Econometrica Supplementary Material

SUPPLEMENT TO “TESTING MODELS OF SOCIAL LEARNING ON NETWORKS: EVIDENCE FROM TWO EXPERIMENTS”

(*Econometrica*, Vol. 88, No. 1, January 2020, 1–32)

ARUN G. CHANDRASEKHAR
Department of Economics, Stanford University, NBER, and J-PAL

HORACIO LARREGUY
Department of Governement, Harvard University

JUAN PABLO XANDRI
Department of Economics, Princeton University

APPENDIX B: BAYESIAN LEARNING ALGORITHM IN INCOMPLETE INFORMATION MODELS

IN THIS APPENDIX, we describe the algorithm for computing the actions in the complete and incomplete information Bayesian model.

B.1. *Setup*

We follow the notation in Osborne and Rubinstein (1994) and Geanakoplos (1994), modeling agents' information in the experiment by means of *dynamically consistent models of action and knowledge* (DCMAK), a natural multi-period generalization of Aumann (1976). Following Geanakoplos (1994), a DCMAK consists of a set of *states of the world* $\omega \in \Omega$, information functions $P_{i,t} : \Omega \to 2^{\Omega}$, and action functions $a_{i,t} : \Omega \to \{0, 1\}$. In what follows, we define these objects for our experimental setup, which we use to calculate the predicted behavior of Bayesian agents $a_{i,t}(\omega)$.

B.2. *States of the World*

In both the complete and incomplete information models, we model agents' information as partitions over $\omega \in \Omega$, where $\omega = (\omega_1, \omega_2 \dots, \omega_n)$ is the vector of agents' initial private information. In the incomplete information model, we model the state of the world as $\omega_i = (s_i, \eta_i)$, where $s_i \in \{0, 1\}$ is the color of the observed ball, and $\eta_i \in \{0, 1\}$ denotes agent i's *type*: she is either a Bayesian type ($\eta_i = 1$) who guesses the most likely state following Bayes's rule, or a DeGroot agent ($\eta_i = 0$) who decides her guess based on an average of her neighbors' and own previous guesses. Both s_i and η_i are drawn i.i.d. across agents and types and signals are independent of each other as well. Bayesian agents have a common prior belief over states $\omega \in \Omega$, conditional on the realization of $\theta \in \{0, 1\}$ (i.e., which bag has been chosen), which we denote by $\rho(\omega \mid \theta)$. Then

$$\rho(s, \eta \mid \theta) := p_\theta^{\sum_j s_j}(1 - p_\theta)^{n-\sum_j s_j}\left[\pi^{\sum_j \eta_j}(1 - \pi)^{n-\sum_j \eta_j}\right], \tag{B.1}$$

Arun G. Chandrasekhar: arungc@stanford.edu
Horacio Larreguy: hlarreguy@fas.harvard.edu
Juan Pablo Xandri: jxandri@princeton.edu

 https://doi.org/10.3982/ECTA14407

where $\pi := \mathbb{P}(\eta_i = 1)$. The set of all type configurations is denoted by $H = \{0, 1\}^n$, and in this model, $\Omega := S \times H = \{0, 1\}^n \times \{0, 1\}^n$.

Let $p_\theta = \mathbb{P}(s_i = 1 \mid \theta)$. In our experiment, $p_\theta = 5/7$ if $\theta = 1$ and $p_\theta = 2/7$ if $\theta = 0$.

B.3. *Recursive Definition of Information and Action Functions*

The function $P_{i,t}(\omega) \subseteq \Omega$ denotes the information set of agent i at round t, under state ω. At round $t = 1$, agent i only observes ω_i out of state ω, and hence, her information set is

$$P_{i,1}(\omega) := \{\omega' \in \Omega : \omega'_i = \omega_i\}. \tag{B.2}$$

In words, the possible states of the world are those compatible with the private information she has received (which includes her signal $s_i \in \{0, 1\}$ and her type).

Based on this information, all agents initially choose to match their signal; that is,

$$a_{i,1}(\omega) := s_i. \tag{B.3}$$

For $t > 1$, we compute $P_{i,t}(\omega)$ and $\mathbf{a}_{i,t}(\omega)$ inductively, for each $\omega \in \Omega$. In our experimental setup, at round t, agent i observes all the actions taken by her neighbors $j \in N(i)$ (including herself) up to $s = t - 1$. Therefore, the states of the world that are consistent with agent i's observations (i.e., her information set) are

$$P_{i,t}(\omega) := \{\omega' \in \Omega : \omega'_i = \omega_i \text{ and } a_{j,s}(\omega') = a_{j,s}(\omega) \text{ for all } j \in N(i), s \leq t - 1\}. \tag{B.4}$$

