UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Estimating diffusion and adoption parameters in networks

New estimation approaches for the latent-diffusion-observed-adoption model

Stephan, L.S.

Publication date
2021

Link to publication

Citation for published version (APA):
Stephan, L. S. (2021). *Estimating diffusion and adoption parameters in networks: New estimation approaches for the latent-diffusion-observed-adoption model*. [Thesis, fully internal, Universiteit van Amsterdam].

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:15 Nov 2026

Chapter 2

The Two-period Estimator

2.1 Introduction

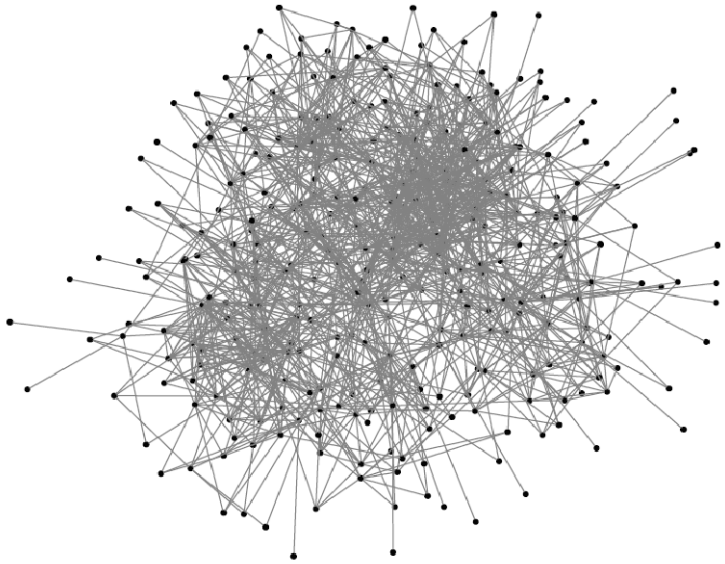
In their prominent paper, Banerjee et al. (2013) investigate how knowledge about a newly initiated micro-finance project spreads through social networks in Indian villages, starting from a small set of initially informed individuals. Upon receiving the information, villagers have the opportunity to take a loan. The authors use data on initially informed individuals and microloan-borrowers provided by the NGO together with village network data accumulated by a survey to draw inference on the hidden information diffusion.

Knowing information diffusion rates is important for investigating potential causes of a sub-optimally low take-up rate. Word-of mouth being a common marketing practice, a low take-up may be caused by low information diffusion rates or by unattractive features of the loan product. Only if we disentangle one from the other, we can improve marketing practices to maximise outreach. Further, since failure to take-up a loan may be a result of an individual choice or a lack of information, once diffusion is accounted for, the determinants of individual loan decisions can be identified. This in turn is crucial if we want to understand, study and model micro-finance demand in order to improve financial products offered.

Yet, quantifying diffusion rates is challenging. As can be seen from Figure 2.1, the village networks are typically dense. This implies that the number of scenarios, how the information could have propagated through the system to create the observed participation pattern, explodes when the time horizon increases.

Banerjee et al. (2013) circumvent this problem by using Simulated Method of Moments (SMM) which consists of creating a large, artificial data-set by simulating the two processes (information diffusion and loan adoption) on the observed village networks for

Figure 2.1: **Village Network Graph (village 52)**



a given parameter combination and choosing the parameter combination that, when averaging over all simulation runs, best replicates the aggregate participation rate.
I instead establish and maximize the probability to observe the individual outcomes. By means of the law of total probability, I construct the log likelihood function dependent on the data and the parameters by integrating out the latent information status variables. If I restrict myself to two time periods, the number of scenarios is manageable, the log likelihood function can be evaluated exactly and the peak can conveniently be identified by grid search. Hence, in the current chapter, I only model two time periods. A larger time horizon will be analysed in subsequent chapters.
My procedure has several unique advantages compared to Banerjee et al. (2013). First, Banerjee et al. (2013) use a two-step procedure in which the adoption parameters are identified from the set of initially informed individuals (IPs) and thereafter diffusion parameters are estimated by SMM estimation. Using the optimality of the ML estimator, I, on the other hand, estimate all parameters jointly and therefore gain in precision. Second, while they match only final aggregate rates, I take the network position of individuals into account. Third, my estimation runs substantially faster and enables me to conveniently apply the Likelihood Ratio (LR) test for hypothesis testing and confidence set constructing from running the model just once, while Banerjee et al. (2013) need to bootstrap their standard errors.
Comparing my estimation results to those stated in Banerjee et al. (2013) or those resulting from applying their algorithm to the shortened time horizon generates new interesting and important insights. In line with Banerjee et al. (2013), I find a signifi-

cantly higher rate of information spread for participants than for non-participants and a significant impact of certain household characteristics. The fact that diffusion rates are relatively high induces me to believe that low final participation rates are primarily a product design not a marketing problem. However, I estimate lower spreading rates for non-participants than the Banerjee et al. (2013) algorithm.

Following Banerjee et al. (2013), I extend the model by allowing individuals to not only pass on information but also to directly impact each others' choices. Since my estimation technique refrains from simulating the information flow, I cannot test for the "endorsement effect" as specified in Banerjee et al. (2013), namely the effect that those previous participants that did transmit the information have on their neighbours' take-up probability. Instead, I estimate the effect of having friends that previously participated (regardless of whether or not these did send the message) on individual adoption probabilities. Doing so, I find that the individuals' participation probabilities significantly increase in the individuals' links to past participants, while Banerjee et al. (2013) find that the number of messages received from participants does not significantly impact the individuals' choices. I thus conclude that increasing initial adoption has a twofold benefit: first, by the high subsequent diffusion rate and second, due to the positive endorsement effect.

Hindus and richer households are less likely to participate. Not including household characteristics leads to a substantial over-estimation of the diffusion rate, in particular for non-participants. Neighbours opting in one after another may erroneously be attributed to a network effect, when in fact it is their similarity in socio-economic characteristics that causes both of them to have higher adoption rates. It is thus crucial to make sure that individuals that are likely to participate upon receiving the information can actually be reached by it.

These findings are particularly important given the growing interest of development economists in network effects. Various qualitative and quantitative studies have highlighted the importance of social ties and personal connections for the spread of new technologies and practices in developing countries. Common findings are that (i) the access to and availability of new technologies alone is insufficient for it to become commonly used and adopted, that (ii) the pattern of the adoption process is correlated with the pattern of information diffusion and that (iii) social networks matter for both, but more so for the adoption process. See, for instance Ryan and Gross (1950), Bandiera and Rasul (2006), Katungi, Edmeades, and Smale (2008) or Todoa, Yadateb, Matousc, and Takahashid (2011). What distinguishes my work from the studies mentioned is that (i) information diffusion is unobserved and (ii) I take into account not just network statistics, but the precise network location of each individual agent. Often, network

data is accumulated in the respective study's data collection procedure, but then not fully used. The gap between the potential usage of data and its actual usage reveals shortcomings and thus provides scope for further research.

While the results from this particular estimation are relevant and interesting, my algorithm, can be used to estimate any situation in which an unobserved network interaction (a latent "signal" spreading through the network) impacts a partly observed individual outcome process (a choice to be made when receiving the signal that is only visible if it leads to a change in outcomes). In development economics, it can be used to investigate the spread of technological innovations and improvements in production techniques, changes in beliefs and household choices such as schooling or vaccination or knowledge spill-over and a change in health or nutrition outcomes. Examples in other areas are inter alia informal lending and bankruptcy, knowledge on job openings and posting of applications, consumer knowledge and purchase decision or opinion transmission and protest participation.

The remainder of the chapter is organized as follows. Section 2 outlines the model and assumptions. Section 3 introduces the estimation technique and compares it with the estimation procedure of Banerjee et al. (2013). Section 4 gives an overview of the models and the hypotheses to be tested. Section 5 discusses the data. Sections 6, 7 and 8 contain the estimation results: I first estimate two simple models that feature homogeneous participation probabilities, I then turn to models including a peer influence (the "endorsement effect") and finally I consider two models take into account household characteristics. Section 9 points out the scope for further research and Section 10 concludes.

2.2 Model and Assumptions

The data provided by the micro-finance organisation and the network survey comprises 37 villages, indexed by $v = 1, ..., 37$.

The N_v individuals in village v are linked in a network, represented by the network adjacency matrix G_v. While villagers $i = 1, ..., N_v$ entertain relationships among themselves, they do not interfere with agents living in other villages. The symmetric matrix G_v comprises binary variables indicating the presence (absence) of a link. If $g_{v,ij} = g_{v,ji} = 1$, we refer to i and j as "neighbors".

The aim is to study the effect of a micro-finance institution entering the community and starting to provide loans to its members. In period $t = 0$, they advertise their loan product to a subset of them (referred to as the "information injection points" or "IPs")

and thereafter rely on word-of-mouth marketing. Having received the information, the IPs can choose to participate. Both the network and the information initiation are exogenous and observed.
Subsequently, two processes occur. First, the IPs can inform their neighbors about the novel opportunity. Second, the newly informed individuals choose whether or not to take up a loan.
The process is modeled using the random matrices Y_v and S_v, containing dummy variables for respectively the participation and information statuses of each of the villagers at each point in time. $Y_{v,i,t} = 1$ ($S_{v,i,t} = 1$) (i.e. the row i, column t element of Y_v (S_v) equals one) implies that individual i in village v participates (is informed) in period t. Accordingly, $S_{v,i,0} = 1$ implies that i is an IP. The distributions of the random matrices Y_v and $S_{v,1}$ depend on the network and the information initiation (as outlined in the next section).
For each village, we observe the network, G_v, the information initialization (i.e. one realization of $S_{v,0}$) and the individual participation decisions over time (i.e. one realization of Y_v). The information statuses of all inhabitants but the IPs are generally unobserved (i.e. $S_{v,1}$ is hidden). In particular, there usually exist various realizations of $S_{1,v}$ that are in accordance with the data at hand.
The following assumptions are made:

Assumption 1: Independence across villages and common data-generating process

Villages are independent entities. In particular, there is no link between any two agents that reside in different communities.
The data of all villages stems from the same underlying data-generating process.

Assumption 2: Exogeneity of the network and the information initiation

The network is exogenous, fixed, observed and not affected by the intervention.[1] The measurement error is negligible.
The choice of the set of IPs is exogenous. In particular, on average, IPs have the same probability to participate as informed non-IPs.

Assumption 3: Timing

[1] It is plausible that the intervention does alter links in the long run, but not in the time horizon that I study.

In period $t = 0$, the IPs are informed. In period $t = 1$, they make their adoption decision and potentially spread the information. In period $t = 2$, the individuals that have been newly informed by the IPs choose whether or not to participate.

Assumption 4: Participation opportunity

An individual can only choose (not) to participate after being informed.

Assumption 5: Participation is an one-time opportunity

In period 2 only newly informed individuals face the participation decision. Having opted in (out), the respective individual will thereafter stay in the set of participants (non-participants).

Assumption 6: Distributional assumption for the outcome

Conditional on being newly informed, the random variables $Y_{v,i,t}$ $\forall v = 1, ..., V;\ i = 1, ..., N_v;\ t = 1, 2$ are independent and follow a Bernoulli distribution with individual specific mean $p_{v,i}$ as specified beneath.

Assumption 7: Information exchange

Individuals can only switch their information status once (from being uninformed ($S_{v,i,t-1} = 0$) to being informed ($S_{v,i,t} = 1$)). Any informed individual can keep on informing her neighbors every period.
The initially informed individuals may transmit the information to their neighbors. Transmitting is independent across senders and receivers.
Let the random variable $I_{v,ij}$ denote the indicator that shows that individual i has sent the information to individual j. Conditional on i being a sender, the variables $I_{v,ij}\, \forall i = 1, ..., N_v; j = 1, ..., N_v; i \neq j; v = 1, ..., V$ are independent and follow a Bernoulli distribution with mean $E[I_{v,ij}] = q_{v,j}$ as specified beneath.

I use a series of nested models to investigate whether and which network effects play a role. The different models distinguish themselves with respect to how I model $p_{v,i}$ and $q_{v,i}$ (i.e. the probability to participate upon being newly informed and the probability to spread information for a particular individual). Precisely, $q_{v,i}$ may depend on the sender's participation status while $p_{v,i}$ can be a function of household characteristics

or previously participating friends. The concrete models are outlined in section four.

2.3 Maximum Likelihood Estimation

2.3.1 Potentially Informed Individuals (PIIs) - the origin of the dimensionality problem

In the following, I drop the subscript v to ease notation. Further, I use uppercase letters to denote random matrices or variables and lowercase letters to denote a concrete realization. Let θ be the vector of model specific parameters that are used to model the probability to send information and the probability to participate upon being newly informed. Using the notation introduced above,

$$\ell(\theta) = Pr\big(Y_1 = y_1, Y_2 = y_2 | S_0 = s_0, G, \theta\big) \tag{2.1}$$

is the probability to observe the actual data at hand, given the (nonrandom) village network, the parameters and the information initiation (i.e. the known set of IPs) and

$$\mathcal{L}(\theta) = \log\ \Big(Pr\big(Y_1 = y_1, Y_2 = y_2 | S_0 = s_0, G, \theta\big)\Big) \tag{2.2}$$

is the village log likelihood function. Note that here Y_t is the vector of observed outcomes in period t and S_0 indicates which individuals belong to the set of IPs.
The probability to observe any particular outcome pattern depends on the (unobserved) information propagation. Knowing the odds of (not) observing somebody as a participant depends on her probability to receive the information which in turn depends on which of her neighbors are informed. In the following, I use the term "information scenario" to denote a unique vector $S_1 = s_1$. The set $\mathbf{S}$ comprises all possible information scenarios. Then the log likelihood function is (by the law of total probability)

$$\mathcal{L}(\theta) = \log\left(\sum_{s \in \mathbf{S}} Pr\Big(Y_1 = y_1, Y_2 = y_2, S_1 \in s | S_0 = s_0, G, \theta\Big)\right) \tag{2.3}$$

In period 1, it is known who is informed at the time of the adoption decisions. In period 2, however, different information scenarios can arise. Let ℓ_t denote the unconditional probability to observe $Y_{1:t} = y_{1:t}$ given the network and the information initiation and let $\ell_{t(s)}$ be the specific probability to observe $Y_{1:t} = y_{1:t}$ conditional on the fact that scenario s has occurred. Then $\Delta\ell_{t(s)}$ is used to denote the change in the likelihood when we consider not just period $t-1$, but also period t conditional again on scenario

s having occurred. The unconditional probability of the outcome can be obtained by the law of total probability, i.e. by establishing the joint probability of the outcome and latent information status variables and integrating out the latter. (I establish the probability to observe the outcome conditional on a specific scenario, weight this term by the scenario's probability to occur and sum over all scenarios.) Further, given a specific scenario has realised, individuals are independent. This yields

$$\ell_2 = \ell_1 \times \sum_{s \in \mathbf{S}} \Delta\ell_{2(s)}, \qquad \Delta\ell_{2(s)} = \prod_{i=1}^{N} \Delta\ell^i_{i,2(s)} \tag{2.4}$$

Where $\Delta\ell_{i,t(s)}$ is used to denote individual $i's$ contribution to the likelihood function in period t given scenario s has occurred. For all non-IPs, the information status at time $t = 2$ is a random variable with mean

$$r_i = P(S_{i,1} = 1 | S_{i,0} = 0, G, \theta) \tag{2.5}$$

which is i's probability to newly receive the information (from any of her neighbors). I can categorise individuals at $t = 1$ into four groups according to what they contribute to $\Delta\ell_2$:

1. **IPs:** $\Delta\ell_{i,2} = 1$
   The IPs have made a choice at $t = 1$ and, according to Assumption 5, these individuals are certain to remain (informed) participants or non participants respectively.

