of binary choice models with social interactions. *Journal of Econometrics*, 140(1):52–75.

Bursztyn, L., Ederer, F., Ferman, B., and Yuchtman, N. (2014). Understanding mechanisms underlying peer effects: Evidence from a field experiment on financial decisions. *Econometrica*, 82(4):1273–1301.

Bursztyn, L., González, A. L., and Yanagizawa-Drott, D. (2020). Misperceived social norms: Women working outside the home in Saudi Arabia. *American Economic Review*, 110(10):2997–3029.

Cai, J., De Janvry, A., and Sadoulet, E. (2015). Social networks and the decision to insure. *American Economic Journal: Applied Economics*, 7(2):81–108.

Calvó-Armengol, A., Patacchini, E., and Zenou, Y. (2009). Peer effects and social networks in education. *Review of Economic Studies*, 76(4):1239–1267.

Carrell, S. E., Page, M. E., and West, J. E. (2010). Sex and science: How professor gender perpetuates the gender gap. *The Quarterly Journal of Economics*, 125(3):1101–1144.

Chambers, C. P., Cuhadaroglu, T., and Masatlioglu, Y. (2023). Behavioral influence. *Journal of the European Economic Association*, 21(1):135–166.

Charness, G., Naef, M., and Sontuoso, A. (2019). Opportunistic conformism. *Journal of Economic Theory*, 180:100–134.

Chung, F. and Lu, L. (2002). Connected components in random graphs with given expected degree sequences. *Annals of Combinatorics*, 6(2):125–145.

de Paula, Á., Richards-Shubik, S., and Tamer, E. (2018). Identifying preferences in networks with bounded degree. *Econometrica*, 86(1):263–288.

Díaz, C., Patacchini, E., Verdier, T., and Zenou, Y. (2021). Leaders in juvenile crime. *Journal of Economic Behavior & Organization*, 192:638–667.

Feigenberg, B., Field, E., and Pande, R. (2013). The economic returns to social interaction: Experimental evidence from microfinance. *Review of Economic Studies*, 80(4):1459–1483.

Galeotti, A., Golub, B., and Goyal, S. (2020). Targeting interventions in networks. *Econometrica*, 88(6):2445–2471.

Geanakoplos, J., Pearce, D., and Stacchetti, E. (1989). Psychological games and sequential rationality. *Games and Economic Behavior*, 1(1):60–79.

Guerra, J.-A. and Mohnen, M. (2022). Multinomial choice with social interactions: Occupations in Victorian London. *The Review of Economics and Statistics*, 104(4):736–747.

Hahn, Y., Islam, A., Patacchini, E., and Zenou, Y. (2020). Friendship and female education: Evidence from a field experiment in Bangladeshi primary schools. *The Economic Journal*, 130(627):740–764.

Herstad, E. I., Houndetoungan, A., and Shin, M. (2026). Identification of heterogeneous peer effects. arXiv:2410.14317.

Houndetoungan, A. (2025). Quantile peer effect models. arXiv:2506.12920.

Houndetoungan, A. (2026). Count data models with heterogeneous peer effects under rational expectations. *Journal of Applied Econometrics*, 41(3):295–309.

Houndetoungan, A. and Lambotte, M. (2026). Asymmetries in peer effects. arXiv:2608.09219.

Hoxby, C. and Weingarth, G. (2005). Taking race out of the equation: School reassignment and the structure of peer effects. Unpublished manuscript, Harvard University.

Islam, A., Vlassopoulos, M., Zenou, Y., and Zhang, X. (2021). Centrality-based spillover effects. Discussion Paper DP16321, Centre for Economic Policy Research (CEPR). Available at SSRN.

Kasahara, H. and Shimotsu, K. (2012). Sequential estimation of structural models with a fixed point constraint. *Econometrica*, 80(5):2303–2319.

Kőszegi, B. and Rabin, M. (2006). A model of reference-dependent preferences. *Quarterly Journal of Economics*, 121(4):1133–1165.

La Cruz, W., Martínez, J. M., and Raydan, M. (2006). Spectral residual method without gradient information for solving large-scale nonlinear systems of equations. *Mathematics of Computation*, 75(255):1429–1448.

Lambotte, M. (2025a). Heterogeneity in peer effects for binary outcomes. arXiv:2511.15891.

Lambotte, M. (2025b). Peer effects in binary outcomes: Strategic complementarity and taste for conformity with endogenous networks. *Journal of Applied Econometrics*, 40(6):608–626.

Lee, L.-f., Li, J., and Lin, X. (2014). Binary choice models with social network under heterogeneous rational expectations. *The Review of Economics and Statistics*, 96(3):402–417.

Lin, Z. and Hu, Y. (2024). Binary choice with misclassification and social interactions, with an application to peer effects in attitude. *Journal of Econometrics*, 238(1):105551.

Liu, X., Patacchini, E., and Zenou, Y. (2014). Endogenous peer effects: local aggregate or local average? *Journal of Economic Behavior & Organization*, 103:39–59.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *Review of Economic Studies*, 60(3):531–542.

Mas, A. and Moretti, E. (2009). Peers at work. *American Economic Review*, 99(1):112–145.

Meng, D. (2025). Targeting network intervention with social norm. *Journal of Economic Theory*, 230:106092.

Menzel, K. (2016). Inference for games with many players. *The Review of Economic Studies*, 83(1):306–337.

Munshi, K. and Rosenzweig, M. (2006). Traditional institutions meet the modern world: Caste, gender, and schooling choice in a globalizing economy. *American Economic Review*, 96(4):1225–1252.

Nakajima, R. (2007). Measuring peer effects on youth smoking behaviour. *The Review of Economic Studies*, 74(3):897–935.

Norros, I. and Reittu, H. (2006). On a conditionally Poissonian graph process. *Advances in Applied Probability*, 38(1):59–75.

Pesendorfer, M. and Schmidt-Dengler, P. (2008). Asymptotic least squares estimators for dynamic games. *The Review of Economic Studies*, 75(3):901–928.

Pesendorfer, M. and Schmidt-Dengler, P. (2010). Sequential estimation of dynamic discrete games: A comment. *Econometrica*, 78(2):833–842.

Rothenberg, T. J. (1971). Identification in parametric models. *Econometrica*, 39(3):577–591.

Shemesh, J., Zapatero, F., and Zenou, Y. (2022). Heterogenous Peer Effects: How Community Connectivity Affects Car Purchases. CEPR Discussion Papers 17306, C.E.P.R. Discussion Papers.

Tao, J. and Lee, L.-f. (2014). A social interaction model with an extreme order statistic. *The Econometrics Journal*, 17(3):197–240.

Tatsi, E. (2015). Endogenous social interactions: Which peers matter? VfS Annual Conference 2015 (Muenster): Economic Development - Theory and Policy 113168, Verein für Socialpolitik / German Economic Association.

Tchuente, G. (2026). Who matters to whom? identifying peer effects with propagation geometry. arXiv:2602.23594.

Topkis, D. (1998). *Supermodularity and Complementarity*. Princeton University Press.

Volpe, O. (2025). Discrete choice with generalized social interactions. Technical report, Working paper, Harvard University.

Yakovlev, E. (2018). Demand for alcohol consumption in Russia and its implication for mortality. *American Economic Journal: Applied Economics*, 10(1):106–49.

Appendix

A Proofs of Main Results

A.1 Existence

In this section we prove the Existence of a Rational Expectations equilibrium of the game using Tarski's Fixed Point Theorem. We first state the Theorem.

Definition A.1 (Tarski's Fixed Point Theorem)**.** Let L be a complete lattice and let $f : L \to L$ be an order-preserving (i.e., monotone) function. Then the set of fixed points of f is a non-empty complete lattice. In particular, f has a least and a greatest fixed point.

Moreover, define the component-wise order as below:

Definition A.2 (Component-wise Order)**.** Let $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^N$ and $\mathbf{y} = (y_1, \ldots, y_n) \in \mathbb{R}^N$. We define the component-wise (partial) order as follows:

- $\mathbf{x} \geq \mathbf{y}$ if and only if $x_i \geq y_i$ for all $i = 1, \ldots, N$,
- $\mathbf{x} > \mathbf{y}$ if and only if $\mathbf{x} \geq \mathbf{y}$ and there exists at least one index j such that $x_j > y_j$.

Note that L := $[0, 1]^N$ is a complete lattice under the component-wise order. For any subset $S \subseteq [0, 1]^N$, define the infimum and supremum component-wise:

$$\inf S = \left(\inf_{x \in S} x_1, \ldots, \inf_{x \in S} x_n\right), \quad \sup S = \left(\sup_{x \in S} x_1, \ldots, \sup_{x \in S} x_n\right)$$

Since each $[0, 1]$ is a complete lattice under the usual order, each component infimum and supremum exists and lies in $[0, 1]$. Hence, $[0, 1]^N$ is a complete lattice under the component-wise partial order. We need to show that $\Gamma(p, \boldsymbol{\theta})$ is a monotone function in p then by Tarski's Fixed Point Theorem, the set of fixed points of Γ would be non-empty and in itself form a complete lattice under pair-wise comparison i.e. there exist a least and greatest fixed point.

We now present the proof of Theorem 1.

Proof of Theorem 1:

We will use Tarski's fixed point theorem to prove existence.

Self-map. Since $F_{\eta_i} \in (0,1)$ for every i, we have $\Gamma(\mathbf{p};\boldsymbol{\theta}) \in (0,1)^N \subset [0,1]^N$, so Γ maps $[0,1]^N$ into itself.

Complete Lattice. The domain $[0,1]^N$, equipped with the component-wise partial order (Definition A.2), is a complete lattice. For any subset $S \subseteq [0,1]^N$, the infimum and supremum are defined component wise and lie in $[0,1]^N$.

Monotonicity. We show that $\Gamma(\cdot;\boldsymbol{\theta})$ is monotone. Let $\mathbf{p}, \mathbf{p}' \in [0,1]^N$ with $\mathbf{p} \leq \mathbf{p}'$ component wise. Consider:

$$\tilde{y}^e_{-i}(\mathbf{p}) = \left(\sum_{j=1}^{N} \tilde{g}_{ij} p_j^{\beta} \right)^{1/\beta}, \quad \text{with } \tilde{g}_{ij} \geq 0.$$

Since $p_j \leq p'_j$ and all $\tilde{g}_{ij} \geq 0$, it follows that:

$$\tilde{y}^e_{-i}(\mathbf{p}) \leq \tilde{y}^e_{-i}(\mathbf{p}').$$

Now, using Assumption 5, we have $\theta_1 d_i + \theta_2 > 0$. Therefore,

$$z_i(\mathbf{p}) = \alpha_i + (\theta_1 d_i + \theta_2)\tilde{y}^e_{-i}(\mathbf{p},\beta) - \frac{\theta_2}{2} \leq z_i(\mathbf{p}') = \alpha_i + (\theta_1 d_i + \theta_2)\tilde{y}^e_{-i}(\mathbf{p}',\beta) - \frac{\theta_2}{2}.$$

Since F_{η_i} is strictly increasing (as the logistic CDF), we have

$$\Gamma_i(\mathbf{p};\boldsymbol{\theta}) = F_{\eta_i}(z_i(\mathbf{p})) \leq F_{\eta_i}(z_i(\mathbf{p}')) = \Gamma_i(\mathbf{p}';\boldsymbol{\theta}).$$

Thus, $\Gamma(\cdot;\boldsymbol{\theta})$ is monotone on $[0,1]^N$. Since Γ is a monotone function from a complete lattice into itself, by Tarski's Fixed Point Theorem, the set of fixed points of Γ is non-empty and forms a complete lattice. Therefore, there exist minimal ($\underline{\mathbf{p}}^*$) and maximal ($\overline{\mathbf{p}}^*$) fixed points with

$$\underline{\mathbf{p}}^* = \Gamma(\underline{\mathbf{p}}^*;\boldsymbol{\theta}) \text{ and } \overline{\mathbf{p}}^* = \Gamma(\overline{\mathbf{p}}^*;\boldsymbol{\theta})$$

□

This result yields implications beyond mere existence. It establishes that the set of fixed points of Γ is a complete lattice in the induced order, with a least and a greatest element. However, it need not be a sublattice of $[0,1]^N$, so the component-wise maximum of two fixed points need not be a fixed point.

A.2 Uniqueness

We first define the logarithmic L_∞ distance, also known as the Thompson metric, used in the uniqueness proof.

Definition A.3 (Logarithmic L_∞ distance)**.** For any $\mathbf{p}, \mathbf{q} \in \mathbb{R}^N_{++}$, define

$$d_{\log,\infty}(\mathbf{p},\mathbf{q}) := \|\log \mathbf{p} - \log \mathbf{q}\|_\infty = \max_{1\le i\le N} \left|\log\left(\frac{p_i}{q_i}\right)\right|.$$

The logarithm is applied component-wise. The Thompson metric measures the largest proportional difference between corresponding components of two strictly positive vectors. We also re-state Assumption 1 of Boucher et al. (2024) which holds here too.

Assumption (Assumption 1 of Boucher et al. (2024))**.** For all $i = 1, 2, \ldots, N$, agent i's social norm $\tilde{y}^e_i : \mathbb{R}^N \to \mathbb{R}$ is monotone increasing, continuous, linearly homogeneous, and normalised such that $\tilde{y}^e_i(\mathbf{1}) = 1$.

The condition in Assumption 6 is *radial* (multiplicative) instead of the usual additive restrictions used in the linear case. An additive condition contracts absolute differences $|p_i - p'_i|$ in beliefs, whereas our radial condition contracts relative differences $|\log(p_i/p'_i)|$ (the Thompson metric). Because the CES norm is scale-free, the relative scale is the natural one. Tchuente (2026, appendix) gives an additive, Lipschitz-based uniqueness proof for the CES norm, but that argument does not generally deliver a contraction when $\beta < 1$, for which the norm need not be globally non-expansive in levels. For $0 \le \beta < 1$, $\partial \tilde{y}^e_{-i}/\partial p_j \propto p_j^{\beta-1}$ diverges as a belief approaches zero, so a Lipschitz bound holds only on beliefs bounded away from zero, with a constant that depends on β and

that range. For $\beta < 0$ a Lipschitz bound holds on the whole domain, but its constant, up to $d_i^{1/|\beta|}$, exceeds one and diverges as $\beta \to 0^-$, so peer effects must be correspondingly smaller. Our radial condition needs no such restriction. Monotonicity and degree-one homogeneity make the CES norm non-expansive in the Thompson metric for every β, and the best-response elasticity vanishes as the norm approaches zero, so uniqueness holds on the entire domain under an explicit condition.

We now present the proof of Theorem 2.

