The Diffusion of a Social Innovation: Executive Stock Options from 1936–2005[1]

Lucas Merrill Brown and H. Peyton Young

January 5, 2016

Abstract. We study the increasing use of stock options to compensate executives in US corporations. As with many technological innovations, the adoption curve exhibits a classic S-shaped pattern: the rate rises slowly at first, then there is a period of rapid acceleration, and finally it tails off as the saturation level is approached. Using a longitudinal data set of Frydman and Saks (2010) supplemented with financial reports compiled by the authors, we argue that the diffusion of options was initially given a jump-start by a change in the tax law, but thereafter it was propelled by a process in which firms learned from the experience of earlier adopters. The notion that options spread primarily through social conformity or 'jumping on the bandwagon' is not borne out by the data.

1. Introduction

Economic growth is increasingly driven by social innovations and 'knowledge assets,' such as new management strategies, innovative financial products, and novel contractual arrangements. Studying how these social innovations spread throughout the economy is therefore of great importance for understanding the propellants of economic growth. To date, however, most of the work on innovation diffusion has concentrated on technological innovations.[2] In this paper we examine a key social innovation – the use of stock options for executive compensation – and analyze the forces that led to their increasing popularity.

[1] The authors thank Carola Frydman and Raven S. Molloy for generously providing access to their historical data, which form the empirical basis of the present paper.

[2] There are some innovations that have both technological and social aspects: a notable example is contraception (Kohler, 1997; Kohler et al. 2001).

It is a surprising fact that stock options have been employed as a compensation tool since the 1930s. However it took many decades before they came into widespread use. As with many other technological innovations, the adoption curve displays a classic S-shaped pattern (see Figure 1). From the 1930s to the 1950s there was a slow upward trend in their usage, but even after 20 years only about 10% of the 101 largest U.S. corporations had adopted them. Following a highly favorable change to their tax treatment in 1951, the adoption level increased sharply from 10% to 25% by 1955. Nevertheless, it took another thirty years for their penetration to reach 75%. We are therefore presented with a puzzle: why did it take so long for the innovation to take hold? And was its gradual acceptance due mainly to social conformity or to an increasing appreciation of its benefits?

In this paper we argue that the use of stock options by U.S. corporations was driven to a significant extent by a process of learning from others. We also show why such a process often takes a long time to build up steam. The essential idea is that the experience of early adopters (the innovators) gradually builds up a store of favorable precedents that lead others to adopt, but since there are relatively few innovators to begin with this phase can take a long time. Once a crucial threshold is reached, however, the process starts to accelerate very rapidly.

Although broadly speaking this S-shaped pattern is shared by other processes, such as contagion and bandwagon models, learning processes have empirically testable features that distinguish them from alternative explanations (see section 3). Learning models also posit that potential adopters respond to favorable outcomes from prior adopters. In the present case we can measure the benefit of adoption as the average change in stock price among adopters relative to the change in stock price among non-adopters. We show that this has a highly significant effect on the subsequent rate of adoption. A similar result holds when we restrict attention to firms in a given line of business, such as manufacturing, retail, or finance: favorable outcomes by prior adopters in a given sector have a significant impact on subsequent rates of adoption by other firms in the same sector. In other words, firms appear to be learning from the experience of firms that are similar to themselves.

Our account is in the same general spirit as that of Griliches (1957), who argued that the rate of adoption of hybrid corn in the midwestern United States was driven by the potential gain in profitability from the innovation compared to the status quo. To demonstrate this point, Griliches analyzed diffusion rates in agricultural regions that differed in their inherent productivity and hence in the potential gain in profitability from switching to the new type of seed. However, he did not have panel data on adoption dates at specific farms, and thus he could not directly test the proposition that farmers were responding to perceived gains in profitability. Our longitudinal data set is sufficiently detailed that we can examine how the lagged outcomes of prior adopters (as measured by changes in their market value compared to non-adopters) affected the subsequent adoption rates by other firms in the same NAICS sector. In this respect our analysis is related to the pioneering work of Ryan and Gross (1943, 1950), who studied the diffusion of hybrid corn using detailed longitudinal information about adopting farmers in rural communities in Iowa. Two key findings of their study were, first, that early adopters acted as experimenters by demonstrating the benefits of adoption to others in the community, and second that later adopters reported that the single most influential source of information was the experience of neighbors who had adopted earlier (Ryan and Gross 1943, 1950). Recent work in this tradition includes Conley and Udry (2010), who study the adoption of pineapple cultivation by farmers in Ghana, and find that the outcomes among prior adopters in one's neighborhood is a key factor in explaining the decision to adopt.

