THE ADOPTION OF NETWORK GOODS:
EVIDENCE FROM THE SPREAD OF MOBILE PHONES IN RWANDA

SUPPLEMENTAL APPENDIX

DANIEL BJÖRKEGREN*

BROWN UNIVERSITY

*E-mail: danbjork@brown.edu, Web: http://dan.bjorkegren.com
Revision January 2, 2018.
I am grateful to Michael Kremer, Greg Lewis, and Ariel Pakes for guidance and encouragement. Thank you to Nathan Eagle for providing access to the data, computing facilities, and helpful conversations. For helpful conversations I am also grateful to Natalie Bau, Arun Chandrasekhar, Brian Dillon, Michael Dinerstein, Ben Golub, Rick Hornbeck, Sabrina Howell, Max Kasy, Divya Kirti, Daniel Pollmann, Martin Rotemberg, Heather Schofield, Mark Shepard, Rob Stewart, Laura Trucco, and Tom Zimmerman. In Rwanda, I thank the staff of my telecom partner, RURA, RDB, CGIS-NUR, NISR, Minicom, PSF, EWSA, and BRD for helpful conversations and access to data. This work was supported by the Stanford Institute for Economic Policy Research through the Shultz Fellowship in Economic Policy.

Contents

Part 1. **Measurement**

S1. General Measurement

This paper reports most prices in United States Dollars; I convert these from Rwandan Francs (RwF) using the mean exchange rate of 552 RwF to the dollar.[1] I adjust prices for inflation using the IMF consumer price series for Rwanda, which averaged 9% inflation per year over this time period.

Accounts. There are 2,092,477 accounts ever referenced in the data, but many do not appear to represent active accounts. For the analysis, I omit the 528,737 accounts that have made fewer than 10 outgoing calls, and 38,679 further accounts for which the time spanned between the first and last observed transaction is less than 90 days (some of these are short term visitors to the country). This results in a sample of 1.5m accounts.

Account openings and closings. I infer an account as opened the date that the first transaction is made from it. Account are not explicitly closed; prepaid accounts that are not topped up regularly are disabled by the operator but can be used again when next topped up. Some accounts cycle through periods of being disabled but many are used again later; for this reason I ignore the possibility of account closure.

Communication graph (social network). Since the decision to communicate over the phone depends on whether it is possible to communicate in person, the measured call graph is conditioned on individuals' geographic locations. If there were internal migration, these locations would change over time, making it difficult to interpret the measured graph. Permanent internal migration is low in Rwanda over this time period (Blumenstock, 2012).

Adopting a phone may transform an individual's social network—they may keep in touch with friends or family living further away, for example. I uncover the communication graph after any transformation associated with adoption: the graph conditional on phone ownership. The inference in this paper remains valid as long as any such transformation is anticipated and coincides with adoption.

Calling Prices. Prior to January 2006, calls were billed by the first minute and each subsequent half minute; after, subscribers could opt in to per second billing (and most quickly did). Modeling the per-minute charges would add significant complexity, so instead I assume these calls were billed at an equivalent per second price, selected to approximate

[1]The exchange rate was relatively stable over the period of data (1.2005-5.2009): average of selling and buying price ranged between 543 and 570 RwF to the dollar, National Bank of Rwanda.

both marginal and average prices. I set the per second price to the equivalent charge under the per minute rate when calls are of length 30 seconds.

Raw coverage. I predict the coverage of mobile phone service at each location and time using tower locations and a elevation map. Tower coordinates (latitude and longitude) for most towers were provided by the operator. For those 12% of towers whose locations are missing from these records, I infer the location based on call handoffs with known towers, using a method detailed in Bjorkegren (2014). I infer the date each tower becomes operational by the date the first transaction that flows through it; I assume that once built, towers are never taken offline. Elevation data is from NASA Shuttle Radar Topographic Mission (SRTM) data, at 90m resolution; I use the version of the data from Jarvis et al. (2008), which has been processed to fill in data gaps.

If I had more information on the towers (specific equipment, tilt, antenna design), it would be possible to precisely predict coverage with the same commercial packages used by operators for coverage planning. As an approximation I predict coverage based on uninterrupted visibility, using the viewshed tool in ArcGIS. Based on the recommendations of the operator's network planner, I assume the antenna on each tower is located 35m above the ground, all antennas are omnidirectional, and that the signal has a maximum range of 15km.[2] I threshold the resulting image so that it indicates whether each location has coverage from at least one tower. This provides a raw coverage map for each month, which is my best estimate of the network availability at each location.

I also omit some features of the market:

Handset sharing. Given the high cost of handsets, sharing is common. 55% of Rwandan phone owners report they allow others to use their handset regularly (Stork and Stork, 2008). An individual may open an account but use it with others' handsets, by inserting their SIM card, but this practice is rare.[3] It is more common that a person borrows another's handset and account.[4] The model assumes that each node in the network represents a unitary entity such as an individual, firm, or household. I assume that this entity weighs

[2]Although the maximum technical range of a GSM tower is 35km, the range in practical use tends to be smaller.

[3]This allows them their own phone number and balance, but it is difficult to receive calls. A representative survey found that fewer than 1% of individuals in 2007 owned SIMs without handsets (Stork and Stork, 2008), and within the phone data on average there are actually 3% more handsets than accounts active in a given month.

[4]This pattern would include the use of payphones that run on the mobile network, which I omit from this analysis. Payphones place approximately 12% of call durations but receive only 0.8%. Because payphones receive so few outgoing calls from the rest of the network, omitting them would have little effect on the preferred usage model which uses outgoing calls.

the communication benefits accruing to the node against the cost of adoption, and that the communication graph is fixed over time. If multiple people use a particular phone, then the node will represent their aggregate demand. The model will correctly account for this demand if the composition of people using a particular phone is fixed over time and the adoption decision takes into account the utility of all users (for example, if the owner internalized the benefits of other users' calls through side payments). If the composition of people using a particular phone changes in response to adoption (say, if a couple initially shares a phone but later obtains separate phones and splits its communication), then the communication graph I estimate will be similar to a weighted average of the underlying networks. In that case, during simulations the nodes will not account for changes in usage when borrowers obtain their own phones, nor coordination of adoption times between the nodes.[5]

SMS and Missed Calls. I do not explicitly model utility from SMS and missed calls. If different relationships use different modes of communication, this omission will underweight the importance of SMS and missed-call relationships in the adoption decision. The data suggests that the different modes pick up slightly different relationships: the correlation between a node's total calls and total SMS is 0.53[6], and the correlation between calls and call attempts within a link is 0.58. The price for sending an SMS is relatively high throughout the period ($0.10, the same as a call of 24 seconds under the lowest peak price) and remains constant from 2005-2009. There appears to be little substitution between communication modes as calling prices change.

Other Omissions. I omit the cost of charging a phone (the four most popular handsets have more than two weeks of battery life on standby). Accounts must be topped up with a minimum denomination of credit (the minimum was $0.90 by the middle of the data); I treat these charges as continuous rather than lumpy.

S2. Household Surveys

In the paper I report background statistics from several household surveys:

[5]Modeling changes in phone sharing would require making assumptions about the set of borrowers for each handset over time, the allocation of utility between owner and borrower, and the hassle cost of borrowing a phone to place a call; these would be difficult to defend.

[6]There are a small number of users who use SMS heavily; to prevent these users from skewing the statistic, I compute the correlation omitting the top 1% of SMS users.

Demographic and Health Survey 2005 and 2010 (DHS). These are representative surveys of 10,272 (2005) and 12,540 (2010) Rwandan households, asking about demographics, ownership of goods, and use of services (electricity, water, phone, radio, television).

Rwandan Household Survey 2005-6 and 2010-11 (EICV 2 and 3). These are representative surveys of 6,900 (2005) and 7,354 (2010) Rwandan households, asking about demographics and consumption.

Research ICT Africa Household Survey 2007-8 (Stork and Stork, 2008). This is a representative survey of 1,078 Rwandan households (201 with phones), asking questions about information technology for one randomly selected household member. Used to provide background on how phones are used in this context throughout the text.

For each survey I apply nationally representative sampling weights.

S3. Inferring Subscriber Locations

The call data reports the location of the cell tower used at the start and end of each call. From the sequence of cell towers used, it is possible to infer an individual's location. At any point during a transaction, a mobile phone handset sends packets of information to one cellular tower, using electromagnetic waves. This tower routes these packets to the rest of the network using either fiber optic cables or a different electromagnetic frequency; the packet is sent to a tower near the receiver and ultimately delivered to the receiver's handset.

Handsets tend to transmit information to a close, unobstructed tower, so that the tower used represents an approximation to the individual's location at that point in time. Calls can bounce between towers due to call traffic, variation in the weather, if a tower is down, or if the handset is in motion. The maximum technical range of a GSM tower is 35 km, but in areas of higher tower density the range is reduced to lower interference.

There is a literature on inferring a subscriber's location based on usage traces (Gonzalez et al., 2008; Isaacman et al., 2010, 2011; Blumenstock et al., 2011). My settings differs from these papers in two ways: the tower network was rapidly expanding, and usage is sparse. I implement a modified version of the 'important places' algorithm as detailed by Isaacman et al. (2011), which for each user identifies one or more important places where they spend time. The paper finds that the identified places were within 3 miles of reported places for 88% of a small validation sample of users in the U.S., with a median error of 0.9 miles. I have modified the algorithm to improve performance in rural areas.

To find the important places for individual i, the algorithm proceeds as follows:

(1) The towers that i has ever used, X_i, are sorted by the number of days i used that tower, d_{ix}

(2) The most used tower forms the start of a new cluster, located at that tower's location.

(3) If the next most used tower falls within a distance threshold of the cluster, it is added to that cluster, and the cluster's location moves to its new centroid (weighted by the days each tower is used). If the tower does not fall within the threshold, it forms a new cluster. The original paper uses a fixed threshold of 1 mile, with which they obtain good results in an urban setting. To allow for good performance in urban and rural areas (high and low tower densities), I compute an adaptive threshold specific to each tower related to the density of towers nearby. In considering the distance from tower x to a cluster, I use a threshold equal to the distance from x to the 9th most distant tower as of May 2009. This adaptive threshold allows the algorithm to smoothly incorporate a large radius of spatial information in rural areas and a narrow radius in urban areas.

(4) The previous step is repeated for each tower: if the nearest cluster is within this tower's threshold, the tower is assigned to that cluster and that cluster's centroid is updated; if the nearest cluster is further away, the tower is assigned to a new cluster.

(5) After all towers have been placed in clusters, each cluster is ranked by the combined days that the individual made calls from that cluster (counting each day only once if transactions were made on multiple towers within that same cluster).

This algorithm has advantages for this setting: it uses the full panel of data, which improves precision when transactions are sparse, and works well with an expanding network: estimates simply become more precise as tower density increases.[7]

S4. Handset Price Index

In order to back out η's in the adoption decision, I create a handset price index.

First, I compile a dataset of each handset model and how its price has changed over time. The country is small (the furthest you can get from the capital is about a 3 hours drive), and handsets all are imported through a small number of distributors, so I assemble one price series for each handset, assuming that at any point in time the price a consumer would pay for a particular handset is uniform across the country.

Price Data from Retailers. Handsets can be purchased through the operator directly or through third parties. I use three sources of historical handset prices. I have two sources of

[7]The determination of clusters could be disturbed if there were measurement error in the tower locations (e.g., a tower placed with error may end up bridging two clusters that should be separate). For this reason, in determining a subscriber's location I ignore the use of towers for which I have only a predicted location.

price data from the operator (retail prices posted on the operator's website from 2004-2012, and the operator's internal sales database from 2005-2012), and a third from an independent shop in Kigali covering sales between 2005-2009. Operator sales records account for only 10% of activated handsets, so independent sellers are important.

The operator's website posted prices of handsets for sale; historical versions were accessed using archive.org. The retail prices posted on the website are taken to be the correct retail prices when published; however, only a small number of handsets were listed and the prices were not updated regularly (the data includes 36 handset models and 44 total observations.). I merge these data to the operator's internal sales records, which list each model sold by the operator each month of each year, with price and quantity sold (including 109 handset models and 1327 total observations). The sales records and website have 18 model-months in common, which can be used to compare the datasets. Although the prices are highly correlated (0.98), retail prices are higher by an average factor of 1.26 (with standard deviation 0.15), suggesting the sales records report an internal accounting below the full retail price paid by consumers.[8] I inflate the internal sales prices by this factor to estimate retail prices. Also, if the sales records report the same handset model sold at different prices in the same month, I use the price that was applied to the largest quantity sold.[9]

The independent shop data includes 66 handset models and 1807 total observations. The shopkeeper who provided the records appears to have written them based partially from archived receipts and partially from recall.[10]

In all three datasets, prices follow a general pattern of decline. Amidst this decline, some of the operator's internal prices bounce around slightly over time. However, the operator's sales represent a small share of the market, and these bounces are not seen in the independent shop's records.[11] If these bounces represent temporary sales, it would be easy for someone to buy a discounted handset from the operator and resell it. I take the cumulative minimum price as a measure of the price level of each handset. This would be invalid if there were temporary factors that drove prices up; however, the bounces appear random. The final

[8]I omit 1 outlying observation for which the sales record lists a higher price than the website.

[9]In some months the same handset model was recorded at different prices; usually a large quantity were noted at a predominant price and then a handful were noted at a different price. This observation, combined with the fact that the operator's website does not suggest the possibility of any price differentiation, lead me to interpret these to be either the result of internal accounting or special purchases.

[10]The shop was one of the earlier handset sellers in Kigali and the staff was knowledgeable about the market when quizzed. This data should be viewed as an expert's estimate at the local price each month.

[11]The independent shop's records show 1,432 model-months with no price change, 269 with decreases, and 6 with increases. The operator's records show 848 model-months with no price change, 187 with decreases, and 146 with increases.

cumulative minimum operator prices correspond closely with the levels reported by the independent shop.[12]

Assembling a Price Index. I use the operator's prices for the model-months covered. I fill in any missing data points with prices from the independent shop.

