Electronic Companion Supplement

EC.1. Mathematical Notation

In Table EC.1 below, we define the terms used in networks. These terms are helpful when we define network properties and in the moment conditions. We illustrate a village network in **??**, with the black nodes indicating leaders.

Table EC.1 Table of Notation

Characteristic	Description	Definition
Nodes Degree	Number of connections (edges) of i	D_i
Edge	Connection between nodes i and j	$e_{ij} \in \{0,1\}$
Adjacency (Edge) Matrix	Connection between nodes i and j	$\mathbf{E}, E_{i,j} \in \{0,1\}$
Node Set	Set of all N nodes in Network	$\mathcal{V} = \{1,2,\ldots,N\}$
Edge Set	Set of all edges in Network	$E = \{(i,j) : e_{ij} = 1\}$
Network Edge Count	Number of undirected connections	$e = \sum_{i\in\mathcal{V},j>i} e_{ij}$
Seeds	Set of all nodes chosen as seeds	$\mathcal{S}$
Adopters	Set of all nodes which have adopted	$\mathcal{A}$
Reachable Set	Nodes with adoption status $s \in \{A, NA\}$ reachable from i in k steps	$E_i^s(k)$
Proportion of adopting neighbors	Fraction of adopting nodes among those reachable from node i in k steps	$z_i(k) = \dfrac{\lvert E_i^A(k)\rvert}{\lvert E_i^A(k)\rvert + \lvert E_i^{NA}(k)\rvert}$
Vector of above	Vector of adopting proportion of neighbors for each node	$z(k) = [z_1(k),\ldots,z_N(k)]$
Minimum Distance	Distance of Shortest Path between i and j	$\delta_{ij} = \min_k s.t. E^k_{(i,j)} > 0$

EC.2. Identification of Leader Fixed Effect

We demonstrate below that the WOM communication probability for leaders q_L is separately identified from the word of mouth communication probability q for non-leaders. While the argument itself is non-parametric and does not rely on a specific functional form, our demonstration model uses a simple parametric representation consistent with the paper. For this argument, we choose to add a leader fixed effect to the simplest model (Model 1) from the paper.

Suppose we had only static adoption data, we would not be able to identify the fixed effect. However, (i) the availability of time series aggregate adoption data and (ii) the presence of multiple networks allows us to identify the fixed "leader" effect.

First, we note that using only the final adoption levels *will not allow* leader fixed effect q_L to be identified separately from just overall propensity to communicate q. Increases in each of these parameters will result in higher final adoption levels in a network. It is straightforward to see that a relatively low level of q in conjunction with a high level of q_L might result in the same adoption level as a high level of q and a low level of q_L.

However, the curvature of the adoption trajectory over time provides variation that permits identification of the leader effect q_L separately from q. Intuitively, if q_L is higher, the adoption trajectory shows a steeper increase in the earlier periods, since only leaders are communicating initially, and only in subsequent periods do non-leaders communicate. Thus, the proportion of communication attributable to leaders is highest at the beginning and decreasing over time. Thus, the impact of a higher q_L will be greatest in earlier periods as opposed to later periods. In contrast, the impact of a higher q will be lower in the initial periods, since few non-leaders are informed, and it has proportionally greater impact on adoption in later periods.

While the above argument is non-parametric and does not rely on specific functional forms for identification, for the purpose of illustration, we use a parametric model below.

Simplified Model

We provide a highly simplified version of the model similar to Model 1 in the paper, for the specific purpose of examining identification and making the required variation transparent. The main features of this model are:

1. A few leader nodes are informed initially (similar to the main model).
2. In each period, each informed node communicates with probability (that depends on the node's leadership status). Thus, non-leaders communicate with probability q and leaders communicate with probability q_L with each of its neighbors. Note that in this simplified model, adoption status *does not* impact communication probability.[11]
3. When nodes are newly informed, they have the ability to adopt a product with probability $\gamma = 0.2$. (We don't have any covariates impacting adoption here, unlike in the main model, and do not require the variation obtainable from these covariates).

The WOM communication probability for node i is specified as:

$$p_i = \begin{cases} q, & \text{if } i \text{ is not a leader} \\ q_L = q + q_\ell, & \text{if } i \text{ is a leader} \end{cases}$$

where q_ℓ is the leader fixed effect. Recall that the leader fixed effect is the difference between the WOM communication probabilities of leaders and non-leaders.

We demonstrate in Figure EC.1 precisely the variation that is required for this identification. There are several sources of possible variation in the network data. First, we observe that both adoption trajectories for (a) $q = 0.01, q_\ell = 0.08, q_L = q + q_\ell = 0.09$ (red curve) and (b) $q = 0.13, q_\ell = -0.12, q_L = q + q_\ell = 0.01$ (green curve) end up after $T = 5$ periods at the same overall adoption

[11] Even though this additional variation based on adoption status might prove useful as a separate source of identification, our identification argument does not require it.

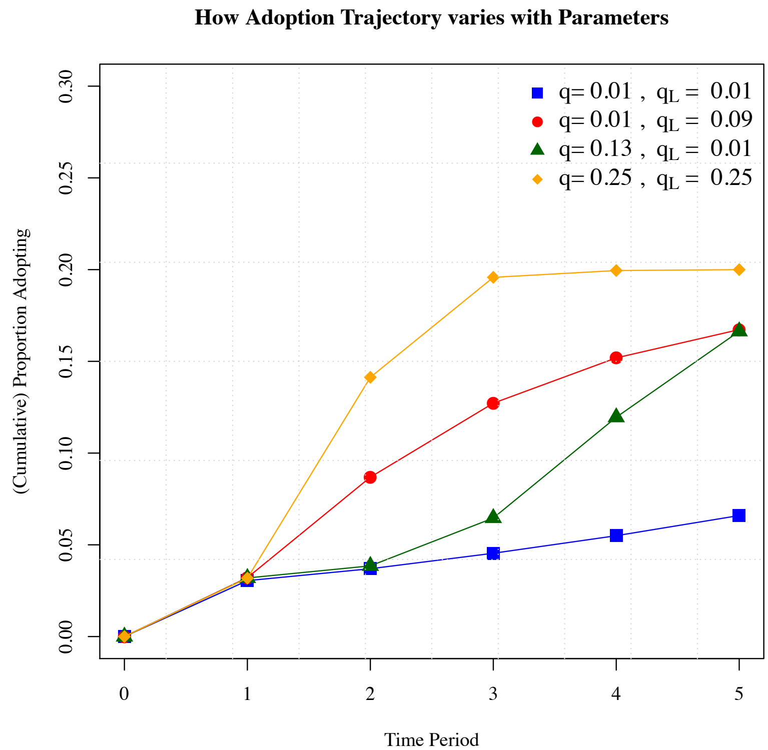


Figure EC.1 **Identification and Adoption Time Trajectory**

level, i.e. 0.165 or 16.5%. Thus, just having the final adoption levels, it would not be possible to separately identify q and q_L.

