SIEPR Discussion Paper No. 16-037

Adoption with Social Learning and Network Externalities

By

Marcel Fafchamps, Mans Soderbom, Monique vanden Boogaart

Stanford Institute for Economic Policy Research
Stanford University
Stanford, CA 94305
(650) 725-1874

Adoption with Social Learning and Network Externalities*

Marcel Fafchamps
Stanford University

Mans Soderbom
University of Gothenburg

Monique vanden Boogaart

April 2016

Abstract

Using 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 network effects: increased usage of the new airtime transfer service by social neighbors predicts a higher adoption probability. We then seek to narrow down the possible sources of these network effects by distinguishing between network externalities and social learning. 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 quality of the product. In contrast, we find that network effects turn negative after first adoption, suggesting that airtime transfers are strategic substitutes among network neighbors.

*We benefitted from comments from Joshua Blumenstock, Jeremy Folz, Paul Frijters, Simon Quinn 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 2015 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 of 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 2009, 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 2013). The fact that all these applications are based on a platform – the mobile phone – originally designed for social communication leaves much room for social networks to affect adoption and usage.

In this paper we examine the adoption of an airtime transfer service in Rwanda using a large administrative dataset from the telecommunication operator. Peer-to-peer transfers of airtime between phone users is a predecessor to the introduction of mobile banking. The main 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, particularly those pertaining to IT technology.

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; Bjorkegren 2015). 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.[1] People talk about new products with others

[1] 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 (e.g., Li and Tan 2016). 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 network effects turn negative after first adoption, suggesting that airtime transfers are strategic substitutes among network neighbors.

This paper complement a large literature documenting the diffusion of new products and behaviors on social networks (e.g., Krystakis and Fowler 2007, 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 presenting the conceptual framework and testing strategy. The information available in the raw data is discussed in Section 3, together with a description of how we use the raw data to construct the variables used in our analysis. Empirical results are presented in Section 4. Section 5 concludes.

2. Conceptual framework

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.[2] 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

[2] 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.

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}.

2.1. 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. 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:

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

where A_{it} is the cumulative number of adopting neighbors between t_i and t, that is:

$$A_{it} \equiv \sum_{s=t_i}^{s=t} \Delta A_{is}$$

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} is the time elapsed between t_i and t, that is:

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

where t_i is the time at which i starts being at risk of being exposed to information about the product's existence.

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

$$\Pr(y_{it+1} = 1 | \{y_{it_i}, ..., y_{it}\} = \{0, ..., 0\}) = [1 - (1-q)^{A_{it}}(1-\theta_v)^{S_{it}}]p_i \tag{2.1}$$

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 with respect to 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:

$$\begin{aligned} \Pr(y_{it+1} = 1 | y_{is} = 1 \text{ for some } s < t) &= p_{i,t+1} \\ &= p_i + \varepsilon_{it+1} \end{aligned} \tag{2.2}$$

Thus once i has learned about the existence of the product, the data generating process shifts from (2.1) to (2.2). 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:

$$\Pr(y_{it+1} = 1 | M_{is} = 1 \text{ for some } s < t) = p_i + \varepsilon_{it+1} \tag{2.3}$$

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.

2.2. 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.[3] 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:

$$h_{it} = \frac{N_{it}^1}{N_{it}} \tag{2.4}$$

The variance of this belief is approximately given by:

$$\begin{aligned} v_{it}^2 &= \frac{1}{N_{it}} \frac{N_{it}^1}{N_{it}} \frac{N_{it} - N_{it}^1}{N_{it}} \\ &= \frac{1}{N_{it}} h_{it}(1 - h_{it}) \end{aligned} \tag{2.5}$$

As sample size increases, h_{it} tends to μ and v_{it}^2 tends to 0.[4]

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

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

In other words, the information people have is, on average, unbiased and the variance of their

[3] 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.