2. **new participants:** $\Delta\ell_{i,2} = r_i p_i$
   These individuals are observed to experience a switch in their outcome, which (according to the modeling assumptions) must be preceded by a switch in their information status.

3. **non-participants out of reach** $\Delta\ell_{i,2} = 1$
   These individuals are too far away from any IP to possibly have received the information and thus certain to be uninformed.

4. **potentially informed individuals (PIIs)**
   Sufficiently close to an IP to potentially receive the information, they yet do not indicate (by participating) that they have done so. As a consequence, while the likelihood contribution for the three groups above is unequivocal, for the PIIs, two scenarios (each leading to a specific likelihood contribution) are indistinguishable in terms of the empirical outcome that they generate.

- **scenario A: newly informed PIIs** $\Delta\ell_{i,2} = P(A)_i = r_i(1-p_i)$
  This group has received the information in the last exchange and opts out this period.
- **scenario B: uninformed PIIs** $\Delta\ell_{i,2} = P(B)_i = (1-r_i)$
  They entered last period's information exchange uninformed and also did not receive the information.

We can thus see that the PIIs are the origin of the dimensionality problem: their information status cannot be derived from the data and thus all different scenarios need to be considered.
Since i is informed as long as *any* of her neighbors sends the information, we can use the counter-factual probability (nobody informs i) to calculate

$$r_i = 1 - \prod_{j=1}^{N} \underbrace{(1 - g_{ij} q_j s_{j,0})}_{\substack{\text{if } j \text{ is an informed neighbor of } i, \\ \text{she does not send the information}}} \tag{2.6}$$

which is one minus the probability that nobody informs i.
Unable to observe S_1, we can not discern whether or not a PII is already informed. The identity of the PIIs is pinned down by the information in s_0 and G: all those individuals that have a positive probability to be informed $r_i > 0$ but do not participate constitute the set of PIIs. For those individuals in this set, the scenarios A: "newly informed and opted out" and B: "uninformed" are observationally equivalent. With two individual scenarios, each of which can be combined with everybody else's information status, the number of potential subsequent information status vectors doubles whenever another individual joins the set of PIIs.
The classification of the agents into the aforementioned categories enables me to straightforwardly obtain a tractable formula for (2.3). For simplicity, let $i = 1, ..., N_{IPP}$ and $j = 1, ..., N_{IPN}$ denote the set of participating IPs and non-participating IPs, respectively. Further, let $k = 1, ..., N_{P2}$ and $l = 1, ..., N_{PII}$ denote the set of period 2 participants and PIIs, respectively.
First observe that for the first three categories of agents, the contribution to the log likelihood function is constant and the same for all scenarios, thus

$$\ell_2 = \ell_1 \times \prod_{k=1}^{N_{P2}} \Delta\ell_{k,2} \times \Big(\sum_{s \in \mathbf{S}} \prod_{l=1}^{N_{PII}} \Delta\ell_{l,2(s)} \Big)$$

Since for any PII both scenarios A and B can be combined with any other PII's scenarios A and B, this leads to

$$\ell_2 = \prod_{i=1}^{N_{IPP}} \ell_{i,1} \times \prod_{j=1}^{N_{IPN}} \ell_{j,1} \times \prod_{k=1}^{N_{P2}} \Delta\ell_{k,2} \times \prod_{l=1}^{N_{PII}} \Big(P(A_l) + P(B_l)\Big)$$

Substituting in the log-likelihood contributions as outlined above I obtain

$$\ell_2 = \underbrace{\prod_{i=1}^{N_{IPP}} p_i}_{\substack{\textit{participating} \\ \textit{IPs}}} \underbrace{\prod_{j=1}^{N_{IPN}} (1-p_j)}_{\substack{\textit{non-} \\ \textit{participating} \\ \textit{IPs}}} \underbrace{\prod_{k=1}^{N_{P2}} r_k p_k}_{\substack{\textit{period 2} \\ \textit{participants}}} \underbrace{\prod_{l=1}^{N_{PII}} \Big(r_l(1-p_l) + (1-r_l)\Big)}_{\textit{PIIs}}$$

2.3.2 Computational Aspects

From the above it is apparent that establishing the log likelihood function is computationally challenging in particular when participation rates are low and the network is dense as this translates into a large number of PIIs. Tractability can be achieved by restricting the time horizon to two periods: in this case, PIIs are endpoints of the information flow, whether or not they are informed has no impact for other individuals' information reception rates and therefore this rate is unique for each individual. When the time horizon is larger, reception rates are scenario specific. For $T = 2$, reception rates are established conditioning on the observed past information status vector S_0, while for $T > 2$ conditioning must be done with respect to the unobserved vector S_1. This implies that each information scenario needs to be considered separately and thereafter, scenario-specific probabilities to observe the data sum up into the likelihood function. Consequently, unlike the $T = 2$ case, the likelihood does not factor into individual-specific components. Instead, individuals contribute differently to each scenario they are involved in.

Even in the case with two time periods, establishing the log likelihood function entails a large number of computations in particular for the richer models. I advocate that using C++ and in particular some specific libraries enables me to execute these computations in reasonable time. The boost library facilitates saving to (reading from) files without having to worry about the dimensions and the nature of the object to be stored. By means of the dropbox uploader, written files can automatically be stored online. Given the easiness of saving (village level) results, the villages can be run in parallel, results can be written to file and later combined. Due to the possibility to store online, there is practically no restriction on the size or number of villages.

Beyond its speed and efficiency, the C++ programming language has the advantage to be completely software independent, since code can be compiled using freeware and thereafter run on any machine. Furthermore, it offers the choice to increase the storage for variables, thus decreasing rounding errors and increasing precision. Despite being fast and efficient, the C++ language is rarely employed in economics, mainly because it requires a high investment in coding. Therefore, I believe the estimation method I provide to be a highly useful and attractive tool. The programs I provide fit a general class of network models and can readily be used, thus relieving applied researchers from the coding work.

2.3.3 Identification of the maximum by grid search

I proceed by evaluating (2.3) over a grid of values, which thereafter is refined around the optimum. There are several reasons for me to opt for the contracting grid search as opposed to analytical or numerical optimisation methods.

First and foremost, grid search is more robust in the sense that it guarantees that the global maximum will be found even if the objective function is ill-behaved. In presence of multiple peaks, iterative maximization algorithms on the other hand may erroneously converge to a local maximum in the neighbourhood of the starting value. Thus if possible, grid search should be preferred and the only argument against it is the computational speed. My efficient algorithm enables me to carry out a grid search in reasonable time, thus making it computationally affordable.

Noting that the objective function is a high order polynomial in the parameters, that changes its entire functional form whenever there is the slightest change in the data, there are further advantages.

Naturally, one might wish to solve the optimisation problem analytically, by establishing the first order conditions (FOCs), equating them to zero and solving for the parameter values. However, an analytical solution to the system of FOCs and thus a closed-form solution for the parameters may not exist. This implies that one must revert to numerical methods to find the parameter values for which the FOCs are as close to zero as possible. This process of substituting parameters with concrete guesses in all FOCs, evaluating the resulting change in the value of the objective function against a stopping criteria, adjusting the guesses and repeating this process potentially numerous times is very slow and demands large amounts of computer memory. Believing grid search to be slower is thus a presupposition rather than a fact in this context. Further, for iterative numerical maximisation methods such as the Newton-Raphson algorithm, the adjustment of the parameter guesses is usually done using the Hessian. These

optimisation techniques work well if the function is approximately quadratic. For the case at hand, however, a second order Taylor expansion provides a poor approximation. The Hessian may even be singular close to being non-invertible thus making Newton-Raphson methods unreliable.
Another argument against an optimisation method that relies on FOC minimization is the potential occurrence of corner solutions for which the FOCs will fail to hold.
Additionally, for numerical or analytical maximisation methods, problems may arise not just in the estimation, but also for inference and hypothesis testing as one relies on asymptotic theory to compute standard errors used for hypothesis testing. In a small sample, this can easily result in false conclusions in particular when the parameters are weakly identified. An alternative is to bootstrap the standard errors. Given the large time and memory requirements to run just one maximisation round, the number of bootstrap repetitions would have to be kept relatively small thus making inference rather unreliable. The grid search on the other hand automatically provides me with confidence sets based on the Likelihood Ratio (LR) statistics, as the log likelihood is evaluated for all values around the peak. A single run is thus sufficient for both estimation and inference, further warranting my choice.
Grid search is furthermore attractive as the sample log likelihood function can provide us with valuable additional information, such as the existence of multiple peaks or flat segments. This, in turn allows me to detect if parameters are weakly identified.
The aforementioned arguments are all relevant for fitting the Banerjee et al. (2013) model to any data-set. For the estimation problem at hand, my procedure has two more additional advantages. First, a relatively wide grid[2] is sufficient in the sense that the value of the log likelihood changes only in digits to the right of the decimal point when the grid is further refined. Second, because the sample log likelihood function over the grid is established as the sum of the village log likelihood functions at the grid point values, it is easier to modify village level data or add villages. In analytical or numerical methods, first the log likelihood function needs to be re-established and then second, derivatives have to be taken again. For the grid search, we simply add the new village to the sample log likelihood and search for the maximal value again.

2.3.4 Confidence Sets from Likelihood Ratio (LR) Tests

Confidence sets are constructed using the LR test. Let $\hat{\theta}$ denote the vector of MLE estimates and C denote the set of grid-points that are inside the $(1-\alpha)\times 100\%$ confidence set. Because the LR test statistic is asymptotically Chi-square distributed

[2] I let all participation parameters vary by 0.1 and diffusion parameters var by 0.01.

with the degrees of freedom equal to the number of model parameters, thus the points in C are identified by the equation

$$\{\tilde{\theta} \in C : -2\big(\mathcal{L}(\tilde{\theta}) - \mathcal{L}(\hat{\theta})\big) \leq crit_{\alpha}\}$$

with $crit_{\alpha}$ the critical value from the Chi-square distribution with degrees of freedom equivalent to the dimension of θ. In words, for any gridpoint $\tilde{\theta} \in C$ we cannot reject the hypothesis that $\theta = \tilde{\theta}$.
The projection-based confidence interval for any given parameter is obtained by projecting the outer borders of the joint confidence set on the respective parameter axis, implying that

$$C_k = \big(\tilde{\theta}_{k,low}, \tilde{\theta}_{k,up}\big), \qquad \tilde{\theta}_{k,low} = min\big(\tilde{\theta}_k \in C\big), \qquad \tilde{\theta}_{k,up} = max\big(\tilde{\theta}_k \in C\big)$$

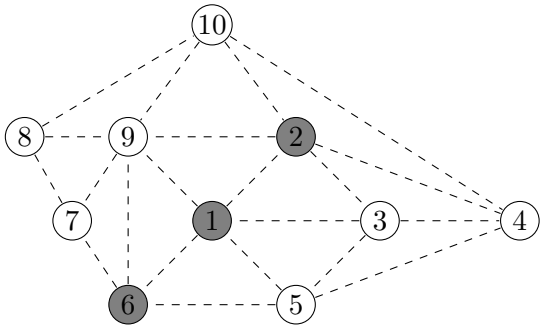
By the grid search, I automatically have the values of the log likelihood function for parameters in the neighbourhood of the peak, so I obtain the LR test confidence set as a byproduct of the procedure.
The projection-based confidence sets are conservative in the sense that their coverage rate is at least $(1-\alpha)*100\%$. This is due to the fact that testing is done jointly. First, for the univariate test, the critical value is lower. Second, if one starts at the MLE and varies only one parameter while leaving the remaining ones constant, then, as long as the co-variation between the parameters is not zero, the critical value will be hit earlier. This is intuitive: with a positive (negative) co-variation in parameters, increasing (decreasing) other parameters can cancel out the effect of increasing the parameter in question. Consequently, I can be sure to cover the confidence interval resulting from univariate testing. My procedure may, on the other hand, lead to under-rejection as the projection-based confidence sets may be larger than the ones resulting from univariate testing.
The alternative to the LR test would be to use the Wald test, which would require evaluating the inverse of the Hessian at the MLE. At best, this would be computationally more demanding without providing any advantages. At worst, the observed Hessian may be (close to being) singular in which case conducting Wald test is infeasible.

2.3.5 Shortcomings of the original estimation procedure

Banerjee et al. (2013) carefully choose and execute their estimation technique (the Simulated Method of Moments - SMM). First, adoption parameters are estimated using the set of IPs, then, these estimated parameter values, together with diffusion

and endorsement parameters drawn from a grid are used to simulate the process of adoption and diffusion, enabling them to calculate simulated moments as averages over simulation runs. Given the empirical moments observed in the data-set, diffusion and endorsement parameters are chosen to minimize the difference between simulated and observed moments.
In this section, I discuss some of the drawbacks of the SMM estimation.

Figure 2.2: **Miniature Village Network**



First, the SMM procedure takes only **average rates** into account and thus neglects the information **who** participates.
Consider the miniature village as depicted in Figure 2.2, with dark dots representing IPs. Assume we observe either agent nine or agent ten as a new participant in the second period. In both cases, village-level participation rates are the same, despite the fact that agent nine (who is connected to three IPs) is more likely to be informed than agent ten (who only has one IP neighbor). Consequently, the first case leads to an under- and the second to an overestimation of the rate of information transmission.
Second, simulating the moments (and in particular bootstrapping their standard errors) is slow and computationally costly. Computation time then restricts the number of simulation runs that can be executed.
For a point on the grid, my procedure runs considerably faster and produces confidence intervals (based on the LR statistics) as a by-product.
Third, in Banerjee et al. (2013) all adoption parameters are estimated solely from the set of IPs, a small (roughly 11 %) subset of the village population. This is in general inefficient and joint estimation should be preferred. Any deviation of the first step estimate from its expected value translates into a bias in the estimate of the diffusion parameters. Although this appears to be a relatively minor issue here, it is a potential cause of bias to be aware of.
Fourth, the estimates are sensitive to the choice of moment conditions and the choice of how to aggregate these moments into the objective function. Although this is asymp-

totically without impact, it may have a non negligible impact in finite samples.
Finally, simulation-based methods are especially advantageous if the aim is to circumvent making a distributional assumption. For the model at hand, however, both the information and the participation status are binary and thus the Bernoulli distribution is apparently plausible.

2.4 Model and Hypothesis Overview

I estimate a series of nested models that provide a wide range of specifications concerning adoption, diffusion and endorsement parameters.
Model one is the most parsimonious specification: it assumes a homogeneous diffusion rate q and a homogeneous adoption rate p. Remembering that the IPs are the initially informed individuals, then let ω_i denote the IP-degree (the number of IP friends) of individual i. For any IP j connected to i, both g_{ij} and $s_{j,0}$ equal one and thus, given that all IPs diffuse information with the same rate, $i's$ information reception rate (r_i) simplifies to $1-(1-q)^{\omega_i}$.
The assumption of homogeneous diffusion is relaxed in model two. Letting H_i denote the headcount of $i's$ friends that are participating IPs, this model presumes that the H_i IP friends of i now spread with probability q_p while the remaining $\omega_i - H_i$ IP friends of i pass on the information with probability q_N, thus individual $i's$ information reception rate is $1-(1-q_N)^{\omega_i-H_i}(1-q_P)^{H_i}$.
Models three and four include the possibility of endorsement (i.e. past participants having an influence beyond messages on their neighbours' adoption decisions). A significant positive endorsement effect implies that individuals who have many friends who already took a loan are more likely to do so themselves. Again, the diffusion rate can either be homogeneous (model three) or vary with the sender's participation status (model four). To incorporate the endorsement effect, I need to model how the number of friends that took up a loan in the past (H_i) impacts the individual's participation probability p_i. A common tool to do so is to run a logit regression using a constant (the chance to participate in absence of participating neighbours) and the headcount of participating neighbours as independent variables. Because a logit regression that only contains a constant is identical with simply directly estimating p, thus model one versus three (two versus four) can be used to test for significance of an endorsement effect by means of the LR test.
Similarly, socio-economic household characteristics can be included in the logit regression. The chosen characteristics for the data at hand are the number of rooms (a wealth

indicator) and the household religion. Let $Rooms_i$ be the number of rooms in i's house and $Hindu_i$ a dummy for i being a Hindu. With model five presuming a homogeneous diffusion rate and model six assuming that participants and non-participants have different diffusion rates, thus model one versus five (two versus six) can be used to test for (joint) significance of these household characteristics.
Table 2.1 gives an overview of the models, the respective adoption and diffusion rates, the induced information reception rates and the set of model parameters.

