Adoption with Social Learning and Network Externalities*

Marcel Fafchamps — Stanford University

Måns Söderbom — University of Gothenburg

Monique van den Boogaart — Uber, San Fransisco

August 2020

Abstract

Based on a large administrate dataset covering the universe of phone calls and airtime transfers in a country over a four year period, we examine the pattern of adoption of airtime transfers over time. We start by documenting strong positive effects of increased usage of the new airtime transfer service by social neighbors on own adoption probability. We narrow down the possible sources of this effect by distinguishing between network externalities and social learning. Within social learning, we differentiate between learning about existence of the new product from learning about its quality or usefulness. We find robust evidence suggestive of social learning both for the existence and the quality of the product. In contrast, we find that the effect of increased usage by social neighbors on own usage turns negative after first adoption, suggesting that airtime transfers are strategic substitutes among network neighbors. We conclude that learning about existence and quality are important mechanisms, while strategic complementarities are not.

*We benefitted from comments from Jeremy Folz, Simon Quinn, Joshua Blumenstock, Paul Frijters and seminar participants at Oxford University, Monash University, the University of Nottingham and the University of Gothenburg, as well as conference participants at the 2016 CSAE Conference, Oxford and at the UNU-WIDER Nordic Conference in Development Economics, Helsinki. We thank Nathan Eagle for making the data available for this study.

1. Introduction

The introduction of IT technology has revolutionized the way many products and services are distributed. This is also true in less developed countries where mobile phones have opened new avenues for the diffusion of information and the adoption of new technologies and services. Examples include: market price information (e.g., Jensen 2007, Aker and Fafchamps 2015, Fafchamps and Minten 2012); agricultural extension services (e.g., Cole and Fernando 2016.); health information; mobile banking (e.g., Jack and Suri 2014); and political elections (e.g., Aker, Collier and Vicente 2017). The fact that all these applications are based on a platform – the mobile phone – originally designed for social communication leaves much room for possible social network effects in adoption and usage.

In this paper we examine the (first) adoption of an airtime transfer service in Rwanda using a large administrative dataset from a monopolistic telecommunication operator.[1] Peer-to-peer transfers of airtime between phone users is a predecessor of the introduction of mobile banking. The only difference is that, when mobile banking is in place, users can redeem airtime for cash from participating agents. The pattern of diffusion of airtime transfers across phone users can therefore be taken as indicative of the likely diffusion of mobile money and other phone-based services. It is also potentially informative about other diffusion processes on social networks.

It has often been observed that the adoption of new products and services, and other behavioral changes, diffuse along social networks (Young 1999, 2009; Jackson and Yariv 2005; Björkegren 2019). What is less clear is why. This paper aims to throw some light on this issue. There are many possible reasons why adoption may spread along social networks. One is that some individuals get to know of a new product.[2] People talk about new products with others

[1] The outcome of interest in the present paper is *first* adoption, (i.e., the first time an individual uses the technology actively). Subsequent to (first) adoption, there is continued useage and non-usage of the technology.

[2] To keep things straightforward, we speak throughout of the adoption of a new product, but the same principles

in their network of acquaintances, so that information about the existence of the new product spreads through social learning (Mobius and Rosenblat 2014). A proportion of those informed of the new product adopt it, and since adoption requires knowing about the new product, adoption is observed to diffuse by social contact, in a way similar to the way an epidemic spreads in a population.

Other forms of social learning are possible as well. For instance, people may learn about the hidden qualities of a new product through usage. The decision to adopt may depend on what people know of these hidden qualities, such as how useful or reliable the new product really is. If too little information is available, risk averse individuals refrain from adopting. It follows that, as people share information about hidden characteristics of the new product along social networks, adoption spreads. The main difference with the first type of social learning is that here more usage by social neighbors provides cumulative information that is valuable for the adoption decision, over and above simply knowing that the product exists.

Diffusion along social networks may also occur for reasons having nothing to do with social learning. One particular case is network externalities or, more precisely, strategic complementarities in adoption decisions (Saloner and Shepard 1995, Jackson and Yariv 2005; Vega-Redondo 2007). If adoption by my social neighbors increases my incentive to adopt, I am more likely to adopt following adoption by my neighbors. This mechanism may arise even when all agents have full information about the existence and qualities of the product, although it may be combined with social learning. The main difference with social learning is that network externalities do not wear off: they continue to reinforce adoption long after any hidden information about the new product would have been learned. Strategic complementarities may arise for many different reasons, some good – the usefulness of the product increases with more widespread usage –

generally apply to the adoption of a new service.

some bad – adoption protects me against some of the negative externalities generated by widespread usage. The canonical example of a strategic complementarity that arises from a negative externality is the installation of a burglar alarm: when I install an alarm, I initially displace crime towards neighbors, which raises their incentive to install a burglar alarm; in equilibrium, everyone incurs the cost of having a burglar alarm but it no longer serves as deterrent (Jackson 2009).

In this paper we seek to identify the respective roles of network externalities and social learning in the adoption of a new service offered to mobile phone users. We also seek to identify the relative importance of social learning about product existence vs. its hidden qualities. To do this, we rely on a large dataset that includes all phone calls made by mobile phone users of a large monopolistic provider in an entire country for a period of four years. While the dataset includes many observations, each observation contains a limited amount of information. We compensate for this to the best of what the data allows by including different types of fixed effects to capture unobserved heterogeneity. We find robust evidence suggestive of social learning both for the existence and the quality of the product. In contrast, we find that the effect of increased usage by social neighbors on own usage turn negative after first adoption, suggesting that airtime transfers are strategic substitutes among network neighbors.

This paper complements a large literature documenting the diffusion of new products and behaviors on social networks (e.g., Centola 2010, Ryan and Tucker 2012, Jack and Suri 2014). Our contribution to this literature is to decompose network effects into different components and to measuring the sign and magnitude of these components. We find that network effects need not be strategic complements, as is commonly assumed in the literature (e.g., Jackson and Yariv 2005, Vega-Redondo 2007). In contrast, we find evidence that networks play a role in the circulation of information. The information effects of social networks have been documented before (e.g., Granovetter 1995, Jensen 2007, Aker 2010, Aker and Fafchamps 2015), but the emphasis has been on the continued informational benefits that networks provide – a form of network externality. We find that, in the case of the diffusion of a new product, the effect of social networks on product adoption and usage are limited in time. These results suggest that network effects in diffusion are driven primarily by the spread of information about the existence and the characteristics of the new product.

The paper is organized as follows. We start in Section 2 by introducing the testing strategy. The conceptual framework behind it is detailed in Appendix A. The information available in the raw data is discussed in Section 3, together with a description of how we construct the variables used in our analysis. Empirical results are presented in Section 4. Section 5 concludes.

2. Testing strategy

The intuition behind our testing strategy can be summarized as follows – see the Appendix A for a formal treatment. Suppose that network effects arise solely due to social learning about the existence and usefulness of a new product. In this case, recent usage by network neighbors predicts first adoption by an individual i: usage by network neighbors generates information that can be passed onto i, thereby increasing the likelihood that i adopts the product too. If using the product conveys full information about its usefulness, after i has used once, recent usage by network neighbors should no longer predict i's own use.

Now suppose instead that network effects are entirely driven by strategic complementarities in usage. In this case, In this case, own usage will co-vary with neighbor usage after first adoption. This observation is the first basis of our testing strategy.

We also wish to distinguish between the two types of social learning: about the existence of a new product; and about the usefulness of the new product. To this effect, we note that existence is known to i as soon as one of i's neighbors reveals the product to i. The researcher observes a signal $M_{it} = 1$ if, at time t, individual i receives unambiguous information about the product's existence, even if i has never used the product yet; $M_{it} = 0$ otherwise. It is then possible to disentangle whether social learning is purely about existence or also about the usefulness of the product: when social learning is purely about product existence, once i has learned about the existence of the product, subsequent usage by network neighbors can no longer predict first adoption by i. In contrast, if social learning is about product quality, usage by network neighbors continues to predict i's first adoption because it accummulates information that can help i decide whether to adopt the product or not. The fact that the two learning models make different predictions makes it possible to test one against the other.

In Appendix A we present a model of social learning that formalizes the above intuition in a clear way. We then use it to derive empirical predictions that can be put to the test using the data at our disposal. It is important to realize that our testing strategy does not necessitate the type of causal identification that is required for impact evaluation. This is certainly true for strategic complementarities: we simply use the fact that, if usage generates strategic complementarities between network neighbors, usage will be correlated between them; in contrast, if strategic complementarities are absent, there is no reason to expect correlation in usage after social learning has ended. Nothing in this strategy relies on identifying who in the network 'causes' others to use the product – this would be futile anyway since causation is mutual.

For strategic learning, we rely on a purely mechanical form of transitivity: for i to receive information from network neighbors, these neighbors have to have been exposed to the information first. Nothing precludes i from actively seeking information from neighbors, or from inciting neighbors to experiment with the product: in each of these cases, i plays a part in the causation process. Our strategy simply relies on the fact that neighbors must first experiment with the product to be able to convey information that helps i decides whether to adopt or not. This is why adoption by network neighbors predicts first adoption by i, without necessarily 'causing' it.

Threats to identification come not from the lack of purely exogenous variation, but rather from the possible existence of confounds, that is, unobservables that predict usage by i and are correlated with usage by network neighbors – but have nothing to do with information transfer or strategic complementarities. The most likely confound is the presence of location- and time-specific advertizing and promotional campaigns. We deal with this issue by demonstrating the robustness of our findings to the inclusion of a wide range of fixed effects. Other possible confounds are discussed in the empirical section.

3. The data

The data we use to test our conceptual framework is administrative data on the usage and diffusion of a mobile phone service entitled ME2U. The service was introduced in Rwanda in September 2006 by the dominant mobile phone operator at the time. This service allows subscribers to transfer airtime to another subscriber at no cost. In February 2010 the operator added the possibility for subscribers to redeem airtime into cash, thereby formally introducing Mobile Money to the country. Over the period of our study, airtime could only be transferred to another subscriber.[3]

Our outcome of interest is the action of sending airtime to another subscriber. From the moment ME2U was introduced in the country, no action was required (e.g., registration or fee) for a subscriber to receive airtime. Hence observing that a subscriber receives airtime at a given point in time does not imply a voluntary decision to use the service. Nonetheless, it does un-

[3] There is some evidence that a small number of subscribers used airtime transfers to retail airtime that they bought in bulk at a discount. We discuss below how we deal with this possibility in our analysis.

ambiguously inform the recipient that peer-to-peer airtime transfers are in existence.[4] Knowing that it is possible to transfer airtime to someone else does not, by itself, confer full information about the usefulness of the service to a particular user. There are many attributes that subscribers may care about, such as easy-of-use, reliability, speed of execution, and protection against abuse or theft. Talking to other users about their experience sending airtime to others may therefore confer useful information to prospective users.

Network externalities may arise once the practice of transferring airtime across subscribers is sufficiently widespread in a particular social or geographical grouping. For instance, it would become easier to solicit small airtime transfers from friends and relatives in order to make a call or send a message, since they would be familiar with how to send airtime. It may also become possible to purchase or otherwise obtain airtime from strangers, e.g., on the bus home. Hence network effects may continue to manifest themselves even after a subscriber is fully acquainted with the service.

