NBER WORKING PAPER SERIES

HETEROGENEITY AND THE DYNAMICS OF TECHNOLOGY ADOPTION

Stephen P. Ryan
Catherine Tucker

Working Paper 17253
http://www.nber.org/papers/w17253

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2011

Financial support from the NET Institute, is gratefully acknowledged. We would also like to thank Dan Ackerberg, Amy Finkelstein, Shane Greenstein, Thierry Magnac, Chris Snyder, Marc Yesman, Andrew Sweeting, and seminar participants at the International Industrial Organization Conference, Boston University, Central European University, Chicago GB, the University of Mannerism, the Minnesota Applied Microeconomics Conference, the Wharton School of the University of Pennsylvania, the Quantitative Marketing and Economics Conference, the Tuck Winter IO Conference, the Fourth IDEI Conference on the Economics of the Software and Internet Industries, MIT, and the NBER Winter IO Meetings. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Heterogeneity and the Dynamics of Technology Adoption
Stephen P. Ryan and Catherine Tucker
NBER Working Paper No. 17253
July 2011
JEL No. C10,L10,O32

ABSTRACT

We estimate the demand for a videocalling technology in the presence of both network effects and heterogeneity. Using a unique dataset from a large multinational firm, we pose and estimate a fully dynamic model of technology adoption. We propose a novel identification strategy based on post-adoption technology usage to disentangle equilibrium beliefs concerning the evolution of the network from observed and unobserved heterogeneity in technology adoption costs and use benefits. We find that employees have significant heterogeneity in both adoption costs and network benefits, and have preferences for diverse networks. Using our estimates, we evaluate a number of counterfactual adoption policies, and find that a policy of strategically targeting the right subtype for initial adoption can lead to a faster-growing and larger network than a policy of uncoordinated or diffuse adoption.

Stephen P. Ryan
MIT Department of Economics
E52-262C
50 Memorial Drive
Cambridge, MA 02142
and NBER
sryan@mit.edu

Catherine Tucker
MIT Sloan School of Management
100 Main Street, E62-533
Cambridge, MA 02142
and NBER
cetucker@mit.edu

1 Introduction

Technological innovation lies at the heart of economic growth, and understanding the diffusion of innovation is a central question in economics. In his work on the diffusion of hybrid corn technology, Griliches (1957) poses three questions which still resonate today: What factors influence the timing of adoption of new technologies? What determines their rates of diffusion? Finally, what factors govern the long-run level of adoption? Griliches, along with other early empirical and theoretical work such as Mansfield (1961) and Rogers (1962), answers these questions by explaining differences in diffusion curves as arising from heterogeneity in user characteristics, such as profitability, cost, and competitive pressure.

Subsequently, Katz and Shapiro (1985) and Farrell and Saloner (1985) highlighted that for network technologies such as telephones and fax machines, a user's payoff from the technology depends on other people also adopting the technology, a phenomenon known as network effects. In a static setting these have shown to be important empirically for understanding levels of technology adoption. For example, Ackerberg and Gowrisankaran (2006) show that in a repeated static game of adoption, network effects matter for the adoption of electronic payments. In marketing, Nair et al. (2004) showed that in a static context network effects mattered for PDA adoption. For these types of technologies, variation in equilibrium beliefs can lead to differences in rates and depth of diffusion even for identical users.

We synthesize these two literatures by constructing a utility-based model of dynamic technology adoption which allows for observable individual heterogeneity in the adoption of a network technology. We examine how heterogeneity, as expressed by differences in adoption costs, network effects, and tastes for a diverse network, influences a network technology's diffusion. We apply our forward-looking expectations model to detailed data on the introduction of a videocalling technology in a multinational bank. This model allows us to quantify the effects of three dimensions of observable individual heterogeneity on network evolution and use, and permits analysis of two common policies for jump-starting network technology diffusion. Our research strategy consists of three steps.

First, we construct a model of dynamic network technology adoption and the subsequent sequence of technology usage. The model addresses two interrelated technological questions: how the installed base of adopters evolves over time, and how agents use the technology after joining the network. People vary in their net fixed costs of adopting the network technology, and weigh the expected present value of joining the network today against the opportunity costs of not adopting. This implies that adoption is an optimal waiting problem. After someone has adopted the communications technology, they decide how to use it. We model the sequence of calls as a function of two factors: the direct utility each person receives from interacting with others, and a desire to interact with different people over a sequence of calls. Our model allows us to provide a rich description of how diversity in the characteristics of network subscribers affects their propensity to adopt.

Second, we estimate our model using detailed data on the diffusion and use of a video-calling technology within a large multinational investment bank. We have detailed data on all 2,169 potential adopters in the firm, from the time that the technology was first offered for installation up to the network's steady state three and half years later. Using data on 463,806 video calls made using this technology, we estimate a model of calling preferences for 64 different types of employees in the firm. The bank deployed the technology in a laissez-faire manner: employees could install the technology at no cost to themselves, but were not compelled to adopt. This means that we can focus our attention on understanding how individual employees within the firm adopted the technology, without having to model the firm's adoption policy. We use recently developed techniques for estimating dynamic games to recover parameters for our model of technology adoption within this firm. Our approach to identification is novel, because our structural model allows us to identify and measure network benefits through *ex post* calling behavior, rather than merely inferring network benefits from correlations in adoption of network technology. A traditional reduced-form approach that considers only adoption decisions would require 64 valid instruments (one for each type of employee in the firm) to use these correlations to measure causal network effects, which is not feasible.

We find significant heterogeneity in adoption costs and usage benefits: Employees in the firm have different tastes for making video calls and adoption costs, depending on their location, job function, and rank. We find that, all else equal, a given function and region in the firm is more likely to call someone similar in the firm. By contrast, employees are more likely to communicate up and down the hierarchy, which supports some of the insights on the literature on the role of hierarchies and communication in firms, such as Garicano (2000) and Radner (1992). Our sequential calling model gives the additional insight that this taste for similarity decreases in the number of times a call is made. Employees therefore have significant positive welfare gains from having access to a diverse network where there are employees of many types for them to call.

Third, we use our estimates to simulate how two broad classes of different technology adoption policies focused on initial adoption could affect the evolution and use of the network over time. These policies represent potential deployment strategies that a firm or network operator can use to avoid sub-optimal diffusion for their technology. Under the first set of policies, the firm targets one type of employee as the initial set of technology adopters. This policy is common in many real-world settings, because firms commonly roll out a new technology in a specific work group, such as among all IT staff, before allowing wider adoption throughout the organization. We compare and contrast multiple different targeted seeding strategies. In the second class of policy, the firm adopts a uniform adoption strategy, where the technology is spread equally across various types in the initial period. Such a policy can be more effective when employees value being able to communicate with a wide variety of other employees. Comparing these two policies to the baseline case of decentralized adoption will allow us to evaluate the extent to which heterogeneity in employee behavior and characteristics must be accounted for in crafting an optimal policy for jump-starting the diffusion of a network technology.

Reflecting the complex interplay between heterogeneity in network effects among employees in the firm and heterogeneity in adoption costs, we find that the policy with correct targeted interventions dominates the uniform adoption policy. We compare numerous targeted policy interventions and find substantial differences in performance: the best performing seeding enjoys a two-thirds performance boost relative to the worse performing seeding. Our simulations emphasize that targeting should be used towards a subtype of employee that has both high adoption costs and large network effects on the adoption of others. By inducing a high-value subtype to enter early, the policymaker triggers a cascade of additional adoptions. This leads to slightly more calls per adopter, and significantly higher overall welfare. Our results show that focusing adoption efforts in a broad area, such as all workers in a region, can have very large relative benefits to the diffuse seeding policy. Our results reinforce the notion that the policymaker must carefully balance adoption costs and network benefits when considering any interventions.

Our paper makes several contributions to the existing literature on technology adoption and network effects that has largely focused on static models due to several challenges. First, in technology adoption models with network effects, the researcher must confront the issue of multiple equilibria that are not necessarily present in other dynamic contexts such as those studied by Gordon (2009) and Misra and Nair (2009). Both Ackerberg and Gowrisankaran (2006) and Rysman (2004) tackle this by estimating which equilibrium out of a limited set is selected. It is also theoretically possible to not limit the set of potential equilibria, and to model explicitly the equilibrium selection process, as in Bajari et al. (2009). However, this approach requires the computation of all equilibria to a system, taking a prohibitive amount of time. The size of the state space is the second difficulty. In the present application, for example, the state space is an indicator function for each employee's adoption status. The naive number of possible combinations of these variables is 2^{2169}, or approximately 10^{602}. Even reducing the size of the state space by grouping agents into subtypes, which discards information about which specific agent has adopted, still results to an impossibly large set of points to compute equilibria over. However, by using the two-step techniques described by Bajari et al. (2007), we circumvent the problem of multiple equilibria and the curse of dimensionality which beset estimation of dynamic technology adoption games since this methodology does not require the researcher to actually calculate equilibria.

The closest paper in approach to our research is Dube et al. (2010) who also use Bajari et al. (2007)'s methodology to estimate a dynamic model of tipping in the video game console market. The key difference between our paper and theirs is that we estimate a rich model of observable consumer heterogeneity and show how this heterogeneity matters for understanding the diffusion process. By contrast, they use a single time-series for the US market, and this aggregate approach allows them to solve a dynamic hardware pricing game. We are also able to use the data on how consumers use the technology to estimate utility and consequently demand for the product. The methodology we present in this paper represents a first attempt in the literature to develop a general approach, albeit with several simplifying assumptions, to estimating demand for network goods in the presence of unobserved heterogeneity.

We also contribute to a more general literature that uses data on ex-post behavior to infer the value of a particular action. For example, Reiss and Spiller (1989); Ellickson and Misra (2010) use ex-post data on revenues to infer fixed costs of entry; Draganska et al. (2009) use ex-post data on demand/prices to infer fixed costs of product introduction; (Keane and Wolpin, 1997) use ex-post data on wages conditional on being in a sector to learn about the value of choosing a job-sector; and Hartmann and Nair (2010) use ex-post data on purchased of tied-complementary goods to learn of the utility from adopting the primary tying good.

The contribution of our research is the use of this ex-post data in inferring and measuring network benefits.

Identifying network effects is problematic.[1] Much of the early empirical work focused on identifying causal network effects, for example Gowrisankaran and Stavins (2004). Tucker (2008) and Tucker (2009) pursue a reduced-form approach using the same data as this paper. The crucial difference between this paper and previous approaches is that we directly model and measure the network benefits through observations on network use after adoption, rather than relying on exogenous shifters of network adoption as a means of causal measurement of imputed network benefits from an installed base. This new approach to identification allows us to disentangle the convoluted effects of heterogeneity in stand-alone use and utility due to the size and diversity of the network. This approach also allows us to explore a broader spectrum of heterogeneity than has previously been considered. Furthermore, since we measure equilibrium network benefits, our structural estimates allow us to set up policy experiments that illuminate how different seeding policies affect network evolution.

Another potentially convoluting factor in estimating demand with network effects are correlated unobservables. For example, there may be some external influence that makes the value of the networking technology exogenously increase in a certain time period. These shocks will induce correlation in the adoption behavior of agents in that period outside that predicted by the fundamentals of the model. Our approach also allows us to account for correlated shocks in adoption behavior. The calling model allows us to recover any network-wide shocks which lead to higher or lower utility of using the network in a given period. Utilizing the panel nature of our calling data, we can recover these shocks and account for their influence in the adoption policy equation.

The paper is organized as follows. Section 2 describes the technology and data used in

[1] Rose and Joskow (1990) discuss how identification is important in generalized diffusion models when trying to put a causal interpretation on a variable such as firm size.

this study. Section 3 lays out a dynamic model of technology adoption choice and subsequent interaction choice. Section 4 discusses our estimation strategy. Section 5 discusses the results of our estimation. Section 6 reports results from a policy experiment to test two alternative technology adoption policies. Section 7 concludes and discusses directions for future work.

2 Technology and Data

We study adoption of a desktop-based videocalling technology within a single multinational bank. The primary benefit of videocalling is that it can improve the effectiveness of oral and written communication by adding visual cues. Older videoconferencing systems failed in part because they were based on expensive and inconvenient videoconferencing rooms. The videocalling technology studied in this paper was attached to an employee's workstation. The end-point technology has three elements: videocalling software, a media compressor, and a camera fixed on top of the computer's monitor.

Three institutional details are central to our analysis. First, videocalling could only be used for internal communication within the firm. This means that we have comprehensive data on both the set of all potential adopters and how the technology was used after adoption.

Second, the bank pursued an unusually *laissez-faire* approach to promotion and adoption of the technology in the firm. After the bank chose this technological standard to conduct internal videocalling, it invested in the basic network components which would form the backbone of the network infrastructure. The bank then publicized the availability of the technology to employees. Each employee independently decided whether and when to order a videocalling unit from an external sales representative. The firm paid for all costs associated with the adoption. The videocalling vendor and bank employees have confirmed that there were no supply constraints which might have restricted adoption. Though such explicitly decentralized adoption is unusual, it is not uncommon for companies to install software or IT equipment for employees and then leave it to the employee's discretion whether or not they use it.

Third, though the firm paid the monetary costs of installing the technology there were still significant non-monetary fixed costs for the employee which play a central role in our analysis. In particular, they had to set aside a morning where there computer would be down, the software installed and they were trained on the device. Interviews with bankers at the firm confirmed that they viewed having their computer out-of-action for a number of hours a large costs, given how quickly financial markets move.

We combine two databases: Personnel information for all employees as of March 2004, and complete records of videocalling adoption and use from the first call in January 2001 through August 2004. There were 2,169 employees who qualified as potential adopters, of whom slightly over 1,600 eventually adopted the videocalling technology.[2] Based on data on each employee's job description, rank, and location, we sorted all potential adopters into 64 broad types. These broad types were demarcated by hierarchy (Associates, Vice-Presidents, Directors, and Managing Directors), function (Administration, Research, Sales, and Trading) and geographical location (Asia, Britain, Europe, and the US).[3]

Figure 1 shows the count of employees by type in the firm. Comparing these results to Figure 2 shows that by contrast the pattern of adoption rates in the firms by August 2004 is highly regular: the higher the employee's rank in the firm, the higher the adoption rate. This variation, and the other variation that we observe in adoption rates across regions and functions, suggests that employees benefit to differing extents from the technology. This drives us to emphasize heterogeneity in our model.

