# Adoption with Social Learning and Network Externalities*

Marcel Fafchamps — Stanford University

Mans Soderbom — University of Gothenburg

Monique vanden Boogaart — University of Oxford

September 2015

**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, suggsting that airtime transfers are strategic substitutes among network neighbors.

*We Nathan Eagle for making the data available for this study.

## 1. Introduction

It has often been observed that the adoption of new products and services, and other behavioral changes, seems to diffuse along social networks (references). What is less clear is why. The purpose of this paper is to throw some light on the issue using a large dataset about the adoption of a new phone service over time. To keep things straightforward, we speak throughout of adoption of a new product, but the same principles generally apply to new services or to the adoption of behavioral changes.

There are many possible reasons why adoption may spread along social networks. One is that some individuals get to know of a new product. People talk about new products with others in their network of acquaintances, so that information about the existence of the new product spreads through social learning. A proportion of those informed of the new product adopt it, and since adoption requires knowing about the new product, adoption is observed to diffuse by social contact, in a way similar to the way an epidemic spreads in a population.

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

Diffusion along social networks may also occur for reasons having nothing to do with social learning. One particular case is network externalities or, more precisely, strategic complementarities in adoption decisions. 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 – 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 (reference).

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). 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 all be strategic complements, are is commonly assumed in the descriptive empirical literature. [MORE LINKS TO THE LITERATURE NEEDED HERE]

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.[1] Further let $T_i$ denote the time at which individual $i$ first uses the product. Finally, let $T$ denote the last data period for which we have information. By definition, $T_i > T$ for an individual who, by time $T$, has not yet used the product.

As we will argue below, usage after adoption provides useful information as well. Usage $y_{it}$ can therefore be divided into two vectors or periods: the time until first usage $\{y_{it_i}, ...y_{iT_i}\}$; and usage after that $\{y_{iT_i+1}, ...y_{iT}\}$. By construction, $\{y_{it_i}, ...y_{iT_i}\}$ is either a sequence of 0's ending

[1] 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.

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 $\theta_{vt}$ 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 $\Delta A_{it}$ 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 $i$. We assume that the researcher observes $\Delta A_{it}$, 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 the two sources of information. If we assume independence between $\theta_{vt}$ 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 $\theta_{vt}$ 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 $i$ begins 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 w.r.t. $S_{it}$ and $A_{it}$ falls with $S_{it}$ and with $A_{it}$. This is because having heard about the product once is enough to know of its existence.

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

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

In the learning about existence model, the certainty that $i$ has learned about the existence of the product immediately shifts the estimation model from (2.1) to (2.2). In contrast, in the learning about quality model, the data generating process remains (2.6) since, by assumption, the person is assumed to know about the product from the beginning. This makes it possible to test the two learning models against each other even in a reduced form.

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:

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

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 $\theta_{vt}$ 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 $\mu$ denote the true probability that the product performs: a high $\mu$ good always performs well, while a low $\mu$ good often performs poorly. Individuals differ in how much they value unobserved quality $\mu$ – more about this later.

We assume that the posterior belief of individual $i$ at time $t$ is simply the sample estimate of the unknown Bernoulli parameter $\mu$ based on the information available to $i$ at time $t$.[2] 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:

$$p_{it} = \frac{N_{it}^1}{N_{it}} \tag{2.3}$$

The variance of this belief is approximately given by:

$$\begin{aligned} v_{it}^2 &= \frac{1}{N_{it}} \frac{N_{it}^1}{N_{it}} \frac{N_{it} - N_{it}^1}{N_{it}} \\ &= \frac{1}{N_{it}} p_{it}(1 - p_{it}) \end{aligned} \tag{2.4}$$

As sample size increases, $p_{it}$ tends to $\mu$ and $v_{it}^2$ tends to 0.[3]

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

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

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

[2] This is not a bona fide Bayesian approach – see Mood, Graybill and Boes (19XX) p. 342 for the correct Bayesian estimator of a Bernouilli parameter. But this simple approach suffices for our purpose.

[3] The above formula for the variance is obtained by combining MGB p. 236 with p. 89.