This sometimes leaves observation gaps after a model was introduced. I fill interior gaps (after the first observed price) with the price from the nearest observed month. Older models are likely still available for purchase somewhere in the market, so I impute trailing gaps (after the last month I observe a price for a given handset) with a price based on predicted decline.[13] Each handset model follows a predictable price path: it is introduced at a high price, and that price is gradually reduced as it becomes obsolete. I estimate this trend, $\log p_t^h = \alpha_h + \beta t + \epsilon_{ht}$, with a fixed effect for each model (where each handset is represented by h) and an exponential time trend (linear in logs). This model fits observed declines with $R^2 = 0.95$, implying an average price decline of $(1 - e^{\beta}) \sim 0.32\%$ per month.[14]

This generates a price series p_t^h for 160 handset models. I create a single price index by weighting by the total quantity of each handset model activated in the data, Q_h: $p_t^{handset} = \frac{\sum_h Q_h \cdot p_t^h}{\sum_h Q_h}$.[15] Many handsets were introduced during this time period; to control for any potential quality differences, in months before a handset model h was available for sale, I assign its weight to the adjacent model of higher quality.[16] I deflate the resulting price series to account for inflation as discussed in S1. Figure S1a shows the resulting price index.

In the adoption model, consumers forecast future prices to decline with an exponential ($E_t p_x^{handset} = \omega^{x-t} p_t^{handset}$, for $\omega = \left(\frac{p_T^{handset}}{p_0^{handset}}\right)^{\frac{1}{T}}$). Figure S1b shows forecasted price series for different origin dates.

[12]The handset shop and operator sales database share 270 observations in common. Among these observations, the correlation in prices between these is 0.76 and mean ratio of prices is 0.98 (standard deviation 0.36).

[13]While models appear enter the market at similar times across the retailers, exit is not as coordinated: the correlation between the first month a model is seen in operator and independent shop records is 0.71 but the correlation between last months is 0.42.

[14]Altogether, in the main period of my data 2005-2009, I have 2108 price observations; I fill in 302 model-months in the interior and 1,242 after. I also extend this series past the end of my data; this extended series is only relevant in simulations if a consumer delays adoption after the end of my data (see discussion in the paper in Section 7). Over the entire period 2005-2012, I have 2,848 price observations; I fill in 469 model-months in the interior and 6,068 after.

[15]I use an independent registry (Mulliner, 2013) to match handsets in the data to their model names. Using this database I am able to match all but 44 359 of the 1 377 836 total handsets used in the call data.

[16]Although the most popular handset models are similar in observed characteristics, there is substantial variation in contemporaneous prices, suggesting differences in unobserved quality. For simplicity, I do not enrich the model to allow for handset model choice. I select the handset h' that was next most costly to h in the first time t when both were available (and thus presumably of higher quality), which is likely the handset that the individual would have chosen in the absence of h. I assign weights in this manner starting backwards from the end of the data towards the beginning.

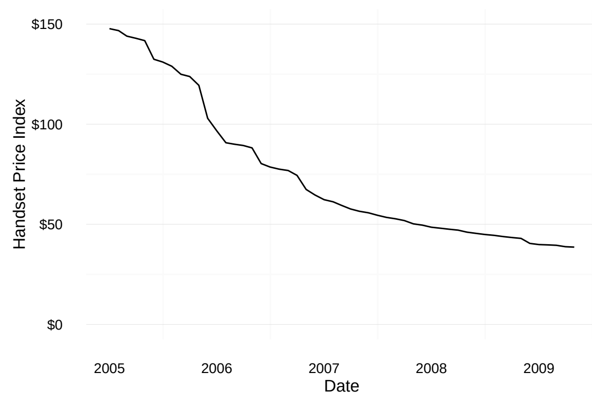
Part 2. Call Model Estimation

S5. Reduced Form Evidence

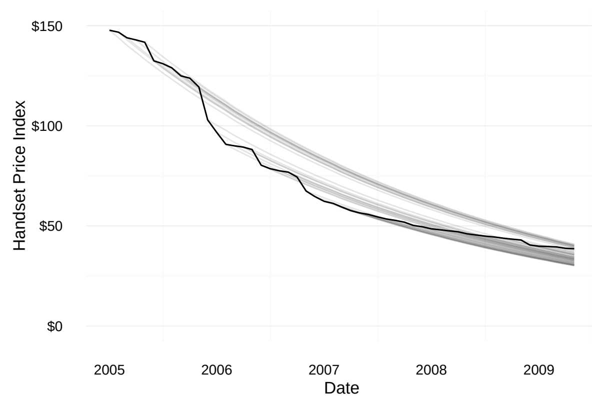
I identify the value of each link based on how communication across that link changes in response to changes in price and coverage. This approach relies on the identification assumption that the latent desire to communicate (ϵ_{ijt}) is uncorrelated with costs (p_t and $\phi_{it}\phi_{jt}$). This could be violated if the latent desire to communicate across a link trended over time for other reasons. Here I consider two potential reasons: potential complementarity or substitutability between links, and that an individual's usage may trend over time (for example, if it takes time to learn to use a phone).

A typical demand model would suggest links are substitutable: when my friend Jacques buys a phone, I may call him more and my brother less. An information sharing model would suggest complementarities: Jacques and my brother may share additional information, and as a result I may call both more. One simple test of dependence is whether the volume of calls across a link changes as more of the sender's and receiver's contacts join the network.

Figure S1. Handset Price Index
(a) True Price Index ($p_x^{handset}$)



(b) Forecasted Price Index ($E_t p_x^{handset}$ for different origin dates t)



To test this, I estimate a simple gravity model, regressing each link's monthly call volume on the sender's and receiver's number of subscribing contacts, controlling for price changes and coverage, and including fixed effects for each link. If links were substitutable, as new contacts join the network a subscriber would reduce calls to existing contacts; barring any confounds this would result in a negative coefficient on number of contacts. Complementarity would result in a positive coefficient. Table S1 presents these results; specification I shows a baseline model without these tests. Specifications II-IV add controls for sender's and receiver's number of contacts; results are consistent with dependence between links being small, and on net complementary.

The next set of specifications test whether an individual's usage trends over time by adding a control for months since the sender adopted. If it takes a new user time to learn to use a phone, I would expect a positive coefficient on this term; alternately if new users excitedly use their phones at the beginning before settling into a pattern of lower use, I would expect a negative coefficient. In all cases the magnitude of the coefficient is small: Specifications V finds a slight positive coefficient; it turns negative when I control for contacts on the network in Specification VI.

These estimates are consistent with these potential confounds being negligible. For the median subscriber, the change in duration associated with the change in time and contacts on the network is roughly 1% of the change associated with the changes in prices and coverage over this time period (+5% associated with the change in contacts, and -4% with months since adoption).

Finally, Specification VII shows that the interaction of sender's and receiver's coverage appears to be the most important term explaining the effect of coverage on calls across a link. To simplify the model, I assume the utility obtained from a contact is independent of the state of other contacts on the network, do not model individual time trends, and model the hassle cost of imperfect coverage as the interaction of sender and receiver's coverage.

S6. Notes on Estimation Procedure

I maximize the log likelihood function using KNITRO, using analytic gradients and hessians, in a two step procedure.[17] Extremely long calls can lead to numerical issues because

[17]Because the subsample used in the first step contains less variation than the full data, two potential issues can arise. First, the subsample is used to estimate the link cohort × average coverage fixed effects. While the subsample includes many of the most common combinations, it does not include all possible combinations (it includes 493/583 possible combinations). For the approximately 0.03% of links whose link cohort × average coverage fixed effect was not estimated in the first step, I interpolate from the nearest estimated fixed effect (nearest in average coverage and then link adoption month). Second, there is an issue with this approach

Table S1. Determinants of Calling

Dependent Variable: Duration *(seconds per month, outgoing)*	I	II	III	IV	V	VI	VII
Price USD/minute	-45.44	-43.04	-43.85	-42.80	-44.82	-42.68	-43.63
(event study)	(0.84)	(0.84)	(0.84)	(0.84)	(0.84)	(0.85)	(0.85)
Sender's Coverage	16.43	12.70	15.11	12.67	13.91	12.97	0.20
	(0.37)	(0.38)	(0.37)	(0.38)	(0.38)	(0.38)	(0.69)
Receiver's Coverage	25.11	22.89	22.50	22.35	22.47	22.64	9.78
	(0.38)	(0.38)	(0.39)	(0.39)	(0.39)	(0.39)	(0.69)
Sender's × Receiver's Coverage							20.75
							(0.94)
# Sender's Subscribed Contacts		0.025		0.023		0.025	0.023
		(0.001)		(0.001)		(0.001)	(0.001)
# Receiver's Subscribed Contacts			0.017	0.005		0.006	0.004
			(0.001)	(0.001)		(0.001)	(0.001)
Months Since Sender's Adoption					0.062	-0.015	
					(0.002)	(0.003)	
$N_{observations}$	44,108,852	44,108,852	44,108,852	44,108,852	44,108,852	44,108,852	44,108,852
N_{links}	1,663,018	1,663,018	1,663,018	1,663,018	1,663,018	1,663,018	1,663,018
R^2	0.0010	0.0011	0.0010	0.0011	0.0010	0.0011	0.0011

All regressions include link, month of year, and price regime fixed effects. Estimates computed using incremental least squares, on a 1% sample of nodes and all their links. The price coefficient is estimated based on an event study around the two price changes, in January 2006 and February 2008, using a two month window before and after. Dummies are included for the other months within each price regime. The top 1% degree nodes have been omitted; their inclusion attenuates the contact coefficients. Standard errors reported in parentheses. R^2's omit contributions of fixed effects.

they result in draws from the extreme tails of the normal distribution, so in the first step I omit the 1% of nodes that have talked to a contact longer than one hour in a given month. In the second step, convergence is less sensitive so I am able to estimate the parameters of all nodes.

The solver may not converge if the problem is initialized to a poor starting point; for this reason in practice I iteratively solve for joint parameters. In iteration k, I solve for joint parameters on a subset of size N^k, obtaining a set of parameters Θ^k. I then select $N^{k+1} > N^k$, and solve for the individual parameters of the additional individuals conditional on Θ^k. These estimates $\tilde{\Theta}^{k+1}$ do not maximize the likelihood for iteration $k+1$ because some of the individual parameters were solved conditioned on other common parameters; however, I use them as a starting point for iteration $k+1$. To guard against finding a local, rather than global, optimum, I optimize from several perturbed starting points.[18]

S7. Monte Carlo Exercise

To test the method, I generate simulated data from the model, and compare estimated results against the results obtained from the true parameters for that test specification.

For each test specification, I specify common parameters and a network with a certain number of nodes, links per node, and months observed across each link. After specifying the network, I draw node and link parameters from a specified distribution. I draw prices and coverage for sender and receiver randomly from the true sequences observed in the data. For each test specification, I run at least 100 simulations. In each simulation, I draw communication shocks and compute observed behavior. I then estimate parameters and compute expected durations and utilities for a random sample of links using Monte Carlo integration.[19]

---

relating to edge cases of coverage. The form of utility function implies that there is a cutoff level of cost (hassle cost of coverage and prices) above which no calls will be placed, regardless of the shock (I discuss this implication further in the functional form Appendix of the main paper). A random subset is unlikely to include the envelope of observations representing the highest cost instances under potential coefficient estimates. If common parameters are estimated off of such a subset, when applied to a set of links with a higher cost instance they could imply that an observed duration has zero likelihood (since the cost lies above the estimated cutoff). This would not be a problem if the full estimation could be done jointly. To correct for this, I constrain the estimation of the subproblem so that it is consistent with a nonzero probability of calling at the envelope of high cost instances.

[18] As the subsample becomes large, KNITRO's performance begins to deteriorate, so I am not able to do a joint estimation with the full sample as a final step.

[19] I estimate each replication 10 times from different perturbed starting points, and take the one that maximizes the likelihood. In most cases the estimates converged to are very close, but in some cases one estimate diverges.

To guide the selection of test networks, I compute the distribution of observations per link in the transaction data, in the first set of columns in Table S2. Most links were available for many months (months where both nodes have subscribed), but most have few months with calls. Links are available for a mean of 32.5 months, and 51.7% are available for two years or more. Links that are available for many months also account for more call duration; 67.6% of total duration is on links that were available for two years or more. However, a large fraction of links have few months with calls: 65.4% had one month with calls, and these account for 14.8% of total durations. Both months with and without calls provide information about parameters: the model explains the former with draws from the communication graph that exceed a cutoff value, and the latter with draws that are below this cutoff.

In the second set of columns in Table S2 I show a distribution of links that I use for Monte Carlo simulation. I avoid simulations with censored observations (months with zero calls) due to a selection issue discussed next.

Selection. Because the network is observed over a finite number of periods, there may be some links where no calls are observed. This is also a concern with the empirical part of this paper: there may be latent links over which I do not observe calls, but calls would have been placed if different shocks had been drawn or if the cost of communication were lower. In the empirics, this is not a problem for counterfactuals that make communication less favorable: under the communication shocks that were realized these latent links would not have become active. It could be a potential issue for counterfactuals that make communication more favorable: the same shock may have led to a link becoming active. Because it would be difficult to justify the assumptions that would be needed to uncover all these links (assumptions on which nodes are linked and the distribution of link strengths), I do not consider what would have happened under alternate communication shocks.

This selection presents an issue in interpreting the Monte Carlo results. The estimation method assumes that the data generating process (DGP) is defined only over links that are observed, but when I simulate data from this data generating process allowing different communication shocks, a link may be observed in some draws and not in others. If I observe the full network in all draws (either the probability of communication across a link is very high, or the number of time periods is very large), then the model I estimate will correspond with the simulation DGP. However, if some links are not observed, then the simulation DGP does not correspond to the one assumed by the estimation procedure. For example, consider a network with 3 nodes A, B, and C, where in the data a call was observed between A and

Table S2. Distribution of Link Observations

T	Distribution in Data: Fraction of links available for T months (total observations)		Distribution in Data: Fraction of links with T months with calls (uncensored observations)		Mimicked Distribution for Monte Carlo Simulation: Fraction of links available for T months (total observations)	Mimicked Distribution for Monte Carlo Simulation: Fraction of links with T months with calls (uncensored observations)
	% Links	% Duration	% Links	% Duration	% Links	% Links
1	0.00	0.00	65.4	14.8	50.0	50.0
2	0.00	0.00	15.1	10.5	16.6	16.6
3	0.02	0.01	6.6	8.1	16.6	16.6
4	0.1	0.05	3.7	6.7		
5	0.5	0.2	2.3	5.5	16.6	16.6
6	0.1	0.0	1.5	4.5		
7	0.4	0.2	1.1	4.0		
8	1.4	0.8	0.8	3.5		
9	1.5	0.9	0.6	3.2		
10	1.6	1.0	0.5	2.9		
11	2.2	1.5	0.4	2.6		
12	2.6	1.8	0.3	2.3		
13	3.3	2.1	0.2	2.2		
14	3.6	2.5	0.2	2.0		
15	4.4	2.7	0.2	1.9		
16	4.3	2.8	0.1	1.7		
17	4.2	2.7	0.1	1.6		
18	3.4	2.4	0.1	1.4		
19	3.7	2.7	0.1	1.5		
20	3.1	2.2	0.1	1.4		
21	3.1	2.1	0.1	1.2		
22	2.7	2.0	0.1	1.1		
23	2.3	1.7	0.05	1.1		
$\geq$24	51.7	67.6	0.3	14.4		

Total months represent months with either censored or uncensored calls; months with calls represents only uncensored months. Computed on 1% sample.