The call database recorded each of the 1.7 million two-way calls from January 2001 to August 2004. There were also 752,055 uses of the technology for its alternative use of TV-watching which we ignore. Of these 1.7 million calls, 1,052,110 actually went through rather

[2]We exclude from our analysis 300 employees who left the firm during our sample period.

[3]The Asia region also includes a small number of isolated offices in other locations.

Figure 1: Distribution of Employees by Type

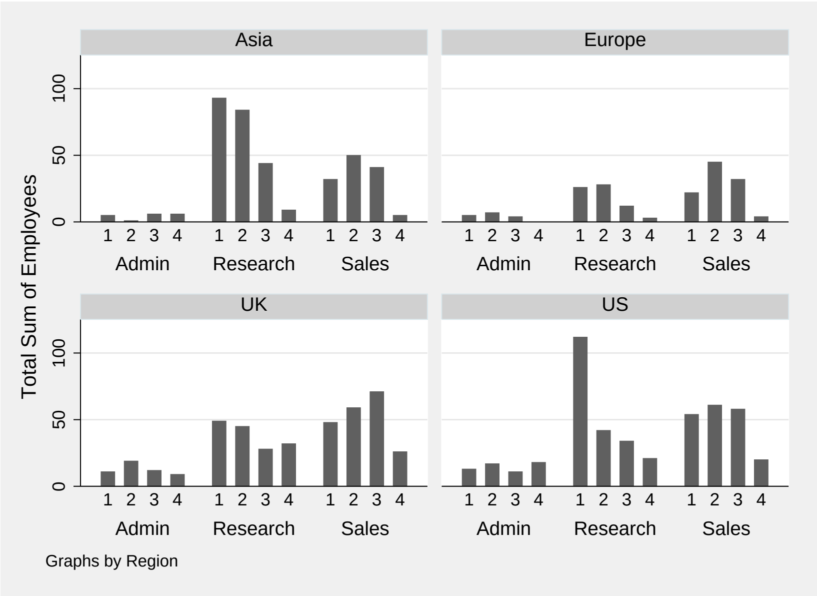
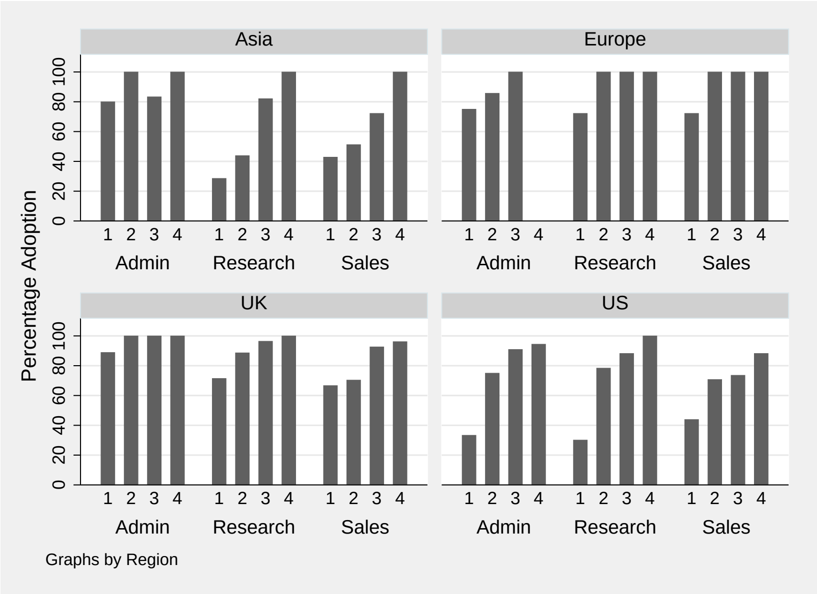


Figure 2: Distribution of Adoption Rates by Type



than being ignored or not answered. For two-way video calls, the database records who made a call to whom, and the timing and length of the call. Each accepted call lasted on average 5 minutes and 46 seconds.[4] We excluded from our data calls which involved an isolated Finance/Credit Analysis division (17 percent of calls) that focused on retail banking rather than investment banking; calls with more than two participants (5 percent of calls);[5] calls made by employees who left the firm (17 percent of calls); and calls that ended in error (14 percent of calls). This left us with a data set of 463,806 person-to-person calls.[6]

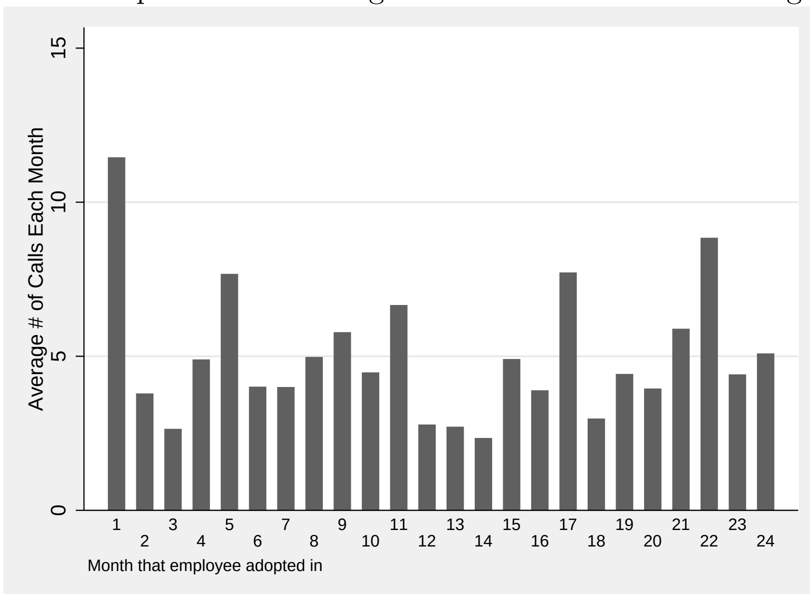
Figures 3 explores the relationship between the number of calls and the timing of adoption. The y-axis shows the average number of calls per month, while the x-axis reflects which month the employee adopted in. It is clear that the relationship is not monotonic. This supports our notion that, conditioning on a specific subtype of agent, variation in adoption timing is driven by variation in the agent's fixed cost and not heterogeneity in calling utility. However, once an individual adopts the technology, we do not see large differences in how they use it over time. Figure 4 shows the number of individuals an employee who had installed the technology in the first three months, called over the following 2 years. It shows that adopters tended to call a relatively stable proportion of users of the network as the network evolved. This suggests against an alternative 'learning story' that might explain delay, as it was not the case that employees adopted, tried out the technology and then stopped using it. Therefore, a central component of our model has to be one that explains the delay in adoption that we observe in the data, but also the relative stability in calling

[4] We do not model the length of calls; low-intensity long calls can be as useful as high-intensity short calls.

[5] Fewer than 5 percent of calls involved more than two people.

[6] Discarding some of the calls may induce bias in our estimated parameters. Removing the Finance/Credit Analysis division could artificially reduce the number of agents in the network, deflating the estimates of the network's worth at any point in time, and as a result will bias the estimates of the fixed costs of adoption downward. Due to the small number of employees who adopted and the fact that only two employees ever called another employee in this division which was focused more on retail banking, this effect is probably small. The same effect is possible with calls from employees who were fired. Though these employees who left the firm made a similar, though slightly lower number of calls to those who did not leave, we are not able to trace out their adoption as we do not have details about their title, function and location within the firm.

Figure 3: Relationship Between Average Number of Calls and Timing of Adoption

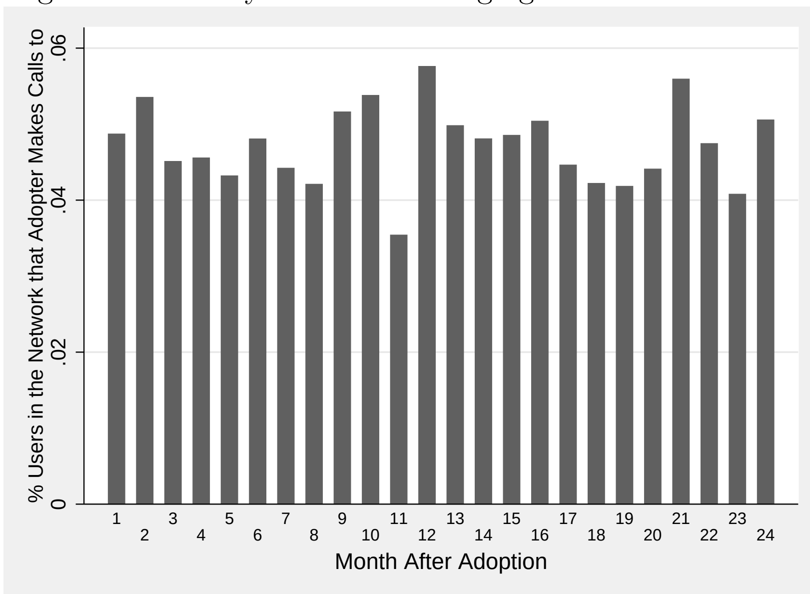


behavior that we observe over time after adoption.

Besides the heterogeneity across types suggested by Figure 2, there may also be heterogeneity within a type. Figure 5 provides further evidence of heterogeneity within our types by graphing the adoption patterns and number of calls for US researchers across different ranks. The left panel shows the cumulative adoption levels for each of the four ranks. Cumulative adoption varies across titles, with Managing Directors adopting the technology at a much higher level than Associates. However, not all employees within each type value videocalling equally; the adoption graphs are not a single stepwise function where all similar employees adopt at the same time. Instead, these graphs imply that any model of adoption in this setting must account for differences in adoption rates both across and within types of employees.

The right panel of Figure 5 shows the average volume of monthly calls by rank over time for US Researchers. Call volumes differ across ranks: For example, Vice Presidents make more calls on average than Directors. Though the data are noisy, call volume appears to grow each month. This is as expected, given that there are more potential receivers for an

Figure 4: Stability in Video-messaging Behavior over Time



employee to call later on. This gives some evidence of a "network effect," in the sense that intensity of use of the technology is growing with the number of adopters.

3 Theoretical Model

Our theoretical model uses a utility-based foundation to rationalize variation in adoption rates and calling patterns across and within types of employees. We characterize the adoption decision as an optimal waiting problem, with each employee joining the videocalling network when the expected benefits of adoption exceed the opportunity cost of non-adoption. Our model is dynamic because each employee computes the expected benefit of joining the network as a sum of utility flows accruing from future network use when making an adoption decision. This utility flow critically depends on expectations about growth of the installed base.

There are three basic ingredients to the model: a set of state variables describing whether each employee has adopted at a given time, payoffs which accrue to each employee as a function of the state variables and their actions, and the process governing changes in the

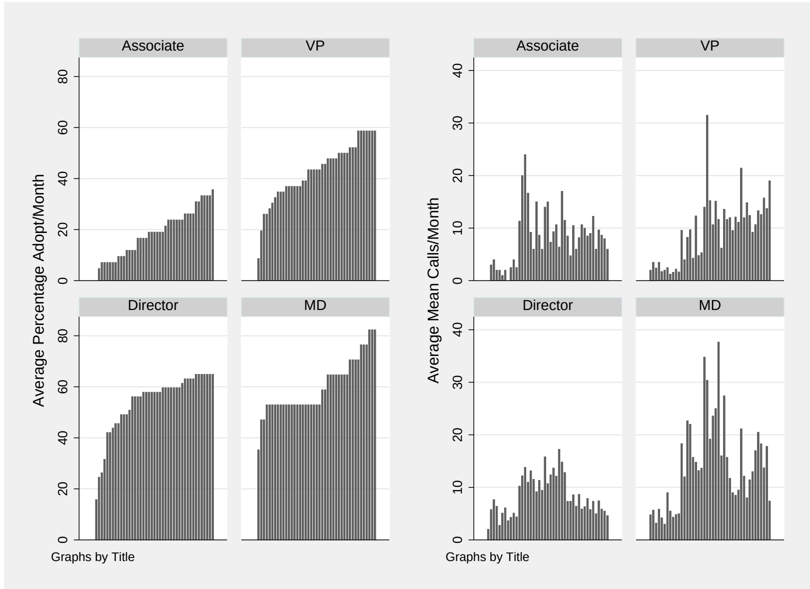


Figure 5: Cumulative Adoption and Call Volume for US Researchers

state space over time.

Before discussing these three components of the model, we note that we restrict the set of Nash equilibria to be Markov-perfect: that is, the current state and current vector of actions are sufficient to characterize the dynamics of the model. This approach has become standard in much of the literature on dynamic games due to the relatively tractable set of implications that it generates, as compared with the the unboundedly complex set of history-dependent strategies. This is due to the fact that in Markovian equilibria players best respond to current actions rather than histories of actions.[7] In our setting, the Markovian assumption implies that adoption behavior is only a function of the current network configuration, which considerably simplifies our analysis.

[7] See Maskin and Tirole (1988) for a discussion of Markovian games and further examples. Also see Maskin and Tirole (2001) for a detailed analysis of Markovian games and justification for focusing on Markovian strategies.

3.1 State Space and Timing

The state space s describes which employees have adopted the technology. Each element of the state space, s_{it}, is an indicator function representing whether each employee i adopts by month t. Employees face an infinite time horizon and discount their payoffs with the same discount factor β. In estimate we need to stipulate the value as it is not separately identified and set β=0.9..