However, their adoptions differ in their time trajectories. For (a) (red curve), with a higher leader fixed effect $q_L = 0.09$, we see the **early period trajectory is steeper** than the case (b) (green curve). On the other hand, with (b), the later period trajectory is steeper than in (a).

In general, for different combinations of (q, q_L) that *obtain the same level of final overall adoption*, the area under the adoption trajectory curve will be greater for combinations of (q, q_L) with higher levels of q_L and lower levels of q.

Does exclusive seeding by leaders help or hinder identification of leader fixed effect? There are two reasons why leader seeding (in contrast to random seeding) is helpful to answering our research question.

First, it may appear that our context in which the firm exclusively used leaders by the for initial seeding makes it more challenging to separately identify the leader fixed effect. But in fact, our explanation above should clarify that this exclusive use of leaders for initial seeding aids identification of the leader fixed effect and allows us to disentangle q_L and q. This is because the exclusive use of leaders for initial seeding guarantees that a higher leader fixed effect will increase the earlier adoption trajectory relative to later. Therefore if the seeding had been random, it would not be feasible to separately identify the effects as one cannot use this identification argument.

Second, leader seeding avoids a specific kind of bias in leader effects. Suppose we only have random seeding, but there are leaders present in the data. If leaders have different (higher or lower) degree on average than others, and if they have differential communication, it would not be possible to identify any leader specific communication effect. For instance, if Leaders have higher degree, the Local friend strategy could result in more leaders on average. The counterfactual results would then be biased to find lower effects for the Local strategy than would be obtained in reality. Due to leader seeding in our data, we can identify and characterize the leader fixed effect (separately from non-leaders), and thus avoid this potential bias.

EC.3. Model Details and Estimation

First, we detail the estimation of the adoption process, followed by the WOM communication process, and finally detail the block bootstrap to obtain standard errors. We simulated $N_{sim} = 150$ diffusion paths with seeds chosen stochastically corresponding to each seeding level and using each of the seeding strategies. The reported WOM communication parameters are based on the average of the simulated diffusion paths.

Adoption Process

The adoption parameter vector is $\beta = (\beta_0, \ldots, \beta_6)$. The logistic regression specification for the adoption decision follows from the utility specification. The log likelihood for household i is $l_i(\beta|X_i)$ and for all households in the network is $l(\beta|X)$

$$l(\beta|X) = \sum_{i=1}^{N} l_i(\beta|X_i) = \sum_{i=1}^{N} \log P(y_i = 1|X_i) = \sum_{i=1}^{N} \log \left[\frac{\exp(\beta X_i)}{1 + \exp(\beta X_i)} \right] \tag{EC.1}$$

$$\tag{EC.2}$$

The adoption process is estimated by maximum likelihood estimation.

WOM Process

Given adoption parameters β, the WOM process is simulated separately for each village network. We track two states for each household: its information state and its adoption state. The information states are *uninformed* (U) and *informed* (I), whereas the adoption states are *Not-adopted* (NA) and *Adopted (A)*. Both the Informed and Adopted states are absorbing states, during which nodes can communicate with their neighbors.

An *informed household* with adoption status $s \in \{NA, A\}$ (i.e. non-adopting or adopting) will communicate with any of its neighbors in a single time period with probability $p^s(D)$. This is a dynamically evolving process over time, and depends on the informed status of all households in the network. We have formalized these details further below using additional notation. Let $p^s(D)$ be the probability that an *informed household* with adoption status at the beginning of time t $s^j(t) \in \{U, NA, A\}$ (i.e. uninformed, non-adopting or adopting) of degree D will communicate with any of its neighbors *in a single time period*. Uninformed households do not communicate. During time period t, an uninformed household i becomes informed if it receives a communication from any of its network neighbors $\mathcal{N}_i$. This event happens with probability $p_{it} = 1 - \Pi_{j \in \mathcal{N}_i}(1 - p^{s_j(t)}(D_j))$.

The WOM process for each of the N_{sim} simulations begins with Step (0) and then proceeds through Steps (1)-(3) for each time period.

(0) Each household (node) in the network is initially in an uninformed (U) information state. In initial period $t = 0$, the seed nodes are chosen in each network based on the seeding strategy. In the actual data, the seed nodes in each village were chosen based on the opinion leadership criterion. In the counterfactual scenarios, seed nodes are chosen based on an alternative strategy (Random, Local Friend etc.). In all cases, the information state of the seed nodes changes from Uninformed (U) $\longrightarrow$ Informed (I).

The following process (1) – (3) process then takes place in each period $t \in \{1, 2, \ldots, T_v\}$ for village v.[12]

(1) Each household that has become informed decides whether to adopt.

(2) Then, an informed household can probabilistically communicate about the microfinance product with each of its network neighbors. The probability of such communication $p^s(D)$ may depend on both its degree D, i.e. the number of neighbors the informed household has, as well as the adoption status $s \in \{A, NA\}$ of the informed household. We separate out the probabilities $p^{NA}(D)$ and $p^A(D)$ as detailed in §3 of the paper.

(3) When this communication takes place, each neighbor receiving information changes its information state from Uninformed (U) $\longrightarrow$ Informed (I). If the neighbor node has already been informed earlier, there is no change in its state.

For each simulation and for each village v, we compute 6 cross-sectional moments according to Table EC.2 at the end of T_v periods of simulation, and 3 time series moments. Thus, for the 43 villages with microfinance adoption, we have $N_{moments} = 9 \times 43 = 301$ moments across the villages. We then minimize the MSM objective function $S(\theta)$ detailed in equation (7) from §3 in the $[0, 1]^K$ region to obtain the probability parameter estimates presented in Table 6 in §4 of the paper. For the MSM objective, we start with the initial weight matrix set to the identity matrix to obtain consistent estimates. Since we obtain standard errors through bootstrap, a consistent estimator is all that is needed.

[12] The number of time periods varies across villages in the data, with a mean of 6.5 and SD of 1.83.

Standard Errors with Bootstrap Estimation

We obtain standard errors for the communication probability parameters using a bootstrap procedure detailed below. First, we obtain $N_R = 2,000$ draws using a random grid for the communication probability vector $\theta = (q_0^{NA}, q_1^{NA}, q_0^A, q_1^A) \in [0,1]^4$. The parameter is characterized appropriately based on the model specification.

We proceed through Steps (a) – (c) below for each of the N_{sim} draws to obtain moments for each village v.

(a): We choose seeds corresponding to the Leader strategy used in the data.

(b): We compute the simulated WOM Process detailed above for T_v periods for each draw of the parameter vector θ.

(c): We use the cross-section and time series adoption status data to compute the moments detailed in Table EC.2 separately for each village.

Compute $B = 10,000$ bootstrap estimates using the moments obtained from the samples above. For $b = 1, 2, \ldots, B$ do Steps (d) – (f) below.

(d): Resample with replacement from moments from the set of villages showing microfinance activity.

(e): Compute the objective function with the resampled moments at each of the N_R points evaluated above.

(f): Choose the parameter vector with the minimum objective as the estimate $\beta^{(b)}$ to be used in the bootstrap.