Proof of Theorem 2:

By Theorem 1, $\Gamma(\cdot)$ has a least fixed point $\underline{\mathbf{p}}^*$ and a greatest fixed point $\overline{\mathbf{p}}^*$ satisfying $\underline{\mathbf{p}}^* \leq \overline{\mathbf{p}}^*$. Because the logistic CDF takes values strictly between zero and one, both fixed points are strictly positive. Applying the CES transformation then implies that the corresponding social norms $\tilde{\mathbf{y}}^e(\cdot)$ at the extreme fixed points are also strictly positive in the absence of isolated agents.

Let $D = d_{\log,\infty}(\overline{\mathbf{p}}^*, \underline{\mathbf{p}}^*)$ denote the Thompson distance between the extreme fixed points. Since both are fixed points,

$$D = \left\| \log \Gamma(\overline{\mathbf{p}}^*; \boldsymbol{\theta}) - \log \Gamma(\underline{\mathbf{p}}^*; \boldsymbol{\theta}) \right\|_\infty . \tag{22}$$

We will show that Assumption 6 forces $D = 0$.

Let $\underline{y}_i = \tilde{y}^e_{-i}(\underline{\mathbf{p}}^*; \beta)$ and $\overline{y}_i = \tilde{y}^e_{-i}(\overline{\mathbf{p}}^*; \beta)$ for the norms at the two extreme fixed points. By Assumption 6 these lie in the interval $[a, b]$ on which the best-response elasticity, ε_i^{BR}, is bounded. From (11), the elasticity of $\Gamma_i(\cdot)$ with respect to the norm of i is,

$$\varepsilon_i^{BR}(\tilde{y}^e_{-i}) = (\theta_1 d_i + \theta_2)\, \tilde{y}^e_{-i} \left[1 - F_\eta\big(\alpha_i - \tfrac{\theta_2}{2} + (\theta_1 d_i + \theta_2)\tilde{y}^e_{-i}\big) \right].$$

By the mean value theorem, there exists a $y_i^\dagger \in [\underline{y}_i, \overline{y}_i] \subseteq [a, b]$ such that

$$\left| \log \Gamma_i(\overline{\mathbf{p}}^*; \boldsymbol{\theta}) - \log \Gamma_i(\underline{\mathbf{p}}^*; \boldsymbol{\theta}) \right| = \varepsilon_i^{BR}(y_i^\dagger) \left| \log \tilde{y}^e_{-i}(\overline{\mathbf{p}}^*; \beta) - \log \tilde{y}^e_{-i}(\underline{\mathbf{p}}^*; \beta) \right|.$$

By Assumption 6 we get that

$$\left|\log \Gamma_i(\overline{\mathbf{p}}^*;\boldsymbol{\theta}) - \log \Gamma_i(\underline{\mathbf{p}}^*;\boldsymbol{\theta})\right| \leq \overline{\varepsilon}^{BR} \left|\log \tilde{y}^e_{-i}(\overline{\mathbf{p}}^*;\beta) - \log \tilde{y}^e_{-i}(\underline{\mathbf{p}}^*;\beta)\right|, \tag{23}$$

where $\overline{\varepsilon}^{BR} = \max_{1\leq i\leq N} \sup_{\tilde{y}^e_{-i}\in[a,b]} \varepsilon_i^{BR}$. Taking the maximum over i and using equation (22) gives

$$D \leq \overline{\varepsilon}^{BR} \left\|\log \tilde{\mathbf{y}}^e(\overline{\mathbf{p}}^*;\beta) - \log \tilde{\mathbf{y}}^e(\underline{\mathbf{p}}^*;\beta)\right\|_\infty. \tag{24}$$

For every i, the definition of D and the ordering of the extremal fixed points implies

$$0 \leq \log\left(\frac{\overline{p}_i^*}{\underline{p}_i^*}\right) \leq D \quad \Longleftrightarrow \quad \underline{\mathbf{p}}^* \leq \overline{\mathbf{p}}^* \leq \exp(D)\,\underline{\mathbf{p}}^*.$$

Applying the CES transformation and using its monotonicity and homogeneous of degree one property gives us,

$$\tilde{\mathbf{y}}^e(\underline{\mathbf{p}}^*;\beta) \leq \tilde{\mathbf{y}}^e(\overline{\mathbf{p}}^*;\beta) \leq \exp(D)\,\tilde{\mathbf{y}}^e(\underline{\mathbf{p}}^*;\beta).$$

Finally, taking component-wise logarithms therefore yields

$$\left\|\log \tilde{\mathbf{y}}^e(\overline{\mathbf{p}}^*;\beta) - \log \tilde{\mathbf{y}}^e(\underline{\mathbf{p}}^*;\beta)\right\|_\infty \leq D. \tag{25}$$

Combining (24) and (25) gives $D \leq \overline{\varepsilon}^{BR} D$. Since Assumption 6 requires $\overline{\varepsilon}^{BR} < 1$, this implies $D = 0$ and hence

$$\overline{\mathbf{p}}^* = \underline{\mathbf{p}}^*.$$

Since the maximal and minimal fixed points coincide, this implies that the set of fixed points of Γ is a singleton. □

Relatedly, we also prove a property concerning the Jacobian of the Best-Response mapping which is used in the proofs in Sections A.3 and A.4. For its proof, we require the following definitions:

Definition A.4 (Spectral Radius)**.** For $A \in \mathbb{R}^{N\times N}$ with (complex) eigenvalues $\lambda_1, \ldots, \lambda_N$, the *spectral radius* is

$$\rho(A) := \max_{1\le k\le N} |\lambda_k|,$$

the largest modulus of its eigenvalues.

Definition A.5 (Spectral Radius: Invertibility, Row-Sum Bound, and Similarity)**.** Let $A \in \mathbb{R}^{N\times N}$.

- *Invertibility (Neumann series):* if $\rho(A) < 1$, then $I - A$ is invertible and $(I - A)^{-1} = \sum_{k=0}^{\infty} A^k$.
- *Row-sum bound:* $\rho(A) \le \|A\|_\infty = \max_{1\le i\le N} \sum_{j=1}^{N} |A_{ij}|$; if $A \ge \mathbf{0}$ entrywise, this equals $\max_i \sum_j A_{ij}$.
- *Similarity invariance:* for any invertible diagonal D, $\rho(D^{-1}AD) = \rho(A)$.

Definition A.6 (Euler's Theorem for Homogeneous Functions)**.** If $g : \mathbb{R}^N_{++} \to \mathbb{R}$ is differentiable and homogeneous of degree one, then $\sum_{j=1}^{N} x_j \frac{\partial g}{\partial x_j}(\mathbf{x}) = g(\mathbf{x})$ for every $\mathbf{x} \in \mathbb{R}^N_{++}$.

Definition A.7 (Partitioned Inverse)**.** Let A be an invertible matrix, partitioned as

$$A = \begin{pmatrix} A_{SS} & A_{ST} \\ A_{TS} & A_{TT} \end{pmatrix},$$

with A_{TT} invertible. Then $A_{SS} - A_{ST}A_{TT}^{-1}A_{TS}$ is invertible, and

$$\left[A^{-1}\right]_{SS} = \left(A_{SS} - A_{ST}A_{TT}^{-1}A_{TS}\right)^{-1}.$$

Lemma 1 (Best-Response Jacobian)**.** *Let $\mathbf{p}^* = \Gamma(\mathbf{p}^*; \boldsymbol{\theta})$ be the equilibrium and $J = \partial\Gamma(\mathbf{p}^*; \theta)/\partial\mathbf{p}^\top$ the Jacobian of the best-response map at $\mathbf{p}^*$. Then under Assumptions 1, 5 and 6, for every set S of individuals,*

$$I - J, \qquad I - J_{SS} \qquad \text{and} \qquad \left[(I - J)^{-1}\right]_{SS}$$

are invertible, where the subscript SS denotes the block with rows and columns in S.

Proof. By Assumption 1, F_η is logistic, so its density at the equilibrium is $p_i^*(1-p_i^*)$. Differentiating Γ_i therefore gives zero rows of J for isolated individuals and

$$J_{ij} = p_i^*(1-p_i^*)(\theta_1 d_i + \theta_2)\frac{\partial \tilde{y}^e_{-i}(\mathbf{p}^*;\beta)}{\partial p_j}, \qquad d_i > 0.$$

By Assumption 5, $\theta_1 d_i + \theta_2 > 0$, and the norm is increasing in each p_j (Assumption 1 of Boucher et al. (2024), restated above), so $J \geq \mathbf{0}$. The similar matrix

$$\widetilde{J} = \operatorname{diag}(\mathbf{p}^*)^{-1} J \operatorname{diag}(\mathbf{p}^*)$$

is therefore nonnegative and has the same spectral radius as J (Definition A.5). The norm is homogeneous of degree one, so, by Euler's theorem (Definition A.6), the row sums of $\widetilde{J}$ are

$$\begin{aligned}\sum_j \widetilde{J}_{ij} &= (1-p_i^*)(\theta_1 d_i + \theta_2)\sum_j \frac{\partial \tilde{y}^e_{-i}(\mathbf{p}^*;\beta)}{\partial p_j} p_j^* \\ &= (1-p_i^*)(\theta_1 d_i + \theta_2)\,\tilde{y}^e_{-i}(\mathbf{p}^*;\beta) = \varepsilon_i^{BR} \leq \bar{\varepsilon}^{BR} < 1\end{aligned}$$

for $d_i > 0$, by (11) and Assumption 6, since the equilibrium norm lies in $[a, b]$. By the row-sum bound (Definition A.5),

$$\rho(J) = \rho(\widetilde{J}) < 1,$$

so $I - J$ is invertible (Definition A.5). For a set S, J_{SS} is similar to $\widetilde{J}_{SS}$, which is nonnegative with row sums no larger than those of $\widetilde{J}$, so $\rho(J_{SS}) < 1$ and $I - J_{SS}$ is invertible. Let T be the complement of S. Since $I - J$ and $I - J_{TT}$ are invertible, Definition A.7 with $A = I - J$ gives

$$\left[(I-J)^{-1}\right]_{SS} = \left(I - J_{SS} - J_{ST}(I-J_{TT})^{-1}J_{TS}\right)^{-1},$$

so the block is invertible. If T is empty, the block is $(I-J)^{-1}$ itself. □

A.3 Identification

Before we prove Theorem 3, we state intermediate definitions and results used for it. First we define the following:

Definition A.8 (Power-Mean Inequality)**.** For positive numbers $p_1, \ldots, p_n$, the power mean $\left(\frac{1}{n}\sum_{j=1}^{n} p_j^t\right)^{1/t}$, read as the geometric mean at $t = 0$, is non-decreasing in t, and strictly increasing unless $p_1 = \cdots = p_n$.

Definition A.9 (Real Analytic Functions)**.** A function f on an open set $U \subseteq \mathbb{R}^n$ is *real analytic* if, near every point of U, it equals a convergent power series. We use two standard facts.

- Sums, products, quotients and compositions of real analytic functions, including exponentials and logarithms, are real analytic where defined.
- *Identity theorem:* if U is connected and $f = 0$ on a non-empty open subset of U, then $f = 0$ on all of U.

We also formally define variation in one's neighbourhood that identifies the parameters.

Definition A.10 (Neighbour Variation)**.** Let $\mathbf{x}$ be the payoff-relevant covariate, and $\mathbf{X}^{-x}$ the other covariates. For a set S of individuals, let $\mathbf{x}_{-S} = (x_k)_{k \notin S}$.

- *Neighbour variation* holds for player i through a neighbour j if, for fixed $(\mathbf{G}, \mathbf{M}, \mathbf{X}^{-x}, \mathbf{x}_{-j})$, the support of x_j contains at least two points. It holds at degree d if it holds at some i with $d_i = d$.
- *Continuous neighbour variation* holds for agent i jointly across all its neighbours if, for fixed $(\mathbf{G}, \mathbf{M}, \mathbf{X}^{-x}, \mathbf{x}_{-N_i})$, the support of $(x_j)_{j \in N_i}$ contains a non-empty open set.

Lemmas 2 and 3 link neighbour variation in the payoff-relevant covariate to the equilibrium choice probabilities. Lemma 2 shows that when a change in x_j raises j's intrinsic utility, no equilibrium probability falls and p_j^* rises strictly. Lemma 3 shows how

much variation in the neighbours' probabilities this creates, holding i's own characteristics fixed. In part (a), continuous variation in the neighbours' covariates moves their probabilities over a whole open set and since the best-response Jacobian is invertible (Lemma 1), the neighbours' probabilities can be varied independently over small ranges. In part (b), a covariate taking three values at one neighbour of an individual with two neighbours changes the gap between the two neighbours' probabilities. The neighbour whose covariate rises gains proportionally more than the other, who responds only through its norm and, by Assumption 6, less than proportionally, hence the spread takes at least two values. Finally, Lemma 4 shows that either kind of variation is enough to tell different values of β apart.

Lemma 2 (Neighbour Variation)**.** *Under Assumptions 1–7, suppose two supported configurations, L and H, differ only in j's payoff-relevant covariate, whose values x_j^L and x_j^H satisfy $\gamma_x x_j^H > \gamma_x x_j^L$. Then their equilibria satisfy*

$$\mathbf{p}^{*,H} \geq \mathbf{p}^{*,L}, \qquad p_j^{*,H} > p_j^{*,L}.$$

Proof. Let Γ^L and Γ^H be the best-response maps of the two configurations. They differ only in j's index, which is higher under Γ^H by $\gamma_x(x_j^H - x_j^L) > 0$. By Assumption 5, $\theta_1 d_k + \theta_2 > 0$ for every k with $d_k > 0$, the norm is increasing in each neighbour probability (Assumption 1 of Boucher et al. (2024), restated in Appendix A.2) and F_η is strictly increasing by Assumption 1. Hence both maps are increasing in $\mathbf{p}$ in the component-wise order (Definition A.2), and

$$\Gamma^H(\mathbf{p}^{*,L}) \geq \Gamma^L(\mathbf{p}^{*,L}) = \mathbf{p}^{*,L}, \qquad \Gamma_j^H(\mathbf{p}^{*,L}) > p_j^{*,L}.$$

Iterating Γ^H from $\mathbf{p}^{*,L}$ therefore gives an increasing sequence, bounded by one. Its limit is a fixed point of Γ^H by continuity and equals $\mathbf{p}^{*,H}$, since the equilibrium is unique under Assumptions 1–6 (Theorem 2). Hence

$$\mathbf{p}^{*,H} \geq \Gamma^H(\mathbf{p}^{*,L}),$$

which gives both inequalities. □

Lemma 3 (Covariates and Neighbour Probabilities)**.** *Under Assumptions 1–7,*

(a) If continuous neighbour variation holds at i, then i's neighbour probabilities $(p_j^*)_{j\in N_i}$ *take every value in some non-empty open set* $\prod_{j\in N_i} I'_j$ *of open intervals.*

(b) If $d_i = 2$*, with neighbours j and l, and, for fixed* $(\mathbf{G}, \mathbf{M}, \mathbf{X}^{-x}, \mathbf{x}_{-j})$*, the support of j's payoff-relevant covariate* x_j *contains at least three points, then the spread* $|\log(p_j^*/p_l^*)|$ *takes at least two values.*

Proof. (a) Let $O = \prod_{j\in N_i} I_j$ be the non-empty open set in Definition A.10, so that, for fixed $(\mathbf{G}, \mathbf{M}, \mathbf{X}^{-x}, \mathbf{x}_{-N_i})$, $(x_j)_{j\in N_i}$ takes every value in O. Changing j's payoff-relevant covariate x_j shifts only j's index, by $\gamma_x x_j$. At each point of O, the equilibrium is unique (Theorem 2) and $I - J$ is invertible (Lemma 1). Since F_η is logistic (Assumption 1), the implicit function theorem shows that $(p_j^*)_{j\in N_i}$ is continuously differentiable in $(x_j)_{j\in N_i}$ on O, with Jacobian

$$\frac{\partial (p_j^*)_{j\in N_i}}{\partial (x_j)_{j\in N_i}^\top} = \gamma_x \left[(I-J)^{-1}\right]_{N_i N_i} \operatorname{diag}\{p_j^*(1-p_j^*)\}_{j\in N_i}.$$

This matrix is invertible, since $[(I-J)^{-1}]_{N_iN_i}$ is invertible by Lemma 1 and $\gamma_x \neq 0$ by Assumption 7. By the inverse function theorem, the image of O contains a non-empty open set $\prod_{j\in N_i} I'_j$ of open intervals.