In the remainder of this section we begin by describing the source and structure of the data used in the analysis. Next we define all the variables used in this study and we explain how they are constructed. Last we present descriptive statistics on the variables used in the empirical section.

3.1. Data source

The data come from a large telecommunications operator. During the period of investigation, this operator enjoyed a quasi-monopoly on mobile phones in Rwanda. Access to the data was granted by Nathan Eagle through remote access to a Northeastern University computer server

[4] On receiving a transfer, the recipient would also receive a message indicating that their airtime balance had been updated. Hence, the recipient would have realized that they got a transfer and thus learned about product existence.

under conditions of strict confidentiality.[5] This is a large dataset comprising multiple computer-generated administrative files. We use two main bodies of data for our analysis: data on airtime transfers; and data on phone calls. The former are used to study first adoption and diffusion; the latter is used to define social networks. The data identifies subscribers through an anonymized identifier based on their phone number/SIM card. The same identifier is used throughout the data. We do not have information on the name or personal characteristics of individual users.[6]

The call data consist of an exhaustive log of all phone-based activity that occurred from the start of 2005 until the end of 2008. It provides information on the time, date, duration, receiver id and sender id for all phone calls made between 2005 and 2008. In total this dataset includes 50 billion transactions relative to approximately 1.5 million subscribers.

Data on calls is matched with a second dataset, from the same source, on usage of the airtime transfer service ME2U. This dataset consists of a log of all mobile-based airtime transfers that occurred between the introduction of the service in September 2006, and December 2008. For each transaction we observe the sender and receiver, the amount sent, and the time stamp (i.e., time and date).[7] We unfortunately do not have any information on the timing or geographical coverage of any promotional campaign that the mobile phone provider may have run. SMS received from the phone company (which may include promotional messages about ME2U) are not included in our data.

After its introduction in September 2006, ME2U usage increased steadily until the 1st of July 2008 when there is a break in the administrative data (see Figure 1).[8] To avoid spurious

[5] If one wishes to use this dataset, please contact Nathan Eagle at nathan@mit.edu.

[6] We cannot rule out that an individual may have multiple phone numbers, or that phone numbers may be transferred across users.

[7] The recipient of an airtime transfer receives a text message informing him/her that airtime has been transferred to their phone. The text message gives the amount transferred and the identity of the person who transferred it. To the best of our knowledge, no information is provided in the SMS on how the recipient can use the service to send airtime to others. But this information is available directly from the provider.

[8] Over the period of our study, there was no mobile money in Rwanda in the sense that is commonly understood, that is, the ability to pay for purchases at affiliated shops and the ability to redeem mobile money for cash from

inference, our analysis is based solely on airtime transfer data between September 2006 and July 2008. During this period, transferring airtime was free, and the number and amount of transfers that a user could send per day was not limited. Receiving or sending airtime could be done without the need to subscribe to the service – ME2U became available to all subscribers immediately after its introduction. The only requirement a user needed to fulfil to use the service is to have sufficient credit on his phone. When a user sends an airtime transfer, the amount sent is deducted from the user's airtime balance, the same balance that is used to make calls or send text messages. Topping up one's balance can be done by buying airtime vouchers from local shops and street vendors. Figure 2 shows how the proportions of adopters and active users in a given week developed over the sampling period.

Since all phone usage is prepaid, topping up by purchasing a voucher is a regular task for all subscribers, irrespective of whether they use ME2U or not. When a transfer is received, the amount is immediately added to the recipient's balance. This airtime can immediately be used to make calls, send airtime to other subscribers, or resell airtime to others. In February 2010 the operator introduced a system by which subscribers could redeem airtime against cash with dedicated agents. During the period covered by our data, such a system had not yet been introduced. We have information on the location of all cell towers in Rwanda during our period of analysis. We can link phone numbers to towers, and thus (crudly) track users' movements in the country. We use this information to control for location in the econometric analysis below.

a network of agents. At the time of our study, agents had not been recruited yet and shops were not signed up by the phone company to accept payment in airtime. This does not mean that people could not barter airtime. Some people figured out that since they could transfer airtime to anyone with a mobile phone, they could also purchase something – or solicit cash – from someone who needed airtime. Being a form of barter exchange, this would require finding someone who (1) happens to want airtime (i.e., 'coincidence of needs'); and (2) with enough trust to engage in a transaction. The introduction of mobile money moved airtime beyond its role of occasional and impractical barter currency.

3.2. Variable definition

Because the number of unique subscribers in the data is extremely large, we only use a randomly selected subset of 5,000 subscribers for our analysis of ME2U adoption and usage.[9] For these subscribers, we observe all their ME2U transfers between the introduction of the service in September 2006, and June 30th 2008. The end-date T is thus the end of June 2008. During our sample window, all transfers were peer-to-peer only

For the purpose of our analysis, we aggregate all phone usage information at the weekly level. This ensures that we take advantage of the detailed time information available in the data while keeping the size of the dataset manageable. For instance, ME2U usage by network neighbors is measured as the total number of neighbors who start using ME2U in a given week. All regressors are lagged by one period (i.e., week). This eliminates the risk of simultaneity bias since actual usage of ME2U by individual i in week t could not have caused usage by network neighbors in the previous week. Lagging regressors does not, of course, eliminate the risk of bias posed by unobserved factors. This issue is discussed more in detail in the empirical section.

We start by defining the dependent variable $y_{i,t}$, which is a dummy that takes value 1 if i has used ME2U in period t, and 0 otherwise. We consider a subscriber to be active from the week he receives or makes his first transaction – e.g., phone call, SMS, or ME2U transaction. This defines t_i, that is, the week from which i is at risk of adopting ME2U. The adoption date T_i for individual i is defined as the week at which the subscriber *sends* his first ME2U transfer. The reason for defining adoption in this way is that sending airtime requires an active decision while receiving a transfer is passive. In order to send a transfer, the subscriber may also need to invest time and effort, e.g., to top up his airtime balance or to learn how to make a transfer.

[9] Limiting our analysis to 5,000 subscribers offers the added advantage that it is extremely unlikely that the dataset used for analysis includes subscribers who belong to the neighborhood of the 5,000 selected subscribers. This further minimizes the risk of reverse causation – see below.

In contrast, the only requirement for a subscriber to receive a ME2U transfer is to have an activated phone number.

We construct the neighborhood of each subscriber as follows. We look in the data for all subscribers who, at some point between January 2005 and June 2008, have a phone contact with i. To be clear, this includes all subscribers in the data, not just those 5,000 subscribers randomly selected for the empirical analysis. We only use call data with a positive duration and from mobile to mobile phone – ME2U cannot be sent to a landline or to an international number.[10] We start from the dataset of all phone calls made between January 2005 and July 2008, and we identify the week in which i and j had their first phone-based contact. When i and j make the first phone call to each other, the network tie g_{ijt} switches from 0 to 1. For the purpose of the econometric analysis we assume that, once connected, i an j stay connected during the span of our analysis. The network ties are thus defined as:

$$g_{ijt} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ had their first phone-based contact in period } s \text{ with } s = t_i, ..., t \\ 0 & \text{otherwise} \end{cases} \tag{3.1}$$

The neighborhood of subscriber i in period t is the union of all the subscribers for which $g_{ijt} = 1$. That is:

$$N_{it}(g) = \{j : g_{ijt} = 1\} \tag{3.2}$$

Next, for each neighbor j of i we collate information on whether j made a ME2U transfer in week t, that is, whether $y_{jt} = 1$. We then construct a variable ΔA_{it} defined as the number of

[10] In addition, call data is missing for October 2006. This means that all variables derived from call data information are missing for that month.

neighbors of i who started sending airtime in week t. Accumulating ΔA_{it} over time yields the cumulative number of adopting neighbors A_{it} of i at week t.

In the conceptual section we introduced a variable M_{it} defined as a signal that i receives at time t that the new service exists. In the empirical implementation of the model, we set $M_{it} = 1$ in the first week that i receives a ME2U transfer. Variable m_{it} permanently switches to 1 once M_{it} has taken value 1. Finally, variable S_{it} is defined as the number of weeks since i started using his SIM-ID – that is, $S_{it} \equiv t - t_i$.

3.3. Descriptive statistics

Our sample consists 5,000 subscribers randomly selected for analysis. Table 1 provides descriptive statistics for the entire sample (column 1); for the subsample of observations before first adoption (column 2); and for the subsample of observations before i receives his first airtime transfer (column 3). The total number of observations is quite large, even when we limit our attention to 5,000 subscribers. We see that the neighborhood of each subscriber is large, as could be expected given our generous definition of social links. There is ample variation in ΔA_{it} and A_{it}, both in the entire sample and in the two subsamples.

4. Empirical results

The first regression model we estimate is eq. (7.12) in Appendix A, using only observations until first adoption. To eliminate the individual fixed effect α_i, we first difference the data.[11] The

[11] Fafchamps, Goyal and van der Leij (2010) estimate a model similar to regression (4.1) with fixed effects instead of taking first differences. They point out that the time structure of the dependent variable – a sequence of 0's ending with a single 1 – generates a spurious correlation between any trending regressor and the dependent variable, and recommend detrending all regressors prior to estimation in order to eliminate this bias. The time structure of the dependent variable in regression (4.1) is similar to theirs, but estimation in first difference de facto eliminates any linear trend in A_{it} and S_{it}. It remains that our findings could be affected by the presence of a quadratic time trend in A_{it}, which would translate in to a linear trend in ΔA_{it}. To investigate whether our results could be affected, we re-estimate regression (4.1) after detrending all first-differenced regressors. Results show absolutely no change in coefficient estimates and standard errors.

estimated model is a linear probability model of the form:

$$\Delta y_{it+1} = \alpha_1 + \alpha_2 \Delta A_{it} + \alpha_3 \Delta(S_{it}^2) + \alpha_4 \Delta(A_{it}^2) + \alpha_5 \Delta(S_{it} A_{it}) + controls + \Delta\varepsilon_{it+1} \tag{4.1}$$

where $\Delta x_t \equiv x_t - x_{t-1}$ by definition of notation and observations up to the first adoption are used, and *controls* is a set of control variables.[12,13] In all regressions below, standard errors are clustered at the district level.

There remains the perennial issue of possible endogeneity of A_{it}. There are several potential sources of endogeneity that we discuss in turn. The first potential source is reflection bias: i influences j and j influences i. To mitigate reflection bias, in the econometric analysis we use the lagged value of A_{it} instead of its contemporaneous value. While this may not be enough to eliminate the bias in general, it should be noted that in the special case where there is no strategic complementarity, there is also no reflection effect and thus no reflection bias. That is, the reflection effect is a pure magnification effect, which does not invalidate a test of the null hypothesis that there are no strategic complementarities.[14] A second potential source of endogeneity is correlated effects: an aggregate shock occurs that makes others and myself more likely to adopt at approximately the same time. An obvious example is a national marketing campaign targeting the entire country in a given month. To address this problem, we include in the *controls* vector separate dummies for each month in the study period (i.e. time dummies). Correlated shocks could also happen at the district level, e.g., because of a location-specific marketing campaign, or because the usefulness of ME2U increases in a district as a result of an exogenous shock such as flood or an earthquake (e.g., Blumenstock, Eagle and Fafchamps 2016).

