Adoption with Social Learning and Network Externalities*

MARCEL FAFCHAMPS[†], MÅNS SÖDERBOM[‡] and MONIQUE VAN DEN BOOGART[£]

[†] *Freeman Spogli Institute for International Studies, Stanford University, 616 Jane Stanford Way, Stanford CA 94305, USA (e-mail: fafchamp@stanford.edu)*

[‡] *Department of Economics, University of Gothenburg, P.O. Box 640, SE 405 30 Gothenburg, Sweden (e-mail: mans.soderbom@economics.gu.se)*

[£] *Uber, 1455 Market St #400, San Francisco, CA 94103, USA (e-mail: moniquevandenboogaart@gmail.com)*

December 2021

Abstract

We examine patterns of adoption of a new airtime transfer service over time. We document a strong positive correlation between own adoption and increased usage of the new airtime transfer platform by social neighbors. We examine the possible sources of this correlation by distinguishing between network externalities that extend after adoption and social learning that stops after adoption. We find no correlation between usage by social neighbors and own usage after first adoption. We conclude that social learning about the existence and quality of a new product platform are important mechanisms in its diffusion.

JEL classification numbers: D83, O12.

Keywords: Social learning, network externalities, airtime transfers.

Word count: 10,606.

*We have benefitted from comments from two anonymous referees, 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 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. Correspondence to Måns Söderbom at mans.soderbom@economics.gu.se .

I. 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 et al., 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, known as ME2U, using a large administrative dataset from a monopolistic telecommunication operator in Rwanda.[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

[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 usage 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 generally apply to the adoption of a new service.

with others 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 (e.g., Bala and Goyal, 1998). The main difference with the first type of social learning is that 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 usage (Saloner and Shepard, 1995; Jackson and Yariv, 2005; Vega-Redondo, 2007). If usage by my social neighbors increases my incentive to use, 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. The main difference with social learning is that network externalities do not wear off: they continue to reinforce usage 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 -- some bad -- using the product protects me against 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 also install a burglar alarm; in equilibrium, everyone incurs the cost of having a burglar alarm but it no longer displaces burglars towards neighbors (Jackson, 2008).

In this paper we seek to identify the roles of social learning for the adoption of a new service offered to mobile phone users. We also examine 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 by including different types of fixed effects to capture unobserved heterogeneity, and by comparing correlation between own usage and neighbor usage during the period leading up to first adoption and after first adoption. We find robust evidence suggestive of social learning both for the existence and the quality of the product. In contrast, we find that increased usage by social neighbors is not correlated with own usage after first adoption. This confirms our interpretation of the positive correlation between first adoption and neighbor usage as evidence of social learning.

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 separately identify the role of social learning in the diffusion process, as distinct from network externalities. We also 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). The information role 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 positive association between own use and neighbors' use is limited in time to the first adoption.

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

II. Testing strategy and identification

A formal presentation of our conceptual framework is presented in detail in Online Appendix Section A. It builds on earlier work by Young (2009) and others (e.g., Jackson, 2008; Kreindler and Young, 2014; Arieli et al. 2020). The intuition behind it can be summarized as follows. 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, once *i* has used the product, recent usage by network neighbors should no longer predict *i*'s own use. Suppose instead that network effects are entirely driven by strategic complementarities in usage. In this case, own usage will continue to co-vary with neighbor usage after first adoption and there should be no difference in this covariation before and after adoption.

[3] The Online Appendix for this article is available here: www.soderbom.net/adoption/adoptionpaper_appx.pdf .

This observation is the basis of our testing strategy: excess covariation between own usage and neighbor usage at the time of adoption is indicative of social learning -- provided possible confounds are controlled for. Another way of stating this is to say that correlation in usage after adoption serves as a placebo: in those observations, social learning is no longer at play and thus there should not be correlation in usage due to social learning. But there will be correlation due to other factors, such as network externalities, correlated tastes resulting from friends' self-selection, and correlated network shocks. It follows that the level of correlation in usage after adoption provides a measure of the combined effect of all these factors. This is what enables us to identify the separate role of social learning.

We also wish to distinguish between 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*. In our data, the researcher observes a signal $M_{it} = 1$ that, at time *t*, individual *i* receives unambiguous information about the product's existence, even though *i* has never used the product yet; $M_{it} = 0$ otherwise. With this signal, it is 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 its existence, 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 accumulates information that can help *i* decide whether to adopt the product or not. The two models of social learning also make different predictions regarding the concavity of the correlation between own adoption and neighbors' usage, predictions that we compare to the data as source of additional evidence. These predictions are not causal, they are just features of the correlation structure that arise in the presence of a certain type of social learning. The fact that the two learning models make different predictions makes it possible to test one against the other.

We now discuss the logic of our empirical identification strategy in more detail. We start by emphasizing that we are not seeking to identify a causal effect of neighbor *j*'s adoption on *i*'s adoption. Much of the literature on network diffusion draws on epidemiological models of contagion in which *j* transmits an illness to *i* through social contact, thereby mechanically `causing' *i* to become ill. In our conceptual framework, we allow *i* to be a rational agent who makes an optimal decision based on the information *i* receives or seeks actively. Even if *i* asks *j* to use ME2U to send him airtime to ascertain the veracity of the service, out test is still valid: the transfer from *j* to *i* serves as a demonstration of the existence of the product. Our testing strategy revolves around comparing *correlation* levels in usage at different points in time, not in identifying causation -- which is always highly problematic in social network data (e.g., Manski, 1993).

The first difference we use for identification is that between correlation in usage *before* and *after* first adoption. After first adoption, learning about the product's existence is no longer an issue and, in our conceptual framework, *i* has accumulated enough positive information about the product to use it. It follows that correlation in usage after first adoption reflects a whole range of factors that we do not seek to disentangle. This includes network externalities creating strategic complementarity -- or strategic substitutes -- in usage, but also shocks to usage that are correlated among socially proximate individuals. For instance, social neighbors may share an attachment to a particular region of Rwanda, and when this region is affected by a negative shock they send airtime to their relatives there (e.g., Blumenstock et al., 2016; Batista and Vicente, 2020). We regard correlation in usage after first adoption as measuring the baseline level of correlation in usage when social learning considerations are absent or negligible. Identification of social learning is achieved by comparing correlation in usage before and after first usage: if correlation is higher before than after, it indicates the presence of social learning. We then compare the propensity to adopt before and after receiving the first ME2U transfer to disentangle learning about the existence of the product from learning about its unobservable qualities, such as reliability. Thus, we do not (and do not seek to) identify network externalities separately from correlated effects. Our identification also does not require that usage by neighbors have an impulse-like diffusion effect on *i*'s adoption: our testing strategy assumes that *i* ultimately decides whether to adopt or not based on available information, some of which *i* may seek out directly from social contacts. Recent usage by neighbors only serves to capture the amount of information about the product that is available from them.