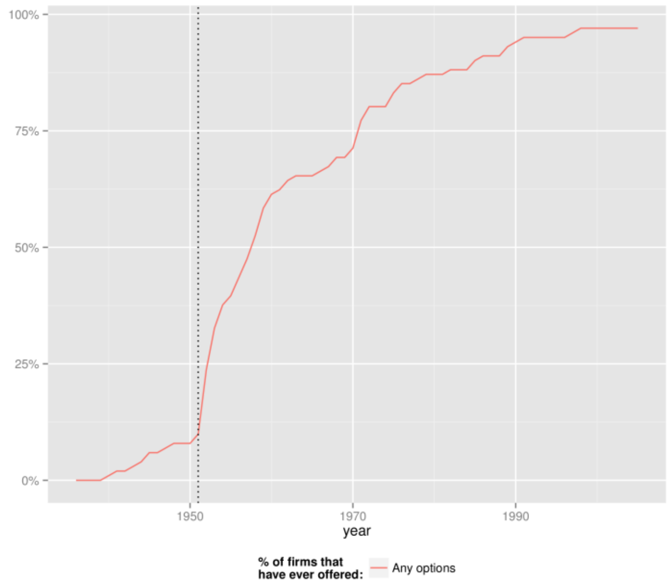

Figure 1. Percentage of the largest 101 U.S. firms (by revenues) that have ever granted options to one or more of their top three executives. The vertical line indicates 1951, when the 1950 tax reforms first applied.

2. Background

The idea that stock options can enhance corporate value has been around for a long time. Indeed, ever since Berle and Means (1932) argued that the separation of ownership and control distorts managers' incentives, there has been a push by corporate observers and academics to more closely align the interests of managers and stockholders through the use of stock grants or options. A stock option gives the employee the right to purchase a certain number of shares at the exercise price. The right expires after a specified period —usually 10 years. The value of an option is calculated using the Black-Scholes formula, which is based on the stock's past volatility.

A number of studies have found that stock-based incentives improve subsequent stock-price performance (Masson 1971; Brickley et al. 1985; Abowd 1990; Yermack 1997). However the question of whether these benefits actually drove their diffusion has not been directly addressed in the prior literature. Indeed a number of authors have suggested that the use of options may have been driven in part by social conformity: as options compensation became more widespread, it took on the character of an industry norm that led other firms to adopt (Hall and Murphy 2003; Levy and Temin 2007; Frydman and Molloy 2010).[3] Although it seems reasonable that social conformity and imitation may have played a role, the central finding of this paper is that learning from the experiences of others was a key factor in the diffusion of options compensation. Specifically, the data suggest that favorable outcomes among early adopters provided the information needed to convince later adopters, and this compounding of information created the 'snowball' effect that one sees in the diffusion process.

Stock options have several characteristics that make them particularly well-suited for studying the mechanics of innovation diffusion. First, grants of stock options are well documented due to disclosure rules by the Securities and Exchange Commission: all stock options given to a firm's top three executives must be included in the company's annual reports. Since these reports are publicly available, other firms can observe which firms have adopted the new technology.

Second, information on the profitability of adopters versus non-adopters is publicly and readily available at frequent intervals. Hence other firms can observe the outcomes of adopting versus non-adopting firms. This public observability provides an appropriate environment for testing learning models that rely on information about the outcomes of others' choices.[4]

[3] Similarly, Murphy (1999) observes that the structure of options compensation seems to have crystallized around certain customary parameters, such as 10-year expiration terms.

[4] When agents can observe whether others have adopted but they cannot see the resulting payoffs, they must infer the benefits of adoption from the fact that others adopted (Banerjee

Third, there is little evidence to suggest that options were substantially refined or improved over time, as is often the case with technological innovations (David 1975; Stoneman 1982). However, their value as a form of compensation is directly affected by the rate at which options income is taxed compared to ordinary salary income. These tax changes are well-documented for the period under consideration. A noteworthy feature is that there was a large one-time jump in the favorable tax treatment of options in 1951, after which their tax advantage (relative to ordinary salary income) steadily eroded. Thus after 1951 the tax benefits of options as a form of compensation steadily decreased even though their use continued to spread rapidly.

Another reason why the benefit of options compensation might have changed over time is the volatility in the stock price. As we previously noted, the Black-Scholes formula implies that the value of an option depends inversely on the volatility of the underlying stock. Moreover, higher volatility undermines the usefulness of options as an incentive, because the stock price becomes a less reliable signal of executive performance. In section 7 we introduce both tax changes and volatility as control variables and show that our results still hold at a high level of statistical significance.