[12] This is similar to a duration model with time-varying regressors estimated in discrete form. Instead of using a maximum likelihood estimator, we opt for a linear probability model so as to be able to remove the individual fixed effect by first-differencing the data. Given the long time series and likely persistence in errors, first differencing is to be preferred to fixed effects.

[13] By construction, $\Delta S_{it} = 1$, so this variable cannot be included as a regressor.

[14] A similar point was made by Moffitt (2003).

To address this concern, we also consider specifications with district-specific time dummies. We also control for correlated effects related to location: adoption patterns may vary across locations for reasons that have nothing to do with network mechanisms per se. For example, it could be that airtime transfers are more useful in urban than in rural settings. We thus include in the set of control variables tower dummy variables indicating the location of i in week t, as well as time-constant district dummies.

We recognize that our set of controls may not fully capture all potentially confounding factors. However, adding the control variables should at least mitigate the bias. Moreover, if the estimates of interest change little as a result of adding controls for observable factors, we might reasonably suppose that the omission of unobservable factors might not cause severe bias. This intuitive idea was formalized by Altonji, Elder and Tabler (2005), and developed further by Oster (2019) who shows how the size of the bias, under certain assumptions, can be inferred from coefficient and R-squared differences across models with different sets of control variables. Drawing on Oster's insights we will report results from specifications with differing sets of control variables, focusing on movements in the effects of interest. It should be noted that our specification (4.1) is nonlinear in the potentially endogenous variable, and there is no straightforward way of making Oster's formal framework applicable to a nonlinear model. However, we will report in an appendix estimates of the unobservable selection effect using Oster's approach and a linear specification. We will also consider results from a two-stage least squares estimator of (4.1).

Coefficient estimates of (4.1) are presented in Table 2. Specification [1] contains no control variables and serves as a benchmark. We see that α_2 is significantly positive, α_4 is significantly negative, and α_5 is significantly positive. Thus the network effects on the probability of first adoption are nonlinear, and dependent on the number of weeks since i started to use his SIM-ID.

Remember that, when social learning is about product existence, the relationship between first adoption and network effects should be strongly concave with respect to A_{it}. In contrast, when social learning is about product quality, this concavity need not be present and may even be reversed. Marginal effects $\partial \Pr / \partial A_{it}$ evaluated at various values of A_{it} are shown below the first-differences (FD) estimates in Table 2. We find that marginal effects are positive throughout, consistent with the presence of network effects. We observe a gradual fall in $\partial \Pr / \partial A_{it}$ as A_{it} increases, as suggested by the negative quadratic term coefficient α_4. This evidence is prima facie consistent with social learning about product existence, although the observed concavity is weaker than that predicted by equation (7.1) (see Appendix A).

Next, we add dummy variables for time, tower, and district to the model. This yields specification [2] in Table 2. The control variables have some explanatory power, as can be seen from the increase in the R-squared, but the coefficients and marginal effects of interest, and their significance levels, change only very marginally. We thus still obsverve a gradual fall a gradual fall in $\partial \Pr / \partial A_{it}$ as A_{it} increases, which is consistent with social learning about product existence. Expanding the set of control variables further, so as to allow for district specific time effects (specification [3]), leads to a small increase the R-squared but only results in trivial changes to the estimates of interest.

The fact the estimated coefficients and marginal effects shown in Table 2 are stable after the inclusion of observable controls, suggests that the bias caused by the omission of unobserved variables is limited. To probe this issue further, we adopt Oster's (2019) method for assessing robustness to selection on unobservables. Appendix B contains a short summary of Oster's method, and defines the relevant parameters; the reader is referred to Oster's paper for details about her approach. In order to use Oster's method, we must omit the terms $\Delta(A_{it}^2)$ and $\Delta(S_{it}A_{it})$ from the specification, yielding a model of the form $\Delta y_{it+1} = a_1 + a_2 \Delta A_{it} + a_3 \Delta(S_{it}^2) +$

controls + $\Delta\varepsilon_{it+1}$. In general, the OLS estimator of a_2 is not a consistent estimator of α_2, but the OLS estimate of a_2 may be a reasonable estimate of the average marginal effect of A_{it} on the likelihood of first adoption. However, we are primarily interested in whether the estimate of a_2 appears robust to selection on unobservables, as this sheds some light on whether the results in Table 2 are robust to selection.

Results are shown in Table B1 in Appendix B. The estimated effect of ΔA_{it} is equal to 0.0056 in the model without control variables and 0.00549 in the model with the full set of control variables included. These estimates are only very marginally lower than the average marginal effects of A_{it} based on the nonlinear model (see Table 2). Under the assumption that unobservable and observable factors are equally related to ΔA_{it} (which implies $\delta = 1$ in Oster's framework), and that $R_{\max}$ (i.e. the R-squared from a hypothetical regression of the dependent variable on the observable and unobservable determinants of the dependent variable) is twice as high as the R-squared for the model with only observed variables included, the bias-adjusted estimate of the effect of ΔA_{it} is equal to 0.0053 (Table B1, col. [3]). Setting $\delta = 2$, implying that the unobservable factors are twice as important determinants of ΔA_{it} as the observable factors, the bias-adjusted effect changes to 0.0051 (col. [4]). Clearly these bias-adjusted estimates differ only trivially from the OLS estimate that is obtained with the model with all controls included (Table B1, col. [2]). We find that the unobservables would have to be more than six times more important than the observables in determining ΔA_{it} for the bias-adjusted effect of ΔA_{it} to be equal to zero, given our assumed value for $R_{\max}$. We conclude from this part of the analysis that our results, for reasonable assumptions regarding the effects of unobservable variables (see Oster, 2019, for a discussion), appear robust.

In Table 3 we present results for a regression model that incorporates the signal m_{it} that ME2U exists (see equation 7.13 in Appendix A). Once again, we eliminate the individual fixed effect α_i by first-differencing the data. The estimated model is a LPM of the form:

$$\Delta y_{it+1} = \alpha_1 + \alpha_2 \Delta A_{it} + \alpha_3 \Delta(S_{it}^2) + \alpha_4 \Delta(A_{it}^2) + \alpha_5 \Delta(S_{it} A_{it}) + \beta_0 \Delta m_{it}$$
$$+\beta_1 \Delta(S_{it} m_{it}) + \beta_2 \Delta(A_{it} m_{it}) + \beta_3 \Delta(S_{it}^2 m_{it}) + \beta_4 \Delta(A_{it}^2 m_{it}) + \beta_5 \Delta(S_{it} A_{it} m_{it})$$
$$+controls + \Delta\varepsilon_{it+1} \tag{4.2}$$

where, as in (4.1), we only include observations up to the first adoption. The table also presents estimates of average marginal effects $\partial \Pr / \partial A_{it}$ evaluated at $m_{it} = 0$ and $m_{it} = 1$.

Effects of A_{it} remain significant throughout, although they are smaller when $m_{it} = 1$ than when $m_{it} = 0$. This is suggestive of a hybrid model in which social learning serves two purposes: circulating information about product existence, and about product quality. Given that network effects remain relatively large even after $m_{it} = 1$ suggests that, of the two, diffusing information about quality accounts for a significant share of social learning effects.[15] Similar to the results in Table 2, our estimates and significance levels of interest hardly change at all as a result of adding control variables to the regression. Results, available on request, from a robustness analysis using Oster's (2019) approach suggest very marginal bias adjustments for reasonable values of δ and $R_{\max}$.

We now seek to rule out that observed network effects on first adoption are purely due to network externalities, not to social learning. To this effect, we estimate a model that includes observations before and after first usage (see equation 7.14 in Appendix A). The model is estimated

[15] As pointed out by an anonymous referee, the frequency of contact could affect the rate of diffusion of information. We can shed some light on this issue empirically. If we add to the set of explanatory variables a variable measuring the total number of phone calls made by and received by i in a given week – a reasonable proxy for frequency of contact - this has a positive effect on first adoption (statistically significant if measured in $t+1$ i.e. contemporaneous with adoption; not quite statistically significant if lagged by one period). More importantly, we find that adding this variable proxying for frequency of contact does not affect the outcome of the test of interest: network effects remain positive and statistically significant even when $m_{it} = 1$ in this case.

in first difference to eliminate unobserved heterogeneity α_i, i.e., it is of the form:

$$\Delta y_{it+1} = \alpha_1 + \alpha_2 \Delta A_{it} + \alpha_3 \Delta(S_{it}^2) + \alpha_4 \Delta(A_{it}^2) + \alpha_5 \Delta(S_{it} A_{it}) + \gamma_0 \Delta z_{it}$$
$$+\gamma_1 \Delta(S_{it} z_{it}) + \gamma_2 \Delta(A_{it} z_{it}) + \gamma_3 \Delta(S_{it}^2 z_{it}) + \gamma_4 \Delta(A_{it}^2 z_{it}) + \gamma_5 \Delta(S_{it} A_{it} z_{it})$$
$$+controls + \Delta \varepsilon_{it+1} \tag{4.3}$$

where all observations are used and $z_{it} = 1$ if subscriber i has used ME2U before time t. Regression results and marginal effects are presented in Table 4. As should be, the α coefficient estimates are very similar to those reported in Table 2. We find that the marginal effect (which is estimated at sample means of the regressors) is much lower after first adoption, which confirms that social learning matters. What is less anticipated is that, after first adoption, network effects are on average negative, implying that, if anything, airtime transfers are strategic substitutes across network neighbors. This result is obtained without and with control variables included (Table 4, specifications [1]-l[3]).

To check the robustness of this finding, we re-estimate (4.3) with the set of control variables expanded to include a measure of the amount received by i. The logic is as follows. We begin by noting that ΔA_{it} captures airtime transfers made by i's network neighbors at time $t-1$. Some of these transfers may have been made to i. If i feels an obligation to reciprocate or pass on the transfers received, we expect to observe a mechanical positive correlation between Δy_{it+1} and ΔA_{it}. If, on the other hand, i receives transfers because he or she is at the receiving end of an altruistic relationship (e.g., a migrant sending remittances to his family, a husband sending airtime to his wife or children) and an airtime transfer is made when the recipient is in need of assistance, Δy_{it+1} and ΔA_{it} may be negatively correlated in the sense that the more i needs assistance, the more he or she receives airtime transfers, hence the larger ΔA_{it}. At the same time, the more i needs assistance, the less i can help others and hence the lower Δy_{it+1} is.

To investigate whether this is what drives the negative $\frac{\partial \Pr(y_{it+1}=1|z_{it}=1)}{\partial A_{it}}$ after first adoption, we reestimate (4.3) adding to the set of explanatory variables the amount of airtime transfers received at t. We have focused on one auxiliary mechanism namely the need for assistance. Consistent with our hypothesis, we find a negative and statistically significant effect of amount received on usage following first adoption (Table 4, specification [4]). However, this addition to the set of explanatory variables doesn't affect the estimated average marginal effect of neighbor usage, which remains negative and statistically significant. From this we conclude that the strategic substitution effect of network neighbors is not simply due to transfers received by i from these network neighbors.