The distribution of $\beta^{(b)}$, with $b = 1, 2, \ldots, B$ provides the bootstrap estimate distribution for computing standard errors.

Moment Conditions for Estimation

In this section, we describe the rationales for the moments listed in Table EC.2 that we use in our estimation. The required mathematical notation is defined in §EC.1.

In general, all moments are informative in the estimation of all parameters. However, the connections between some moments and parameters are more intuitive. The time series moments, and more generally the temporal trajectory are especially important for identification when there are differential effects for leaders. We describe the moments and the obvious associated links with parameters below.

First, we detail the cross-sectional moments MC1 to MC6. (MC1) is the proportion of seeds that have adopted. Since the seeds are guaranteed to be informed outside the WOM process, this allows us to estimate the parameters impacting adoption probability without relying on the communication process. In contrast, (MC2) is the proportion of households with no adopting

Table EC.2 List of Moments.

Symbol	Description	Definition
MC1	Proportion of seeds adopting	$\frac{\lvert\mathcal{S}\cap\mathcal{A}\rvert}{\lvert\mathcal{S}\rvert}$
MC2	Proportion of households with no adopting neighbors who have adopted	$\frac{\sum_{i\in\mathcal{A}}\mathbf{I}\left[\mathcal{N}_i\cap\mathcal{A}=\phi\right]}{\sum_{i\in\mathcal{V}}\mathbf{I}\left[\mathcal{N}_i\cap\mathcal{A}=\phi\right]}$
MC3	Proportion of neighbors of adopting seeds who have adopted	$\frac{\bigcup_{j\in\mathcal{S}\cap\mathcal{A}}\lvert\mathcal{N}_j\cap\mathcal{A}\rvert}{\bigcup_{j\in\mathcal{S}\cap\mathcal{A}}\lvert\mathcal{N}_j\rvert}$
MC4	Proportion of neighbors of non-adopting seeds who have adopted	$\frac{\bigcup_{j\in\mathcal{S}\cap\mathcal{V}\setminus\mathcal{A}}\lvert\mathcal{N}_j\cap\mathcal{A}\rvert}{\bigcup_{j\in\mathcal{S}\cap\mathcal{V}\setminus\mathcal{A}}\lvert\mathcal{N}_j\rvert}$
MC5	Covariance between a household's adoption and average adoption of their first degree neighbors	$cov\left(y,z(1)\right)$
MC6	Covariance between a household's adoption and average adoption of their second degree neighbors	$cov\left(y,z(2)\right)$
MTτ	Cumulative adoption upto time τ (Time series moment)	$y_\tau=\frac{1}{N}\sum_{j=1}^{N}y_{j\tau}$

neighbors who adopt, which allows us to match a non-adopter's communication likelihood, because such an adopting household could only have received information from neighbors, all of whom are non-adopters.

(MC3) is the proportion of neighbors of adopting seeds who have adopted. This moment most closely connects to the WOM probability of adopters, since the neighbors of seeds have a high probability of receiving information from the seeds. With (MC4), the proportion of nodes that are neighbors of non-adopting seeds who adopt. The focus here is primarily on parameters q_0^{NA} and q_1^{NA}. With low probability, it becomes less likely that neighbors of non-adopting seeds would adopt (all else being equal).

(MC5) and (MC6) captures the relationship between adoption by a focal household and its first and second degree neighbors. This is particularly important in networks where there is a significant region (or sub-network) that is uninformed. In such regions of the network, both a focal node and its neighbors will have zero adoption, which results in a perfect correlation. Observe that in such a case, (MC2) and (MC4) are not informative since the moment will have values exactly zero for such sub-networks. Thus (MC5) and (MC6) can also be viewed as characterizing the limits of the WOM process.

Overall, we need to have moments that match global network-level measures, e.g. (MC1) that focuses on overall adoption. It is also critically important to incorporate moments that match local network structure, allowing these connections to have a strong impact on the adoption process, which is what distinguishes the network approach from the Bass model.

The time series moments (MTτ) matches the cumulative overall adoption in each time period τ period within each village. This is the typical data used in estimation of aggregate Bass-like diffusion models. These moments helps us to estimate the time-path of the diffusion process. In each period of the model, based on the network structure and the diffusion of the information process, we have different number of households which potentially become informed and therefore have the opportunity to make adoption choices.

We detail the sensitivity of parameter estimates to moments using the methodology of Andrews et al. (2017) in §EC.5.6.

EC.3.1. Model Fit

Additional Model Fit Metrics We next evaluate the fit of these models below using 3 additional measures. The metrics used for fit are detailed below:

1. First, we regress the actual adoption rate during each time period in the data (as dependent variable) against the simulated adoption rate obtained from the model, similar to what Banerjee et al. (2013) present in Table 2 of their paper. The intercept terms are found to be not significant, and the coefficient of interest across all models indicate that the model is able to capture and characterize the essential dynamics of the process. If the coefficient of simulated adoption is close to 1, that would indicate a good fit.
2. Next, we examine typical fit measure like **RMSE** (root mean squared error) and **MAPE** (Mean Absolute Percent / Proportion Error). Lower values of these measures indicate better fit.

We find that the model fit is consistent with the original paper for in-sample fit (see Table 2 of Banerjee et al. (2013)). We then examine out of sample fit by estimating our preferred models using 85% of the villages, and holding the remaining 15% of the sample as holdout. We find that the out of sample fit is not significantly worse than in sample fit, indicating the models do not suffer from an obvious overfitting problem. Banerjee et al. (2013) do not provide *out of sample fit* in their paper.

Table EC.3 provides the in-sample and out-of-sample fit for our preferred models. We note that the coefficients on simulated adoption for both in-sample and out-of-sample are between 0.87 and 0.89. The RMSE and MAPE measures are similar for both of our chosen models, and it is useful to verify that the out-of-sample fit is not much worse than in-sample fit. If out-of-sample were indeed much worse, then we should be concerned about the model overfitting the data.

EC.4. Counterfactuals

We detail first the implementation of each of the strategies, and then performance of the strategies under different models and at different seeding levels.

EC.4.1. Seeding Strategy Implementation

Table EC.4 provides specific implementation details for each of the seeding strategies we consider.