Table 2.1: **Model Overview**

Model 1: Homogeneous Adoption and Diffusion	$p_i = p \forall i$ $q_i = q \forall i$ $r_i = 1 - (1-q)^{\omega_i}$	$\theta = p, q$
Model 2: Homogeneous Adoption, Differentiated Diffusion	$p_i = p \ \forall i$ $q_i = q_P \ if \ Y_{i,1} = 1$ $q_i = q_N \ if \ Y_{i,1} = 0$ $r_i = 1 - (1-q_N)^{\omega_i - H_i}(1-q_P)^{H_i}$	$\theta = p, q_N, q_P$
Model 3: Endorsed Adoption, Homogeneous Diffusion	$p_i = \frac{exp(\alpha+\gamma H_i)}{1+exp(\alpha+\gamma H_i)}$ $q_i = q \ \forall i$ $r_i = 1 - (1-q)^{\omega_i}$	$\theta = \alpha, \gamma, q$
Model 4: Endorsed Adoption, Differentiated Diffusion	$p_i = \frac{exp(\alpha+\gamma H_i)}{1+exp(\alpha+\gamma H_i)}$ $q_i = q_p \ if \ Y_{i,1} = 1$ $q_i = q_N \ if \ Y_{i,1} = 0$ $r_i = 1 - (1-q_N)^{\omega_i - H_i}(1-q_P)^{H_i}$	$\theta = \alpha, \gamma, q_N, q_P$
Model 5: Socioeconomically Determined Adoption, Homogeneous Diffusion	$p_i = \frac{exp(\alpha+\beta_1 Rooms_i+\beta_2 Hindu_i)}{1+exp(\alpha+\beta_1 Rooms_i+\beta_2 Hindu_i)}$ $q_i = q \ \forall i$ $r_i = 1 - (1-q)^{\omega_i}$	$\theta = \alpha, \beta_1, \beta_2, q$
Model 6: Socioeconomically Determined Adoption, Differentiated Diffusion	$p_i = \frac{exp(\alpha+\beta_1 Rooms_i+\beta_2 Hindu_i)}{1+exp(\alpha+\beta_1 Rooms_i+\beta_2 Hindu_i)}$ $q_i = q_p \ if \ Y_{i,1} = 1$ $q_i = q_N \ if \ Y_{i,1} = 0$ $r_i = 1 - (1-q_N)^{\omega_i - H_i}(1-q_P)^{H_i}$	$\theta = \alpha, \beta_1, \beta_2, q_N, q_P$

As the models are nested, the LR statistic can be used to test three hypotheses.

- **Test 1: Homogeneous versus differentiated diffusion**
  $H_0 : q_N = q_P$
  "Participants and non-participants spread the information with the same rate."

  - **Test 1.1.:** The test is performed assuming a homogeneous adoption rate, i.e. using model one (restricted model) versus model two (unrestricted model)
  - **Test 1.2.:** The test is performed in a model with endorsed adoption i.e. using model three (restricted model) versus model four (unrestricted model)
  - **Test 1.3.:** The test is performed in a model with socioeconomically determined adoption, i.e. using model five (restricted model) versus model six (unrestricted model)

- **Test 2: Homogeneous versus endorsed adoption**
  $H_0 : \gamma = 0$
  "The number of her friends that previously participated has no effect on an individual's participation probability."

  - **Test 2.1.:** The test is performed assuming homogeneous diffusion rates, i.e. using model one (restricted model) versus model three (unrestricted model)
  - **Test 2.2.:** The test is performed in a model that allows for differentiated diffusion, i.e. using model two (restricted model) versus model four (unrestricted model).

- **Test 3: Homogeneous versus socio-economically determined adoption**
  $H_0 : \beta_1 = \beta_2 = 0$
  "The number of rooms and the household religion have no significant impact on the household's adoption rate."

  - **Test 3.1.:** The test is performed assuming homogeneous diffusion, i.e. using model one (restricted model) versus model five (unrestricted model).
  - **Test 3.2.:** The test is performed in a model that allows for differentiated diffusion, i.e. using model two (restricted model) versus model six (unrestricted model)

The decision rule is given by

$$\text{reject } H_0 \text{ if } -2\big(\mathcal{L}(\theta_{H0}) - \mathcal{L}(\theta_{H1})\big) > crit_{95}$$

with $crit_{95}\%$ being the critical value from the chi-square distribution with one (test one and two) or two (test three) degrees of freedom. $\mathcal{L}(\theta_{H0})$ is the value of the log likelihood function evaluated at the restricted model (the specific restriction being imposed by the hypothesis) and $\mathcal{L}(\theta_{H1})$ is the value of the log likelihood function evaluated at the unrestricted model. Hypothesis one and two imply that one model parameter is restricted and consequently, the unrestricted model contains the same parameters as the restricted one plus one additional parameter, which is the one to be tested. Hypothesis three restricts two parameters and thus the restricted and unrestricted model share all but two parameters.

2.5 Data

The data stems from a large-scale study on micro-finance adoption conducted in rural Indian villages. To accumulate this data, researchers from Stanford and MIT cooperated with a non-governmental organisation (NGO) that started providing group loans to villagers that previously had no exposure to micro-finance. A subset of villagers (the IPs) were initially informed and thereafter, the organization relied on word-of-mouth marketing. Two studies (Banerjee et al. (2013) and Jackson et al. (2012)) were published and the two data-sets were made available. Since each data-set only contains the data insofar as it is used for the publication of the respective paper, for my empirical analysis, the two data-sets were merged. The time series of (village-level) aggregate participation rates at different points in time were taken from Banerjee et al. (2013). This data-set also contains the household identification (id) numbers of the initially informed individuals (IPs) and listed the individuals that (at some point in time) took up a micro-loan. The data from Jackson et al. (2012) provided the information on the ordering (succession) in which the participating households signed up for micro-finance group loans. I thus match the information on the aggregate timing from Banerjee et al. (2013) with the information on the individual ordering from Jackson et al. (2012). In absence of knowledge of the time that passes between the intake of the groups from Jackson et al. (2012), I stick to the timing as provided by the NGO: upon inclusion of each cohort, the village's overall participation rate was calculated and compared to the aggregate rate that was calculated by the organization in order to determine how many cohorts had been formed at each point in time.
The result of the merge is a data-set that contains the time series of household par-

ticipation decisions over a village specific time horizon (T_v) [3] : For each village, we model the time horizon from period $t = 0$ (the initially informed individuals receive the information) to period $t = T_v$ (no further change in overall participation rates is observed and all participating cohorts are included). For all villages except 31, overall participation rates increase to the final level within a few time periods and as a consequence, $T_v \leq 4 \ \forall v$.[4] Computation becomes very challenging when more than two time periods are considered and in this case, establishing the exact log likelihood function is computationally infeasible such that one needs to establish and maximize an approximate log likelihood function. This will be done in chapter three. For the current chapter, I conduct an exact maximum likelihood estimation and consequently limit my attention to the first two time periods.

While the choice to restrict the time horizon to two periods was mainly driven by computational aspects, in practice, there are also various advantages of an early analysis. First, detecting project shortcomings early implies a timely adjustment to potential shortcomings. Second, for micro-finance projects, early take-up is particularly crucial: late participation implies that financial resources are unused and staff members are disengaged, early lending on the other hand frees up funds again faster and enables staff to accumulate important experiences. Finally, enlarging the time horizon also entails a larger possibility of information spread outside of the network, making it harder to quantify the pure network effects.

While the data does provide me with the identity of the initially informed individuals, the time series of participation decisions and the overall participation rates that is successively measured by the micro-loan provider, it is not unequivocal *how much* time has actually passed between two measurements and how many rounds of information exchange have taken place. I assume that the micro-loan provider did measure the participation at each point in time, i.e. between any two measurements, there is always one round of information exchange and participation decisions. This assumption is strong, but necessary for the purpose of my estimation strategy. Since my time horizon is only two periods, I thus restrict my attention to the first and the second measurement.

Knowing that the average participation rates only increase from the first to the second (all villages), the second to the third (the majority of villages) and at most the third

[3] Whereas for some villages successively entering the micro-finance cohorts enumerated in Jackson et al. (2012) produced the identical (villages 2, 4, 23, 31, 51, 73) or nearly identical (villages 1, 3, 12, 20, 24, 25, 29, 32, 36, 42, 45, 47, 50, 52, 55, 57, 59, 65, 67, 68, 72) participation rates as the overall rates observed in Banerjee et al. (2013), for other villages (villages 19, 21, 33, 39, 43, 46, 62, 70, 71, 75) the final rates of the Jackson et al. (2012) cohort data are more than 5% below the village-level rates in Banerjee et al. (2013).

[4] Including more time periods would only be beneficial for village 31 but substantially increase the computational burden

to the fourth (roughly one third of the villages) measurement, but thereafter stay constant, it would thus, according to my assumptions, be sufficient to model the first four time periods.

Banerjee et al. (2013) only make use of the last (i.e. tenth) measurement provided by the organization. Arguing that the villages have been offered loans over five to eight quarters at the time the data accumulation was finished, Banerjee et al. (2013) presume that there are five to eight information exchanges. Consequently, they presume five to eight rounds of information exchanges also for those villages for which the aggregate participation rate (as measured by the organization) already reaches its final level at the second measurement. I model the first two time periods and postpone the modeling of a larger time horizon to subsequent chapters.

Besides the participation decisions and the information initiation, the data-set contains different measures of the village network and household characteristics.

The following are the questions by which the network data is elicited:

Table 2.2: **Network Measurement Overview**

Question no. (identifier)	**Survey Question**
Q1 (borrowmoney)	"If you suddenly needed Rs. 50 for a day, who would you ask?"
Q2 (giveadvice)	"Who comes to you for advice?"
Q3 (helpdecision)	"If you had to make a difficult personal decision, who would you ask for advice?"
Q4 (keroricecome)	"Who would come to you if he/she needed to borrow kerosene or rice?"
Q5 (keroricego)	"If you needed to borrow kerosene or rice, to whom would you go?"
Q6 (lendmoney)	"Who do you trust enough so that if he/she needed Rs 50 for a day, you would lend it to him/her?"
Q7 (medic)	"If you had a medical emergency and were alone at home, who would you ask for help in getting to a hospital?"
Q8 (nonrel)	"Name four non-relatives whom you speak to the most"
Q9 (rel)	"Name four close relatives whom you speak to the most"
Q10 (templecompany)	"Do you visit temple/mosque/church? Do you go with anyone else?"
Q11 (visitcome)	"Who visits your house in his or her free time?"
Q12 (visitgo)	"In your free time, whose house do you visit?"

Several features of the observed data are at odds with the Banerjee et al. (2013) modeling assumptions:

1. I observe that individuals other than the IPs participate in period $t = 1$. This violates assumption four, since non-IP individuals can not possibly be informed in period one.
   For non-IP individuals that are recorded to participate in period 1, participation is delayed.[5]

2. I observe that IPs participate at later points in time. This violates assumption five since IPs have their chance to choose only in period one. I thus presume that all IPs that participate do so in the first period.

3. I observe that some individuals participate despite being too far away from the information to be informed in the respective time period, which again violates assumption four.
   If individuals are recorded as participants despite being too far away from the information to be reached by it, their participation is delayed to the time period following the first point in time at which they can be reached.

4. Not all participants of the Banerjee et al. (2013) data-set are also listed as participants in the Jackson et al. (2012) data-set.
   In case of doubt, I assume that the data-set Jackson et al. (2012) is relevant for enumerating participants (since for those appearing only in Banerjee et al. (2013) I have no information on the timing.)[6]

5. Among the participants are isolated agents which again represents a violation of assumption four. Isolated agents are excluded from the analysis.

In the data-sets at hand, the network is uncovered by means of a survey in which a subset of villagers is asked to enumerate others with whom she engages in regular activities. With twelve activities in total, the authors take *any* nomination as evidence for the existence of a link and further assume that nominations do not need to be reciprocated. By the same token, having a network survey that only includes a subset of the villagers is viewed to be un-problematic by the authors of Banerjee et al. (2013) and Jackson et al. (2012).
Due to the multitude of possible realizations of S_1 that are in accordance with the data,

[5] I then assume that these individuals participate in the time period that directly follows the first information exchange in which they can be informed

[6] Another possibility (for the future) would be to append them as a last cohort and assume they enter after everybody else

the authors refrain from seeking to maximize (or even to establish) the log likelihood function. Instead, they opt for a two-step procedure in which first adoption parameters are estimated solely from the set of the IPs and diffusion parameters are recovered by SMM: the estimates of the adoption parameters from the first step are, together with value of the diffusion and endorsement parameters drawn from the grid, used to simulate the information diffusion 75 times per village. As explained above, the village aggregate participation rates only change two to four times to reach their final level and stay constant thereafter. The authors nonetheless assume that there are five to eight periods of information exchanges and participation decisions. Since only the final participation rates and the final empirical rates are compared, the authors thus ignore that these final rates are reached in just two to four steps. This is equivalent to assuming that interim, unobserved participation rounds occur. By basically mimicking the information exchange, they thus circumvent the problem of the information status being unobserved. Averaging over simulation runs lead to village-level simulated final moments, averaging over villages to the corresponding simulated sample moments. The criterion function to be minimized is then simply the inner product of the sample moment vector. The algorithm is outlined in Appendix 1.

2.6 Models with Homogeneous Adoption Rates

If participation probabilities are constant, the only way in which agents differ from one another is by their link portfolio. Therefore (2.3) simplifies substantially.

$$\ell_2 = \underbrace{\prod_{i=1}^{N_{IPP}} p_i}_{\substack{\textit{participating} \\ \textit{IPs}}} \underbrace{\prod_{j=1}^{N_{IPN}} (1-p_j)}_{\substack{\textit{non-} \\ \textit{participating} \\ \textit{IPs}}} \underbrace{\prod_{k=1}^{N_{P2}} r_k p_k}_{\substack{\textit{period 2} \\ \textit{participants}}} \underbrace{\prod_{l=1}^{N_{PII}} \Big(r_l(1-p_l) + (1-r_l) \Big)}_{\textit{PIIs}}$$

$$= p^{N_{IPP}} (1-p)^{N_{IPN}} \prod_{k=1}^{N_{P2}} r_k p \prod_{l=1}^{N_{PII}} \Big(r_l(1-p) + (1-r_l) \Big)$$

$$= p^{N_{IPP}} (1-p)^{N_{IPN}} \prod_{k=1}^{N_{P2}} r_k p \prod_{l=1}^{N_{PII}} (1 - r_l p)$$

The only individual-specific terms in this expression are the information reception rates. Knowing this will reduce the computational burden substantially.

2.6.1 Model 1: Homogeneous Adoption and Diffusion

Knowing from section four that $r_i = 1 - (1-q)^{\omega_i}$, agents i and j have the same probability to be informed if they have the same number of IP-friends ($\omega_i = \omega_j$). Agents can be segregated into groups according to their number of IP-friends and their participation status. The group's likelihood contribution is then simply the contribution of one of its members, raised to the power of the group size. If there are $m = 1, ..., M$ different realizations of ω in the sample then agents can be sorted into groups such that $\omega_i = \omega_m \ \forall i \in m$. Within each group, the agents can be split into participants and PIIs. Let N_{mp} (N_{mn}) denote the number of participants (PIIs) in group m, then the likelihood function is

$$\ell_2 = p^{N_{IPP}}(1-p)^{N_{IPN}} \prod_{m=1}^{M} (r_m p)^{N_{mp}} (1 - r_m p)^{N_{mn}}$$

With

$$r_m = 1 - (1-q)^{\omega_m}$$

This means that (2.3) is tractable even if N grows very large, as long as the maximal number of IP-friends that agents can have is bounded. These considerations enable me to establish the village log likelihood function with a few computations, even when the network is large and dense.