If we allow agents to hold a prior belief $p_{i0}$, 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:

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

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

We do not observe $p_{i0}$ 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:

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

As with uninformed priors, beliefs $p_{it}^b$ tend to $\mu$ over time, but they show some persistence around initial priors.[4]

Having modelled learning, we now turn to adoption. We start without prior beliefs. We assume that individuals differ in the threshold value of $\mu$ that they require before adopting. At first glance, it seems that we could simply assume that people adopt if their estimate of $\mu$

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

is larger than some value $\tau_i$ 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 $p_{i1} = 1 \geq \tau_i$ for any $\tau_i$. This is clearly an unappealing decision rule because an estimate of $\mu$ 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.[5] We have:

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

where $R$ is a risk aversion parameter and $\tau_i$ is now a threshold value of expected utility. Since we do not observe $p_{it}$ and $v_{it}^2$ directly, we replace them by formulas (2.3) and (2.4) above and we get:

$$\Pr(y_{it+1} = 1|\{y_{it_i}, ..., y_{it}\} = \{0, ..., 0\}) = \Pr\left((\mu - \tau_i) - R\frac{\mu(1-\mu)}{n_{it}} \geq -e_{it+1}\right) \tag{2.5}$$

Equation (2.5) 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 $\mu$ is not higher than $\tau_i$ for everyone.

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

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

[5] Another possibility would be to express the threshold as a percentile in the posterior distributio, e.g., require that the median of $p_{it}$ is at least 60%, say. This is more intellectually satisfying but seems harder to implement analytically.

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 $\theta_v$,[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 $\theta_v$ 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]

$$n_{it} = \theta_v S_{it} + qA_{it} + u_{it} \text{ with } u_{it}\tilde{}(0, v^2) \tag{2.7}$$

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

$$\Pr(y_{it+1} = 1 | \{y_{it_i}, ..., y_{it}\} = \{0, ..., 0\}) = \Pr\left((\mu - \tau_i) - R\frac{\mu(1-\mu)}{\theta_v S_{it} + qA_{it}} \geq -e_{it+1}\right) \tag{2.8}$$

Equation (2.8) 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 (2.6) by its value given by (2.7). Our earlier observation remains valid: with strong prior beliefs, the variance term that multiplies $R$ in equation (2.6) can initially be quite small. If the prior belief $p_{i0}$ 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 $\tau_i$, 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 (2.6) and adoption will spread among individuals with a sufficiently high valuation $\tau_i$ 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 $\tau_i$; social learning no longer matters. Formally we have:

$$\Pr(y_{it+1} = 1 | y_{is} = 1 \text{ for some } s \leq t) \ = \Pr\left((\mu - \tau_i) \geq -e_{it+1}\right) \tag{2.9}$$

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 (2.6) 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 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. 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.

### 2.3. Individual experimentation

So far we have ignored the possibility that an individual may experiment with a product in order to learn its quality. If the cost of a single usage is large, such experimentation may be strongly discouraged. But if the cost of a single usage is small, it may be optimal for an individual to experiment individually rather than waiting to learn from others.

Formally introducing experimentation into our model goes well beyond the scope of this paper. But it is fairly straightforward to see how it would affect our predictions so far. If social learning is non-existent, the only way people can learn about the quality of the product is to experiment themselves. The presence of social learning, however, generates an option value of waiting: individual may simply free-ride on the experimentation of others, and wait to receive information through their network (e.g., Foster and Rosenzweig 1996).

How does this observation affect model predictions? Intuitively, individuals who are socially well connected can anticipate to receive a lot of information through their network and thus have a high option value of waiting; in contrast, individuals who are not well connected have a low value of waiting and thus are more likely to experiment by themselves.[8] It follows that, other things being equal, high degree individuals have a lower $\Pr(y_{it} = 1|\{y_{it_i}, ..., y_{it-1}\} = \{0, ..., 0\})$

[8] This is demonstrated formally by LeDuc (2015) for the case when social learning is free and network information decays rapidly, i.e., does not travel more than distance 1.

than low degree individuals.

Let the network degree of individual $i$ at $t$ be denoted as $B_{it}$. This degree includes all direct neighbors of $i$, irrespective of whether they have already adopted or not. We expect $\Pr(y_{it} = 1)$ to be a non-increasing function of $B_{it}$, e.g.:

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

Experimentation does not indicate 'adoption' in the usual sense: experimentation can reveal to the individual that the product is not suitable for them, i.e., that its quality falls below $\tau_i$ – in which case they would never use it again. This means that people who adopt early because of a low $B_i$ are also more likely to abandon the product after having used it once. Thus we have:

$$\Pr(y_{it+1} = 1 | y_{is} = 1 \text{ for some } s \leq t) \text{ is increasing in } B_{it}$$

These predictions are empirical testable if the researcher has information about $B_{it}$.