Table EC.3 Main Models: In Sample and Out of Sample Model Fit Measures

	In Sample Fit		Out of Sample Fit	
	$\mathbf{M_2^{E=1}}$	$\mathbf{M_2^{E=0}}$	$\mathbf{M_2^{E=1}}$	$\mathbf{M_2^{E=0}}$
Intercept	0.002	0.000	-0.002	-0.001
	(0.02)	(0.02)	(0.02)	(0.02)
Simulated Adoption	0.874	0.89	0.875	0.87
	(0.097)	(0.098)	(0.096)	(0.1)
RMSE	0.067	0.067	0.069	0.069
MAPE ($\times$100%)	0.379	0.372	0.395	0.406

Table EC.4 Seeding Strategies and Implementation

Category	**Strategy**	**Implementation Procedure (for each of m seeds)**	**Information Required**
Random	Random	Select node at random from list as seed.	Randomly sampled subset of list of individuals (or Complete List)
Friend	Local Friend	Select node at random from list. Obtain one randomly chosen friend of node as a seed.	Randomly sampled subset of list of individuals + Obtain random friend
Leader	(Firm's) Leader	Select node from list of leaders indicated by firm	List of Leaders (where leadership is specific to domain)
	Like Leader	Select leader node ℓ at random. Select the non-leader node most similar to ℓ in terms of network properties‡.	List of leaders + Entire Social Network (Adjacency Matrix **E**)
Hybrid	Friend of Leader (Weak Hybrid)	Select a random **leader** from list of leaders. Obtain one randomly chosen friend of this leader as a seed.	List of leaders + Obtain **random** friend
	Leader Friend of Leader (Strong Hybrid)	Select a random **leader** from list of leaders. Obtain one randomly chosen friend who is also a **leader** to be seed.	List of leaders + List of **leader** friends of each leader
Network Information Strategies	Top Degree	Select a **node** randomly from list of top degree nodes (We specify this as the top 15% most highly connected nodes.	Degree of each node in the network.
	Top Diffusion	Select a **node** randomly from list of top diffusion centrality nodes. Diffusion Centrality is defined as $DC = [\sum_{t=1}^{\tau} q^t \mathbf{E^t}] \cdot \mathbf{1}$ where **E** is the adjacency matrix and **1** is the column vector of 1s. As suggested in Banerjee et al. (2013), we set $q = \dfrac{1}{\lambda_1}$ where λ_1 is the greatest eigenvalue of the adjacency matrix.	Full Adjacency matrix **E** (who is connected to whom) and the number of periods for diffusion τ.

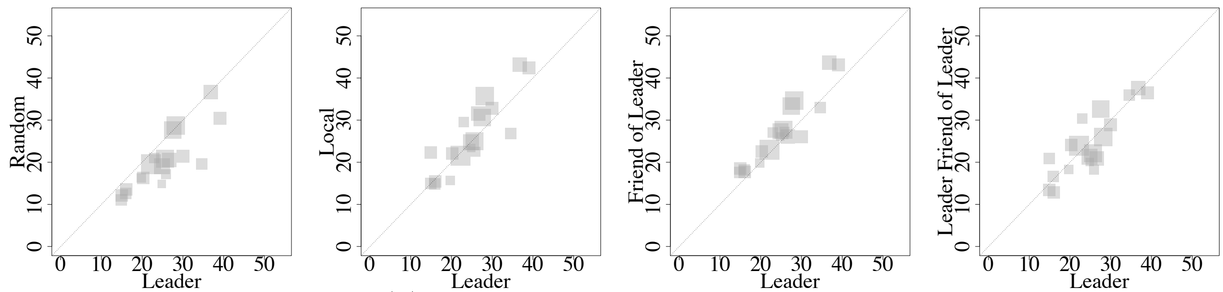
‡ : Similarity between nodes in network position could be implemented using the following centrality metrics (among others): degree, eigenvector, Bonancich power centrality

EC.4.2. Comparison of Strategies

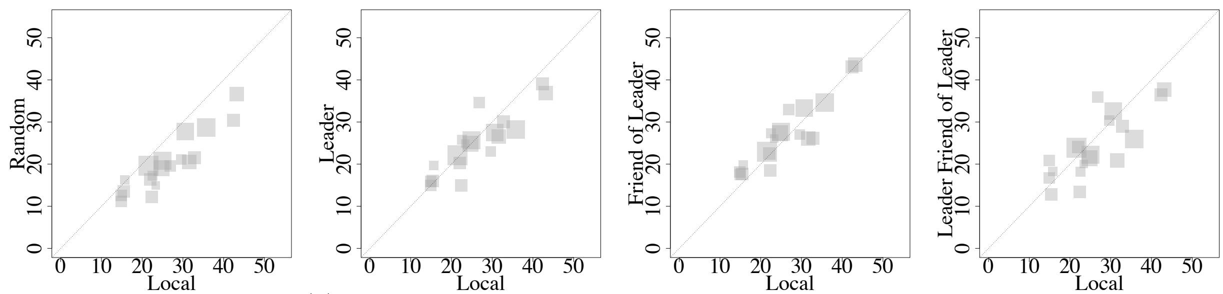
Figure EC.2, shows the performance of the strategies pairwise, where performance is measured by the proportion of informed households in each counterfactual strategy evaluation.

Figure EC.2 Comparison of Strategies across Villages (1% of Households Seeded). (% Informed Households)

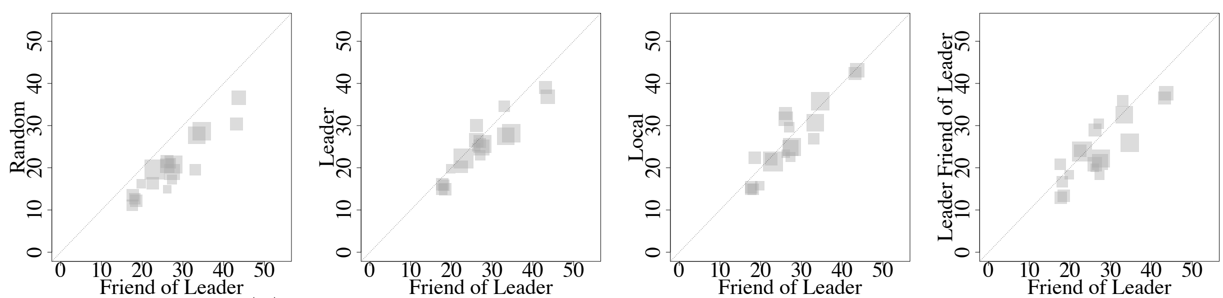
(a) **(Firm's) Leader versus Other Strategies**



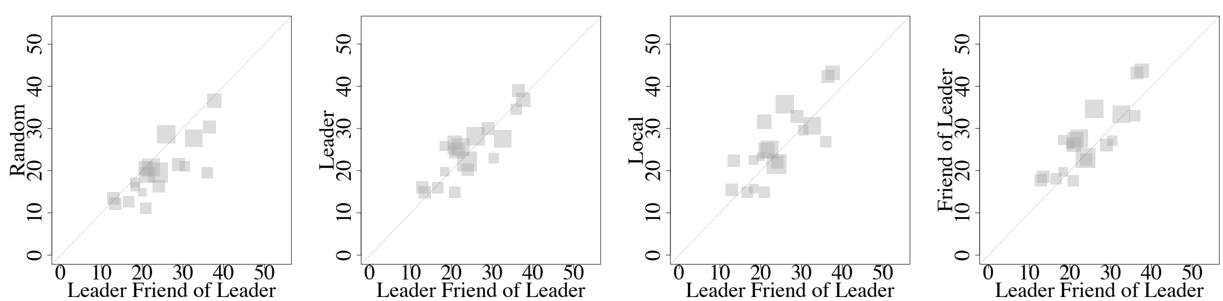
(b) **Local versus Other Strategies**



(c) **Friend of Leader versus Other Strategies**



(d) **Leader Friend of Leader versus Other Strategies**



Note: Each data point (square) is a village network in all panels. The size of the shape is proportional to the size of the village (number of households). Darker colors indicate overlap between villages.