Figure 2.3 shows the estimated adoption and diffusion rates for the different networks defined using the questions enumerated above. Clearly, the estimated participation rate (p) is stable across all networks, while the estimated diffusion rate (q) varies quite a bit. This is not surprising, as the link portfolios of both PIIs and participants are determining q and this is precisely what varies from one network measure to another. Note that the sparser networks may imply that participants lack IP connections and thus have to be excluded from the estimation, the result being that different questions generate different final adoption rates. The occurrence of these "IP-hermits" can be interpreted as an indication that there are even many more links through which the information could spread that the survey fails to capture. Figures 2.4 and 2.5 show the 95% (blue) and 99% (green) confidence sets for the union of all questions and for each question separately. As expected, the confidence sets indicate that the two parameter estimates are negatively correlated. The confidence sets of q tends to range over a quite substantial part (roughly about slightly less than one third) of the parameter space. As shown in the Appendix, the variance of q is large if the second derivative of the log likelihood function with respect to q is small in absolute value. For p on

the other hand, a small second derivative with respect to q implies a small variance. The intuition is that in dense networks, there are more possibilities for information propagation implying less precise estimation of q.

The confidence sets are largest for question 10, which is the question that implies the sparsest network. The confidence set being so large in this case is mainly a small sample effect: as more individuals are not linked to IPs at all, thus fewer individuals (participants or PIIs) can be used for the estimation. A smaller sample size then naturally translates into a larger variance for both p and q.

Figure 2.3: **Estimates for Model 1: Homogeneous Adoption and Diffusion**

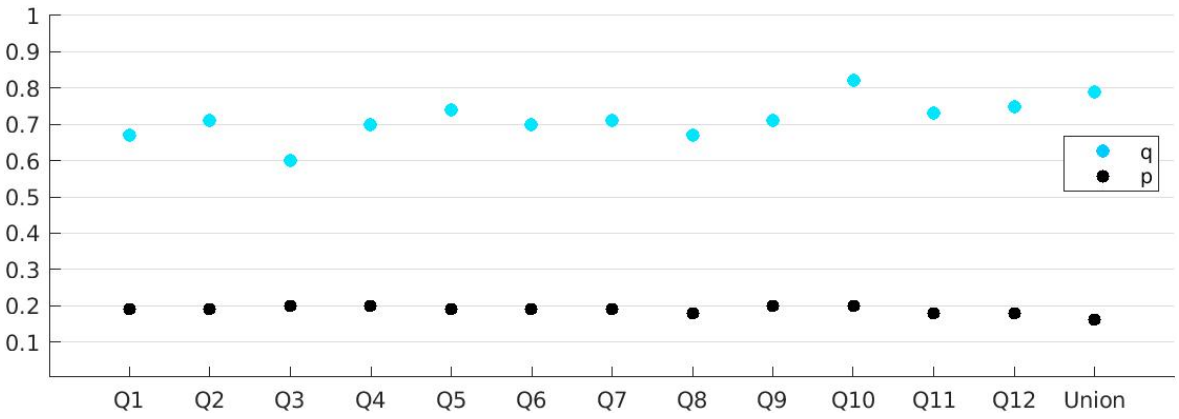


Figure 2.4: **Sample Log-Likelihood Surface and Parameter Confidence Sets for Model 1: Homogeneous Adoption and Diffusion**
(Blue: 95% Confidence Set, Green: 99% Confidence Set Based on the LR Test) for the Union of all Network Measures

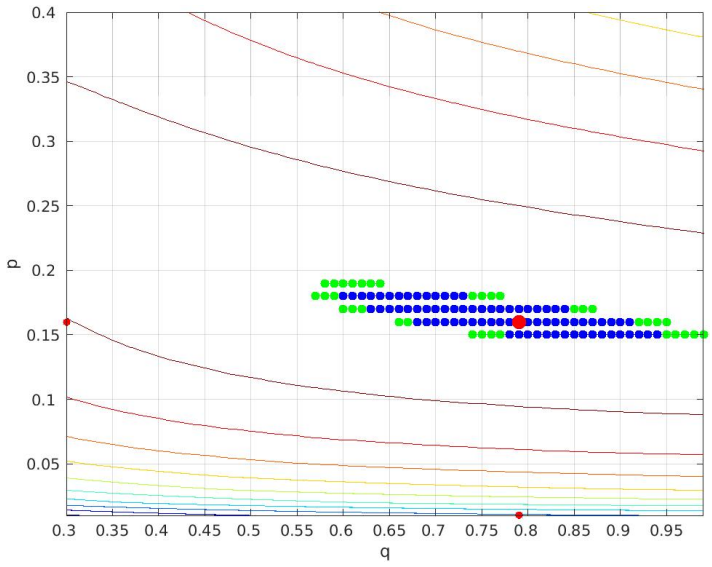
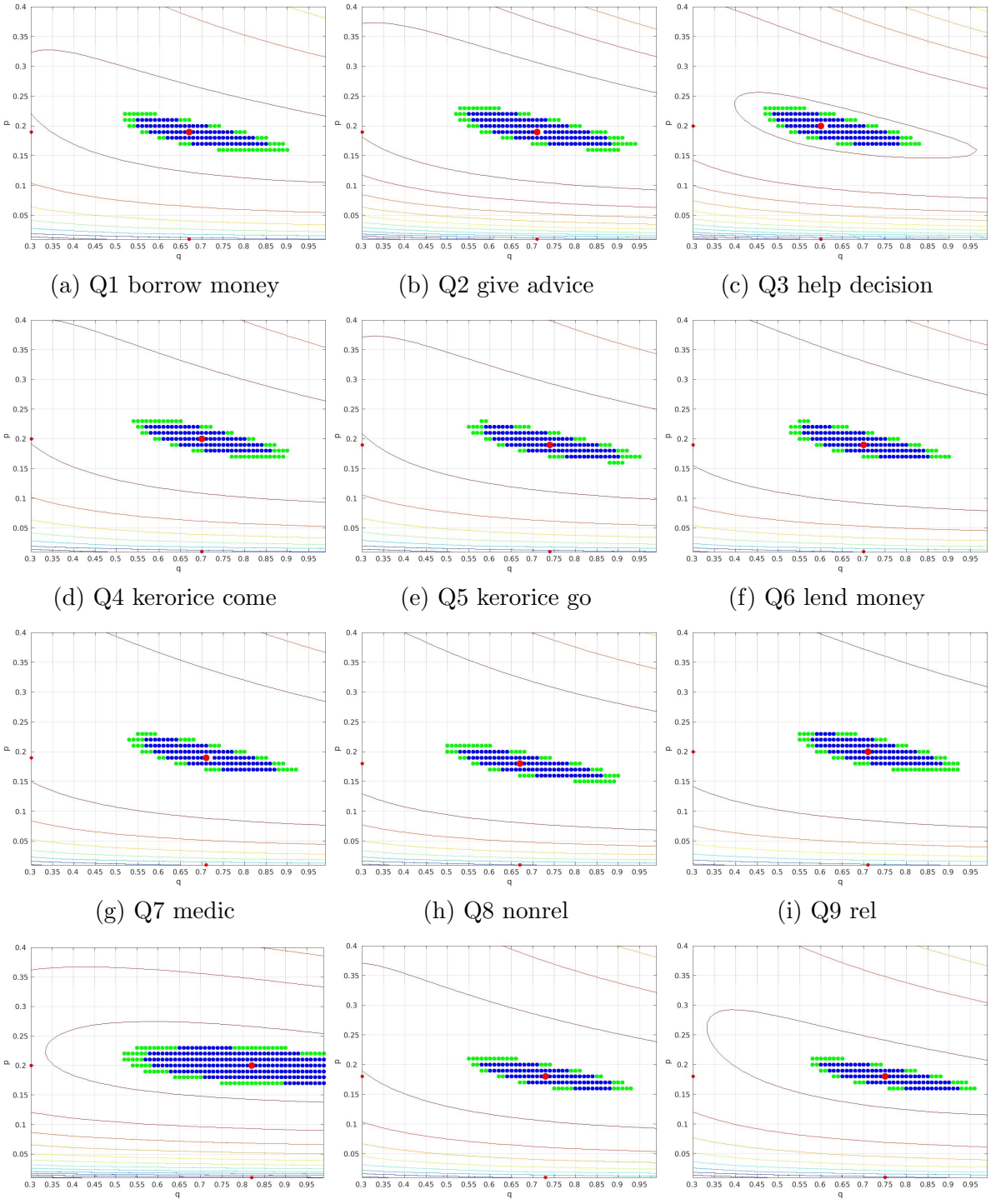


Figure 2.5: **Sample Log-Likelihood Surfaces and Parameter Confidence Sets for Model 1: Homogeneous Adoption and Diffusion**
(Blue: 95% Confidence Set, Green: 99% Confidence Set Based on the LR Test) for the Twelve Network Measures (Questions)



(a) Q1 borrow money (b) Q2 give advice (c) Q3 help decision

(d) Q4 kerorice come (e) Q5 kerorice go (f) Q6 lend money

(g) Q7 medic (h) Q8 nonrel (i) Q9 rel

(j) Q10 temple company (k) Q11 visit come (l) Q12 visit go

2.6.2 Model 2: Homogeneous Adoption, Differentiate Diffusion

Model two takes into account that participating IPs may spread the information with a different rate than non-participating IPs. The two probabilities are denoted q_N and q_P, respectively. This model still assumes that, once informed, all individuals have the same probability to participate.

Because of this differentiated diffusion, grouping agents needs to be done according to both ω and H: only those agents that have the same number of IP friends in total and the same number of participating IP friends do have the same information reception rates. Thus there will be more groups and less agents within each group. The likelihood is then

$$\ell_2 = p^{N_{IPP}}(1-p)^{N_{IPN}} \prod_{m=1}^{M} (r_m p)^{n_{mp}} (1 - r_m p)^{n_{mn}},$$

$$r_m = 1 - (1 - q_N)^{\omega_m - H_m}(1 - q_P)^{H_m}.$$

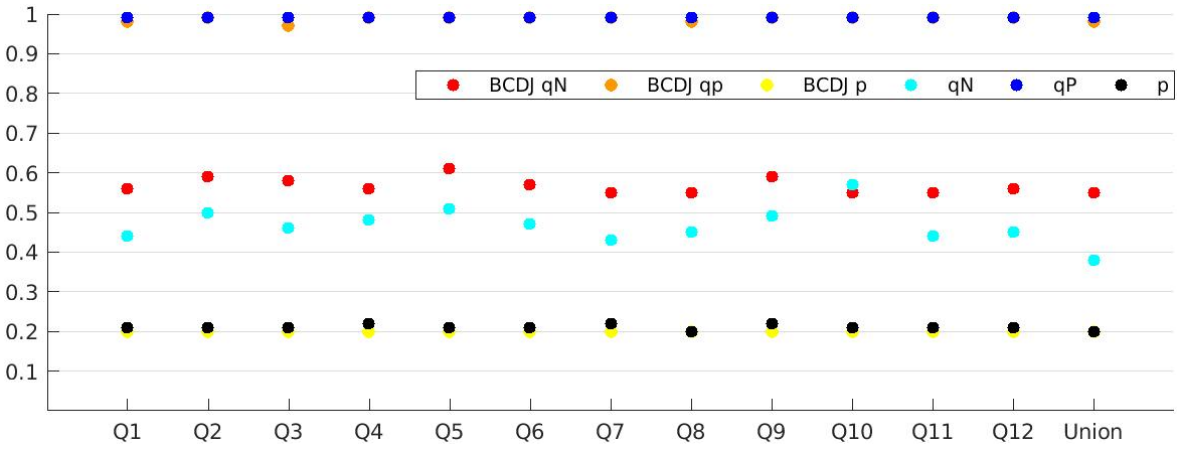
Figure 2.6 depicts the estimates of the model that now incorporates three parameters. Additionally, the estimates that result from applying the Banerjee et al. (2013) algorithm to the shortened time horizon are shown. The algorithm is outlined in the Appendix.

The most important observation that can be made is that the more dense the network, the further apart the estimates that result from the two procedures. The plot of the diffusion rate from model two (Figure 2.6) shows that the diffusion dynamics are largely determined by non-participants, while participants are practically sure to pass the information. With negligible variation in p, q_N together with the distribution of the IP degrees in the first wave are the driving force for changes in the final adoption rates. Given the large discrepancy between q_N and q_P, increasing early participation is essential for word-of-mouth marketing to be efficient: with non-participants being much less likely to pass on the information, the set of potential future borrowers (i.e. those individuals that do face a choice in the second period) increases substantially when there are more participating IPs.

Applying the Banerjee et al. (2013) algorithm often leads to a slight underestimation in p (which is set to the IP-participation rate of 0.2) and sometimes also to a slight underestimation of q_P (Q1, Q3, Q8, Union). More importantly, however, q_N is substantially over-estimated. This is even more the case for the denser networks (observe that the largest discrepancy between the two estimates of q_N occurs for the Union which, per definition, produces the most dense network). Strikingly, this over-estimation of q_N

does not seem to be driven by an erroneous first-step estimate of p (observe that for the Union network, p coincides with the IP-participation rate). In the sense that a very low information diffusion rate for non-participants implies a marketing problem, incorrectly estimating this parameter can have severe consequences for practitioners. Since there are three model parameters, features of the confidence sets are best illustrated by holding one parameter constant and scattering the point in the confidence sets for various values of the remaining two parameters. Figures 2.7, 2.8, and 2.9 illustrate the 95 % confidence sets varying q_N and q_P while holding p constant. These are cuts through the cloud of points in the confidence set. From these figures, we can see a clear negative correlation between q_N and p: when p goes up, the confidence set shifts to the left. We can also see that a particularly large or small p implies little variation in q_P: the first and last confidence sets extend over a substantially smaller range of the q_P-axis. Finally, higher q_P-values generally also feature more variation in q_N. Parameter estimation is the least precise for intermediate p values.

Figure 2.6: **Estimates for Model 2: Homogeneous Adoption, Differentiated Diffusion**



BCDJ q_N, BCDJ q_P and BCDJ p are the estimated diffusion rates for non-participants, the estimated diffusion rate for participants and the adoption rate respectively that result from applying the Banerjee et al. (2013) algorithm to the two-period time horizon.
q_N, q_P nd p are the ML estimates of the diffusion rates for non-participants, the diffusion rate for participants and the adoption rate respectively.