For this strategy to work, a number of possible confounds must be controlled for that could induce a higher correlation in usage at the time of *i*'s first adoption. The first -- and potentially most important -- confound is the existence of trends in the data: our data covers the spread of a new service throughout the country, and this mechanically creates a correlation between likelihood of adoption by *i* and usage by others: as more people start using ME2U, more people adopt it, i.e., adoption follows a logistic curve (e.g., Bass, 1969; Björkegren, 2019). To correct for this, we include a large number of time dummies in the estimation and we rely on the fact that adoption is staggered over a long period of time. Because eventual adopters do not all adopt around the same time, this means that adoption is observed at many different points in time, and a different time dummy controls for any common trend at each of these points in time.

Secondly, people's time-invariant tastes and proclivities may be correlated with those of their neighbors, e.g., because of homophily in the self-selection of friends and social neighbors. This includes people's propensity to adopt new things and the utility they may derive from ME2U. To address this concern, we remove individual fixed effects by first-differencing our regressions of interest. This means that identification only relies on variation in the usage of ME2U by neighbors over time -- not on the absolute level of their usage. Third, *i*'s decision to adopt at time *t* may in return affect *j*'s own usage at *t* -- e.g., because *i*'s adoption provide *j* with information about ME2U. To avoid this type of reverse feedback, we lag the value measuring neighbor usage instead of using its contemporaneous value.

Fourth, neighbors' tastes -- and thus their propensity to use and adopt ME2U -- may vary in ways that are correlated, either because of a common local trend (e.g., the product naturally spreads faster in a particular area) or because of location-specific common shocks (e.g., a marketing campaign targeting a particular geographical area through billboards or radio ads). We control for these factors by introducing a very large number of time dummies interacted with location fixed effects. This soaks up local trends as well as local marketing shocks. Because we observe first adoption by different individuals over a very long period of time, it is unlikely that these shocks would coincide only with the timing of adoption. Even if they did, however, as long as they continue to occur after first adoption, they are included in subsequent correlation in usage, which we net out to identify the pre-adoption increase in correlation due to social learning. The last possible confound is the existence of marketing campaigns, such as viral marketing, that target individuals not by geographical location but via their social network. To the extent that these campaigns target ME2U users as seeds for information diffusion, they are subsumed by our conceptual framework as another form of social learning. Although we have no information suggesting that such campaigns were used in Rwanda during our study period, we nonetheless provide evidence regarding this possibility using Oster's (2019) method, as detailed in the empirical section.

To summarize, identification is achieved by comparing correlation in usage between *i* and his or her neighbors at different points of *i*'s adoption process. We are not identifying or measuring causal effects in the usual sense. Instead, we allow for a certain level of baseline correlation to be present, due to network externalities, correlated effects, or any other source. Our tests of social learning rely entirely on the presence or absence of an excess correlation -- above that default level -- around the time of first usage.

III. 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.[4]

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 unambiguously inform the recipient that peer-to-peer airtime transfers are in existence.[5] 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 ease-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.

[4] There is some evidence that a small number of subscribers used airtime transfers to retail airtime that they bought at a discount. We have therefore dropped all observations that were above the 99th percentile for total amount of calls, amount of airtime sent or amount of airtime received.

[5] 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.

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. These network effects would naturally continue to manifest themselves 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.

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 under conditions of strict confidentiality.[6] This is a large dataset comprising multiple computer-generated administrative files. Using this dataset, Blumenstock et al. (2010) and Blumenstock and Eagle (2012) document that phone owners in Rwanda over the period covered by the data are considerably wealthier, better educated, and more predominantly male than the general population. Further evidence that phone owners are wealthier than the average Rwandan is provided by Blumenstock et al. (2015) and Blumenstock (2018).

We use two main bodies of data for our analysis: data on airtime transfers; and data on phone calls. The former data are used to study first adoption and diffusion; the latter are used

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

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.[7] The data on transfers was used by Blumenstock et al. (2016) to show that individuals make transfers and calls to people affected by disasters.

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 are 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).[8] 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).[9] To avoid spurious

---

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

[8] 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.

[9] 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 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 happens to want airtime and has enough trust to engage in a

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 cell towers, and thus (crudely) track users' movements in the country. We use this information to control for location in the econometric analysis below.

Variable definition

transaction. The introduction of mobile money moved airtime beyond its role of occasional and impractical barter currency.

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. 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, which eliminates the risk of simultaneity bias since actual usage of ME2U by individual i in week t could not have caused previous usage by network neighbors. 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_{it}, 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. 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 and j stay connected during the span of our analysis. The network ties are thus defined as:

$$g_{ijt} = \begin{cases} 1 & \text{if } i \,\&\, j \text{ had their first phone-based contact in period } s,\ s=t_i,\ldots,t \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

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

$$N_{it}(g) = \{j : g_{ijt} = 1\} \quad (2)$$

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 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, $S_{it} = t - t_i$.

Regression models

[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.

Our regression analysis is organized around three regression models. Regression (3) investigates the shape of the correlation between adoption by *i* and past adoption by *i*'s neighbors. It corresponds to the reduced form of models (A1) and (A7) in Online Appendix Section A.[11]

$$\Pr\left(y_{i,t+1} = 1 \middle| \left\{y_{i,t_i}, \dots, y_{it}\right\} = \{0, \dots, 0\}\right) = \cdots$$

$$= \alpha_i + \alpha_1 S_{it} + \alpha_2 A_{it} + \alpha_3 S_{it}^2 + \alpha_4 A_{it}^2 + \alpha_5 S_{it} A_{it} + controls + \varepsilon_{i,t+1} \qquad (3)$$

By itself regression (3) does not demonstrate the presence of social learning because estimates will combine learning mechanisms with the baseline correlation between own usage and past usage by social neighbors. But it allows us to examine the shape of the relationship between adoption and network usage, which is in itself informative about social learning -- see Online Appendix Section A for details.