Clearly, we have $P_{i,t}(\omega) \subseteq P_{i,t-1}(\omega)$ for all $i, \omega \in \Omega$ (i.e., $P_{i,t}(\cdot)$ corresponds to a filtration).[1] The round-t action function $a_{i,t}(\omega)$ is then given by

$$a_{i,t}(\omega) := \begin{cases} \mathbf{1}\left\{I_{i,t}(\omega) > \frac{1}{2}\right\} & \text{if } I_{i,t}(\omega) \neq \frac{1}{2}, \\ a_{i,t-1}(\omega) & \text{if } I_{i,t}(\omega) = \frac{1}{2}, \end{cases} \tag{B.5}$$

where $I_{i,t}(\omega)$ is the "belief index" at state ω, which depends on the agents' type. If agent i is Bayesian (i.e., under the complete information model, or if $\eta_i = 1$ in the incomplete information model), then $I_{i,t}(\omega) := \mathbb{P}(\theta = 1 \mid P_{i,t}(\omega))$, which is calculated using Bayes rule conditioning on the event $P_{i,t}(\omega)$:

$$\mathbb{P}(\theta = 1 \mid P_{i,t}(\omega)) := \frac{\sum_{\omega' \in P_{i,t}(\omega)} \rho(\omega' \mid \theta = 1)}{\sum_{\omega' \in P_{i,t}(\omega)} \rho(\omega' \mid \theta = 1) + \rho(\omega' \mid \theta = 0)}. \tag{B.6}$$

When i is not Bayesian at ω, then $I_{i,t}(\omega) := \sum_{j=1}^{n} T_{ij} a_{j,t-1}(\omega)$, where $[T_{ij}]_{ij}$ are the DeGroot weights.

[1]We can also define $P_{i,t}$ recursively, starting at $P_{i,1}$ as in (B.2), and for $t \geq 1$ let $P_{i,t}(\omega) := P_{i,t-1}(\omega) \cap \{\omega' \in \Omega : a_{j,t-1}(\omega') = a_{j,t-1}(\omega') \text{ for all } j \in N(i)\}$.

B.4. *Numerical Implementation*

The algorithm used is based on the inductive step defined above, calculating iteratively the objects $P_{i,t}(\omega)$ and $a_{i,t}(\omega)$ for all i, t and ω.

ALGORITHM 1—Bayesian Learning Algorithm:
Inputs:
1. An n-person network $G = (V, E)$ with adjacency matrix $A_{n\times n}$;
2. A row stochastic matrix of DeGroot weights $T_{n\times n}$; and
3. Probability $\pi \in [0, 1]$.

Output: Information and action functions $P_{i,t}(\omega)$ and $a_{i,t}(\omega)$.
Step 1: Initialize algorithm by defining:
1. State space $\Omega = S \times H = \{\omega = (s, \eta)$ where $s \in S := \{0, 1\}^n$, $\eta \in H := \{0, 1\}^n\}$;
2. Measures $\rho(\omega \mid \theta) = \rho(s, \eta \mid \theta)$ according to (B.1), for $\theta \in \{0, 1\}$; and
3. Information functions $P_{i,t}(\omega)$ and actions $a_{i,t}(\omega)$ according to (B.2) and (B.3) for all $i = 1, \dots n$ and $\omega \in \Omega$.

Step $t > 1$: Given $(P_{i,s}(\omega), a_{i,s}(\omega))_{i=1,\dots,n,s=1,\dots,t-1,\omega\in\Omega}$, calculate $P_{i,t}(\omega)$ and $a_{i,t}(\omega)$ for all i and $\omega \in \Omega$ according to (B.4) and (B.5), where $I_{i,t}(\omega) = \mathbb{P}(\theta = 1 \mid P_{i,t}(\omega))$ if $\eta_i = 1$ and $I_{i,t}(\omega) = \sum_j T_{ij} a_{j,t}(\omega)$ if $\eta_i = 0$.

It is worth noting that an alternative way of modeling the knowledge structure is by including the true state θ in the description of the state of the world; that is, define $\omega = (\theta, s)$ in the complete information case, and $\omega = (\theta, s, \eta)$ in the incomplete information case, which would need the definition of just one common prior $\rho(\omega)$, instead of having to define it conditional on θ. While this would perhaps be a better fit for most epistemic models, the description of the algorithm is slightly easier in our model, given the fact that $\omega = s$ in the complete information model and $\omega = (s, \eta)$ in the incomplete information models are, respectively, sufficient statistics for the actions sequence of players, since θ is never in any information set of any of the players, significantly reducing the relevant state space. In fact, these are the minimal state spaces we can consider, exactly because of sufficiency.