B but not between A and C. In my estimated network, there will be no link between A and C, and zero probability of a call between them. There will be a link between A and B. However, if I generate data from the estimated network, in some draws there will be no call between A and B, so the link is omitted in the generated data. When a link is omitted, it may affect the individual level parameters (shock mean μ_i, shock standard deviation σ_i, and probability of no call at any price $(1-q_i)$. The values of these parameters generating the data in the simulation would differ from that measured in estimation on links that had high enough draws to be observed (the conditioned distribution). If I was attempting to recover the latent network, then the simulation DGP is the relevant truth. However, it does not represent the DGP that corresponds to the network with latent links removed, which is what I aim to estimate; that DGP would be difficult to simulate because it assumes a communication shocks are distributed lognormally after being conditioned on observing a call.

Because it would be difficult to interpret a comparison between a DGP generating low probability links and an estimated model conditioned on links being observed, I use a different strategy to test the procedure. I simulate shock distributions with high means, for which the simulation DGP begins to match the estimated model. These correspond with higher probabilities of calls than I observe in the data. So, T months drawn from a high shock distribution will provide more information about parameters than T months drawn from a low distribution, and would overstate the performance that could be expected with actual data. To account for the increased amount of information per month, I simulate from fewer months than is observed in the data. For example, in the data I may observe a link with a low distribution for 50 months, where calls are observed in only 5 of those months. As an analogue to this scenario, I may simulate a link with a high distribution for 5 months, so that calls are observed in all 5 months. Both scenarios have the same number of parameters and the same number of months with calls observed, but the analogue is conservative in the sense that the 45 censored observations in the real world scenario provide more information about parameters.[20]

Results. I present the results of these tests in two tables: Table S3 for one set of parameters and Table S4 for a second set. Both present results for a high shock distribution (high means ($\mu_i \sim Triangular[5,6,7]$), moderate standard deviations ($\sigma_i \sim Triangular[0.5,1,1.5]$), and no censoring that is independent of cost ($q_i = 1$)). In all results I omit the link fixed effect terms ($\mu_{\max(x_i,x_j),\overline{\phi_{it}\phi_{jt}}} = 0$), as these are small in number and do not grow with the size of the sample. These simulations use a network of $N = 50$ nodes.

[20]Performance between the real world scenario and its analogue may also differ if the procedure has differential performance for high and low shock distributions, or due to deviation between the model specification and the real world DGP.

The columns in Table S3 show three different tests. The first two tests show how the performance of the method is affected by panel length T. Because in the data, few links are active for this many periods, I provide these nodes smaller networks: each has $D = 30$ links. Test 1 shows results from a panel of length $T = 50$, which is approximately the length of the data. Since many links are not active that many months, Test 2 is a more restricted panel with $T = 5$. Test 3 is extremely conservative, and mimics the distribution of observations in the uncensored empirical distribution (the equivalent of ignoring data from months without calls), allowing each node to have $D = 60$ links. Nodes have 10 links observed for 5 periods, 10 observed for 3, 10 observed for 2, and 30 observed for 1. (A comparison to the empirical distribution is shown in Table S2.)

The top two sections of Table S3 shows results on parameters. I present true values, the mean deviation of estimates from the true value (empirical bias), and the standard deviation of estimates. A first observation is that there appears to be negative bias in the estimate of α, which grows larger as the number of observations decreases (up to -51% in Test 3), and smaller amounts of bias in other parameters.

However, the main goal of the call model is to describe expected utilities (to compute consumer welfare) and durations (to compute revenue), and how these quantities change when prices and coverage are changed (to compute changes in individual decisions). Thus the primary test of the estimation procedure is how it performs on levels of these two quantities, and differences associated with changes in prices and coverage. To gauge how well the estimation recovers these quantities I have evaluated these quantities starting at the initial price (54.3 U.S. cents per minute, in 2005 dollars) and a level of coverage close to the initial median (70%), and how the quantities change either as price is reduced to the final price (11.0 U.S. cents per minute, in 2005 dollars) or coverage is increased to the approximate final median coverage (90%). These statistics are presented in the third section of each table. Apparent bias in these quantities of interest is much lower; in Tests 1 and 2 it is negligible. In the very conservative Test 3, I do find some bias: an average of 9% bias in expected durations and 15% in utilities; there is less bias in changes (7% and 10% respectively). However, the bias is only 21% the size of the standard deviation of these quantities across simulation draws.

Table S4 repeats these tests for different choices of the common parameters, and finds very similar results (average bias of 10% in expected durations, 13% in utilities, 7% in changes in durations, and 11% in changes in utilities; the bias is also 21% the size of the standard deviation).

I interpret these simulations as a worst case scenario for bias, and their results to show that even in this worst case, bias is limited. To the extent my method introduces bias, it would also be revealed when I evaluate model fit when using the data. Bias would show up in two places: biases in expected durations would be observed in the evaluation of model fit in Section S10 of this Supplemental Appendix. Biases in levels of expected utilities would show up when I compare the estimate of price sensitivity implied by the call model to that implied by the adoption decision (in Section 6 of the paper). These empirical comparisons also suggest that any bias is limited.

S8. Additional Estimates

I include individual fixed effects. These estimates are shown in Figure S2A(iii). There is wide variation among earlier adopters (some are very talkative and some are less talkative); among later adopters there is less variation.

However, not all of an individual's links may be available at the point they adopt: some may become available only later, after their contacts adopt. If an individual's early links differ systematically from their late links, then selection would confound a naïve estimate of price and coverage sensitivity.

To control for the difference in early and late adopting links, in addition to individual fixed effects I include link cohort times average coverage fixed effects. These fixed effect estimates are presented in Figure S2B. Link cohort is represented on the X axis and the average coverage across the link is represented by color.

The plot shows that the fixed effects pick up systematic selection. Conditional on an individual's overall fixed effect, the links that became active later and those that had higher coverage on average tend to be more talkative.

S9. Comparative Statics

I interpret these parameters using comparative statics in Table S5. For example, the top panel shows results for full coverage and the lowest price observed in the data: for the median link, the probability of making in a given month is 0.37, and conditional on making a call, the expected duration for that month is 84 seconds. The expected monthly cost of communicating across the median link is $0.07 and the link provides an expected utility of $0.27. The median individual has 61 links.

Table S3. Monte Carlo Test of Call Model Estimator: Parameters 1

All tests: $N = 50$		**Test 1:** $T = 50$, $D = 30$			**Test 2:** $T = 5$, $D = 30$			**Test 3:** Mimic uncensored empirical distribution $D = 60$		
Mean fraction of replications links observed		1.000			1.000			0.934		
Common Parameters		True	Mean Estimate	SD_{sim}	True	Mean Estimate	SD_{sim}	True	Mean Estimate	SD_{sim}
γ		2.300	2.283	0.011	2.300	2.240	0.035	2.300	2.176	0.038
α		100.000	89.596	2.496	100.000	69.649	6.739	100.000	49.532	5.016
β_{cost}		0.120	0.121	0.002	0.120	0.120	0.006	0.120	0.113	0.007
$\beta_{cost} \cdot \beta_{coverage}$		-0.600	-0.603	0.027	-0.600	-0.614	0.073	-0.600	-0.575	0.096
Communication Graph	True	True Mean	Mean Deviation	Mean SD_{sim}	True Mean	Mean Deviation	Mean SD_{sim}	True mean	Mean Deviation	Mean SD_{sim}
μ_i	$\sim Triangular[5, 6, 7]$	6.000	-0.060	0.399	5.999	-0.239	0.409	5.945	-0.438	0.385
σ_i	$\sim Triangular[0.5, 1, 1.5]$	0.981	-0.003	0.196	1.014	-0.076	0.201	1.012	-0.131	0.184
q_i	1.000	1.000	-0.014	0.019	1.000	-0.038	0.048	1.000	-0.032	0.040
Expected utilities and durations of randomly selected links		True Mean	Mean Deviation	SD_{sim}	True Mean	Mean Deviation	SD_{sim}	True Mean	Mean Deviation	SD_{sim}
$ED_{p_{high},\phi_i=0.7,\phi_j=0.7}$	seconds	86.926	1.108	38.948	90.053	-1.680	39.597	85.086	7.858	40.408
$ED_{p_{low},\phi_i=0.7,\phi_j=0.7}$		140.195	2.434	57.497	144.603	-0.849	59.125	137.219	11.847	59.980
ΔED price		53.269	1.325	19.470	54.550	0.831	20.584	52.133	3.989	20.682
$ED_{p_{high},\phi_i=0.9,\phi_j=0.9}$		110.168	1.596	47.039	113.921	-1.117	48.287	107.882	9.487	49.097
ΔED coverage		23.242	0.488	9.659	23.868	0.563	10.356	22.796	1.629	10.631
$EU_{p_{high},\phi_i=0.7,\phi_j=0.7}$	USD	0.429	0.003	0.219	0.449	-0.011	0.227	0.421	0.072	0.247
$EU_{p_{low},\phi_i=0.7,\phi_j=0.7}$		1.217	0.012	0.552	1.261	-0.021	0.566	1.190	0.139	0.591
ΔEU price		0.788	0.010	0.336	0.812	-0.010	0.344	0.769	0.067	0.350
$EU_{p_{high},\phi_i=0.9,\phi_j=0.9}$		0.715	0.006	0.344	0.745	-0.009	0.360	0.701	0.103	0.385
ΔEU coverage		0.286	0.003	0.130	0.296	0.002	0.140	0.280	0.031	0.147

Results of Monte Carlo simulations; results reported for at least 100 data sets generated with the stated parameters. Communication graph parameters are held fixed across each of these data sets. N is the total number of nodes, D the degree of each node, and T the number of time periods. Results that mimic the uncensored portion of the empirical distribution have 60 links total; 30 observed for 1 period, 10 observed for 2, 10 observed for 3, and 10 observed for 5. When applied to data, links also have many more censored observations and more nodes are included in estimation. SD_{sim} represents the standard deviation of the simulated estimates. Expected utilities and durations are computed for a sample of one link per node, randomly selected, using Monte Carlo integration with at least 1,000 draws.

TABLE S4. Monte Carlo Test of Call Model Estimator: Parameters 2

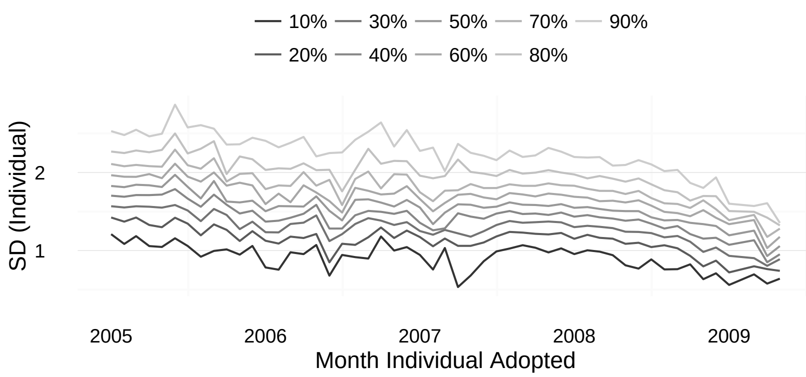
All tests: $N = 50$		**Test 1:** $T = 50$, $D = 30$			**Test 2:** $T = 5$, $D = 30$			**Test 3:** Mimic uncensored empirical distribution $D = 60$		
Mean fraction of replications links observed		1.000			0.993			0.848		
Common Parameters		True	Mean Estimate	SD_{sim}	True	Mean Estimate	SD_{sim}	True	Mean Estimate	SD_{sim}
γ		2.600	2.589	0.010	2.600	2.553	0.035	2.600	2.243	0.028
α		160.000	148.986	3.465	160.000	125.036	8.998	160.000	97.473	8.145
β_{cost}		0.180	0.181	0.001	0.180	0.181	0.004	0.180	0.179	0.005
$\beta_{cost} \cdot \beta_{coverage}$		−0.900	−0.899	0.020	−0.900	−0.899	0.073	−0.900	−0.900	0.079
Communication Graph	True	True Mean	Mean Deviation	Mean SD_{sim}	True Mean	Mean Deviation	Mean SD_{sim}	True mean	Mean Deviation	Mean SD_{sim}
μ_i	$\sim Triangular[5, 6, 7]$	6.012	−0.028	0.397	6.077	−0.182	0.408	5.979	−0.134	0.379
σ_i	$\sim Triangular[0.5, 1, 1.5]$	0.973	0.015	0.199	1.004	−0.058	0.209	1.033	−0.100	0.196
q_i	1.000	1.000	−0.022	0.032	1.000	−0.057	0.073	1.000	−0.045	0.058
Expected utilities and durations of randomly selected links		True Mean	Mean Deviation	SD_{sim}	True Mean	Mean Deviation	SD_{sim}	True Mean	Mean Deviation	SD_{sim}
$ED_{p_{high},\phi_i=0.7,\phi_j=0.7}$	seconds	22.656	0.428	10.829	25.151	−2.091	11.027	58.859	7.030	32.726
$ED_{p_{low},\phi_i=0.7,\phi_j=0.7}$		49.625	0.608	18.061	53.526	−3.071	18.599	140.163	12.558	62.585
ΔED price		26.969	0.180	7.441	28.375	−0.980	7.818	81.304	5.528	30.507
$ED_{p_{high},\phi_i=0.9,\phi_j=0.9}$		34.982	0.465	14.275	38.117	−2.628	14.610	94.622	9.596	46.283
ΔED coverage		12.327	0.036	3.696	12.966	−0.537	3.850	35.763	2.566	14.431
$EU_{p_{high},\phi_i=0.7,\phi_j=0.7}$	USD	0.055	0.001	0.031	0.062	−0.007	0.032	0.138	0.021	0.094
$EU_{p_{low},\phi_i=0.7,\phi_j=0.7}$		0.308	0.005	0.132	0.338	−0.025	0.135	0.831	0.089	0.425
ΔEU price		0.253	0.004	0.102	0.276	−0.018	0.104	0.692	0.068	0.333
$EU_{p_{high},\phi_i=0.9,\phi_j=0.9}$		0.138	0.002	0.068	0.154	−0.015	0.070	0.361	0.048	0.212
ΔEU coverage		0.084	0.001	0.037	0.092	−0.007	0.038	0.223	0.027	0.120

Results of Monte Carlo simulations; results reported for at least 100 data sets generated with the stated parameters. Communication graph parameters are held fixed across each of these data sets. N is the total number of nodes, D the degree of each node, and T the number of time periods. Results that mimic the uncensored portion of the empirical distribution have 60 links total; 30 observed for 1 period, 10 observed for 2, 10 observed for 3, and 10 observed for 5. When applied to data, links also have many more censored observations and more nodes are included in estimation. SD_{sim} represents the standard deviation of the simulated estimates. Expected utilities and durations are computed for a sample of one link per node, randomly selected, using Monte Carlo integration with at least 1,000 draws.