3. Contagion models

Contagion or bandwagon models posit that an innovation spreads like an epidemic: a non-adopter becomes 'infected' with a certain probability when he comes into contact with a previous adopter. The rate of infection is not modelled explicitly: it may result from pressure to conform, from an inclination to imitate others, or from some form of information contagion. Pressure to adopt may also arise from outside sources, such as

---

1993; Bikhchandani et al. 1993). This type of herding process is based on different premises than the models considered here.

mass advertising. The combination of these internal and external sources of pressure leads to a family of models that were first popularized by Bass (1969) as a model of innovation diffusion.[5]

Let $\lambda > 0$ denote an *internal influence* parameter: the rate at which a non-adopter is persuaded to adopt by meeting a prior adopter within the group. Let $\gamma \geq 0$ denote an *external influence* parameter: the rate at which each non-adopter is convinced to adopt by external sources. Let $p(t)$ be the proportion of adopters at time *t*. Then the evolution of $p(t)$ is described by the ordinary differential equation

$$\dot{p}(t) = (\lambda p(t) + \gamma)(1 - p(t)) . \qquad (1)$$

The solution is

$$p(t) = [1 - \beta\gamma e^{-(\lambda+\gamma)t}] / [1 + \beta\lambda e^{-(\lambda+\gamma)t}], \quad \beta > 0 . \qquad (2)$$

The choice of β is determined by the initial conditions; in particular if $p(0) = 0$ then $\beta = 1/\gamma$.

When γ = 0 we obtain a logistic curve, which is the functional form employed by Griliches in his study of hybrid corn. The key idea in his paper was to demonstrate that the speed of adjustment λ is an increasing function of the potential gain in profit from switching to the new technology.

[5] The 'Bass model' is widely used in marketing and sociology. For surveys and critiques of this approach see Sultan et al. (1990), Bass et al. (1994), Bass and Wind (1995), Stoneman (2002), and Rogers (2003). There is another class of models that is based on the idea that people have different degrees of susceptibility to social influence (Schelling 1971,1978; Granovetter 1978; Granovetter and Soong 1988; Macy 1991; Valente 1995, 1996, 2005; Dodds and Watts 2004, 2005; Lopez-Pintado and Watts 2008). These models are quite varied and we shall not attempt to test them here.

There is a substantial literature arguing that the Bass diffusion model can be made to fit diffusion data with a judicious choice of λ and γ (for an overview see Rogers (2003)). In the present case, however,we find that this family of models does not do a good job of explaining the data. Consider the following discrete-time analog of (1):

$$\begin{aligned}\Delta p(t) &= p(t) - p(t-1) = (\lambda p(t-1) + \gamma)(1 - p(t-1)) \\ &= \gamma - \lambda(p(t-1))^2 + (\lambda - \gamma)p(t-1)\end{aligned} \tag{3}$$

Thus the model predicts that $p(t)$ should be quadratic in the prior-period's level of adoption, and the coefficient on the quadratic term should be negative (since λ is assumed to be positive). To mitigate the effect of the tax change in 1951, we examined this prediction for the period 1952-2005 when taxes on options were stable or somewhat increasing relative to cash compensation. As shown in Table 1 the regression coefficient on *p(t-1)* is *positive* and highly significant, which is inconsistent with the predictions of a contagion model.

	Coefficient	S.E.	*t*-value	*p*-value
(Intercept)	0.1141	0.0197	5.80	0.0000
$p(t-1)^2$	0.1000**	0.0438	2.28	0.0267
$p(t-1)$	-0.2112	0.0605	-3.49	0.0010

Table 1. Testing the contagion model on adoption data from 1952–2005. The coefficient on the quadratic term is positive and highly significant, whereas the contagion model predicts that it should be negative.

A second test of the contagion model is based on the behavior of the quantity $H(t) = \frac{\dot{p}(t)}{p(t)(1-p(t))}$, which is the rate of new infections amongst the uninfected *per currently infected agent*. This is known as the *relative hazard rate* (Young 2009). It can be shown that in any contagion model $H(t)$ must be nonincreasing; moreover this holds even when the population is heterogeneous with respect to the contagion parameters λ and γ (Young, 2009).

This prediction is a not supported by the data either. Regression analysis shows that after 1951 there are two periods -- 1960–1971 and 1976–1990 -- during which $H(t)$ was significantly and steadily increasing (see Table 2).[6] These periods of growth in the relative hazard rate are contrary to the predictions of the contagion model for any distribution of the underlying parameters.