(b) There are three supported configurations that differ only in x_j, and, since $\gamma_x \neq 0$ (Assumption 7), $\gamma_x x_j$ takes a different value at each. It suffices to show that $\log(p_l^*/p_j^*)$ is decreasing in $\gamma_x x_j$.

Take two of these configurations, L and H, with $\gamma_x x_j$ higher at H. By Lemma 2, $\mathbf{p}^{*,H} \geq \mathbf{p}^{*,L}$ with $p_j^{*,H} > p_j^{*,L}$. The norm is increasing and homogeneous of degree one (Assumption 1 of Boucher et al. (2024)), and, by Definition A.3,

$$\mathbf{p}^{*,L} \leq \mathbf{p}^{*,H} \leq \exp\left(d_{\log,\infty}(\mathbf{p}^{*,H}, \mathbf{p}^{*,L})\right) \mathbf{p}^{*,L},$$

so, for every k with $d_k > 0$,

$$0 \le \log \tilde{y}^e_{-k}(\mathbf{p}^{*,H}; \beta) - \log \tilde{y}^e_{-k}(\mathbf{p}^{*,L}; \beta) \le d_{\log,\infty}(\mathbf{p}^{*,H}, \mathbf{p}^{*,L}).$$

For $k \neq j$ with $d_k > 0$, k's index changes only through its norm, and both norms lie in the interval $[a, b]$ of Assumption 6, which is common to all supported configurations. By the mean value theorem,

$$\log \frac{p^{*,H}_k}{p^{*,L}_k} \le \bar{\varepsilon}^{BR}\, d_{\log,\infty}(\mathbf{p}^{*,H}, \mathbf{p}^{*,L}), \qquad k \neq j,$$

and the left side is zero for isolated k. If the largest increase $d_{\log,\infty}(\mathbf{p}^{*,H}, \mathbf{p}^{*,L})$ occurred at some $k \neq j$, the last display would give $d_{\log,\infty}(\mathbf{p}^{*,H}, \mathbf{p}^{*,L}) \le \bar{\varepsilon}^{BR} d_{\log,\infty}(\mathbf{p}^{*,H}, \mathbf{p}^{*,L})$, so the distance would be zero, contradicting $p^{*,H}_j > p^{*,L}_j$. It therefore occurs at j:

$$d_{\log,\infty}(\mathbf{p}^{*,H}, \mathbf{p}^{*,L}) = \log \frac{p^{*,H}_j}{p^{*,L}_j}.$$

For $k = l$, this gives

$$\log \frac{p^{*,H}_l}{p^{*,L}_l} \le \bar{\varepsilon}^{BR} \log \frac{p^{*,H}_j}{p^{*,L}_j} < \log \frac{p^{*,H}_j}{p^{*,L}_j},$$

so $\log(p^*_l/p^*_j)$ is lower at H than at L. Applied to each pair of the three configurations, this shows that $\log(p^*_l/p^*_j)$ takes three different values. At most two of them can have the same absolute value, so the spread $|\log(p^*_j/p^*_l)|$ takes at least two values. □

Lemma 4 (Power Means)**.** *Let $d_i \ge 2$, let the neighbour probabilities $(p_j)_{j \in N_i}$ range freely over $(0, 1)$, and let β, β' be finite and κ_0, κ_1 constants.*

(a) If $\tilde{y}^e_{-i}(\mathbf{p}; \beta) = \kappa_0 + \kappa_1 \tilde{y}^e_{-i}(\mathbf{p}; \beta')$ for every $(p_j)_{j \in N_i}$ in some non-empty open set $\mathcal{V} \subseteq (0, 1)^{d_i}$, then $\kappa_0 = 0$, $\kappa_1 = 1$ and $\beta = \beta'$.

(b) If $d_i = 2$, with neighbours j and l, and $\tilde{y}^e_{-i}(\mathbf{p}; \beta) = \kappa_1 \tilde{y}^e_{-i}(\mathbf{p}; \beta')$ at two different spreads $|\log(p_j/p_l)|$, then $\kappa_1 = 1$ and $\beta = \beta'$.

Proof. (a) For finite β, the norm is built from sums, exponentials and logarithms of p_j, hence it is real analytic on the connected open set $(0,1)^{d_i}$ (Definition A.9). Therefore, the difference $\tilde{y}^e_{-i}(\mathbf{p};\beta) - \kappa_0 - \kappa_1 \tilde{y}^e_{-i}(\mathbf{p};\beta')$ is also real analytic. By the hypothesis of (a), it is zero on $\mathcal{V}$, and by the identity theorem, it is zero on all of $(0,1)^{d_i}$. Thus,

$$\tilde{y}^e_{-i}(\mathbf{p};\beta) = \kappa_0 + \kappa_1 \tilde{y}^e_{-i}(\mathbf{p};\beta') \qquad \text{for every } (p_j)_{j\in N_i} \in (0,1)^{d_i}.$$

Replace $\mathbf{p}$ by $\lambda\mathbf{p}$, with $\lambda \in (0,1)$. The norm is homogeneous of degree one (Assumption 1 of Boucher et al. (2024)), so

$$\lambda \tilde{y}^e_{-i}(\mathbf{p};\beta) = \kappa_0 + \lambda\kappa_1 \tilde{y}^e_{-i}(\mathbf{p};\beta') = \kappa_0 + \lambda\big(\tilde{y}^e_{-i}(\mathbf{p};\beta) - \kappa_0\big),$$

so $\kappa_0 = \lambda\kappa_0$ and $\kappa_0 = 0$. If all p_j equal some q, both norms equal q, so $\kappa_1 = 1$. If the p_j are not all equal, the norm is strictly increasing in its exponent (Definition A.8), so $\tilde{y}^e_{-i}(\mathbf{p};\beta) = \tilde{y}^e_{-i}(\mathbf{p};\beta')$ gives $\beta = \beta'$.

(b) For finite $\beta \neq 0$, factoring out $(p_j p_l)^{\beta/2}$ gives

$$\frac{p_j^\beta + p_l^\beta}{2} = (p_j p_l)^{\beta/2}\,\frac{(p_j/p_l)^{\beta/2} + (p_l/p_j)^{\beta/2}}{2} = (p_j p_l)^{\beta/2}\cosh\Big(\tfrac{\beta}{2}\log\frac{p_j}{p_l}\Big),$$

and taking the power $1/\beta$ and then logarithms gives, since cosh is even,

$$\log \tilde{y}^e_{-i}(\mathbf{p};\beta) = \tfrac{1}{2}\log(p_j p_l) + \frac{1}{\beta}\log\cosh\Big(\tfrac{\beta}{2}\Big|\log\frac{p_j}{p_l}\Big|\Big).$$

At $\beta = 0$ the last term is read as zero, its limit, which gives the geometric mean. The first term does not depend on β, so the norm depends on β only through the spread $|\log(p_j/p_l)|$. Both norms are positive, so $\kappa_1 > 0$, and taking logarithms of the identity gives, at both values,

$$\log\kappa_1 = \frac{1}{\beta}\log\cosh\Big(\tfrac{\beta}{2}\Big|\log\frac{p_j}{p_l}\Big|\Big) - \frac{1}{\beta'}\log\cosh\Big(\tfrac{\beta'}{2}\Big|\log\frac{p_j}{p_l}\Big|\Big).$$

The derivative of the right side with respect to the spread is

$$\tfrac{1}{2}\Big[\tanh\Big(\tfrac{\beta}{2}\Big|\log\frac{p_j}{p_l}\Big|\Big)-\tanh\Big(\tfrac{\beta'}{2}\Big|\log\frac{p_j}{p_l}\Big|\Big)\Big].$$

If $\beta \neq \beta'$, it has the sign of $\beta-\beta'$ at every positive spread, so the right side is strictly monotone in the spread and cannot take the same value at two different spreads. Hence $\beta=\beta'$, and then $\kappa_1=1$. □

We are now ready to prove Theorem 3.

Proof of Theorem 3:

Suppose two admissible parameter vectors

$$(\xi,\gamma,\delta,\theta_1,\theta_2,\beta),\qquad(\xi',\gamma',\delta',\theta_1',\theta_2',\beta')$$

generate the same conditional choice probabilities $\mathbf{p}^*$ at every supported configuration, the first being the true parameter vector. Since F_η is strictly increasing (Assumption 1), their indices coincide. For $d_i>0$, this gives

$$\begin{aligned}0 = {} & M_i(\xi-\xi')+X_i(\gamma-\gamma')-\frac{\theta_2-\theta_2'}{2}\\ & +(\theta_1 d_i+\theta_2)\tilde{y}^e_{-i}(\mathbf{p}^*;\beta)-(\theta_1' d_i+\theta_2')\tilde{y}^e_{-i}(\mathbf{p}^*;\beta'),\end{aligned}\tag{26}$$

and for $d_i=0$ it gives

$$0=M_i(\xi-\xi')+X_i(\gamma-\gamma')+\delta-\delta'.\tag{27}$$

Neighbour variation. Suppose neighbour variation holds at degree d, with individual i and neighbour j. Then two supported configurations, L and H, differ only in j's payoff-relevant covariate, with $\gamma_x x_j^H>\gamma_x x_j^L$. Since x_j does not enter i's own index, differencing (26) at i between them gives

$$\begin{aligned}&(\theta_1 d+\theta_2)\big[\tilde{y}^e_{-i}(\mathbf{p}^{*,H};\beta)-\tilde{y}^e_{-i}(\mathbf{p}^{*,L};\beta)\big]\\ &\quad=(\theta_1' d+\theta_2')\big[\tilde{y}^e_{-i}(\mathbf{p}^{*,H};\beta')-\tilde{y}^e_{-i}(\mathbf{p}^{*,L};\beta')\big].\end{aligned}\tag{28}$$

By Lemma 2, $\mathbf{p}^{*,H} \geq \mathbf{p}^{*,L}$ with $p_j^{*,H} > p_j^{*,L}$. In the no-peer model this is immediate, since only p_j^* changes. The norm is strictly increasing in each neighbour probability, so both brackets in (28) are strictly positive. We now prove each part separately.

Part (i). Under non-zero peer effects and continuous neighbour variation at an individual i with $d_i \geq 2$, i's choices over the open set of neighbour probabilities from Lemma 3(a) identify β and $\theta_1 d_i + \theta_2$. Neighbour variation at a second degree then separates θ_1 from θ_2, and full rank identifies the remaining coefficients.

Curvature. By (i)(a) and Lemma 3(a), with i's degree and own characteristics held fixed, the neighbour probabilities $(p_j^*)_{j\in N_i}$ take every value in a nonempty open set $\mathcal{V} \subseteq (0,1)^{d_i}$, and (26) holds at each of these configurations. Since $\theta_1 d_i + \theta_2 > 0$ by Assumption 5, dividing (26) by it gives

$$\tilde{y}^e_{-i}(\mathbf{p};\beta) = -\frac{M_i(\xi-\xi') + X_i(\gamma-\gamma') - (\theta_2-\theta_2')/2}{\theta_1 d_i + \theta_2} + \frac{\theta_1' d_i + \theta_2'}{\theta_1 d_i + \theta_2}\,\tilde{y}^e_{-i}(\mathbf{p};\beta')$$

for every $(p_j)_{j\in N_i} \in \mathcal{V}$, where the two fractions are constants. By Lemma 4(a), the first fraction is zero, the second equals one, and $\beta = \beta'$. Hence

$$\theta_1 d_i + \theta_2 = \theta_1' d_i + \theta_2'.$$

Remaining coefficients. Apply (28) at the degree $d \neq d_i$ in (i)(b). Since $\beta = \beta'$, its two brackets coincide and are positive, so

$$\theta_1 d + \theta_2 = \theta_1' d + \theta_2'.$$

Together with the curvature step, and because $d \neq d_i$, this gives $\theta_1 = \theta_1'$ and $\theta_2 = \theta_2'$. The peer terms and the conformity constants in (26) then cancel, so, together with (27), every individual satisfies

$$M_i(\xi-\xi') + X_i(\gamma-\gamma') + \mathbf{1}\{d_i = 0\}(\delta-\delta') = 0, \tag{29}$$

and the full-rank condition in Assumption 7 gives $\xi = \xi'$, $\gamma = \gamma'$ and $\delta = \delta'$. This proves part (i).

Part (ii). Without continuous variation, we use individuals with one and two neighbours. At degree one, the norm equals the neighbour's probability for every β, so neighbour variation at $d = 1$ identifies $\theta_1 + \theta_2$, and full rank among these individuals identifies γ, and ξ up to $\theta_2/2$. At degree two, the different spreads from Lemma 3(b) identify β and $2\theta_1 + \theta_2$, which separates θ_1 from θ_2, and full rank then identifies the remaining coefficients.