### 2.4. 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. Some of these externalities may occur geographically, others may be specific to the social network. 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 these strategic complementarities and social learning is that the effect of social learning disappears after $i$ has used the product at least once; the effect of other strategic complementarities do not. This simple observation forms the basis of our identification strategy between social learning and other network externalities.

### 2.5. 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 (2.6) 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 (2.6) is similar and can be written as:

$$\Pr(y_{it+1} = 1|\{y_{it_i}, ..., y_{it}\} = \{0, ..., 0\}) = \alpha_i + \alpha_1 S_{it} + \alpha_2 A_{it} + \varepsilon_{it+1} \quad (2.10)$$

$$\Pr(y_{it+1} = 1|\{y_{it_i}, ..., y_{it}\} = \{0, ..., 0\}) = \alpha_i + \alpha_1 S_{it} + \alpha_2 A_{it} + \alpha_3 S_{it}^2 + \alpha_4 A_{it}^2 + \alpha_5 S_{it} A_{it} + \varepsilon_{it+1} \quad (2.11)$$

Model (2.10) is a simple linear approximation of the two structural models (2.1) and (2.6). Parameter $\alpha_i$ captures variation in product usefulness across individuals. With any social learning we expect $\alpha_2 > 0$ on average. We also expect $\alpha_1 > 0$: the likelihood of adoption should increase over time as more information about the product becomes more generally available, form within and outside the social network. In regression model (2.11) 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 $\alpha_5 < 0$.[9]

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:

$$\begin{aligned} \Pr(y_{it+1} = 1 | \{y_{it_i}, ..., y_{it}\} = \{0, ..., 0\}) &= \alpha_i + \alpha_1 S_{it} + \alpha_2 A_{it} + \alpha_3 S_{it}^2 + \alpha_4 A_{it}^2 + \alpha_5 S_{it} A_{it} \\ &+ \beta_1 S_{it} m_{it} + \beta_2 A_{it} m_{it} + \beta_3 S_{it}^2 m_{it} + \beta_4 A_{it}^2 m_{it} + \beta_5 S_{it} A_{it} m_{it} + \varepsilon_{it+1} \end{aligned} \quad (2.12)$$

with $m_{it} = 1$ if $M_{is} = 1$ for some $s \leq t$, and $= 0$ otherwise. As before $\alpha_i$ captures variation in product usefulness across individuals. If the true model is social learning only about existence, then all $\beta$'s should be equal to 0. If the true model is only social learning about quality, then all $\beta$'s should be equal to the corresponding $\alpha$'s. If we reject both hypothesis – 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 $i$

[9] The sign of the cross term is because information from the network is less valuable if the person has already received many signals from non-network sources.

has already used the product prior to period $t$. The estimated model is of the form:

$$\Pr(y_{it+1} = 1) = \alpha_i + \alpha_1 S_{it} + \alpha_2 A_{it} + \alpha_3 S_{it}^2 + \alpha_4 A_{it}^2 + \alpha_5 S_{it} A_{it} + \gamma_1 S_{it} z_{it} + \gamma_2 A_{it} z_{it} + \gamma_3 S_{it}^2 z_{it} + \gamma_4 A_{it}^2 z_{it} + \gamma_5 S_{it} A_{it} z_{it} + \varepsilon_{it+1} \quad (2.13)$$

Unlike models (2.11) and (2.12), regression model (2.13) 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 $\gamma_2 = \gamma_4 = \gamma_5 = 0$. 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 (2.13) 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.[10]

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

[10] 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

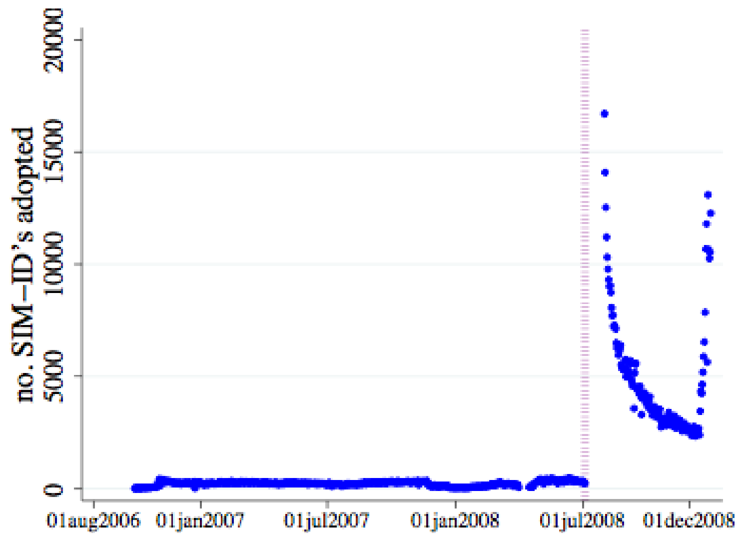
The data come from a large telecommunications operator. During the period of investigation, this operator enjoyed a quasi-monopoly on mobile phones in Rwanda. Access to the data was granted by Nathan Eagle through remote access to a Northeastern University computer server under conditions of strict confidentiality.[11] 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.[12]

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

[11] If one wishes to use this dataset, please contact Nathan Eagle at nathan@mit.edu.
[12] 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).[13]

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.