Next, we examine the consistency of relative performance of the various seeding strategies across villages. **??** provides an overall comparison of the 4 strategies with the adoption levels of Leader, Local Friend and Hybrid strategies plotted against one another. We find that both Local Friend and Friend of Leader consistently perform better on adoption relative to Random as all villages fall above the diagonal. In contrast, while Leader is better than Random for most villages, it is worse for some villages, as shown by the points that fall above the diagonal in the top-left panel.

Moreover, the villages where the Leader strategy performs especially well are smaller (fewer households). In terms of the hybrid strategies, we find that the weak hybrid Friend of Leader strategy mostly outperforms Leader, but it does not do better than Local Friend overall. The strong hybrid Leader Friend of Leader actually performs worse than the Local Friend and weak hybrid strategy. In many villages, it performs worse than the Leader strategy as well.

EC.4.3. Leverage Under Different Models

We examine how the number of seeds impacts the performance of different seeding strategies in the counterfactual across the full set of model specifications. We examine seeding at the level of 0.5%, 1%, and 5% to understand how the level of seeding affects relative benefits of our friendship paradox strategies. The results for different seeding levels are detailed in Table EC.5.

A few observations are relevant here:

(a) The (firm's) leader strategy typically (but not always) outperforms the random node strategy for any combination of model / (#seeds)

(b) The friendship paradox based Local strategy achieves higher performance (leverage) than the firm's leader strategy under all of the model specifications.

(c) The weak hybrid Friend of Leader seeding strategy achieves better performance than Local strategy in most model specifications. However, the strong hybrid Leader Friend of Leader strategy seems to consistently underperform the (firm's) leader strategy.

(d) The "Like Leader" strategy performs very similar to Firm's leader (within 2-3% of the leverage metric).

(e) Leverage for all counterfactual strategies decreases as the number of seeds increases.

Table EC.5 Leverage for Counterfacual Strategies

(a) Seeding at 5% of number of nodes

	$\mathbf{M}_1^{E=0}$	$\mathbf{M}_2^{E=0}$	$\mathbf{M}_3^{E=0}$	$\mathbf{M}_4^{E=0}$	$\mathbf{M}_1^{E=1}$	$\mathbf{M}_2^{E=1}$	$\mathbf{M}_3^{E=1}$	$\mathbf{M}_4^{E=1}$	$\mathbf{M}_B^{E=0}$	$\mathbf{M}_B^{E=1}$
Local	1.06	1.04	1.03	1.03	1.05	1.06	1.04	1.04	1.02	1.05
(Firm's) Leader	1.03	1.03	1.00	0.99	1.02	1.04	1.00	1.01	1.01	1.02
Like Leader	1.04	1.02	1.01	1.02	1.03	1.04	1.02	1.02	1.01	1.02
Friend of Leader	1.05	1.05	1.03	1.03	1.05	1.06	1.04	1.04	1.03	1.04
Leader Friend of Leader	1.02	1.02	1.00	1.00	1.02	1.04	1.01	1.02	1.01	1.03
Top Degree	1.09	1.07	1.05	1.06	1.08	1.09	1.05	1.05	1.04	1.07
Top Diffusion	1.07	1.06	1.04	1.05	1.06	1.07	1.04	1.05	1.03	1.06

(b) Seeding at 1% of number of nodes

	$\mathbf{M}_1^{E=0}$	$\mathbf{M}_2^{E=0}$	$\mathbf{M}_3^{E=0}$	$\mathbf{M}_4^{E=0}$	$\mathbf{M}_1^{E=1}$	$\mathbf{M}_2^{E=1}$	$\mathbf{M}_3^{E=1}$	$\mathbf{M}_4^{E=1}$	$\mathbf{M}_B^{E=0}$	$\mathbf{M}_B^{E=1}$
Local	1.16	1.21	1.15	1.18	1.19	1.19	1.17	1.16	1.15	1.20
(Firm's) Leader	1.07	1.10	1.00	1.01	1.11	1.08	1.00	1.02	1.10	1.13
Like Leader	1.09	1.11	1.07	1.09	1.12	1.13	1.09	1.08	1.09	1.12
Friend of Leader	1.18	1.22	1.18	1.18	1.21	1.19	1.18	1.15	1.16	1.22
Leader Friend of Leader	1.10	1.12	1.03	1.03	1.11	1.13	1.03	1.02	1.11	1.12
Top Degree	1.39	1.37	1.30	1.33	1.37	1.43	1.32	1.30	1.30	1.39
Top Diffusion	1.35	1.36	1.30	1.32	1.39	1.40	1.32	1.30	1.31	1.37

(c) Seeding at 0.5% of number of nodes

	$\mathbf{M}_1^{E=0}$	$\mathbf{M}_2^{E=0}$	$\mathbf{M}_3^{E=0}$	$\mathbf{M}_4^{E=0}$	$\mathbf{M}_1^{E=1}$	$\mathbf{M}_2^{E=1}$	$\mathbf{M}_3^{E=1}$	$\mathbf{M}_4^{E=1}$	$\mathbf{M}_B^{E=0}$	$\mathbf{M}_B^{E=1}$
Local	1.30	1.25	1.26	1.26	1.25	1.27	1.23	1.24	1.21	1.25
(Firm's) Leader	1.11	1.11	1.02	1.01	1.16	1.19	1.00	1.11	1.13	1.13
Like Leader	1.14	1.12	1.18	1.15	1.17	1.14	1.12	1.16	1.16	1.15
Friend of Leader	1.31	1.25	1.28	1.24	1.27	1.34	1.22	1.28	1.24	1.25
Leader Friend of Leader	1.18	1.12	1.06	1.06	1.15	1.23	1.02	1.12	1.12	1.21
Top Degree	1.56	1.54	1.52	1.48	1.54	1.62	1.43	1.51	1.47	1.55
Top Diffusion	1.56	1.54	1.49	1.51	1.61	1.60	1.48	1.52	1.49	1.54

EC.5. Alternative Models and Robustness Checks

We consider different models of WOM communication and seeding to assess if our key claims are sensitive to model specification. Here we consider three models of WOM communication. In §EC.5.1, as a basic benchmark, we consider a single source—advertising type, non-network model where information is not transmitted through the social network, but all households receive information from a central single source (perhaps the firm). Assessing the relative fit of this model with respect to our preferred network based communication model can clarify the importance of modeling information transmission through social networks before even assessing the role of seeding strategies.[13] In §EC.5.2, we consider a broadcast model where in the first period seeds conduct a village-wide meeting in which information about the microfinance program can be broadcast to all those who attend. This model is motivated by a meeting process that the firm encouraged the seeds to conduct and is described in Banerjee et al. (2013). In §EC.5.4, we consider a process where leaders have a certification impact—their adoption or support may increase persuasion even outside of their networks. This model goes beyond the WOM effects of leaders in allowing differential impact based on source of certification. Finally, in §EC.5.3 we also consider a seeding process where not just the firm chosen seeds are leaders but a random sample from certain select occupations that are considered as leaders. In this process, leaders are present both among seeds and non-seeds, but all seeds are randomly chosen members of these occupations.