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

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

$$\begin{aligned}
q_{it}^b &= \alpha\frac{\gamma}{\gamma + n_{it}} + \mu\frac{n_{it}}{\gamma + n_{it}} + e_{it}^b \text{ with } e_{it}^b\,\tilde{}\,(0, \sigma_{it}^2) \\
\sigma_{it}^2 &= \frac{1}{\gamma + n_{it}}(\alpha\frac{\gamma}{\gamma + n_{it}} + \mu\frac{n_{it}}{\gamma + n_{it}})(1 - \alpha\frac{\gamma}{\gamma + n_{it}} - \mu\frac{n_{it}}{\gamma + n_{it}})
\end{aligned}$$

As with uninformed priors, beliefs h_{it}^b tend to μ over time, but they show some persistence around initial priors.[5]

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.

[5] The variance σ_{it}^2 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_{it}^2 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_{it}^2 increases. Eventually posterior beliefs settle on μ and the variance falls, dominated now by N_{it}.

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:

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

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_{it}^2 directly, we replace them by formulas (2.4) and (2.5) above and we get:

$$\Pr(y_{it+1} = 1|\{y_{it_i}, ..., y_{it}\} = \{0, ..., 0\}) = \Pr\left((\mu - \tau_i) - R\frac{\mu(1-\mu)}{n_{it}} \geq -e_{it+1}\right) \tag{2.6}$$

Equation (2.6) 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:

$$\Pr(y_{it+1} = 1|\{y_{it_i}, ..., y_{it}\} = \{0, ..., 0\}) = \tag{2.7}$$
$$\Pr\left(\alpha\frac{\gamma}{\gamma + n_{it}} + \mu\frac{n_{it}}{\gamma + n_{it}} + R\frac{1}{\gamma + n_{it}}(\alpha\frac{\gamma}{\gamma + n_{it}} + \mu\frac{n_{it}}{\gamma + n_{it}})(1 - \alpha\frac{\gamma}{\gamma + n_{it}} - \mu\frac{n_{it}}{\gamma + n_{it}}) \geq \tau_{it} - e_{it+1}^b\right)$$

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,[6] 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:[7]

$$n_{it} = \theta_v S_{it} + qA_{it} + u_{it} \text{ with } u_{it}\tilde{}(0, v^2) \tag{2.8}$$

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

$$\Pr(y_{it+1} = 1 | \{y_{it_i}, ..., y_{it}\} = \{0, ..., 0\}) = \Pr\left((\mu - \tau_i) - R\frac{\mu(1-\mu)}{\theta_v S_{it} + qA_{it} + u_{it}} \geq -e_{it+1}\right) \tag{2.9}$$

Equation (2.9) 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 (2.7) by its value given by (2.8). Our earlier observation remains valid: with strong prior beliefs, the variance term that multiplies R in equation (2.7) 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

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

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

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 (2.7) 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:

$$\Pr(y_{it+1} = 1 | y_{is} = 1 \text{ for some } s \leq t) \; = \Pr\left((\mu - \tau_i) \geq -e_{it+1}\right) \tag{2.10}$$

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 (2.7) 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 (2.3) 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.

2.3. 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.

2.4. Testing strategy

We are now ready to put all these predictions together in the form of a regression model. To recap, if network effects are purely due to social learning, then they disappear after first usage.

If they are purely due to other strategic complementarities, the data generating process should be the same before and after first adoption. To distinguish between the two types of social learning, we need to observe a signal M_{it} which equal 1 when i is unambiguously informed of the product's existence – even though i has not adopted it. If such signal is observed by the researcher, identification between the two comes from the following observation: when social learning is purely about product existence, once i has learned about the existence of the product, the data generating process immediately shifts from (2.1) to (2.2). In contrast, if social learning is about product quality, the data generating process remains (2.7) until first adoption. This makes it possible to test the two learning models against each other in reduced form.

The reduced form for models (2.1) and (2.7) is similar and can be written as:

$$\Pr(y_{it+1} = 1|\{y_{it_i}, ..., y_{it}\} = \{0, ..., 0\}) = \alpha_i + \alpha_1 S_{it} + \alpha_2 A_{it} + \varepsilon_{it+1} \tag{2.11}$$

$$\Pr(y_{it+1} = 1|\{y_{it_i}, ..., y_{it}\} = \{0, ..., 0\}) = \alpha_i + \alpha_1 S_{it} + \alpha_2 A_{it} + \alpha_3 S_{it}^2 + \alpha_4 A_{it}^2 + \alpha_5 S_{it} A_{it} + \varepsilon_{it+1} \tag{2.12}$$

Model (2.11) is a simple linear approximation of the two structural models (2.1) and (2.7). 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 (2.11) this means $\alpha_2 > 0$. We also the marginal effect of S_{it} to be positive – which implies $\alpha_1 > 0$ in equation (2.11). 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 (2.12) 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 $\alpha_5 < 0$.[8] We have include

[8] 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.