Network externalities are typically believed to generate strategic complement effects. How could airtime transfers be strategy substitutes after first adoption? It is difficult to say for sure from the data at our disposal. But strategic substitution effects have been discussed in the theoretical literature on networks (e.g., Jackson 2009, Bramoullé, Kranton and d'Amours, 2014) and evidence of network strategic substitutes has been provided in the case of the adoption of business practices (e.g., Fafchamps and Söderbom, 2014). In our context, strategic substitutes may arise from free-riding. To illustrate, suppose i has two network neighbors j and k. If j has given airtime to k at time t, there is less pressure on i to give at time $t+1$. Individual i may feel exonerated even if k is not a direct neighbor of i. This may be what explains why neighbors of individuals who send transfers send fewer transfers themselves.

Whatever the reason for strategic substitution effects, the main lesson we draw from our analysis is that, prior to first adoption, networks serve an important social learning role. Moreover, given the presence of negative externalities, the importance of social learning may be underestimated by regressions (4.1) and (4.2). For instance, if we combine the two estimates from the column 3 of Table 4, we would conclude that $\frac{\partial \Pr(y_{it+1}=1|z_{it}=0)}{\partial A_{it}}$ underestimates the network effect of social learning by 39% (i.e., 0.0051/(0.0051+0.0033) - 1). The results in the other columns of Table 4 also suggest a significant underestimation of social learning from models (4.1) and (4.2).

4.1. Robustness to endogeneity: Two-stage least squares estimates

Recall that our empirical specifications take the form $\Delta y_{it+1} = \alpha \Delta X_{it} + controls + \Delta\varepsilon_{it+1}$, where X_{it} is a vector containing linear, quadratic and interacted terms of A_{it}. If there are strategic complementarities to usage, the effect of A_{it} is positive. However, as noted above, unobservable factors could lead to bias in the OLS estimator, due to confoundedness. A reasonable benchmark case of confoundedness is contemporaneous positive correlation between A_{it} and ε_{it}, i.e. $E(A_{it}\varepsilon_{it}) > 0$. This could represent a case where, for example, there is an unobserved positive shock to the usefulness of the technology that is common across all neighbors in the network. It should be noted, however, that this doesn't necessarily imply $E(\Delta A_{it}\Delta\varepsilon_{i,t+1}) > 0$. Indeed, $E(A_{it}\varepsilon_{it}) > 0$ could imply $E(\Delta A_{it}\Delta\varepsilon_{i,t+1}) < 0$, in which case our estimator of the effect of interest may actually be downward biased. In this section we probe the issue of confoundedness further using a two-stage least squares (2SLS) approach. We do not have extraneous instruments, but we can still obtain 2SLS estimates that can be informative of the relative importance of confounders. It can be noted, for example, that under the assumption that $E(\Delta A_{i,t-1}\Delta\varepsilon_{i,t+1}) = 0$, the model is identified even if $E(A_{it}\varepsilon_{it}) \neq 0$, provided that $\Delta A_{i,t-1}$is a relevant instrument. In our instrumental variable approach we thus use lagged values of the change in neighbor usage, and similarly lagged values of the change in quadratic and interacted terms of A_{it}, as instruments, whilst treating all variables containing A_{it} are as econometrically endogenous. We consider our three main models, i.e. first adoption, first adoption distinguishing prior exposure to the technology, and usage after first adoption (in all cases with the full set of control variables included).

Results are shown in Table 5. For the model of first adoption, α_2 is significantly positive, α_4 is significantly negative, and α_5 is significantly negative. We find that marginal effects are positive throughout, and we observe a gradual fall in $\partial \Pr / \partial A_{it}$ as A_{it} increases. These results are thus qualitatively the same as those obtained by means of OLS (Table 2, specification [1]). We can reject the null hypothesis that the model is underidentified at the 5% level, indicating that the instruments are relevant.[16] We conclude that there is no evidence that the OLS estimates of the model of first adoption are upward biased. For the model of first adoption distinguishing prior exposure to the technology (specification [2]), there are six endogenous variables (since A_{it} enters several interaction terms), so we need six relevant instruments for identification. We cannot reject the null hypothesis that this model is underidentified, suggesting that we do not have six relevant instruments in this case. The average marginal effect of A_{it} is not significant when $m_{it} = 0$, and further analysis of the data confirms that the instruments are indeed weak in the subsample of observations for which $m_{it} = 0$. We conclude that the results for specification [2] should be interpreted with caution. We have also estimated this model using a control function appproach, which is more parsimonious (and more restrictive) than 2SLS, and may be more efficient (see Wooldridge, 2015, for a discussion).[17] The results, which are available on request, appear to be more robust than the 2SLS estimates, and do not suggest that the OLS estimates in Table 3 are upward biased.[18] Finally, for the model of first adoption and subsequent usage

[16] We use Stata and the command ivreg2 to obtain 2SLS estimates. The underidentification test reported as part of the ivreg2 output is an LM test of whether the equation is identified, i.e., that the excluded instruments are relevant.

[17] The control function approach involves adding the residual from a first-stage regression of ΔA_{it} on all exogenous variables to the set of explanatory variables, and estimating the augmented model by means of OLS. This is a way of "controlling for" the endogeneity of ΔA_{it} using the residual from the first stage. With this approach, we only require $\Delta A_{i,t-1}$ to be a relevant instrument, which is clearly less demanding than the requirements underlying 2SLS.

[18] We obtain strong evidence that $\Delta A_{i,t-1}$ is a relevant instrument. The average marginal effect of A_{it} is

(specification [3] in Table 5), we can reject the null hypothesis that the model is underidentified, indicating that the instruments are relevant. We obtain a positive and statistically significant average effect of A_{it} on first adoption, and a negative and significant average effect on usage following first adoption. Both results are consistent with the OLS results reported earlier in this paper.

5. Conclusion

This study is based on a large administrate dataset covering the universe of phone calls and airtime transfers in an entire country over a four year period. We examine the pattern of adoption of a new phone service over time. This phone service, called ME2U, allows a phone user to transfer airtime from their phone to someone else's. This early form of mobile money was introduced in Rwanda in 2005 by the then de facto monopolist in cell phone services. As a result, we observe the entire universe of peer-to-peer airtime transfers that took place in Rwanda over a four year period.

We start by documenting strong network effects on adoption of the new service: increased usage of ME2U by social neighbors predicts a higher probability of transferring airtime to another user. We then seek to narrow down the possible sources of these network effects by distinguishing between social learning and strategic complementarities in the use of the service. Within social learning, we also seek to differentiate between learning about existence of the new product from learning about its quality or usefulness. We find robust evidence suggestive of social learning both for the existence and the reliability or usefulness of the new service. In contrast, we find that network effects turn negative after first adoption, suggesting that airtime transfers are strategic substitutes among network neighbors, rather than strategic complements. This implies

estimated at 0.010 (z-value 10.6) when $m_{it} = 0$, and 0.008 (z-value 10.2) when $m_{it} = 1$ (the standard errors underlying these results have not been corrected to take into account that the residual is a generated regressor).

that the positive network effects observed in the adoption decision are entirely due to social learning and are not driven by strategic complementarities in usage. Our results thus provide useful insights in the process by which products and services may diffuse on social networks. In our study, learning about existence and quality are important mechanisms, while strategic complementarities are not. It would be interesting to investigate similar mechanisms for other types of services and products, but we leave this for future research.

6. References

Aker, J. C. (2010). "Information from Markets Near and Far: Mobile Phones and Agricultural Markets in Niger," *American Economic Journal: Applied Economics*, 2: 46–59.

Aker, Jenny C. and Marcel Fafchamps (2015). "Mobile Phone Coverage and Producer Markets: Evidence from West Africa", *World Bank Economic Review*, 29(2): 262-92

Aker, Jenny C., Paul Collier and Pedro Vicente (2017). "Is Information Power? Using Mobile Phones and Free Newspapers during an Election in Mozambique," *Review of Economics and Statistics*, 99(2): 185-200.

Altonji, J., Elder, T., and Taber (2005). "An Evaluation of Instrumental Variable Strategies for Estimating the Effects of Catholic Schooling," *Journal of Human Resources*, 40, 791–821.

Bass, Frank (1969). "A New Product Growth for Model Consumer Durables," Management Science 15(5), Theory Series, pp. 215-227 .

Björkegren, Daniel (2019). "The Adoption of Network Goods: Evidence from the Spread of Mobile Phones in Rwanda", *The Review of Economic Studies*, 86(3): 1033-1060.

Blumenstock, Joshua, Nathan Eagle and Marcel Fafchamps (2016). "Airtime Transfers and Mobile Communications: Evidence in the Aftermath of a Natural Disaster", *Journal of Development Economics*, 120: 157-181.

Bramoullé, Yann, Rachel Kranton and Martin D'Amours (2014). "Strategic Interaction and Networks," *American Economic Review*, 104(3): 898-930, March

Centola, Damon (2010). "The Spread of Behavior in an Online Social Network", *Science*, 329: 1194-7

Cole, Shawn A. and A. Nilesh Fernando (2016). "The Value of Advice: Evidence from the Adoption of Agricultural Practices", Harvard University (mimeograph)

Fafchamps, Marcel, Sanjeev Goyal and Marco van der Leij (2010). "Matching and Network Effects", *Journal of the European Economic Association*, 8(1): 203-31.

Fafchamps, Marcel and Måns Söderbom (2014). "Network Proximity and Business Practices in African Manufacturing," *World Bank Economic Review*, 28(1): 99-129

Fafchamps, Marcel and Bart Minten (2012). "Impact of SMS-Based Agricultural Information on Indian Farmers", *World Bank Economic Review*, 26(3): 383-414

Granovetter, Mark (1995). *Getting a Job: A Study of Contacts and Careers*, University of Chicago Press, Chicago

Jack, W. and T. Suri (2014). "Risk Sharing and Transactions Costs: Evidence from Kenya's Mobile Money Revolution," *American Economic Review*, 104: 183–223

Jackson, Matthew O. (2009). *Social and Economic Networks*, Princeton University Press, Princeton, 2008

Jackson, Matthew O. and Leeat Yariv (2005) "Diffusion on Social Networks", *Economie Publique*, 16(1): 3-16

Jensen, R. (2007). "The Digital Provide: Information (Technology), Market Performance, and Welfare in the South Indian Fisheries Sector," *Quarterly Journal of Economics*, 122: 879–924.

Mobius, Markus and Tanya Rosenblat (2014). "Social Learning in Economics", *Annual*

Review of Economics, 6: 827-847, August

Moffitt, Robert (2003). "Policy Interventions, Low-Level Equilibria, and Social Interactions," in S. N. Durlauf and H. P. Young, *Social Dynamics*, Cambridge, Massachusetts: MIT Press.

Mood, Alexander M., Franklin A. Graybill and Duane C. Boes (1974). *Introduction to the Theory of Statistics*, Third edition, McGraw-Hill, New York

Oster, Emily (2019). "Unobservable Selection and Coefficient Stability: Theory and Evidence," *Journal of Business & Economic Statistics*, 37(2), 187-204.

Ryan, S. P. and C. Tucker (2012). "Heterogeneity and the dynamics of technology adoption," *Quantitative Marketing and Economics*, 10: 63–109.