APPENDIX C: IMPLICATIONS FOR REAL-WORLD NETWORKS

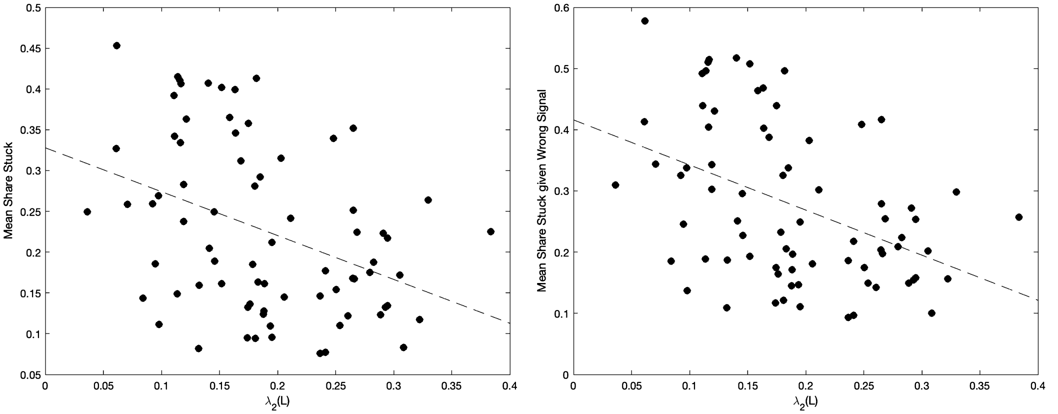
The above results show that whether asymptotic efficiency is reached or not depends on the structure of networks in question. In this section, we explore real-world network data to assess whether the problems due to coarse DeGroot learning might be a concern in real-world network settings.

We consider data from Banerjee, Chandrasekhar, Duflo, and Jackson (2019) Wave 2 sample consisting of detailed network data in 75 villages in Karnataka, India. We use graphs constructed from the links through which information is transmitted between households in the networks.

We use the results in Section 2.4.2 and Corollary 1. For every graph G in the sample, we compute the second eigenvalue of the Laplacian: $\lambda_2(L(G))$. Recall that if $\lambda_2(L(G)) > \frac{1}{2}$, then the graph cannot have any clans.[2]

We then simulate a learning model. We assume every agent is DeGroot operating in our coarse communication environment where agents can only communicate their best

[2]We use the bound, as counting the number of clans is an NP-hard problem.



(a) Share stuck versus $\lambda_2(L(G))$

(b) Share stuck of those with initial incorrect signal $1-\theta$ versus $\lambda_2(L(G))$

FIGURE 4.—Share of households stuck in each village plotted against $\lambda_2(L(G))$ which bounds the conductance of the graph. Larger values of $\lambda_2(G)$ correspond to greater expansiveness. We present results from 200 simulations per village, with $p = 0.6$ and $T = 200$.

guesses in every period. This is motivated both by the fact that we have found a very low share of Bayesian agents in our experiment in the Indian village context but also by the fact that simulating the incomplete information model is an NP-hard problem and we have a large number of agents, rendering this infeasible. The coarse DeGroot model here sets $p = 0.6$ and we run 200 simulations per village.

Figure 4 presents the results. First, observe that no value of $\lambda_2(L(G))$ exceeds 0.4, let alone $\frac{\sqrt{2}}{2}$, and so every village can have at least a clan. In fact, the values of $\lambda_2(L(G))$ can be quite low, and while this does not guarantee clan presence of course, it is suggestive.

Second, we find that the share stuck is high. Panel A shows that the fraction of villagers stuck can range as high as over 45% for villages with very low expansiveness, and, across the range of expansiveness that we see in the data, the average share of nodes stuck remains between 20% and 25%. Importantly, the share of villagers stuck is decreasing in the expansiveness of the village network.

Third, we look at the share of agents stuck among those who initially received an incorrect signal. This directly measures the share of agents who failed to learn. We see that this can be up to nearly a 60% share stuck, and the average is around 30%. As before, the share of villagers that fail to learn is decreasing in the expansiveness of the village network.

Taken together, the results suggest that real-world networks have significant clan presence. Furthermore, nodes have a large propensity to fall into misinformation traps, and especially so when village networks have low expansiveness.

APPENDIX D: CONSISTENCY OF STRUCTURAL ESTIMATION

D.1. *Setup*

There are V villages, each with n individuals who are arranged in a network. Our asymptotic sequence will take $V \to \infty$.[3]

[3]In what follows, we use the terminology of our experiment in India but we could just as well have v index session with a total of V sessions.

Every network of n individuals will play a learning game as follows. Each of n individuals has a type (Bayesian or DeGroot), so $\eta_i \in \{\mathrm{B}, \mathrm{D}\}$. This type is drawn i.i.d. with probability $\pi = \mathbb{P}(\eta_i = \mathrm{B})$ before the start of the game. This is commonly known by all agents.[4]

Our goal is to estimate π from the data generated in our experiment.