Degree-one. Apply (28) at $d = 1$, with the individual i and sole neighbour j in (ii)(a). For every β, $\tilde{y}^e_{-i}(\mathbf{p}; \beta) = p_j$, so (28) reads

$$\left[(\theta_1 + \theta_2) - (\theta'_1 + \theta'_2)\right]\left(p_j^{*,H} - p_j^{*,L}\right) = 0.$$

By Lemma 2, the probability difference is strictly positive. Therefore

$$\theta_1 + \theta_2 = \theta'_1 + \theta'_2. \tag{30}$$

Own coefficients up to the conformity constant. By (30), equation (26) at individuals with $d_i = 1$ reduces to

$$M_i(\xi - \xi') + X_i(\gamma - \gamma') - \frac{\theta_2 - \theta'_2}{2} = 0.$$

Each row of $\mathbf{M}$ has a single one, so $M_i\mathbf{1} = 1$. Since, by (ii)(b), $[\,\mathbf{M}\ \ \mathbf{X}\,]$ restricted to these individuals has full column rank, we obtain

$$\gamma = \gamma', \qquad \xi - \frac{\theta_2}{2}\mathbf{1} = \xi' - \frac{\theta'_2}{2}\mathbf{1}. \tag{31}$$

Hence $M_i\xi + X_i\gamma - \theta_2/2 = M_i\xi' + X_i\gamma' - \theta'_2/2$ for every connected individual, and (26) reduces to

$$(\theta_1 d_i + \theta_2)\tilde{y}^e_{-i}(\mathbf{p}^*; \beta) = (\theta'_1 d_i + \theta'_2)\tilde{y}^e_{-i}(\mathbf{p}^*; \beta'), \qquad d_i > 0. \tag{32}$$

Curvature from the spread. By Lemma 3(b), the individual i in (ii)(c), with neighbours j and l, has different spreads $|\log(p_j^*/p_l^*)|$ at two supported configurations. Since $2\theta_1+\theta_2 > 0$ by Assumption 5, (32) gives, at both,

$$\tilde{y}^e_{-i}(\mathbf{p}^*;\beta) = \frac{2\theta_1' + \theta_2'}{2\theta_1 + \theta_2}\,\tilde{y}^e_{-i}(\mathbf{p}^*;\beta').$$

By Lemma 4(b), $\beta = \beta'$ and $2\theta_1 + \theta_2 = 2\theta_1' + \theta_2'$.

Remaining coefficients. Together with (30), the curvature step gives $\theta_1 = \theta_1'$ and $\theta_2 = \theta_2'$. As in part (i), (29) then holds for every individual, and the full-rank condition in Assumption 7 gives $\xi = \xi'$, $\gamma = \gamma'$ and $\delta = \delta'$. This proves part (ii).

Part (iii). When $\theta_1 = \theta_2 = 0$, i's choice does not respond to a change in a neighbour's covariate, so an observationally equivalent alternative cannot have peer effects either. Full rank then identifies (ξ, γ, δ), while β does not enter the model.

Let $\theta_1 = \theta_2 = 0$, and apply (28) at the degree $d > 0$ in part (iii). Its left side is zero and its right bracket is positive, so

$$\theta_1' d + \theta_2' = 0.$$

By Assumption 5, peer effects satisfy $\theta_1' d + \theta_2' > 0$ at every supported $d > 0$, so $\theta_1' = \theta_2' = 0$. Equations (26) and (27) then reduce to (29) for every individual. Since $[\,\mathbf{M}\;\; \mathbf{1}_{d=0}\;\; \mathbf{X}\,]$ has full column rank by Assumption 7, $\xi = \xi'$, $\gamma = \gamma'$ and $\delta = \delta'$. The norm does not enter any index, so every admissible finite β' is observationally equivalent. This proves part (iii). □

A.4 Optimal targeting

We first state the KKT conditions that we use in the proof for optimal targeting.

Definition A.11 (Karush–Kuhn–Tucker Conditions)**.** Let $V : \mathbb{R}^N \to \mathbb{R}$ be differentiable. At any local maximiser $\boldsymbol{\tau}^*$ of V on $\{\boldsymbol{\tau} \geq \mathbf{0} : \sum_j \tau_j = B\}$ there exists $\lambda \in \mathbb{R}$ with $\partial V/\partial \tau_i = \lambda$ when $\tau_i^* > 0$ and $\partial V/\partial \tau_i \leq \lambda$ when $\tau_i^* = 0$.

We now provide the results on optimal targeting for the generalised norm for the problem discussed in Section 7.

Proof of Proposition 1:

Consider the Lagrangian

$$\mathcal{L} = \frac{1}{N}\sum_j p_j(\boldsymbol{\tau};\beta) - \lambda\left(\sum_j \tau_j - B\right), \qquad \boldsymbol{\tau} \geq \mathbf{0}.$$

By the Karush–Kuhn–Tucker conditions (Definition A.11), at a differentiable optimum there is λ with

$$\frac{1}{N}\sum_j \frac{\partial p_j}{\partial \tau_i} \begin{cases} = \lambda, & \tau_i^* > 0, \\ \leq \lambda, & \tau_i^* = 0. \end{cases}$$

It remains to evaluate the marginal effects $\partial p_j/\partial \tau_i$ at the post-policy equilibrium.

For finite β the CES norm is continuously differentiable at the strictly positive equilibrium probabilities, and the logistic density is $p_i(1-p_i)$. Differentiating each connected agent's equilibrium condition (20) in $\boldsymbol{\tau}$ gives

$$(I - J)\, d\mathbf{p} = \operatorname{diag}\{p_i(1-p_i)\}\, d\boldsymbol{\tau}, \qquad J_{ij} := p_i(1-p_i)(\theta_1 d_i + \theta_2)\,\frac{\partial \widetilde{y}^e_{-i}(\mathbf{p};\beta)}{\partial p_j},$$

with J the best-response Jacobian in $\mathbf{p}$ (zero rows for isolated agents). By Lemma 1, applied with intrinsic utilities $\alpha_i + \tau_i$, $I - J$ is invertible.

The implicit function theorem then gives $\partial \mathbf{p}/\partial \boldsymbol{\tau}^\top = (I-J)^{-1}\operatorname{diag}\{p_i(1-p_i)\}$, so the marginal effect on average adoption of a transfer to agent i is

$$\frac{1}{N}\sum_j \frac{\partial p_j}{\partial \tau_i} = \frac{p_i(1-p_i)}{N}\left[(I-J)^{-\top}\iota\right]_i.$$

Substituting into the first-order conditions gives the stated result. □

B Estimation

We make the following regularity assumption in line with Lin and Hu (2024), required for the asymptotic properties of the NPL estimator and for the proof of Theorem 1 of Lin and Hu (2024).

Assumption 8 (A8). 1. Θ is compact, θ_0 is an interior point of Θ and $\mathcal{P}$ is a compact and convex subset of $(0,1)^N$.

2. $(\theta_0, \mathbf{p}^*)$ is an isolated population NPL fixed point, i.e. it is unique or there is an open ball around it that does not contain any other fixed point in Λ_{0N}.

3. The operator $\Gamma_0(\mathbf{p}) - \mathbf{p}$ has a non-singular Jacobian matrix at $\mathbf{p}^*$.

4. $V_1(\theta_0)$ and $V_2(\theta_0)$ as defined in equation (18) are non-singular matrices with $V_1(\theta_0)$ additionally being negative definite.

5. The family $\{q_i(\theta, \mathbf{p}) : \theta \in \Theta\}$ is a Vapnik-Chervonenkis (VC) class of functions.

6. $\mathbb{E}\left[|q_i(\theta, \mathbf{p}, y)| \,|\, \cdot\right]$ and $|\mathbb{E}\left[q_i(\theta, \mathbf{p}, 1) - q_i(\theta, \mathbf{p}, 0)\right]|$ are bounded by a constant;

7. The class $\{q_i(\theta, \mathbf{p}, y)\}$ is Donsker with respect to the distribution of $(Z(\mathbf{p}))_i$ for $y = 1$ or $y = 0$, with a square-integrable function.

Remark. Parts (1)-(4) are standard regularity conditions that ensure the NPL estimator converges to the true parameter and its asymptotic variance is well defined. Parts (5)-(7) allow the use of uniform Law of Large Numbers and Central Limit Theorem results for network games from Menzel (2016). Together, these assumptions guarantee consistency and asymptotic normality of the NPL estimator in our setting. Since β ranges over the real line, we take its parameter space to be the extended line $[-\infty, \infty]$, whose endpoints are the max- and min-aggregator limits $\beta \to \pm\infty$, to which the CES norm extends continuously. This set is compact, so Part (1) holds. In estimation, β is profiled over a grid that includes these endpoints, as described in Section 5.

Aguirregabiria and Mira (2007) provide an algorithm (NPL algorithm) which is a recursive version of the two step PML. Starting with an initial random guess for the probabilities, the *K-stage* estimator is defined as

$$\hat{\theta}_K = \arg\max_{\theta \in \Theta} Q_N(\theta, \hat{\mathbf{p}}_{K-1}) \tag{33}$$

and the probabilities $\{\hat{\mathbf{p}}_K : K \geq 1\}$ are obtained recursively as

$$\hat{\mathbf{p}}_K = \Gamma(\hat{\mathbf{p}}_{K-1}, \hat{\theta}_K). \tag{34}$$

The algorithm terminates when the difference between two consecutive probabilities is less than some tolerance. However, as shown in later works such as Kasahara and Shimotsu (2012) and Aguirregabiria and Marcoux (2021), the NPL algorithm has finite-sample convergence issues especially in the case of multiple fixed points. While Kasahara and Shimotsu (2012) propose a relaxation of the NPL algorithm by replacing the fixed-point mapping with a Cobb–Douglas-type combination of the mapping and the current probabilities which performs better, it requires knowing the population Jacobian and probabilities else can have similar convergence issues. Other methods include solving the non-linear system of equations subject to the fixed point constraint which requires the calculation of the Jacobian at every step using Newton's method which can be computationally costly.

Thus, for the estimation of parameters and probabilities using the NPL estimator, we employ the spectral algorithm proposed by Aguirregabiria and Marcoux (2021). This approach addresses the convergence issues present in the original NPL algorithm of Aguirregabiria and Mira (2007) and eliminates the need for arbitrary selection of the tuning parameter α as required in Kasahara and Shimotsu (2012), while maintaining computational efficiency. The spectral method (see La Cruz et al. (2006) for technical details) approximates the Jacobian matrix using a scalar value. This scalar is typically computed as a weighted average of the minimum and maximum eigenvalues of the Jacobian of the mapping function $\hat{\phi}(\hat{\mathbf{p}}) = \hat{\Gamma}(\hat{\mathbf{p}}) - \hat{\mathbf{p}}$. These can also be easily estimated

using in-built solvers in R where α_k at every step k is given by:

$$\alpha_k = \frac{\Delta\hat{\mathbf{p}}_k' \Delta\hat{\mathbf{p}}_k}{\Delta\hat{\mathbf{p}}_k' \Delta\hat{\phi}(\hat{\mathbf{p}}_k)}, \tag{35}$$

where $\Delta\hat{\mathbf{p}}_k = \hat{\mathbf{p}}_k - \hat{\mathbf{p}}_{k-1}$ and in every iteration, the conditional choice probabilities are updated using:

$$\hat{\mathbf{p}}_{k+1} = \hat{\mathbf{p}}_k - \alpha_k \hat{\phi}(\hat{\mathbf{p}}_k). \tag{36}$$

C Simulation Set-up and Discussion

We present the set-up of the Monte Carlo simulations in this section. Networks are generated from the Chung–Lu model (Chung and Lu, 2002). Each node i draws a fitness $w_i \sim \text{Gamma}(\text{shape} = k,\ \text{scale} = \mu/k)$ with $\mathbb{E}[w_i] = \mu$, and nodes i and j are linked independently with probability $1 - \exp\big(-w_i w_j / \sum_l w_l\big)$, giving expected degree μ. This is the Norros–Reittu (Norros and Reittu, 2006) form. The degree then has variance $\mu + \mu^2/k$, so smaller k produces more degree heterogeneity and more isolated nodes. We set $\mu = 6$ and $k = 1.5$ to match the mean degree (5.75) and share of isolated households (9.3%) of the microfinance network in Section 6 (Table E.12). For each $N \in \{500, 1000, 2000, 4000\}$ we fix the Monte Carlo number to 500. We draw two covariates independently from $\mathcal{N}(0, 2^2)$ and $\mathcal{N}(0, 1^2)$ with coefficients $\gamma = (-0.2, 0.1)$. The peer parameters are $\theta_1 = 0.15$ and $\theta_2 = 1$, with the isolated-household dummy set to $\delta = -0.7$. Since the strength of identification depends on adoption rates, we try two values of the common intercept, (-1.70 and -1.30). We vary $\beta \in \{-8.3,\ 3.9,\ 22.3\}$, which covers interior values, and also consider the extreme values $\beta \in \{-386,\ 371.8\}$, which fall in the flat-tail regime.

For a given N the network is fixed across replications. In each replication we solve the equilibrium fixed point for the true choice probabilities, draw logistic preference shocks to realise the actions $\mathbf{y}$, and estimate both the CES and the restricted linear model with the algorithm of Section 5. All simulation tables are in Appendix E. For each design they report the median and mean of $\hat{\beta}$, the share of tail solutions, the mean and standard deviation of the peer effect estimates under both norms, the LR test of $\beta = 1$, the bias and MSE under both norms, and the 90% Wald coverage under the CES norm.

Discussion: The adoption rate matters for identification. In the low-adoption design (intercept -1.70), the CES estimate of θ_2 is 0.16 for $\beta = -8.3$ and 0.33 for $\beta = 3.9$ at $N = 500$, and still 0.70 and 0.78 at $N = 4000$, whereas in the high-adoption design it is within 0.07 of its true value at $N = 4000$ in every design. At $N = 4000$, the median $\hat{\beta}$ is -3.1 and -5.8 for $\beta = -8.3$, 3.9 and 4.0 for $\beta = 3.9$, and 21.5 and 22.7 for $\beta = 22.3$, in the low- and high-adoption design respectively. The share of tail solutions falls from

37–69% at $N = 500$ to 10–60% at $N = 4000$, and because of these draws $\hat{\beta}$ has a standard deviation across replications of about 70 to 165, which distorts its mean. The coverage of $\hat{\theta}_1$ rises with N to about 82% at $\beta = 22.3$, stays at 85 to 90% at $\beta = 3.9$, and falls to about 80% at $\beta = -8.3$. When $\beta = 3.9$, imposing linearity matters less: at $N = 4000$ the linear model over-estimates complementarity by about a quarter (0.19 against 0.15), and the LR test rejects linearity in 27 to 53% of draws. When $\beta = 22.3$, by contrast, the test rejects in about half of the draws even at $N = 500$.

For the extreme values (Tables E.7–E.11), at $N = 4000$ the median $\hat{\beta}$ has the right sign but lies only between -6.6 and -2.4 for $\beta = -386$ and between 54 and 70 for $\beta = 371.8$. For $\beta = 371.8$, the CES estimates of θ_1 and θ_2 stay close to their true values, while the linear model misses both by 65% or more, and the LR test rejects $\beta = 1$ in almost every draw. For $\beta = -386$, both models under-estimate θ_1, the linear model by about twice as much, and the test rejects in only 12 to 27% of draws.