Saloner, G. and A. Shepard (1995): "Adoption of Technologies with Network Effects: An Empirical Examination of the Adoption of Automated Teller Machines," *RAND Journal of Economics*, 26: 479–501.

Suri, Tavneet and William Jack (2014). "Risk Sharing and Transaction Costs: Evidence from Kenya's Mobile Money Revolution", *American Economic Review*, 104 (1): 183-223, January

Vega-Redondo, Fernando (2007). *Complex Social Networks*, Cambridge University Press, Cambridge.

Wooldridge, Jeffrey M. (2015). "Control Function Methods in Applied Econometrics," Journal of Human Resources 50(2): 420-445.

Young, H. Peyton (1999). "Diffusion in Social Networks," Papers 2, Brookings Institution - Working Papers.

Young, H. Peyton (2009). "Innovation Diffusion in Heterogeneous Populations: Contagion, Social Influence, and Social Learning," *American Economic Review*, 99(5): 1899-1924, December.

Figure 1: Number of SIM-ID's adopted: August 2016 – December 2008

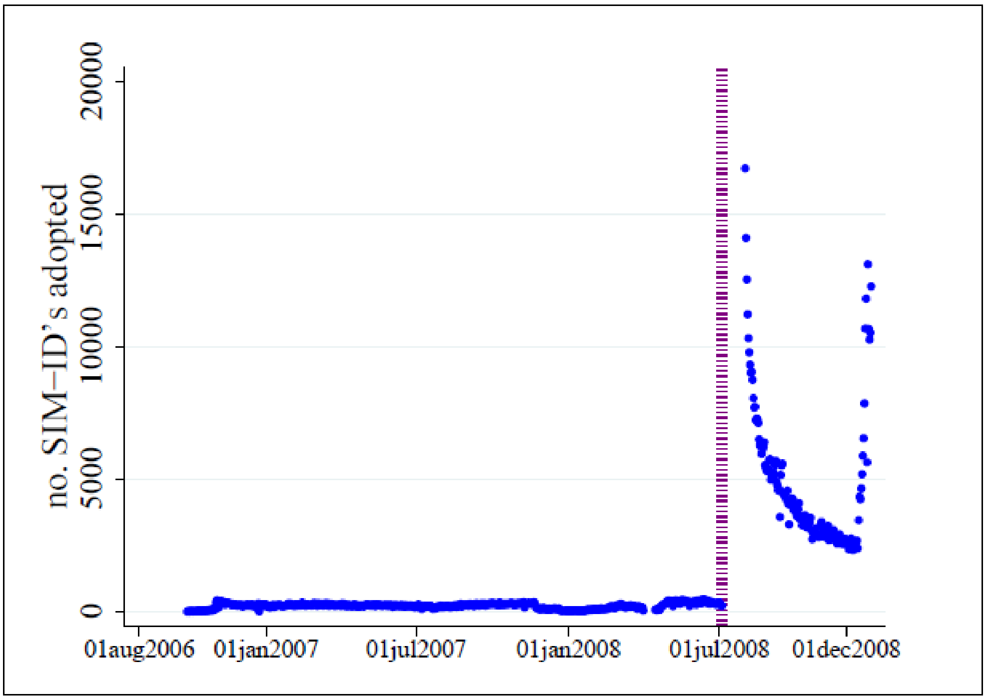

Figure 2: Proportion of adopters and users

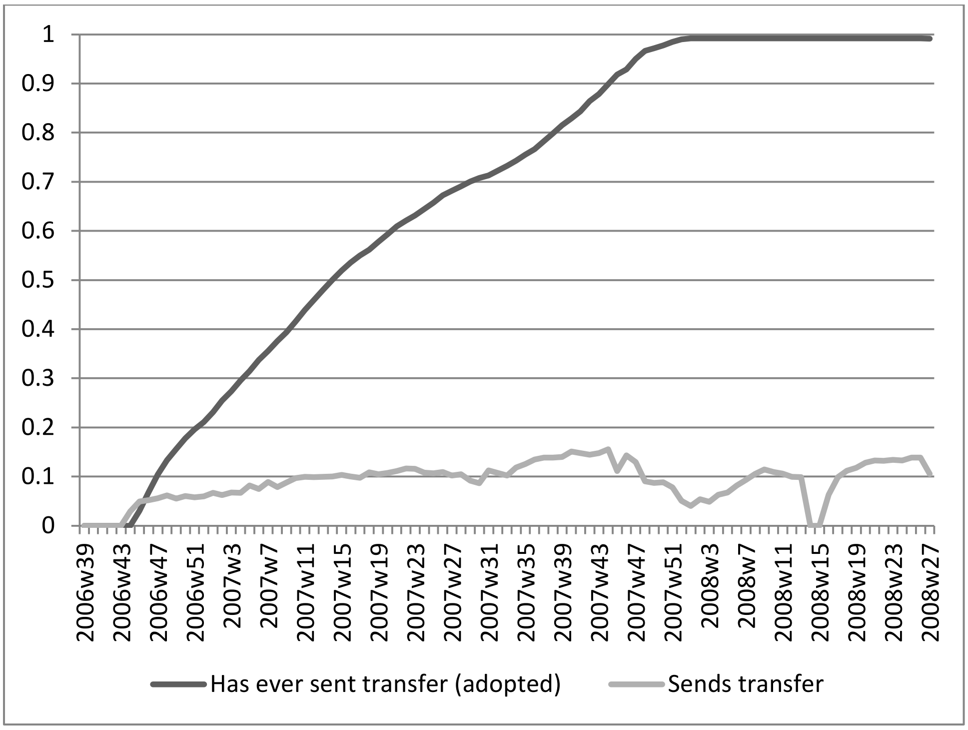

Table 1
Summary statistics

	(1) Full sample			(2) Before adoption			(3) Before adoption & 1st in-transfer		
	Mean	Median	Std dev	Mean	Median	Std dev	Mean	Median	Std dev
N(it)	264	202	243	102	66	115	89.3	57	99.3
A(it)	73.3	51	76.3	24.9	13	33.7	20.6	11	27.8
ΔA(it)	1.71	1	1.95	1.55	1	1.92	1.45	1	1.82
Amount received, if positive	457	200	1478	642.3	223	2682			
Number of transfers received	0.12	0	0.55	0.05	0	0.35			
Number of neighbors from whom i received a transfer	0.09	0	0.36	0.04	0	0.21			
Amount sent, if positive	589	200	1735						
Number of transfers sent	0.19	0	1.52						
Number of neighbors to whom i sent a transfer	0.15	0	0.97						
Number of phone calls by i	21.9	12	31.3	17.7	10	26.8	16.5	9	25.1
Weeks with SIM card	43.4	42	25	20	16	14.7	17.3	14	12.3
Observations	376372			91889			71835		

Table 2
First Adoption: First Difference Estimates

	(1)			(2)			(3)		
FD Estimates	Coef.	s.e.	Coef/s.e.	Coef.	s.e.	Coef/s.e.	Coef.	s.e.	Coef/s.e.
ΔA(it)	0.0023602	0.0007445	3.17	0.0018	0.00077	2.35	0.0018	0.0007723	2.35
ΔS(it)^2	0.0000827	0.000043	1.92	-0.00076	7.1E-05	-10.7	-0.00074	7.8E-05	-9.51
ΔA(it)^2	-0.0000332	3.33E-06	-9.99	-3.2E-05	5.0E-06	-6.44	-0.000031	4.8E-06	-6.39
Δ[A(it)S(it)]	0.0002695	0.000029	9.31	0.00029	4.3E-05	6.81	0.00028	4.5E-05	6.26
Marginal effects of A(it), at different levels of A(it)									
A(it) = 0	0.0078	0.0004	17.55	0.0077	0.0005	14.4	0.0075	0.0005	14.5
A(it) = 20	0.0065	0.0004	15.65	0.0064	0.0004	14.8	0.0063	0.0004	15.1
A(it) = 40	0.0051	0.0004	12.13	0.0051	0.0004	12.4	0.0050	0.0004	13.1
A(it) = 60	0.0038	0.0005	8.05	0.0039	0.0005	8.01	0.0038	0.0004	8.49
A(it) = 80	0.0025	0.0006	4.5	0.0026	0.0006	4.23	0.0026	0.0006	4.50
A(it) = 100	0.0011	0.0006	1.78	0.0013	0.0008	1.70	0.0013	0.0007	1.85
Average marginal effect of A(it)									
A(it) = 25.4	0.0061	0.0004	14.83	0.0061	0.0004	14.46	0.0059	0.0004	14.92
Controls									
year x month		N			Y			N	
district		N			Y			N	
tower		N			Y			Y	
year x month x district		N			N			Y	
R-squared		0.007			0.037			0.044	
Observations		96,266			96,266			96,266	

Note: Standard errors are clustered at the district level (M=27). Marginal effects are evaluated at sample means of regressors (in levels).

Table 3

Generalized First Adoption Model: First Difference Estimates

	(1)			(2)			(3)		
FD Estimates	Coef.	s.e.	Coef/s.e.	Coef.	s.e.	Coef/s.e.	Coef.	s.e.	Coef/s.e.
ΔA(it)	0.0035	6E-04	5.45	0.0028	6E-04	4.62	0.0029	6E-04	4.63
ΔS(it)^2	0.0002	4E-05	5.49	-0.0006	5E-05	-12.56	-0.0006	6E-05	-10.98
ΔA(it)^2	0.0000	3E-06	-10.5	-0.00003	4E-06	-8.02	-0.00003	4E-06	-7.72
Δ[A(it)S(it)]	0.0002	2E-05	8.3	0.0002	3E-05	7.90	0.0002	3E-05	7.12
Δm(it)	0.0699	3E-02	2.32	0.0237	3E-02	0.72	0.0255	3E-02	0.78
Δ[m(it) x S(it)]	0.0102	3E-03	3.69	0.0144	3E-03	4.32	0.0141	3E-03	4.23
Δ[m(it) x A(it)]	-0.0051	5E-04	-9.82	-0.0050	7E-04	-7.60	-0.0050	7E-04	-7.56
Δ[m(it) x S(it)^2]	-0.0002	6E-05	-3.36	-0.0003	8E-05	-3.96	-0.0003	8E-05	-3.88
Δ[m(it) x A(it)^2]	0.0000	3E-06	1.17	0.000003	3E-06	0.85	0.000003	3E-06	0.83
Δ[m(it) x A(it) x S(it)]	0.0001	3E-05	4.08	0.0001	3E-05	3.89	0.0001	3E-05	3.90
Average marginal effects of A(it)									
m(it) = 0	0.0058	0.0005	11.88	0.0059	0.0005	12.71	0.0058	0.0004	13.18
m(it) = 1	0.0035	0.0004	7.98	0.0037	0.0005	8.01	0.0036	0.0004	8.04
Controls									
year x month		N			Y			N	
district		N			Y			N	
tower		N			Y			Y	
year x month x district		N			N			Y	
R-squared		0.012			0.042			0.049	
Observations		96,266			96,266			96,266	

Note: Standard errors are clustered at the district level (M=27). Marginal effects are evaluated at sample means of regressors (in levels).