Each employee is endowed with a set of characteristics which describe their role in the firm, that are both exogenous and time invariant. We denote the vector of characteristics of employee i by:

$$x_i = \{\hat{e}^r, \hat{e}^f, \hat{e}^t\}, \tag{1}$$

where each $\hat{e}$ is a 1×4 unit vector representing the region, function, and title of each employee. We order alphabetically: The regions as Asia, Britain, Europe, and the United States; and the functions as Administration, Research, Sales, and Trading. We order the titles by importance in the firm: Associate, Vice President, Director, and Managing Director. For example, a Vice President of Administration in Europe would be represented as $x_i = \{(0,0,1,0),(1,0,0,0),(0,1,0,0)\}$.

3.2 Per-Period Payoffs: Communications utility

In each period, an employee who has adopted the technology makes as many video calls as she desires to any other employees in the network. To capture the benefits of these interactions, we develop a simple model of a utility-based sequence of interdependent choices. Specifically, we model the utility employee i obtains from making the k-th call in a calling sequence to employee j as a function of both caller and receiver characteristics and the set of previous calls already made in that month.[8]

[8]Tucker (2008) explores the differences between modeling calls as one-way or two-way process and presents evidence that for modeling purposes the directionality of calls is not empirically important.

In period t, the connection utility of agent i calling agent j as the k-th call in sequence is:

$$U_{ijkt} = \underbrace{\theta_1 + \theta_2'\Gamma}_{\delta_1} - \underbrace{\theta_3'\eta_{ik} + \theta_4(k-1)}_{\delta_2} + \xi_t + \epsilon_{ijk} \tag{2}$$

This utility function is composed of three parts: a static connection utility, δ_1; a composition component, δ_2 that depends on the set of previous calls made in the current period; and a stochastic shock to utility, $\xi_m + \epsilon_{ijk}$, which is composed of two terms. The first term, ξ_t, captures correlated shocks that make the network particularly attractive or unattractive to use in period t. The second term, ϵ_{ijk}, is an idiosyncratic error term which is distributed Type-I extreme value with unit variance, and represents connection-specific shocks to utility unobserved by the econometrician.[9] The utility of not making a call is normalized to zero.

The static component, δ_1, is composed of a constant, θ_1, and the Γ function which governs the potential caller-receiver interactions. The constant θ_1 determines a baseline utility that influences how many calls a given employee will make in any period. Each element of $\Gamma = (\gamma_{ij}^r, \gamma_{ij}^f, \gamma_{ij}^t)$, γ_{ij} is a vector defined by the interaction of each characteristic e_i of the caller in Equation 1 with the corresponding characteristic e_j of the receiver:

$$\gamma_{ij} = (\hat{e}_{i1}\hat{e}_{j1}, \ldots, \hat{e}_{i4}\hat{e}_{j1}, \hat{e}_{i1}\hat{e}_{j2}, \ldots, \hat{e}_{i4}\hat{e}_{j2}, \hat{e}_{i1}\hat{e}_{j3}, \ldots, \hat{e}_{i4}\hat{e}_{j3}, \hat{e}_{i1}\hat{e}_{j4}, \ldots, \hat{e}_{i4}\hat{e}_{j4}). \tag{3}$$

Intuitively, the Γ function zeros out all the interaction terms which are not relevant for the connection between two given employees. In the terminology of Jackson and Wolinsky (1996), $\theta_2'\Gamma$ measures the "link synergy" between two types of employees.

The second component of the utility function, δ_2, reflects changes to the utility of a connection as the employee makes additional calls. The first component, $\theta_3'\eta_{ik}$, reflects the intuition that employees may get diminishing value from calling the same group of employees

[9] Our specification does not explicitly account for intertemporal correlation in calling patterns, as in the case of two coworkers who work on the same project and who repeatedly call each other.

repeatedly. We account for the number of previous calls to a characteristic within a sequence using the function η_{ik}, which is a 12×1 vector counting the number of times employee i has made calls to each of the 12 possible employee characteristics in the previous $k-1$ calls. The term $\theta_3'\eta_{ik}$ captures these effects by allowing the marginal utility of calling employee j to depend on the number of previous calls to other employees with similar characteristics. The second component of δ_2, $\theta_4(k-1)$, shifts the marginal utility of making any calls linearly in the number of calls previously made in the current month. This captures the idea that the opportunity cost of using the videocalling technology is increasing due to the need to attend to other work-related activities.

3.3 Generating a Calling Sequence

Each employee makes a sequence of calls, denoted by Ω_t, during calling period t. Indexing agent calls in increasing order by k starting at $k = 0$, the agent maximizes utility by either making a video call to another worker in the network or choosing the outside option, which gives a utility of zero. The utility of the k-th choice is then:

$$U_{ikt} = \max\{0, \max_{j \in N_t}\{U_{ijkt}\}\}, \tag{4}$$

where U_{ijkt} is the utility of calling worker j, and N_t denotes the set of agents who have adopted the technology at time t. The connection utility depends on k is two ways: first, the connection utility may be shifted by decay parameters due to previous calls to other workers with similar characteristics as worker j, and second, the error term in the utility of each connection is drawn anew for each k choice in the sequence.

Each employee makes calls until the best marginal call has a negative utility. The length of the sequence is then determined as:

$$K = \{K \in \mathbb{N}_0 : \max_{j \in N} U_{ijK} > 0\}, \tag{5}$$

where $\mathbb{N}_0$ is the set of non-negative integers.

The error term, ϵ_{ijk}, helps rationalize why employees do not make the same number of calls each period to the same set of receivers, and also helps explain why network benefits increase as the network grows. The addition of a marginal adopter is important to the installed base for two reasons: first, that new adopter may be of a different type than currently exists in the network. This means there are new possibilities for connection synergies between that employee and the installed base. Second, there is also one more draw from the set of stochastic connection utilities. This extra draw is important because the calling sequence that results from the optimization problem in Equation 4 is driven by order statistics: the expected value of the maximum over random utilities is increasing in the number of potential receivers. Therefore, the more employees there are in a network, the higher the number of expected calls, even if they all share the same type. We note that this is a generic property of random utility models, and is not specific to the distribution of the error term used here. The marginal increase in the expected maximum also decreases as the network grows, which we think is also an intuitive implication of network growth with stochastic connection utilities.[10]

Our model presumes that the agents are myopic, in that they do not anticipate future calls when making current calling choices. Ideally, given the intertemporal nature of calling patterns in the data, one would model the call choice through a dynamic model. However, we feel two considerations motivate the use of a myopic model. The first is that our calling model is consistent with a world where agents do not know ahead of time what calls they will need to make in the future. If agents in the firm make calls to resolve questions regarding current workloads, it is plausible that they will not know who they may need to call in future periods for different tasks. More to the point, given the pressures involved in working at an

[10] Our model is agnostic about the identity of agents from two subtypes who call each other. Our calling model is consistent with the notion that two different agents repeatedly call each other, but is not as efficient as an estimator which explicitly accounts for such autocorrelation in the error structure.

international bank, it seems reasonable to assume that agents do not care about the specifics of how many calls they will make in the future and to whom—they simply make calls out of present necessity, without any strategic view of the future. This view of how the technology is used in each calling period is not incompatible with a dynamic model of adoption—the adoption model simply says that agents understand that the benefits of using the technology accrue over time and grow with network size.

The second consideration is that if one wants to model the calling process as a fully dynamic process, then it is necessary to take a stand on how frequently the error terms are renewed.[11] Our myopic model maintains that calling shocks are drawn anew after each completed phone call. In a dynamic model, it is necessary to define how frequently these updates take place to identify when a call could have happened but did not. If we take the shortest interval in the data as the minimum time between successive calls, this defines an upper bound on the number of possible calls in a calling period. One could then take this model to the data by simulating the calling sequence up to the point where the agent no longer has expected utility above zero of making future calls.

Philosophically, this approach is appealing from the perspective of allowing agents to be forward-looking in both their technology adoption and use. However, it is significantly more complex than the simple myopic model presented here, both in terms of modeling and in ease of estimation. Furthermore, it is unclear how the empirical content of the more complex model would differ significantly from the myopic model. Since the decay parameters in the utility function are linear, any re-ordering of the calls in a given sequence obtains the same overall level of utility, and thus the dynamics of this choice are unimportant. Both models match the number and order of calls; since the decay parameters are estimated conditional on making a call and the prior calling history, the decay parameters would be qualitatively similar across both approaches. The number of calls would mechanically appear to be lower

[11] In principle, one would also impose a discount factor within the calling period.

in the dynamic model; however, this would only imply a lower level of the intercept. Given that the continuation value of making calls in the future contains more draws of the error term, this adjustment in the intercept would simply counteract that additional utility. The level of utility in the dynamic model would differ, but this is not critical since the model only considers relative utility between choices. Therefore, though we recognize that a fully dynamic model of calling would be preferable and more internally we consistent, we feel that practically the loss from using a myopic calling model is not great from an economic perspective.

3.4 Transitions Between States: Technology Adoption

The second component of our model is whether employees who are outside the network adopt. At the beginning of each period, every employee who has not already adopted the videocalling technology can do so. Adoption is instantaneous and the employee is able to make calls immediately. We assume that it is not possible to divest the technology. This seems reasonable, given that the option value of holding the technology is always positive in our model, and we did not observe any divestitures in our data.

If an employee adopts, she can expect to use that technology to communicate with others in the network, both today and in the future. The value function for adopters is:

$$V_i(s_t, s_{i,t} = 1) = E\left[U(\Omega_{it}) + \tau_i - F_i + \beta V_i(s_{t+1}; s_{i,t+1} = 1)\right]. \tag{6}$$

We can write the value function for each potential adopter as:

$$\begin{aligned} &V_i(s_t, s_{i,t} = 0) = \\ &\qquad \max\left\{E\left[U(\Omega_{it}) + \tau_i - F_i + \beta V_i(s_{t+1}; s_{i,t+1} = 1)\right], \beta E\left[V_i(s_{t+1}; s_{i,t+1} = 0)\right]\right\}, \end{aligned} \tag{7}$$

where expectations are taken with respect to that employee's beliefs about both how the network is going to evolve in all future periods and the associated distribution of future calling utilities. As in Farrell and Saloner (1985), the benefits of adopting the videocalling technology consist of both the network benefit derived from the stream of expected discounted calling utilities, $E\left[U(\Omega_{it})+\beta V_i(s_{t+1};s_{i,t+1}=1)\right]$, and the stand-alone benefits (such as watching television), denoted by τ_i. Without loss of generality, we set $\tau_i=0$, since the stand-alone benefits and adoption costs are not separately identified in the model.[12] If the employee does not adopt the technology, she receives the expected discounted continuation value. The employee solves an optimal waiting problem, adopting the technology when the benefits exceed the opportunity cost of adopting in a future period.

The cost of adopting the technology to the employee are the time spent setting it up and learning how to use it. The firm bears all monetary costs. To reflect this installation cost, we assume that adopters have to pay a one-time up-front fixed cost of F_i. F_i is drawn from a distribution that is known to all employees. We assume that F_i is time-invariant and the realized value is known only to the employee.

The employee has three dynamic considerations when evaluating how the network may evolve. First, an employee with a large F_i has an incentive to wait for the installed base s_t to grow and cover the net fixed costs. Second, employees may anticipate that their adoption now may spur other employees to adopt in future periods. Farrell and Saloner (1985) explain how this forward-looking behavior may help reduce the coordination failure. Forward-looking behavior could just nudge inframarginal non-adopters towards adoption without visible effect, or it could potentially generate an entire cascade of adoptions. Third, employees receive option value from postponing adoption. Even though all employees have

[12]This is true even if τ_i is time-varying. The reason is that we can write the expected discounted stream of stand-alone benefits as $\bar{\tau}=\sum_{t=0}^{\infty}E[\tau_{it}]$. Note, however, that this results in the same formulation as Equation 7 since the stand-alone benefits are a stochastic stream of benefits that accrue to the agent regardless of the evolution of the state space or actions by the agent. Therefore, the stream of payoffs can be replaced by the expected discounted value, and it is clear from Equation 7 that only the difference τ_i-F_i is identified at the point that the agent adopts the technology.

rational expectations about the expected evolution of the network, the presence of private information in the fixed costs of joining the network implies variance in who actually joins in any period. The resolution of this uncertainty over time creates the option value of delaying adoption.[13]

3.5 Equilibrium and Network Evolution

Adoption in Equation 7 depends critically on each employee's beliefs about how the network is going to evolve in all future periods. As discussed above, we assume that players only condition on the current state vector when making choices, which gives rise to a Markov perfect Nash equilibrium in network evolution. Equilibrium obtains when all employees have beliefs which ensure that no employee has an incentive to change their action or beliefs in response. Without formally deriving any properties of the equilibria of our model, we note that such models typically involve a large set of admissible equilibrium beliefs. As is discussed in the next section, we do not need to solve for the equilibria of our model in order to estimate its underlying structural primitives.

3.6 Heterogeneity in Calling Parameters

We note that we do not allow the coefficients on the utility parameters to vary in an unobservable fashion across the population of agents. This is a very strong assumption, as *a priori* it is reasonable to assume that the usefulness of the videocalling technology varies within our 64 identifiable groups. The reason for this modeling choice is that this unobserved heterogeneity produces a significant selection problem in the adoption equation. Agents with a sufficiently advantageous combination of high calling marginal utilities and low adoption costs will adopt the technology at any given point. This generates a selected sample, particularly with early adopters, and as a result the distribution of marginal calling utilities

[13] We present a two-period, two-agent model in Section A.1 of the appendix to illustrate analytically why agents may find it optimal to wait as uncertainty is resolved. While our model has many agents and many time periods, the intuition of the simple model extends to our more complex setting.

in that sample is not representative of the distribution of calling utilities in the firm as a whole. Without further assumptions, this implies that we are not able to disentangle the two distributions when considering the adoption decision of agents. The typical solution to this problem is to assume exclusion restrictions across the adoption and calling equations. Under sufficient support conditions, one drives the adoption probabilities of a subgroup to one, which then reveals the unconditional distribution of marginal utilities in their post-adoption behavior. However, in our application such excluded variables are not apparent to us, and as a result we have decided to model the calling utilities as homogenous in the population of potential adopters. As a result, our results are biased toward overstating the value of the videocalling to the firm as a whole, as the lowest-utility agents are precisely those that will not have adopted the technology.