At time 0, there is a vector of binary signals $s = (s_1, \dots, s_n)$ drawn i.i.d. conditional on the state ($\theta \in \{0, 1\}$). Agents are trying to learn θ. The signals are distributed

$$s_i = \begin{cases} \theta & \text{with probability } p = 5/7, \\ 1 - \theta & \text{with probability } 1 - p. \end{cases}$$

The agents are engaging in a learning task wherein in every period, given the history, they take their best guess about the state of the world (1 or 0). Agents observe all their own previous actions as well as those of their network neighbors from prior periods. The type space here is therefore the cross between agent type (Bayes or DeGroot) and signal endowment. Let $\omega = (\eta, s)$. Note that the most information an agent could theoretically use to assess the value of θ is $(s_1, \dots, s_n)$.

In every period τ, there is an action taken by i, $a^\star_{i\tau}$. The type of the agent and the history determine the action. Given a history $A^{t-1} = (a_{i\tau})_{i=1,\tau=1}^{n,t-1}$, there is a prescribed action under the model of behavior which can depend on the agent's type η_i, the history of observed play, and the prior probability that an agent is Bayesian:[5]

$$a^\star_{it}\big(A^{t-1}; \eta, \pi\big).$$

Then, given the prescribed option, the observed data for the econometrician (and agents) are

$$a_{it} = \begin{cases} a^\star_{it} & \text{with probability } 1 - \epsilon, \\ 1 - a^\star_{it} & \text{with probability } \epsilon, \end{cases}$$

for any $t = 2, \dots, T$. Note that the history is the history of observed actions, which can differ from the prescribed action. We assume that this mistake is not internalized by agents. For the network-level approach, we can take any $T \geq 3$, whereas for the individual-level approach, assume $T = 3$.[6]

The matrix $A^T_v = [a_{it,v}]$ is the data set for a given village v. Suppressing v until it is needed, the likelihood is

$$\mathcal{L}\big(\pi, \epsilon; A^T\big) = \mathbb{P}\big(A^T | \pi, \epsilon\big) = \mathbb{P}\big(a_T | A^{T-1}, \pi, \epsilon\big) \cdot \mathbb{P}\big(a_{T-1} | A^{T-2}, \pi, \epsilon\big) \cdots \mathbb{P}(a_1 | \pi, \epsilon).$$

Notice that $\mathbb{P}(a_1|\pi)$ and $\mathbb{P}(a_2|\pi)$ are both independent of π, because they are independent of η: in period 1, every agent plays their signal, and in period 2, every agent plays the majority (subject to a fixed tie breaking rule).

[4]Note that DeGroot agents are mechanical and do not use this information so it really matters for Bayesian agents.

[5]In the network-level approach, this is $a^\star_{it}(A^{t-1}; \eta, \pi) = a^\star_{it}((a_{i0})_{i=1}^n; \eta, \pi)$, and in the individual-level approach, this is $a^\star_{it}(A^{t-1}; \eta, \pi) = a^\star_{it}(A_i^{t-1}; \eta_i, \pi)$.

[6]As discussed in the text, we say that the model is defined until the first t at which some i encounters a zero probability information set, which we denoted as $T^\star$. This cannot happen at $T = 3$, so for simplicity, consider this to be the case which defines a valid sample from which to construct a consistent estimator.

D.2. *Estimation of* ϵ

Observe that for any graph v for any node i such that the majority of their neighbors and their own signal is unique, both the Bayes and DeGroot models, irrespective of π, prescribe the majority. Therefore, recalling that $N_i^* = \{j : g_{ij} = 1\} \cup \{i\}$,

$$\widehat{\epsilon} := \frac{\sum_v \sum_j \mathbf{1}\{a_{j2} \neq \text{majority}(a_{j1} : j \in N_i^*)\} \cdot \mathbf{1}\{\text{unique majority}(a_{j1} : j \in N_i^*)\}}{\sum_v \sum_j \mathbf{1}\{\text{unique majority}(a_{j1} : j \in N_i^*)\}}.$$

By standard arguments, $\widehat{\epsilon} \longrightarrow_p \epsilon$ and $\sqrt{V}\frac{\widehat{\epsilon}-\epsilon}{\widehat{\epsilon}(1-\widehat{\epsilon})} \rightsquigarrow \mathcal{N}(0,1)$, since this is just a set of Bernoulli trials.

D.3. *Estimation of* π

For simplicity of exposition, we take ϵ as known, though in practice this will be a two-step estimator.

We can now consider

$$\mathcal{L}(\pi; A^T, \epsilon) = \prod_{t=3}^{T} \mathbb{P}(a_t | A^{t-1}, \pi, \epsilon).$$