	Coefficient	S.E.	*t*-value	*p*-value
(Intercept)	-26.5294	11.6794	-2.27	0.0465
Year	0.0135**	0.0059	2.28	0.0459

Dependent variable is the relative hazard rate $H(t)$. Data from 1960 to 1971.

	Coefficient	S.E.	*t*-value	*p*-value
(Intercept)	-19.2684	8.8474	-2.18	0.0484
Year	0.0098**	0.0045	2.19	0.0477

Dependent variableis the relative hazard rate $H(t)$. Data from 1976 to 1990.

Table 2. Estimation of the coefficient on the relative hazard rate on two time intervals identified by the Chow test for structural breaks.

By contrast, learning models are compatible with an increasing hazard rate, and also with a positive quadratic dependence on the prior period level of adoptions (Young 2009). As we shall explain in the next section, the key reason why learning models differ from contagion is that they typically undergo an explosive acceleration phase as the accumulated evidence from prior adopters convinces more and more firms that the innovation really is beneficial.

[6] These two intervals are identified by the Chow test for structural breaks with p-values less than 0.03.

4. Learning by observation

A key feature of most innovations is that their benefits only become clear after they have been tried repeatedly. Thus even if an innovation has positive benefits on average, the realizations by early adopters may exhibit a great deal of variation. Agents who are risk averse, or who start with pessimistic prior beliefs, will need to see many favorable outcomes before adopting themselves. It is therefore natural to expect that agents will differ in the amount of information that they need to see before adopting. It turns out that such learning processes have distinctive characteristics that set them apart from contagion models (Young 2009).

To see the distinction between the two approaches, it will be helpful to make a number of simplifying assumptions. Suppose that there are N members of the population, and that each adopter generates one observable outcome per unit time period. Assume further that the outcomes from prior adopters are equally informative, and that every non-adopter can observe all of these outcomes.[7] Modelled as a continuous-time process,

the expected number of observations generated from the beginning of the process to time t is given by the expression:

$$N\int_0^t p(s)ds\,. \qquad (4)$$

Let us posit that for every agent i there is a critical *information threshold* -- a positive integer N_i -- such that i is likely to adopt after seeing at least N_i outcomes by prior

[7] A different class of models arises if agents are situated in a social network and can observe only the outcomes of their neighbors (Blume 1993; 1995; Valente 1995; Bala and Goyal 1998; Young 2003; Chatterjee and Xu 2004; Cowan and Jonard 2004; Jackson and Rogers 2007; Jackson and Yariv 2007; Vega-Redondo 2007; Golub and Jackson 2010; Jackson 2010; Conley and Udry 2010; Banerjee et al. 2013; Kreindler and Young 2014; Centola and Baronchelli 2015).

adopters, but is not likely to adopt after seeing fewer than N_i outcomes. Not all outcomes need be favorable; indeed they will tend to be dispersed given that the process is stochastic. However, as the number of realizations grows, the observed mean will be positive with increasing probability (given our assumption that the innovation is superior to the status quo), and the variance around the observed mean will decrease.[8] Thus an agent who is risk averse, or holds pessimistic prior beliefs, will eventually be persuaded that it is worthwhile to adopt given sufficient evidence.

For the sake of concreteness assume that the payoff gain from the innovation (relative to the status quo) is described by a normally distributed random variable X with mean $\mu > 0$, that is, the expected net benefit from adopting is m per period. A risk neutral individual would adopt if he knew that μ is positive, but ex ante he is uncertain about the true value of the mean. Suppose that individual i has a prior belief μ_{i0} about the value of μ. If μ_{i0} is positive he is ready to adopt with no further evidence, whereas if μ_{i0} is negative he would prefer to wait.[9] Consider an individual i who observes n independent draws with mean $\bar{x}$. In a normal-normal updating model her posterior beliefis given by an expression of form

$$\tilde{\mu}_i = \frac{n\bar{x} + \tau_i \mu_{i0}}{n + \tau_i}, \qquad (5)$$

where $\tau_i > 0$ reflects the *inflexibility* of her beliefs (deGroot 1970). The quantity $\pi_i = \max\{0, -\mu_{i0} / \mu\}$ can be interpreted as i's initial level of *pessimism*. If $\mu_{i0} > 0$ then i already believes that the innovation is a good thing (i is an optimist). A pessimistic

[8] The variance of the mean will be a decreasing function of n even if the realizations are correlated, so long as they are not perfectly correlated.