Table 4
Adoption & subsequent usage: First Difference Estimates

	(1)			(2)			(3)			(4)		
	Coef.	s.e.	Coef/s.e.	Coef.	s.e.	Coef/s.e.	Coef.	s.e.	Coef/s.e.	Coef.	s.e.	Coef/s.e.
ΔA(it)	0.0049	0.0007	6.9	0.0034	0.0007	4.95	0.0033	0.0007	4.91	0.0033	0.0007	4.92
ΔS(it)^2	0.0005	3E-05	17.26	0.0001	4E-05	2.26	0.0001	4E-05	2.09	0.0001	4E-05	2.18
ΔA(it)^2	-1E-05	3E-06	-4.28	-1E-05	3E-06	-4.23	-1E-05	3E-06	-4.29	-1E-05	3E-06	-4.24
Δ[A(it) x S(it)]	4E-05	2E-05	1.74	0.0001	2E-05	3.5	0.0001	2E-05	3.50	0.0001	2E-05	3.53
Δ[z(it) x S(it)]	-0.0189	0.0016	-11.69	-0.0475	0.0024	-19.44	-0.0480	0.0026	-18.52	-0.0474	0.0026	-18.44
Δ[z(it) x A(it)]	-0.0098	0.0013	-7.61	-0.0078	0.0012	-6.5	-0.0077	0.0012	-6.48	-0.0075	0.0012	-6.33
Δ[z(it) x S(it)^2]	-0.0005	3E-05	-16.34	-0.0001	4E-05	-3.8	-0.0001	4E-05	-3.57	-0.0001	4E-05	-3.65
Δ[z(it) x A(it)^2]	1E-05	3E-06	4.84	1E-05	3E-06	4.58	1E-05	3E-06	4.64	1E-05	3E-06	4.59
Δ[z(it) x A(it) x S(it)]	-2E-05	2E-05	-0.9	-0.0001	2E-05	-2.66	-0.0001	2E-05	-2.70	-0.0001	2E-05	-2.74
Amount_received(it)										4E-06	3E-06	1.38
z(it) x Amount_received(it)										-1E-05	5E-06	-3.04
Average marginal effects of A(it)												
z(it) = 0	0.0049	0.0004	11.49	0.0052	0.0005	11.4	0.0051	0.0004	11.44	0.0051	0.0004	11.46
z(it) = 1	-0.0038	0.0006	-6.01	-0.0033	0.0006	-5.3	-0.0033	0.0006	-5.22	-0.0032	0.0006	-4.97
Controls												
year x month		N			Y			N			N	
district		N			Y			N			N	
tower		N			Y			Y			Y	
year x month x district		N			N			Y			Y	
R-squared		0.006			0.007			0.008			0.008	
Observations		361,616			361,616			361,616			361,616	

Note: Standard errors are clustered at the district level (M=27). Marginal effects are evaluated at sample means of regressors (in levels).

Table 5
Robustness to Endogeneity: First Difference 2SLS Estimates

	(1)			(2)			(3)		
FD Estimates	Coef.	s.e.	Coef/s.e.	Coef.	s.e.	Coef/s.e.	Coef.	s.e.	Coef/s.e.
ΔA(it)*	0.0069	0.0017	4.13	0.0163	0.0116	1.41	0.0076	0.0016	4.68
ΔS(it)^2	-0.0008	0.0001	-12.06	0.0014	0.0012	1.16	0.0001	0.0001	2.19
ΔA(it)^2*	-0.00005	4E-06	-10.23	0.0001	0.0002	0.38	-2E-05	8E-06	-2.49
Δ[A(it)S(it)]*	0.0003	4E-05	9.65	-0.0012	0.0010	-1.21	0.0001	3E-05	2.36
Δm(it)				0.0711	0.0898	0.79			
Δ[m(it) x S(it)]				-0.0231	0.0339	-0.68			
Δ[m(it) x A(it)]*				0.0177	0.0157	1.12			
Δ[m(it) x S(it)^2]				-0.0024	0.0017	-1.44			
Δ[m(it) x A(it)^2]*				-0.0002	0.0002	-1.02			
Δ[m(it) x A(it) x S(it)]*				0.0020	0.0013	1.47			
Δ[z(it) x S(it)]							-0.0335	0.0034	-9.78
Δ[z(it) x A(it)]*							-0.0154	0.0021	-7.36
Δ[z(it) x S(it)^2]							-0.0002	0.0001	-3.26
Δ[z(it) x A(it)^2]*							2E-05	8E-06	2.95
Δ[z(it) x A(it) x S(it)]*							-0.0001	3E-05	-1.84

Note: Standard errors are clustered at the district level (M=27). All variables containing A(it) are treated as endogenous, as indicated by *. Lagged differenced values of endogenous explanatory variables are used as instruments. The table continues on the next page.

Table 5 continued

	(1)			(2)			(3)		
Marginal effects of A(it), evaluated at sample means of the regressors									
A(it) = 0	0.0137	0.0015	9.35						
A(it) = 20	0.0119	0.0014	8.77						
A(it) = 40	0.0101	0.0013	7.98						
A(it) = 60	0.0082	0.0012	6.95						
A(it) = 80	0.0064	0.0011	5.66						
A(it) = 100	0.0046	0.0011	4.14						
m(it) = 0				-0.0055	0.0081	-0.68			
m(it) = 1				0.0413	0.0064	6.47			
z(it) = 0							0.0082	0.0015	5.55
z(it) = 1							-0.0062	0.0012	-5.31
Underidentification test (p-value)		0.022			0.237			0.0147	
Controls									
tower		Y			Y			Y	
year x month x district		Y			Y			Y	

Note: Standard errors are clustered at the district level (M=27). All variables containing A(it) are treated as endogenous, as indicated by *. Lagged differenced values of endogenous explanatory variables are used as instruments.

7. Appendix A. Conceptual framework

The purpose of this Appendix is to provide a theoretical framework to support our empirical analysis. The focus of our attention is adoption, that is, the first usage of a new product or service by someone who has not used it before. We are interested in how social networks influence adoption. To formalize this process, let $y_{it} = \{0, 1\}$ be a dichotomous variable equal to 1 if individual i uses the product at time t, and 0 otherwise. We think of time as a sequence of time intervals, i.e., our model is in discrete time. Adoption describes the first time at which $y_{it} > 0$ for individual i. Let t_i denote the time at which individual i becomes 'at risk' of adopting the product.[19] Further let T_i denote the time at which individual i first uses the product. Finally, let T denote the last data period for which we have information. By definition, $T_i > T$ for an individual who, by time T, has not yet used the product.

As we will argue below, usage after adoption provides useful information as well. Usage y_{it} can therefore be divided into two vectors or periods: the time until first usage $\{y_{it_i}, ...y_{iT_i}\}$; and usage after that $\{y_{iT_i+1}, ...y_{iT}\}$. By construction, $\{y_{it_i}, ...y_{iT_i}\}$ is either a sequence of 0's ending with a single 1, or a string of 0's (for someone who never adopts). The length of each of the two i vectors varies across individuals.

We are interested in identifying predictors of y_{it} that depend on the adoption and usage behavior of the social neighbors of i. To do so effectively, we present a few simple concepts before articulating our testing strategy. We first discuss social learning, before introducing network externalities. We assume throughout that the researcher has information about y_{it}.

[19] This can be the time at which the new product is introduced, or the time at which i acquires a device for which product is useful.

7.1. Social learning about product existence

There is much to learn from simple models of social learning. Let us first focus on information about the existence of the product. We then turn to information about the qualities of the product. We end with a short discussion of experimentation, which is adoption purely for the purpose of eliciting information about product quality. The focus of this section is to use simple models to develop intuition about social learning that we can then take to the data.

Learning about the existence of the new product closely resembles a contagion process (see e.g. Bass, 1969). Without information about the existence of the product, the agent simply cannot adopt. Hence having been exposed to information about the product is a necessary condition for adoption. This information can come from two sources: (1) information received from various sources outside the social network (e.g., ads on billboard, radio, TV, junk mail, or newspaper); and (2) information received from the social network (e.g., friends, relatives, co-workers).

Let θ_{vt} denote the probability of receiving information from outside the social network in location v at time t. We take this probability as given and we do not seek to model its determinants. But we think of it as having a strong local component, capturing the local nature of advertisement coverage.

A simple model for the probability of receiving information from a social source at time t can be formulated as:

$$\Pr(i \text{ receives information from network at } t+1) = 1 - (1-q)^{\Delta A_{it}}$$

where ΔA_{it} is the number of neighbors of i who have started using the product in period t – and thus have become aware of its existence and can relay this information to i, something each of them does with probability q. We assume that the researcher observes ΔA_{it}, or a close proxy. The cumulative probability that i has received information about the existence of the product is thus an increasing and convex function of the cumulative number of i's neighbors who have adopted at t – and thus could have passed information about the product to i with probability q during that time period.

Let us now combine the two sources of information. If we assume independence between θ_{vt} and the signal received from each neighbor, the probability of *not* being informed within period t is $(1-\theta_{vt})(1-q)^{\Delta A_{it}}$. Now let us assume that, once i is informed that the product exists, i adopts with probability p_i. This is the probability of usage in any given period, conditional on knowing about the product. For some individuals this probability is low; for others it is high.

Over time the likelihood of having heard of the product increases. Formally, the probability of *not* having heard of the product between time t_i and t is:

$$\Pr = \prod_{s=t_i}^{t} (1-\theta_{vs})(1-q)^{\Delta A_{is}} = (1-q)^{A_{it}} \prod_{s=t_i}^{t} (1-\theta_{vs})$$

where $A_{it} \equiv \sum_{s=t_i}^{s=t} \Delta A_{is}$ is the cumulative number of adopting neighbors between t_i and t, and t_i is the time at which i starts being at risk of being exposed to information about the product's existence. If θ_{vt} is constant over time for location v, the formula simplifies to:

$$\Pr = (1-q)^{A_{it}}(1-\theta_v)^{S_{it}}$$

where $S_{it} \equiv t - t_i$ is the time elapsed between t_i and t.

The probability that agent i adopts the product at time t is the probability that he has been informed times p_i:

$$\Pr(y_{it+1} = 1 | \{y_{it_i}, ..., y_{it}\} = \{0, ..., 0\}) = [1 - (1-q)^{A_{it}}(1-\theta_v)^{S_{it}}]p_i \tag{7.1}$$

Adoption can take place even for someone who has no social neighbors, or whose neighbors have not adopted. The model predicts that the likelihood of adoption increases in a systematic fashion over time, without or without adopting neighbors. This is a mechanical effect: as time passes, the agent has more and more chances of hearing about the product. The probability of first adoption increases with time since inception S_{it} and with A_{it}, although in both cases the effect is concave: the derivative of the probability of adoption w.r.t. S_{it} and A_{it} falls with S_{it} and with A_{it}. This is because having heard about the product once is enough to know of its existence.