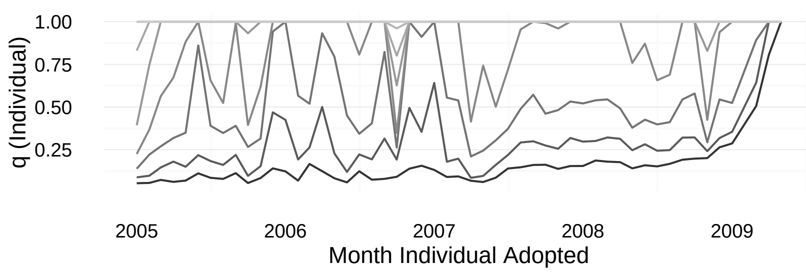
FIGURE S2. Estimated Parameters

A. Quantiles of Individual Parameters by Adoption Month

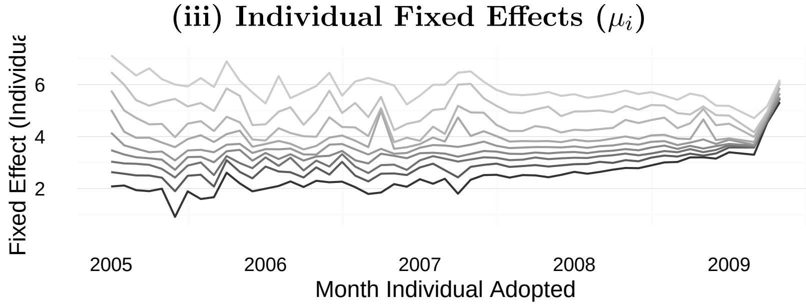
(i) Standard Deviation of Shock Distribution (σ_i)



(ii) 1-Probability of No Call at Any Price (q_i)



(iii) Individual Fixed Effects (μ_i)



B. Link Cohort × Average Coverage Fixed Effects ($\mu_{\max(x_i,x_j),\overline{\phi_{it}\phi_{jt}}}$)

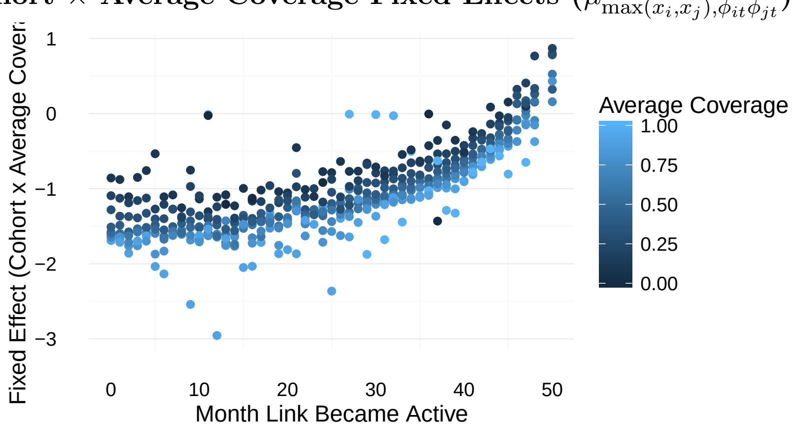


Table S5. Call Model Comparative Statics

Price	Coverage	Expected:	Quantile of Parameters (μ_{ij}, σ_i, q_i)	5%	25%	50%	75%	95%
Lowest	100%	Duration conditional on call		1.3 sec	31.0 sec	84.0 sec	4.6 min	19.0 min
		Probability of call		0.00	0.02	0.37	0.64	0.73
		Charge	$	0.00	0.00	0.07	0.38	1.77
		Hassle cost of imperfect coverage	$	0.00	0.00	0.00	0.00	0.00
		Net utility	$	0.00	0.00	0.20	1.30	6.42
Lowest	70%	Duration conditional on call		0.9 sec	28.3 sec	73.4 sec	3.9 min	15.8 min
		Probability of call		0.00	0.01	0.30	0.58	0.69
		Charge	$	0.00	0.00	0.05	0.29	1.39
		Hassle cost of imperfect coverage	$	0.00	0.00	0.08	0.49	2.31
		Net utility	$	0.00	0.00	0.10	0.74	3.74
Highest	100%	Duration conditional on call		0.6 sec	25.4 sec	62.3 sec	3.2 min	12.6 min
		Probability of call		0.00	0.01	0.22	0.50	0.63
		Charge	$	0.00	0.00	0.12	0.87	4.32
		Hassle cost of imperfect coverage	$	0.00	0.00	0.00	0.00	0.00
		Net utility	$	0.00	0.00	0.04	0.35	1.85

Comparative statics are shown for links with shock mean μ_{ij}, shock variance σ_i, and cost-independent censoring parameter q_i set to the given quantile from the data. Computed on on 1% random subsample of nodes. Hassle cost of imperfect coverage is defined in the model section of the paper.

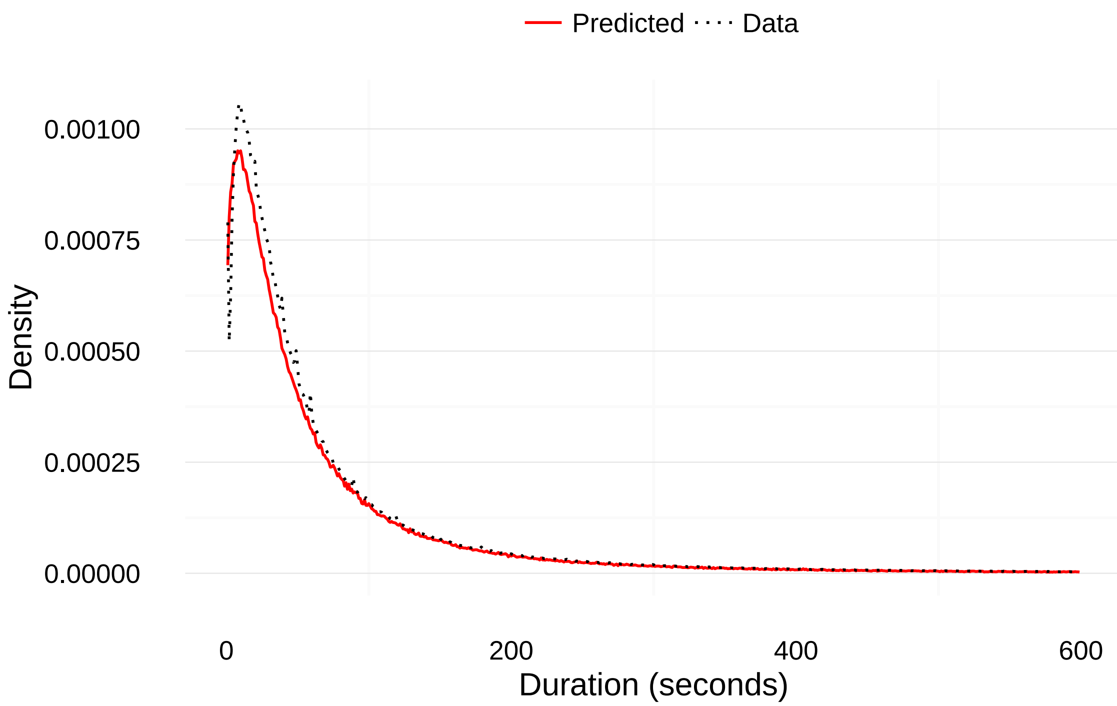
S10. Fit

As described in the paper, the call model has two goals: to uncover from observed durations and costs the underlying conditional distributions, and to translate these durations and costs into utilities. Since the data cannot directly distinguish between the shape of the utility function and the distribution of shocks, I narrow the choice of utility function using theoretical restrictions and then select a distribution of shocks that matches the data well.

I evaluate fit in two ways. In this section, I compare the durations generated by the model against those that are observed. In the paper I also compare the utility implied by the call model to the utility implied by the adoption decision.

Since the model generates a distribution of durations, I compare different margins of the data to model estimates. The ideal test would be out of the sample I used to estimate the model, but since I estimate node- and link-specific parameters, I cannot perform a completely out of sample test. Instead, I use a subset of the data that did not contribute

FIGURE S3. Call Model: Fit of Duration Distribution



Months without call:	Data	93.4%
	Predicted	93.9%

Computed on over 1m draws from a random subsample of links.

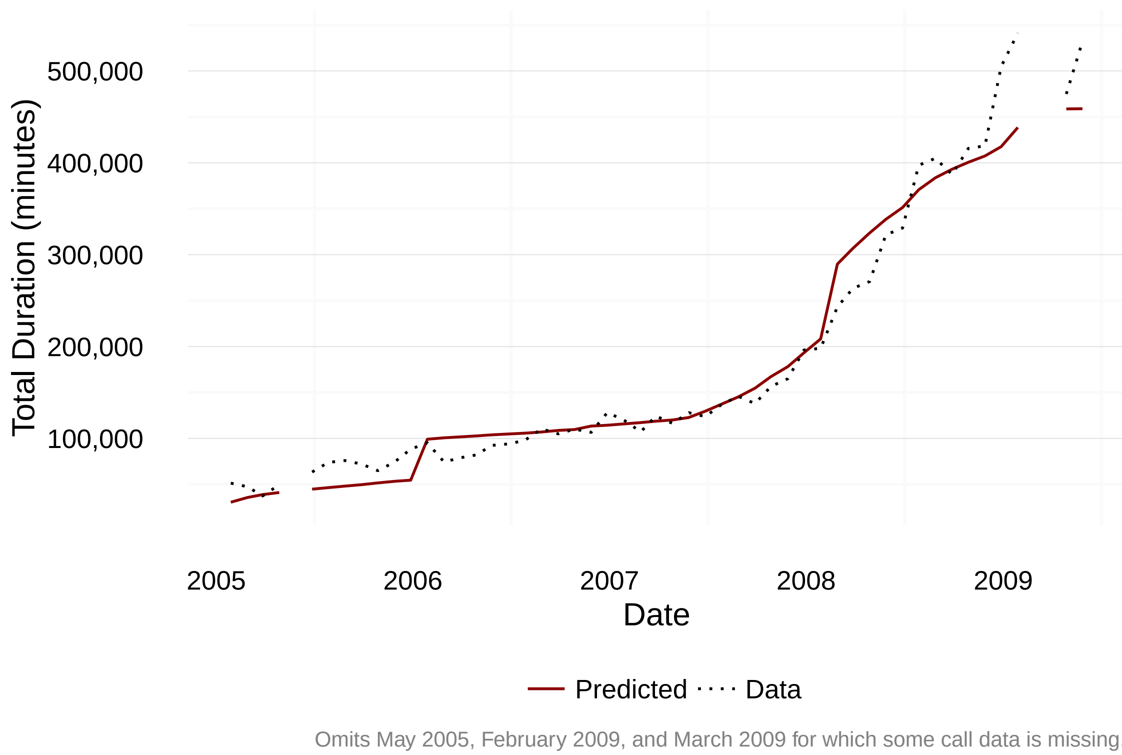
to the estimation of the common parameters, so that at least the common parameters are estimated out of sample.

Overall, the model has a very slight amount of positive bias in estimated durations: the ratio between total expected duration and total observed duration is 1.04.

There is substantial heterogeneity in the network, including nodes that are outliers in many dimensions. Because I attempt to model the full network, some dimensions of fit are sensitive to these outliers; I note these when presenting model fit.

By duration of call. The fit of the duration distribution is shown in Figure S3. The fit is helped by the large number of parameters estimated, but the choice of functional form is still important, as is evident from the predicted distribution's slight systematic deviations from the data.

Figure S4. Total Duration by Date



Randomly selected 1% subsample of nodes, omitting one outlier node.

Most of the analysis, however, does not use the entire distribution of durations but rather the average duration spoken each month across each link, which I test below with different cuts of the data:

By month. As time passes, prices decline and coverage improves, leading more individuals to adopt, and individuals to talk more over each link. In Figure S4, I compare predicted and actual call volumes by month. This margin holds fixed each individual's adoption date to isolate the contribution of the call model. The correlation between predicted and actual total duration is 0.95, though the model predicts a sharper response to the price changes in January 2006 and February 2008.

By month and link start date. Here I isolate within-link changes by comparing the model fit over each link by the date each link became active (as soon as both parties have adopted). Results are shown in Figure S5. Although levels are affected by the full set of parameters, the changes in duration arise solely from changes in coverage and calling prices, which are controlled primarily by the two associated sensitivity parameters (estimated on a separate sample). As with the previous margin, the model predicts a sharper response to price decreases than is seen in the data, but it does capture general trends. One weakness is that durations in the last two months of data, April and May 2009, tend to be underestimated, particularly for links joining in 2006.[21]

By node. I compare the average monthly predicted versus observed duration for each node, in Figure S6. The ratio of expected node duration to actual duration is 1.06, the Pearson correlation is 0.90 and the Spearman rank correlation coefficient is 0.98.[22]

Part 3. Adoption

S11. Notes on Adoption Model

Additional Fees. Before April 2005 there was a monthly fee. Before June 2007, subscribers had to top up their balance with a minimum of $4.53 every 30 days to keep their account active. If the subscriber incurred charges less than this amount, the leftover balance would accumulate. This would be a binding restriction for most subscribers, as most spend less than this amount and accrued balances could not be cashed out. I model the hassle of having to accumulate balance in this way as an extra fee of half the extra top up amount, based on the average duration talked with the contacts on the network during that month. I include these costs in estimation and simulation, in both utility and revenues.