4 Estimation

A simulated method of moments approach, where we search for parameters that match the simulated moments from the model to their empirical counterparts, is infeasible in our setting. Such an approach requires solving the dynamic model of Section 3 for each iteration of a nonlinear optimization program. Unfortunately, the computational burden of this approach is astronomical, as we would have to solve for the fixed point of an extremely large system of nonlinear equations.

To circumvent this problem, we follow Bajari et al. (2007), who advocate a two-step approach for estimating dynamic games. The intuition of their approach is that we can let the employees in the firm solve that dynamic program for us, rather than calculating it ourselves. Under the assumption that the employees optimize their adoption decision as in Equation 7, we find parameters such that their observed behavior is optimal. In the first step, we recover reduced-form policy functions which describe the equilibrium strategies followed by each employee as a function of the state vector. In the second step, we project these functions onto our dynamic model of technology adoption choice and usage. In this manner, we recover consistent estimates of the underlying parameters which govern the process of network evolution and utilization.

There are two separate policy functions in the first stage. The first policy function describes how the network will be used by employees who have adopted the technology. We develop a "simulated sequence estimator" to estimate the calling utility parameters in Equation 4 defining how employees use the videocalling technology. The second reduced-form policy function describes the propensity to join the network, given the number and composition of current users.

4.1 Comparison to BBL

It is worth comparing and contrasting our approach to the typical two-stage approach in BBL. Like BBL, in a first stage we estimate reduced-form policy functions that arise from a complicated dynamic programming problem to circumvent the need to solve for the equilibrium of the model. In contrast with BBL, our approach does not require us to perturb these estimated policy functions in a second stage, as all of the important parameters in our model can be recovered using threshold conditions from the underlying dynamic model. This difference arises since we do not have "dynamic" parameters in the payoff function, in the sense of Ryan (2010), who estimates parameters such as investment costs. As a result, our estimation approach is considerably more direct.

4.2 Simulated Sequence Estimator

Our goal is to estimate the utility calling parameters which govern how employees use the network once they adopt the videocalling technology. For a given calling sequence, Ω, of length K, the simulated sequence estimator splits the calling sequence problem into two parts by exploiting the following identity:

$$Pr(\Omega, K) = Pr(\Omega|K)Pr(K) \tag{8}$$

$$\ln Pr(\Omega, K) = \ln Pr(\Omega|K) + \ln Pr(K). \tag{9}$$

The simulated sequence estimator first estimates the composition of the call and then estimates the parameters which determine the number of calls.[14] This separation greatly simplifies our estimation, as connection utility parameters do not depend on parameters governing the number of calls an agent makes.

4.2.1 Connection Utility Parameters

The assumption that the error term in Equation 2 is distributed type-I extreme value generates a logit probability of observing a call from employee i to employee j as the k-th call of a sequence in period t:

$$Pr(\Omega_{ijkt}; s_t, \theta_2, \theta_3) = \frac{\exp(\bar{U}_{ijk}(\theta_2, \theta_3))}{\sum_{j' \in s_t} \exp(\bar{U}_{ij'k}(\theta_2, \theta_3))}, \tag{10}$$

where $\bar{U}_{ijk} = U_{ijk} - \epsilon_{ijk}$. The outside option does not enter the probability of a call as it usually does in discrete choice models, as we are conditioning on the length of the sequence. Computationally, we find parameters to maximize the probability of observing each call in the sequence in that order. We use the order of calls within the sequence to identify the parameters governing the taste for diversity δ_2, as the conditional probability of each call in the sequence depends on the calls made before it. Specifically, the relative frequency with which we observe two calls to the same subtype in a given sequence identifies δ_2 for that subtype. We note that these shifters only enter into the connection utility of other agents

[14] In earlier versions of this paper, we discussed Monte Carlo evidence that suggested this simulated sequence estimator performs well in small samples, and that joint estimation is badly biased in small samples. These results are available upon request from the authors.

with those characteristics. For example, if a worker has called Administration once in the past, this shifts the utility of all other agents with the Administration characteristic by the relevant shifter while leaving all the other agents with different functions unchanged. We apply this estimator to all of the video calls made by employees in this firm during the last three months of our data.

In the first step of our estimator we maximize the following likelihood function:

$$\max_{\{\theta_2,\theta_3\}} \sum_{t=1}^{T} \sum_{i=1}^{N_t} \sum_{k=1}^{K_{it}} \ln Pr(\Omega_{ijkt}; s_t, \theta_2, \theta_3), \tag{11}$$

where t indexes time, N_t is the number of agents in the network, s_t is the state of network, K_{it} is number of calls in sequence Ω_{it} by worker i. We search for the parameters governing the connection utilities in order to maximize the predicted probabilities of the observed sequences by each agent in the network at each point in time.

4.2.2 Sequence Length Parameters

The second step in the simulated sequence estimator recovers the parameters which govern the length of the sequences. To solve for these parameters, we use a simulated method of moments approach. We use the calling parameters we found in step one to generate U independent calling sequences by repeatedly simulating the process defined in Equation 4 for each employee in each month for the installed base at any time. We then compute the expected sequence length by averaging over these simulated sequences:

$$\hat{K}_{it}(s_t; \theta_1, \theta_4, \xi_t) = \frac{1}{U} \sum_{u=1}^{U} \left| \Omega_{itu}(s_t; \theta_1, \theta_4, \xi_t) \right|, \tag{12}$$

where $|\Omega_{itu}(s_t;\theta_1,\theta_4,\xi_t)|$ is the length of the u-th simulated calling sequence for employee i in period t. We then perform the following minimization program:

$$\min_{\{\theta_1,\theta_4,\xi\}} \sum_{t=1}^{T} \sum_{i=1}^{N_t} \left(\hat{K}_{it}(s_t;\theta_1,\theta_4,\xi_t) - |\Omega_{it}| \right). \tag{13}$$

We search for parameters such that we match the length of observed calling sequences against the sequence lengths predicted by the process in Equation 4.

4.2.3 Correlated Unobservables

We note that our approach allows for the estimation of correlated unobservables in usage utility through the ξ_t shocks. These shocks can be consistently recovered from panel variation in the average number of calls made by the network in different periods, controlling for the observable characteristics of the network. Our baseline specification allows for time-specific shocks, although we also consider more disaggregated shocks which shift the overall level of calling utility by function, region, and title. We use the estimates of these shocks to control for correlated unobservables in the adoption equation, which we in the next section.

However, it is important to also mention that there are other potential ways of accounted for correlated shocks and unobserved heterogeneity. For example, the more tractable nature of their setting means that Keane and Wolpin (1997); Ellickson and Misra (2010); Hartmann and Nair (2010) control directly for unobserved heterogeneity is in the error term of agent usage decisions which may be correlated with the normally distributed unobserved fixed cost component.

4.3 Adoption Policy and the Evolution of the Network

The second policy function that we estimate governs the choice of videocalling technology adoption. We focus our attention on estimating the proportion of each subtype that adopts the technology in each period, as this is sufficient for characterizing both the evolution of the network and the probability of any given agent adopting in a given period. In general, this policy is a function of the current state of the network, the lagged state of the network, and the time-specific unobserved shock to the network utility. The current state influences adoption as the utility of the technology depends on the size and composition of the current network. Lagged adoption controls for the selected sample of potential adopters after the first period, since our model implies that agents who have not adopted previously will have higher fixed costs of adoption than those who have. Combined with the fact that the value of adopting the network is always increasing in our model, the value of the network in the last period is greater than any periods previous to that. Therefore, in order to account for the selection in the distribution of fixed adoption costs, it is sufficient to only keep track of how valuable the network was in the last period. The reduced form adoption policy is also a function of the shock to the network utility, as agents are more likely to adopt the technology when the utility of adoption is particularly favorable. Ideally, one would estimate the policy function using a nonparametric estimator of those three variables. However, due to data constraints we restrict our policy functions to belong to a parametric family. We estimate the proportion of adopters of employee type m as a function of current and lagged state variables:

$$Proportion(adopt_m = 1; s_t, s_{t-1}, \lambda) = \lambda_1' x_m + \lambda_2'(x_m \otimes \nu_t) + \lambda_3'(x_m \otimes \nu_{t-1}) + \lambda_4 \xi_t, \quad (14)$$

where x_m is defined in Equation 1, ν_t is a 12×1 vector enumerating the counts of employee characteristics currently present in the installed base, the operator $\otimes$ represents element-wise multiplication, and $\lambda_4 \xi_t$ captures the effect of the correlated unobservable in period t calling utility described above.

The functional form of this policy function is guided by our model of adoption. First, λ_1 allows for the possibility that different employee groups in the firm have different propensities to join the network. Second, we have assumed that the fixed cost of adoption is employee-specific private information that is time-invariant. Since expected calling utility is weakly increasing in the size of the network, this assumption implies that the proportion of people within an employee type who adopt is nondecreasing in the size of the network. Therefore, we restrict the coefficients on these state variables, λ_2, to be positive. Another way of stating this restriction is that the probability of adoption by a given worker should not decrease if there is an exogenous expansion of the network's installed base.

4.4 Selection in the Adoption Policy Function

We include lagged state variables to correct for the fact that the distribution of adoption costs changes over time due to selection. The technology adoption process described in Equation 7 produces a selected sample of potential adopters after the first period. The reason is that lowest-cost adopters choose the technology in the first period, removing the lower left tail of the distribution of adoption costs. Furthermore, since the value of the network is non-decreasing over time, due to the positive benefits of more users combined with the assumption that no one can ever discard the technology, the threshold value at which users adopt the technology is also non-decreasing in time. This implies that the threshold for adoption in the last period is a sufficient statistic for capturing the degree of truncation in the current period's distribution of fixed costs. In the reduced form, we capture this effect by including one-month lagged state variables and restricting the coefficients on these state variables, λ_3, to be negative.[15]

The inclusion of $\lambda_4 \xi_t$ captures the role of correlated shocks in network utility in the

[15] To illustrate this point further, suppose we exogenously placed two observably identical agents each into one of two networks, where one network has an initially larger installed base than the other. Suppose that neither agent adopts the technology in the first period, and that both networks grow to the same size in the second period. The probability of seeing the agent in the initially smaller network adopt is now higher than the agent who did not adopt in the initially larger network, as the adoption benefits in that network were larger, and therefore the agent has revealed themselves to be of at least as high a type as the agent who did not adopt in the small network. This intuition translates into our restriction on the sign of λ_3 in the reduced-form adoption policy function.

adoption policy function. In principle, one could consistently recover this function nonparametrically; however, limited data concerns lead us to estimate a simple level shifter.[16]

4.5 Estimating Net Fixed Costs of Adoption

Once we have estimated the policy functions governing adoption and use of the videocalling technology, it is possible to estimate the net fixed costs of adoption. Rearranging Equation 7, the necessary and sufficient condition for adoption at time t is:

$$F_i \leq E\left[U(\Omega_{it}) + \beta\left(V_i(s_{t+1}; s_{i,t+1} = 1) - V_i(s_{t+1}; s_{i,t+1} = 0)\right)\right] = \bar{F}_t, \tag{15}$$

where $\bar{F}_t$ denotes the threshold at which a worker is indifferent to adopting the technology at time t. We assume that F_i is normally distributed with mean μ and variance σ^2, with associated cumulative distribution function $\Phi(x; \mu, \sigma^2)$. Therefore, the probability that employee has a draw of F_i low enough to induce adoption is:

$$Pr(adopt_{it}) = \Phi\left(E\left[U(\Omega_{it}) + \beta\left(V_i(s_{t+1}; s_{i,t+1} = 1) - V_i(s_{t+1}; s_{i,t+1} = 0)\right)\right] | F_i \geq \bar{F}_{t-1}; \mu, \sigma^2\right). \tag{16}$$

With exception of the mean and variance parameters of $\Phi(\cdot)$, the terms in Equation 16 are either known or computable. We can calculate $Pr(adopt_{it})$, the empirical probability of adoption, from Equation 14. The first set of policy functions gives an estimate of $U(\Omega_{it})$, the expected calling utility, for any configuration of the network. The second set of policy functions describe how that network evolves over time as a function of current and lagged state variables. In combination, these policy functions allow us to simulate the evolution of the network and compute $EV_i(s_{t+1}; s_{i,t+1} = 1)$, the expected present discounted utility of joining the network in this period.

[16] We have experimented with several different specifications, including ones where the shock and lagged adoption variables are interacted. Unfortunately, data limitations preclude the use of more complex functions of the shock and the state variable. In all specifications we tested, the estimated coefficient on the shock was statistically insignificant.

Computing the expected value of not joining the network in this period is a little more involved. In principle, one needs to solve out an infinite series of nested dynamic programming problems, starting at a time infinitely far in the future and working backward, solving Equation 7 at each point in time. However, suppose there is a month T at which the network has stopped growing. We can compute $EV_i(s_T)$ because adoption depends on whether or not the expected benefits of joining the network exceed F_i. When the network does not grow then employees have no reason to delay adoption until a future period; an employee either adopts now or never adopts. This terminal value then allows us to solve the value function backwards to the current time period, approximating $EV_i(s_t)$. This seems reasonable because most uncertainty about the evolution of the system has been resolved by period 10; we use this time horizon in computing the continuation value of not joining in the current period.

For $t > 0$, the population of potential adopters has a distribution of adoption costs that is truncated below by the $\bar{F}_{t-1}$. To emphasize this, we write the probability in Equation 16 as being conditional on the fixed cost draw for worker i being at least as large as the threshold in the previous period, $\bar{F}_{t-1}$. As discussed above, this is the reason why we estimate the reduced-form adoption policy function as a function of both current and lagged state variables.