[9] For simplicity we shall suppose that there are no switching costs.

individual will wait to adopt until his posterior belief $\tilde{\mu}_i$ is positive, that is, until $\bar{x} > (\tau_i \pi_i / n)\mu$. Since the expected value of $\bar{x}$ equals μ and the standard deviation decreases as $1/\sqrt{n}$, this event is likely if $n > \tau_i \pi_i$ and becomes more likely the larger n is. We may therefore think of $N_i = \tau_i \pi_i$ as the "critical information threshold" for *i*.

In what follows it will be convenient to consider the normalized threshold $r_i = N_i / N$, which we shall call *i*'s *resistance level*. From the preceding it follows that *i* is likely to adopt when

$$r_i \leq \int_0^t p(s)ds. \tag{6}$$

In general the information thresholds N_i will depend on various factors in addition to agents' prior beliefs, such as idioscyncratic costs of adoption, degree of risk aversion, and so forth. Whatever the sources, it is reasonable to assume that the numbers N_i, and likewise the resistances r_i, will be heterogeneously distributed in the population of potential adopters.

Fix the population size N and let $F(r)$ denote the cumulative distribution function of the resistance levels $r \in [0,1]$. $F(0)$ is a measure of the optimists who are prepared to adopt without further evidence (these are the "innovators" in the terminology of Rogers (2003)). We assume henceforth that $F(0) > 0$, for otherwise the process cannot get started. Let the parameter $\lambda > 0$ represent the *speed of adjustment*, that is, the rate at which potential adopters actually do adopt once their resistance level has been passed. (Thus $1 - \lambda$ is the degree of inertia.) Then the instantaneous rate of new adoptions can be written as follows:

$$\dot{p}(t) = \lambda[F(\int_0^t p(s)\,ds) - p(t)]. \qquad (7)$$

This expression states that the rate of new adoptions at time t is proportional to the number who are ready to adopt (i.e., whose resistance level has been passed by time t) but have not yet adopted.

Unlike contagion models the process is driven forward by the cumulative level of prior adoptions $\int_0^t p(s)ds$ and not just on the current level of adoptions $p(t)$. This leads to a compounding effect: as information accumulates from early adopters, more individuals adopt, which enlarges the base of information, which leads still more individuals to adopt and so on.

The basic logic can be illustrated by the following example. Consider the distribution of individuals shown in Figure 2. There is a relatively small group of innovators (who have no resistance), a somewhat larger group of early adopters who have some resistance, a still larger group with even more resistance, and so forth.

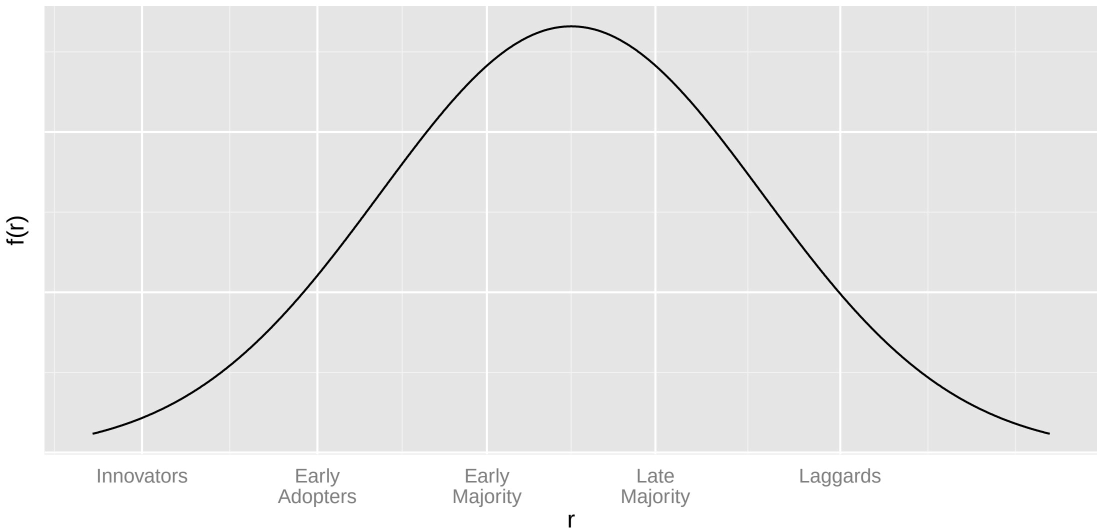

Figure 2. Illustrative distribution of resistance among potential adopters adapted from Rogers (2003).

Assume for the sake of illustration that $F(r)$ is normal with mean 3, standard deviation 2, and that $\lambda = 0.1$. The equation of motion defined in (7) traces out the adoption curve shown in Figure 3. The very steep acceleration phase occurs when the accumulated information from the innovators is enough to convince the early adopters, who together with the innovators generate enough information to convince the 'early majority' and so on. After the mode of the distribuiton is reached, the innovation continues to spread but at a decelerating rate.