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 (2.2) 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:

$$\begin{aligned} \Pr(y_{it+1} &= 1 | \{y_{it_i}, ..., y_{it}\} = \{0, ..., 0\}) = \alpha_i + \alpha_1 S_{it} + \alpha_2 A_{it} + \alpha_3 S_{it}^2 + \alpha_4 A_{it}^2 + \alpha_5 S_{it} A_{it} \\ &+ \beta_0 m_{it} + \beta_1 S_{it} m_{it} + \beta_2 A_{it} m_{it} + \beta_3 S_{it}^2 m_{it} + \beta_4 A_{it}^2 m_{it} + \beta_5 S_{it} A_{it} m_{it} + \varepsilon_{it+1} \end{aligned} \quad (2.13)$$

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.

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:

$$\begin{aligned}\Pr(y_{it+1} &= 1) = \alpha_i + \alpha_1 S_{it} + \alpha_2 A_{it} + \alpha_3 S_{it}^2 + \alpha_4 A_{it}^2 + \alpha_5 S_{it} A_{it} + \gamma_0 z_{it} \\ &+\gamma_1 S_{it} z_{it} + \gamma_2 A_{it} z_{it} + \gamma_3 S_{it}^2 z_{it} + \gamma_4 A_{it}^2 z_{it} + \gamma_5 S_{it} A_{it} z_{it} + \varepsilon_{it+1} \end{aligned} \tag{2.14}$$

Unlike models (2.12) and (2.13), regression model (2.14) 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 (2.14) allows us to test this as well.

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.[9]

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. 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

[9] 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.

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 under conditions of strict confidentiality.[10] 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 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.[11]

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).

After its introduction in September 2006, ME2U usage increased steadily until the 1st of

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

[11] We cannot rule out that an individual may have multiple phone numbers, or that phone numbers may be transferred across users. We come back to this issue in the empirical section.

July 2008 when there is a break in the administrative data (see Figure 1). To avoid spurious 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.

Since all phone usage in Rwanda 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. For information, we give in Appendix Figure A1 the location of all cell towers in Rwanda during our period of analysis.

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.[12] For these subscribers, we observe all their ME2U transfers between the introduction of the service in

[12] 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.

September 2006, and June 30th 2008. The end-date T is thus the end of June 2008.

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 – more below. As indicated in the conceptual section, all regressors are lagged – by one 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. 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. 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.[13] 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 $g_{i,j,t}$ 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:

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

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 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, $S_{it} \equiv t - t_i$.

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

3.3. Descriptive statistics

We now provide summary statistics for the variables used in the analysis. Remember that these variables relate to the 5,000 subscribers randomly selected for analysis. Table 1 provides descriptive statistics for all the variables; Tables 2 and 3 provide the same information, but split between before and after i receives his first airtime transfer.

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 before and after i receives his first airtime transfer, to hope achieving identification.

Table 3.1: Summary statistics

Variable	**Mean**	**Std. Dev.**	**Number of zero value**	**N**
N_{it}	507.376	466.274	859	395507
ΔA_{it}	1.678	1.945	13114	390515
A_{it}	71.053	75.857	16347	395507
S_{it}	41.843	25.819	5000	400507

Table 3.2: Summary statistics before signal (m_{it}=0)

Variable	**Mean**	**Std. Dev.**	**Number of zero value**	**N**
N_{it}	475.151	457.827	793	282004
ΔA_{it}	1.553	1.883	101701	277140
A_{it}	61.164	71.366	15790	282004
S_{it}	38.617	26.165	4936	286940

Table 3.3: Summary statistics after signal (m_{it}=1)

Variable	**Mean**	**Std. Dev.**	**Number of zero value**	**N**
$N_{i,t}$	587.441	477.304	66	113503
ΔA_{it}	1.983	2.058	29413	113375
$A_{i,t}$	95.624	80.933	557	113503
$S_{i,t}$	49.993	22.99	64	113567