In the text of the paper, we had used the best fitting model of adoption (Table 5). In §EC.5.5, we report the alternative models of adoption we considered and their relative fit with respect to the chosen model. Finally, we report parameter sensitivity to the different moments used in the estimation §EC.5.6.

EC.5.1. Single Source Model

We consider a single source—advertising type, non-network model here as a null benchmark model. In this model, information is not transmitted through the social network, but all households receive information from a central single source (perhaps the firm). In **each period**, the source transmits information to each household with probability θ—could be thought of as a household seeing an ad. Informed households then have a chance to adopt. The adoption model is identical to the main model in the paper. Informed households do not communicate any information to other household in this single source model.

Observe that this model is parametrized by only one parameter θ. The estimated value θ on an average set of villages is $\hat{\theta} = 0.46$. Using this model, we evaluate the in-sample and out of sample fit (using 15-20% of the villages as a hold out sample). Similar to Banerjee et al. (2013), we regress the real adoption data on the simulated adoption trajectory derived from the model.

There are a few observations:

1. The coefficient of simulated adoption share based on the estimated null model are not statistically significant for either the in sample or out of sample adoption share regressions. Thus, this model does not have any predictive powerin explaining true adoption.

The finding above is not surprising, since a null model must lead to a concave cumulative adoption curve over time (since there are fewer households that have not been informed over time), whereas

[13] We thank a reviewer for suggesting this benchmark

Table EC.6 Null Single Source Model: Fit Measures

	Dependent variable:	
	In Sample Villages Adoption Share	Out of Sample Villages Adoption Share
Constant	0.060	0.017
	(0.064)	(0.197)
Simulated Adoption share	0.847	1.143
	(0.544)	(1.669)
RMSE	0.112	0.076
MAPE ($\times$100%)	0.38	0.30
R^2	0.047	0.055
Note:		*p<0.1; **p<0.05; ***p<0.01

a network based model is more consistent with the S-shaped curve for cumulative adoption, similar to the classic model of Bass (1969).

Overall, the empirical evidence of Table EC.6 does not support the single source non-network model, since this model does not capture the primary data patterns of adoption across the village networks.

EC.5.2. Broadcast during Initial Period

Our model proposed that information about microfinance propagates through word of mouth over the social network. We consider a benchmark (null) model where the information is *broadcast* to households initially in period 0 at a meeting, where the attendance at the meeting is probabilistic. In such a model, information flows directly from a common source to any of the households in the network (subject to their attendance at the meeting), and the *structure of the social network is not relevant for this initial communication.* After this initial broadcast, regular WOM communication occurs through the social network in subsequent periods. As in our main models, informed households have the opportunity to make an adoption decision, whereas non-informed households cannot do so.

We explain why modeling the initial broadcast mechanism would only strengthen our qualitative conclusions about the relative superiority of local friend seeding. In a model with the broadcast mechanism, we should attribute *some* part of the adoption in early periods to that meeting rather than organic household-to-household word of mouth. This implies that the word-of-mouth driven trajectory would be *even lower* in earlier periods, which further implies that the leader fixed effect would be more negative. Taking this logic to the counterfactuals where non-leaders are chosen as seeds, we would therefore see a further increase in relative performance of the local friend seeding strategy and other non-leader strategies when compared to the leader strategy that generates the data.

We now demonstrate this argument by estimating the model that allows for such an initial broadcast and performing the counterfactual. To reiterate, we restate the modeling assumptions.

- In period 0, the leaders invite members of the village to an initial meeting where they explain the microfinance product. For each household, the probability of attending the initial meeting is γ. We refer to this as the initial broadcast effect.
- In addition to the above, regular word of mouth communication happens through the social network, as we have specified in the main model.
- We consider the case when there are separate leader fixed effects, and the case where this effect is absent.

There are a few points to consider here. First, there is the question of separate identification of an initial broadcast effect from the leader fixed effect. Here, the separation is possible because the broadcast is a one-time initial event, whereas the impact of the leader fixed effect continues beyond the initial period. Thus, if we have 3 or more periods, we can identify both effects. Next, a higher level of leader fixed effect (q_ℓ) will lead to more friends of leaders being informed (relative to non-friends), and leads to higher adoption among friends of leaders. In contrast a higher value of initial broadcast effect γ informs households who are not friends of leaders, leading to higher adoption among that group.

Second, if the probability of attending the initial meeting is very high, $\gamma \approx 1$, then there is little role for the network in communication. The model is then similar to the single source null model of §EC.5.1. More generally, the higher the broadcast effect, the less important are the structure of network connections. Third, in the counterfactual where we choose non-leaders as seeds, we might expect this initial broadcast to be less likely or absent. When leaders have a unique ability to do bring about such a broadcast that non-leaders do not possess, then in the counterfactual, we would set the broadcast parameter $\gamma = 0$. Of course, this assumption *stacks the deck against* any of our proposed strategies, but we include it to show that our strategies still perform better than the leader strategy.

Table EC.7 details the parameter estimates from the initial broadcast model. We find that the results are qualitatively very similar to that of our main model, and quantitatively the relative magnitudes and ordering between the parameters are also the same. For instance, non-adopters communicate less than adopters, and degree is negatively correlated to probability of communicating with a network neighbor. In the model with Leader fixed effects, the leader fixed effect is negative but not significant. We also find the initial broadcast effect, which represents the probability that each household attended the initial meeting to be $\gamma = 0.013$ or $\gamma = 0.079$ depending on whether the model includes leader fixed effects or not.

Table EC.7 Initial Broadcast Model Estimates

Parameter	Symbol	Model Specification: Estimates (Standard Errors) $\mathbf{M_5^{L=0}}$ (No Leader FE)	$\mathbf{M_5^{L=1}}$ (with Leader FE)
Non-adopter lowest degree	q_{min}^{NA}	0.103 (0.062)	0.073 (0.062)
Non-adopter highest degree	q_{max}^{NA}	0.063 (0.087)	0.071 (0.089)
Adopter lowest degree	q_{min}^{A}	0.401 (0.034)	0.392 (0.090)
Adopter highest degree	q_{max}^{A}	0.314 (0.100)	0.259 (0.129)
Leader Effect	q_ℓ	– –	-0.001 (0.080)
Initial Broadcast Effect	γ	0.013 (0.030)	0.079 (0.030)

Table EC.8 Leverage with Initial Broadcast Model

Strategy	*Seeding at:*	No Leader FE 0.50%	1.00%	5.00%		With Leader FE 0.50%	1.00%	5.00%
Local Friend		1.31	1.21	1.10	1.31	1.21	1.10	
(Firm's) Leader		1.20	1.15	1.06	1.20	1.15	1.06	
Like Leader		1.18	1.13	1.06	1.18	1.13	1.06	
Hybrid Strategies:								
Friend of Leader		1.32	1.22	1.10	1.32	1.22	1.10	
Leader Friend of Leader		1.21	1.14	1.07	1.21	1.14	1.07	

Table EC.8 reports how the counterfactual strategies perform relative to random seeding using the estimates from the Initial broadcast model. The ratio (leverage) of 1 indicates that the strategy performs just as well as random.