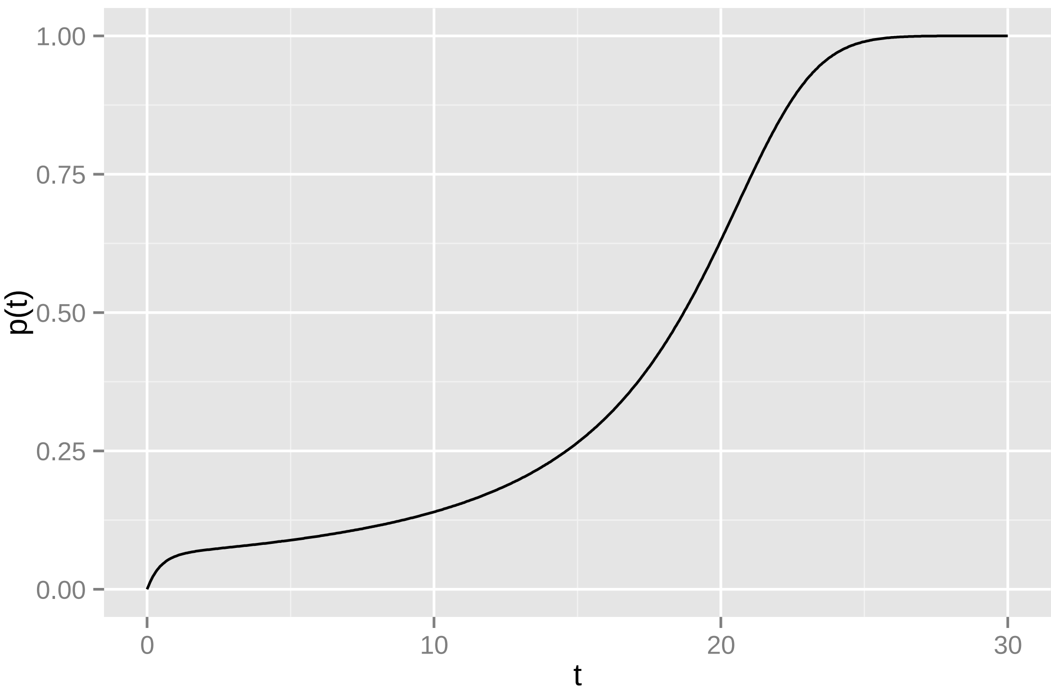

Figure 3: Adoption curve generated by equation (7) when $\lambda = 0.1$ and the CDF is normal with mean 3 and standard deviation 2.

5. Estimating the learning model

In this section we shall estimate the parameters of the learning model using the options diffusion data of Frydman and Saks (2010). We proceed under the assumption that F is a (truncated) cumulative normal distribution. Although we cannot test for the best functional form directly, the normal distribution is broadly consistent with the prior literature about the distribution of resistance in the population (see for example Rogers 2003). Since this family of distribution functions is nonlinear we cannot use OLS. Instead we shall estimate the mean and variance of F together with the coefficients of various control variables using a maximum likelihood approach. This model gives a very good fit as we shall subsequently show.

Let y_t be the observed number of first-time adoptions in year t, and let N be the size of the population of potential adopters. Thus y_t/N is a discrete-time approximation of $\dot{p}(t)$. The likelihood function must reflect the fact that the values of y_t are non-negative integers. For this purpose we assumed that the y_t are generated by a negative binomial model whose parameters we estimated from the data (for details see Appendix A.3).

Table 3 shows the results of the estimation procedure, with the weight on the CDF estimated simultaneously with the standard deviation and mean. Standard errors are Huber-White robust standard errors. Maximum-likelihood estimation converges on a CDF with a mean of 0.4644 and a standard deviation of 0.3036. The weight on the CDF is positive at a high level of significance.

Name	Coefficient	S.E.	*t*-value	*p*-value
CDF of integral of prev. adoptions (%)	7.11***	1.69	4.20	0.00
Level of previous adoptions (%)	-7.47***	3.42	-2.18	0.03
Year (scaled)	0.25	2.22	0.11	0.91

Table 3. Estimation of the learning model, where the dependent variable is the number of firms who first adopt options compensation in a given year. The negative binomial model has estimated coefficient $\alpha^{-1} = 5.16$ (S.E. = 4.66).