D Belief-Aggregated Perceived Social Norms

In this section, we discuss Assumption 3, under which agents evaluate the social norm at their beliefs about peers' actions, so that these beliefs enter utility directly. The standard alternative is expected utility, in which agent i takes the expectation of the utility in (2) over the realised actions of its peers. Since the shocks are independent across agents (Assumption 1), each peer j then adopts independently with probability p_j.

Under expected utility, agent i compares the expected utilities of the two actions. For a given realisation of peers' actions, differencing (2) across $y_i \in \{0, 1\}$ gives

$$\begin{aligned} U_i(1, y_{-i}, \mathbf{G}) - U_i(0, y_{-i}, \mathbf{G}) &= \alpha_i + \theta_1 d_i \, \tilde{y}_{-i}(\beta) - \frac{1}{2}\theta_2 \Big[\big(1 - \tilde{y}_{-i}(\beta)\big)^2 - \tilde{y}_{-i}(\beta)^2 \Big] + \epsilon_i(1) - \epsilon_i(0) \\ &= \alpha_i + \theta_1 d_i \, \tilde{y}_{-i}(\beta) + \theta_2 \Big(\tilde{y}_{-i}(\beta) - \frac{1}{2} \Big) - \eta_i, \end{aligned}$$

since $(1 - \tilde{y}_{-i})^2 - \tilde{y}_{-i}^2 = 1 - 2\tilde{y}_{-i}$. The squared norm enters both utilities with the same coefficient and cancels, exactly as in (7), so the difference is linear in $\tilde{y}_{-i}(\beta)$. Peers' actions are independent of η_i by Assumption 1, so taking expectations over them shows that agent i adopts if and only if

$$\alpha_i + \theta_1 d_i \, \mathbb{E}\big[\tilde{y}_{-i}(\beta)\big] + \theta_2 \Big(\mathbb{E}\big[\tilde{y}_{-i}(\beta)\big] - \frac{1}{2} \Big) \;\geq\; \eta_i,$$

which is (7) with $\tilde{y}^e_{-i}(\beta)$ replaced by $\mathbb{E}[\tilde{y}_{-i}(\beta)]$. Only the mean of the realised norm matters. Its second moment $\mathbb{E}[\tilde{y}_{-i}(\beta)^2]$, which would carry the variance, drops out.

At $\beta = 1$ the two coincide, since the norm is the weighted average of peers' actions and, by linearity of expectations, $\mathbb{E}[\tilde{y}_{-i}(1)] = \sum_j \tilde{g}_{ij} p_j = \tilde{y}^e_{-i}(1)$, the benchmark of Brock and Durlauf (2001a). For $\beta \neq 1$ they differ for any agent with at least two peers. The CES norm is convex in peers' actions for $\beta > 1$ and concave for $\beta < 1$, and $\tilde{y}^e_{-i}(\beta)$ evaluates it at the expected actions p_j, so Jensen's inequality gives

$$\tilde{y}^e_{-i}(\beta) \leq \mathbb{E}\big[\tilde{y}_{-i}(\beta)\big] \;\text{ for } \beta > 1, \qquad \tilde{y}^e_{-i}(\beta) \geq \mathbb{E}\big[\tilde{y}_{-i}(\beta)\big] \;\text{ for } \beta < 1.$$

We adopt the norm at beliefs for three reasons. First, it is tractable. The expected realised norm sums the CES norm over all 2^{d_i} action profiles of i's peers and has no closed form, whereas $\tilde{y}^e_{-i}(\beta)$ is closed-form and differentiable in $\mathbf{p}$ and β, which the equilibrium, identification and estimation results use. Second, it keeps β informative with binary actions. Since $y_j^\beta = y_j$, the realised norm is a power of the share of adopting peers when $\beta > 0$, and when $\beta < 0$ it is one if every peer adopts and zero otherwise. Every $\beta < 0$ then gives the same expected realised norm, $\prod_{j \in N_i} p_j$, so under expected utility negative values of β cannot be told apart. Beliefs lie in $(0, 1)$, so p_j^β varies with β over its whole range and different values of β imply different norms. Third, it has a behavioural foundation. As discussed in the remark following Assumption 3, utility that depends directly on beliefs places the model among psychological games (Geanakoplos et al., 1989; Battigalli and Dufwenberg, 2009), whose equilibrium requires beliefs to be correct. Here β is the rule by which an agent aggregates its beliefs about individual peers into a perceived norm. A misspecified β therefore distorts the norm even when the beliefs are correct, unlike in Bursztyn et al. (2020), where the beliefs themselves are wrong.

E Additional Tables and Figures

E.1 Simulation Results

These tables report the simulation results discussed in Subsection 5.2.

	$\hat{\beta}$			CES		Linear		LR test	
N	Median	Mean	% tail	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_1$	$\hat{\theta}_2$	Statistic	Rejection
				$\beta = -8.3$ **(adoption** = 12.1%**)**					
500	1.19	14.23	69%	0.04	0.16	0.02	0.15	0.72	2%
		(73.58)		(0.38)	(1.89)	(0.24)	(1.80)		
1000	-1.17	15.06	67%	0.12	0.40	0.06	0.33	0.94	5%
		(94.43)		(0.26)	(1.76)	(0.15)	(1.67)		
2000	-1.62	14.05	61%	0.12	0.58	0.07	0.53	1.02	8%
		(99.34)		(0.17)	(1.66)	(0.10)	(1.53)		
4000	-3.05	2.63	60%	0.13	0.70	0.07	0.58	1.39	15%
		(101.26)		(0.13)	(1.44)	(0.07)	(1.31)		
				$\beta = 3.9$ **(adoption** = 14.1%**)**					
500	0.99	16.17	58%	0.20	0.33	0.16	0.21	0.86	5%
		(95.35)		(0.30)	(1.78)	(0.17)	(1.78)		
1000	1.65	18.87	43%	0.22	0.56	0.19	0.52	1.16	11%
		(103.61)		(0.19)	(1.59)	(0.09)	(1.52)		
2000	3.01	21.16	39%	0.18	0.73	0.18	0.67	1.22	12%
		(102.28)		(0.13)	(1.41)	(0.06)	(1.34)		
4000	3.93	19.31	28%	0.17	0.78	0.19	0.85	2.17	27%
		(107.83)		(0.10)	(1.08)	(0.04)	(1.03)		
				$\beta = 22.3$ **(adoption** = 26.3%**)**					
500	13.41	46.02	43%	0.21	0.95	0.32	0.57	3.84	57%
		(90.91)		(0.12)	(0.92)	(0.05)	(0.98)		
1000	17.01	73.33	38%	0.18	1.02	0.30	0.59	7.51	83%
		(124.32)		(0.06)	(0.60)	(0.03)	(0.67)		
2000	15.74	76.04	27%	0.16	1.04	0.29	0.66	11.85	94%
		(137.95)		(0.04)	(0.48)	(0.02)	(0.53)		
4000	21.46	91.64	28%	0.16	0.99	0.29	0.56	26.66	100%
		(143.76)		(0.02)	(0.29)	(0.01)	(0.35)		

Notes: Peer estimates are means over 500 replications, with standard deviations across replications in parentheses. We report the median and mean $\hat{\beta}$, since extreme tail draws distort the mean. Likelihood Ratio (LR) test columns report the mean test statistic and the rejection rate for $H_0 : \beta = 1$ at 10% using χ^2_1. % tail is the share flagged as tail solutions. Adoption is the mean share of adopting households in the simulated data.

Table E.1: Simulation Results with Intercept = −1.70: Point Estimates for CES Elasticity and Peer Effects (CES vs. Linear Norm), LR Test Statistic and Rejection Rates.

	$\hat{\beta}$			CES		Linear		LR test	
N	Median	Mean	% tail	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_1$	$\hat{\theta}_2$	Statistic	Rejection
$\beta = -8.3$ **(adoption** = 18.3%**)**									
500	0.00	18.44	64%	0.10	0.37	0.05	0.26	0.90	5%
		(97.91)		(0.22)	(1.78)	(0.13)	(1.67)		
1000	-2.56	1.00	56%	0.14	0.71	0.08	0.56	1.34	15%
		(101.13)		(0.15)	(1.48)	(0.08)	(1.42)		
2000	-3.93	1.43	54%	0.13	0.87	0.07	0.80	1.60	19%
		(113.71)		(0.11)	(1.29)	(0.05)	(1.18)		
4000	-5.83	-21.69	48%	0.14	0.93	0.08	0.83	2.47	32%
		(127.77)		(0.09)	(1.00)	(0.04)	(0.97)		
$\beta = 3.9$ **(adoption** = 24.6%**)**									
500	2.36	24.84	40%	0.19	0.59	0.18	0.59	1.27	13%
		(120.66)		(0.14)	(1.48)	(0.06)	(1.34)		
1000	3.74	24.82	30%	0.17	0.90	0.19	0.88	1.89	26%
		(95.79)		(0.09)	(0.98)	(0.03)	(0.93)		
2000	3.97	31.71	19%	0.15	0.97	0.18	0.93	2.00	27%
		(113.28)		(0.06)	(0.79)	(0.03)	(0.79)		
4000	4.01	30.33	10%	0.15	0.97	0.19	0.98	4.11	53%
		(95.83)		(0.04)	(0.48)	(0.02)	(0.52)		
$\beta = 22.3$ **(adoption** = 37.7%**)**									
500	12.51	58.49	37%	0.20	0.95	0.25	0.55	3.23	50%
		(107.67)		(0.09)	(0.89)	(0.04)	(0.89)		
1000	16.47	81.64	37%	0.17	0.98	0.25	0.51	5.84	75%
		(129.76)		(0.06)	(0.61)	(0.03)	(0.58)		
2000	19.38	91.21	32%	0.16	1.03	0.24	0.54	9.51	91%
		(145.64)		(0.03)	(0.48)	(0.02)	(0.44)		
4000	22.70	105.69	32%	0.15	0.99	0.24	0.51	19.20	98%
		(164.55)		(0.02)	(0.30)	(0.01)	(0.30)		

Notes: Peer estimates are means over 500 replications, with standard deviations across replications in parentheses. We report the median and mean $\hat{\beta}$, since extreme tail draws distort the mean. Likelihood Ratio (LR) test columns report the mean test statistic and the rejection rate for $H_0 : \beta = 1$ at 10% using χ^2_1. % tail is the share flagged as tail solutions. Adoption is the mean share of adopting households in the simulated data.

Table E.2: Simulation Results with Intercept = −1.30: Point Estimates for CES Elasticity and Peer Effects (CES vs. Linear Norm), LR Test Statistic and Rejection Rates.

	θ_1				θ_2			
	CES		Linear		CES		Linear	
N	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
				$\beta = -8.3$ **(adoption** = 12.1%**)**				
500	-0.11	0.15	-0.13	0.08	-0.84	4.26	-0.85	3.96
1000	-0.03	0.07	-0.09	0.03	-0.60	3.44	-0.67	3.22
2000	-0.03	0.03	-0.08	0.02	-0.42	2.93	-0.47	2.56
4000	-0.02	0.02	-0.08	0.01	-0.30	2.16	-0.42	1.89
				$\beta = 3.9$ **(adoption** = 14.1%**)**				
500	0.05	0.09	0.01	0.03	-0.67	3.62	-0.79	3.79
1000	0.07	0.04	0.04	0.01	-0.44	2.71	-0.48	2.53
2000	0.03	0.02	0.03	0.01	-0.27	2.06	-0.33	1.90
4000	0.02	0.01	0.04	0.00	-0.22	1.22	-0.15	1.08
				$\beta = 22.3$ **(adoption** = 26.3%**)**				
500	0.06	0.02	0.17	0.03	-0.05	0.86	-0.43	1.15
1000	0.03	0.00	0.15	0.02	0.02	0.36	-0.41	0.62
2000	0.01	0.00	0.14	0.02	0.04	0.23	-0.34	0.40
4000	0.01	0.00	0.14	0.02	-0.01	0.09	-0.44	0.32

Notes: Bias = mean$(\hat{\theta} - \theta)$ and MSE = mean$((\hat{\theta} - \theta)^2)$ over 500 replications. True $\theta_1 = 0.15$ and $\theta_2 = 1$. Adoption is the mean share of adopting households in the simulated data.

Table E.3: Simulation Results with Intercept = −1.70: Bias and mean squared error of the peer-effect estimates under the CES and linear norms.

	θ_1				θ_2			
	CES		Linear		CES		Linear	
N	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
			$\beta = -8.3$	**(adoption** = 18.3%**)**				
500	-0.05	0.05	-0.10	0.03	-0.63	3.54	-0.74	3.31
1000	-0.01	0.02	-0.07	0.01	-0.29	2.27	-0.44	2.21
2000	-0.02	0.01	-0.08	0.01	-0.13	1.67	-0.20	1.43
4000	-0.01	0.01	-0.07	0.01	-0.07	1.01	-0.17	0.97
			$\beta = 3.9$	**(adoption** = 24.6%**)**				
500	0.04	0.02	0.03	0.00	-0.41	2.35	-0.41	1.97
1000	0.02	0.01	0.04	0.00	-0.10	0.97	-0.12	0.88
2000	0.00	0.00	0.03	0.00	-0.03	0.63	-0.07	0.63
4000	-0.00	0.00	0.04	0.00	-0.03	0.23	-0.02	0.27
			$\beta = 22.3$	**(adoption** = 37.7%**)**				
500	0.05	0.01	0.10	0.01	-0.05	0.80	-0.45	0.99
1000	0.02	0.00	0.10	0.01	-0.02	0.38	-0.49	0.58
2000	0.01	0.00	0.09	0.01	0.03	0.23	-0.46	0.40
4000	0.00	0.00	0.09	0.01	-0.01	0.09	-0.49	0.33

Notes: Bias = mean($\hat{\theta} - \theta$) and MSE = mean($(\hat{\theta} - \theta)^2$) over 500 replications. True $\theta_1 = 0.15$ and $\theta_2 = 1$. Adoption is the mean share of adopting households in the simulated data.

Table E.4: Simulation Results with Intercept = −1.30: Bias and mean squared error of the peer-effect estimates under the CES and linear norms.