Figure 2.7: **LR 95 % Confidence Set for Model 2: Homogeneous Adoption, Differentiate Diffusion for the Union of all Network Measures, the Estimate of Banerjee et al. (2013) in Red**
The confidence sets are three-dimensional and thus plotted holding one parameter (p) constant while varying the remaining two parameters

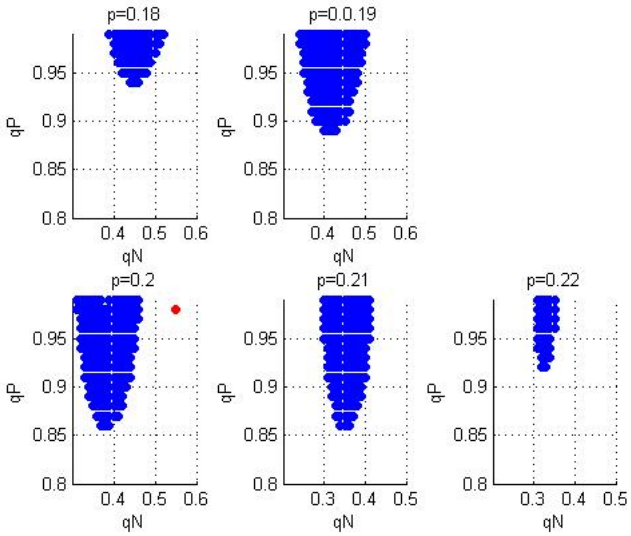
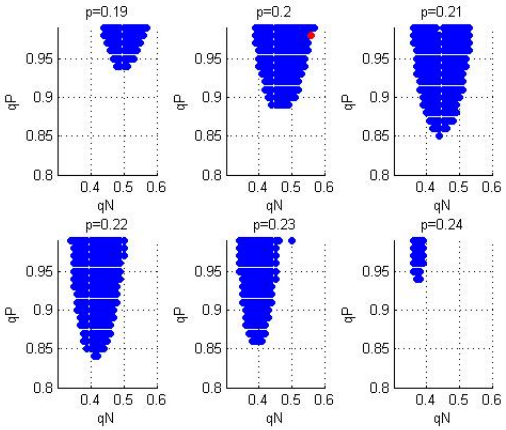
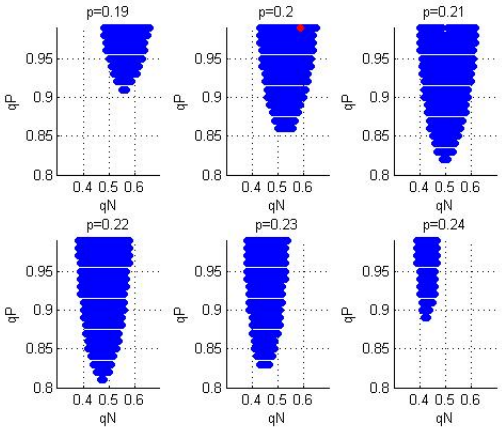


Figure 2.8: **LR 95 % Confidence Sets for Model 2: Homogeneous Adoption, Differentiate Diffusion for Different Network Measures, the Estimate of Banerjee et al. (2013) in Red**
The confidence sets are three-dimensional and thus plotted holding one parameter (p) constant while varying the remaining two parameters

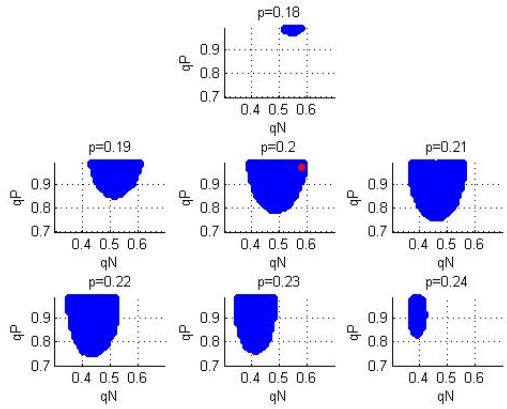
(a) Q1 borrow money



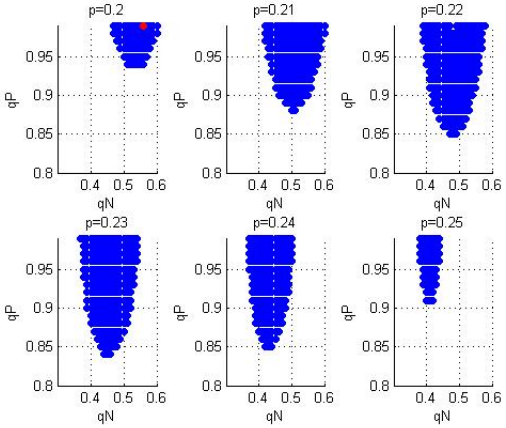
(b) Q2 give advice



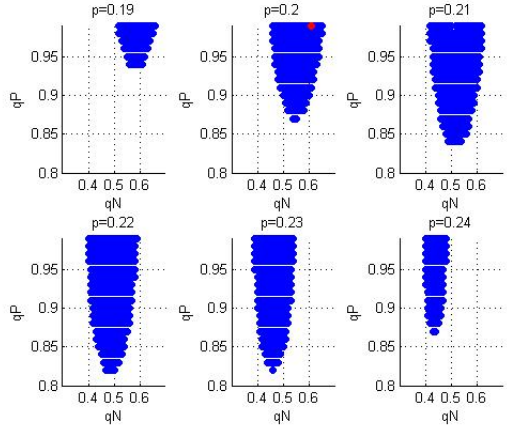
(c) Q3 help decision



(d) Q4 kerorice come



(e) Q5 kerorice go



(f) Q6 lend money

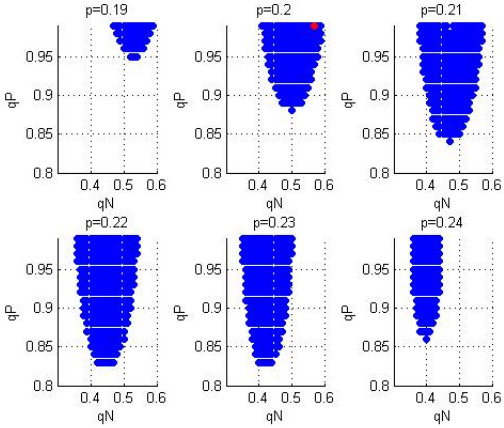
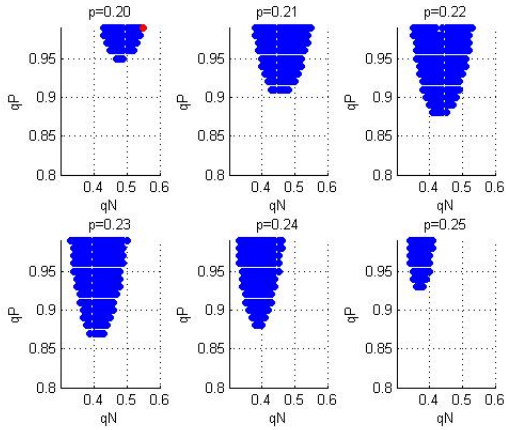
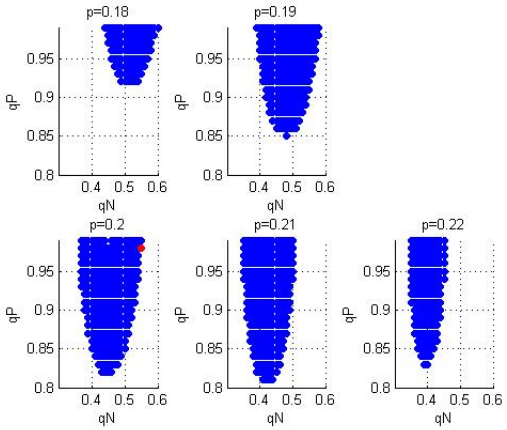


Figure 2.9: **LR 95 % Confidence Sets for Model 2: Homogeneous Adoption, Differentiate Diffusion for Different Network Measures, the Estimate of Banerjee et al. (2013) in Red**
The confidence sets are three-dimensional and thus plotted holding one parameter (p) constant while varying the remaining two parameters

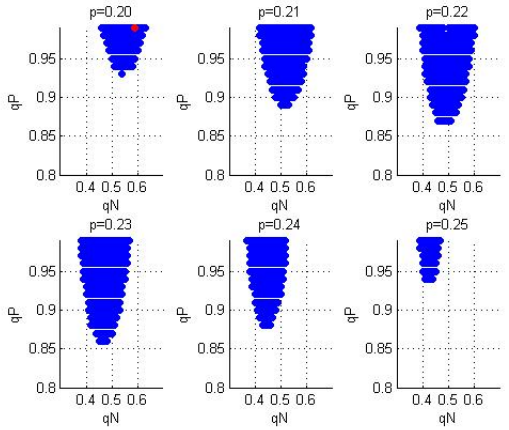
(a) Q7 medic



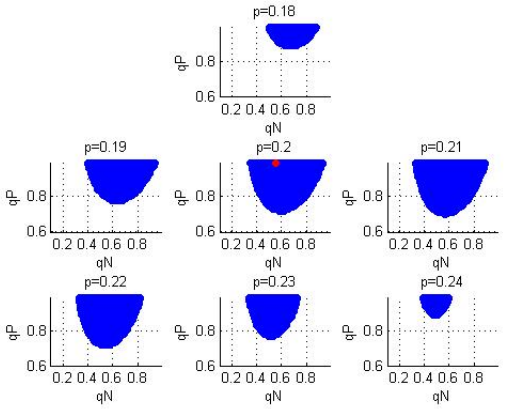
(b) Q8 nonrel



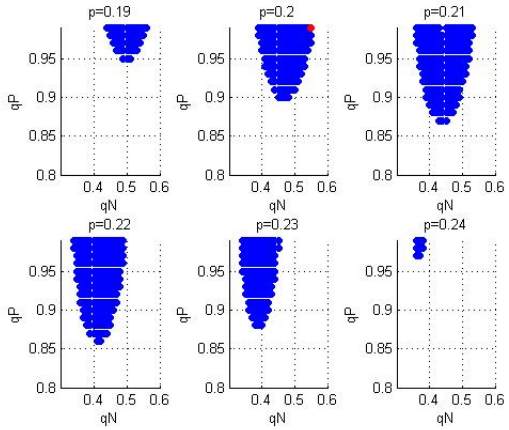
(c) Q9 rel



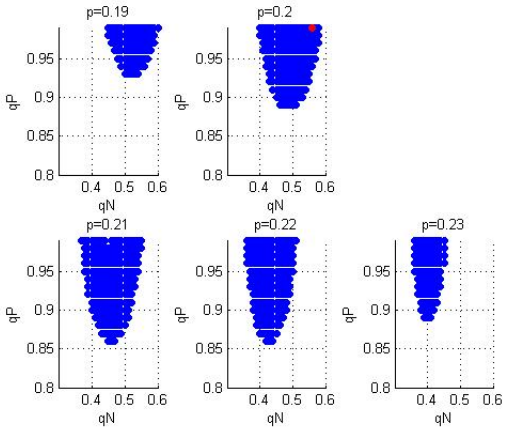
(d) Q10 temple company



(e) Q11 visit come



(f) Q12 visit go



2.6.3 Hypothesis Testing

Model one and two differ from one another concerning the diffusion rate, which is assumed to be independent of the sender's participation decision in model one. To evaluate whether model two with its additional parameter truly fits the data better, I conduct the statistical hypothesis test 1.1. the hypothesis being that differentiated diffusion does not lead to a significant improvement in model fit. I employ the LR test for model one, in which I apply the restriction that $q_N = q_P$ against model two, the unrestricted model. Not only can I reject the hypothesis of a homogeneous diffusion rate for all questions at the 95% significance level, but moreover, the difference in the log-likelihood function is very large making rejection unequivocal.

Table 2.3: **Test 1.1.: Homogeneous Versus Differentiated Diffusion Given Homogeneous Adoption Using Model One (Restricted Model) Versus Model Two (Unrestricted Model)**

	M1 $H_0 : q_N = q_P$ $\mathcal{L}(\theta_0)$	M2 $H_1 : q_N \neq q_P$ $\mathcal{L}(\theta_1)$	LR statistic	$\chi^2(1)$ crit. value 3.841
Q1	-1548.99	-1522.52	52.94	reject
Q2	-1225.49	-1209.22	32.54	reject
Q3	-1289.43	-1274.31	30.24	reject
Q4	-1502.58	-1479.28	46.6	reject
Q5	-1521.97	-1504.6	34.74	reject
Q6	-1470.82	-1448.52	44.6	reject
Q7	-1458.61	-1428.92	59.38	reject
Q8	-1638.2	-1616.94	42.52	reject
Q9	-1251.51	-1229.83	43.36	reject
Q10	-635.011	-631.037	7.948	reject
Q11	-1712.34	-1684.76	55.16	reject
Q12	-1761.88	-1738.6	46.56	reject
Qall	-2288.16	-2257.2	61.92	reject

2.7 Models with Endorsement Effect

The baseline assumptions imply that participation is purely an individual decision and that friends matter only in so far as that they may pass on the news about the opportunity. One can, however, also think that individuals are influenced by their

friends' actual choices. To test for this "endorsement effect", the number of friends that already took up a loan in the last period must be included when modelling the participation probability (p_i), which now is individual specific. As introduced in section four, this number is denoted H_i.

Note that I distinguish between this lagged endorsement effect from contemporaneous coordination. In the former case, only friends' *past* decisions matter for each individuals current choice. In the latter case, individuals that are entitled to choose coordinate with one another, which creates a system of simultaneous equations challenging to estimate. In this chapter, I limit myself to the lagged endorsement effect. This case is un-problematic. Due to our baseline assumptions, any individual will face the choice only once and therefore, H_i is predetermined with respect to $Y_{i,2}$.

In model three and four, H_i, defined as the number of neighbours of individual i who took up a loan in the past, is included as a regressor in the logit model determining individual participation, i.e.

$$p_i = P(Y_{i,2} = 1 | S_{i,1} = 1, S_{i,0} = 0) = \frac{exp(\alpha + \gamma H_i)}{1 + exp(\alpha + \gamma H_i).} \tag{2.7}$$

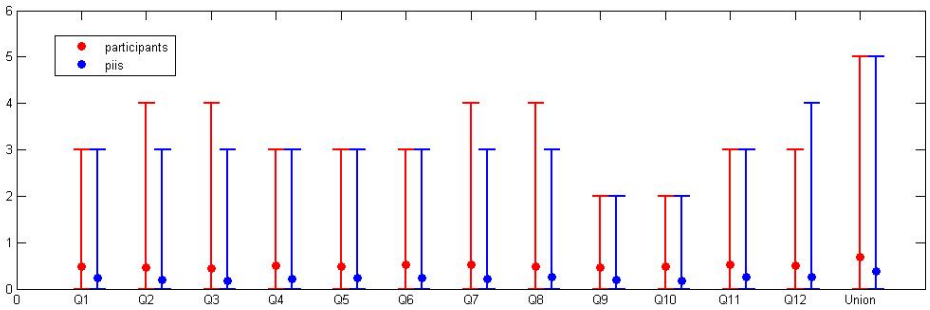
Here, α represents the participation probability for an individual who is not linked to any participating IP, while γ indicates the extent to which being friends with an additional IP-participant alters the adoption choice, ceteris paribus.

Figure 2.10 illustrates the distribution of H_i for period-two participants and PIIs. The dots correspond to the average, the bars indicate the maximum number of participating friends observed. We can discern that indeed for all questions, PIIs have on average less participating neighbors than participants. This effect is not caused by large outliers in the distribution of H_i over participants: indeed, for most questions, the minimum and maximum of H_i observed in the sample are equivalent for participants and PIIs. This permits me to conclude that not just some, but most participants have more participating IP-friends than most PIIs. This thus hints towards the presence of a positive endorsement effect.

Before statistically estimating and testing this effect, I briefly discuss the channels by which it may arise.

First, a positive endorsement coefficient can result from the fact that participants transmit a positive opinion about the product to their peers and advise them to take a loan. Even in absence of a recommendation from a participating friend, information recipients, disposing over little financial education, may deduct information on the quality of the product from peers choices and mimic them. They may also face lower than average cost of participation for instance because they can join their friends to visit the

Figure 2.10: **The Distribution of Participating IP-friends for non-IP Participants and PIIs**



Number of friends that took up a loan in the last period for participants (red) and PIIs (blue). The bars stretch from the minimum to the maximum with dots representing the average

branch or they can be visited by loan officers that frequent their friends' businesses. Furthermore, knowing that their friends started or enlarged a business by means of a micro-loan, they may seize new opportunities for their own business by increased cooperation and trade. Finally, newly informed individuals may feel peer pressure to "do what everybody around does" in order to belong to the "insider group". Also, they may wish to signal, by means of taking up a loan, that they are actually able to enlarge their business, which increases their reputation.