Regression (4) separates the pre-period adoption into the period before *i* receives his/her first ME2U transfer, and the period after that:

$$\Pr\left(y_{i,t+1} = 1 \middle| \left\{y_{i,t_i}, \dots, y_{it}\right\} = \{0, \dots, 0\}\right) = \cdots$$

$$= \alpha_i + \alpha_1 S_{it} + \alpha_2 A_{it} + \alpha_3 S_{it}^2 + \alpha_4 A_{it}^2 + \alpha_5 S_{it} A_{it} + \dots$$

$$\dots + \beta_0 m_{it} + \beta_1 S_{it} m_{it} + \beta_2 A_{it} m_{it} + \beta_3 S_{it}^2 m_{it} + \beta_4 A_{it}^2 m_{it} + \beta_5 S_{it} A_{it} m_{it} + \dots$$

$$\dots + controls + \varepsilon_{i,t+1}. \qquad (4)$$

Its purpose is to test whether the correlation between adoption by *i* and past adoption by *i*'s neighbors falls after *i* receives a transfer. If social learning only relates to the existence of the

[11] Parameter α_i captures variation in product usefulness across individuals. With any social learning we expect adoption or usage by neighbors to be correlated with own adoption, *i*.e., $\partial \Pr\left(y_{i,t+1} = 1\right) / \partial A_{it} > 0$. We also expect adoption to increase with S_{it} because the likelihood of adoption should increase over time as more information about the product becomes available from within and outside the social network. Cross-terms are included to test the concavity of the relationship with respect to S_{it} and A_{it} as predicted by social learning about product existence.

product, then the correlation between i's adoption and A_{it} should fall to its baseline level after m_{it} becomes 1.[12]

Regression (5) combines pre- and post-adoption data on usage by i. The estimated model is of the form:

$$\Pr(y_{i,t+1} = 1) = \alpha_i + \alpha_1 S_{it} + \alpha_2 A_{it} + \alpha_3 S_{it}^2 + \alpha_4 A_{it}^2 + \alpha_5 S_{it} A_{it} + \gamma_0 z_{it} + \cdots$$

$$\ldots + \gamma_1 S_{it} z_{it} + \gamma_2 A_{it} z_{it} + \gamma_3 S_{it}^2 z_{it} + \gamma_4 A_{it}^2 z_{it} + \gamma_5 S_{it} A_{it} z_{it} + \ldots$$

$$\ldots + controls + \varepsilon_{i,t+1}. \qquad (5)$$

where $z_{it} = 1$ if i has already used the product prior to period t, and 0 otherwise. In the absence of social learning, we expect no excess correlation between usage by i and past usage by social neighbors, i.e., we should observe that $\gamma_2 = \gamma_3 = \gamma_4 = \gamma_5 = 0$. If social learning is the only source of network correlation, we should observe that:

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

which is guaranteed if $\gamma_2 = -\alpha_2$, $\gamma_4 = -\alpha_4$, and $\gamma_5 = -\alpha_5$. If both learning and complementarities are present, we should observe:

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

In our data, 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 -- between individuals who interact over the phone. This is true even if phone interactions are complemented by face-to-face exchanges.

[12] This follows from the fact that the reduced form model for (A2) in the Online Appendix Section A is of the form: $\Pr(y_{i,t+1} = 1 | \{y_{i,t_i}, \ldots, y_{it}\} = \{0, \ldots, 0\}, M_{is} = 1 \text{ for some } s \leq t) = \alpha_i + \varepsilon_{i,t+1}$.

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 et al. (2010) offer an elaborate treatment of the question of the embeddedness of an 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. On the other hand, if interactions in the acquaintance network are correlated with interactions in the phone network in terms of their information/network effects, then our estimated coefficients capture the joint influence of both types of social interactions, which is ideal for us. Either way, our estimation approach is robust to unobserved social interactions.

IV. Empirical results

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* has adopted or received their 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.

The first regression we estimate is (3) -- with a few modifications. To eliminate the individual fixed effect α_i and any persistent common shocks, all variables are first differenced.[13] As shown in Online Appendix Section B, first-differencing creates a mechanical negative correlation between $\Delta y_{i,t+1}$ and ΔA_{it} in the presence of correlated effects in the form of contemporaneous or two-period common shocks.[14] To eliminate these correlated effects, we use $\Delta A_{i,t-2}$ in lieu of ΔA_{it}.[15] Since the dependent variable is measured at *t*+1, this implies that a three-period lag is used. Since adoption by neighbors proxies for information on the existence and quality of ME2U, lagging the adoption variable preserves

[13] Fafchamps et al. (2010) estimate a model similar to regression (6) 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 our regressions 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 (6) after detrending all first-differenced regressors. Results show absolutely no change in coefficient estimates and standard errors.

[14] This negative correlation arises in the presence of a common shock c_{t+1} affecting $y_{i,t+1}$ and $A_{i,t+1}$ in the same period. After first differencing both variables, and lagging ΔA_{it}, we have $\Delta y_{i,t+1}$ being correlated with $c_{t+1} - c_t$ while ΔA_{it} is correlated with $c_t - c_{t-1}$. Hence $\Delta y_{i,t+1}$ and ΔA_{it} become negatively correlated due to the presence of common shock c_t appearing with opposite signs. Using $\Delta A_{i,t-1}$ instead eliminates this problem. By extension, using $\Delta A_{i,t-2}$ also eliminates any mechanical correlation arising from two-period common shocks. See Online Appendix Section B for details.

[15] For reference, we also report in the Online Appendix Section D estimation results using $\Delta A_{i,t-1}$ (corrects for contemporaneous common shocks only) and $\Delta A_{i,t-3}$ (corrects for three-period common shocks).

this interpretation since information accumulates over time. The estimated model is a thus linear probability model expressed in first-differences:

$$\Delta y_{i,t+1} = \alpha_1 + \alpha_2 \Delta A_{i,t-2} + \alpha_3 \Delta S_{it}^2 + \alpha_4 \Delta A_{i,t-2}^2 + \alpha_5 \Delta\left(S_{it} A_{i,t-2}\right) + \cdots$$

$$+ controls + \Delta\varepsilon_{i,t+1} \qquad (6)$$

where: $\Delta X_{it} = X_{it} - X_{i,t-1}$ by definition of notation; only observations up to the first adoption by *i* are used; and *controls* is a set of control variables.[16] By construction, $\Delta S_{it} = 1$, so this variable cannot be included as a regressor. In all regressions, standard errors are clustered at the district level.

The purpose of estimating (6) is to document whether there is positive correlation between *i*'s adoption and the past usage of *i*'s neighbors, and what shape this correlation has. We later compare it with the same correlation measured after adoption, to net out baseline correlation in usage and identify social learning. Coefficient estimates of (6) are presented in Table 2. Specification [1] contains no control variables and serves as benchmark. We see that α_2 and α_4 are significantly negative while α_5 is significantly positive. The relationship between $A_{i,t-2}$ and the probability of first adoption is thus nonlinear and depends 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 neighbors' usage should be strongly concave with respect to $A_{i,t-2}$. In contrast, when social learning is about product quality, this concavity need not be present and may even be reversed. Marginal 'effects' $\partial \Pr\left(y_{i,t+1} = 1\right) / \partial A_{i,t-2}$ evaluated at various values of $A_{i,t-2}$ are shown below the first-differences (FD) estimates in Table 2. We find that $\partial \Pr\left(y_{i,t+1} = 1\right) / \partial A_{i,t-2}$ is positive for $A_{i,t-2} \leq 60$, consistent with the presence of a positive correlation in this range. We observe a

[16] 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 preferred to fixed effects.

gradual fall in $\partial\Pr(y_{i,t+1} = 1)/\partial A_{i,t-2}$ as $A_{i,t-2}$ increases, as suggested by the negative quadratic term coefficient α_4. This evidence is prima facie consistent with social learning about product existence.