[13] For some poorly understood reason, the data shows an explosion of usage immediately after July 1st 2008, followed by a steady decline. For this reason, we ignore all ME2U data after that date.

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.[14] 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.

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.

[14] 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.

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.[15] 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:

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

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

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

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

### 3.3. Descriptive statistics

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 $\Delta A_{it}$ 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 |
| $\Delta A_{it}$ | 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 |
| $\Delta A_{it}$ | 1.553 | 1.883 | 101701 | 277140 |
| $A_{it}$ | 61.164 | 71.366 | 15790 | 282004 |
| $S_{it}$ | 38.617 | 26.165 | 4936 | 286940 |

## 4. Empirical results

The first regression model we estimate is (2.11), using only observations until first adoption. To eliminate the individual fixed effect $\alpha_i$, we first difference the data. The estimated model is thus of the form:

$$\Pr(\Delta y_{it+1} = 1 | \{y_{it_i}, ..., y_{it}\} = \{0, ..., 0\}) = \alpha_1 + \alpha_2 \Delta A_{it} + \alpha_3 \Delta(S_{it}^2) + \alpha_4 \Delta(A_{it}^2) + \alpha_5 \Delta(S_{it} A_{it}) + \Delta\varepsilon_{it+1} \tag{4.1}$$

where $\Delta x_t \equiv x_t - x_{t-1}$ by definition of notation. We have $\Delta S_{it} = 1$ by construction. Coefficient estimates are presented in Table 4. We see that $\alpha_2$ and $\alpha_3$ are both significantly positive while $\alpha_4$ 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

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 |
| $\Delta A_{it}$ | 1.983 | 2.058 | 29413 | 113375 |
| $A_{i,t}$ | 95.624 | 80.933 | 557 | 113503 |
| $S_{i,t}$ | 49.993 | 22.99 | 64 | 113567 |

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 $\alpha_4$. 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 (2.12). Once again, we eliminate the individual fixed effect $\alpha_i$ by first-differencing the data. The estimated model is thus of the form:

$$\begin{aligned}\Pr(\Delta y_{it+1} &= 1|\{y_{it_i}, ..., y_{it}\} = \{0, ..., 0\}) = \alpha_1 + \alpha_2 \Delta A_{it} + \alpha_3 \Delta(S_{it}^2) + \alpha_4 \Delta(A_{it}^2) + \alpha_5 \Delta(S_{it} A_{it}) \\ &\beta_1 \Delta(S_{it} m_{it}) + \beta_2 \Delta(A_{it} m_{it}) + \beta_3 \Delta(S_{it}^2 m_{it}) + \beta_4 \Delta(A_{it}^2 m_{it}) + \beta_5 \Delta(S_{it} A_{it} m_{it}) + \Delta\epsilon_{it+1} \quad (4.2)\end{aligned}$$

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 (2.13) in the same data. The model is estimated in first difference to eliminate unobserved heterogeneity $\alpha_i$, i.e., it is of the form:

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

where $z_{it} = 1$ 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 $\alpha$ 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 8. 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 $\Delta A_{it}$ 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 $\Delta y_{it+1}$ and $\Delta A_{it}$. 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, $\Delta y_{it+1}$ and $\Delta A_{it}$ may be negatively correlated in the sense that the more $i$ needs assistance, the more he or she receives airtime transfers, hence the larger $\Delta A_{it}$. At the same time, the more $i$ needs assistance, the less $i$ can help others and hence the lower $\Delta y_{it+1}$ 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.

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. 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.

## 6. References TO BE FILLED IN AND EXTENDED

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 (2011). "Risk and Reciprocity Over the Mobile Phone Network: Evidence from Rwanda," CSAE Working Paper Series 2011-19, Centre for the Study of African Economies, University of Oxford.

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

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

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

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

van den Boogaart, Monique (2013). "Thesis title", Department of Economics, Oxford University (unpublished PhD thesis)

**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).