4. Empirical results

The first regression model we estimate is (2.12), using only observations until first adoption. To eliminate the individual fixed effect α_i, we first difference the data. The estimated model is a linear probability model of the form:

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

where $\Delta x_t \equiv x_t - x_{t-1}$ by definition of notation and observations up to the first adoption are used.[14] We have $\Delta S_{it} = 1$ by construction. Coefficient estimates are presented in Table 4. Standard errors are clustered at the district level. We see that α_2 and α_3 are both significantly positive while α_4 is significantly negative. Remember that, when social learning is about product existence, the relationship between 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.

To investigate this, we report in Table 5 the marginal effect $\partial \Pr / \partial A_{it}$ evaluated at various values of A_{it}. 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 much weaker than that predicted by equation (2.1).

In Table 6 we present coefficient estimates for regression model (2.13). Once again, we eliminate the individual fixed effect α_i by first-differencing the data. The estimated model is a

[14] 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.

LPM of the form:

$$\Delta y_{it+1} = \alpha_1 + \alpha_2 \Delta A_{it} + \alpha_3 \Delta(S_{it}^2) + \alpha_4 \Delta(A_{it}^2) + \alpha_5 \Delta(S_{it} A_{it}) + \beta_0 \Delta m_{it}$$

$$+\beta_1 \Delta(S_{it} m_{it}) + \beta_2 \Delta(A_{it} m_{it}) + \beta_3 \Delta(S_{it}^2 m_{it}) + \beta_4 \Delta(A_{it}^2 m_{it}) + \beta_5 \Delta(S_{it} A_{it} m_{it}) + \Delta\varepsilon_{it+1} \quad (4.2)$$

where, as in (4.1), we only include observations up to the first adoption. In Table 7 we present estimates of marginal effects $\partial \Pr / \partial A_{it}$ evaluated for $m_{it} = 0$ and $m_{it} = 1$. Network effects remain significant throughout, although they are significantly 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 large even after $m_{it} = 1$ suggests that, of the two, diffusing information about quality accounts for a large share of social learning effects.

We now seek to rule out that observed network effects on adoption are purely due to network externalities, not to social learning. To this effect, we estimate model (2.14) in the same data. The model is estimated in first difference to eliminate unobserved heterogeneity α_i, i.e., it is of the form:

$$\Delta y_{it+1} = \alpha_1 + \alpha_2 \Delta A_{it} + \alpha_3 \Delta(S_{it}^2) + \alpha_4 \Delta(A_{it}^2) + \alpha_5 \Delta(S_{it} A_{it}) + \gamma_0 \Delta z_{it}$$

$$+\gamma_1 \Delta(S_{it} z_{it}) + \gamma_2 \Delta(A_{it} z_{it}) + \gamma_3 \Delta(S_{it}^2 z_{it}) + \gamma_4 \Delta(A_{it}^2 z_{it}) + \gamma_5 \Delta(S_{it} A_{it} z_{it}) + \Delta\varepsilon_{it+1} \quad (4.3)$$

where all observations are used and $z_{it} = 1$ if subscriber i has used ME2U before time t. Regression results are presented in Table 8. Marginal effects estimated at the sample mean are presented in Table 9. As should be, the α coefficient estimates are very similar to those reported in Table 4, and the average marginal effect is similar as well. We find that the marginal effect estimated at the sample mean 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.

To check the robustness of this finding, we re-estimate (4.3) in two alternative ways. Results are presented in columns 2 and 3 of Tables 8 and 9. We start in column 2 by adding a time trend to the regression. The concern is that the usage of airtime transfers by network neighbors may be varying over time in a way that is correlated with a time trend. Omitting this trend may result in a spurious negative correlation between neighbor usage and own usage that varies systematically before and after first adoption. We do find evidence of a time trend in airtime transfer usage – the trend coefficient is strongly statistically significant. But this has little effect on coefficient estimates and on marginal effects estimated at the mean: $\frac{\partial \Pr(y_{it+1}=1|z_{it}=1)}{\partial A_{it}}$ goes up a bit, but remains significantly negative. Similar results are obtained if we use time dummies instead of a linear time trend in the first difference regression (4.3).