Next, we add dummy variables for time, cell tower, and district as controls to regression (6). This yields specification [2] in Table 2. The control variables have some explanatory power, as can be seen from the increase in the R^2. But the coefficients of interest and their significance levels change little. We still observe a gradual fall in $\partial\Pr(y_{i,t+1} = 1)/\partial A_{i,t-2}$ as $A_{i,t-2}$ increases, which is prima facie consistent with social learning about product existence -- except that the magnitude of the estimated marginal effect is larger. Expanding the set of control variables further to allow for district-specific time fixed effects in specification [3] leads to a small increase in the R^2, but only results in trivial changes to the coefficient estimates of interest.

As explained in Section 2, the dummies added in specification [2] and [3] aim to control for a variety of correlated shocks among social neighbors. The fact that the estimated coefficients of interest change little between specifications suggest that they may also be robust to other correlated shocks that we do not observe. This intuitive idea, initially formalized by Altonji et al. (2005), was developed by Oster (2019) who shows how, under certain assumptions, the size of the omitted variable bias can be inferred from coefficient and R^2 differences across models with a widening set of control variables. Online Appendix Section C contains a short summary of Oster's method and defines the relevant parameters. We now use Oster's (2019) method to formally assess the sensitivity of our coefficients of interest to omitting unobserved shocks from the analysis. Since the method only applies to linear models, we drop $\Delta A_{i,t-2}^2$ and $\Delta(S_{it}A_{i,t-2})$ from the specification and estimate a regression of the form $\Delta y_{i,t+1} = a_1 + a_2\Delta A_{i,t-2} + a_3\Delta S_{it}^2 + controls + \Delta\varepsilon_{i,t+1}$. This is equivalent to focusing on the average value of our object of interest, $\partial \Pr(y_{i,t+1} = 1)/\partial A_{i,t-2}$.

We start by noting that going from a regression without controls (Table 2, column 1) to one with some controls (column 2) raises the average value of $\partial \Pr(y_{i,t+1} = 1)/\partial A_{i,t-2}\}$, which indicates a negative correlation between unobservables and the regressor of interest. Applying the Oster method to this case would imply that, if anything, omitting unobservables biases the coefficient of interest downwards. There is, however, a small reduction in $\partial \Pr(y_{i,t+1} = 1)/\partial A_{i,t-2}$ between columns (2) and (3) of Table 2. We therefore apply the Oster method to that case. Results are shown in Online Appendix Table C1. The estimated coefficient of $\Delta A_{i,t-2}$ is equal to 0.0041 in the partially controlled model and 0.0039 in the model with our full set of controls. These estimates are marginally higher than the average $\partial \Pr(y_{i,t+1} = 1)/\partial A_{i,t-2}$ reported in Table 2. Let $R_{\max}$ denote the R^2 from a hypothetical regression of the dependent variable on the observable and unobservable determinants of the dependent variable. Then, under the assumption that unobservable and observable factors are equally related to $\Delta A_{i,t-2}$ (i.e., $\delta = 1$ in Oster's framework) and that $R_{\max}$ is twice the value of R^2 in the model with all observables, the bias-adjusted estimate of a_2 is 0.0031 (Table C1, col. [3]). If we want to be more conservative and set $\delta = 2$, the bias-adjusted estimate of a_2 is 0.0023 (col. [4]). The unobservables would have to be more than four times more important than the observables in determining $\Delta A_{i,t-2}$ for the bias-adjusted coefficient of $\Delta A_{i,t-2}$ to be equal to zero, given our assumed value for $R_{\max}$. This is a very conservative estimate since, without any controls, the estimate of a_2 is actually lower than that reported in column (1) of Online Appendix Table C1.

Results for regression (4) are presented in Table 3. As for Table 2, the individual fixed effect α_i is eliminated by first-differencing the data and the key variable of interest $\Delta A_{i,t-2}$ is double-lagged to net out correlated effects. The estimated model is thus a differenced linear probability model of the form:

$$\Delta y_{i,t+1} = \alpha_1 + \alpha_2 \Delta A_{i,t-2} + \alpha_3 \Delta S_{it}^2 + \alpha_4 \Delta A_{i,t-2}^2 + \alpha_5 \Delta\left(S_{it} A_{i,t-2}\right) + \ldots$$

$$\ldots + \beta_1 \Delta(S_{it} m_{it}) + \beta_2 \Delta\left(A_{i,t-2} m_{it}\right) + \beta_3 \Delta(S_{it}^2 m_{it}) + \beta_4 \Delta\left(A_{i,t-2}^2 m_{it}\right) + \ldots$$

$$\ldots + \beta_5 \Delta\left(S_{it} A_{i,t-2} m_{it}\right) + controls + m_{it} \times controls + \Delta\varepsilon_{i,t+1} \qquad (7)$$

where, as in (6), we only include observations up to the first adoption.[17] All control variables are interacted with m_{it}, as denoted by $m_{it} \times controls$.

Similar to Table 2, our estimates and significance levels of interest hardly change when adding controls.[18] Table 3 presents estimates of the average $\partial \Pr\left(y_{i,t+1} = 1\right)/\partial A_{i,t-2}$ evaluated separately at $m_{it} = 0$ and $m_{it} = 1$, that is, before and after receiving a ME2U transfer. We see that the average $\partial \Pr\left(y_{i,t+1} = 1\right)/\partial A_{i,t-2}$ remains significant throughout and is significantly larger when $m_{it} = 1$ than when $m_{it} = 0$. This confirms that, in this context, social learning circulates information about the existence of the new service and may also circulate relevant information about the quality or usefulness of the service. Whether or not the latter effect is present depends on how $\partial \Pr\left(y_{i,t+1} = 1\right)/\partial A_{i,t-2}$ varies before and after adoption by *i*: if $\partial \Pr\left(y_{i,t+1} = 1\right)/\partial A_{i,t-2}$ falls to zero after *i* has made a transfer with ME2U, it is reasonable to conclude that social learning extends to information about the quality of the product -- and that, in our context, social learning essentially disappears after first use.