It is useful to expand the term noting that $A^1 = s$,

$$\mathbb{P}(a_t | A^{t-1}, \pi, \epsilon) = \prod_{i=1}^{n} \mathbb{P}(a_{it} | A^{t-1}, \pi, \epsilon) = \prod_{i=1}^{n} \sum_{\eta} \mathbb{P}(a_{it} | A^{t-1}, \eta, \pi, \epsilon) \mathbb{P}(\eta | \pi)$$

by independence and then

$$\begin{aligned}\mathbb{P}(a_{it} | A^{t-1}, \eta, \pi, \epsilon) &= \mathbf{1}\{a_{it} = a_{it}^\star\} \mathbb{P}(a_{it} = a_{it}^\star | a_{it}^\star(A^{t-1}), A^{t-1}, \eta, \pi, \epsilon) \mathbb{P}(a_{it}^\star | A^{t-1}, \eta, \pi, \epsilon) \\ &\quad + \mathbf{1}\{a_{it} \neq a_{it}^\star\} \mathbb{P}(a_{it} \neq a_{it}^\star | a_{it}^\star(A^{t-1}), A^{t-1}, \eta, \pi, \epsilon) \mathbb{P}(a_{it}^\star | A^{t-1}, \eta, \pi, \epsilon) \\ &= \mathbf{1}\{a_{it} = a_{it}^\star(A^{t-1}; \eta, \pi)\} \cdot (1-\epsilon) + \mathbf{1}\{a_{it} \neq a_{it}^\star(A^{t-1}; \eta, \pi)\} \cdot \epsilon.\end{aligned}$$

Let $x_{it} = \mathbf{1}\{a_{it} = a_{it}^\star(A^{t-1}; \eta, \pi)\}$, which computes whether the observed action matches that which was prescribed by the model given the history, type vector, and parameter value. So,[7]

$$\mathbb{P}(a_t | A^{t-1}, \pi, \epsilon) = \prod_{i=1}^{n} \sum_{\eta} (1-\epsilon)^{x_{it}} \epsilon^{1-x_{it}} \mathbb{P}(\eta | \pi).$$

[7]It is worth noting that if we could pass the logarithm, then this is

$$\sum_{t=3}^{T} \sum_{i=1}^{n} \sum_{\eta} \left(x_{it}[A^{t-1}; \eta, \pi] \log(1-\epsilon) + (1 - x_{it}[A^{t-1}; \eta, \pi]) \log \epsilon\right) \cdot \mathbb{P}(\eta | \pi),$$

and for small ϵ this is a reweighted divergence.

Recalling $x_{it} = x_{it}[A^{t-1}; \eta, \pi]$, we can consider the log likelihood for a given v,

$$\ell_v(\pi; A^T, \epsilon) = \sum_{t=3}^{T}\sum_{i=1}^{n} \log\left\{\sum_{\eta}(1-\epsilon)^{x_{it}[A^{t-1};\eta,\pi]}\epsilon^{1-x_{it}[A^{t-1};\eta,\pi]} \cdot \mathbb{P}(\eta|\pi)\right\}.$$

And, since villages are independent, the full log likelihood is

$$\ell(\pi; A^T, \epsilon) = \sum_{v=1}^{V}\sum_{t=3}^{T}\sum_{i=1}^{n} \log\left\{\sum_{\eta}(1-\epsilon)^{x_{it}[A^{t-1};\eta,\pi]}\epsilon^{1-x_{it}[A^{t-1};\eta,\pi]} \cdot \mathbb{P}(\eta|\pi)\right\}.$$

Then, let us define

$$\log f(A_v|\pi) := \sum_{t=3}^{T}\sum_{i=1}^{n} \log\left\{\sum_{\eta}(1-\epsilon)^{x_{it}[A^{t-1};\eta,\pi]}\epsilon^{1-x_{it}[A^{t-1};\eta,\pi]} \cdot \mathbb{P}(\eta|\pi)\right\}.$$

D.4. *Consistency of* $\widehat{\pi}$

In what follows, for simplicity assume that ϵ is known. Standard arguments will allow us to extend the below to joint consistency. Now we demonstrate that

$$\widehat{\pi} := \operatorname*{argmax}_{\pi} \widehat{Q}_V(\pi) := \operatorname*{argmax}_{\pi} \frac{1}{V}\sum_{v} \log f(A_v|\pi)$$

generates a consistent estimator of π. The limit object is $Q_0(\pi) := \mathbb{E}[\log f(A_v|\pi)]$.

PROPOSITION D.1: *Under the above assumptions,* $\widehat{\pi} \to_p \pi_0$ *as* $V \to \infty$.

PROOF: This serves only as a sketch, but follows the arguments of Theorem 2.1 in Newey and McFadden (1994). First, by the arguments of Lemma 2.2 in Newey and McFadden (1994), there is a unique maximum of $Q_0(\pi)$ at the true value π_0, since

$$Q_0(\pi_0) - Q_0(\pi) = \mathbb{E}_{\pi_0}\left[-\log\frac{f(A_v|\pi)}{f(A_v|\pi_0)}\right] > -\log \mathbb{E}_{\pi_0}\left[\frac{f(A_v|\pi)}{f(A_v|\pi_0)}\right] = 0$$

by the information inequality.

Second, we can take compactness as given since $\pi \in [0, 1]$.