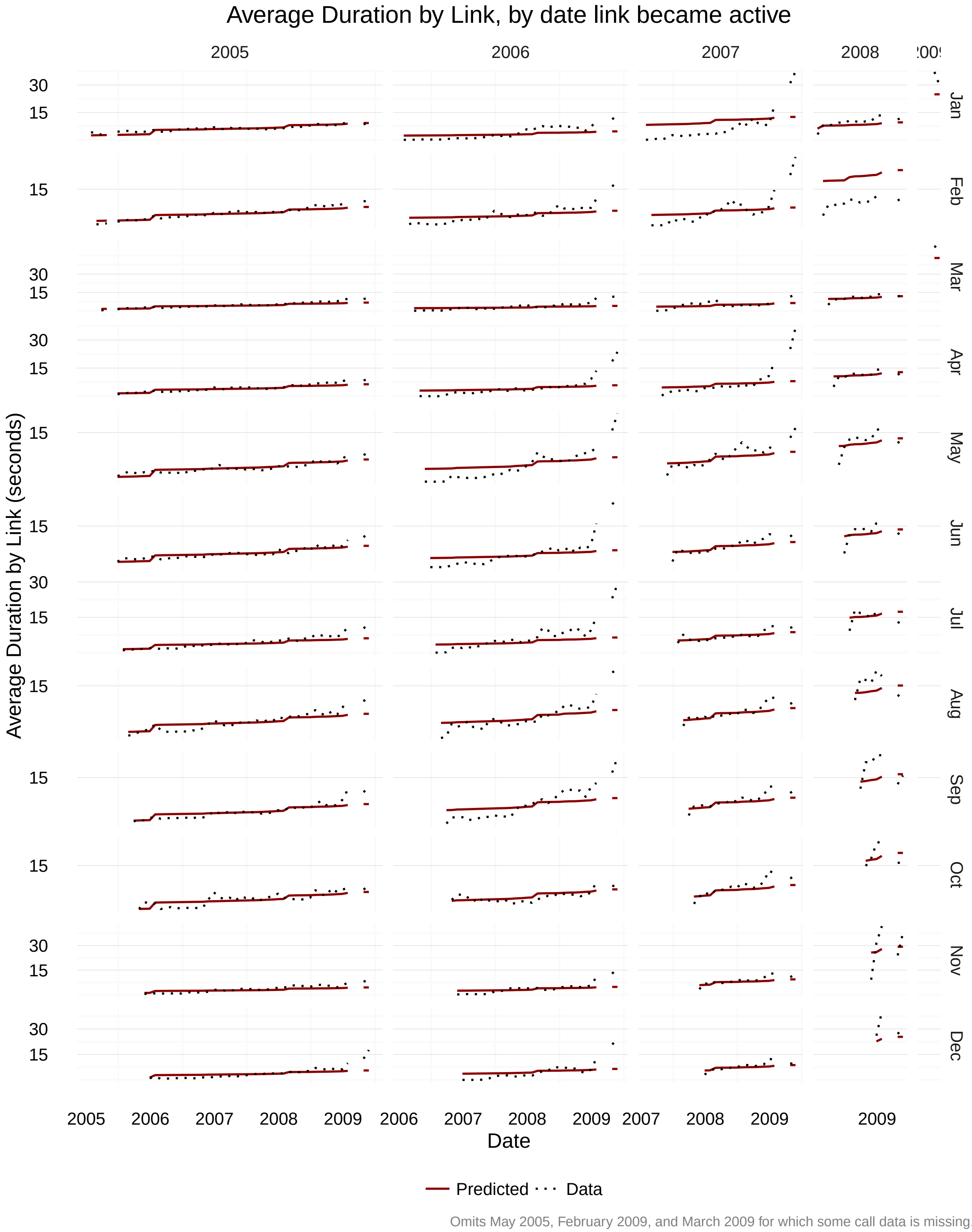
Extrapolation after End of Data. The calling data ends at $T = 53$. For $t \leq T$, I use the formulation of expected utility as described by the model. However, counterfactuals may induce an individual to delay adoption after T. Although I do not report any outcomes after T, individuals make forward looking decisions. Thus, for simulation I set $\bar{T} = 89$, three years beyond the end of the data, which corresponds to the last month I have handset price data. For $t > T$, there is not enough data to completely populate this model, so I use aggregate data on the expansion of the network. I assume the utility is a multiple of the utility from the last period, where the factor γ_t is derived from the increase in adoption from regulator statistics.

$$u_{it} = \gamma_t \cdot u_{iT} \quad \text{for all } t > T$$

[21]It is not clear what drives the increased durations observed in those months. One omission is that there appear to be many towers built during that time, but because of issues with the tower data I only consider coverage improvements up until January 2009.

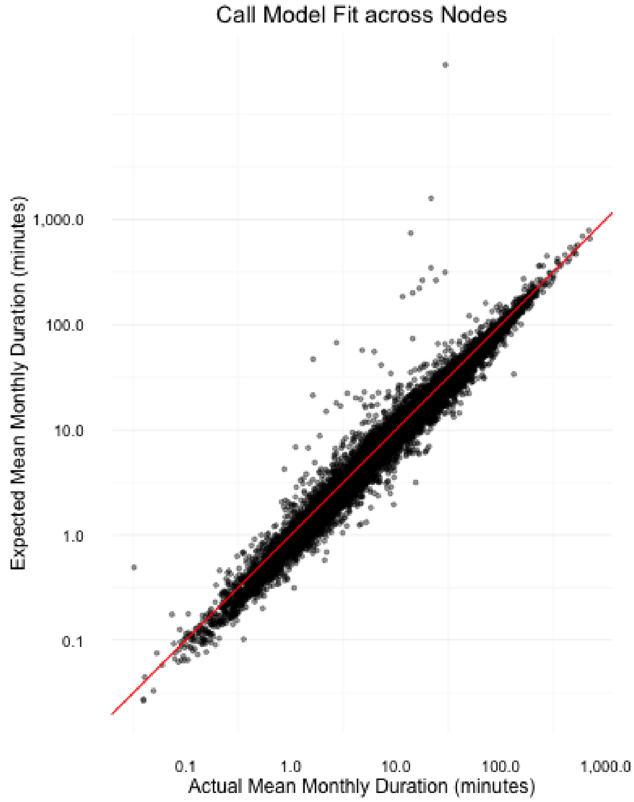
[22]When computing the Pearson correlation I omit one outlier node (which is visible in Figure S6). A high correlation is not surprising given that I estimate three parameters of the shock distribution for each node. There are a small number of outlier nodes for which the fit is poor; one of the reasons this can arise is that I was not able to obtain high enough numerical precision in the tails of the normal distribution to obtain reliable convergence when a link has a very long call.

FIGURE S5. Fit by Month and Start Date



Each cell represents average monthly usage of links that became active in the year represented by the column and the month represented by the row. Randomly selected 1% subsample of nodes, omitting one outlier node.

Figure S6. Fit by Node (log-log)



This assumption is needed because the full network of calls after T is not observed; it implies that the increase in benefits accrues in the same proportion to each node in the network. I focus on how benefits improve due to increased adoption. I fit a sigmoid function to the total number of mobile subscribers as measured by the regulator, assuming a saturation point of 70%, resulting in predicted subscribers at month t of $\left|\hat{S}_t\right| = \frac{0.7 \cdot 11,000,000}{1+e^{-0.052(t-72.8)}}$. I then compute $\gamma_t = \frac{|\hat{S}_t|}{|\hat{S}_T|}$, the proportional increase in subscribers over the number of subscribers in the last period with full data. This factor overstates the increase in benefits in two ways: at high levels of penetration the marginal benefit of an additional subscriber is likely declining, and the additional subscriptions measured by the regulator double count individuals who hold accounts with multiple operators, which becomes an issue after the third operator joins and the market becomes more competitive. It understates the increase in benefits in that it does not account for price declines associated with increasing competition nor additional services introduced later (mobile money was introduced in 2010). I assume that γ_t becomes stable in year 2025.

Simulation Method. The method is described briefly in Appendix C; I add additional notes here. To speed convergence, at each step k I use the path defined by x_j^k for individuals j that have reoptimized in this step and x_j^{k-1} for individuals who have not yet reoptimized in this step, in the same manner as the Gauss-Seidel method. The equilibrium identified may be sensitive to the order that individuals reoptimize when simulating policies with nonmonotonic effects (shifting some individuals' adoption forwards and others backwards), such as some of the taxation counterfactuals. I tested sensitivity by comparing a solved equilibrium to one solved with agents optimizing in reverse order and found small changes likely arising from rounding error (0.2% of nodes had different adoption months, averaging to an average difference in adoption month of -0.0003).

S12. Monte Carlo Exercise

In this section I evaluate the performance of the adoption model.

In my method, the expected utility i obtains from adopting at date x is given by Equation 5 in the text. After expected utilities are estimated from calling decisions, I use the adoption decision to back out bounds on the nonnetwork value of having a handset (individual types η_i). I then simulate a low and a high equilibrium, using bounds in η and a pessimistic or optimistic starting adoption path. I evaluate policy impacts by reporting how the low equilibrium and high equilibrium change in a counterfactual environment.

There are two potential concerns with this approach. First, it could be that the method performs poorly even when my model is the true model. One might be concerned that the method will fail if there is correlation the types of connected individuals (η_i). Or it could be that the change in the low and high equilibrium are not indicative of the changes that one would see at equilibria between them, including the true equilibrium that would result in a counterfactual. Second, my model may not be the true model. Particularly, my model does not allow for time-varying idiosyncratic shocks to individual utility. This would be a concern if individuals face credit constraints when purchasing a handset, for example.

To test these concerns, I simulate a data based on a model that can include idiosyncratic shocks, and report how the performance of the estimated model compares. I use the empirically defined network and the expected utilities derived from calling decisions, and assume that the true adoption model is:

$$E_t U_i^{x_i}(\boldsymbol{x}_{G_i}) = \delta^{x_i}\left[\sum_{s\geq x_i}^{\infty} \delta^{s-x_i} E u_{is}(p_s, \boldsymbol{\phi}_s, \boldsymbol{x}_{G_i}) - E_t p_{x_i}^{handset} + \eta_i\right] + \nu_{it}$$

for values of η_i and error structure ν_{it}. I allow the error structure to follow an AR(1), $\nu_{it} = \rho\nu_{it-1} + e_{it}$, and assume that individuals know the stream of errors at time zero.

I subsample 10% of the nodes in the observed network and include all of their outgoing links. For each replication, I use the utilities $Eu_{ij}(p_t, \boldsymbol{\phi}_t)$ estimated in the call model for the observed network, draw random values for each type: $\eta_i \sim N(0, \sigma_\eta^2)$ and shock: $e_{it} \sim N(0, \sigma_{perturb}^2)$, and then simulate an equilibrium.[23] In some specifications I allow an individual's draws to be correlated with those of their contacts.[24] I set the standard deviation of individual types to be similar to the empirical distribution (a standard deviation of $\sigma_\eta = \$400$ for the lifetime stream, or $\sigma_\eta \cdot (1-\delta) = \$3.36$ per month), and vary the standard deviation of perturbations $\sigma_{perturb}^2$. I simulate a new equilibrium using the procedure developed in the paper, with $w = 0$. Nodes outside the 10% sample maintain their actual adoption months; nodes within the sample start from a random candidate adoption path $\boldsymbol{x}^0$. I report results for this subsample.[25]

The tests range from no time-varying shocks (as assumed by the model), up to very large shocks (a monthly standard deviation of $\sigma_{perturb}^2 = \$50$, representing 1.4 months of per capita consumption for the average phone owning Rwandan in 2010). Because I am ultimately interested in how well the model predicts the effects of counterfactuals, I simulate counterfactuals under a true model with time-varying perturbations and compare the results to those obtained by the assumed model. I draw random types and shocks, and simulate the true equilibrium. From this simulated equilibrium I back out bounds on the types η_i that would be implied by the model without time-varying perturbations. I then simulate the results that would be obtained by this model under the lower bound equilibrium $\underline{\Gamma}$ and the upper bound equilibrium $\bar{\Gamma}$. Since my main outcomes are the impact of policies, I evaluate the model's performance in predicting changes from a counterfactual to the baseline. Since I am only allowing 10% of the network to adjust, the impact of these counterfactuals are likely to be attenuated in these simulations. Because of this attenuation, I use a counterfactual

[23]For simplicity I leave out extra fees that are incorporated into the main estimation procedure.

[24]Because it would be difficult to draw from a multivariate normal distribution that includes the full covariance structure, I approximate this distribution as follows. I draw a node at random, and then identify all of his contacts for whom I have not yet drawn a type/shock. Among this subset of nodes, I draw types/shocks from a multivariate normal distribution with the specified correlation between each pair of contacts (including connections between contacts). I then mark all of these nodes as completed, draw a new uncompleted node, and repeat.

[25]For a given amount of computational capacity, I face a tradeoff between (a) the size of the subsample, (b) the number of different parameters to test from, and (c) the number of replications to perform for each parameter set. For these Monte Carlo tests I found it most informative to use a 10% sample, using several different parameter values, and two replications.

that would induce a stronger response among this 10% subset; specifically, I consider the counterfactual of the operator not building the 30 lowest revenue rural towers.

I present simulation results in Table S6. I compare the true impact to the change in the bounds, so this tests both the relevance of these changes in considering policy impact as well as the effect of the structure of types and shocks. As the magnitude of time-varying shocks increases, these shocks explain more of adoption choices and the effect of the expansion attenuates slightly. Although the estimates I present are changes in bounds and not bounds in changes, I find nonetheless that the true coverage expansion impact lies between the largest and smallest change for up to moderate shocks ($1 and below). For large shocks (above $2), the model predicts smaller changes than would be observed.

These results are similar regardless whether individual types or shocks are correlated with their neighbors, and whether or not shocks are persistent ($\rho > 0$).

I interpret these results to suggest that the adoption model estimated without time-varying perturbations does quite well at capturing the effect of policy changes on adoption even when the underlying data generating process includes small perturbations. If the underlying data generating process includes larger shocks, at least for the coverage counterfactual tested here, my method tends to report attenuated policy impacts.

S13. Details on Robustness Check of Network Value

I use the adoption decision as a check on the estimate of sensitivity to handset prices. Here I describe the construction of the three instruments used and present evidence on validity.

Incidental coverage instrument. Towers are powered by electric lines or with generators. It is much cheaper to operate towers on the electric grid, and as a result the proximity to an electric grid is an important determinant of tower placement. However, while proximity to the grid directly affects the location of the towers themselves, given Rwanda's hilliness it is not the best measure of the resulting coverage. Instead I compute an incidental coverage map: the coverage that would result from building towers along the full network of power lines.[26] These areas of the country had a higher ex-ante probability of receiving coverage because of the interaction between their geographic features and the existing electric grid. One obvious concern is that areas closer to power lines will have higher incidental coverage, and these areas may differ for other reasons that violate the exclusion restriction (for one,

[26] I use a GIS layer of the electric grid as of 2008 provided by Rwanda's Energy, Water and Sanitation Authority. I only consider segments of the grid providing low enough voltage that they could be used to power a cellular tower.