To this effect, we now estimate regression model (5) -- again in first differences to eliminate unobserved heterogeneity α_i and with $\Delta A_{i,t-2}$ double-lagged to eliminate correlated effects:

---

[17] We exclude the observations for which $\Delta m_{it} = 1$ from the estimation sample. This procedure ensures that, provided a full set of $m_{it} \times controls$ are included, the results from regression (7) will be equivalent to results obtained from separate regressions of $\Delta y_{i,t+1}$ on $\Delta A_{i,t-2}, \Delta S_{it}^2, \Delta A_{i,t-2}^2, \Delta\left(S_{it} A_{i,t-2}\right)$ and $controls$, for the subsamples with $m_{it} = 0$ and $m_{it} = 1$. Hence, Δm_{it} is not included as a regressor in (7).

[18] Additional results from a robustness analysis using Oster's (2019) approach, available on request, again yields a very small bias adjustments for reasonable values of δ and $R_{\max}$.

$$\Delta y_{i,t+1} = \alpha_1 + \alpha_2 \Delta A_{i,t-2} + \alpha_3 \Delta S_{it}^2 + \alpha_4 \Delta A_{i,t-2}^2 + \alpha_5 \Delta\left(S_{it} A_{i,t-2}\right) + \ldots$$

$$\ldots + \gamma_1 \Delta(S_{it} z_{it}) + \gamma_2 \Delta\left(A_{i,t-2} z_{it}\right) + \gamma_3 \Delta(S_{it}^2 z_{it}) + \gamma_4 \Delta\left(A_{i,t-2}^2 z_{it}\right) + \ldots$$

$$\ldots + \gamma_5 \Delta\left(S_{it} A_{i,t-2} z_{it}\right) + controls + z_{it} \times controls + \Delta\varepsilon_{i,t+1} \qquad (8)$$

where z_{it}=1 if subscriber *i* has sent airtime to someone else with ME2U before time *t*. For the estimation of (8), all observations are used – pre-adoption and post-adoption – except those for which Δz_{it}=1 or $\Delta z_{i,t-1}$=1 (i.e., just after first-adoption).[19] All control variables are interacted with z_{it}, which implies that the estimates of the α coefficients based on (8) will be numerically identical to those based on (6). Hence, the coefficients $\gamma_2, \gamma_4, \gamma_5$ determine whether the covariation between own usage and neighbor usage after adoption differs from the covariation before first adoption, which is the basis of our testing strategy.

Coefficient estimates and corresponding estimates of $\partial \Pr\left(y_{i,t+1} = 1\right) / \partial A_{i,t-2}$ are presented in Table 4. The coefficients $\gamma_2, \gamma_4, \gamma_5$ are all statistically significant, and the average $\partial \Pr\left(y_{i,t+1} = 1\right) / \partial A_{i,t-2}$ is much lower after first adoption -- slightly negative in fact – than before adoption. The difference in $\partial \Pr\left(y_{i,t+1} = 1\right) / \partial A_{i,t-2}$ is statistically significant. Hence, social learning matters: the correlation between $y_{i,t+1}$ and $A_{i,t-2}$ during the period leading up to adoption is not just due to network externalities. Combined with earlier results from Table 3, the evidence is thus suggestive of a hybrid model in which social learning serves two purposes: circulating information about product existence, and about product quality. That the average $\partial \Pr\left(y_{i,t+1} = 1\right) / \partial A_{i,t-2}$ remains larger after $m_{it} = 1$ than after $z_{it} = 1$ further indicates that, of the two, diffusing information about quality accounts for a significant share of social learning.[20]

[19] We exclude the observations in the period immediately after first adoption since, under two-period common shocks, $\Delta y_{i,t+1}$ in this period likely correlates with the common shocks that drive first adoption.

[20] 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. A reasonable proxy for frequency of contact is the total number of phone calls made by and received by *i* in a given week. If we add this variable (dated t-2 and expressed in first differences) to the set of explanatory variables in regression (4), we obtain a positive but not

What is less anticipated is that, after first adoption, $\partial \Pr(y_{i,t+1} = 1)/\partial A_{i,t-2}$ is on average negative, suggesting that, if anything, airtime transfers are strategic substitutes across network neighbors. Network externalities are typically taken as synonymous to strategic complement. 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, 2008; Bramoullé et al., 2014) and evidence of strategic substitutes in networks has been provided in the case of the adoption of business practices (e.g., Fafchamps and Söderbom, 2014). This result, however, disappears once we include control variables (see Table 4, specifications [2]-[3]), so we should perhaps not pay it too much attention. Whatever the reason for the slightly negative but mostly non-significant $\partial \Pr(y_{i,t+1} = 1)/\partial A_{i,t-2}$ after adoption, the main lesson is that, prior to first adoption, networks serve an important social learning role, and that social learning includes learning about the existence of the product as well as its quality.

As robustness check, we estimate the regressions in Tables 2, 3, and 4 using $\Delta A_{i,t-1}$ instead of $\Delta A_{i,t-2}$. As demonstrated in Appendix B, using $\Delta A_{i,t-1}$ eliminates contemporaneously correlated effects, but not two-period common shocks. If the results obtained using $\Delta A_{i,t-1}$ are similar to those obtained with $\Delta A_{i,t-2}$, this would confirm our findings while suggesting that two-period common shocks are not an issue. But if $\partial \Pr(y_{i,t+1} = 1)/\partial A_{i,t-1}$ is found to be significantly negative after first adoption, this could signal the presence of two-period common shocks. Results from regressions with $\Delta A_{i,t-1}$ are shown in Online Appendix Section D. The results in Tables D.1 and D.2 are qualitatively similar to the corresponding Tables 2 and 3, except that the estimated marginal effects

quite statistically significant correlation with first adoption. More importantly, we find that adding this variable proxying for frequency of contact does not affect the test of interest: $\partial \Pr(y_{i,t+1} = 1)/\partial A_{i,t-2}$ remains positive and statistically significant even when $m_{it} = 1$.

$\partial \Pr(y_{i,t+1} = 1)/\partial A_{i,t-1}$ are somewhat larger. The main difference arises in Table D.4, where we find a significantly negative $\partial \Pr(y_{i,t+1} = 1)/\partial A_{i,t-1}$ after first adoption by i. While we cannot entirely rule out the possibility that it captures strategic substitution in ME2U usage, the fact that this negative correlation disappears when using $\Delta A_{i,t-2}$ is also consistent with the possible presence of two-period common shocks. To ensure that this possible source bias does not affect our results, we prefer the more conservative estimates presented in Tables 2, 3 and 4. We have also estimated our regressions using $\Delta A_{i,t-3}$ (i.e. a four-period lag). The results, which are shown in Tables D.4-D.6 in the Online Appendix, are consistent with the main results reported above: prior to first adoption, network adoption correlates positively with own adoption; after first adoption, the correlation between network adoption and own usage is either zero or slightly negative.

IV. Conclusions

This study is based on a large administrative 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.