Third, the objective is continuous in π with probability 1. To see this, notice that $\mathbb{P}(\eta|\pi)$ is continuous in the parameter since it consists of binomial draws with probability π. Further, $x_{it}[A^{t-1}; \eta, \pi]$ is continuous a.e. in π because it is a step function.

Last, we need to establish that the finite sample objective function converges uniformly in probability to its limit. To show that, we argue that $\widehat{Q}_V(\pi) := \frac{1}{V}\sum_v \log f(A_v|\pi)$ is stochastically equicontinuous and converges pointwise. Pointwise convergence is self-evident. To show stochastic equicontinuity, we check the Holder inequality which is a sufficient condition. Consider any two π so we that have

$$\log\left\{\sum_\eta (1-\epsilon)^{x_{it}[A^{t-1};\eta,\pi]}\epsilon^{1-x_{it}[A^{t-1};\eta,\pi]}\cdot\mathbb{P}(\eta|\pi)\right\} \\ -\log\left\{\sum_\eta (1-\epsilon)^{x_{it}[A^{t-1};\eta,\pi']}\epsilon^{1-x_{it}[A^{t-1};\eta,\pi']}\cdot\mathbb{P}\big(\eta|\pi'\big)\right\},$$

which is

$$\left|\log\left\{\frac{\sum_\eta (1-\epsilon)^{x_{it}[A^{t-1};\eta,\pi]}\epsilon^{1-x_{it}[A^{t-1};\eta,\pi]}\cdot\mathbb{P}(\eta|\pi)}{\sum_\eta (1-\epsilon)^{x_{it}[A^{t-1};\eta,\pi']}\epsilon^{1-x_{it}[A^{t-1};\eta,\pi']}\cdot\mathbb{P}\big(\eta|\pi'\big)}\right\}\right| \\ \leq 0+\left|\pi-\pi'\right|\left|\frac{\frac{\partial}{\partial\pi}\left\{\sum_\eta (1-\epsilon)^{x_{it}[A^{t-1};\eta,\pi]}\epsilon^{1-x_{it}[A^{t-1};\eta,\pi]}\cdot\mathbb{P}(\eta|\pi)\right\}}{\sum_\eta (1-\epsilon)^{x_{it}[A^{t-1};\eta,\pi]}\epsilon^{1-x_{it}[A^{t-1};\eta,\pi]}\cdot\mathbb{P}(\eta|\pi)}\right|.$$

Then,

$$\frac{\partial}{\partial\pi}\left\{\sum_\eta (1-\epsilon)^{x_{it}[A^{t-1};\eta,\pi]}\epsilon^{1-x_{it}[A^{t-1};\eta,\pi]}\cdot\mathbb{P}(\eta|\pi)\right\} \\ =\sum_\eta\left[\frac{\partial}{\partial\pi}(1-\epsilon)^{x_{it}[A^{t-1};\eta,\pi]}\right]\epsilon^{1-x_{it}[A^{t-1};\eta,\pi]}\cdot\mathbb{P}(\eta|\pi) \\ +\left\{\sum_\eta (1-\epsilon)^{x_{it}[A^{t-1};\eta,\pi]}\left[\frac{\partial}{\partial\pi}\epsilon^{1-x_{it}[A^{t-1};\eta,\pi]}\right]\cdot\mathbb{P}(\eta|\pi)\right\} \\ +\sum_\eta (1-\epsilon)^{x_{it}[A^{t-1};\eta,\pi]}\epsilon^{1-x_{it}[A^{t-1};\eta,\pi]}\cdot\frac{\frac{\partial}{\partial\pi}\mathbb{P}(\eta|\pi)}{\mathbb{P}(\eta|\pi)}\cdot\mathbb{P}(\eta|\pi).$$

Then, the first two terms are 0 a.e. and therefore certainly bounded by 1, and the final term is just

$$\mathbb{E}\big[(1-\epsilon)^{x_{it}[A^{t-1};\eta,\pi]}\epsilon^{1-x_{it}[A^{t-1};\eta,\pi]}\cdot\text{Score}(\eta|\pi)\big]\leq n2^n,$$

which is a constant since n is fixed. This follows from

$$\mathbb{E}\big[(1-\epsilon)^{x_{it}[A^{t-1};\eta,\pi]}\epsilon^{1-x_{it}[A^{t-1};\eta,\pi]}\cdot\text{Score}(\eta|\pi)\big] \\ =\sum_\eta (1-\epsilon)^{x_{it}[A^{t-1};\eta,\pi]}\epsilon^{1-x_{it}[A^{t-1};\eta,\pi]}\cdot\big(z\pi^{z-1}(1-\pi)^{n-z}+(n-z)\pi^z(1-\pi)^{n-z-1}\big)\leq n2^n.$$

So, we have a parameter-independent bound that satisfies the Holder condition. *Q.E.D.*