		θ_1			θ_2		
N	% tail	All	Interior	Tail	All	Interior	Tail
			$\beta = -8.3$ **(adoption** = 12.1%**)**				
500	69%	92%	95%	90%	99%	100%	98%
1000	67%	88%	88%	88%	97%	99%	96%
2000	61%	83%	85%	81%	96%	99%	93%
4000	60%	72%	80%	66%	88%	92%	86%
			$\beta = 3.9$ **(adoption** = 14.1%**)**				
500	58%	90%	90%	90%	97%	100%	96%
1000	43%	83%	86%	80%	97%	99%	93%
2000	39%	81%	89%	69%	94%	98%	87%
4000	28%	73%	88%	35%	89%	94%	74%
			$\beta = 22.3$ **(adoption** = 26.3%**)**				
500	43%	79%	70%	91%	94%	96%	91%
1000	38%	80%	77%	86%	96%	96%	97%
2000	27%	75%	76%	74%	91%	91%	90%
4000	28%	74%	82%	56%	91%	91%	92%

Notes: Coverage over 500 replications, split by the estimator's tail flag. % tail is the share flagged as tail solutions where the profile likelihood is flat in β. Adoption is the mean share of adopting households in the simulated data.

Table E.5: Simulation Results with Intercept = −1.70: 90% Wald Coverage of Peer Effect Estimates under CES Norm.

		θ_1			θ_2		
N	% tail	All	Interior	Tail	All	Interior	Tail
			$\beta = -8.3$ **(adoption** = 18.3%**)**				
500	64%	85%	92%	81%	96%	99%	94%
1000	56%	81%	86%	77%	93%	97%	90%
2000	54%	71%	79%	65%	89%	93%	85%
4000	48%	72%	78%	66%	89%	94%	83%
			$\beta = 3.9$ **(adoption** = 24.6%**)**				
500	40%	81%	88%	70%	91%	96%	84%
1000	30%	71%	85%	40%	93%	97%	85%
2000	19%	74%	86%	18%	92%	96%	76%
4000	10%	79%	86%	15%	88%	91%	62%
			$\beta = 22.3$ **(adoption** = 37.7%**)**				
500	37%	78%	72%	89%	94%	93%	96%
1000	37%	81%	79%	84%	93%	92%	95%
2000	32%	78%	82%	70%	90%	91%	89%
4000	32%	72%	81%	53%	91%	90%	93%

Notes: Coverage over 500 replications, split by the estimator's tail flag. % tail is the share flagged as tail solutions where the profile likelihood is flat in β. Adoption is the mean share of adopting households in the simulated data.

Table E.6: Simulation Results with Intercept = −1.30: 90% Wald Coverage of Peer Effect Estimates under CES Norm.

	$\hat{\beta}$		CES		Linear		LR test	
N	Median	Mean	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_1$	$\hat{\theta}_2$	Statistic	Rejection
Intercept = −1.70								
β = −386.0 **(adoption** = 11.8%**)**								
500	1.37	16.29	-0.01	0.12	-0.02	0.16	0.71	2%
		(80.29)	(0.40)	(1.89)	(0.26)	(1.82)		
1000	-0.95	12.76	0.08	0.47	0.03	0.37	0.89	5%
		(90.56)	(0.27)	(1.74)	(0.16)	(1.68)		
2000	-1.27	15.11	0.07	0.60	0.03	0.51	0.93	5%
		(95.19)	(0.18)	(1.69)	(0.11)	(1.57)		
4000	-2.41	7.14	0.09	0.67	0.04	0.56	1.23	12%
		(116.13)	(0.13)	(1.49)	(0.08)	(1.34)		
β = 371.8 **(adoption** = 31.6%**)**								
500	15.90	61.81	0.23	0.91	0.33	0.37	4.35	62%
		(107.10)	(0.10)	(0.86)	(0.04)	(0.87)		
1000	24.09	105.22	0.19	1.00	0.31	0.37	8.52	88%
		(153.18)	(0.05)	(0.58)	(0.03)	(0.58)		
2000	38.54	133.98	0.17	1.01	0.30	0.37	14.82	99%
		(171.59)	(0.03)	(0.42)	(0.02)	(0.41)		
4000	54.31	174.38	0.17	0.97	0.30	0.32	30.44	100%
		(192.00)	(0.02)	(0.28)	(0.01)	(0.30)		
Intercept = −1.30								
β = −386.0 **(adoption** = 17.6%**)**								
500	0.47	20.93	0.06	0.36	0.02	0.25	0.85	5%
		(97.68)	(0.24)	(1.80)	(0.15)	(1.72)		
1000	-1.45	9.06	0.09	0.69	0.04	0.56	1.13	9%
		(95.48)	(0.15)	(1.54)	(0.09)	(1.46)		
2000	-2.33	3.78	0.08	0.83	0.04	0.74	1.38	16%
		(117.73)	(0.11)	(1.39)	(0.06)	(1.26)		
4000	-6.55	-17.64	0.10	0.97	0.05	0.81	2.05	27%
		(118.75)	(0.08)	(1.07)	(0.04)	(1.02)		
β = 371.8 **(adoption** = 41.9%**)**								
500	20.44	84.23	0.21	0.93	0.27	0.41	3.52	52%
		(126.48)	(0.08)	(0.90)	(0.04)	(0.83)		
1000	34.56	120.03	0.18	0.97	0.26	0.35	6.39	80%
		(156.14)	(0.05)	(0.59)	(0.03)	(0.54)		
2000	46.45	143.58	0.17	1.01	0.26	0.36	11.11	95%
		(171.19)	(0.03)	(0.47)	(0.02)	(0.40)		
4000	70.22	176.18	0.16	0.99	0.26	0.35	21.49	99%
		(186.75)	(0.02)	(0.30)	(0.01)	(0.29)		

Notes: Peer estimates are means over 500 replications, with standard deviations across replications in parentheses. We report the median and mean $\hat{\beta}$, since extreme tail draws distort the mean. Likelihood Ratio (LR) test columns report the mean test statistic and the rejection rate for $H_0 : \beta = 1$ at 10% using χ^2_1. Adoption is the mean share of adopting households in the simulated data.

Table E.7: Simulation Results for Extreme β: Point Estimates for CES Elasticity and Peer Effects (CES vs. Linear Norm), LR Test Statistic and Rejection Rates.

	θ_1				θ_2			
	CES		Linear		CES		Linear	
N	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
			$\beta = -386.0$	**(adoption** = 11.8%**)**				
500	-0.16	0.18	-0.17	0.10	-0.88	4.34	-0.84	4.02
1000	-0.07	0.08	-0.12	0.04	-0.53	3.32	-0.63	3.20
2000	-0.08	0.04	-0.12	0.02	-0.40	3.01	-0.49	2.70
4000	-0.06	0.02	-0.11	0.02	-0.33	2.33	-0.44	1.99
			$\beta = 371.8$	**(adoption** = 31.6%**)**				
500	0.08	0.02	0.18	0.03	-0.09	0.74	-0.63	1.16
1000	0.04	0.00	0.16	0.03	0.00	0.33	-0.63	0.74
2000	0.02	0.00	0.15	0.02	0.01	0.18	-0.63	0.56
4000	0.02	0.00	0.15	0.02	-0.03	0.08	-0.68	0.56

Notes: Bias = mean$(\hat{\theta} - \theta)$ and MSE = mean$((\hat{\theta} - \theta)^2)$ over 500 replications. True $\theta_1 = 0.15$ and $\theta_2 = 1$. Adoption is the mean share of adopting households in the simulated data.

Table E.8: Simulation Results with Intercept = −1.70 and Extreme β: Bias and mean squared error of the peer-effect estimates under the CES and linear norms.

	θ_1				θ_2			
	CES		Linear		CES		Linear	
N	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
			$\beta = -386.0$	**(adoption** = 17.6%**)**				
500	-0.09	0.06	-0.13	0.04	-0.64	3.64	-0.75	3.51
1000	-0.06	0.03	-0.11	0.02	-0.31	2.46	-0.44	2.31
2000	-0.07	0.02	-0.11	0.02	-0.17	1.95	-0.26	1.65
4000	-0.05	0.01	-0.10	0.01	-0.03	1.14	-0.19	1.07
			$\beta = 371.8$	**(adoption** = 41.9%**)**				
500	0.06	0.01	0.12	0.02	-0.07	0.82	-0.59	1.04
1000	0.03	0.00	0.11	0.01	-0.03	0.34	-0.65	0.71
2000	0.02	0.00	0.11	0.01	0.01	0.22	-0.64	0.56
4000	0.01	0.00	0.11	0.01	-0.01	0.09	-0.65	0.50

Notes: Bias = mean$(\hat{\theta} - \theta)$ and MSE = mean$((\hat{\theta} - \theta)^2)$ over 500 replications. True $\theta_1 = 0.15$ and $\theta_2 = 1$. Adoption is the mean share of adopting households in the simulated data.

Table E.9: Simulation Results with Intercept = −1.30 and Extreme β: Bias and mean squared error of the peer-effect estimates under the CES and linear norms.

		θ_1			θ_2		
N	% tail	All	Interior	Tail	All	Interior	Tail
		$\beta = -386.0$ **(adoption** = 11.8%**)**					
500	68%	92%	94%	91%	99%	100%	99%
1000	70%	87%	87%	86%	99%	100%	98%
2000	63%	80%	86%	77%	97%	99%	95%
4000	62%	69%	70%	68%	90%	94%	88%
		$\beta = 371.8$ **(adoption** = 31.6%**)**					
500	44%	69%	54%	89%	92%	92%	91%
1000	44%	75%	61%	93%	94%	93%	95%
2000	46%	75%	59%	95%	93%	94%	91%
4000	51%	77%	62%	91%	90%	88%	92%

Notes: Coverage over 500 replications, split by the estimator's tail flag. % tail is the share flagged as tail solutions where the profile likelihood is flat in β. Adoption is the mean share of adopting households in the simulated data.

Table E.10: Simulation Results with Intercept = −1.70 and Extreme β: 90% Wald Coverage of Peer Effect Estimates under CES Norm.

		θ_1			θ_2		
N	% tail	All	Interior	Tail	All	Interior	Tail
		$\beta = -386.0$ **(adoption** = 17.6%**)**					
500	68%	83%	90%	79%	98%	99%	97%
1000	59%	75%	78%	73%	92%	96%	90%
2000	57%	64%	64%	63%	89%	93%	85%
4000	57%	66%	59%	71%	89%	92%	86%
		$\beta = 371.8$ **(adoption** = 41.9%**)**					
500	47%	72%	58%	88%	92%	91%	94%
1000	52%	76%	58%	92%	94%	91%	96%
2000	50%	77%	60%	93%	92%	91%	92%
4000	53%	76%	62%	88%	89%	86%	92%

Notes: Coverage over 500 replications, split by the estimator's tail flag. % tail is the share flagged as tail solutions where the profile likelihood is flat in β. Adoption is the mean share of adopting households in the simulated data.

Table E.11: Simulation Results with Intercept = −1.30 and Extreme β: 90% Wald Coverage of Peer Effect Estimates under CES Norm.

E.2 Microfinance Estimation Tables

In this section we present the tables related to the empirical exercise using the microfinance dataset of Banerjee et al. (2013). We report the summary statistics, the full pooled estimates (with and without the isolated-household dummy), the segment composition and the choice of wealth proxy, the full segment estimates for both partitions, and the subsidy-recipient characteristics for the illustrative heterogenous and rich segments.

Variable	Mean	Std. dev.	Min	Max
Outcome				
Adoption (took up a BSS loan)	0.17	0.38	0	1
Social network				
Network degree	5.75	4.73	0	48
Isolated household (deg = 0)	0.09	0.29	0	1
Household characteristics				
Number of rooms	2.44	1.27	0	18
Number of beds	0.76	1.11	0	20
Non-Hindu head of household	0.05	0.22	0	1
Scheduled caste or tribe	0.33	0.47	0	1
Owns a latrine	0.26	0.44	0	1
Village leader household	0.12	0.33	0	1
Rents (rather than owns) home	0.06	0.23	0	1

Notes: 6,733 households in 29 villages from the 2006 baseline survey of Banerjee et al. (2013), measured before the BSS microfinance programme entered any village. Adoption is subsequent participation in the programme. Network degree is the number of households linked by a reported visiting tie; binary indicators take values in $\{0, 1\}$, so their standard deviation is $\sqrt{\bar{p}(1-\bar{p})}$. Covariate coding follows the household survey. All controls enter the private payoff alongside village fixed effects.

Table E.12: Summary statistics for the Indian Microfinance Dataset of Banerjee et al. (2013).

	With isolated dummy		Without isolated dummy	
	Non-linear	Linear	Non-linear	Linear
Endogenous peer effects				
Strategic complementarity (θ_1)	0.053**	0.180***	0.417***	0.245***
	(0.023)	(0.032)	(0.068)	(0.028)
Conformity (θ_2)	2.064***	1.538***	-0.725**	-0.182
	(0.445)	(0.501)	(0.351)	(0.334)
CES elasticity (β)	491.5†	1	$\to -\infty$†	1
	(722.2)			
Own effects				
Number of rooms	-0.064*	-0.063**	-0.065**	-0.066**
	(0.033)	(0.032)	(0.033)	(0.033)
Number of beds	-0.087**	-0.082*	-0.085*	-0.088**
	(0.044)	(0.042)	(0.043)	(0.043)
Non-Hindu head of household	0.869***	0.877***	1.221***	1.147***
	(0.149)	(0.154)	(0.163)	(0.154)
Scheduled caste or tribe	0.375***	0.348***	0.465***	0.444***
	(0.069)	(0.072)	(0.077)	(0.077)
Owns a latrine	-0.365***	-0.348***	-0.381***	-0.375***
	(0.091)	(0.090)	(0.093)	(0.093)
Village leader household	0.562***	0.567***	0.588***	0.580***
	(0.095)	(0.094)	(0.097)	(0.097)
Rents (rather than owns) home	0.026	0.028	0.019	0.010
	(0.143)	(0.138)	(0.145)	(0.143)
Isolated-household dummy	-1.033***	-0.894***	–	–
	(0.189)	(0.214)		
Village fixed effects	Yes	Yes	Yes	Yes
Num. obs.	6,733	6,733	6,733	6,733
LR test $\beta = 1$ (stat, p)	4.63 (0.031)		8.74 (0.003)	

Notes: All 29 villages pooled (6,733 households) with village fixed effects. CES columns leave the CES curvature β free; linear columns impose $\beta = 1$. † marks a flat-tail fit, where β is numerically indistinguishable with a flat profile likelihood and the presented value is the refined optimum in the flat region. θ_1 is complementarity and θ_2 conformity. Standard errors in parentheses, stars are Wald tests against zero, $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The LR statistic tests $\beta = 1$ against a χ^2_1 reference. CES SEs without the isolated-household dummy condition on the fixed minimum norm; uncertainty about β and norm selection is excluded.

Table E.13: Microfinance pooled sample estimates: With and Without the isolated-household dummy, under the CES and linear ($\beta = 1$) norms.