Our testing strategy revolves around comparing correlations in usage before and after first adoption. As discussed above, for this strategy to work, a number of possible confounds must be controlled for that could induce a higher correlation in usage at the time of *i*'s first adoption. We control for trends, time-invariant taste differences, reverse feedback, and local correlated shocks to the benefits of using the technology. We also show that our identification strategy is robust to the presence of unobserved shocks that are common to everyone in *i*'s network, provided these do not extend for more than two periods. While we cannot rule out the possibility that unobserved common shocks are even more persistent, it should also be noted that the inclusion of fixed effects goes some way towards controlling for highly persistent unobserved shocks. Overall, we think it is plausible to argue that our identification is robust to potentially relevant unobserved common shocks.

We find robust evidence suggestive of social learning for both the existence and the reliability or usefulness of the new service. In contrast, we find that the correlation between own use and usage by network neighbors essentially disappears after first adoption. The positive network correlation observed in the adoption decision can thus be attributed to social learning, and is not driven by strategic complementarities in usage. All regression results reported in the paper are based on specifications in which neighbor adoption is lagged by three periods. We obtain qualitatively similar results if we use a lag length of two, or four, periods.

Our results thus provide useful insights into the process by which products and services 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.

References

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

Aker, J.C. and Fafchamps, M. (2015). 'Mobile Phone Coverage and Producer Markets: Evidence from West Africa', *World Bank Economic Review*, Vol. 29, pp. 262-292.

Aker, J.C., Collier, P., and Vicente, P.C. (2017). 'Is Information Power? Using Mobile Phones and Free Newspapers during an Election in Mozambique', *Review of Economics and Statistics*, Vol. 99, pp. 185-200.

Altonji, J., Elder, T., and Taber, C.R. (2005). 'An Evaluation of Instrumental Variable Strategies for Estimating the Effects of Catholic Schooling', *Journal of Human Resources*, Vol. 40, pp. 791-821.

Arieli, I., Babichenko, Y., Peretz, R., and Young, H.P. (2020). 'The Speed of Innovation Diffusion in Social Networks', *Econometrica*, Vol. 88, pp. 569-594.

Bala, V. and Goyal, S. (1998). 'Learning from Neighbours', *Review of Economic Studies*, Vol. 65, pp. 595-621.

Batista, C., and Vicente, P.C. (2020), 'Adopting Mobile Money: Evidence from an Experiment in Rural Africa', *AEA Papers and Proceedings*, Vol. 110, pp. 594-598.

Bass, F. (1969). 'A New Product Growth for Model Consumer Durables', *Management Science* Vol. 15, Theory Series, pp. 215-227.

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

Blumenstock, J.E., Gillick, D., and Eagle, N. (2010). 'Who's calling? Demographics of mobile phone use in Rwanda', *Transportation* Vol. 32, pp. 2-5.

Blumenstock, J.E. and Eagle, N. (2012). 'Divided we call: disparities in access and use of mobile phones in Rwanda', *Information Technologies & International Development*, Vol. 8, pp. 1-16.

Blumenstock, J.E., Cadamuro, G., and On, R. (2015). 'Predicting Poverty and Wealth from Mobile Phone Metadata', *Science*, Vol. 350, pp. 1073-1076.

Blumenstock, J.E. (2018). 'Estimating economic characteristics with phone data', *AEA Papers and Proceedings* Vol. 108, pp. 72-76.

Blumenstock, J.E., Eagle, N. and Fafchamps, M. (2016). 'Airtime Transfers and Mobile Communications: Evidence in the Aftermath of a Natural Disaster', *Journal of Development Economics*, Vol. 120, pp. 157-181.

Bramoullé, Y., Kranton, R. and D'Amours, M. (2014). 'Strategic Interaction and Networks', *American Economic Review*, Vol. 104, pp. 898-930.

Centola, D. (2010). 'The Spread of Behavior in an Online Social Network', *Science*, Vol. 329, pp. 1194-1197.

Cole, S.A. and Fernando, A.N. (2016). 'The Value of Advice: Evidence from the Adoption of Agricultural Practices', unpublished manuscript, Harvard University.

Fafchamps, M., Goyal, S. and van der Leij, M. (2010). 'Matching and Network Effects', *Journal of the European Economic Association*, Vol. 8, pp. 203-231.

Fafchamps, M. and Söderbom, M. (2014). 'Network Proximity and Business Practices in African Manufacturing', *World Bank Economic Review*, Vol. 28, pp. 99-129.

Fafchamps, M. and Minten, B. (2012). 'Impact of SMS-Based Agricultural Information on Indian Farmers', *World Bank Economic Review*, Vol. 26, pp. 383-414.

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

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

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

Jackson, M.O. and Yariv, L. (2005) 'Diffusion on Social Networks', *Economie Publique*, Vol. 16, pp. 3-16.

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

Kreindler, G.E. and Young, H.P. (2014). ‘Rapid Innovation Diffusion in Social Networks’, *PNAS*, Vol. 111(Supplement 3), pp. 10881-10888.

Manski, C.F. (1993). ‘Identification of Endogenous Social Effects: The Reflection Problem’, *The Review of Economic Studies*, Vol. 60, pp. 531-542.

Mobius, M. and Rosenblat, T. (2014). ‘Social Learning in Economics’, *Annual Review of Economics*, Vol. 6, pp. 827-847.

Oster, E. (2019). ‘Unobservable Selection and Coefficient Stability: Theory and Evidence’, *Journal of Business & Economic Statistics*, Vol. 37, pp. 187-204.

Ryan, S. P. and Tucker, C. (2012). ‘Heterogeneity and the dynamics of technology adoption’, *Quantitative Marketing and Economics*, Vol. 10, pp. 63-109.

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

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

Young, H.P. (1999). ‘Diffusion in Social Networks’, Papers 2, Brookings Institution - Working Papers.

Young, H.P. (2009). ‘Innovation Diffusion in Heterogeneous Populations: Contagion, Social Influence, and Social Learning’, *American Economic Review*, Vol. 99, pp. 1899-1924.