But some of the aforementioned channels may also translate into a negative endorsement effect: the opinion transmitted may be negative, for instance if previous participants are disappointed. Additionally, individuals may refrain from taking a micro-loan since they anticipate that their friends, who took a loan, have more financial resources and may informally lend some to them.

The endorsement effect for which Banerjee et al. (2013) test is conceptually slightly different from the one I consider. Since the information diffusion is simulated, it is possible for them to count the number of participants that *did actually* transmit the information. For each link, information transmission is randomly determined thus all of the indicator variables I_{ij} are thereafter treated as observed. Consequently, for each simulation run, a certain fraction out of an individual's participating friends has transmitted the information. Then this number is multiplied by γ drawn from the grid and thereafter, together with α estimated from the IPs, included in the logit equation predicting the individual's participation probability. This probability then gives the (individual and simulation run specific) treshold for randomly determining the individual's participation decision. Once all individual participation decisions are simulated, this leads to one realization of the sample moment per simulation run from which the

average over simulation runs (the simulated moment) can be calculated. The value of γ, q_N and q_P that minimize the distance between empirical and simulated moments is chosen. Doing so, they do not find a significant endorsement effect.
Unable to know who has sent the information and who has not, I simply include the *total* number of participating IP-friends, which includes those who sent the message *and* those who did not.

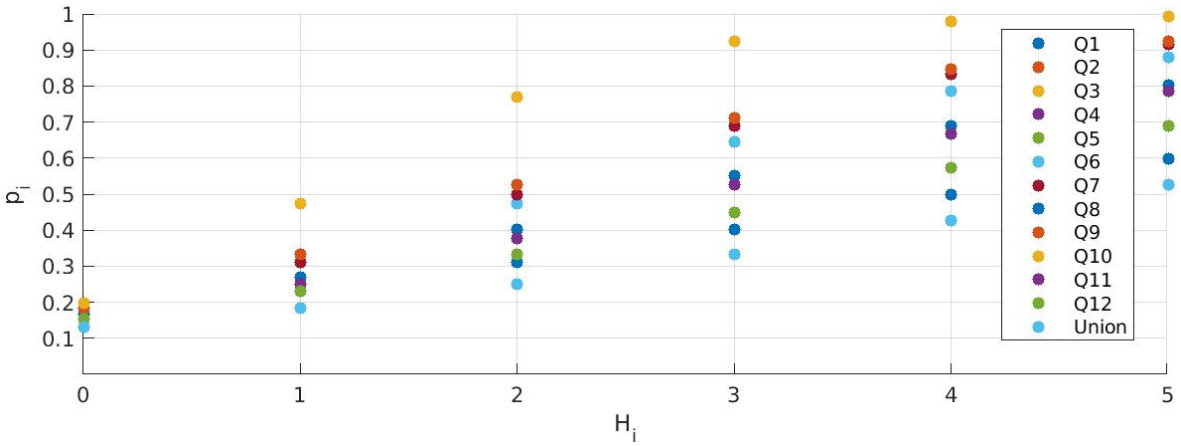
2.7.1 Model 3: Endorsed Adoption, Homogeneous Diffusion

Table 2.4 tabulates the estimation results for the endorsed adoption, homogeneous diffusion model. The confidence intervals are calculated as described above. First, I identify any point on the parameter grid for which, on the basis of the LR test, the hypothesis that the estimates equal this point could not be rejected. These points define a set and, for each parameter, I find the minimal and maximal value within this set, which gives me the univariate projection-based confidence intervals.
The endorsement effect is positive, substantial in magnitude and significant. Individuals lacking IP friends have a much smaller probability to take a loan upon being informed. Besides, I see a decrease in q for the sparse and increase in q for the dense networks. Figure 2.11 illustrates the predicted participation probabilities as a function of the number of friends that participated previously. These are calculated using equation (2.7), by substituting in the MLE for α and γ. For all questions, p_i increases linearly in H_i to finally reach a level that for most questions is close to one for those individuals linked to five participating IPs. This implies thus that these individuals almost surely participate. In practical terms, this implies that the micro-finance lender should heavily invest in convincing IPs to participate.

Table 2.4: **Estimates and Confidence Intervals for Model 3: Endorsed Adoption, Homogeneous Diffusion**

	α			γ			q		
question	$\hat{\alpha}$	95% confidence set		$\hat{\gamma}$	95% confidence set		$\hat{q}$	95% confidence set	
Q1	-1.6	-1.8	-1.4	0.6	0.4	0.8	0.67	0.48	0.89
Q2	-1.5	-1.7	-1.3	0.7	0.4	1.1	0.64	0.46	0.84
Q3	-1.5	-1.7	-1.3	0.8	0.5	1.2	0.56	0.4	0.73
Q4	-1.5	-1.7	-1.3	0.8	0.5	1	0.63	0.48	0.84
Q5	-1.5	-1.7	-1.3	0.7	0.4	0.9	0.64	0.5	0.91
Q6	-1.5	-1.7	-1.4	0.7	0.5	1	0.61	0.47	0.85
Q7	-1.6	-1.8	-1.4	0.8	0.5	1	0.68	0.49	0.88
Q8	-1.6	-1.8	-1.4	0.4	0.3	0.6	0.66	0.49	0.91
Q9	-1.5	-1.6	-1.3	0.8	0.5	1.2	0.66	0.47	0.84
Q10	-1.4	-1.6	-1.2	1.3	0.3	2.2	0.62	0.36	0.99
Q11	-1.7	-1.8	-1.5	0.6	0.4	0.7	0.75	0.53	0.94
Q12	-1.7	-1.8	-1.5	0.5	0.4	0.7	0.79	0.58	0.99
Qall	-1.9	-2	-1.8	0.4	0.3	0.5	0.94	0.70	0.99

Figure 2.11: **Predicted Participation Probabilities for Varying Numbers of Previously Participating Friends for Model 3: Endorsed Adoption, Homogeneous Diffusion**



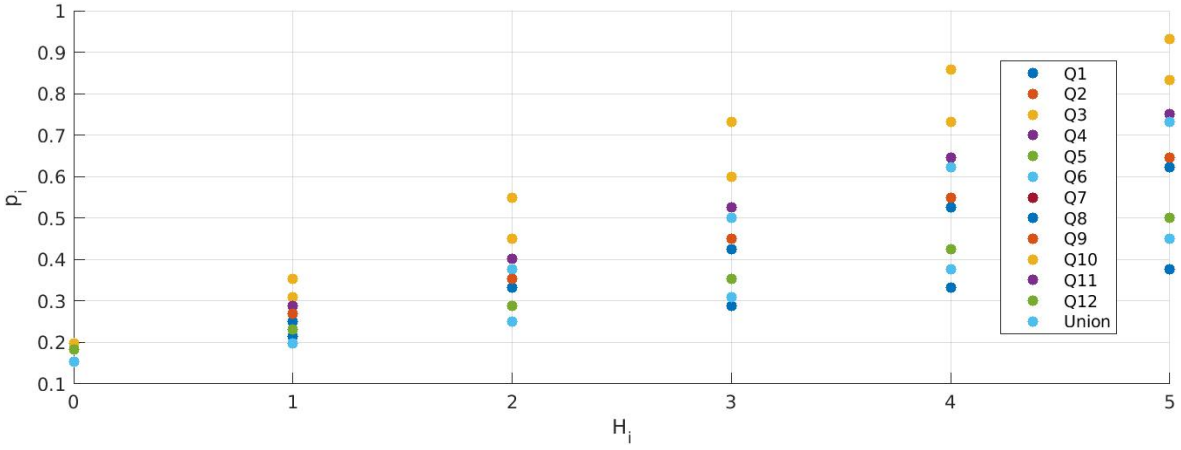
2.7.2 Model 4: Endorsed Adoption, Differentiated Diffusion

In model four, additional to the endorsement effect, I allow for the possibility that participants and non participants transmit the information with different rates. This thus has the effect that a past participant alters her neighbors' outcomes in two ways: first by providing the information and second by impacting her borrowing choice. As can be seen from Table 2.5, both of these effects are positive and significant.
Accordingly, I find a drop in the participation probability for those newly informed individuals that are not linked to any participating IP and an increase for those with one or more participating IP friend(s). Further, I also see that q_N increases in all while q_P is reduced in some cases. This is intuitive: having a lower probability to opt in when informed, those that have no participating friends must be more likely to receive the information. The confidence sets vary substantially in size (in the sense of gridpoints included) being most narrow for the densest networks (the union). It is in particular q_P and γ that feature a large variance and are thus weakly identified.
As mentioned above, in model four, participating friends have a twofold effect on the individual: first, they pass on the information with a higher probability, second, they nudge information recipients into participation. The first effect is decreasing in H_i, we can thus speak of a "diminishing return to participating friends". The reason for this is that it is sufficient for i to receive the message just once, receiving it multiple times does not provide any additional benefit. The new friend has a benefit to i only in the event that "all previous friends fail to send the information" - an event that clearly becomes less and less likely the more previous friends i has. The effect of H_i on p_i on the other hand is also decreasing, but much less so than the first. In particular, each additional friend has the same (linear) impact on the individual's odds ratio.

Table 2.5: **Estimates and Confidence Intervals for Model 4: Endorsed Adoption, Differentiated Diffusion**

	α			γ			qN			qP		
		95%			95%			95%			95%	
	$\hat{\alpha}$	conf.	set	$\hat{\gamma}$	conf.	set	$q\hat{N}$	conf.	set	$q\hat{P}$	conf.	set
Q1	-1.5	-1.7	-1.3	0.4	0.1	0.6	0.52	0.37	0.72	0.97	0.67	0.99
Q2	-1.4	-1.7	-1.2	0.4	0.1	0.9	0.53	0.39	0.79	0.91	0.53	0.99
Q3	-1.4	-1.7	-1.2	0.6	0.2	1.2	0.49	0.36	0.72	0.7	0.39	0.99
Q4	-1.4	-1.7	-1.2	0.5	0.2	0.9	0.54	0.41	0.78	0.86	0.54	0.99
Q5	-1.5	-1.7	-1.3	0.5	0.2	0.8	0.61	0.43	0.85	0.86	0.54	0.99
Q6	-1.5	-1.7	-1.3	0.5	0.2	0.8	0.55	0.38	0.76	0.88	0.56	0.99
Q7	-1.4	-1.6	-1.2	0.4	0.2	0.8	0.49	0.37	0.72	0.99	0.68	0.99
Q8	-1.5	-1.7	-1.3	0.2	0	0.4	0.5	0.36	0.71	0.99	0.69	0.99
Q9	-1.4	-1.6	-1.2	0.4	0.1	0.9	0.55	0.4	0.76	0.99	0.62	0.99
Q10	-1.4	-1.6	-1.2	0.8	-0.1	1.5	0.6	0.32	0.99	0.89	0.33	0.99
Q11	-1.5	-1.7	-1.3	0.3	0.1	0.5	0.52	0.37	0.73	0.99	0.73	0.99
Q12	-1.5	-1.7	-1.3	0.3	0.1	0.5	0.55	0.4	0.8	0.99	0.73	0.99
Qall	-1.7	-1.8	-1.5	0.3	0.1	0.4	0.58	0.37	0.81	0.99	0.81	0.99

Figure 2.12: **Predicted Participation Probabilities for Varying Numbers of Previously Participating Friends for Model 4: Endorsed Adoption, Differentiate Diffusion**



2.7.3 Hypothesis Testing

Estimation of models one to four allows for performing tests one and two.

Tables 2.6 indicates that the hypothesis of homogeneous diffusion can be rejected in

all but two cases. This thus indicates that indeed participants fulfill the twofold role of submitting information and inducing others to borrow.
Table 2.7 and 2.8 provide strong evidence for the existence of a positive endorsement effect: for any network measure, the hypothesis of no endorsement can be rejected. Including differentiated diffusion decreases the value of the LR statistic testing for $\gamma = 0$, however, the value is still largely in the rejection region.

Table 2.6: **Test 1.2.: Homogeneous Versus Differentiated Diffusion With Endorsed Adoption Using Model Three (Restricted Model) Versus Model Four (Unrestricted Model)**

	M3 $H_0 : qN = qP$ $\mathcal{L}(\theta_0)$	M4 $H_1 : qN \neq qP$ $\mathcal{L}(\theta_1)$	LR stat.	$\chi^2(1)$ crit. value 3.841
Q1	-1526.85	-1517.19	19.32	reject
Q2	-1207.4	-1203.7	7.4	reject
Q3	-1267.98	-1266.52	2.92	not reject
Q4	-1471.42	-1467.94	6.96	reject
Q5	-1498.44	-1495.9	5.08	reject
Q6	-1444.74	-1440.06	9.36	reject
Q7	-1428.74	-1419.31	18.86	reject
Q8	-1624.87	-1614.96	19.82	reject
Q9	-1229.31	-1223.28	12.06	reject
Q10	-628.118	-627.706	0.824	not reject
Q11	-1689.26	-1679.16	20.2	reject
Q12	-1740.3	-1732.71	15.18	reject
Qall	-2258.04	-2248.65	18.78	reject

Table 2.7: **Test 2.1.: Homogeneous Versus Endorsed Adoption Given Homogeneous Diffusion Using Model One (Restricted Model) Versus Model Three (Unrestricted Model)**

	M1 $H_0 : \gamma = 0$ $\mathcal{L}(\theta_0)$	M3 $H_1 : \gamma \neq 0$ $\mathcal{L}(\theta_1)$	LR stat.	$\chi^2(1)$ crit. value 3.841
Q1	-1548.99	-1526.85	44.28	reject
Q2	-1225.49	-1207.4	36.18	reject
Q3	-1289.43	-1267.98	42.9	reject
Q4	-1502.58	-1471.42	62.32	reject
Q5	-1521.97	-1498.44	47.06	reject
Q6	-1470.82	-1444.74	52.16	reject
Q7	-1458.61	-1428.74	59.74	reject
Q8	-1638.2	-1624.87	26.66	reject
Q9	-1251.51	-1229.31	44.4	reject
Q10	-635.011	-628.118	13.786	reject
Q11	-1712.34	-1689.26	46.16	reject
Q12	-1761.88	-1740.3	43.16	reject
Qall	-2288.16	-2258.04	60.24	reject

Table 2.8: **Test 2.2.: Homogeneous Versus Endorsed Adoption Given Differentiated Diffusion Using Model Two (Restricted Model) Versus Model Four (Unrestricted Model)**

	M2 $H_0 : \gamma = 0$ $\mathcal{L}(\theta_0)$	M4 $H_1 : \gamma \neq 0$ $\mathcal{L}(\theta_1)$	LR stat.	$\chi^2(1)$ crit. value 3.841
Q1	-1522.52	-1517.19	10.66	reject
Q2	-1209.22	-1203.7	11.04	reject
Q3	-1274.31	-1266.52	15.58	reject
Q4	-1479.28	-1467.94	22.68	reject
Q5	-1504.6	-1495.9	17.4	reject
Q6	-1448.52	-1440.06	16.92	reject
Q7	-1428.92	-1419.31	19.22	reject
Q8	-1616.94	-1614.96	3.96	reject
Q9	-1229.83	-1223.28	13.1	reject
Q10	-631.037	-627.706	6.662	reject
Q11	-1684.76	-1679.16	11.2	reject
Q12	-1738.6	-1732.71	11.78	reject
Qall	-2257.2	-2248.65	17.1	reject

2.8 Models including Co-variates

Besides different measures of the village network, the information initiation and the participation decisions, we also have data on household characteristics.[7] [8]. For each household, we dispose over information on the religion of the household head, the roof type, the number of rooms, the number of beds, the house ownership status (owned, shared, rented, leased or government provided) as well as whether the household disposes over private, government provided or no electricity and whether the household had access to their own, shared or no latrine.