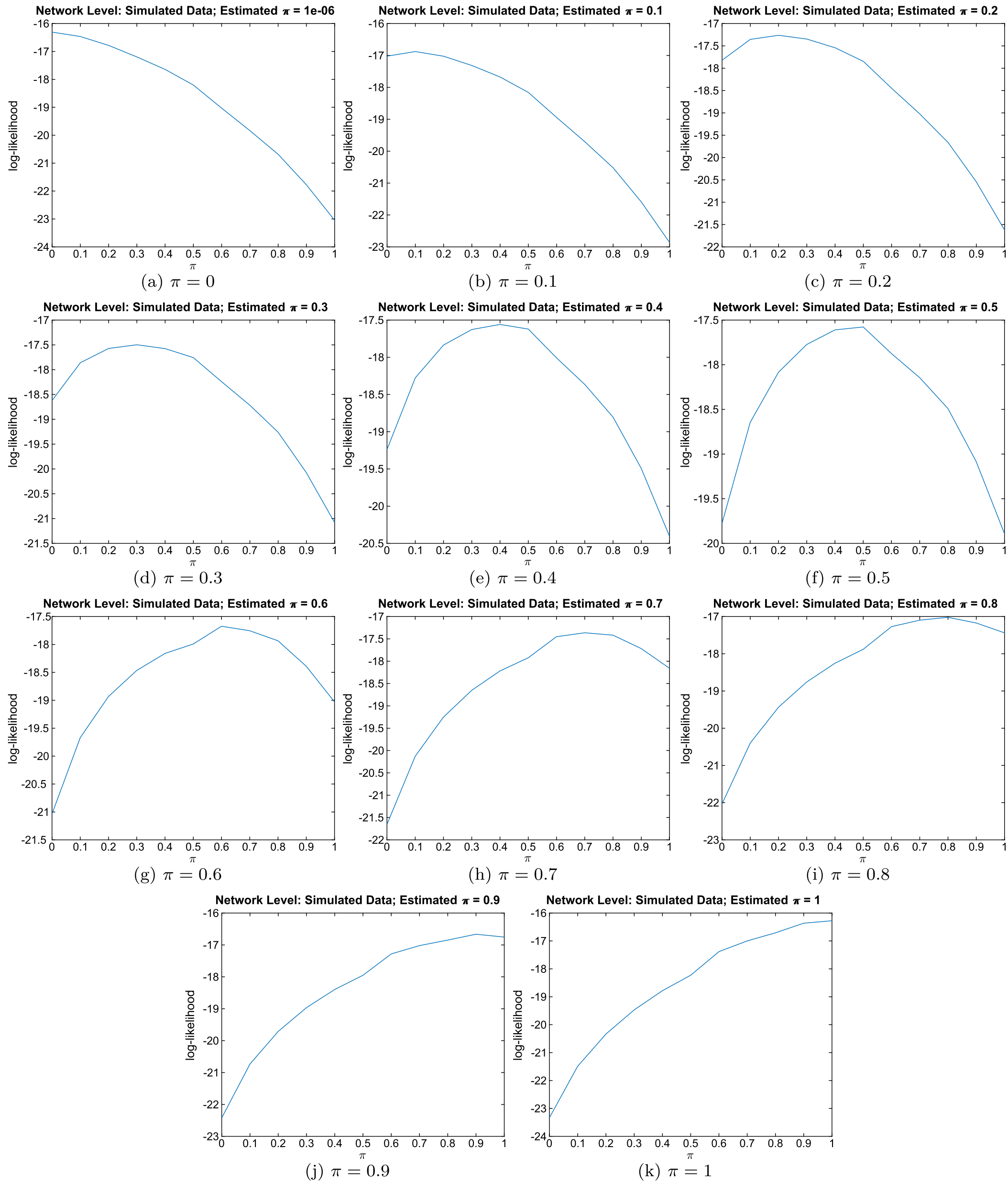
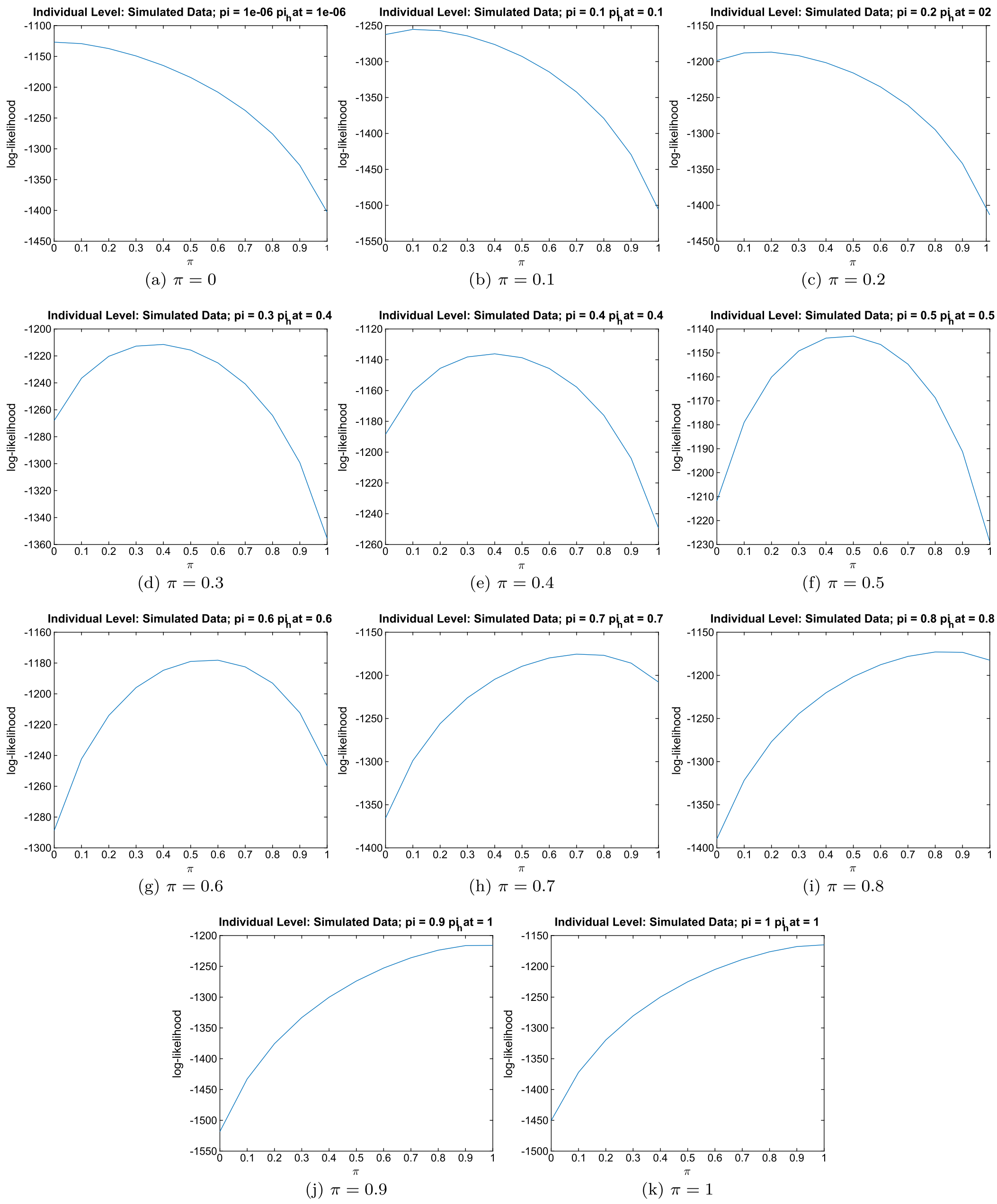