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

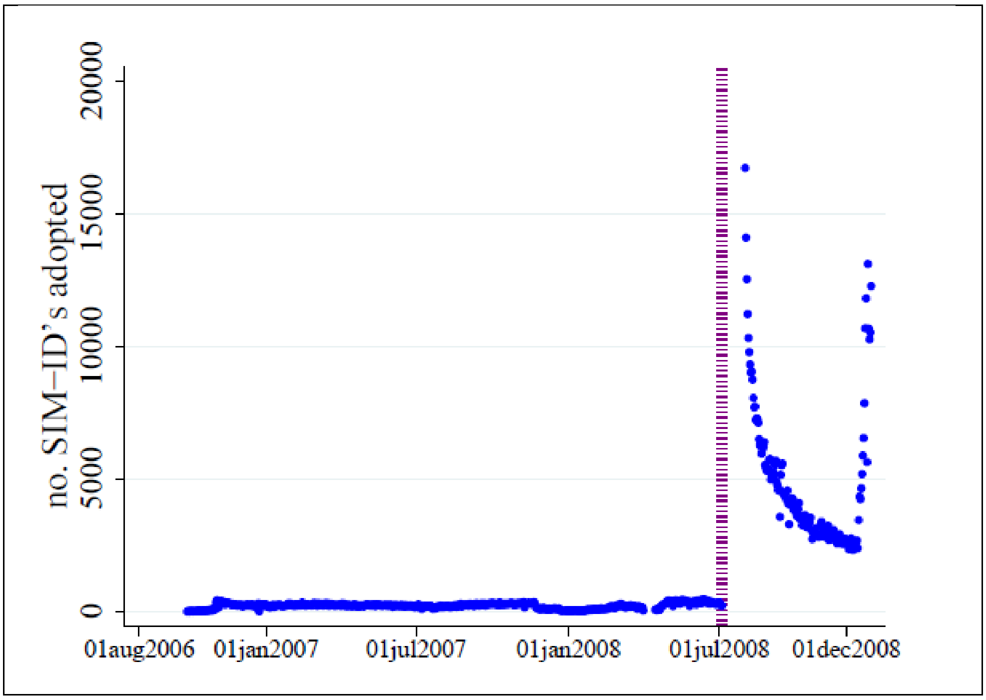

Figure 2. Proportion of adopters and users

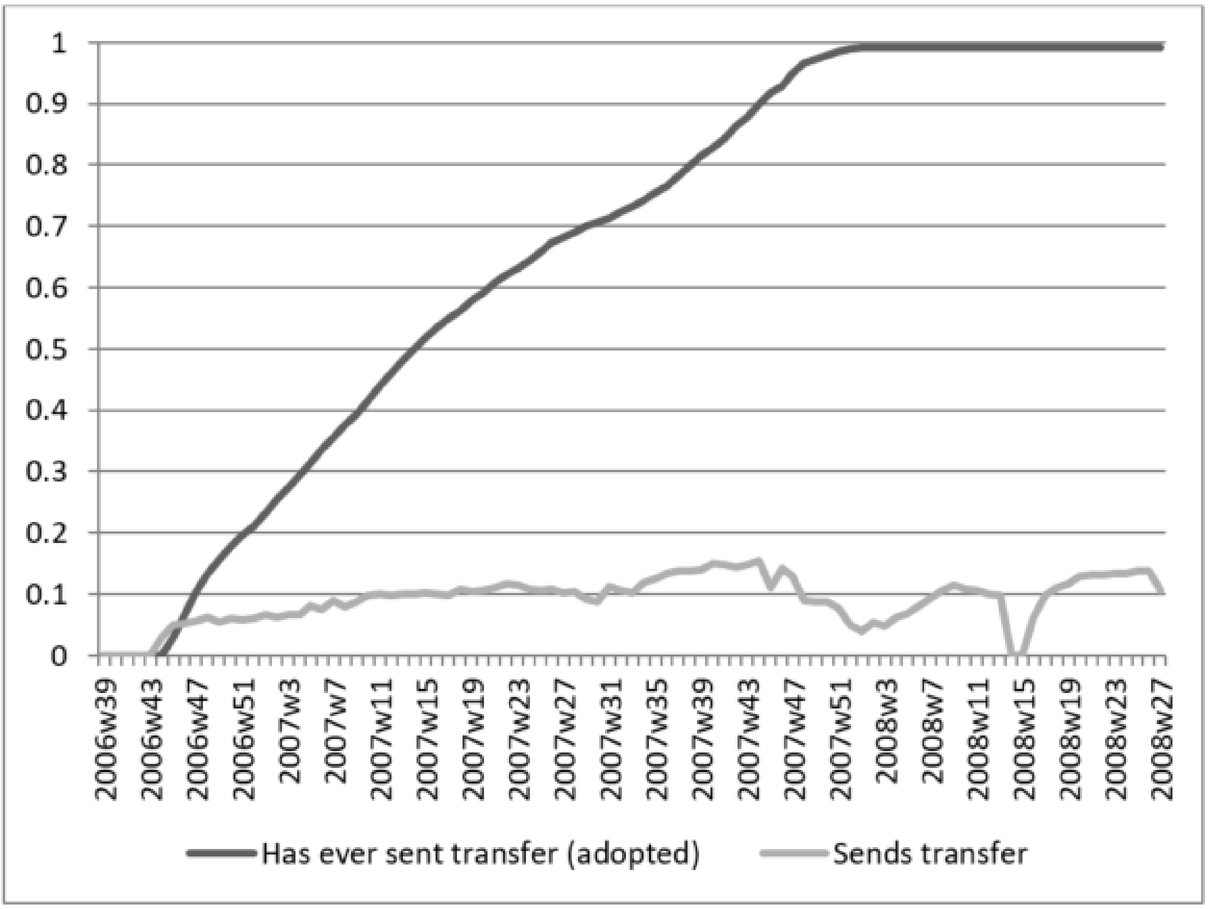

TABLE 1
Summary statistics

	(1) Full sample			(2) Before first transfer sent			(3) Before first transfer sent or received		
	Mean	Median	Std Dev	Mean	Median	Std Dev	Mean	Median	Std Dev
N(it)	273	212	244	113	78	119	100.7	70	103
A(it)	76.0	54	76.8	27.8	16	35.0	23.6	14	29.2
ΔA(it)	1.72	1	1.94	1.66	1	1.96	1.61	1	1.89
Weeks with SIM card	43.9	42	24.4	20.8	17	14.5	18.2	15	12.1
Number of phone calls by i	22.2	12	31.5	19.3	11	27.7	18.4	11	26.3
Number of transfers received	0.12	0	0.55	0.06	0	0.39			
Number of neighbors from whom i received a transfer	0.09	0	0.36	0.05	0	0.24			
Amount received, if positive	456	200	1480	623	225	2421			
Number of transfers sent	0.19	0	1.55						
Number of neighbors to whom i sent a transfer	0.15	0	0.99						
Amount sent, if positive	583	200	1707						
Observations		357,947			87,563			69,154	

Note: Weekly observations, at the individual (phone number) level.