To recover the parameters underlying the distributions of net fixed costs, we estimate Equation 16 by forming the following moment:

$$\min_{\{\mu,\sigma^2\}} \sum_M \sum_T \sum_i (Pr(adopt_{itm}; s_t, s_{t-1}) - \Phi\left(\cdot; \mu_m, \sigma_m^2\right)), \tag{17}$$

where $\Phi(\cdot)$ is the conditional probability defined in Equation 16. This is the probability that a given agent of subtype m adopts the technology at time t, given the state of the network in this period and last. We have written the adoption probability as a function of current and lagged state variables to emphasize the fact that the distribution of fixed costs is truncated after the first period. We also index μ_m and σ_m^2 by m to highlight that we estimate the fixed cost distributions separately for each type of employee in the firm.

4.6 Multiple Equilibria

A central challenge for the network effects literature is dealing with multiple equilibria. One advantage of our empirical approach is that we recover the equilibrium actually played in the data. Furthermore since employees make choices within a single network, it follows that only one equilibrium is being played. To our knowledge, this is unique among applications of the BBL framework, as we do not have to confront the possibility of multiple equilibria across markets, as in Ryan (2010).

4.7 Identification

We have two sets of parameters to identify in the model: calling utilities and the distributions of adoption fixed costs. The calling utility parameters in Equation 10 are identified in the cross-section of calls made by agents in a single calling period. These parameters do not depend on the fixed costs of adopting the technology, and can be considered separately. A necessary and sufficient condition for the identification of the connection utilities in Equation 11 is that we observe connections between all different types during a sample period. The corresponding condition for the decay parameters in Equation 13 is that we observe all possible two-call sequences in the data. Both of these conditions are satisfied in our data, as we have a fully saturated call network in the later months of the sample.

Identification of the distributions of fixed cost parameters in Equation 16 follows directly from timing assumptions in the model. We assume that agents pay a setup cost to adopt the technology and then receive a stand-alone use from the technology in perpetuity. As discussed above, these two parameters may be combined into a single net adoption cost parameter at the time of adoption. Given the adoption policy function and calling utilities, we can calculate the expected value of joining the network in any period. A necessary and sufficient condition for identifying the mean and variance of the normally-distributed fixed costs of adoption is that for each subtype the econometrician observes the proportion of agents who adopt for two different states of the network. This condition is satisfied in our sample by the first two periods.

5 Results

This section reports estimates for parameters for the utility from making calls and the distribution of the net fixed costs of adoption that govern the transition between states. We first discuss the fits of the adoption policy function.

5.1 Adoption Policy Function

Table 7 reports statistics for the fit of our first stage regressions for the adoption policy function. Column (1) reports the fit statistics and specification of the regression where we assume that agents behave identically. In this specification there are only four parameters: the total size of the network, the total lagged size of the network and respective squared terms. Recall that we restrict the coefficients on the current state variables to be positive and the coefficients on the lagged variables to be negative, as discussed in Section 4.4 above. This model fits the data fairly poorly, with an R-squared of 0.16. Column (2) reports the fit statistics and specification when we allow adoption patterns to vary across subtypes. This specification results in a dramatically improved fit, with an R-squared of over 0.98.[17]

For robustness purposes, we also examined several specifications of the adoption policy function where we allowed for the network-wide time-specific aggregate utility shock to influence adoption behavior. In the simplest specification, we added xi_t as a level shifter to

[17]We do not report standard errors, as the asymptotic distribution for inequality-constrained estimators with many parameters on the boundary is a frontier area of research. We note that this will bias the standard errors on the distributions of adoption costs downward, as they are functions of the adoption policy function. See Andrews and Guggenberger (2009) for a discussion of possible solutions in the case of one parameter on the boundary, and why intuitive solutions such as the bootstrap are inconsistent.

the second specification. The estimated coefficient was -0.0041, with a standard error of 0.0030, which rendered the estimated coefficient statistically insignificant from zero at the five percent level. We also experimented with more complex specifications, such as allowing the shock to interact with lagged adoption variables in order to allow the effect of the shock to vary with the size of the extant network, but these models did not give statistically significant results. We also estimated the policy function with time-specific dummies directly. An F-test on the resulting coefficients failed to reject the hypothesis that the coefficients were jointly zero. In sum, the empirical evidence suggests that correlated unobservables had a weak effect on the adoption behavior of agents. Economically, this also makes sense, as the adoption of a durable good is a long-run decision and the period-to-period variance in the usage of the technology is relatively small.

We conclude that our model does a good job of fitting the evolution of adopters over time, and that accounting for heterogeneity across subtypes results in a much improved fit over a homogenous adoption model. In order to disentangle whether this is driven by heterogeneity in usage benefit or adoption cost, we next turn to the results of our calling model.

5.2 Call Utilities

We use observations on 463,806 calls from February 2001 to August 2004 to estimate the calling utility parameters in Equation 2. Tables 8 through 11 display the results.

Table 8 illustrates that generally employees prefer to call other employees within their region. The only exception is UK-based employees, who prefer to call other employees from Europe. Given that this within-region propensity is larger for employees in the US and Asia, we speculate that the propensity to call within regions could be influenced by time zones. Employees' work hours in the US and Asia barely overlap, but the work hours of British and European employees do.

Table 9 illustrates that employees prefer to call employees in similar functions to themselves. Connections with administration are also valued relatively highly. It is interesting to note that the lowest connection utility is between trading and research, which may reflect institutional boundaries between these two components of the bank due to potential for conflicts of interest. The highest connection utility is between sales to sales, followed by sales to adminstration. Trading tends to value the network the lowest, and the other groups also tend to value traders on the lower end of connection utilities. This may be due to the special nature of the trading desk, and its orientation towards external markets rather than other operations within the bank.

Most of the theoretical literature on hierarchies and firm organization pose abstract models of why the need to process information may lead a firm to organize itself into a hierarchy. See for example Radner (1992), Radner (1993), Van Zandt (1999) and Garicano (2000). These theories predict that communication in a firm will be predominantly directed up and down a hierarchy. By contrast, our results on calling preferences across the hierarchy in Section 10 suggest a more nuanced pattern of communication. Managing Directors are most likely to call each other and less likely to call employees further down the ladder of command. Other employees appear to have similar preferences for calling other employees in similar positions in the hierarchy or one above them (two steps above them in the case of VPs and managing directors). However, they are less likely to call employees either lower in the hierarchy than they are or in most cases step removed above them in the hierarchy. These results augur against the technology being used successfully for monitoring, but instead suggest that it is being used to exchange information about tasks assigned to one layer of the hierarchy or occasionally gathering information from a superior one rung up in the hierarchy. Across groups, Vice Presidents have the lowest average utility from using the network, followed by Managers. The lowest and highest tiers of the organization both have higher average utility from using the network than the tiers in the middle.

One of the auxiliary aims of this paper is to provide some empirical evidence on communication patterns within a firm. Lack of data has meant that most of the literature on hierarchies and firm organization is theoretical. The estimates presented in tables 10 through 11 have an advantage over existing empirical research on the organization of firms such as Garicano and Hubbard (2007) and Rajan and Wulf (2006), namely that we study and model actual communication flows. This means that we are able to provide evidence on whether hierarchies are fulfilling the communications role assigned to them by theory. The obvious caveat is that to get this level of detail in data (like Baker et al. (1994)), we must study the internal communications using one technology in one firm.

The results for the parameter $\hat{\theta}_3$, which captures the role of the dynamic decay rates, are displayed in Table 11. The taste for diversity is strong.[18] The decay rates are large enough to have a significant effect at the margin of calling the same group twice in a row, especially with respect to employees at the associate level, employees in research, and employees in Asia. We speculate that this reflects the fact that these employees are more on the periphery of the firm and that their roles are more to provide one-time information than engaging in consistent exchange of information.

To assess the overall fit of our model, we have tabulated the predicted number of calls under our estimated model against their empirical counterparts in Table 12. In the interest of clarity, instead of tabulating all possible call proportions for each subtype we break down the calls by region, function, and title. As the table shows, our model does an exceptionally good job of matching the average number of calls between aggregated types. Standard errors on the predictions were calculated using a bootstrap and are generally tight. The uncertainty of the predicted fits is small, and for all groups the difference in the two quantities is not statistically significant. We also ran a likelihood ratio test for the hypothesis that all parameters are jointly equal to zero, and found that it was overwhelmingly rejected at the

[18] The likelihood ratio test for the hypothesis that the decay rates are all jointly zero is overwhelming rejected at the 0.001 level.

0.001 level.

5.3 Net Fixed Costs of Adoption

Tables 13 through 16 display our fixed costs estimates. The costs of adoption vary a great deal across the hierarchy and function in the four regions. Generally, fixed costs decline the higher an employee is in the hierarchy. However, in Europe managing directors actually had the highest net fixed costs of adoption. This may be because the managing directors with the greatest operational responsibilities tended to be located in Europe and as a result the value of their time was high. Generally, employees in sales, trading and research in the US and Asia had the highest net fixed costs of adoption. Administrators had lower net costs of adoption. This may be because such administrators were in peripheral positions and therefore had lower time costs.

Comparing these results with the results for calling choices in Tables 8 through 10, shows that it is not the case that the employees whom most employees preferred to call had the lowest net fixed costs of adoption. Instead, in the case of Managing Directors of Research in Europe, callers received high utilities from calling them, but they also had some of the relatively highest fixed adoption costs.[19]

The scale of the fixed costs is relatively small for most users. In Table 11, we compute the net present value for adopters under the actual diffusion policy used by the firm. We find that the first quartile had a net value of 25, which includes the fixed cost component, compared to average fixed costs which ranged from 0 to 2.5 for most subtypes. Therefore, the fixed costs reduced the level of utility associated with the network by roughly ten to fifteen percent for users at the first quartile of utility. However, for many users, the draw on fixed cost is negative, implying that they would have adopted the technology even without any other users in order to utilize its stand-alone benefits.

[19] These results on decay rates rely on the assumption that the appropriate calling period is a calendar month. We have contrasted the results obtained from using a calendar month with the results of using quarters, weeks, or individual days. We found no evidence that our choice of period influenced our results.

The standard errors on the fixed costs of adoption understate the variance of these estimates. These estimates depend on the adoption policy function, and therefore are less precisely estimated than indicated here. However, for the reasons discussed above, we treat the adoption policy function as perfectly known, and do not correct the second-stage errors.

5.4 Sensitivity of Results to the Discount Factor

To test the sensitivity of our results with respect to our choice of the discount factor, we estimated the model with the discount factor ranging from 0 to 0.992 (roughly corresponding to a yearly discount factor of 0.9). The results for one subtype is given in Table 1. Upon first review, the results are surprising: for most of the range of the discount factor, the net effect on the estimated fixed costs is minimal. This is due to two countervailing forces in the model. The first is that, holding the optimal period of adoption constant for a given draw of F_i, increasing the discount factor increases the value of adoption. This effect pushes the distribution of fixed costs upward, since the adoption rate holds steady as the value of adoption rises, which implies that the fixed costs must be increasing, as well. On the other hand, increasing the value function increases the value of waiting until the future—this is most easily seen in the case where $\beta = 0$, where there is no value in waiting for the future. This has the effect of increasing the opportunity cost of adopting now, which reduces the expected value of current adoption. This results in a decrease in the distribution of fixed costs. The results in Table 1 show that these two forces roughly counterbalance each other over a wide the range of the discount factor, up to 0.8 or so.

After that point, the agents are all sufficiently patient that there are significant opportunity costs to adopting right away. In this range, for this subtype, high adoption rates imply that the average fixed costs are negative; that is, agents highly value the use of the network for its stand-alone uses. It could be also that these agents are just enthusiastic about the idea of having a new technology at their disposal. In any case, increasing the discount factor

Table 1: Fixed Costs by Discount Rate

Discount Factor (β)	**Fixed Cost Mean**	**Fixed Cost Variance**
0.0	-403.57	7.16E7
0.1	-415.98	7.10E7
0.2	-381.76	7.16E7
0.3	-395.81	7.02E7
0.4	-393.44	7.15E7
0.5	-372.47	7.03E7
0.6	-355.19	6.67E7
0.7	-340.30	6.17E7
0.8	-206.02	3.47E7
0.9	-516.36	5.33E7
0.992	-2102.11	6.24E7

towards one increases that opportunity cost, which means the agents must be taking draws from an increasingly favorable (negative) cost distribution.

Economically, it is hard to say what the right discount factor should be. However, it is worth noting that the policy functions for adoption are estimated directly from the data; irrespective of the discount factor we impose on the model, the counterfactual distributions of network evolutions under different seeding policies will remain the same. This is one benefit of using perturbations to the state space, as opposed to counterfactuals where one had to re-solve the model. Therefore, the primary role of the discount factor is to weigh the relative size of adoption costs against the usage utilities. Part of the fundamental identification problem of the discount factor comes into play here—it is impossible to distinguish agents with high fixed costs and discount factors close to one from agents with low fixed costs and low discount factors. Given that our results in the counterfactual are denominated in utility terms, we have no economic criterion to judge the discount factor against in order to determine if it is reasonable. For these reasons, we conclude that our primary findings are not sensitive to the discount factor, in the sense that our results do not depend on it in a direct fashion.

6 Policy Experiments

6.1 Motivation

Carr (2003) documents that the typical company spends 3.7 percent of its revenues on IT. A challenge for managers is to ensure that their employees actually use the firm's technology investment to its full advantage. Therefore, a practical use of any model of IT technology adoption when simulating counterfactuals is examining different policies a firm or policy-maker might employ to encourage the spread and use of the new IT technology.

One noted limitation of our approach is that any policy experiments are limited to perturbations of the state space. This means that we cannot study policies or counterfactuals that change the underlying primitives of the model and thus require re-solving for the equilibrium policy functions. The reason is that we cannot solve for even one of the potential equilibria that agents might play in counterfactual policy regimes, such as would result from giving workers direct subsidies to alleviate their adoption costs. However, it is possible to conduct counterfactual exercises that only affect the state space, as our policy functions are consistent descriptions of equilibrium behavior for all possible states of the network. The key assumption that we make in the simulations below is that the same equilibrium is played in these counterfactuals as was played in the original network evolution. Fortunately, our setting is well-suited to the limitations of a BBL approach, in that perturbations of the initial state space are of crucial interest to firms and policy-makers who want to seed a network optimally.