Table S6. Monte Carlo Test of Adoption Model

Perturbation SD (monthly)	**Correlation of type with neighbor (η_i)**	**Correlation of shock with neighbor (e_{it})**	**Persistence (ρ)**	**Policy Impact** (Difference in adoption month Counterfactual-Baseline) True	Mean Estimate $\underline{\Gamma}$	$\bar{\Gamma}$
\$0.00	0	0	0	0.37	0.46	0.31
\$0.25	0	0	0	0.36	0.42	0.29
\$0.50	0	0	0	0.36	0.39	0.28
\$1.00	0	0	0	0.36	0.36	0.27
\$1.00	0	0.8	0	0.35	0.36	0.27
\$1.00	0.8	0	0	0.36	0.36	0.28
\$1.00	0	0	0.5	0.36	0.35	0.27
\$1.00	0.8	0.8	0.5	0.37	0.36	0.29
\$2.00	0	0	0	0.35	0.32	0.26
\$2.00	0	0	0.5	0.34	0.31	0.25
\$5.00	0	0	0	0.35	0.28	0.23
\$10.00	0	0	0	0.33	0.25	0.22
\$20.00	0	0	0	0.30	0.21	0.19
\$50.00	0	0	0	0.27	0.16	0.15

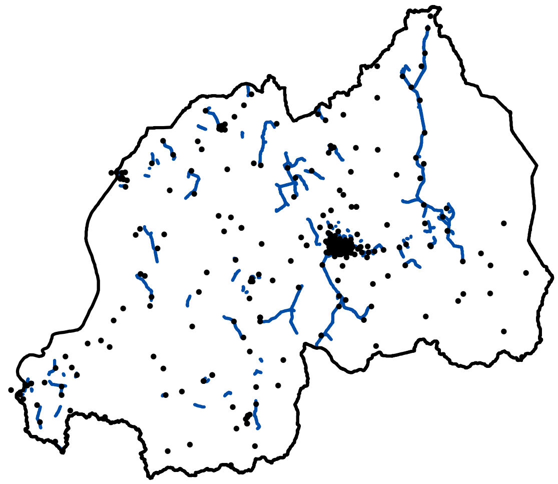
Results from Monte Carlo test of adoption model, using 10% of nodes, under two replications for each trial. The lifetime standard deviation in individual type is set to \$400. Simulation test uses, and assumes econometrician knows, true value of $\beta_{cost} = 0.2$.

households are more likely to have electricity). For this reason, I use variation in incidental coverage only for locations further than 5km from the grid. For locations within 5km of the grid, I set the instrument's value to the mean value of incidental coverage outside the buffer region. The resulting instrument picks up incidental coverage based on geographical idiosyncrasies, such as whether they are on a hillside facing towards or away from a power line, for households further than 5km from the electric grid. See Figure S7 for a visual of the construction of the instrument.

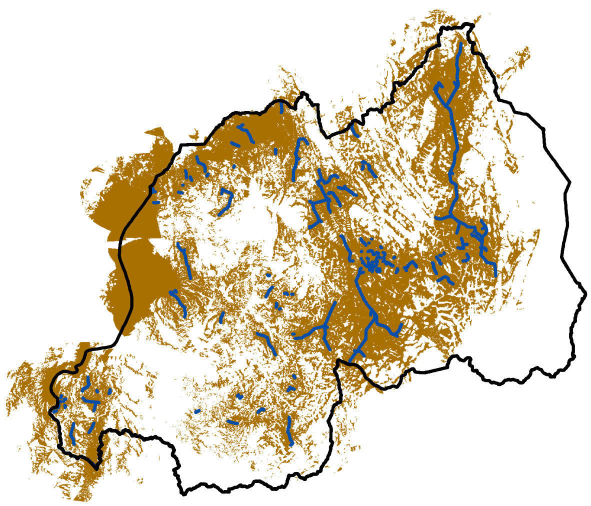
Fraction of contacts receiving subsidized handsets. The Rwandan government allocated subsidized handsets to rural areas in the first few months of 2008. I consider an account as subsidized if it was activated during the first four months of 2008 and its mode handset was the subsidized model, which was otherwise rare in the country at the time. There are 41,225 such accounts. Then, for every individual, I compute the fraction of contacts that received subsidized handsets. Imagine two individuals who have yet to subscribe, who do

Figure S7. Incidental Coverage from Electric Grid

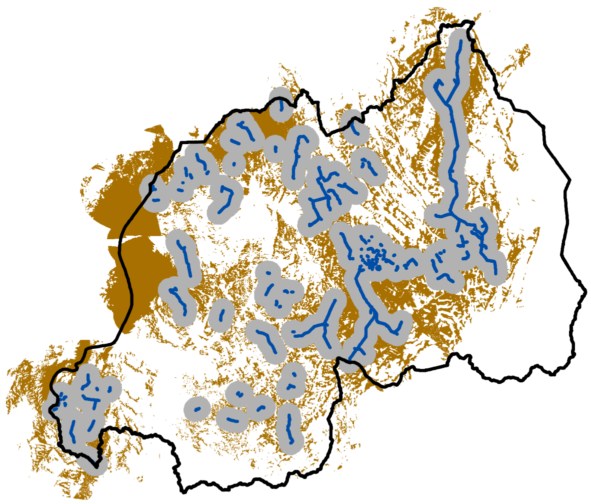
(a) Locations of electric grid and towers, January 2009



(b) Areas that would receive coverage if towers were built along full extent of electric grid



(c) Incidental coverage instrument, with 5km buffer around electric grid removed



I only consider segments of the grid with low enough voltage that they could be used to power a cellular tower.

not themselves receive a subsidy. The subsidy represents a shock that induces a fraction of their contacts to join. The one that has a higher fraction of contacts affected by the subsidy will receive a larger shock to the utility of being on the network. For individuals that had subscribed before the subsidy, the effect is ambiguous, because a higher fraction of contacts who are subsidized also implies a higher fraction of contacts who wait to join the network. For this reason I use variation in this instrument only for individuals subscribing after the beginning of the subsidy period in January 2008. Because rural residents may find it most costly to adopt a handset, I use only variation within groups of individuals who have a similar fraction of rural contacts.[27]

Tests. In order for the instruments to be valid, they must induce variation in the utility at adoption but be uncorrelated with the unobserved idiosyncratic benefit of being on the network (η_i—the exclusion restriction). Note that I observe a lot—every individual use of the phone. The idiosyncratic benefit would pick up differences in individuals' average valuations for calling, differences in the utility of owning a handset independent of the calling decision (such as SMS), or forecast errors in the utility of joining the network.

Incidental coverage is positively correlated with coverage, especially later in the data when more of the rural network has been rolled out (0.08 in 2005 and 0.49 in 2009). The fraction of contacts subsidized is positively correlated with the total number of contacts subscribing over the months of the subsidy (0.14).

In Table S7, I present correlations that measure mechanisms that I assume are excluded. Since I do not have standard characteristics for subscribers, I derive metrics from transaction data to describe channels that should be excluded. The first three columns represent correlations for the two instruments. As a comparison test, I include two more columns representing correlations with coverage at the beginning and end of the data; coverage itself may fail the exclusion restriction because the operator is more likely to build towers in locations where individuals receive more idiosyncratic benefit from the network.

First, I consider measures of network structure. Individuals with different network structure may receive different benefits of being on the network: a trader with many dispersed contacts may receive a different utility than a mother in a rural area communicating with a few, well-connected family members. I present results for the number of contacts (degree) as well as for the clustering coefficient (the fraction of a node's neighbors who are themselves

[27] For each decile of fraction of rural contacts $x \in \{0...9\}$, I construct an instrument. For subscribers within the decile, the instrument equals the fraction of contacts receiving a subsidy. For subscribers outside the decile, the instrument is set to the mean of those within the decile.

connected), both measured using the final network revealed through the end of the data, by which point coverage had expanded. I find that both measures are most correlated with coverage in January 2009 (0.11 and -0.15): individuals with higher coverage tend to have larger, more dispersed networks. The correlations with the instruments are much lower.

Last, I consider the quality of handset used, which is likely correlated with the unobserved benefit of adoption. In the model in the paper, for simplicity I do not consider differences in handset models, but for the majority of subscribers I know both the model of handset used and the price series specific to that model. To compare handset quality, I measure the price of each subscriber's chosen handset model as of the same date, January 2009. As shown in the last two columns, there is a correlation between coverage and this measure of handset quality (0.12 in 2005 and 0.10 in 2009): individuals who have higher coverage also have higher quality handsets. The correlation between this measure and the instruments is smaller: it is quite small for incidental coverage (0.02); it is larger for the fraction of contacts subsidized (-0.07): individuals who have many contacts receiving subsidized handsets tend to have slightly lower quality handsets.

Part 4. **Robustness of Simulation Results**

S14. If Incoming Calls Are Valued

The robustness check in Appendix B compares the value implied by calls to that implied by adoption, which suggests that setting $w = 1$ would overcount utility. As a robustness check I also estimate and simulate the model under the assumption that incoming calls are valued the same as outgoing calls ($w = 1$).

Results are shown in Table S8. The first panel presents parameter estimates. If $w = 1$, then the individual types needed to rationalize adoption are negative for over 50% of users: the median user receives [$0.71, $1.64] less utility per month than implied by the call model.

The second panel of Table S8 presents simulation results under the assumption that $w = 1$. Consumer surplus is roughly twice as large as the preferred specification. The effects of counterfactuals that change the cost of placing calls across links (e.g., coverage, usage tax) tend to be amplified, as an increase in usage will make the network more valuable for both senders and receivers. The effect of counterfactuals that change the cost of adoption (handset tax) tend to be muted: the model roughly doubles the contribution of usage utility, so that handset prices are given much less role in influencing adoption.

TABLE S7. Correlations with Excluded Mechanisms by Instrument

Correlation	Instruments: Incidental Coverage	Instruments: Fraction contacts subsidized	Comparison: Coverage January 2005	Comparison: Coverage January 2009
Number of contacts (Degree)	-0.02	-0.01	0.03	0.11
Clustering coefficient	0.01	0.02	-0.11	-0.15
Price of handset model purchased, as of January 2009	0.02	-0.07	0.12	0.10
N^*	280,533	452,211	1,503,369	1,503,369
Sample	Primary location ≥ 5 km from electric grid	Subscribing after January 2008	All	All

All correlations have a p-value of 0.00. *: I can match the specific handset model a node is affiliated with to a price for 960,854 nodes. Correlations with handset price are computed on this subset.

Stepping through the table, rural coverage expansion again is again unprofitable, but improves welfare more. The handset tax has a smaller effect, with an average welfare cost per dollar raised of \$0.91 or \$1.14 (vs. \$2.95 or \$3.11 in the preferred specification). However, it again is larger than a naïve estimate that neglects network effects, of \$0.71 or \$0.77 (vs. \$1.22 or \$1.06 in the preferred specification). Under full passthrough, the usage tax has larger effects, with an extremely high average welfare cost per dollar raised of \$3.61 or \$4.57 (vs. \$2.67 or \$2.47 in the preferred specification). This is again larger than the naïve estimate of \$3.02 or \$2.76 (vs. \$2.06 or \$2.00 in the preferred specification). The effects under no passthrough are very similar.

S15. MYOPIC MODEL: IF INDIVIDUALS DO NOT CONSIDER THE FUTURE

The main model assumes that individuals forecast future changes in prices, coverage, and their contacts on the network. They discount future utility it at a rate δ derived from the

Table S8. If Incoming Calls are Valued (w=1): Estimation and Simulation Results

Implied η_i's

	Quantile:		**5%**	**25%**	**50%**	**75%**	**95%**	**Number**
$w = 0$ (preferred)	$\bar{\eta}_i \cdot (1-\delta)$	\$	-4.18	0.01	1.07	1.56	4.22	0.8m
	$\underline{\eta}_i \cdot (1-\delta)$	\$	-5.88	-0.69	0.71	1.39	3.77	0.8m
$w = 1$	$\bar{\eta}_i \cdot (1-\delta)$	\$	-12.33	-3.63	-0.71	0.92	3.81	0.8m
	$\underline{\eta}_i \cdot (1-\delta)$	\$	-15.84	-5.37	-1.64	0.50	3.22	0.8m

Simulation Results		Model	Revenue ($m)		Consumer	Avg. Welfare Cost per Dollar of
			Telecom	Government	Surplus ($m)	Public Funds ($)
Baseline (Tax: 23% Usage, 48% Handset)		$w = 0$	[165.06, 187.39]	[65.29, 73.08]	[243.55, 269.79]	-
		$w = 1$	[151.66, 190.68]	[60.49, 73.11]	[474.90, 579.01]	-
Coverage Expansion: Effect of Removal		$w = 0$	-0.09, -0.11	-0.03, -0.03	-0.36, -0.37	-
		$w = 1$	-0.11, -0.09	-0.04, -0.03	-0.80, -0.72	-
Handset Tax: Effect of Removal						
Complete Passthrough	Total Effect	$w = 0$	14.66, 17.49	-12.07, -12.34	20.96, 20.90	2.95, 3.11
		$w = 1$	3.35, 5.00	-14.65, -15.12	9.96, 12.28	0.91, 1.14
	Proximal Only	$w = 0$	7.06, 6.89	-14.34, -15.51	10.48, 9.55	1.22, 1.06
		$w = 1$	2.64, 3.22	-14.86, -15.65	7.98, 8.77	0.71, 0.77
Usage Tax: Effect of Removal						
Complete Passthrough	Total Effect	$w = 0$	57.81, 62.72	-48.43, -55.16	71.62, 73.30	2.67, 2.47
		$w = 1$	46.11, 77.18	-44.42, -55.68	114.09, 177.22	3.61, 4.57
	Proximal Only	$w = 0$	45.43, 50.67	-48.72, -55.41	55.12, 60.16	2.06, 2.00
		$w = 1$	39.58, 45.00	-44.64, -56.21	95.13, 110.23	3.02, 2.76
No Passthrough	Total=Proximal	$w = 0$	49.30, 55.97	-49.30, -55.97	0.00, 0.00	1.00, 1.00
		$w = 1$	45.30, 56.96	-45.30, -56.96	0.00, 0.00	1.00, 1.00

real interest rate in Rwanda. To gauge the importance of future looking behavior, I consider the results if instead consumers use a myopic model, where they consider only the current level of utility provided by the network.