Segment	Villages	N	Caste min. share	Beds	Latrine	Rooms	Avg. degree	Adoption
Caste × Beds partition								
Het/P	8	1,616	0.514	0.529	0.190	2.271	5.55	19.0%
Het/R	6	1,387	0.397	0.865	0.252	2.539	6.11	20.3%
Hom/P	7	1,728	0.230	0.490	0.282	2.396	5.51	15.0%
Hom/R	8	2,002	0.223	1.110	0.308	2.557	5.86	16.0%
Caste × Latrine partition								
Het/P	8	1,691	0.480	0.655	0.157	2.428	6.26	16.8%
Het/R	6	1,312	0.433	0.723	0.298	2.352	5.24	23.2%
Hom/P	7	1,718	0.228	0.767	0.175	2.480	6.00	12.4%
Hom/R	8	2,012	0.224	0.870	0.399	2.485	5.44	18.3%

Notes: Het/Hom = above/below the median village caste-minority share (low minority share = homogeneous); P/R = below/above the median on the wealth variable naming the partition. Columns 4–7 are village-population means of the household covariates; "Avg. degree" is mean network degree and "Adoption" the microfinance take-up rate.

Table E.14: Segment composition for the Caste × Beds and Caste × Latrine partitions.

Partition	Min. centroid dist.	Mean centroid dist.	Min. dist. (all X)	Smallest Seg.
Caste × Beds	1.26	1.72	1.37	1,387
Caste × Latrine	1.19	1.64	1.51	1,312

Notes: Centroid distances are between segment means, standardised. Larger is better separated. The two proxies are close with beds giving the larger minimum and mean separation on the segmenting variables and the larger smallest segment, latrine the larger minimum separation over the full covariate vector.

Table E.15: Separation of the two candidate wealth proxies.

	Het/P		Het/R		Hom/P		Hom/R	
	NL	L	NL	L	NL	L	NL	L
Endogenous peer effects								
Strategic complementarity (θ_1)	0.287**	0.150**	0.013	0.136**	0.506***	0.209***	0.076	0.230***
	(0.125)	(0.067)	(0.038)	(0.061)	(0.176)	(0.065)	(0.064)	(0.062)
Conformity (θ_2)	-0.317	0.638	2.680***	1.488	-1.192	0.224	1.742	1.390
	(1.565)	(1.337)	(0.833)	(0.945)	(1.045)	(1.068)	(1.589)	(1.128)
CES elasticity (β)	-83.1†	1	491.5†	1	-296.7†	1	787.7†	1
	(126.2)		(1182.7)		(436.7)		(13855.2)	
Own effects								
Number of rooms	-0.058	-0.058	-0.063	-0.061	-0.061	-0.058	-0.035	-0.038
	(0.069)	(0.069)	(0.060)	(0.060)	(0.086)	(0.086)	(0.057)	(0.054)
Number of beds	-0.182	-0.176	-0.069	-0.072	-0.242*	-0.250*	-0.057	-0.055
	(0.116)	(0.114)	(0.087)	(0.084)	(0.130)	(0.130)	(0.058)	(0.054)
Non-Hindu head of household	1.085**	0.950**	0.645***	0.700***	2.110***	1.824***	0.640**	0.633**
	(0.441)	(0.396)	(0.201)	(0.252)	(0.439)	(0.519)	(0.321)	(0.271)
Scheduled caste or tribe	0.117	0.110	0.581***	0.567***	0.943***	0.842***	-0.013	-0.023
	(0.152)	(0.141)	(0.155)	(0.176)	(0.164)	(0.183)	(0.145)	(0.135)
Owns a latrine	-0.153	-0.148	-0.475**	-0.451**	-0.485**	-0.481**	-0.343**	-0.331**
	(0.198)	(0.194)	(0.200)	(0.197)	(0.227)	(0.227)	(0.148)	(0.146)
Village leader household	0.485**	0.489**	0.445**	0.464**	0.657***	0.651***	0.701***	0.684***
	(0.193)	(0.191)	(0.208)	(0.206)	(0.211)	(0.209)	(0.166)	(0.171)
Rents (rather than owns) home	0.185	0.169	0.002	0.009	-0.374	-0.328	0.174	0.188
	(0.303)	(0.299)	(0.277)	(0.281)	(0.348)	(0.333)	(0.243)	(0.243)
Isolated-household dummy	-0.510	-0.840*	-1.135***	-0.870**	0.323	-0.229	-1.061**	-0.946**
	(0.517)	(0.443)	(0.340)	(0.403)	(0.494)	(0.450)	(0.497)	(0.447)
Village fixed effects	Yes		Yes		Yes		Yes	
Num. obs.	1,616		1,387		1,728		2,002	
LR test $\beta = 1$ (stat, p)	1.81 (0.179)		3.39 (0.066)		1.56 (0.212)		0.75 (0.387)	

Notes: Segments are Het/Hom (heterogeneous vs homogeneous on caste-minority share) × P/R (poor vs rich), split at village-level medians of the Caste × Beds partition. NL leaves the CES curvature β free; L imposes $\beta = 1$. † marks a flat-tail fit (β is numerically indistinguishable; the value is the refined estimate in the plateau). θ_1 is complementarity, θ_2 conformity. Standard errors in parentheses where computed; stars are Wald tests vs zero, $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The LR statistic tests $\beta = 1$ against a χ^2_1 reference. A dash marks a control dropped for lack of within-segment variation.

Table E.16: Segment estimates Caste × Beds partition: CES (NL, β free) and linear (L, $\beta = 1$) with the isolated-household dummy.

	Het/P		Het/R		Hom/P		Hom/R	
	NL	L	NL	L	NL	L	NL	L
Endogenous peer effects								
Strategic complementarity (θ_1)	0.155	0.105*	0.321**	0.142**	0.351*	0.168**	0.232***	0.248***
	(0.135)	(0.060)	(0.155)	(0.066)	(0.209)	(0.075)	(0.070)	(0.057)
Conformity (θ_2)	2.747***	3.083***	-0.057	1.019	4.035*	3.376***	0.779	0.671
	(1.062)	(0.837)	(1.997)	(1.172)	(2.155)	(1.190)	(1.126)	(0.895)
CES elasticity (β)	-0.97	1	-282.8†	1	-4.05	1	2.00	1
	(4.3)		(756.4)		(5.9)		(2.1)	
Own effects								
Number of rooms	-0.010	-0.010	-0.123*	-0.116*	0.016	0.017	-0.114*	-0.114*
	(0.060)	(0.060)	(0.067)	(0.067)	(0.067)	(0.069)	(0.058)	(0.058)
Number of beds	-0.143*	-0.140*	-0.075	-0.074	-0.021	-0.021	-0.083	-0.082
	(0.084)	(0.084)	(0.115)	(0.111)	(0.062)	(0.063)	(0.069)	(0.068)
Non-Hindu head of household	1.354**	1.065***	0.734**	0.638**	–	–	1.063***	1.082***
	(0.644)	(0.323)	(0.351)	(0.261)			(0.411)	(0.327)
Scheduled caste or tribe	0.303**	0.290**	0.281	0.254*	0.384**	0.377**	0.291**	0.293**
	(0.153)	(0.127)	(0.189)	(0.153)	(0.159)	(0.148)	(0.146)	(0.139)
Owns a latrine	-0.268	-0.276	-0.340*	-0.325*	-0.459*	-0.537**	-0.347**	-0.347**
	(0.222)	(0.225)	(0.186)	(0.172)	(0.241)	(0.233)	(0.144)	(0.144)
Village leader household	0.422**	0.420**	0.546***	0.542***	0.653***	0.629***	0.649***	0.650***
	(0.195)	(0.189)	(0.205)	(0.203)	(0.216)	(0.204)	(0.162)	(0.166)
Rents (rather than owns) home	0.463	0.436	-0.220	-0.220	0.438	0.394	-0.162	-0.161
	(0.311)	(0.290)	(0.279)	(0.273)	(0.396)	(0.391)	(0.220)	(0.218)
Isolated-household dummy	-1.837***	-1.957***	-0.245	-0.592	-1.713*	-1.508***	-0.572	-0.547
	(0.503)	(0.453)	(0.502)	(0.377)	(0.900)	(0.568)	(0.410)	(0.370)
Village fixed effects	Yes		Yes		Yes		Yes	
Num. obs.	1,691		1,312		1,718		2,012	
LR test $\beta = 1$ (stat, p)	0.67 (0.412)		3.16 (0.075)		5.23 (0.022)		0.35 (0.556)	

Notes: Segments are Het/Hom (heterogeneous vs homogeneous on caste-minority share) × P/R (poor vs rich), split at village-level medians of the Caste × Latrine partition. NL leaves the CES curvature β free; L imposes $\beta = 1$. † marks a flat-tail fit (β is numerically indistinguishable; the value is the refined estimate in the plateau). θ_1 is complementarity, θ_2 conformity. Standard errors in parentheses where computed; stars are Wald tests vs zero, $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The LR statistic tests $\beta = 1$ against a χ^2_1 reference. A dash marks a control dropped for lack of within-segment variation.

Table E.17: Segment estimates Caste × Latrine partition: CES (NL, β free) and linear (L, $\beta = 1$) with the isolated-household dummy.

	Het/P		Het/R		Hom/P		Hom/R	
	NL	L	NL	L	NL	L	NL	L
Endogenous peer effects								
Strategic complementarity (θ_1)	0.316**	0.207***	0.353**	0.208***	0.354	0.226***	0.286***	0.275***
	(0.144)	(0.061)	(0.162)	(0.050)	(0.286)	(0.056)	(0.077)	(0.060)
Conformity (θ_2)	-1.354	-1.179	-0.854	0.255	-0.557	-0.189	-0.781	-0.718
	(0.973)	(0.861)	(0.808)	(0.626)	(0.963)	(0.634)	(0.840)	(0.722)
CES elasticity (β)	-491.5†	1	-80.6†	1	-5.06	1	0.00	1
	(531.4)		(109.9)		(25.1)		(2.4)	
Own effects								
Number of rooms	-0.061	-0.064	-0.064	-0.069	-0.059	-0.056	-0.040	-0.040
	(0.071)	(0.070)	(0.061)	(0.061)	(0.086)	(0.086)	(0.057)	(0.056)
Number of beds	-0.183	-0.178	-0.078	-0.072	-0.244*	-0.252*	-0.057	-0.057
	(0.116)	(0.114)	(0.086)	(0.085)	(0.138)	(0.130)	(0.056)	(0.056)
Non-Hindu head of household	1.216***	1.221***	0.880***	0.771***	2.030***	1.951***	0.894***	0.893***
	(0.398)	(0.380)	(0.312)	(0.278)	(0.490)	(0.429)	(0.307)	(0.281)
Scheduled caste or tribe	0.142	0.150	0.776***	0.673***	0.919***	0.876***	-0.013	-0.011
	(0.162)	(0.161)	(0.185)	(0.167)	(0.180)	(0.171)	(0.162)	(0.161)
Owns a latrine	-0.170	-0.170	-0.484**	-0.466**	-0.483**	-0.490**	-0.344**	-0.344**
	(0.199)	(0.197)	(0.208)	(0.200)	(0.222)	(0.227)	(0.154)	(0.154)
Village leader household	0.485**	0.496***	0.481**	0.460**	0.658***	0.657***	0.717***	0.716***
	(0.192)	(0.192)	(0.214)	(0.212)	(0.211)	(0.209)	(0.169)	(0.172)
Rents (rather than owns) home	0.177	0.147	0.020	0.034	-0.360	-0.342	0.173	0.170
	(0.305)	(0.302)	(0.299)	(0.292)	(0.343)	(0.333)	(0.254)	(0.253)
Village fixed effects	Yes		Yes		Yes		Yes	
Num. obs.	1,616		1,387		1,728		2,002	
LR test $\beta = 1$ (stat, p)	4.12 (0.042)		2.13 (0.144)		1.06 (0.303)		0.15 (0.695)	

Notes: Segments are Het/Hom (heterogeneous vs homogeneous on caste-minority share) × P/R (poor vs rich), split at village-level medians of the Caste × Beds partition. NL leaves the CES curvature β free; L imposes $\beta = 1$. The isolated-household dummy is excluded. † marks a flat-tail fit (β is numerically indistinguishable; the value is the refined estimate in the plateau). θ_1 is complementarity, θ_2 conformity. Standard errors in parentheses where computed; stars are Wald tests vs zero, $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The LR statistic tests $\beta = 1$ against a χ^2_1 reference. A dash marks a control dropped for lack of within-segment variation.

Table E.18: Segment estimates Caste × Beds partition: CES (NL, β free) and linear (L, $\beta = 1$) without the isolated-household dummy.

	Het/P		Het/R		Hom/P		Hom/R	
	NL	L	NL	L	NL	L	NL	L
Endogenous peer effects								
Strategic complementarity (θ_1)	0.338***	0.248***	0.319*	0.188***	0.560**	0.228***	0.294***	0.291***
	(0.075)	(0.045)	(0.169)	(0.059)	(0.247)	(0.076)	(0.065)	(0.051)
Conformity (θ_2)	0.001	0.104	-0.732	-0.377	-0.017	-0.061	-0.294	-0.287
	(0.753)	(0.582)	(1.229)	(0.828)	(0.735)	(0.711)	(0.807)	(0.612)
CES elasticity (β)	-2.28	1	-491.5†	1	-298.2†	1	0.76	1
	(2.0)		(545.4)		(191.6)		(2.8)	
Own effects								
Number of rooms	-0.027	-0.027	-0.125*	-0.119*	0.020	0.026	-0.114*	-0.114*
	(0.064)	(0.063)	(0.069)	(0.069)	(0.073)	(0.073)	(0.059)	(0.059)
Number of beds	-0.157*	-0.155*	-0.078	-0.078	-0.023	-0.025	-0.084	-0.085
	(0.086)	(0.086)	(0.114)	(0.113)	(0.068)	(0.067)	(0.070)	(0.070)
Non-Hindu head of household	1.640***	1.505***	0.795***	0.775***	–	–	1.281***	1.280***
	(0.455)	(0.417)	(0.302)	(0.266)			(0.357)	(0.290)
Scheduled caste or tribe	0.460***	0.482***	0.314*	0.323*	0.575***	0.589***	0.337**	0.335**
	(0.150)	(0.145)	(0.181)	(0.166)	(0.155)	(0.170)	(0.149)	(0.143)
Owns a latrine	-0.307	-0.325	-0.353**	-0.348**	-0.548**	-0.576**	-0.357**	-0.357**
	(0.232)	(0.238)	(0.179)	(0.175)	(0.244)	(0.261)	(0.144)	(0.145)
Village leader household	0.384*	0.385*	0.550***	0.555***	0.666***	0.665***	0.665***	0.664***
	(0.202)	(0.200)	(0.206)	(0.206)	(0.219)	(0.217)	(0.165)	(0.166)
Rents (rather than owns) home	0.482	0.462	-0.234	-0.245	0.433	0.400	-0.186	-0.187
	(0.305)	(0.298)	(0.285)	(0.281)	(0.429)	(0.421)	(0.222)	(0.221)
Village fixed effects	Yes		Yes		Yes		Yes	
Num. obs.	1,691		1,312		1,718		2,012	
LR test $\beta = 1$ (stat, p)	4.29 (0.038)		4.50 (0.034)		3.90 (0.048)		0.07 (0.796)	

Notes: Segments are Het/Hom (heterogeneous vs homogeneous on caste-minority share) × P/R (poor vs rich), split at village-level medians of the Caste × Latrine partition. NL leaves the CES curvature β free; L imposes $\beta = 1$. The isolated-household dummy is excluded. † marks a flat-tail fit (β is numerically indistinguishable; the value is the refined estimate in the plateau). θ_1 is complementarity, θ_2 conformity. Standard errors in parentheses where computed; stars are Wald tests vs zero, $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The LR statistic tests $\beta = 1$ against a χ^2_1 reference. A dash marks a control dropped for lack of within-segment variation.