As discussed by Liebowitz and Margolis (1994), network owners can prevent coordination failure if they induce enough initial participation. In the presence of network effects which are heterogeneous in interactions, however, the question of which participants should be induced to adopt the technology initially is more complex. It is not clear whether a network operator should focus inducements on users who are unlikely to adopt the network and whose adoption therefore may most shift expectations of its evolution, or whether the network operator should target users who have the largest network benefits for others. It is also not clear whether the network operator should strive to have diversity amongst its initial users or whether it should target a specific group.

Therefore, we evaluate several permutations of two possible “rule of thumb” technology management policies: a targeted policy where agents with some characteristics in common are seeded into the network, and a uniform policy where a few employees from every subtype join the network. These policies assume that the firm will install the physical hardware and provide whatever training is necessary to overcome the net fixed costs of adoption for a selected set of employees under each policy.

The first policy we consider is where the firm picks one subtype to adopt/test the technology first. This resembles the way that many firms roll out new IT technologies. IT managers usually pick this initial seed from employees who are similar by virtue of their function and location. Therefore we conduct a policy experiment where the starting network is seeded with all 112 research associates located in the US. This group constitutes the single largest subtype within the firm, and may be considered a natural place to seed the network, as employees in the US generally have high adoption costs.

The second policy takes a diffuse approach to adoption. Here the firm spreads 112 installations across the entire set of subtypes. The idea here is that diversity increases the value of the network, and that seeding the initial network with a broad range of types may most efficiently jump-start the growth of the network. Given there are 64 subtypes for 112 installations, there will be 16 groups which start with only one employee. We choose these groups to be the ones located in the US. Since adoption was decentralized in our context, we interpret “adoption” in our data as the equivalent of the more general idea of “activation,” the active usage of a new technology by an employee.

We should note though, that while the counterfactuals are of managerial interest they do involve perturbations of the state space to states that were not visited in the data. Consequently, when assuming the policy functions apply in these previously unvisited states we are relying heavily on the parametric assumptions implied in our estimating equations.

6.2 Implementation

In each counterfactual simulation, we start by seeding the initial network in accordance with the desired policy. Starting at time zero, the network is then simulated forward for thirty-six months. This amount of time is sufficient to allow the network to achieve a steady state where it is no longer growing. Also, the discounted present value of utility of months more than 36 periods from the first period is very small for the discount rate of 0.9 that we use. To simulate the evolution of the network, we draw uniform random variables for each potential adopter, and check these against each employee's corresponding subtype-specific policy function. If the policy function indicates that the employee will join the network, we draw a sunk entry cost from the associated truncated normal distribution in Equation 16. After determining the evolution of the network in that period, we then calculate the sum of expected utilities for all employees in the network. This calculation is simplified by the fact that it is possible to do this on a subtype basis, rather than employee by employee. The results of the two policy experiments and a baseline comparison against the empty starting network are shown in Table 17. Figures 6, 7 and 8 contrast the results graphically for the total adoption, calls and average utility. We compute standard errors by simulating the outcomes for each counterfactual 100 times.[20]

Figure 6 plots the size of the network over time. Adoption is initially very rapid before tapering off. The plot reveals that the targeted policy results in a significantly larger long-run network size, with a slightly smaller network under the uniform policy and a much smaller network under the baseline case. The difference between the uniform and the targeted

[20] As noted above, we do not account for statistical error in the adoption policy function. The standard errors capture variation in outcomes due to statistical error the estimated calling parameters and structural errors in the calling functions.

Figure 6: Counterfactual Adoption

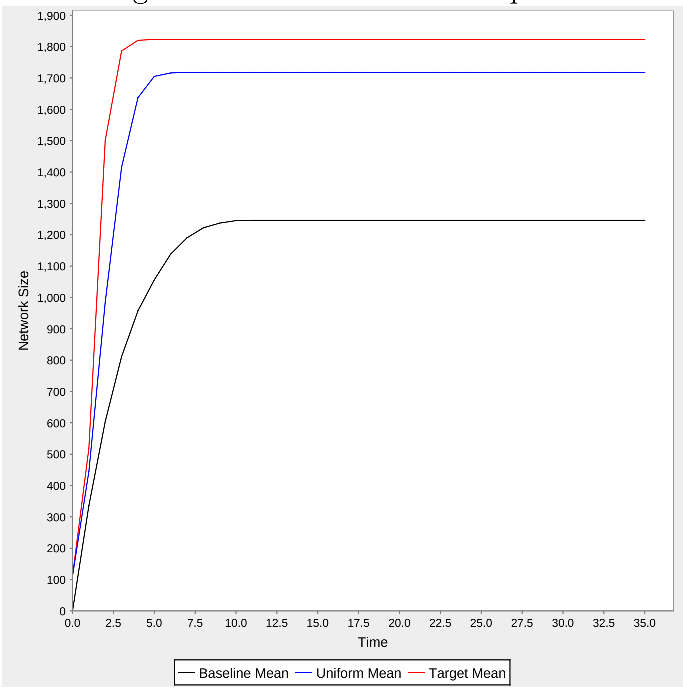


Figure 7: Counterfactual Calls

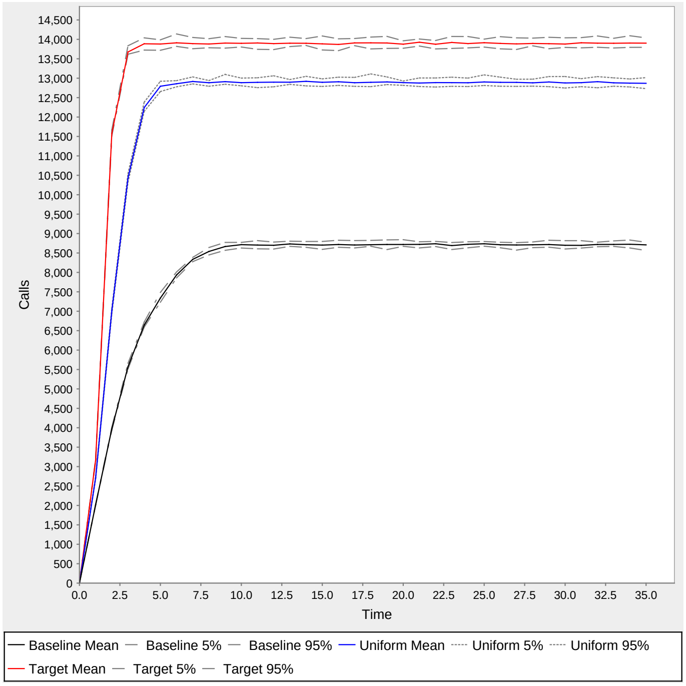
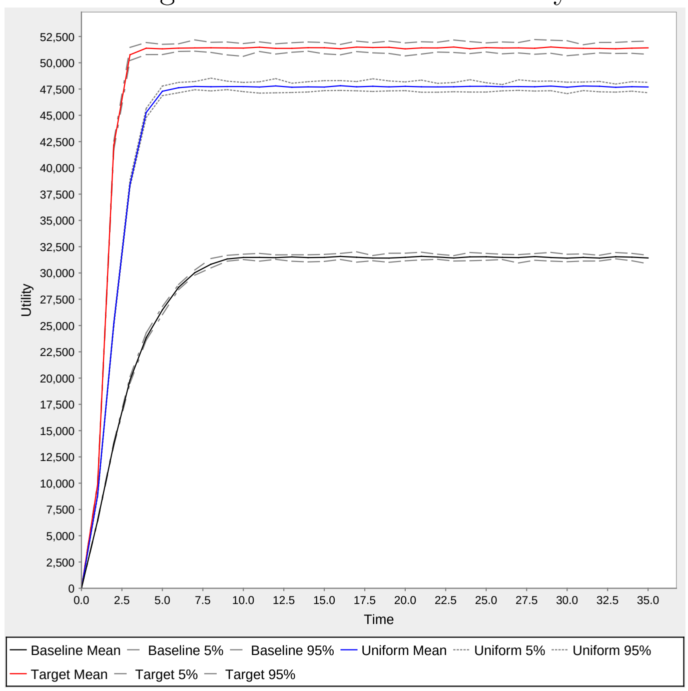


Figure 8: Counterfactual Utility



policy is interesting because they both show rapid initial gains, but the targeted policy has significantly higher adoption in the few periods after the initial seeding, and that this difference persists in the long run. Both the targeted and uniform policy show a slight S-shape in the adoption rate. This is the classical shape of diffusion, as initial adopters cause a cascade of secondary adoption in later periods. Over time, the remaining potential adopters are increasingly a selected sample, and the adoption rate eventually tails off to zero after the fifth month in both policies. This is contrast to the uncoordinated policy, which has much slower growth that continues until the tenth month. This graphical evidence suggests that a uniform policy may not be as beneficial as the firm would hope for.

A similar pattern is reflected in Figure 7, which shows the number of calls per period in the three simulations. The dashed lines indicate the 90 percent confidence intervals for each series. While the targeted and uniform policies look similar in the first few periods, the targeted policy dominates the uniform policy, driven by the higher adoption rates. The plot of per-period utilities, shown in Figure 8, also illustrates this general pattern.

The first line of Table 17 compares the average number of monthly phone calls per user across the three policies. Across each specification, the undiscounted average number of calls in each month is roughly similar, with slightly higher amounts in the baseline and targeted policy than in the uniform policy. As a check on the model, it is reassuring that the average number of calls in the baseline case (7.14) compares favorably to the true monthly average in the data (7.00).

The next line compares the maximal size of the network across the three policies. The baseline case is very close to the actual number of adopters used in our analysis (1,300). The maximum number of adopters is considerably higher in the targeted case than in the baseline or uniform cases. This occurs because other employees find it valuable when this group adopts, but the group's high net fixed costs of adoption prevent them from adopting in the baseline case. Consequently, targeting this group has a large effect on employees' adoption both directly and because it changes expectations about how the network will evolve. The results for the uniform policy suggest that a broad-based adoption process may be inefficient, since it does not target employees with high net fixed costs of adoption and consequently does not alter expectations of how the network will evolve.

We calculate the expected discounted monthly utility for each subtype across the three policies. Both the targeted and uniform policies result in higher levels of utility across various quartiles of the distribution of utilities than the baseline case. This is due to the fact that there are more adopters in both of these networks, and therefore there are more opportunities for positive utility network interactions. The utility for employees which results from the targeted policy is higher than the baseline case, and moderately higher than the uniform case. Targeting increases present discounted utility by over 11% (discounted at $\beta = 0.9$) for the median type relative to the baseline. This increase is also reflected across the other quartiles of the utility distribution. If the objectives of the firm are positively correlated with the utility of the employees, then this policy results in a significantly positive effect from the firm's perspective. In addition the utility gains appear to shift the utilities equally across subtypes in the firm, even in the targeted case. This illustrates that in this setting, the utility benefits of changing the number of people in the network by targeting those with high net fixed costs outweigh trying to encourage diversity in the network.

Figure 8 shows an interesting feature of our model. Initially, the per-period utility is high, as the earliest adopters of the network have a combination of the highest connection utilities and low adoption costs. This large amount of surplus decreases in subsequent periods, particularly in the targeted policy, as more marginal adopters pay fixed costs to join the network. The combination of these two factors leads to the marginal per-period utility decreasing over a short time frame in the early stages of the network.

The last two panels in the table illustrate inter-temporal differences in adoption rates and network usage. We assume that, everything else being equal, the firm would prefer to have a given number of phone calls or employees in the network sooner rather than later. We report the discounted sums of users who have adopted the network in a month and the number of calls they have made, using two contrasting potential monthly discount rates for the firm. The differences are quite stark: the uniform policy makes marginal improvements over the baseline case, while the targeted policy dominates along both dimensions. When $\beta = 0.9$, user counts in the targeted policy increased by 65 percent and calls increased by more than 79 percent over the baseline case. These gains are significant even compared to the uniform policy; the targeted policy had 11 percent more adopters and 13.7 percent more calls. The differences are similar under the discount rate of $\beta = 0.99$.

In an investment bank where the opportunity cost of time is high, these results suggest that there are significant gains to be made from targeting the right group of agents in the early stages of a network technology diffusion. Our results also highlight that the policymaker has to carefully weigh the tradeoffs between adoption costs and network benefits, as our uniform policy results demonstrate that the overall network utility could be lower than the baseline case when non-discriminatory policies are used. We have assumed throughout this discussion that it is costless to the firm, although not the employee, to place employees in the installed base in the first period, and that such policies can be accomplished by fiat without any compensation towards the employees concerned. If employees had to be compensated, the targeted policy might prove more expensive to the firm relative to more diffuse policies merely because of the higher net fixed costs involved in the targeted policy.

These results suggest several other policy experiments. One natural question to ask is the extend to which the specific subtype influences long-run network evolution and use. To answer this, we ran several counterfactuals where we seeded different groups systematically. We took all subtypes with 20 or more workers and seeded 20 of that type in the initial network. The results of this experiment are reported in Table 2. For clarity, we focus on network utility as the outcome of interest. As the results show, the highest network utility is achieved by seeding 20 Director-level workers in Sales in the US. The difference across seeding subtypes is substantial; the best utility is a little under two-thirds better than the worst possible seeding. Interestingly, Sales Directors in the US are high value adoptees to the rest of the network, as evidenced by the high connection utilities for calling a Director, yet these agents have relative low fixed costs. The key is that this subtype induces large amounts of adoption relative to other subtypes, and this drives their high utility values.