Further, we observe that:

- Leverage reduces as the seeding proportion increases, similar to the main model.
- At all seeding proportions $(0.5\%, 1\%, 5\%)$, the results show that incorporating the initial broadcast effect does not change the relative performance of the Random, Local and Leader seeding strategies, and this holds *with* or *without* leader fixed effects.

EC.5.3. Leader Based on Occupation

In our main model, we made the assumption that the set of households with leader fixed effects is the same as the seed set, as all of the seeds were considered leaders in their villages. But as per Banerjee et al. (2013), the microfinance firm chose its seeds based on whether they were in certain "leader" occupations (e.g., teachers, shopkeepers, business owners etc). However, not all households with those occupational characteristics were chosen as seeds.

It may then be reasonable to consider a specification where the seeds are assumed to be randomly sampled from those working these selected occupations. In this case, the leader fixed effect should be associated with all members belonging to these occupations, *whether they were used as seeds or not*.

We estimate such a model and obtain parameter estimates $\theta = (q^{NA}_{min}, q^{NA}_{max}, q^{A}_{min}, q^{A}_{max}, q_l) = (0.074, 0.056, 0.424, 0.344, -0.016)$. We then run counterfactuals under different seeding strategies as before.

The counterfactual results presented in Table EC.9 show that the results are quantitatively similar and qualitatively identical. The results for all the strategies differ from the main results of Table 10 because the parameter estimates used above are different.

Table EC.9 Leverage for Counterfactual Strategies

Strategy \ Seeding Level	0.5%	1%	5%
Leader by Occupation	1.17	1.13	1.03
Local Friend	1.27	1.19	1.05
Friend of Leader	1.27	1.24	1.05
Leader Friend of Leader	1.16	1.15	1.04
Like Leader	1.16	1.11	1.04
Top Degree	1.53	1.40	1.08
Top Diffusion	1.53	1.37	1.07

EC.5.4. Leader Certification Effect

One concern is if the leader seeds chosen by the firm had an additional certification effect beyond their differential WOM communication that we have already modeled as the leder fixed effect. The original study seeded information with leaders who were pre-defined. Thus, there might be a question of whether such an effect may be present in the counterfactual, where seeding is not focused on leaders. Such a *"leader certification effect"* effect posits that *households might be more likely to adopt if they hear through word of mouth that a leader has certified or endorsed the product.*

We detail two arguments below to demonstrate that this concern is unlikely to hold in the counterfactual. First, for this concern to be valid, none of the seeds recommended by the counterfactual strategy should overlap with the "leader characteristics" of seeds chosen in the original study. To the extent, those choices were made based on certain occupations and other characteristics, this is unlikely that our seeding strategies did not have overlap with the chosen occupations.

Second, we quantify the overlap in leaders (which are originally chosen by the firm) across the strategies, and demonstrate that such leaders are chosen even in the counterfactuals, although more under some strategies than others.

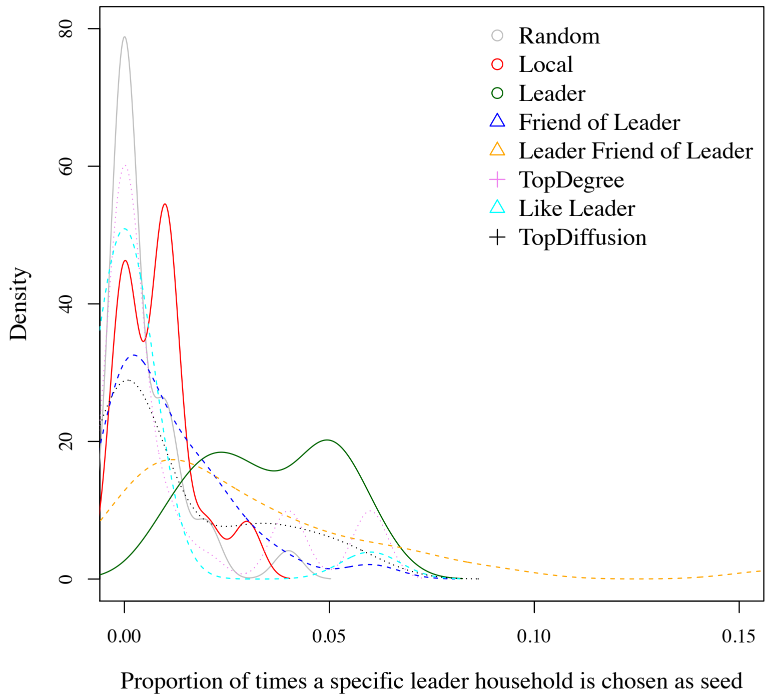


Figure EC.3 Seeding Overlap (1% seeding)

We detail the degree of overlap with a density plot in Figure EC.3 that details the probability that any given leader household will be chosen under each of the counterfactual strategies.[14] We focus on seeding at 1%, the results are qualitatively similar at other levels. There are a few noteworthy observations here. First, we do expect that almost all leader households have a higher probability of being chosen as seeds under the Leader strategy, since the seeding is limited to leader households here. Second, as expected, the Random strategy (in gray) has the lowest probabilities of these leader households being selected. Third, observe that leader households have a non-zero probability of being chosen under each of the counterfactual strategies, notably the Local Friend strategy. Finally, each of the other (non-random) counterfactual strategies have a higher probability of leader households being selected compared to the random strategy. More specifically, observe that the comparison of interest is the difference between the Leader strategy (in green) and the Local Friend strategy (in red).

Overall, we believe that above result indicates that our counterfactual outcomes are unlikely to be biased due to the potential for some *unobserved leader certification* ability in the chosen seeds.

[14] The main strategies (Random, Local Friend and Leader) are in solid lines, whereas the hybrid strategies are in dashed lines and the network information strategies are dotted line format.

EC.5.5. Generalizing the Adoption Model

The adoption model in the main paper was chosen based on model fit across models that incorporated household data, as well as network characteristics of households derived from the social network within each village. We detail the results in Table EC.10. The results suggest that the household characteristics including number of rooms and beds and indicator for electricity are informative for adoption, and produce a better model fit, as measured by the Akaike Information Criterion (AIC). The AIC is a well regarded measure of fit that is commonly used since it balances model complexity with a likelihood based model fit, unlike measures like likelihood and Pseudo R^2.

Specifically, as we add more predictors (electricity and latrine) in moving from model (1) to model (2), the fit as measured by AIC increases. Similarly in adding Rooms and Beds per capita to obtain model (3), the fit improves. However, we find that the household's home ownership and roof type are not significant predictors of microfinance adoption and the models (4) and (5) that include these variables are worse in terms of AIC than model (3).