Once the product has been used once, i may continue using it with a certain probability. But if the only source of network effects is social learning about the existence of the product, the probability of usage after first adoption is no longer a function of the number of adopting neighbors. Formally we have:

$$\Pr(y_{it+1} = 1 | y_{is} = 1 \text{ for some } s < t) \ = p_i + \varepsilon_{it+1} \tag{7.2}$$

Thus once i has learned about the existence of the product, the data generating process shifts from (7.1) to (7.2). An identical prediction is made if the researcher observes a signal M_{it} that is equal to 1 when individual i has unambiguously been made aware of the existence of the new product, and 0 otherwise:

$$\Pr(y_{it+1} = 1 | M_{is} = 1 \text{ for some } s < t) \ = p_i + \varepsilon_{it+1} \tag{7.3}$$

To recap, when network neighbors circulate information about product existence and nothing more, the probability of adoption increases in the number of adopting neighbors, but at a decreasing rate. After first adoption or after becoming aware of the product, subsequent usage does not depend on the number of adopting neighbors.

7.2. Social learning about product quality

We get different predictions if social learning is about product quality. In this case, the decision to adopt at time t depends not on the probability of receiving a signal within a given time interval, but rather on the cumulative information about the product received up to time t.

To keep the same notation, let θ_{vt} now denote the probability that individual i receives an independent signal about the quality of the product at time t. This probability can vary over time t and across locations v. To keep things simple, let us assume that this signal takes only two values, 0 and 1, i.e., a bad signal or a good signal. Let μ denote the true probability that the product performs: a high μ good always performs well, while a low μ good often performs poorly. Individuals differ in how much they value unobserved quality μ – more about this later.

We assume that the posterior belief h_{it} of individual i at time t is simply the sample estimate of the unknown Bernoulli parameter μ based on the information available to i at time t.[20] Let N_{it} be the number of signals received by i at up to t and let N_{it}^1 be the number of signals with value 1, i.e., the number of good signals. We have:

$$h_{it} = \frac{N_{it}^1}{N_{it}} \tag{7.4}$$

[20] This is simplified Bayesian approach – see Mood, Graybill and Boes (1974) p. 342 for the correct Bayesian estimator of a Bernouilli parameter. But this simple approach suffices for our purpose.

The variance of this belief is approximately given by:

$$v_{it}^2 = \frac{1}{N_{it}} h_{it}(1 - h_{it}) \tag{7.5}$$

As sample size increases, h_{it} tends to μ and v_{it}^2 tends to 0.[21]

Since we do not observe what signal people observe, we never know what N_{it}^1 is. But we can write:

$$h_{it} = \mu + e_{it} \text{ with } e_{it} \sim (0, \mu(1 - \mu)/N_{it})$$

In other words, the information people have is, on average, unbiased and the variance of their beliefs shrinks over time.

If we allow agents to hold a prior belief h_{i0}, this belief can be regarded as coming from a sample of observations N_{i0} that we do not observe. The point estimate of this belief marks how biased the prior belief is, and the size of the sample determines how confident the agent is in his prior belief. This can be formalized as follows:

$$\begin{aligned} h_{i0} &= \frac{N_{i0}^1}{N_{i0}} \\ h_{it}^b &= \frac{N_{i0}^1 + N_{it}^1}{N_{i0} + N_{it}} = h_{i0} \frac{N_{i0}}{N_{i0} + N_{it}} + h_{it} \frac{N_{it}}{N_{i0} + N_{it}} \\ v_{it}^2 &= \frac{1}{N_{i0} + N_{it}} h_{it}^b (1 - h_{it}^b) \end{aligned}$$

where h_{it}^b now denotes the posterior belief of agent i at t.

We do not observe h_{i0} and N_{i0}. If we let the number of signals received be denoted n_{it},

[21] The above formula for the variance is obtained by combining Mood et al. (1974) p. 236 with p. 89.

beliefs can be written as following a model of the form:

$$\begin{aligned} q^b_{it} &= \alpha\frac{\gamma}{\gamma+n_{it}} + \mu\frac{n_{it}}{\gamma+n_{it}} + e^b_{it} \text{ with } e^b_{it}\tilde{}(0,\sigma^2_{it}) \\ \sigma^2_{it} &= \frac{1}{\gamma+n_{it}}(\alpha\frac{\gamma}{\gamma+n_{it}} + \mu\frac{n_{it}}{\gamma+n_{it}})(1-\alpha\frac{\gamma}{\gamma+n_{it}} - \mu\frac{n_{it}}{\gamma+n_{it}}) \end{aligned}$$

As with uninformed priors, beliefs h^b_{it} tend to μ over time, but they show some persistence around initial priors.[22]

Having modelled learning, we now turn to adoption. We start without prior beliefs. We assume that individuals differ in the threshold value of μ that they require before adopting. At first glance, it seems that we could simply assume that people adopt if their estimate of μ is larger than some value τ_i with $0 < \tau_i < 1$. This decision rule, however, is too crude. It predicts that people adopt after a single good signal since, in that case, their posterior belief is $h_{i1} = 1 \geq \tau_i$ for any τ_i. This is clearly an unappealing decision rule because an estimate of μ based on a single observation is very imprecise. To capture this intuition in the simplest possible way, we posit that the expected utility of adoption $E[U_{it}(y_{it} = 1)|\omega_{it}]$ can be written as a mean-variance form. We have:

$$y_{it+1} = 1 \text{ iff } h_{it} - Rv^2_{it} \geq \tau_i$$

where R is a risk aversion parameter and τ_i is now a threshold value of expected utility. Since we do not observe h_{it} and v^2_{it} directly, we replace them by formulas (7.4) and (7.5) above and

[22] The variance σ^2_{it} is not monotonic over time, however. Intuition is as follows. Imagine the agent starts with a strong prior far from μ (a strong prior means N_{i0} is large). Initially σ^2_{it} is quite small because it is dominated by the strong prior. As more information is revealed, posterior beliefs are progressively pulled away from prior h_{i0} and σ^2_{it} increases. Eventually posterior beliefs settle on μ and the variance falls, dominated now by N_{it}.

we get:

$$\Pr(y_{it+1}=1|\{y_{it_i},...,y_{it}\}=\{0,...,0\})=\Pr\left((\mu-\tau_i)-R\frac{\mu(1-\mu)}{n_{it}}\geq -e_{it+1}\right) \tag{7.6}$$

Equation (7.6) shows that the probability of adoption increases with n_{it}. The intuition is straightforward: the variance term shrinks and vanishes at the limit, and this raises the expected utility of adoption for some people. Not everybody adopts, however, because μ is not higher than τ_i for everyone.

We can now generalize the above to the case where people hold prior beliefs. We now have:

$$\Pr(y_{it+1}=1|\{y_{it_i},...,y_{it}\}=\{0,...,0\})= \tag{7.7}$$
$$\Pr\left(\alpha\frac{\gamma}{\gamma+n_{it}}+\mu\frac{n_{it}}{\gamma+n_{it}}+R\frac{1}{\gamma+n_{it}}(\alpha\frac{\gamma}{\gamma+n_{it}}+\mu\frac{n_{it}}{\gamma+n_{it}})(1-\alpha\frac{\gamma}{\gamma+n_{it}}-\mu\frac{n_{it}}{\gamma+n_{it}})\geq\tau_{it}-e^b_{it+1}\right)$$

To close the model, we need to stipulate the data generating process of n_{it}, the number of signals received. In practice, we do not observe n_{it} but, by analogy with the previous subsection, we expect it to be an increasing function of time since inception S_{it} and of the number of adopting neighbors A_{it}. To show this formally, let us assume that in each period individual i receives a signal from outside his network with a constant location-specific probability θ_v,[23] and with probability q individual i receive a signal from any newly adopting neighbor. The expected number of signals received at time t is a sum of two binomial processes. The average number of signals received outside the network up to time is given by a binomial process with parameter θ_v and S_{it}, and is simply $\theta_v S_{it}$. The average number of signals from the networks is qA_{it}. Thus we have:[24]

[23] To keep the algebra simple and derive the intuition clearly, we ignore here the possibility of a time-varying signal probability.

[24] Where, given our assumptions, v^2 can in principle be calculated from the variance formula for binomial distributions.

$$n_{it} = \theta_v S_{it} + qA_{it} + u_{it} \text{ with } u_{it}\tilde{}(0, v^2) \tag{7.8}$$

Without prior beliefs, the probability of adoption can thus be written:

$$\Pr(y_{it+1} = 1 | \{y_{it_i}, ..., y_{it}\} = \{0, ..., 0\}) = \Pr\left((\mu - \tau_i) - R\frac{\mu(1-\mu)}{\theta_v S_{it} + qA_{it} + u_{it}} \geq -e_{it+1} \right) \tag{7.9}$$

Equation (7.9) shows that the probability of first adoption is monotonically increasing in S_{it} and A_{it}.

The probability of adoption with prior beliefs is similarly obtained by replacing n_{it} in equation (7.7) by its value given by (7.8). Our earlier observation remains valid: with strong prior beliefs, the variance term that multiplies R in equation (7.7) can initially be quite small. If the prior belief h_{i0} is high and its variance v_{i0}^2 is small, individual i will adopt immediately. The social learning model therefore predict that individuals with strong optimistic priors adopt early. So doing, they receive information about the quality of the product, information that they may circulate among their social circle. If the information is sufficiently bad, i.e., if revealed quality is less than τ_i, early adopters will abandon the new product, and the information that diffuses among the social network will discourage adoption by others. If the information is sufficiently good, its diffusion in the network will progressively raise posterior beliefs according to equation (7.7) and adoption will spread among individuals with a sufficiently high valuation τ_i for the product. Because the accumulation of information eventually reduces the variance of posterior beliefs, adoption is an increasing function of the information received, and thus of the number of adopting neighbors.

What happens after an individual has adopted the product once? In the context of our empirical application, it is natural to assume that usage reveals a lot of relevant information about the product. To capture this idea in a stylized way, let us imagine that using the product once perfectly reveals the quality of the product. It follows that usage is now driven by τ_i; social learning no longer matters. Formally we have:

$$\Pr(y_{it+1} = 1 | y_{is} = 1 \text{ for some } s \leq t) = \Pr\left((\mu - \tau_i) \geq -e_{it+1}\right) \tag{7.10}$$

which does not depend on time or adopting neighbors.

What happens if individual i is observed to receive an unambiguous signal revealing the existence of the product? In this case, this signal does not, by itself, dispel uncertainty about the quality of the product and thus should not eliminate the role of social learning in reducing uncertainty about the net benefit of adoption. In other words, adoption continues to follow equation (7.7) after $M_{it} = 1$. This is different from what happens when social learning only affects knowledge about the existence of the product, and thus provides a way of identifying which type of social learning is present in the data.

To summarize, when social learning is purely about product quality, the likelihood of adoption is predicted to increase over time as the number of adopting neighbors rises, irrespective of whether the individual received a signal about product existence or not, that is, whether $M_{is} = 1$ or not. After first adoption, however, the role of social learning essentially disappears and the probability of continued usage is no longer a function of the number of adopting neighbors. In contrast, if social learning is solely about product existence, the data generating process switches to (7.3) after $M_{is} = 1$. This makes it possible to test the two learning models against each other even in a reduced form. If social learning combines both elements, then we expect the coefficient of A_{it} to be significantly lower after $M_{is} = 1$, but to remain positive until first adoption.