I consider a model where i adopts if the current utility provided by the network (with the current level of prices, coverage, and the current contacts who are on the network) exceeds some cutoff. i adopts at the first period t where the expected utility flow minus the cost of the handset exceed the individual's cutoff type ($\tilde{\eta}_i$):

$$\frac{Eu_{it}(p_t, \boldsymbol{\phi}_t, \boldsymbol{x}_{G_i})}{1-\delta} - p_t^{handset} \geq \tilde{\eta}_i \tag{S1}$$

The discount factor δ enters this expression, but serves just to relate the flow utility to the fixed cost of a handset. (In the full model it also controls how much weight is placed on future flow utilities.) Given this model and the call utility estimates, individual $i's$ cutoff type can be backed out from the equivalent of Equation 10 in the paper. That i adopted at x_i implies the following bounds:

$$\begin{aligned} \bar{\tilde{\eta}}_i &= \frac{Eu_{ix_i}(p_{x_i}, \boldsymbol{\phi}_{x_i}, \boldsymbol{x}_{G_i})}{1-\delta} - p_{x_i}^{handset} \\ \underline{\tilde{\eta}}_i &= \frac{Eu_{ix_i-1}(p_{x_i-1}, \boldsymbol{\phi}_{x_i-1}, \boldsymbol{x}_{G_i})}{1-\delta} - p_{x_i-1}^{handset} \end{aligned}$$

I back out these cutoff types for each individual, and then simulate adoption according to Equation S1.

This model implies that consumers optimize a utility function different from the full model. To make results comparable to the full model, I take the equilibrium behavior resulting from the myopic model (adoption times $\boldsymbol{x}$) and compute the utility for that behavior under the full model. Results are presented in Table S9. The first panel shows the estimates of $\tilde{\eta}$ compared to estimates of η from the full model. To make results comparable, I report the monthly flow utility whose discounted stream would add up to the cutoff: $\tilde{\eta} \cdot (1-\delta)$. The estimates suggest the median consumer adopts when the current utility provided by the network, received in perpetuity, exceeds the price of a handset by between \$-0.46 and \$-0.35. If consumers do consider the future in their decision, $\tilde{\eta}$ will factor in both the fact that the utility of using the network will improve in the future (tends to decrease $\tilde{\eta}$ relative to η) and the option value of future handset price declines (tends to increase $\tilde{\eta}$).

The second panel of Table S9 compares simulation results from the myopic model to the full model presented in the paper. Results are broadly similar. Counterfactuals that affect the cost of communication across links (coverage and usage tax) yield very close results; expectations matter less for these since these costs are not sunk. Counterfactuals that affect the cost of adoption (handset tax) differ more: adoption costs are sunk, so expectations of future changes have a larger effect on the adoption decision.

Stepping through the table, I find that building the last 10 towers again improves welfare, by a slightly larger amount, and is slightly less profitable. The handset tax imposes less of a welfare cost (\$1.44 or \$1.52 per dollar of funds raised, vs. \$2.95 or \$3.11 in the full model), but again this is far larger than a naïve estimate based on the proximal effect (\$0.63 or \$0.62 per dollar raised, vs. \$1.22 or \$1.06 in the full model). Results for the usage tax are very similar under the myopic and full models, for both complete and no passthrough.

S16. Incomplete Network, and Robustness to End Point of Data

As described in Section 5 of the paper, I do not necessarily observe $\bar{G}$, the full communication graph of Rwanda. I observe individuals who subscribed by the end of my data ($S_T \subseteq N$) and infer that i and j are linked if i has called j by period T. That is, I observe the subgraph $G^T \subseteq \bar{G}^T \subseteq \bar{G}$. This may have several effects.

Missing links. Missing links could potentially affects my estimates and all counterfactuals. Based on a representative household survey, 90% of subscribers report that most calls are to family and friends, and 94% report that the main purpose of the last 10 calls was social (Stork and Stork, 2008). However, because I infer links from calls, there may be subscribers who i would potentially call, but who he did not call by the time my data ends ($j \in \bar{G}^T \backslash G^T$). The omission of these latent links does not introduce a clear bias in overall call utility, due to the law of iterated expectations: while I observe only links $ij \in G^T$ that receive a call before T, their shock distributions ϵ_{ijt} are estimated conditional on receiving a call before T. Both forces are counterbalanced in computing the utilities u_{it}; their net effect will depend on the curvature of the functions.[28] In simulations this will also tend to make late adopting individuals a bit too sensitive to the adoption of their realized contacts, however I expect this effect to be small.

Relatedly, one of the benefits of owning a phone is the option value of being able to place calls, which may be valued even if the option is not realized. An extreme example would be a phone purchased solely for emergency use, which provides expected utility even though it may never be used. Since the utility computed in this model relies on realized calls, any option value would necessarily be underweighted in the call model, though it would be captured in an individual's type η_i in the adoption decision.[29] This omission is likely to be

[28]One exception is if the counterfactual makes calling across each link ij more attractive. Then the utilities will be underestimated, because this would undo the counterbalance: some links that I do not observe would become more attractive. The only counterfactuals that I run that do this are the usage tax removal under complete passthrough; this would suggest my results there are likely to underestimate the welfare cost of usage taxes.

[29]It would be possible to include utility from nodes that are on the network but for which no calls have been realized, but this would require a careful decision about which nodes provide option value and which do not.

TABLE S9. Myopic Model: Estimation and Simulation Results

Adoption Model Estimates	**Quantile:**		**5%**	**25%**	**50%**	**75%**	**95%**	**Number**
Full Model (Original Estimates)	$\bar{\eta}_i \cdot (1-\delta)$	$	-4.18	0.01	1.07	1.56	4.22	0.8m
	$\underline{\eta}_i \cdot (1-\delta)$	$	-5.88	-0.69	0.71	1.39	3.77	0.8m
Myopic Model	$\bar{\bar{\eta}}_i \cdot (1-\delta)$	$	-6.17	-1.53	-0.35	0.16	0.94	0.8m
	$\tilde{\underline{\eta}}_i \cdot (1-\delta)$	$	-6.81	-1.76	-0.46	0.11	0.91	0.8m

Simulation Results		Model	Revenue ($m) Telecom	Revenue ($m) Government	Consumer Surplus ($m)	Avg. Welfare Cost per Dollar of Public Funds ($)
Baseline (23% Usage Tax, 48% Handset Tax)		Full	[165.06, 187.39]	[65.29, 73.08]	[243.55, 269.79]	-
		Myopic	[167.82, 179.00]	[65.80, 69.65]	[249.75, 263.69]	-
Rural Coverage Expansion: Effect of Removal		Full	-0.09, -0.11	-0.03, -0.03	-0.36, -0.37	-
		Myopic	-0.16, -0.15	-0.06, -0.05	-0.44, -0.45	-
Handset Tax: Effect of Removal	**Passthrough**					
Complete Passthrough	Total Effect	Full	14.66, 17.49	-12.07, -12.34	20.96, 20.90	2.95, 3.11
		Myopic	6.89, 7.59	-14.24, -14.55	13.69, 14.45	1.44, 1.52
	Proximal Only	Full	7.06, 6.89	-14.34, -15.51	10.48, 9.55	1.22, 1.06
		Myopic	2.27, 2.09	-15.62, -16.19	7.61, 7.90	0.63, 0.62
Usage Tax: Effect of Removal						
Complete Passthrough	Total Effect	Full	57.81, 62.72	-48.43, -55.16	71.62, 73.30	2.67, 2.47
		Myopic	59.92, 68.30	-48.80, -52.20	70.91, 78.75	2.68, 2.82
	Proximal Only	Full	45.43, 50.67	-48.72, -55.41	55.12, 60.16	2.06, 2.00
		Myopic	47.80, 49.24	-49.26, -52.62	56.55, 57.46	2.12, 2.03
No Passthrough	Total Effect=Proximal Only	Full	49.30, 55.97	-49.30, -55.97	0.00, 0.00	1.00, 1.00
		Myopic	50.13, 53.47	-50.13, -53.47	0.00, 0.00	1.00, 1.00

less problematic than it might be in other settings: Rwanda has little in the way of formal emergency response; emergency calls are likely to be directed to close contacts.

Missing nodes. For counterfactuals that make adoption less attractive (such as the coverage counterfactuals), nodes outside the subgraph S^T would choose to adopt weakly later than they did in reality, so that the estimates I derive are unaffected by the omitted nodes $N \backslash S_T$.

However, counterfactuals that make adoption more attractive may be affected by missing nodes. The taxation counterfactuals make adopting more attractive, and thus could potentially induce a node i that I don't observe in my data (because $x_i > T$) to move its adoption forward (so that in some counterfactual, $x_i^* \leq T$). That is, the effects of the counterfactual would spill over across the boundary of the network I observe. This would be akin to a counterfactual of speeding up a film using fast forward: if the original film is 5 minutes long, you may run out of tape at 3 minutes: the film could not tell you what happens after those 3 minutes.

However, for these counterfactuals there is a time period $\bar{t} < T$ for which the counterfactual variation is a subset of the variation I observe in the data, and the network is no more desirable to any individual i than it was during the period I observe. To assess potential bias from observing a subset of the full network, I estimate and simulate the model using all of the data, through T, and then report results only up to the period $\bar{t}$. This would be akin to pausing the fast-forwarded film before it runs out of tape (though the counterfactual has more nuanced effects).

In the data, individuals are exposed to a sustained drop in real handset prices over 53 months. In the taxation counterfactuals, I evaluate the effect of eliminating handset taxes, which has the effect of tracing through the same variation in handset prices in 27 months.

Robustness Exercise. As a robustness exercise, I take the simulation results from the paper and present how they would differ for subsets and subperiods of the data ($t \leq \bar{t}$ for different values of $\bar{t} \leq T$). I structure the exercise to isolate the effect of considering a subset of nodes, and of cutting off the horizon of the problem.

For each potential end period $\bar{t}$ and outcome Y_{it}, I consider three objects:

The total outcome through $\bar{t}$, for all nodes:[30]

[30]For outcomes that are themselves a function of the end period, I set the end period to the last period computed (e.g., for net utility, I assume that consumers sell back handsets for the prevailing price at $\bar{t}$ for Equations S2 and S3).

(S2) $$\sum_{t=0}^{\bar{t}} \sum_{i \in S_T} Y_{it}$$

The total outcome through $\bar{t}$, for nodes who had adopted by $\bar{t}$ in the data:

(S3) $$\sum_{t=0}^{\bar{t}} \sum_{i \in S_{\bar{t}}} Y_{it}$$

The total outcome through the end of the data (T), for nodes who had adopted by $\bar{t}$ in the data:

(S4) $$\sum_{t=0}^{T} \sum_{i \in S_{\bar{t}}} Y_{it}$$

The difference between Equations S2 and S3 captures the effect of observing an incomplete graph: the difference between the results we would have gotten had we only observed the network through $\bar{t}$, and those we get observing the network through $T \geq \bar{t}$.

However, this is a dynamic problem: consumers, firms, and governments invest in adoption to obtain future benefits. Restricting the period of observation to $\bar{t}$ not only affects how complete the graph is, but also how much future benefits are taken into account. Taxing handsets may appear more desirable if you ignore future benefits. The difference between Equations S3 and S4 reveals these dynamic effects. Both consider the same subset of nodes, but Equation S3 evaluates their outcome only through $\bar{t}$ and Equation S4 evaluates them through T.

Results for utility, revenue, government revenue, and the implied marginal cost of funds are shown in Figures S8 and S9. The left columns present the full effect of removing handset taxes (with both proximal and ripple effects), and the right columns present the naïve counterfactual only considering proximal effects (not allowing effects to ripple through the network). The estimate from the paper is shown by the dot at the final month T, and month 27 is denoted by the red line. To the left of the red line, the counterfactual remains within the variation in the data; to the right, the handset price is below that observed in the data, and thus there may be nodes outside of the data that would have also adopted. The spell of adoptions following month 30 is greatly affected by the taxes, so an estimate that considered only through month 27 would likely underestimate the effects of the tax.

I make three observations:

Restricting the sample to a shorter period omits future usage and thus tends to undercount the cost of handset taxes (in the graph, this is seen as the blue line [Equation S2] in almost all cases lies above the others [particularly the dashed green line of Equation S3], for outcomes as well as the marginal cost of funds). For outcomes through the end of the data, the lowest estimated marginal cost of funds across subsamples of nodes that had adopted by $\bar{t}$ is 2.30 in the low equilibrium and 1.76 in the high equilibrium.

Relative to the dynamic effects, the effect of considering the subgraph that had adopted by $\bar{t}$ is small (the solid red line [Equation S2] vs. the dashed green line [Equation S3]). The omission of these nodes tends to lower the effect of the tax: the marginal cost of funds tends to be lower in just the subgraph (except for a spike in the high equilibrium around months 20-35).

In all subsets, the full equilibrium cost of taxation is higher than the naïve estimate, particularly after the spell of adoption following month 30 (in the bottom graphs for marginal cost of funds, the lines in the left graph are above those in the right).

I interpret these results to suggest that my primary estimates are likely to provide a lower bound of the welfare cost of taxing the growing telecom network, because these taxes affect both future periods that are not in my data ($t > T$) as well as nodes that are not in my data. That the policy conclusion is stable when considering only earlier adopters also suggests that my results are not very sensitive to missing links, since later adopters are likely to have more missing links (to nodes they would like to call, but haven't called by T).