A few observations are relevant here. First, we note that pseudo-R^2 values are not comparable to R^2 values for linear models, and are typically much lower, as noted by Guadagni and Little (1983). Second, the adoption model is estimated using data from seeded leader household. The microfinance product is a financial one and is typically not targeted at leader households, but other households close to the bottom of the income pyramid. Therefore, the explanatory power of observables is typically lower for leader households. Finally, though the ability to predict whether any particular household will adopt is low, the cumulative adoption by integrating the adoption probabilities across all households should have less variance and therefore better fit. Many of these diffusion paths are equivalent in terms of the overall adoption since households are more likely to be connected to households that are similar to them in terms of network position.

Table EC.10 Comparison of Adoption Models

	Dependent variable:				
	Adoption				
	(1)	(2)	(3)	(4)	(5)
Constant	-0.560^{***} (0.149)	-1.292^{***} (0.320)	-1.210^{***} (0.322)	-1.353^{***} (0.436)	-1.664^{***} (0.612)
Number of Rooms	-0.196^{***} (0.058)	-0.161^{***} (0.059)	0.007 (0.085)	0.005 (0.085)	0.004 (0.086)
Number of Beds	-0.108^{*} (0.061)	-0.063 (0.062)	-0.283^{**} (0.143)	-0.284^{**} (0.143)	-0.258^{*} (0.144)
(No) Electricity		0.174 (0.122)	0.156 (0.123)	0.158 (0.123)	0.128 (0.125)
(No) Latrine		0.160^{**} (0.080)	0.179^{**} (0.080)	0.181^{**} (0.080)	0.163^{**} (0.082)
Rooms Per Capita			-1.023^{***} (0.392)	-1.015^{***} (0.393)	-0.989^{**} (0.394)
Beds Per Capita			1.147^{*} (0.656)	1.153^{*} (0.655)	1.075 (0.667)
Own Home				0.147 (0.299)	0.142 (0.304)
Roof Type 1					0.729 (0.736)
Roof Type 2					0.507 (0.442)
Roof Type 3					0.330 (0.443)
Roof Type 4					0.397 (0.453)
Roof Type 5					0.150 (0.482)
Roof Type 6					1.974 (1.561)
Roof Type 7					2.109 (1.497)
Roof Type 8					-11.716 (375.669)
$PseudoR^2$	0.02	0.026	0.032	0.032	0.037
Observations	1,140	1,140	1,140	1,140	1,140
Log Likelihood	-610.256	-606.784	-603.093	-602.970	-599.684
Akaike Inf. Crit.	1,226.513	1,223.567	1,220.185	1,221.940	1,231.368

Note: *p<0.1; **p<0.05; ***p<0.01.

The log-likehood LL_0 for a model with just the intercept is -622.878 and Pseudo $R^2 = 1 - \dfrac{LL}{LL_0}$.

EC.5.6. Parameter Sensitivity to Moments

We have explored a wide variety of alternative model specifications and assumption for both the adoption and word of mouth communication processes. However, an interested reader might be interested in testing robustness to an alternative they have in mind that might be quite different. This issue is explored in detail by Andrews et al. (2017), who provide a unified framework to help make structural (and other) models more transparent so that readers can easily evaluate sensitivity to assumptions. They recommend providing a sensitivity matrix (Λ) that allows us to evaluate how violations of specific moment conditions can change the model parameter estimates.

One might view the approach of Andrews et al. (2017) as complementary to specifying different models to demonstrate robustness, which we have also done. Their point is that since it is impossible to test *all potential alternative models*, providing the sensitivity matrix allows any interested reader to determine how each of the data moments contribute to parameter estimates. Following Andrews et al. (2017), we report the sensitivity matrix Λ in Table EC.11.

Similar to the applications presented in Andrews et al. (2017), we scale the values so that the sensitivity values correspond to a 1% change in each moment condition. These results can be helpful in evaluating the sensitivity of each parameter on each of the moment conditions used in estimation.

First, observe that we use simulated method of moments (SMM), which is a (simulated) version of Generalized Method of Moments (GMM) to obtain the parameter estimates. This allows for the form of the sensitivity matrix

$$\Lambda = -(G'WG)^{-1}G'W$$

where G is the Jacobian corresponding to the moment conditions $g(\theta)$ and W is the weighing matrix used in the GMM estimation. The main result of Andrews et al. (2017) is that the asymptotic bias of local violations of the moment conditions is then given as

$$\mathbf{E}(\tilde{\theta}) = \Lambda \mathbf{E}(g)$$

so that knowing Λ allows us to determine how violations of the moment conditions g translate into differences in parameter estimates.

The sensitivity matrix corresponding to the baseline model with the leader fixed effect are detailed in Table EC.11. For each parameter, the table shows the sensitivity of parameter estimates to violations of the moment conditions. We have 9 moments that are each present across all the villages in the data, and in the model the moments are generated from the parameter values. Overall, there are 6 cross sectional moments, and 3 time series moments (period 1, period 2 and period 3).

Table EC.11 Plug-in Sensitivity Λ for Model with Leader FE Effect

	Cross Sectional						*Time Series*		
Moment # $\rightarrow$	M1	M2	M3	M4	M5	M6	M7	M8	M9
q^{NA}_{min}	0.009	0.016	0.004	0.004	0.005	0.003	0.001	0.022	0.011
q^{A}_{min}	0.003	0.005	0.001	0.001	0.002	0.002	0.000	0.006	0.003
q^{NA}_{max}	0.008	0.050	0.005	0.001	0.018	0.014	0.000	0.090	0.037
q^{A}_{max}	0.013	0.012	0.011	0.013	0.004	0.011	0.002	0.003	0.005
q_ℓ	0.001	0.006	0.001	0.005	0.001	0.002	0.000	0.003	0.002

To understand the table, consider for example the parameter q^{NA}_{max}, which represents the word of mouth communication probability for (high degree) non-adopters. First, observe that the cross-sectional moment this parameter is most sensitive to is moment 2, which is the proportion of households with no adopting neighbors who have adopted. Similarly we find that the parameter is not sensitive at all to moment 7, which is the initial adoption. This is consistent because households adopting in the first period are unlikely to be hearing about it from non-adopters. In the earliest period, only the seeds have a chance to adopt, and the seeds are directly informed by the firm, so word of mouth among non-adopters is unlikely to play a role. In contrast, the adoption in periods 2 and later are relevant to the parameter, especially in contrast to the initial period. If initial period adoption is low (consider the extreme case of zero adoption in the initial period), then adoption in later periods must be driven by word of communication, which then informs both parameters q^{NA}_{min} and q^{NA}_{max}.

Similarly, we find that for the leader fixed effect, moment 2 and moment 4 are most important. The time series moment 7 (adoption in period T=1) **does not contribute any information** about the leader fixed effect. This is consistent with our intuition since the leader fixed effect is communication by the leaders to their friends, and thus, the level of period 1 adoption will not be informative of how much communication has occurred through the network by leaders. In contrast, period 2 adoption (moment 8) is important since we would find a greater jump in the early adoption trajectory when the leader fixed effect is greater.

Overall, this method provides transparency in illustrating what variation in the moments is driving the parameter estimates of the model.