A second interesting topic to explore is the role the number of agents seeded into the network has on long-run network growth. First, we varied the number of initial adopters from zero to 100, uniformly distributing the seeded workers throughout the firm. The utilities of the resulting networks are presented in Table 3. Overall network utility grows with the number of workers initially seeded into the network, but the growth is rather low. At 65 initial adopters every subtype has at least one seeded adopter, which results in an increase in utility over the baseline of over 100,000 utils, or over 40 percent. This is a significant

Table 2: Network Utilities One Year After Seeding 20 Workers

Region	Title	Function	Network Utility
Asia	Vice President	Administration	266,849.80
Asia	Vice President	Research	257,080.72
Asia	Vice President	Sales	253,001.53
Asia	Director	Administration	309,967.69
Asia	Director	Research	298,555.90
Asia	Director	Sales	299,426.68
Asia	Managing Director	Administration	273,550.49
Asia	Managing Director	Research	252,785.45
Asia	Managing Director	Sales	254,716.33
United Kingdom	Associate	Research	260,160.79
United Kingdom	Vice President	Administration	254,416.56
United Kingdom	Vice President	Research	250,206.44
United Kingdom	Vice President	Sales	250,113.94
United Kingdom	Vice President	Trading	275,821.72
United Kingdom	Director	Administration	309,138.89
United Kingdom	Director	Research	343,376.38
United Kingdom	Director	Sales	340,110.65
United Kingdom	Director	Trading	321,198.94
United Kingdom	Managing Director	Administration	253,721.39
United Kingdom	Managing Director	Research	250,060.15
United Kingdom	Managing Director	Sales	250,212.80
United Kingdom	Managing Director	Trading	275,887.08
Europe	Vice President	Research	250,191.21
Europe	Director	Administration	299,534.88
Europe	Director	Research	299,965.80
Europe	Director	Sales	299,618.54
Europe	Managing Director	Administration	253,847.37
Europe	Managing Director	Research	249,561.62
USA	Associate	Administration	345,669.57
USA	Associate	Research	344,246.34
USA	Associate	Sales	341,537.41
USA	Associate	Trading	351,176.05
USA	Vice President	Administration	345,530.20
USA	Vice President	Research	334,717.58
USA	Vice President	Sales	348,860.19
USA	Director	Administration	366,138.38
USA	Director	Research	370,563.93
USA	Director	Sales	399,592.45
USA	Managing Director	Administration	333,723.75
USA	Managing Director	Research	339,190.16
USA	Managing Director	Sales	347,054.78
USA	Managing Director	Trading	346,596.58

increase, although one would suspect that targeting a particular group may increase that gain due to the fact that some subtypes are inherently more valuable to potential adopters than others.

To evaluate this hypothesis, we ran a second seeding experiment where we put increasing numbers of the most valuable subtype, Directors of Sales in the US, in the network, as seen in Table 4. The results are dramatically larger—the utility achieved with a uniform seeding of 100 agents is surpassed by the targeted policy with less than 20 workers of the this subtype. The utilities also show a pronounced nonlinearity: initially, the first five workers in the network have a large effect, increasing utility by approximately 44,000 utils. However, the second five increases utility by even more, raising marginal utility by approximately 49,000 utils. After that inflection point, utility increases at a decreasing rate, finally tailing off to a marginal increase of 5,000 utils between 50 and 55. These results illustrate both that the firm must be careful not to seed the network with too few agents at the beginning in order to achieve optimal network growth, and that choosing the right group to start the technology off with can have a profound effect on the value of the network.

One drawback to the analysis above is that the central planner may not know the ex-ante benefits across different subtypes in the firm. In lieu of knowing ex-ante which subtype would the most valuable to enabling the growth of the network, the firm may have to rely on less nuanced guesses about the best groups to seed the initial network with. For example, the firm may put all the workers of a certain region in the network initially. Regional seeding makes sense in that workers will all roughly be working at similar times, and thus is a natural place to start. In this vein, we ran seeding experiments where we seeded the network with all agents in a given region of the firm. We report the results for the US in Tables 5 and 6. To compare with the above results, we first seeded the network with 0 to 55 workers from the US in increments of 5 workers. We then repeated the experiment by seeding the network with 0 to 669 (the total number of workers in the US) in increments of 100 workers

Table 3: Network Utilities by Number of Uniformly Seeded Adopters

Initial Number	Network Utility
0	248,292.04
5	267,612.69
10	291,207.33
15	309,452.59
20	314,628.03
25	321,564.61
30	322,039.49
35	322,277.12
40	330,762.62
45	331,969.05
50	332,662.52
55	341,864.82
60	343,698.41
65	350,379.03
70	368,178.25
75	379,340.18
80	388,578.02
85	390,850.76
90	392,701.73
95	394,559.95
100	396,919.76

Table 4: Network Utilities by Seeded Number of Directors of Sales in the US

Initial Number	Network Utility
0	248,292.04
5	292,422.14
10	341,151.65
15	382,536.97
20	415,631.26
25	440,257.31
30	461,321.33
35	477,435.77
40	491,312.54
45	503,497.68
50	515,638.63
55	520,548.49

Table 5: Network Utilities by Seeded US Workers

Initial Number	Network Utility
0	248,292.05
5	267,612.69
10	291,207.34
15	309,452.60
20	326,928.05
25	348,192.93
30	360,276.44
35	370,568.54
40	385,642.11
45	394,318.59
50	401,622.94
55	411,227.65

to examine larger-scale effects.

Comparing the results from the first 55 workers to the Directors of Sales in the US suggest that the uniform adoption policy has a substantially lower overall level of network utility. There is an 9.2 percent difference even at 5 workers, which increases to 26 percent by 55 workers. However, the policy still does well compared to the global case. Note that the results between the firm-wide uniform seeding and the US-only uniform seeding are identical up through the first 15 workers, since the uniform seeding case starts with the US workers first when filling out the network. Once the two networks diverge, the gains from a US-only policy are significant: the gains are more than 20 percent by the 50th worker (which is only the 34th different worker in the two networks). While the gap narrows by the 100th worker, the difference is still over 9 percent.

In sum, these results suggest that concentrating resources even at a larger aggregate level can still have significant gains on diffuse adoption policies. While we have emphasized differences across seeding policies, we also note that all of the policies that we have considered lead to much higher network utilities than the laissez-faire policy that the firm actually followed. Even small network seedings lead to substantial increases in network adoption and use. Our results suggest that the firm missed a significant opportunity to encourage growth

Table 6: Network Utilities by Seeded US Workers

Initial Number	Network Utility
0	248,292.05
100	435,267.83
200	459,552.95
300	476,613.84
400	490,796.13
500	503,783.75
600	512,795.48
669	516,573.11

and use of its videocalling network by following a hands-off approach.

7 Conclusion

This paper synthesizes an older literature that explained diffusion patterns by user heterogeneity, and a newer literature on network effects that emphasizes the interdependence of technology adoption. We estimate a model of forward-looking technology adoption and the subsequent sequence of usage that explicitly models heterogeneity over adoption costs, network effects and usage behavior. We estimate this model using data on 463,806 calls made after the introduction of a videocalling technology in a large investment bank. We quantify how different types of heterogeneity affect network evolution and use, and analyze several variants of two common policies which are used to jump-start network technology diffusion. Our results strongly favor a policy of targeted seeding, especially if the firm can identify a subgroup in the firm that has high network benefits to other potential adopters, even if they do not have high fixed costs of adoption. The effects of properly-targeted seeding are large: seeding the network by just five people in the beginning can have as large as a 17 percent increase in overall network utility compared to a policy of hands-off diffusion. However, there is large heterogeneity in the effectiveness of different targeted seeding campaigns: The best performing targeted campaign outperforms the worse performing campaign by two-thirds. This suggests that an optimal strategy for a firm in such an environment may be to expend resources identifying which groups in the firm are high value adopters, as the long-run payoff to such investments can be large.

References

Ackerberg, D. A. and G. Gowrisankaran (2006, Autumn). Quantifying equilibrium network externalities in the ACH banking industry. *RAND Journal of Economics 37*, 738–61.

Andrews, D. and P. Guggenberger (2009). Asymptotic size and a problem with subsampling and with the m out of n bootstrap. *Econometric Theory 25*.

Bajari, P., C. L. Benkard, and J. Levin (2007). Estimating Dynamic Models of Imperfect Competition. *Econometrica 75*, 1331–70.

Bajari, P., H. Hong, and S. Ryan (2009). Identification and estimation of a normal form game. University of Minnesota Working Paper.

Baker, G., M. Gibbs, and B. Holmstrom (1994). The internal economics of the firm: Evidence from personnel data. *The Quarterly Journal of Economics 109*(4), 881–919.

Carr, N. G. (2003). It doesn't matter. *Harvard Business Review 81*(5), 41–49.

Draganska, M., M. Mazzeo, and K. Seim (2009, June). Beyond plain vanilla: Modeling joint product assortment and pricing decisions. *Quantitative Marketing and Economics 7*(2), 105–146.

Dube, J.-P. H., G. J. Hitsch, and P. K. Chintagunta (2010). Tipping and Concentration in Markets with Indirect Network Effects. *Marketing Science 29*(2), 216–249.

Ellickson, P. and S. Misra (2010). Enriching interactions: Incorporating revenue and cost data into static discrete games. *Mimeo, University of Rochester*.

Farrell, J. and G. Saloner (1985). Standardization, compatibility, and innovation. *RAND Journal of Economics 16*, 70–83.

Garicano, L. (2000). Hierarchies and the organization of knowledge in production. *Journal of Political Economy 108*(5), 874–904.

Garicano, L. and T. N. Hubbard (2007, February). Managerial leverage is limited by the extent of the market: Hierarchies, specialization, and the utilization of lawyers' human capital. *Journal of Law & Economics 50*(1), 1–43.

Gordon, B. R. (2009). A Dynamic Model of Consumer Replacement Cycles in the PC Processor Industry. *Marketing Science*, mksc.1080.0448.

Gowrisankaran, G. and J. Stavins (2004). Network externalities and technology adoption: lessons from electronic payments. *RAND Journal of Economics 35*(2), 260–276.

Griliches, Z. (1957). Hybrid corn: An exploration in the economics of technological change. *Econometrica 25*(4), 501–522.

Hartmann, W. R. and H. S. Nair (2010). Retail Competition and the Dynamics of Demand for Tied Goods. *Marketing Science 29*(2), 366–386.

Jackson, M. and A. Wolinsky (1996). A Strategic Model of Social and Economic Networks. *Journal of Economic Theory 71*, 44–74.

Katz, M. L. and C. Shapiro (1985). Network externalities, competition, and compatibility. *American Economic Review 75*(3), 424–40.

Keane, M. P. and K. I. Wolpin (1997, June). The career decisions of young men. *Journal of Political Economy 105*(3), 473–522.

Liebowitz, S. and S. E. Margolis (1994, Spring). Network externality: An uncommon tragedy. *Journal of Economic Perspectives 8*(2), 133–150.

Mansfield, E. (1961, October). Technical change and the rate of imitation. *Econometrica 29*(4), 741–66.

Maskin, E. and J. Tirole (1988, May). A theory of dynamic oligopoly, i: Overview and quantity competition with large fixed costs. *Econometrica 56*(3), 549–69.

Maskin, E. and J. Tirole (2001, October). Markov perfect equilibrium: I. observable actions. *Journal of Economic Theory 100*(2), 191–219.

Misra, S. and H. Nair (2009). A structural model of sales-force compensation dynamics: Estimation and field implementation. Mimeo, Stanford.

Nair, H., P. Chintagunta, and J.-P. Dub (2004, March). Empirical analysis of indirect network effects in the market for personal digital assistants. *Quantitative Marketing and Economics 2*(1), 23–58.

Radner, R. (1992, September). Hierarchy: The economics of managing. *Journal of Economic Literature 30*, 1382–1415.

Radner, R. (1993, September). The organization of decentralized information processing. *Econometrica 61*, 1109–1146.

Rajan, R. and J. Wulf (2006, November). The flattening firm: Evidence from panel data on the changing nature of corporate hierarchies. *Review of Economics and Statistics 88*(4), 759–773.

Reiss, P. C. and P. T. Spiller (1989, October). Competition and entry in small airline markets. *Journal of Law & Economics 32*(2), S179–202.

Rogers, E. M. (1962, February). *Diffusion of Innovations.* Free Press.

Rose, N. L. and P. L. Joskow (1990, Autumn). The diffusion of new technologies: Evidence from the electric utility industry. *RAND Journal of Economics 21*(3), 354–373.

Ryan, S. P. (2010). The Costs of Environmental Regulation in a Concentrated Industry. MIT Working Paper.

Rysman, M. (2004, April). Competition between networks: A study of the market for yellow pages. *Review of Economic Studies 71*(2), 483–512.

Tucker, C. (2008, December). Identifying formal and informal influence in technology adoption with network externalities. *Management Science 54*(12), 2024–2039.

Tucker, C. (2009, January). Network stability and network effects. Mimeo, MIT.

Van Zandt, T. (1999). Real-time decentralized information processing as a model of organizations with boundedly rational agents. *Review of Economic Studies 66*(3), 633–58.

Table 7: First Stage Policy Function Specification and Fit

	Homogeneity Specification	Heterogeneity Specification
Lagged installed base variables	Yes	Yes
Group Dummies	No	Yes
Installed base squared terms	Yes	Yes
Separate Installed Base terms for each group	No	Yes
Observations	2709	2709
R-Squared	0.164	0.986
Adjusted R-Squared	0.163	0.986

We restricted the signs to be positive on current installed base and negative on lagged installed base; 74 of the 96 coefficients had binding constraints.