To pick the variables that significantly impact the participation decision, I first conduct a logit regression using only the set of IPs, regressing participation on all household characteristics. As the IPs are known to be informed and by assumption have the same adoption probability as informed non-IPs, this enables me to identify those household

[7]The data-set also includes individual characteristics, but we refrain from using them firstly because only one credit per household can be disbursed and secondly because only a subset of 46.30 % of the household did participate in the individual survey.

[8]The variable "caste" was excluded as it was missing for some entire villages

characteristics that have a significant impact for the adoption decision. Out of all aforementioned characteristics, only the religion of the household head and the number of rooms in the house are significant. Consequently, these two covariates are included in model five and six. Model five again features a heterogeneous diffusion rate while model six features differentiated diffusion rates.
In model five and six, as stated above, β_1 is the coefficient of the number of rooms in the household and β_2 is the coefficient of the Hindu-dummy variable.
Households dispose over zero[9] to twenty rooms. Ninety nine percent of households have only six rooms or less.
Not including covariates can lead to confounding due to homophily: we observe individual i and her neighbour, individual j to opt in one shortly after the other and we attribute this to a network effect that i exercises on j when in reality it is their characteristics that cause both of them to opt in. Since they may select friends with similar characteristics (homophily) a cluster of friends may be more or less likely to opt in, not because of network effects or mutual influence through the network, but because of similarities in the covariates that determine both participation and linking. In presence of this preferential attachment effect, household characteristics are correlated across neighbours.
This indeed appears to be the case, as can be seen from Table 2.9. Theoretically, the sample realizations of the household characteristics should be independent across households, and the data appears to reconfirm that this, as correlation in the two characteristics for all individuals is very close to zero. Observe, however, that if two individuals are neighbors, the covariance in their characteristics is non-negligible and positive. Neighbours are more likely to both deviate from the mean in one and the same direction. This is an evidence for homophily linking.

2.8.1 Model 5: Socio-economically determined Adoption, Homogeneous Diffusion

Table 2.10 shows the estimates and confidence sets. The coefficient of the number of rooms is β_1, β_2 is the coefficient of the dummy variable indicating that the household belongs to the Hindu religion. Both household characteristics have a negative impact on the household's probability to take up a loan, indicating that both Hindus and individuals living in houses with more rooms are ceteris paribus less likely to participate. As the number of rooms can be perceived as a wealth proxy, this can be a positive

[9]It is not clear to me what it implies if a household indicates to dispose over a accommodation with zero rooms, but this is a frequent data entry

Table 2.9: **Average Observed Covariances in Relevant Household Characteristics**

	Covariance in *Hindu*	Covariance in *Rooms*
all individuals	-0.000172047	-0.00731055
neighbours in the Q1-network	0.0210039	0.157914
neighbours in the Q2-network	0.0230822	0.170005
neighbours in the Q3-network	0.0249542	0.160597
neighbours in the Q4-network	0.0238819	0.153676
neighbours in the Q5-network	0.0240561	0.183825
neighbours in the Q6-network	0.021048	0.156206
neighbours in the Q7-network	0.024215	0.225827
neighbours in the Q8-network	0.017342	0.141868
neighbours in the Q9-network	0.0333316	0.239923
neighbours in the Q10-network	0.0288934	0.298586
neighbours in the Q11-network	0.0220211	0.189858
neighbours in the Q12-network	0.0213379	0.180338
neighbours in any network	0.0218962	0.135264
	Sample Variance in *Hindu*	Sample Variance in *Rooms*
	0.0366867	1.55047

message as the organization succeeds in reaching poorer households. The introduction of household covariates leads to a drop in q in the range of one to nine percent. The drop in q lets me conclude that there could indeed be confounding due to homophily in the estimation of model one: erroneously, we concluded that it is the messaging between two participants that caused them to opt in shortly after one another while indeed it may be their common characteristic that have caused both, the link between them and their interest in micro-finance. Once we drop the assumption of homogenous participation rate, higher probabilities to participate due to (similar) characteristics of IP-participants and their second wave participating friends are accounted for and the substantially lowered q reveals the true diffusion rate.

The confidence interval for the wealth proxy (*Rooms*) is very narrow, indicating that the effect is present albeit small. The confidence set for the *Hindu* dummy on the other hand is relatively wide and in some cases includes zero.

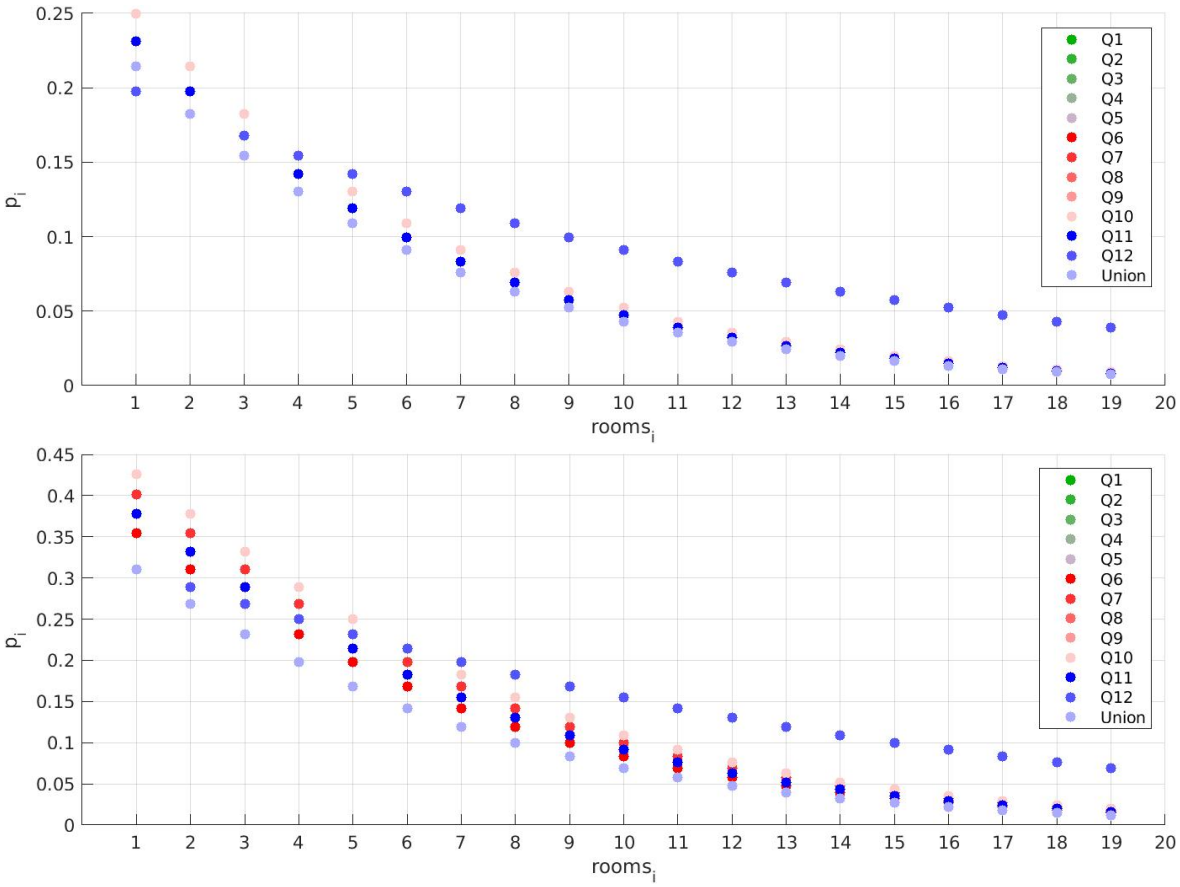
Figure 2.13 shows the predicted participation probabilities as a function of the rooms in the house for Hindus and households belonging to a different religion, ceteris paribus. Again, as in Table 2.10, the estimates (thus the predictions) resulting from different networks often overlap, thus less than thirteen points are visible. I discern that Hin-

dus indeed start at a much lower rate. Thereafter, for both Hindus and non-Hindus, the participation probabilities decrease, reaching a level very close to zero when the household has sixteen or more rooms.

Table 2.10: **Estimates and Confidence Intervals for Model 5: Socio-economically Determined adoption, Homogeneous Diffusion**

	α			β_1			β_2			q		
		95%			95%			95%			95%	
	$\hat{\alpha}$	conf.	set	$\hat{\beta}_1$	conf.	set	$\hat{\beta}_2$	conf	set	$\hat{q}$	conf.	set
Q1	-0.4	-1.2	0.2	-0.2	-0.2	-0.1	-0.6	-1.1	0	0.65	0.48	0.84
Q2	-0.1	-1	0.6	-0.2	-0.3	-0.1	-0.8	-1.5	-0.1	0.63	0.49	0.87
Q3	-0.3	-1	0.5	-0.2	-0.3	-0.1	-0.6	-1.2	0.1	0.58	0.45	0.79
Q4	-0.2	-0.9	0.4	-0.2	-0.3	-0.1	-0.7	-1.3	-0.1	0.65	0.51	0.84
Q5	-0.2	-1	0.4	-0.2	-0.3	-0.1	-0.7	-1.3	-0.1	0.65	0.52	0.9
Q6	-0.4	-1.2	0.2	-0.2	-0.2	-0.1	-0.5	-1.1	0.1	0.63	0.49	0.85
Q7	-0.2	-0.9	0.4	-0.2	-0.3	-0.1	-0.7	-1.3	-0.1	0.64	0.5	0.87
Q8	-0.7	-1.3	0.2	-0.1	-0.2	-0.1	-0.6	-1.2	0	0.66	0.47	0.86
Q9	-0.3	-1	0.4	-0.2	-0.3	-0.1	-0.6	-1.2	0.1	0.68	0.52	0.88
Q10	-0.1	-1	0.5	-0.2	-0.3	-0.1	-0.8	-1.5	-0.1	0.77	0.57	0.9
Q11	-0.3	-1.1	0.4	-0.2	-0.3	-0.1	-0.7	-1.4	0	0.66	0.49	0.83
Q12	-0.7	-1.2	0.3	-0.1	-0.2	-0.1	-0.6	-1.3	0.1	0.74	0.52	0.91
Qall	-0.6	-1	0.2	-0.2	-0.3	-0.1	-0.5	-1.2	0.1	0.66	0.5	0.84

Figure 2.13: **Predicted Participation Probabilities for Varying Numbers of Rooms for Model 5: Socio-economically Determined Adoption, Homogeneous Diffusion**



2.8.2 Model 6: Socio-economically determined Adoption, Differentiated Diffusion

Model six also allows for heterogeneous diffusion rates. The results are tabulated in table 2.11. While diffusion rates change little as compared to the more parsimonious model two, still covariates are jointly significant. *Rooms* again features a small and *Hindu* a substantially larger confidence set. Since the estimates do not change much, I refrain from plotting the predicted participation probabilities again.

Table 2.11: **Estimates and Confidence Intervals for Model 6: Socio-economically Determined Adoption, Differentiated Diffusion**

	α			β_1			β_2		
		95%			95%			95%	
	$\hat{\alpha}$	conf.	set	$\hat{\beta}_1$	conf.	set	$\hat{\beta}_2$	conf	set
Q1	-0.4	-1.2	0.2	-0.2	-0.3	-0.1	-0.4	-1.1	0.2
Q2	-0.2	-1	0.6	-0.2	-0.3	-0.1	-0.7	-1.4	0
Q3	-0.5	-1.2	0.5	-0.2	-0.3	-0.1	-0.4	-1.1	0.3
Q4	-0.2	-0.9	0.4	-0.2	-0.3	-0.1	-0.6	-1.2	0.1
Q5	-0.3	-1	0.3	-0.2	-0.3	-0.1	-0.6	-1.2	0.1
Q6	-0.4	-1.2	0.3	-0.2	-0.3	-0.1	-0.4	-1.1	0.2
Q7	-0.2	-0.8	0.4	-0.2	-0.3	-0.1	-0.6	-1.2	0
Q8	-0.7	-1.3	0.1	-0.1	-0.2	-0.1	-0.5	-1.1	0.1
Q9	-0.3	-1	0.3	-0.2	-0.3	-0.1	-0.5	-1.1	0.2
Q10	-0.1	-1	0.9	-0.2	-0.4	-0.1	-0.8	-1.7	0.2
Q11	-0.3	-1.1	0.3	-0.2	-0.3	-0.1	-0.6	-1.1	0.1
Q12	-0.6	-1.2	0.2	-0.1	-0.2	-0.1	-0.5	-1.1	0.1
Qall	-0.5	-1.2	-0.1	-0.2	-0.2	-0.1	-0.4	-0.9	0.2

	qN			qP		
		95%			95%	
	$q\hat{N}$	conf.	set	$q\hat{P}$	conf.	set
Q1	0.41	0.32	0.59	0.99	0.76	0.99
Q2	0.51	0.35	0.68	0.99	0.72	0.99
Q3	0.47	0.32	0.64	0.99	0.64	0.99
Q4	0.47	0.35	0.62	0.99	0.76	0.99
Q5	0.53	0.37	0.69	0.99	0.72	0.99
Q6	0.44	0.33	0.61	0.99	0.75	0.99
Q7	0.43	0.31	0.58	0.99	0.8	0.99
Q8	0.46	0.33	0.63	0.99	0.73	0.99
Q9	0.48	0.4	0.65	0.99	0.79	0.99
Q10	0.57	0.28	0.73	0.99	0.6	0.99
Q11	0.45	0.4	0.59	0.99	0.78	0.99
Q12	0.45	0.34	0.63	0.99	0.78	0.99
Qall	0.36	0.28	0.54	0.99	0.78	0.99

2.8.3 Hypothesis Testing

Having estimated model five and six, I can conduct the tests 1.3, 3.1 and 3.2. In line with the findings from above, the hypothesis of a homogeneous diffusion rate can be rejected. Additionally, the hypothesis that the coefficients of both *Rooms* and *Hindu* are zero can be rejected for all questions. This induces me to conclude that indeed the richer models provide a significant improvement over the more parsimonious ones.

Table 2.12: **Test 1.3.: Homogeneous Versus Differentiated Diffusion Given Socio-economically determined Adoption Using Model Five (Restricted Model) versus Model Six (Unrestricted Model)**

	M5 $H_0 : qN = qP$ $\mathcal{L}(\theta_0)$	M6 $H_1 : qN \neq qP$ $\mathcal{L}(\theta_1)$	LR stat.	$\chi^2(1)$ crit. value 3.841
Q1	-1535.12	-1509.65	50.94	reject
Q2	-1210.9	-1195.8	30.2	reject
Q3	-1271.43	-1258.61	25.64	reject
Q4	-1485.59	-1463.54	44.1	reject
Q5	-1505.6	-1489.96	31.28	reject
Q6	-1458.7	-1436.93	43.54	reject
Q7	-1440.57	-1412.54	56.06	reject
Q8	-1626.86	-1606.8	40.12	reject
Q9	-1236.51	-1216.03	40.96	reject
Q10	-623.88	-619.863	8.034	reject
Q11	-1696.38	-1670.64	51.48	reject
Q12	-1748.84	-1726.78	44.12	reject
Qall	-2272.29	-2241.79	61	reject

Table 2.13: **Test 3.1.: Homogeneous Versus Socio-economically Determined Adoption Given Homogeneous Diffusion Using Model One (Restricted Model) against Model Five (Unrestricted Model)**

	M1 $H_0 : \beta_1 = \beta_2 = 0$ $\mathcal{L}(\theta_0)$	M5 H_1 $\mathcal{L}(\theta_1)$	LR stat.	$\chi^2(2)$ crit. value 5.991
Q1	-1548.99	-1535.12	27.74	reject
Q2	-1225.49	-1210.9q	29.18	reject
Q3	-1289.43	-1271.43	36	reject
Q4	-1502.58	-1485.59	33.98	reject
Q5	-1521.97	-1505.6	32.74	reject
Q6	-1470.82	-1458.7	24.24	reject
Q7	-1458.61	-1440.57	36.08	reject
Q8	-1638.2	-1626.86	22.68	reject
Q9	-1251.51	-1236.51	30	reject
Q10	-635.011	-623.88	22.262	reject
Q11	-1712.34	-1696.38	31.92	reject
Q12	-1761.88	-1748.84	26.08	reject
Qall	-2288.16	-2272.29	31.74	reject

Table 2.14: **Test 3.2.: Homogeneous Versus Socio-economically Determined Adoption Given Differentiated Diffusion Using Model Two (Restricted Model) against Model Six (Unrestricted Model)**

	M2	M6	LR	$\chi^2(2)$
	$H_0 : \beta_1 = \beta_2 = 0$	H_1	stat.	crit. value
	$\mathcal{L}(\theta_0)$	$\mathcal{L}(\theta_1)$		5.991
Q1	-1522.52	-1509.65	25.74	reject
Q2	-1209.22	-1195.8	26.84	reject
Q3	-1274.31	-1258.61	31.4	reject
Q4	-1479.28	-1463.54	31.48	reject
Q5	-1504.6	-1489.96	29.28	reject
Q6	-1448.52	-1436.93	23.18	reject
Q7	-1428.92	-1412.54	32.76	reject
Q8	-1616.94	-1606.8	20.28	reject
Q9	-1229.83	-1216.03	27.6	reject
Q10	-631.037	-619.863	22.348	reject
Q11	-1684.76	-1670.64	28.24	reject
Q12	-1738.6	-1726.78	23.64	reject
Qall	-2257.2	-2241.79	30.82	reject

2.9 Limitations

All estimations are done under the assumption of network exogeneity, hence questioning this assumption naturally implies questioning the results.