Table E.19: Segment estimates Caste × Latrine partition: CES (NL, β free) and linear (L, $\beta = 1$) without the isolated-household dummy.

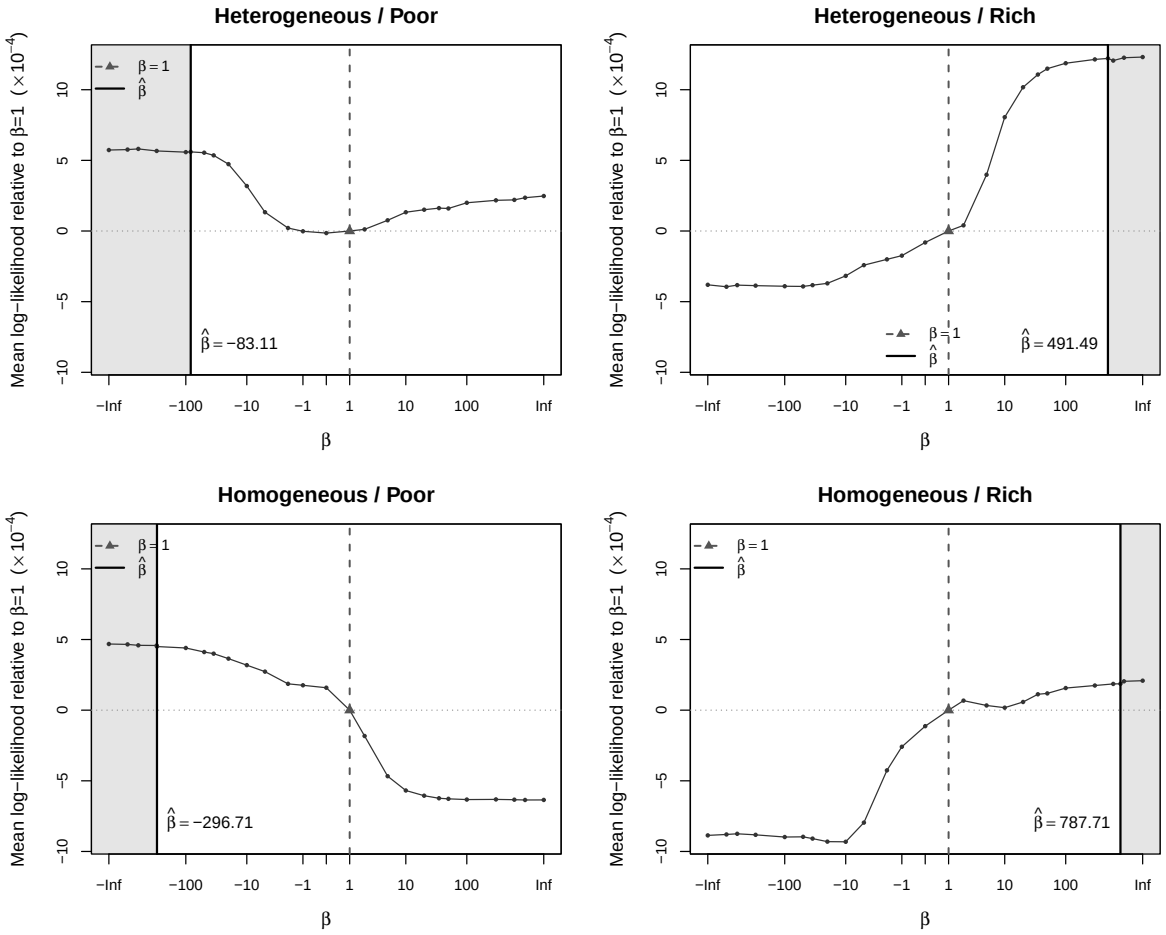
Recipient group	N	Degree	Rooms	Beds	Non-Hindu	SC/ST	Latrine	Leader
Caste × beds: Het/R								
CES only	10	16.10	2.00	1.10	20.0%	60.0%	10.0%	40.0%
Linear only	18	15.72	3.22	1.33	0.0%	61.1%	33.3%	5.6%
Both	5	23.40	3.00	1.00	0.0%	80.0%	0.0%	40.0%
Neither	1,354	5.85	2.53	0.86	5.8%	39.1%	25.3%	11.2%
Caste × latrine: Het/R								
CES only	92	6.04	3.40	1.16	13.0%	51.1%	57.6%	3.3%
Linear only	23	13.61	2.30	0.57	21.7%	69.6%	34.8%	26.1%
Both	1	12.00	4.00	2.00	0.0%	100.0%	100.0%	0.0%
Neither	1,196	5.01	2.27	0.69	10.1%	42.1%	27.5%	12.5%

Notes: Entries are group means; binary characteristics are percentages. A household is a recipient if its subsidy exceeds 10^{-6}. Each segment receives a budget of 20 utility-index units, and each policy uses its own estimates. The CES estimates are flat-tail fits ($\hat{\beta} = 491.5$ for beds and $\hat{\beta} = -282.8$ for latrine).

Table E.20: Household characteristics by subsidy-recipient group in the Het/R segments.

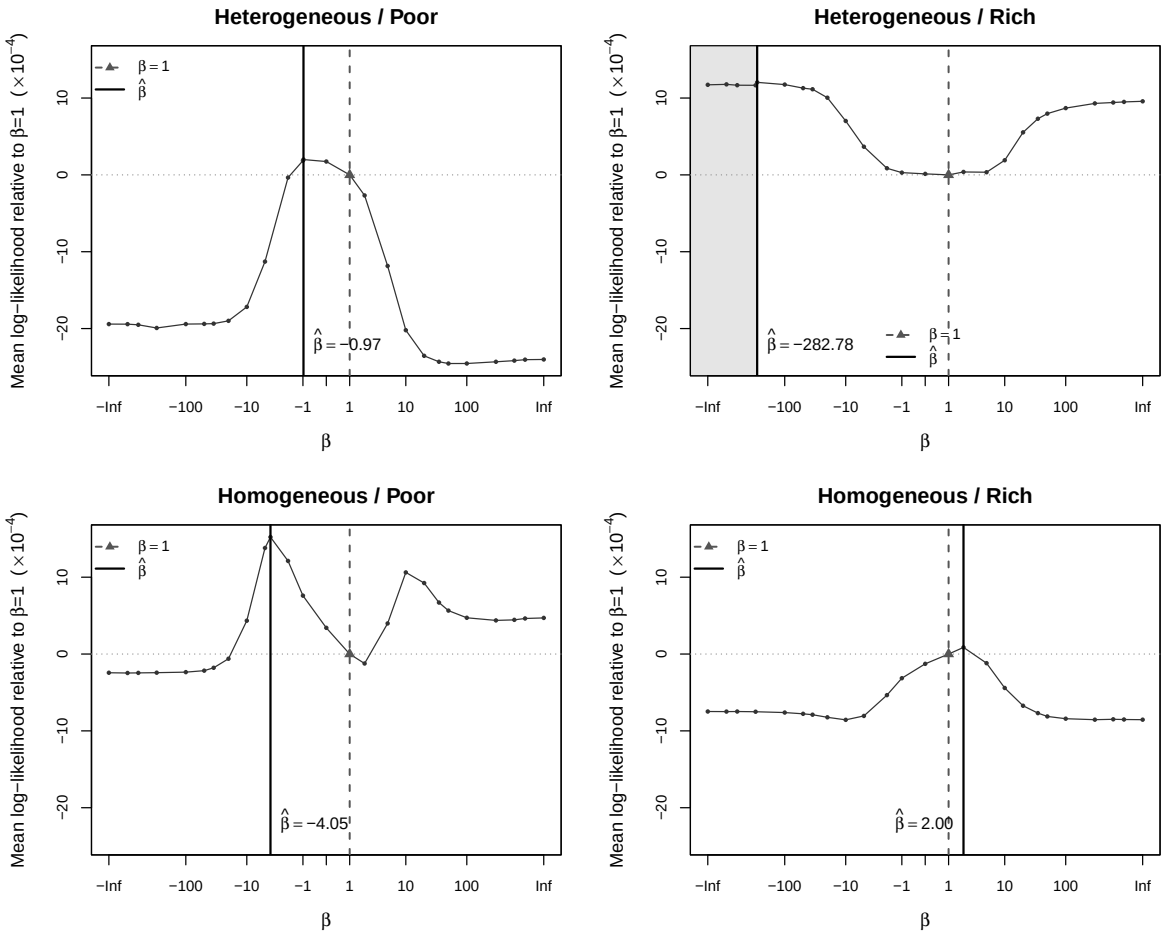
E.3 Microfinance Estimation Figures

This section presents the figures for the empirical exercise, plotting the profiled log-likelihood in β by segment for both partitions and the CES-versus-linear optimal-subsidy comparison for the illustrative Het/R segments.



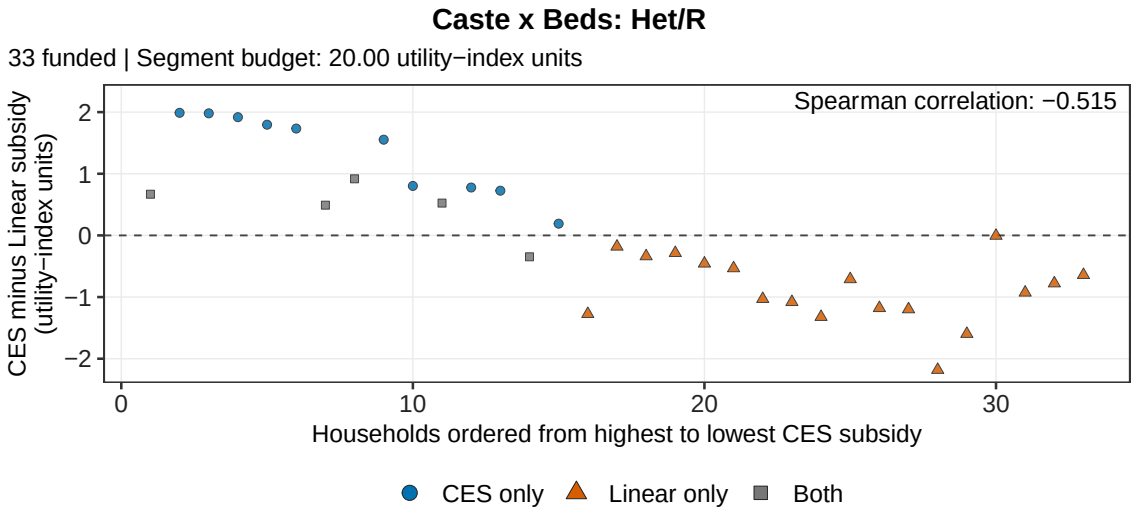
Notes: The y-axis is the Log-Likelihood of the unrestricted CES model relative to the linear case. Shaded region denotes the flat tail-region where β is numerically indistinguishable.

Figure E.1: Profiled log-likelihood in β by segment, Caste × Beds.

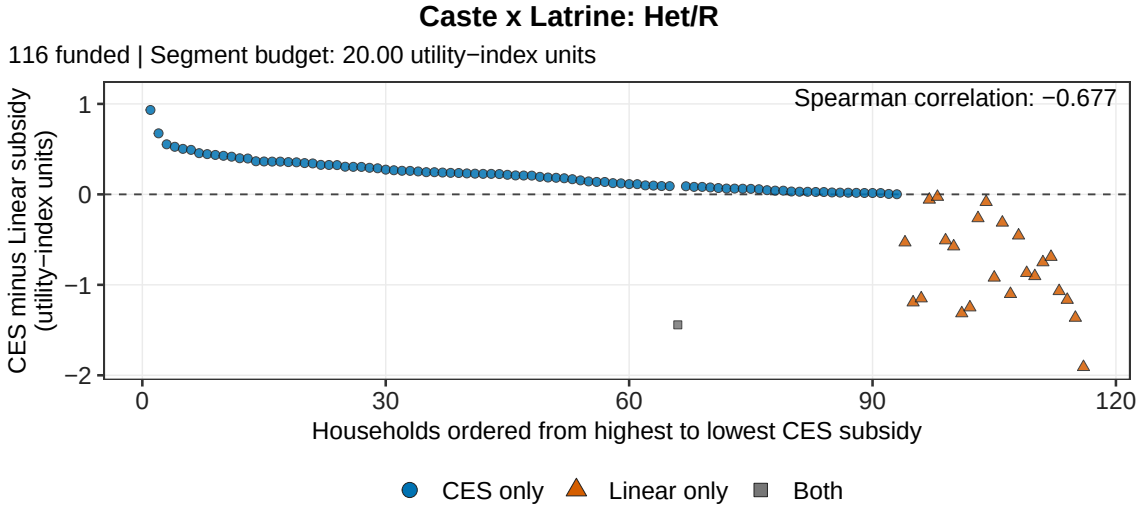


Notes: The y-axis is the Log-Likelihood of the unrestricted CES model relative to the linear case. Shaded region denotes the flat tail-region where β is numerically indistinguishable.

Figure E.2: Profiled log-likelihood in β by segment, Caste × Latrine.



Notes: Linear adoption rises by 1.35 percentage points (0.33 direct, 1.01 indirect); CES adoption rises by 13.29 points (0.33 direct, 12.96 indirect).



Notes: Linear adoption rises by 1.05 percentage points (0.37 direct, 0.68 indirect); CES adoption rises by 0.91 points (0.28 direct, 0.63 indirect).

Figure E.3: Household subsidy differences under CES and Linear norms for the two Het/R segmentations. Each model uses its own estimates and stationary allocation. Within each segment, households funded by either model are ordered by CES subsidy; Spearman correlations are computed over this funded union, excluding households unfunded under both. Each segment receives $B = 20$ utility-index units. Direct gains hold baseline peer probabilities fixed; indirect gains are the remaining equilibrium response.