7.3. Network externalities and strategic complementarities

Social learning can be seen as a network externality: individuals benefit from the information accumulated and shared by others. We have shown that social learning generates a correlation between neighbors' adoption and own adoption by individual i. There are many other network externalities that do not involve learning. Since we do not have any information to further disentangle different types of strategic complementarities, we need not discuss them in more detail. The main distinction between strategic complementarities and social learning is that the effect of social learning disappears after i has used the product at least once, while the effect of other strategic complementarities does not. This simple observation forms the basis of our identification strategy between social learning and other network externalities.

7.4. Estimation

We now demonstrate how these ideas can be turned into an estimation strategy. The reduced form for models (7.1) and (7.7) is similar and can be written as:

$$\Pr(y_{it+1} = 1|\{y_{it_i}, ..., y_{it}\} = \{0, ..., 0\}) = \alpha_i + \alpha_1 S_{it} + \alpha_2 A_{it} + \varepsilon_{it+1} \quad (7.11)$$

$$\begin{aligned} \Pr(y_{it+1} &= 1|\{y_{it_i}, ..., y_{it}\} = \{0, ..., 0\}) \quad (7.12) \\ &= \alpha_i + \alpha_1 S_{it} + \alpha_2 A_{it} + \alpha_3 S_{it}^2 + \alpha_4 A_{it}^2 + \alpha_5 S_{it} A_{it} + \varepsilon_{it+1} \end{aligned}$$

Model (7.11) is a simple linear approximation of the two structural models (7.1) and (7.7). Parameter α_i captures variation in product usefulness across individuals. With any social learning we expect the marginal effect adopting neighbors to be positive, i.e., $\frac{d\Pr}{dA_{it}} > 0$. In equation (7.11) this means $\alpha_2 > 0$. We also the marginal effect of S_{it} to be positive – which implies $\alpha_1 > 0$ in equation (7.11). This is because the likelihood of adoption should increase over time as more information about the product becomes available from within and outside the social network. In regression model (7.12) we have included extra terms to test the concavity of the relationship with respect to S_{it} and A_{it} as predicted by social learning about product existence. This concavity can be investigated by testing $\alpha_3 < 0, \alpha_4 < 0$ and $\alpha_5 < 0$.[25] We have include error terms to reflect the possibility that adoption probabilities may vary across individuals over time – more about this in the empirical section.

In contrast, the reduced form model for (7.2) is of the form:

$$\Pr(y_{it+1} = 1 | \{y_{it_i}, ..., y_{it}\} = \{0, ..., 0\}, M_{is} = 1 \text{ for some } s \leq t) = \alpha_i + \varepsilon_{it+1}$$

It is therefore easy to test one model against the other by estimating a regression model of the form:

$$\begin{aligned} \Pr(y_{it+1} &= 1 | \{y_{it_i}, ..., y_{it}\} = \{0, ..., 0\}) = \alpha_i + \alpha_1 S_{it} + \alpha_2 A_{it} + \alpha_3 S_{it}^2 + \alpha_4 A_{it}^2 + \alpha_5 S_{it} A_{it} \\ &+ \beta_0 m_{it} + \beta_1 S_{it} m_{it} + \beta_2 A_{it} m_{it} + \beta_3 S_{it}^2 m_{it} + \beta_4 A_{it}^2 m_{it} + \beta_5 S_{it} A_{it} m_{it} + \varepsilon_{it+1} \end{aligned} \quad (7.13)$$

with $m_{it} = 1$ if $M_{is} = 1$ for some $s \leq t$, and $= 0$ otherwise. As before α_i captures variation in product usefulness across individuals. If the true model is social learning only about existence, then all β's should be equal to minus the corresponding α's, so that the sum of the two equals 0. If the true model is only social learning about quality, then all β's should be equal to 0. If we reject both hypotheses – and the total marginal effect of S_{it} and A_{it} on the dependent variable is smaller when $m_{it} = 1$ – it means that the true model is a hybrid of the two forms of social learning.

[25] The sign prediction on the cross term $S_{it} A_{it}$ arises because information from the network is less valuable if the person has already received many signals from non-network sources.

A similar approach can be used to test the presence of network externalities and strategic complementarities driven by factors other than social learning. Identification is achieved simply by noting that social learning stops once i has adopted, while other network externalities continue having an influence on usage even after i is familiar with the product and its characteristics.

Formally, let $z_{it} = 1$ if $y_{is} = 1$ for some $s < t$, and 0 otherwise. In other words, $z_{it} = 1$ if i has already used the product prior to period t. The estimated model is of the form:

$$\begin{aligned} \Pr(y_{it+1} &= 1) = \alpha_i + \alpha_1 S_{it} + \alpha_2 A_{it} + \alpha_3 S_{it}^2 + \alpha_4 A_{it}^2 + \alpha_5 S_{it} A_{it} + \gamma_0 z_{it} \\ &+ \gamma_1 S_{it} z_{it} + \gamma_2 A_{it} z_{it} + \gamma_3 S_{it}^2 z_{it} + \gamma_4 A_{it}^2 z_{it} + \gamma_5 S_{it} A_{it} z_{it} + \varepsilon_{it+1} \qquad (7.14) \end{aligned}$$

Unlike models (7.12) and (7.13), regression model (7.14) includes observations before and after first adoption. If there is no social learning, network effects should be the same before and after first adoption, i.e., we should observe that $\gamma_2 = \gamma_3 = \gamma_4 = \gamma_5 = 0$. If there are no network effects other than social learning, then we should observe that whatever network effects were present before first adoption should cancel out after first adoption, i.e., that:

$$\frac{\partial \Pr(y_{it+1} = 1 | z_{it} = 0)}{\partial A_{it}} > 0 = \frac{\partial \Pr(y_{it+1} = 1 | z_{it} = 1)}{\partial A_{it}}$$

which is guaranteed if $\gamma_2 = -\alpha_2$, $\gamma_4 = -\alpha_4$ and $\gamma_5 = -\alpha_5$. If the data generating process is characterized by a combination of social learning and strategic complementarities, then we should observe that:

$$\frac{\partial \Pr(y_{it+1} = 1 | z_{it} = 0)}{\partial A_{it}} > \frac{\partial \Pr(y_{it+1} = 1 | z_{it} = 1)}{\partial A_{it}} > 0$$

Estimating model (7.14) allows us to test this as well.

In our empirical implementation, we only observe social network activity taking place over the phone. We do not observe other forms of social interaction. This nonetheless does not invalidate the application of the above model. First, we study the diffusion of a service only available on mobile phones. It is therefore reasonable to assume that information transmission or network effects are more relevant – and thus more likely to occur – with individuals with whom one interacts over the phone. This is true even if phone interactions are complemented by face-to-face exchanges.

Second, the phone interaction network that we observe is embedded into the denser network of social interactions – i.e., if two individuals interact on the phone, they are by definition interacting socially. This has beneficial implications for identification. Fafchamps, Goyal and van der Leij (JEEA 2010) offer an elaborate treatment of the question of the embeddedness of the observed network into a broader network of acquaintances. Their logic is the following. They observe the co-authorship network between economists and they wish to test whether two individuals i and j who have never coauthored before are more likely to coauthor if their respective past coauthors start to collaborate. This is then extended to the coauthors of coauthors, etc. They argue that, because of embeddedness, social distance in the coauthor network is an upper bound on social distance in the denser network of social interactions. Consequently, if the upper bound falls (distance falls in the coauthor network) then the average distance in the social network also falls. Applied to our setting this means that the interactions that we observe – e.g., i calling j or receiving airtime from j – are a subset of all the interactions between i and j. The key here is that if we do observe an interaction in phone network, then certainly an interaction took place in the larger network of social acquaintances since it contains the phone network.

It follows that if there are interactions in the social network that are not observed in the phone network, and these additional interactions are uncorrelated with those in the phone network from the point of view of information diffusion/network effects, then they simply enter the error term. These interactions create noise that reduces the precision of our estimates, but the dataset is large enough to cope with this problem. If additional interactions in the social network are correlated with interactions in the phone network in terms of their information/network effects, then our estimated coefficients capture the joint effect of both types of social interactions, which is ideal for us. Either way, our estimation approach is robust to unobserved social interactions.

8. Appendix B. Oster's (2019) approach for assessing bias posed by unobservable selection

Oster (2019) shows how the size of the bias posed by unobservable selection, under certain assumptions, can be inferred from coefficient and R-squared differences across models with different sets of control variables. Adopting Oster's notation, let the parameter δ denote the proportional selection relationship. If unobservable and observable factors are equally related to treatment, $\delta = 1$; if unobservable are more strongly related to treatment than observable factors, $\delta > 1$; and if observable factors are more strongly related to treatment than observables, $\delta < 1$. Further, let $R_{\max}$ denote the R-squared from a hypothetical regression of the dependent variable on the treatment variable and the observable and unobservable determinants of the dependent variable. For a model that is linear in a single treatment variable, Oster shows how the bias on the treatment coefficient obtained from a regression where observable but not unobservable factors are included can be written as approximately equal to $\delta \left[\beta^0 - \widetilde{\beta}\right] \frac{\left[R_{\max} - \widetilde{R}\right]}{\widetilde{R} - R^0}$, where β^0 denotes the coefficient resulting from the short regression of the dependent variable on the treatment variable with observable control variables excluded; R^0 is the R-squared from the short regression; $\widetilde{\beta}$ is the coefficient resulting from the regression with observable control variables included, and $\widetilde{R}$ is the R-squared from that regression. Clearly, the bias in $\widetilde{\beta}$ can be severe if: unobservable factors are strongly related to treatment (in which case δ is high); if the treatment coefficient changes considerably as a result of the addition of observable control variables (in which case $\left[\beta^0 - \widetilde{\beta}\right]$ is high) while at the same time the R-squared doesn't move much (in which case $\widetilde{R} - R^0$ is low); and/or if the unobservable factors (would) have considerable explanatory power (in which case $R_{\max} - \widetilde{R}$ is high). Of course, neither δ nor $R_{\max}$ is observable, but the bias formula above is nevertheless useful as it enables researchers to quantify the bias for specific values of δ and $R_{\max}$. Clearly, if there is no movement in the treatment coefficient as we move from the short regression to the regression with observable controls included, Oster's framework implies that there is no bias, regardless of the values of δ and $R_{\max}$.

Table B1
First Adoption: Robustness to selection on unobservables

	(1) Linear model: Uncontrolled	(2) Linear model: Controlled	(3) Bias adjusted β	(4) Bias adjusted β	(5) δ for $\beta=0$
ΔA(it)	0.0056	0.00549	0.0053	0.0051	0
s.e.	0.00031	0.00032	--	--	--
R-squared	0.005	0.0427			
δ			1.0	2.0	6.167
R_{max}			0.085	0.085	0.085
Controls					
ΔS(it)^2	Y	Y			
Tower	N	Y			
year x month x district	N	Y			
Observations	96,266	96,266			

Note: Columns (1) and (2) show results for a linear specification of the form Δy(i,t+1) = β*ΔA(it) + controls + Δε(i,t+1). Standard errors are clustered at the district level (M=27). Columns (3)-(4) show bias-adjusted estimates of β, based on the approach developed by Oster (2019). Column (5) shows the value of δ for which β = 0, again based on Oster (2019). Oster's approach is not suitable for specifications where the potentially endogenous explanatory variable enters nonlinearly (as in Table 2), hence we consider linear specifications for the analysis of robustness to selection on unobservables.