If, for instance, there are unobserved variables relevant for the individuals' linking decisions that also impact the participation decision, the network is not truly exogenous, but correlated with the error in the participation equation. If, for example, we assume a clustering based on some unobservable variable that impacts the adoption decision, we would expect an over-estimation of γ, as individuals in a cluster that features a higher (lower) probability to participate will also always have more (less) participating friends. For β_1 and β_2 a bias could occur if household wealth and religion are correlated with the cluster specific fixed effect in participation.

This, however, is an issue for any estimation technique applied to this data that assumes exogeneity of the network, thus also for the SMM estimation in Banerjee et al. (2013).

Further, it is quite plausible to assume that individuals who are neighbors and who face

the choice at a same point in time impact one another. If present, contemporaneous correlation creates a system of simultaneous equations challenging to estimate.
Table 2.15 provides some evidence that indeed contemporaneous correlation may be present in the data. For all non-IPs, I tabulate the number of friends that participate in period two. We can see clearly that individuals that themselves opt in at time $t = 2$ also have many more friends participating at the same point in time, thus making contemporaneous impact plausible.
I am planning to address both of the aforementioned issues (network endogeneity and contemporaneous correlation) in my follow-up work.

Table 2.15: **Average Number of Non-IP Participating Friends for Non-IP Individuals**

	non-IP non-participants	non-IP participants
Q1	0.332067	0.711469
Q2	0.214482	0.493631
Q3	0.209275	0.448763
Q4	0.334595	0.765346
Q5	0.330072	0.762005
Q6	0.328364	0.72129
Q7	0.299708	0.573788
Q8	0.437663	0.869274
Q9	0.240396	0.412633
Q10	0.0317262	0.0990766
Q11	0.465549	1.01413
Q12	0.474755	1.0089
Union	0.951434	1.68076

2.10 Conclusion

The chapter proposes a new method to estimate the diffusion model from Banerjee et al. (2013). The algorithm is faster and more efficient than the estimation technique used by the original authors. A series of nested models are analyzed and compared, yielding important insights. First, from my evidence, I can conclude that low participation rates are a product design, rather than a marketing problem. Hindus and rich households are particularly unlikely to take-up a micro-loan. In all models, non participants are significantly less likely to pass on the information than participants. Further, I find evidence of a positive and substantial endorsement effect, i.e. previous participants impact their neighbours' adoption probabilities positively. Given the limitations of the word-of-mouth marketing, increasing participation among initially informed (IP)

individuals is a challenge just as important as choosing the right IPs as a high initial adoption leads to twofold benefits of higher diffusion and a positive endorsement effect. In the next two chapters, larger time horizons are considered. Further extensions allow for simultaneous coordination between individuals that face a choice at the same point in time.

2.11 Appendix

2.11.1 The Banerjee et al. (2013) Algorithm for Two Time Periods

Following the notation in the original paper, I use the subscript $r = 1, .., R$ when I refer to the moments or the data for village r, ι_r a vector of ones of same size as the adjacency matrix of village r. Further, the subscript $s = 1, ..., S$ indicates that the respective moment results from simulation run s.
The original estimation technique is a two-step procedure. First, the set of all IPs is used to identify adoption parameters (α in model 1,2,3,4 and α, β_1, β_2 in model 5,6) by means of a logit regression.[10] As the original paper does not take the time dimension into account, the authors assume that any village only reaches the final participation rate after 5-8 time periods. In the reported participation rates, however, I observe that for all 37 villages in our sample, the participation rates ceases to increase after $T = 2$ (13 villages), $T = 3$ (13 villages) or $T = 4$ (11 villages). For my algorithm, I take into account the participation rates at $t = 1$ (IP-participation) and $t = 2$ (first wave of non-IP participation). As a consequence, to facilitate comparison, I assume that there is just one information exchange.
The algorithm from Banerjee et al. (2013) proceeds as follows. Disposing over the first stage estimates, diffusion parameters (all models) and the endorsement parameter (model 4,5) are found by grid-search. For each point on the grid and each simulation run three steps were carried out

1. ip participation: random draw from Bernoulli distribution with mean $\hat{p}_i$ for each PII (i participates if $\varepsilon_i < \hat{p}_i$) where $\hat{p}_i$ is the predicted participation probability for individual i calculated using the first-step estimates

2. - information exchange model (1,3,5) random draw from a Bernoulli distribution with mean q for each link (i informs j if $\varepsilon_{i,j} < q$)

[10] Note that Banerjee et al. (2013) include a larger set of household covariates

- information exchange model (2,4,6) if i is a participant, random draw from a Bernoulli distribution with mean q_P for each of $i's$ links (i informs j if $\varepsilon_{i,j} < q_P$) else random draw from a Bernoulli distribution with mean q_N for each of $i's$ links (i informs j if $\varepsilon_{i,j} < q_N$)

3.
- non-IP participation model 1,2,5,6: random draw from a Bernoulli distribution with mean $\hat{p}_i$ for each informed agent (j participates if $\varepsilon_j < \hat{p}_i$) where $\hat{p}_i$ is the individual-specific participation probability calculated using the first-step estimates (α in model 1,2 and α, β_1, β_2 in model 5,6)
- non-IP participation model 3,4: calculation of the number of participating friends that in the previous step submitted the information: Random draw from Bernoulli distribution with mean $\hat{p}_i$ for each informed agent (j participates if $\varepsilon_j < \hat{p}_i$), where again $\hat{p}_i$ is the individual-specific participation probability calculated using the first-step estimates ($\hat{\alpha}$) and the value of γ drawn from the grid and multiplied by the headcount of past participants that did sent the message.

Finally, the moments for round s and village r (denoted by $m_{r,s}(q)$ or respectively $m_{r,s}(q_N, q_P)$) are calculated ($S = 75$). The simulated moments for village r are the average of the S runs

$$m_{r,sim}(q) = \frac{1}{S}\sum_{s=1}^{S} m_{r,s}(q)$$

$$m_{r,sim}(q_N, q_P) = \frac{1}{S}\sum_{s=1}^{S} m_{r,s}(q_N, q_P)$$

or respectively

$$m_{r,sim}(q_N, q_P, \gamma) = \frac{1}{S}\sum_{s=1}^{S} m_{r,s}(q_N, q_P, \gamma)$$

Disposing over the empirical moments ($m_{r,emp}$), this leads to the vector of deviations of the simulated from the observed moments (denoted $d_r(q)$ and $d_r(q_N, q_P)$ for each village

$$d_r(q) = m_{r,sim}(q) - m_{r,emp}$$

$$d_r(q_N, q_P) = m_{r,sim}(q_N, q_P) - m_{r,emp}$$

$$d_r(q_N, q_P, \gamma) = m_{r,sim}(q_N, q_P, \gamma) - m_{r,emp}$$

and averaging over villages leads to the vector of sample deviations

$$D(q) = \frac{1}{R}\sum_{r=1}^{R} d_r(q)$$
$$D(q_N, q_P, \gamma) = \frac{1}{R}\sum_{r=1}^{R} d_r(q_N, q_P)$$
$$D(q_N, q_P) = \frac{1}{R}\sum_{r=1}^{R} d_r(q_N, q_P, \gamma)$$

Then the optimal grid point is chosen that the inner product of this vector is minimized.[11]

$$\hat{q} = min_q \; D'D \qquad \text{model (1,5)}$$
$$\widehat{qN}, \widehat{qP} = min_{qN,qP} \; D'D \qquad \text{model (2,6)}$$
$$\hat{q}, \hat{\gamma} = min_{q,\gamma} \; D'D \qquad \text{model (3)}$$
$$\widehat{qN}, \widehat{qP}, \hat{\gamma} = min_{q_N,q_P,\gamma} \; D'D \qquad \text{model (4)}$$

Banerjee et al. (2013) propose a set of moments, which I slightly adapt to suit the shortened time horizon. If there is only one round (as opposed to five or eight rounds) of information exchange, the ratio of IPs to other individuals used for estimation (PIIs and participants) is larger. Since IPs cannot provide any information that is useful for identification of diffusion and endorsement parameters, my suspicion is that including them in the moments used to estimate the latter (as done in Banerjee et al. (2013)) will introduce noise. To verify this, for each moment suggested in Banerjee et al. (2013) I include a pooled version in which all households contribute and another in which only non-IP households contribute. Finally, a moment taking into account links of lengths two is not very informative in the $T = 2$ case, hence this moment was excluded. This leads to the following twelve moments:[12]

1. The share of households who participate
2. The share of leaders who participate
3. The share of households with no participating neighbour who participate

[11] Banerjee et al. (2013) also use a weighting matrix in their objective function, which is obtained in a two-step procedure

[12] These moments correspond to my interpretation of how the original moments should be adapted to suit the shortened time horizon

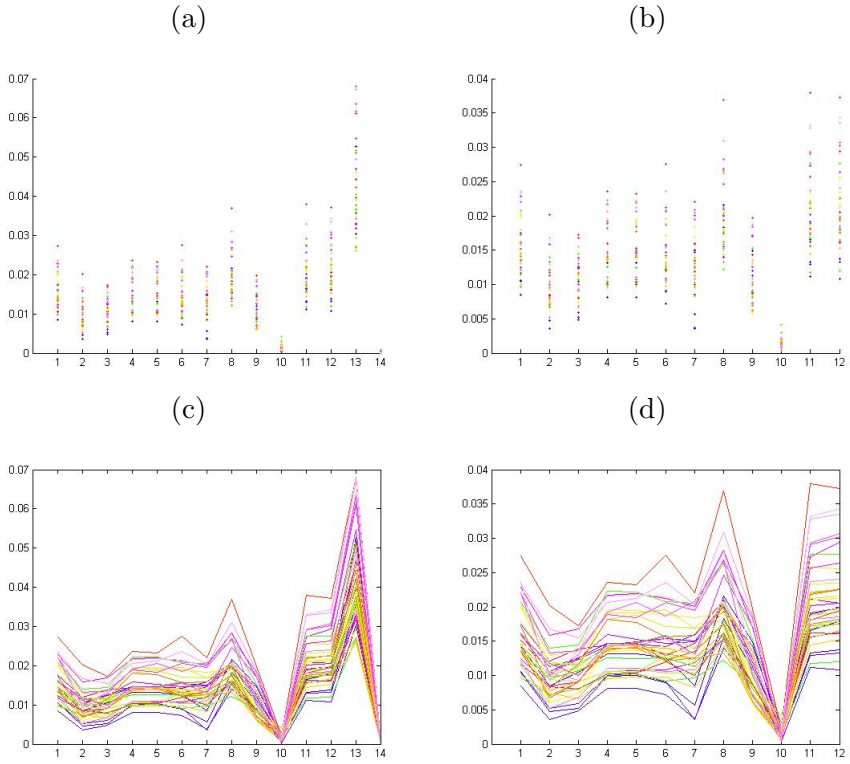
4. The share of non-IP households with no participating neighbour who participate
5. The share of households with at least one participating neighbour who participate
6. The share of non-IP households with at least one participating neighbour who participate
7. The share of households with no neighbour who is a participating IP who participate
8. The share of non-IP households with no neighbour who is a participating IP who participate
9. The share of households with at least one neighbour who is a participating IP who participate
10. The share of non-IP households with at least one neighbour who is a participating IP who participate
11. The covariance of household participation with with the share of its neighbours who participate
12. The covariance of non-IP household participation with with the share of its neighbours who participate

In this chapter, I used the best performing moment choice (in the sense of being closest to the MLE). These are moments eight and ten. Various other combination were tried, but indeed including the even moments worsens performance. However, in any case results all show the same patterns and remain close to one another.

2.11.2 Appendix 2: Densities and Similarity Indices (Jacquard)

The density of a network is defined as the proportion of potential links that do actually exist. As we can see in Figure 2.14, the villages all exhibit the same pattern: questions that lead to a dense (sparse) network do so in all villages (observe that the villages specific lines in plot (c) and (d) hardly ever intersect. This reconfirms us in treating the village data-sets as resulting from the same underlying data-generating process. Furthermore, on the basis of the Jacquard index, average overlap of links between questions is small.

Figure 2.14: Network densities of individual villages for question 1 to 12 as well as the Union andt he Intersection



2.11.3 Appendix 4: Parameter Variance in the Homogeneous Adoption and Diffusion Model

The asymptotic variance of the MLE is given by the inverse of the Fisher Information Matrix, which is the negative of the expected value of the Hessian

$$Var(\theta) = \mathcal{I}^{-1}N^{-1} = -N^{-1}E[\mathcal{H}^{-1}]$$

The observed Fisher Information given the particular data sample is

$$-\mathcal{H}(\hat{\theta})$$

Thus, the observed variance in p and q are

$$Var(q) = -N^{-1}\frac{\frac{\partial^2\mathcal{L}}{\partial^2 q}}{\frac{\partial^2\mathcal{L}}{\partial^2 p}\frac{\partial^2\mathcal{L}}{\partial^2 q} - \left(\frac{\partial^2 Q}{\partial p \partial q}\right)^2}$$

$$Var(q) = -N^{-1}\frac{\frac{\partial^2\mathcal{L}}{\partial^2 p}}{\frac{\partial^2\mathcal{L}}{\partial^2 p}\frac{\partial^2\mathcal{L}}{\partial^2 q} - \left(\frac{\partial^2 Q}{\partial p \partial q}\right)^2}$$

From which it is apparent that a second derivative of the log likelihood function with respect to q that is small in absolute value translates into a small variance in p and a large variance in q. In the most parsimonious model, this second derivative is

$$\frac{\partial^2 \mathcal{L}}{\partial^2 q} = \sum_{k=1}^{N_{P2}} \frac{r_k \frac{\partial^2 r_k}{\partial^2 q} - (\frac{\partial r_k}{\partial q})^2}{r_k} - \sum_{l=1}^{N_{PII}} \frac{(1 - r_l p)\frac{\partial^2 r_l}{\partial^2 q} p - (\frac{\partial r_l}{\partial q} p)^2}{(1 - r_l p)^2}$$

if ω_i is relatively high, then r_i is high, $\frac{\partial r}{\partial q}$ is small and positive and $\frac{\partial^2 r}{\partial^2 q}$ is small and negative. As a consequence, the second derivative is small in absolute value if participants are well-connected to the IPs, while non-participants have much less IP friends.