Consistent with the predictions of the learning model, the weight on the CDF of the integral of prior adoptions is *positive* and highly significant, whereas the coefficient on the level of adoptions in the prior period is *negative* and highly significant.[10]

The mean and standard deviation that are estimated for the CDF of the integral of previous adoptions lead to a reasonably shaped density curve, as shown in Figure 4. Those individuals with adoption thresholds less than or equal to zero could be considered "innovators" who believe in the utility of the innovation without having to observe prior adoptions.

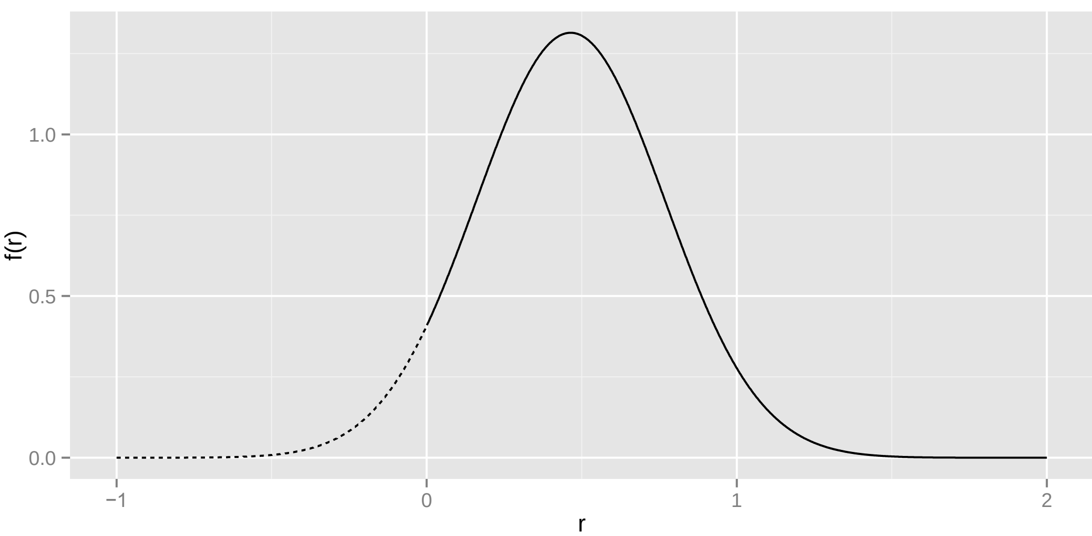

Figure 4: Graph of the normal CDF as estimated by the maximum likelihood procedure. The estimated mean is .46 (0.13) and the standard deviation is .30 (0.08).

[10] The inclusion of a dummy variable for the year (1951) in which the tax advantage of options increased does not appreciably change these results.

The estimated model yields an adoption curve that matches the observed adoption curve quite closely, as shown in Figure 5.

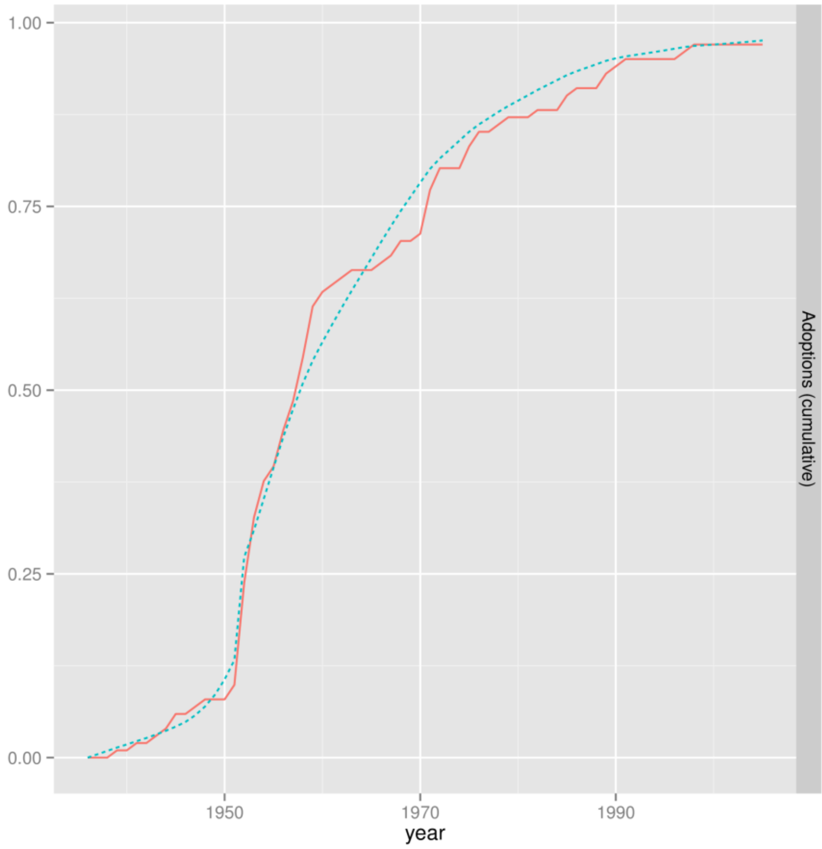

Figure 5. Actual versus estimated adoptions for the learning model, using the results in Table 3. The dotted line is the estimate, the red line is the actual.