Table 8: Static Interactions of Caller and Receiver Regions on Calling Choice

Variable	**Mean**	**StdDev**
Asia to UK	-0.8459	0.0462
Asia to Europe	-0.5961	0.0472
Asia to USA	-1.4801	0.0525
UK to Asia	-0.7401	0.0517
UK to UK	0.2476	0.0147
UK to Europe	0.7176	0.0202
UK to USA	-0.3558	0.0269
Europe to Asia	-1.6014	0.0544
Europe to UK	-0.1944	0.0155
Europe to Europe	0.3768	0.0159
Europe to USA	-1.9384	0.0408
USA to Asia	-3.4672	0.0542
USA to UK	-2.7070	0.0358
USA to Europe	-3.3635	0.0602
USA to USA	-2.2053	0.0389

Table 9: Static Interactions of Caller and Receiver Functions on Calling Choice

Variable	**Mean**	**StdDev**
Administration to Research	-1.7410	0.0435
Administration to Sales	-1.1825	0.0242
Administration to Trading	-1.4164	0.0400
Research to Administration	-0.4711	0.0853
Research to Research	-0.0308	0.0584
Research to Sales	-0.6255	0.0686
Research to Trading	-1.2338	0.0811
Sales to Administration	0.5644	0.0289
Sales to Research	-0.2044	0.0350
Sales to Sales	0.5580	0.0236
Sales to Trading	-0.0444	0.0378
Trading to Administration	-0.9153	0.0485
Trading to Research	-2.0034	0.0724
Trading to Sales	-1.1906	0.0526
Trading to Trading	-0.0858	0.0558

Table 10: Static Interactions of Caller and Receiver Titles on Calling Choice

Variable	**Mean**	**StdDev**
Associate to VP	0.2185	0.0602
Associate to Director	0.1600	0.0492
Associate to Managing Director	0.1445	0.0519
Vice President to Associate	-2.3617	0.1029
Vice President to VP	-2.1242	0.0999
Vice President to Director	-2.1704	0.1035
Vice President to Managing Director	-1.9089	0.0982
Director to Associate	-1.2756	0.0444
Director to VP	-1.1918	0.0246
Director to Director	-0.7578	0.0180
Director to Managing Director	-0.3251	0.0179
Managing Director to Associate	-0.6432	0.0501
Managing Director to VP	-0.2667	0.0381
Managing Director to Director	0.5131	0.0296
Managing Director to Managing Director	1.4038	0.0234

Table 11: Decay Rates by Receiver Characteristic

Variable	**Mean**	**StdDev**
Intercept	0.3936	0.0254
N	-0.9043	0.0048
decay Asia	-0.2393	0.0049
decay UK	-0.0549	0.0016
decay Europe	-0.0641	0.0010
decay USA	-0.1111	0.0003
decay Admin	-0.1211	0.0036
decay Research	-0.1220	0.0025
decay Sales	-0.0663	0.0014
decay Trading	-0.0814	0.0004
decay Associate	-0.2559	0.0070
decay Vice President	-0.0884	0.0015
decay Director	-0.0473	0.0004
decay Managing Director	-0.0509	0.0005

Table 12: Actual and Predicted Calling Patterns

Sender	Receiver	Actual	Predicted	Std.Dev.
Asia	Asia	806	799.39	(19.59)
Asia	UK	749	742.88	(20.40)
Asia	Europe	378	375.59	(17.05)
Asia	USA	277	292.14	(15.58)
UK	Asia	650	636.87	(26.05)
UK	UK	6783	6841.66	(55.16)
UK	Europe	5757	5753.12	(49.56)
UK	USA	2564	2522.35	(40.92)
Europe	Asia	367	344.76	(16.17)
Europe	UK	6923	6938.76	(49.79)
Europe	Europe	5967	5981.19	(51.07)
Europe	USA	476	468.29	(19.04)
USA	Asia	259	249.14	(16.58)
USA	UK	2128	2117.50	(36.68)
USA	Europe	405	427.36	(19.62)
USA	USA	4149	4147	(37.43)
Admin	Admin	3836	3813.96	(37.55)
Admin	Research	672	661.90	(19.65)
Admin	Sales	1735	1758.11	(33.96)
Admin	Trading	1026	1035.03	(28.10)
Research	Admin	712	729.95	(25.50)
Research	Research	3512	3508.63	(40.72)
Research	Sales	1617	1619.55	(34.82)
Research	Trading	610	592.86	(18.94)
Sales	Admin	1752	1728.76	(36.06)
Sales	Research	2096	2085.94	(39.32)
Sales	Sales	7278	7299.37	(50.07)
Sales	Trading	2334	2345.93	(40.16)
Trading	Admin	732	734.26	(23.92)
Trading	Research	714	694.28	(23.07)
Trading	Sales	1781	1852.29	(39.63)
Trading	Trading	8231	8177.18	(44.56)
Associate	Associate	304	315.25	(15.30)
Associate	Vice President	958	953.96	(23.95)
Associate	Director	725	716.83	(22.48)
Associate	Managing Director	395	395.96	(17.82)
Vice President	Associate	785	807.74	(30.06)
Vice President	Vice President	3439	3442.31	(43.58)
Vice President	Director	3490	3467.91	(46.41)
Vice President	Managing Director	1837	1833.04	(38.66)
Director	Associate	708	723.07	(25.50)
Director	Vice President	2575	2546.50	(41.69)
Director	Director	4320	4281	(56.37)
Director	Managing Director	4426	4478.43	(50.24)
Managing Director	Associate	419	399.97	(17.72)
Managing Director	Vice President	1940	1949.13	(37.57)
Managing Director	Director	4853	4816.82	(53.49)
Managing Director	Managing Director	7464	7510.08	(53.70)

Table 13: Fixed Costs by Function and Title for Asia

Subtype	Mean	StdDev	Variance	StdDev
Administration				
Associate	0.310	0.043	1.255	0.002
Vice President	-0.677	0.044	0.970	0.023
Director	0.693	0.038	1.074	0.009
Managing Director	-0.006	0.042	1.218	0.011
Research				
Associate	2.221	0.017	0.535	0.021
Vice President	2.193	0.016	0.558	0.020
Director	1.576	0.024	1.072	0.016
Managing Director	0.727	0.015	1.061	0.003
Sales				
Associate	1.868	0.021	0.873	0.017
Vice President	1.737	0.019	0.964	0.014
Director	0.963	0.016	1.009	0.004
Managing Director	0.190	0.017	1.248	0.003
Trading				
Associate	2.079	0.019	0.680	0.020
Vice President	1.533	0.019	1.098	0.010
Director	1.188	0.018	0.967	0.003
Managing Director	0.550	0.014	1.132	0.049

Table 14: Fixed Costs by Function and Title for United Kingdom

Subtype	Mean	StdDev	Variance	StdDev
Administration				
Associate	1.306	0.054	0.953	0.024
Vice President	0.629	0.041	1.093	0.024
Director	0.739	0.038	1.062	0.009
Managing Director	-0.400	0.046	1.094	0.021
Research				
Associate	1.772	0.020	0.945	0.016
Vice President	0.699	0.013	1.070	0.003
Director	0.387	0.017	1.248	0.002
Managing Director	-0.003	0.017	1.203	0.007
Sales				
Associate	1.332	0.015	0.943	0.003
Vice President	0.208	0.016	1.250	0.003
Director	0.530	0.016	1.191	0.057
Managing Director	0.151	0.015	1.245	0.004
Trading				
Associate	1.221	0.016	0.963	0.003
Vice President	0.585	0.013	1.100	0.004
Director	0.537	0.018	1.171	0.061
Managing Director	-0.137	0.017	1.176	0.007

Table 15: Fixed Costs by Function and Title for Europe

Subtype	Mean	StdDev	Variance	StdDev
Administration				
Associate	0.810	0.036	1.043	0.008
Vice President	0.862	0.042	1.032	0.010
Director	0.480	0.047	1.226	0.041
Managing Director	2.540	0.007	0.128	0.001
Research				
Associate	1.543	0.020	1.092	0.012
Vice President	1.113	0.019	0.980	0.003
Director	0.459	0.021	1.239	0.003
Managing Director	0.242	0.019	1.243	0.004
Sales				
Associate	1.824	0.023	0.901	0.018
Vice President	1.080	0.015	0.984	0.003
Director	0.591	0.015	1.097	0.003
Managing Director	-0.031	0.017	1.205	0.007
Trading				
Associate	1.612	0.020	1.047	0.014
Vice President	1.119	0.018	0.978	0.003
Director	0.959	0.016	1.010	0.004
Managing Director	-0.088	0.019	1.194	0.008

Table 16: Fixed Costs by Function and Title for USA

Subtype	Mean	StdDev	Variance	StdDev
Administration				
Associate	1.376	0.056	0.964	0.055
Vice President	0.969	0.034	1.007	0.008
Director	0.692	0.033	1.074	0.008
Managing Director	-0.042	0.037	1.208	0.011
Research				
Associate	2.348	0.014	0.394	0.017
Vice President	1.270	0.011	0.958	0.002
Director	0.648	0.011	1.084	0.003
Managing Director	0.169	0.012	1.237	0.003
Sales				
Associate	1.724	0.011	0.975	0.008
Vice President	1.334	0.009	0.944	0.002
Director	1.014	0.011	0.997	0.002
Managing Director	0.693	0.007	1.072	0.002
Trading				
Associate	2.191	0.013	0.553	0.017
Vice President	2.002	0.015	0.759	0.015
Director	1.708	0.016	0.984	0.011
Managing Director	0.853	0.009	1.033	0.002

Table 17: Policy Experiment Results

Variable	**Baseline**	**Targeted**	**Uniform**
Average Number of Calls	6.97	7.59	7.46
	(0.0317)	(0.0368)	(0.0319)
Maximum number of Adopters	1,246	1,823	1,718
Present Discounted Value utility (mean over all subtypes)	25.31	28.46	27.39
	(0.140)	(0.169)	(0.149)
Present Discounted Value utility (median subtype)	24.95	27.75	27.09
	(0.196)	(0.173)	(0.175)
Present Discounted Value utility (25% subtype)	16.23	18.57	18.07
	(0.136)	(0.181)	(0.191)
Present Discounted Value utility (75% subtype)	34.47	38.67	36.70
	(0.516)	(0.291)	(0.189)
Discounted Value to Firm with $\beta = 0.9$			
Present Discounted Monthly Users	8,877	14,664	13,164
Present Discounted Calls	61,441	110,225	96,917
	(275)	(538)	(391)
Discounted Value to Firm with $\beta = 0.99$			
Present Discounted Monthly Users	33,985	51,987	48,185
Present Discounted Calls	236,914	394,866	359,519
	(1,078)	(1,917)	(1,523)

The final number of adopters is the data is 1,300, while the average number of actual calls per month is 7.00.

A Appendix

A.1 Simple Model of Technology Adoption

Consider the following simple model of technology adoption. There are two periods, $t = \{1, 2\}$. Suppose that there are two agents, each of whom receives a draw of adoption costs from a common distribution, F. If both agents have joined the network, they call each other and obtain utility U per period. At the beginning of each period, each agent can join the network and make a call to the other agent if they are also in the network. Assume there is no discounting. Let $s_i = 1$ if agent i has adopted the technology and zero otherwise. Denote the probability that agent i joins at time t by P_{it}.

In the second period, there are three possible configurations of agent adoption carried over from the first period: neither has adopted, one has adopted, and both have adopted. If both have adopted, then the agents call each other and obtain a payoff of U each. If one has adopted, consider the choice problem facing the other agent:

$$\max \{U - F_i, 0\}$$

The probability that the agent will adopt in the second period conditional on the other agent adopting in the first period is given by $P_{i2}(s_j = 1)$. The distribution that agent i projects over the agent j's fixed costs associated with this probability is different than F, since the agent j already revealed information about their draw on F by not adopting in the first period.

If neither agent has adopted in the first period, the choice problem facing each agent is:

$$\max \{P_{j2}U - F_i, 0\} .$$

The probability that agent i adopts in the second period conditional on the other agent adopting in the first period is given by $P_{i2}(s_j = 0)$. The distribution associated with this probability is also different than F for the same reason as above.

In the first period, each agent must decide whether to adopt or wait. If the agent adopts, the expected payoff is given by:

$$A = P_{j1}2U + (1 - P_{j1})P_{j2}(s_i = 1)U - F_i.$$

The first term is the value accruing to the agent from the possibility of the other agent adopting early, and receiving $2U$ as a result. The second term is the possibility that the other adopts j only in the second period, with a payout of U.

If the agent waits, the payoff is given by:

$$B = P_{j1} \max \{U - F_i, 0\} + (1 - P_{j1}) \max \{P_{j2}(s_i = 0)U - F_i, 0\}.$$

Given these two payoffs in the first period, the agent makes an optimal choice of whether to adopt or not:

$$\max \{A, B\}.$$

Clearly the agent will never adopt if $F_i \geq 2U$. For the case where $F_i \leq U$, we have the following comparison:

$$\max \{P_{j1}2U + (1 - P_{j1})P_{j2}(s_i = 1)U - F_i, P_{j1}(U - F_i) + (1 - P_{j1})P_{j2}(s_i = 0)(U - F_i)\}.$$

The agent will adopt in the first period if and only if:

$$F_i \leq \frac{P_{j1}2U + (1 - P_{j1})P_{j2}(s_i = 1)U - P_{j1}U - (1 - P_{j1})P_{j2}(s_i = 0)U}{1 - P_{j1} - (1 - P_{j1})P_{j2}(s_i = 0)}$$

Factoring common terms, this simplifies to:

$$F_i \leq \frac{P_{j1} + (1 - P_{j1})(P_{j2}(s_i = 1) - P_{j2}(s_i = 0))}{1 - P_{j1} - (1 - P_{j1})P_{j2}(s_i = 0)} \cdot U = \bar{F},$$

where we denote the threshold value at which the agent is indifferent to adoption as $\bar{F}$. The associated conditional probabilities of adoption in the second period are:

$$P_{i2}(s_j = 1) = Pr(F_i \leq U | F_i \geq \bar{F}),$$

and

$$P_{i2}(s_j = 0) = Pr(F_i P_{j2}(s_i = 0) \leq U | F_i \geq \bar{F}).$$

Therefore, the potential for revelations about the distribution of F generates "learning" or "cascading" adoption behavior, where agents find it advantageous to wait to adopt until after seeing how their colleagues behave. In the model, we estimate that the fact that each of the 64 sub-types of agents have a different F increases the aggregate uncertainty within the network beyond that suggested in this simplified model.