FIGURE S8. Effect of Removing Handset Tax: Robustness to End Date (low equilibrium)

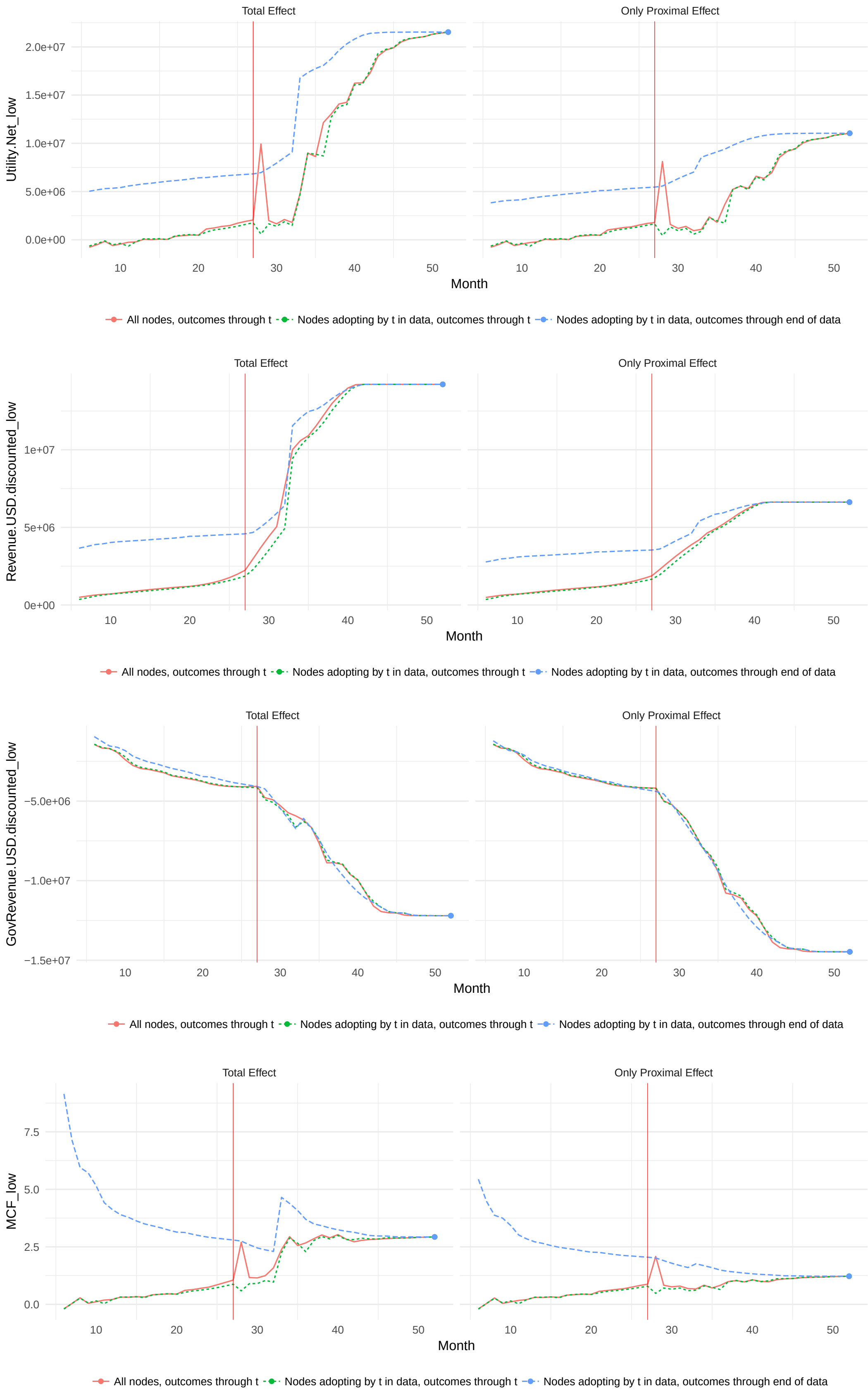
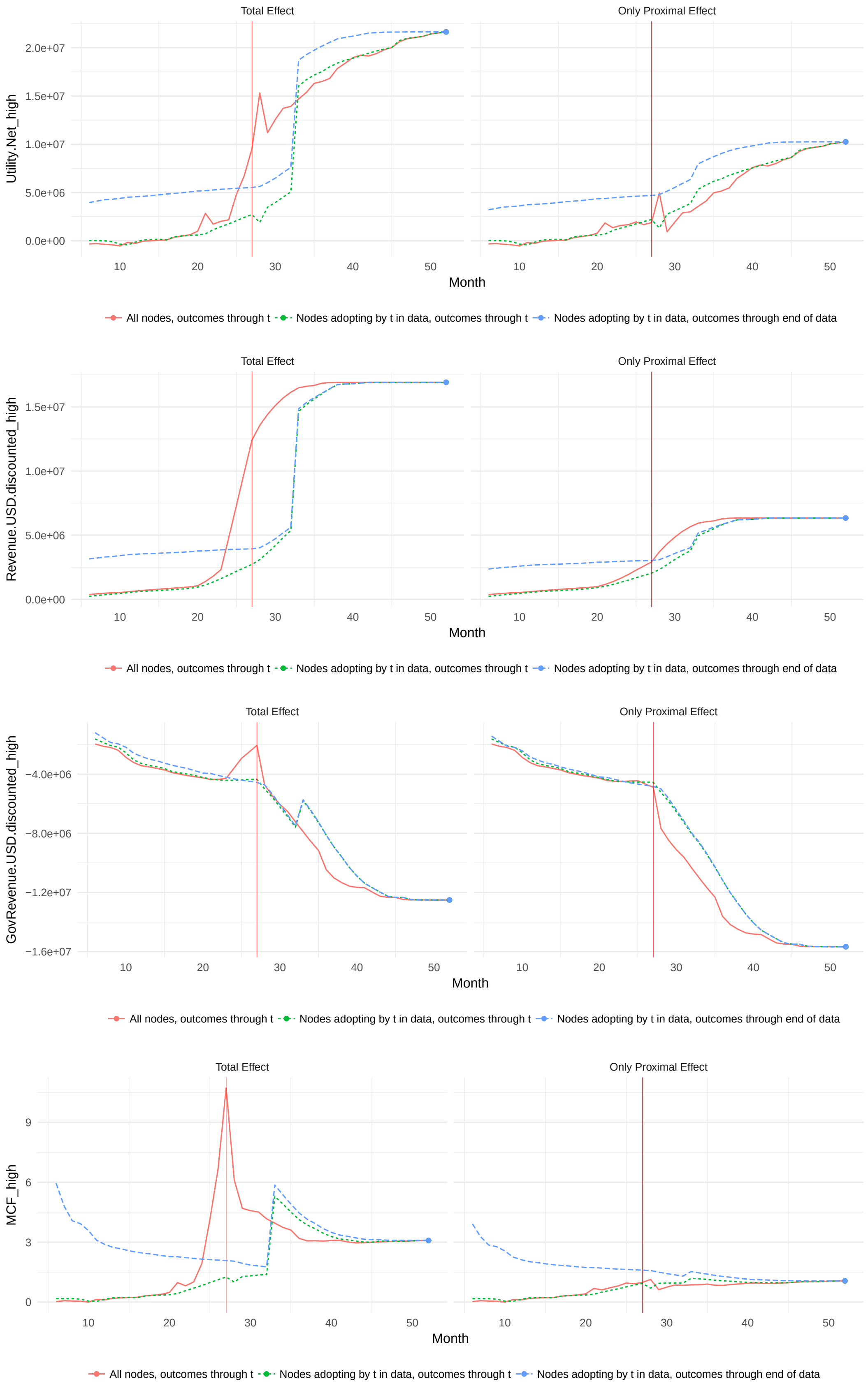


FIGURE S9. Effect of Removing Handset Tax: Robustness to End Date (high equilibrium)



S17. Supply Side Responses

The paper does not have a full model of firm optimization, and instead considers policy deviations. Here I evaluate the potential impact of (a) firm costs other than operating towers, and (b) supply side responses to policy.

A more explicit model of firm profits is given by:

$$\pi(\mathbf{p}_t|\mathbf{p}_t^{handset}, \tau_{it}^{usage}, \mathbf{z}) = R_F^{\Gamma}(\mathbf{p}_t|\mathbf{p}_t^{handset}, \tau_{it}^{usage}, \mathbf{z}) - C_F(\mathbf{z}) - B(\mathbf{p}_t|\mathbf{p}_t^{handset}, \tau_{it}^{usage}, \mathbf{z})$$

given calling prices $\mathbf{p}_t$, a path of handset prices $\mathbf{p}_t^{handset}$ (inclusive of handset taxes), usage taxes τ_{it}^{usage}, and a tower rollout plan $\mathbf{z}$. R_F^{Γ} represents firm revenues, and C_F represents the cost of a particular tower rollout plan (defined in Section 8). However, there may be additional costs of operating the network, such as staff, central operations, and equipment in addition to towers, which I do not observe, given by the function B. If B is constant across my counterfactuals, then it will cancel out when I consider a change in profits between two policies. However, a counterfactual that reduces the usage of the network (such as reducing the expansion of rural coverage) could potentially lower B, or a counterfactual that increases usage (such as reducing taxes) may increase it.

I can gauge the shape of B by considering actions that the firm chose not to make in the baseline environment. Assume the firm takes handset prices, taxes, and the coverage obligation (and the rollout plan $\mathbf{z}$) as given. Then the firm solves:

$$\max_{\mathbf{p}_t} \pi(\mathbf{p}_t|\mathbf{p}_t^{handset}, \tau_{it}^{usage}, \mathbf{z})$$

If in the baseline environment, the firm optimized prices, then if I find an alternative price sequence $\mathbf{p}_t'$ that returns higher revenues:

$$R_F^{\Gamma}(\mathbf{p}_t'|\mathbf{p}_t^{handset}, \tau_{it}^{usage}, \mathbf{z}) > R_F^{\Gamma}(\mathbf{p}_t|\mathbf{p}_t^{handset}, \tau_{it}^{usage}, \mathbf{z})$$

the firm would have implemented the alternate sequence unless this potential revenue gain were offset by at least as much cost:

$$\begin{aligned} &B(\mathbf{p}_t'|\mathbf{p}_t^{handset}, \tau_{it}^{usage}, \mathbf{z}) - B(\mathbf{p}_t|\mathbf{p}_t^{handset}, \tau_{it}^{usage}, \mathbf{z}) \\ &> R_F^{\Gamma}(\mathbf{p}_t'|\mathbf{p}_t^{handset}, \tau_{it}^{usage}, \mathbf{z}) - R_F^{\Gamma}(\mathbf{p}_t|\mathbf{p}_t^{handset}, \tau_{it}^{usage}, \mathbf{z}) \end{aligned}$$

Thus I can use unexploited revenue generating opportunities to bound costs.

Bounding Costs. First, I use this idea to bound the effect of omitting costs. In the baseline environment, I vary the calling price to be different multiples of the observed price, simulate the resulting equilibrium, and observe the change in revenue. Results are reported in the first revenue column of Table S10. A small increase in price decreases revenue, but decreases in price raise revenue. If the firm is profit maximizing, these potential revenue increases must have been offset by at least as large costs, from servicing the larger amounts of usage.

I can use this to evaluate the potential omission of costs from the tax counterfactuals. The counterfactual $0.81\mathbf{p}_t$ corresponds with the prices that consumers would face if the government eliminated usage taxes and the firm passed these savings completely through to consumers. Since the firm chose not to lower prices to this level on its own, I infer that doing so would have incurred costs of at least \$6.55m (lower equilibrium) or \$5.20m (upper). If we take these as point estimates, the corrected welfare cost per dollar of public funds moves very slightly—to \$2.53 (\$2.37) rather than \$2.67 (\$2.47) as reported in Table 4. However, these are lower bounds on costs. Under the extreme assumption that all potential revenue generated by eliminating usage taxes would be absorbed by higher operation costs, the welfare cost drops to \$1.48 (\$1.33)—which is still higher than alternative instruments cited by Auriol and Warlters (2012). Since the firm is not in control of the handset price, for the potential cost impact of a handset tax I can only repeat the last part of this exercise. If all of the revenue the firm would gain from eliminating the handset tax were absorbed in additional operation costs, the welfare cost per dollar raised would fall from \$2.94 (\$3.11) to \$1.74 (\$1.69)—still high.

On the other hand, the coverage counterfactual reduced usage, and thus could have lowered costs beyond the cost of operating towers. As described in the paper such a change would suggest that I underestimate the cost of the coverage obligation to the firm. However, that counterfactual suggests small effects and thus limited scope for such changes.

Supply Side Responses. Second, I use this technique to consider potential supply side responses. I compare the revenue change of moving the calling price in the baseline environment to the revenue change in a counterfactual environment (either no coverage obligation, or no handset tax), across the three Revenue columns in Table S10. If the revenue change associated with setting a different usage price depends on the policy environment, it is likely that changing the policy environment would induce a supply side response. The results suggest that the revenue impact of changing the usage price does not change much based on

TABLE S10. Gauging Potential Supply Side Responses

	Calling Prices	Revenue Baseline	Revenue No Coverage Obligation	Revenue No Handset Taxes	Level is equivalent to
Baseline	$\mathbf{p}_t$	[165.06, 187.39]	[164.96, 187.29]	[179.71, 204.89]	
Impact	$1.02\mathbf{p}_t$	-0.21, -2.07	-0.20, -2.06	0.17, -1.68	
	$\mathbf{p}_t$	0.00, 0.00	0.00, 0.00	0.00, 0.00	
	$0.81\mathbf{p}_t$	6.55, 5.20	6.57, 5.23	6.56, 7.36	(removing usage tax)

the policy environment. I interpret these results to suggest that these policies are unlikely to induce the firm to significantly change the usage price.[31]

References

REFERENCES

AURIOL, E. AND M. WARLTERS (2012): "The marginal cost of public funds and tax reform in Africa," *Journal of Development Economics*, 97, 58–72.

BJORKEGREN, D. (2014): "Inferring Missing Cell Tower Locations using Call Handoffs," *White Paper*.

BLUMENSTOCK, J. (2012): "Inferring patterns of internal migration from mobile phone call records: evidence from Rwanda," *Information Technology for Development*, 18, 107–125.

BLUMENSTOCK, J., N. EAGLE, AND M. FAFCHAMPS (2011): "Risk Sharing Over the Mobile Phone Network: Evidence from Rwanda," *Working Paper*.

GONZALEZ, M. C., C. A. HIDALGO, AND A.-L. BARABASI (2008): "Understanding individual human mobility patterns," *Nature*, 453, 779–782.

ISAACMAN, S., R. BECKER, R. CACERES, S. KOBOUROV, M. MARTONOSI, J. ROWLAND, AND A. VARSHAVSKY (2011): "Identifying Important Places in People's Lives from Cellular Network Data," in *Pervasive Computing*, ed. by K. Lyons, J. Hightower, and E. Huang, Springer, vol. 6696 of *Lecture Notes in Computer Science*, 133–151.

ISAACMAN, S., R. BECKER, R. CACERES, S. KOBOUROV, J. ROWLAND, AND A. VARSHAVSKY (2010): "A tale of two cities," in *Proceedings of the Eleventh Workshop on Mobile Computing Systems*, New York, NY, USA: ACM, HotMobile '10, 19–24.

[31]This is valid under the assumption that the cost associated with servicing a different usage price would not change substantially between these environments.

Jarvis, A., H. Reuter, A. Nelson, and E. Guevara (2008): "Hole-filled seamless SRTM data V4," *CIAT*.

Mulliner, C. (2013): "Public Research TAC Database," *Website*.

Stork, C. and M. Stork (2008): "ICT Household Survey Methodology & Fieldwork," Tech. Rep. 1, Research ICT Africa.