In column 3 we add controls for the transfers 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) with four additional regressors: a time trend, as in column 2; the number of transfers received at t, the amount of airtime transfers received at t, and the number of neighbors from whom i received a transfer at t. Coefficient estimates are significant but their interpretation is somewhat confusing. Two of the coefficients are negative, in agreement with our conjecture above, indicating that when i receives more transfers from more people, he or she is less likely to transfer airtime to others during the next period. The third coefficient, however, (amount received) is positive, indicating the opposite effect. More importantly, the estimate marginal effect $\frac{\partial \Pr(y_{it+1}=1|z_{it}=1)}{\partial A_{it}}$ remains negative and significant – and the change in magnitude relative to column 2 is relatively small (e.g., from -0.0030 to -0.0028). 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 – and either reciprocated or not in the subsequent period.

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 2008, Bramoulle, 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 Soderbom 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. Another possibility is when individuals i and j (e.g., spouses, friends, or relatives) team up to purchase airtime in bulk to get a quantity discount – and subsequently transfer airtime to each other. In this case too, if i does purchase the airtime on her phone before transferring it to network neighbors, observing that i transfers airtime is negatively correlated with j transferring airtime.

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 1 of Table 9, 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 74% (i.e., -0.00356/0.00483). Comparisons made using the other two columns are slightly lower, but continue to suggest a significant underestimation of social learning from models (4.1) and (4.2)

5. Robustness analysis

Fafchamps, Goyal and Vander Leij (2010) estimate a model similar to regression (4.1) in fixed effect instead of first difference. 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. They 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 are affected, we re-estimate regression (4.1) after detrending all first-differenced regressors. Results show absolutely no change in coefficient estimates and standard errors.[15]

There remains the perennial issue of possible endogeneity of A_{it}. Since our analysis is based on complete administrative data obtained directly from the phone provider, it is less susceptible

[15] Except for the fact that $\Delta(S_{it}^2)$ drops out of the estimation since, by construction, it is linear in time because S_{it} is linear in time.

to measurement error, which is arguably the most common source of endogeneity affecting survey data. This notwithstanding, there remain other potential sources of endogeneity that we discuss in turn. The first potential source is reflection bias: i influences j and j influences i. To eliminate this type of simultaneity bias, in our analysis we have used the lagged value of A_{it} instead of its contemporaneous value. Experimentation with different lag lengths leaves the results unaffected. The second potential source of endogeneity is network self-selection: I create new links to adopters when I am considering adopting myself. To obviate this possibility, we have defined i's network as including the phone numbers with which i had a phone contact at any time during the entire study period, that is, between January 2005 and June 2008. This rules out any time variation in network neighborhood that is correlated with i's time-varying propensity to adopt.

A third 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 investigate this possibility, we reestimate model (4.1) with separate dummies for each month in the study period. The results are presented in Table 10. For comparison purposes, the first column reproduces the results from Tables 4 (first panel) and 5 (lower panel). Comparable results with month dummies are presented in column (2). The month dummies are often significant, suggesting the presence of aggregate shocks, but we find little change in the estimated coefficients of ΔA_{it} and $(\Delta A_{it})^2$. We similar find little change in marginal effects, reported in the lower panel of the Table.

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).

To address this concern, we reestimate (4.1) with district-specific dummies for each month of the study. The results are shown in Table 10, column (3). Again we find little change in the estimated coefficients of ΔA_{it} and $(\Delta A_{it})^2$, and hardly any change in estimated marginal effects.

In Tables 11 and 12 we do the same thing for the results reported in Tables 6 to 9. Here too we find little if any change in estimated coefficients and marginal effects. From this we conclude that our main results of interest are robust to correcting for possible endogeneity to the extent allowed by the data.

6. Conclusion

In this study we use 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 network externalities and social learning. 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. All results are robust to the inclusion of district-specific month dummies.