FIGURE 5.—Objective functions for MLEs of π for simulated data generated at various π (network-level estimation).

D.5. *Simulations*

We now show that, if we generate data with parameters (π, ϵ), we can use our estimator to recover both parameters. Figures 5 and 6 show the results. We have generated data with $\epsilon = 0.13$ (the level estimated in both data sets) and $\pi \in \{0, 0.1, \dots, 0.9, 1\}$. We show that across the board, the objective function is maximized exactly at the right parameter value in both the network- and individual-level estimations in Figures 5 and 6, respectively.



(a) $\pi = 0$ (b) $\pi = 0.1$ (c) $\pi = 0.2$

(d) $\pi = 0.3$ (e) $\pi = 0.4$ (f) $\pi = 0.5$

(g) $\pi = 0.6$ (h) $\pi = 0.7$ (i) $\pi = 0.8$

(j) $\pi = 0.9$ (k) $\pi = 1$

FIGURE 6.—Objective functions for MLEs of π for simulated data generated at various π (individual-level estimation).

REFERENCES

AUMANN, R. J. (1976): "Agreeing to Disagree," *The Annals of Statistics*, 1236–1239. [1]

BANERJEE, A., A. G. CHANDRASEKHAR, E. DUFLO, AND M. O. JACKSON (2019): "Using Gossips to Spread Information: Theory and Evidence From Two Randomized Controlled Trials," *The Review of Economic Studies*. [3]

GEANAKOPLOS, J. (1994): “Common Knowledge,” *Handbook of Game Theory with Economic Applications*, 2, 1437–1496. [1]
NEWEY, W. K. AND D. MCFADDEN (1994): “Large Sample Estimation and Hypothesis Testing,” in *Handbook of Econometrics*, Vol. 4. Elsevier, 2111–2245. [7]
OSBORNE, M. J., AND A. RUBINSTEIN (1994): *A Course in Game Theory*. Cambrdige, MA: MIT Press. [1]

Co-editor Fabrizio Zilibotti handled this manuscript.

Manuscript received 16 May, 2016; final version accepted 13 May, 2019; available online 14 August, 2019.