TABLE 2

First adoption

	(1)	(2)	(3)
ΔA(i,t-2)	-0.00226	-0.00176	-0.00188
	(0.00048)***	(0.00054)***	(0.00052)***
ΔS(it)^2	-0.00001	-0.00078)	-0.00077
	(0.00004)	(0.00006)***	(0.00007)***
ΔA(i,t-2)^2	-0.00004	-0.00004	-0.00004
	(3E-06)***	(4E-06)***	(4E-06)***
Δ[A(i,t-2)S(it)]	0.00036	0.00038	0.00037
	(0.00002)***	(0.00003)***	(0.00003)***
Marginal effects of A(i,t-2), at different levels of A(i,t-2)			
A(i,t-2) = 0	0.0046	0.0054	0.0052
	(0.0005)***	(0.0006)***	(0.0007)***
A(i,t-2) = 20	0.0032	0.004	0.0038
	(0.0004)***	(0.0005)***	(0.0005)***
A(i,t-2) = 40	0.0018	0.0025	0.0024
	(0.0004)***	(0.0004)***	(0.0004)***
A(i,t-2) = 60	0.0004	0.0011	0.0010
	(0.0003)	(0.0003)***	(0.0003)***
A(i,t-2) = 80	-0.001	-0.0003	-0.0004
	(0.0004)***	(0.0004)	(0.0003)
Marginal effect of A(i,t-2) at means of A(i,t-2) and S(it)			
A(i,t-2) = 23.4	0.0029	0.0037	0.0036
	(0.0004)***	(0.0005)***	(0.0005)***
Fixed effects			
Year x month	N	Y	N
District	N	Y	N
Cell tower	N	Y	Y
Year x month x district	N	N	Y
R-squared	0.01	0.03	0.04
Observations	87,563	87,563	87,563

Notes: The dependent variable is Δy(i,t+1). The estimation method is OLS. An intercept is included in all regressions. Standard errors are clustered at the district level (M=27). Marginal effects are evaluated at sample means of regressors (in levels). *** p<0.01 ** p<0.05 * p<0.10.

TABLE 3

Generalized first adoption model

	(1)	(2)	(3)
ΔA(i,t-2)	0.00062	0.0006	0.0004
	(0.00054)	(0.00053)	(0.00057)
ΔS(it)^2	0.00039	-0.00025	-0.00029
	(0.00005)***	(0.00007)***	(0.00007)***
ΔA(i,t-2)^2	-4E-06	-5E-06	-4E-06
	(0.00001)	(0.00001)	(0.00001)
Δ[A(i,t-2)S(it)]	0.00008	0.0001	0.00011
	(0.00004)*	(0.00005)*	(0.00006)*
Δ[m(it) x S(it)]	0.06303		
	(0.00573)***		
Δ[m(it) x A(i,t-2)]	-0.00737	-0.00726	-0.0071
	(0.00169)***	(0.00142)***	(0.00153)***
Δ[m(it) x S(it)^2]	-0.00095	-0.00136	-0.00123
	(0.00004)***	(0.00018)***	(0.00022)***
Δ[m(it) x A(i,t-2)^2]	-0.00004	-0.00005	-0.00005
	(0.00002)**	(0.00002)**	(0.00002)**
Δ[m(it) x A(i,t-2) x S(it)]	0.0004	0.00051	0.00051
	(0.00007)***	(0.00009)***	(0.0001)***
Marginal effect of A(i,t-2) at means of A(i,t-2) and S(it)			
m(it) = 0	0.00171	0.00208	0.00197
	(0.00045)***	(0.00042)***	(0.00042)***
m(it) = 1	0.00425	0.00756	0.00749
	(0.00125)***	(0.00133)***	(0.00136)***
Marginal effects	-0.00253	-0.00549	-0.00552
difference†	(0.00134)*	(0.00135)***	(0.00134)
Fixed effects			
Year x month	N	Y	N
District	N	Y	N
Cell tower	N	Y	Y
Year x month x district	N	N	Y
R-squared	0.01	0.05	0.06
Observations	86,378	86,378	86,378

Notes: The dependent variable is Δy(i,t+1). The estimation method is OLS. An intercept is included in all regressions. Standard errors are clustered at the district level (M=27). Marginal effects are evaluated at sample means of regressors (in levels). Datapoints for which Δm(it)=1 (i.e. where m(it) switches from 0 to 1) are excluded for these estimations. Δ[m(it) x S(it)] is collinear with the fixed effects in (2) and (3), and is therefore excluded from these specifications. *** p<0.01 ** p<0.05 * p<0.10.

† This is equal to the marginal effect at m(it)=0 minus the marginal effect at m(it)=1.

TABLE 4

Adoption and subsequent usage

	(1)	(2)	(3)
ΔA(i,t-2)	-0.00226	-0.00176	-0.00188
	(0.00048)***	(0.00054)***	(0.00052)***
ΔS(it)^2	-0.00001	-0.00078	-0.00077
	(0.00004)	(0.00006)***	(0.00007)***
ΔA(i,t-2)^2	-0.00004	-0.00004	-0.00004
	(3E-06)***	(4E-06)***	(4E-06)***
Δ[S(it) x A(i,t-2)]	0.00036	0.00038	0.00037
	(0.00002)***	(0.00003)***	(0.00003)***
Δ[z(it) x S(it)]	-0.04137		
	(0.00198)***		
Δ[z(it) x A(i,t-2)]	0.00186	0.00171	0.00184
	(0.00054)***	(0.00058)***	(0.00056)***
Δ[z(it) x S(it)^2]	0.00006	0.00084	0.00083
	(0.00004)	(0.00007)***	(0.00008)***
Δ[z(it) x A(i,t-2)^2]	0.00004	0.00004	0.00003
	(3E-06)***	(5E-06)***	(4E-06)***
Δ[z(it) x A(i,t-2) x S(it)]	-0.00036	-0.00038	-0.00037
	(0.00002)***	(0.00004)***	(0.00004)***
Marginal effect of A(i,t-2) at means of A(i,t-2) and S(it)			
z(it) = 0	0.0029	0.0037	0.0036
	(0.0004)***	(0.0005)***	(0.0005)***
z(it) = 1	-0.0006	-0.0002	-0.0002
	(0.0003)**	(0.0003)	(0.0003)
Marginal effects difference†	0.0036	0.0039	0.0038
	(0.0005)***	(0.0006)***	(0.0006)***
Fixed effects			
Year x month	N	Y	N
District	N	Y	N
Cell tower	N	Y	Y
Year x month x district	N	N	Y
R-squared	0.01	0.01	0.01
Observations	357,947	357,947	357,947

Notes: The dependent variable is Δy(i,t+1). The estimation method is OLS. An intercept is included in all regressions. Standard errors are clustered at the district level (M=27). Marginal effects are evaluated at sample means of regressors (in levels). Datapoints for which Δz(it)=1 and Δz(i,t-1)=1 (i.e. the period when z(it) switches from 0 to 1 and the subsequent period) are excluded for these estimations. Δ[z(it) x S(it)] is collinear with the fixed effects in (2) and (3), and is therefore excluded from these specifications. *** p<0.01 ** p<0.05 * p<0.10.

† This is equal to the marginal effect at z(it)=0 minus the marginal effect at z(it)=1.