Our results provide useful insights in the process by which products and services diffuse on social networks, particularly those involving IT technology. Observing correlated patterns of adoption and usage is typically interpreted as symptomatic of network externalities, without specifying what these externalities may be. We unpack this black box to distinguish social learning from strategic complementarity in usage. We find strong evidence of the former, but no evidence of latter – if anything, usage is a strategic substitute between network neighbors. This leads to two final observations. First, we should stop assuming, as we often do when considering network effects, that interactions on networks are synonymous with strategic complements. In our application, strategic substitutes are a more natural alternative. Secondly, the presence of strategic substitution effects does not rule out positive welfare effects from social interaction. Given how rapidly airtime transfer services spread across Rwanda – and given that usage peaked after national disasters (e.g., Blumenstock, Eagle and Fafchamps 2016) – we strongly suspect that the introduction of the service generated large welfare gains, irrespective of usage.

7. 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 (2013). "Is Information Power? Using Mobile Phones and Free Newspapers during an Election in Mozambique", IGC Working Paper

Bjorkegren, Daniel (2015). "The Adoption of Network Goods: Evidence from the Spread of Mobile Phones in Rwanda", Department of Economics, Brown University (mimeograph)

Blumenstock, J. and N. Eagle (2012). "Divided We Call: Disparities in Access and Use of Mobile Phones in Rwanda", *Information Technology and International Development*, 8(2): 1-16

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* (forthcoming)

Bramoulle, 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

Christakis, Nicholas A. and James H. Fowler (2007). "The Spread of Obesity in a Large Social Network over 32 Years", *New England Journal of Medecine*, 357:370-379, July 26

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

Economides, Nicholas and Przemyslaw Jeziorski (2016). "Mobile Money in Tanzania", Stern School of Business, New York University (mimeo)

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

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

Foster, Andrew D and Mark R. Rosenzweig (1996). "Technical Change and Human-Capital Returns and Investments: Evidence from the Green Revolution," *American Economic Review*, 86(4): 931-53, September

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.

Li, Wei and Xu Tan (206). "Boundedly Rational Learning in Social Networks", University of Washington (mimeo).

Mobius, Markus and Tanya Rosenblat (2014). "Social Learning in Economics", *Annual Review of Economics*, 6: 827-847, August

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

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 van den Boogaart, Monique (2013). "Social Networks and Technology Adoption: Evidence from a mobile money service in Rwanda", Department of Economics, Oxford University (unpublished PhD thesis)

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

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.

Table 4

First Adoption: First Difference Estimates

	Coef.	s.e.	Coef/s.e.
ΔA(it)	0.0049613	0.000714	6.95
ΔS(it)^2	0.0005408	3.04E-05	17.81
ΔA(it)^2	-0.0000126	2.95E-06	-4.27
Δ[A(it)S(it)]	0.0000402	2.36E-05	1.71
Observations	92,700		

Note: Standard error are clustered at the district level (M=27).

Table 5

First Adoption: Marginal effect of A(it), for different A(it)*

A(it)	m.eff.	s.e.	Coef/s.e.
0	0.005731	0.00055	10.42
20	0.005227	0.000484	10.81
40	0.004722	0.00044	10.74
60	0.004217	0.000425	9.92
80	0.003713	0.000443	8.39
100	0.003208	0.000489	6.56

* Based on results in Table 2.

Evaluated at sample means of regressors.

Note: Standard error are clustered at the district level (M=27).

Table 6

Generalized First Adoption Model: First Difference Estimates

	Coef.	s.e.	Coef/s.e.
ΔA(it)	0.005118	0.000709	7.22
ΔS(it)^2	0.000532	3.13E-05	17.01
ΔA(it)^2	-1.3E-05	3.92E-06	-3.28
Δ[A(it) x S(it)]	3.55E-05	2.46E-05	1.44
Δm(it)	0.018369	0.021977	0.84
Δ[m(it) x S(it)]	0.002727	0.002179	1.25
Δ[m(it) x A(it)]	-0.00115	0.000425	-2.72
Δ[m(it) x S(it)^2]	-2.7E-05	4.07E-05	-0.67
Δ[m(it) x A(it)^2]	9.93E-07	2.82E-06	0.35
Δ[m(it) x A(it) x S(it)]	1.93E-05	2.14E-05	0.9
Observations	92,700		

Note: Standard error are clustered at the district level (M=27).

Table 7
First Adoption: Marginal effect of A(it)*

	m.eff.	s.e.	Coef/s.e.
m(it)=0	0.005	0.0005	10.8
m(it)=1	0.004	0.0005	8.7

* Based on results in Table 4
Evaluated at sample means of regressors.
Note: Standard error are clustered at the district level (M=27).

Table 8
Adoption & subsequent usage: First Difference Estimates

	Specification (1)			Specification (2)			Specification (3)		
	Coef.	s.e.	Coef/s.e.	Coef.	s.e.	Coef/s.e.	Coef.	s.e.	Coef/s.e.
ΔA(it)	0.004961	0.000714	6.95	0.005047	0.000695	7.26	0.005075	0.000706	7.19
ΔS(it)^2	0.000541	3.04E-05	17.81	0.000373	3.14E-05	11.87	0.000352	3.18E-05	11.08
ΔA(it)^2	-1.3E-05	2.95E-06	-4.27	-1.3E-05	2.99E-06	-4.33	-1.3E-05	2.98E-06	-4.47
Δ[A(it) x S(it)]	4.02E-05	2.36E-05	1.71	4.45E-05	2.35E-05	1.9	4.75E-05	2.36E-05	2.02
Δ[z(it) x S(it)]	-0.01854	0.001507	-12.31	-0.02594	0.001834	-14.15	-0.02731	0.001877	-14.54
Δ[z(it) x A(it)]	-0.00949	0.001291	-7.35	-0.00897	0.001232	-7.29	-0.00887	0.001252	-7.09
Δ[z(it) x S(it)^2]	-0.00052	3.04E-05	-17.11	-0.00049	3.04E-05	-16.11	-0.00048	3.13E-05	-15.29
Δ[z(it) x A(it)^2]	1.42E-05	2.90E-06	4.89	1.42E-05	2.92E-06	4.85	1.38E-05	2.89E-06	4.76
Δ[z(it) x A(it) x S(it)]	-2.3E-05	2.39E-05	-0.94	-2.8E-05	2.36E-05	-1.17	-2.6E-05	2.41E-05	-1.08
Time (week number)				0.000408	1.99E-05	20.57	0.000447	1.99E-05	22.5
log(amount received + 1)							0.002777	0.000641	4.34
Number of transfers received							-0.00498	0.001963	-2.54
Number of neighbors from whom i received a transfer							-0.01283	0.003589	-3.58
Observations	371,785			371,785			361,616		

Note: Standard error are clustered at the district level (M=27).

Table 9

First Adoption: Marginal effect of A(it), before & after first adoption

	Specification (1)			Specification (2)			Specification (3)		
	m.eff.	s.e.	Coef/s.e.	m.eff.	s.e.	Coef/s.e.	m.eff.	s.e.	Coef/s.e.
z(it)=0	0.0048358	0.0004096	11.81	0.00506	0.000416	12.18	0.005161	0.000418	12.34
z(it)=1	-0.0035638	0.0006329	-5.63	-0.00303	0.000622	-4.87	-0.00282	0.000652	-4.33

* Based on results in Table 6

Evaluated at sample means of regressors.

Note: Standard error are clustered at the district level (M=27).

Table 10
First Adoption: First Difference Estimates and Marginal Effects

	(1)		(2)		(3)	
FD estimates	Coef.	Coef/s.e.	Coef.	Coef/s.e.	Coef.	Coef/s.e.
ΔA(it)	0.0049613	6.95	0.0043119	6.23	0.0042423	6.38
ΔS(it)^2	0.0005408	17.81	-0.0003368	-8.09	-0.0003276	-7.88
ΔA(it)^2	-0.0000126	-4.27	-0.0000111	-3.13	-0.0000110	-3.05
Δ[A(it)S(it)]	0.0000402	1.71	0.0000653	2.28	0.0000647	2.15
Marginal effects of A(it), at different levels of A(it)						
A(it) = 0	0.0057	10.42	0.0056	10.13	0.0055	10.36
A(it) = 20	0.0052	10.81	0.0051	10.93	0.0050	11.32
A(it) = 40	0.0047	10.74	0.0047	11.07	0.0046	11.57
A(it) = 60	0.0042	9.92	0.0042	10.03	0.0042	10.43
A(it) = 80	0.0037	8.39	0.0038	8.11	0.0037	8.3
A(it) = 100	0.0032	6.56	0.0033	6.11	0.0033	6.16
year x month dummies	N		Y		N	
year x month x district dummies	N		N		Y	
Observations	92,700		92,700		92,700	

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

Table 11
Generalized First Adoption Model: First Difference Estimates and Marginal Effects

	(1)		(2)		(3)	
FD estimates	Coef.	Coef/s.e.	Coef.	Coef/s.e.	Coef.	Coef/s.e.
ΔA(it)	0.0051	7.22	0.0044	6.66	0.0043	6.82
ΔS(it)^2	0.0005	17.01	-0.0003	-7.68	-0.0003	-7.34
ΔA(it)^2	-1.3E-05	-3.28	-1.1E-05	-2.81	-1.1E-05	-2.61
Δ[A(it)S(it)]	3.6E-05	1.44	0.0001	2.24	0.0001	2.02
Δm(it)	0.0184	0.84	-0.0167	-0.79	-0.0163	-0.77
Δ[m(it) x S(it)]	0.0027	1.25	0.0058	2.88	0.0059	2.87
Δ[m(it) x A(it)]	-0.0012	-2.72	-0.0009	-2.12	-0.0010	-2.29
Δ[m(it) x S(it)^2]	-2.7E-05	-0.67	-0.0001	-2.34	-0.0001	-2.4
Δ[m(it) x A(it)^2]	9.9E-07	0.35	3.9E-07	0.14	-2.1E-07	-0.08
Δ[m(it) x A(it) x S(it)]	1.9E-05	0.9	1.9E-05	0.84	2.5E-05	1.09
Marginal effects of A(it)						
m(it)=0	0.0051	10.75	0.0050	10.95	0.0049	11.34
m(it)=1	0.0044	8.72	0.0044	8.85	0.0043	9.17
year x month dummies	N		Y		N	
year x month x district	N		N		Y	
Observations	92,700		92,700		92,700	

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

Table 12
Adoption & subsequent usage, with additional controls

	(1)		(2)		(3)		(4)	
FD estimates	Coef.	Coef/s.e.	Coef.	Coef/s.e.	Coef.	Coef/s.e.	Coef.	Coef/s.e.
ΔA(it)	0.00334	4.77	0.00346	4.87	0.00326	4.75	0.00339	4.86
ΔS(it)^2	0.00011	2.95	0.00008	2.17	0.00011	2.82	0.00008	2.07
ΔA(it)^2	-0.00001	-4.5	-0.00001	-4.64	-0.00001	-4.59	-0.00001	-4.75
Δ[A(it) x S(it)]	0.00009	3.74	0.00009	3.78	0.00009	3.77	0.00009	3.81
Δ[z(it) x S(it)]	-0.04424	-19.7	-0.04628	-20.45	-0.04493	-18.38	-0.04692	-19.25
Δ[z(it) x A(it)]	-0.00742	-6.07	-0.00742	-5.97	-0.00736	-6.08	-0.00738	-5.97
Δ[z(it) x S(it)^2]	-0.00017	-4.7	-0.00015	-3.94	-0.00017	-4.5	-0.00014	-3.8
Δ[z(it) x A(it)^2]	0.00001	4.95	0.00001	4.88	0.00002	5.06	0.00001	4.99
Δ[z(it) x A(it) x S(it)]	-0.00007	-2.94	-0.00007	-2.76	-0.00007	-2.97	-0.00007	-2.79
log(amount received + 1)			0.00247	3.88			0.00246	3.86
Number of transfers received			-0.00483	-2.52			-0.00486	-2.53
Number of neighbors from whom i received a transfer			-0.01167	-3.25			-0.01173	-3.28
Marginal effects of A(it)								
z(it)=0	0.00515	11.92	0.00527	12.09	0.00508	12.16	0.00520	12.36
z(it)=1	-0.00312	-4.99	-0.00294	-4.48	-0.00313	-4.91	-0.00295	-4.41
year x month dummies	Y		Y		N		N	
year x month x district dummies	N		N		Y		Y	
Observations	371,785		371,785		361,616		361,616	

Note: Standard errors are clustered at the district level (M=27).
Marginal effects are evaluated at sample means of regressors.
z(it)=0 prior to adoption, z(it)=1 after adoption