6. Estimating the potential benefits from adoption

In the learning model discussed above, firms are assumed to adopt if the number of realized outcomes from prior adopters is sufficiently high given their initial resistance level. The integral under the adoption curve up to time *t* is a measure of the cumulative number of such realizations. Thus far we have shown that the cumulative number of realizations – the amount of information -- does in fact have explanatory power. One could ask, however, whether there is *direct* evidence that the realized net benefits have an effect on the subsequent adoption rate. The question is how net benefits should be measured. Given that the exercise value of an option is determined by the underlying stock price, and given that the purpose of options is to motivate executives to increase the stock price, a natural measure of 'net benefit' is the percentage change in the stock market value of adopting firms, minus the percentage change in the stock market value of nonadopting firms over some suitable period. When we include this variable in the estimations we obtain further evidence that firms are in fact reacting to the financial benefits accruing to prior adopters.

To be specific, let us define the market value of firm *i* at the end of a given year *t* to be the share price multiplied by the number of outstanding shares.[11] Let $\Delta m_{i,t}$ denote the percentage change in the market value of firm *i* from the end of year $t-1$ to the end of year *t*. Let Δm_t be the average of the values $\Delta m_{i,t}$ over all firms *i* who were adopters by the end of year $t-1$, less the average of the values $\Delta m_{j,t}$ over all firms *j* who were non-adopters in both year $t-1$ and in year *t*. Finally, let B_t^k be the average of the values Δm_t over the previous *k* years: $B_t^k = (1/k)\sum_{s=t-k+1}^{t} \Delta m_s$ We shall take this as our measure of the *net benefit* from adopting. Our hypothesis is that firms are increasingly likely to adopt when B_t^k is high over a substantial number of periods *k*. Table 5 shows that the net benefit when measured over the previous five years has a significant impact on the probability of adoption.[12]

[11] The market value of the sampled firms was accessed through the Center for Research in Security Prices (CRSP). http://wrds-web.wharton.upenn.edu/wrds/ds/crsp/index.cfm.

[12] When k = 1 the coefficient on change in market value is positive but not significant.

	Coefficient	S.E.	*t*-value	*p*-value
CDF of integral of prev. adoptions (%)	3.0228***	0.4930	6.13	0.0000
Level of previous adoptions (%)	-5.1269***	1.8394	-2.79	0.0053
Year (scaled)	-0.1941	2.2403	-0.09	0.9309
Net benefit (5 yr avg), adopters vs non-adopters	1.9576***	0.7643	2.56	0.0104

Table 4. Estimation of the learning model when the net benefit of adoptions (averaged over the previous five years) is included. The estimated mean of the normal CDF is .46 (0.13) and the standard deviation is .30 (0.08).

7. The effect of volatility and tax rates.

In this section we consider two exogenous variables that affect the value of options compensation: i) marginal tax rates on options versus cash compensation and ii) volatility of stock prices. The impact of changes in the tax code on options have been discussed by several authors, including Madeo and Omer (1994), Goolsbee (2000), and Hall and Liebman (2000), but they did not employ the dataset that we do here. To the best of our knowledge the impact of volatility on the rate of options diffusion has not been examined elsewhere.

7.1. Tax rates on options. Prior to 1951, executives could only receive 'non-qualified' stock options, which were taxed as labor income at a marginal rate of nearly 90% for the highest income bracket (Washington and Rothschild 1962). By comparison, the highest bracket for capital gains was approximately 25%. Through the introduction of 'qualified stock options', the 1950 Revenue Act changed the classification of stock options from labor income to capital gains income, thus reducing the marginal rate for top earners by

65%. In 1976, the Tax Reform Act eliminated qualified stock options (Noble 1981). The ban expired in 1981, but a cap of $100,000 was placed on the value of qualified stock that could be awarded to any one executive (Hall and Liebman 2000; Frydman and Molloy 2011). Thus for the executives of large corporations the effective marginal rate on options compensation was essentially the same as for cash compensation. The difference between the highest marginal tax on options and on cash compensation over the period is shown in Figure 6.

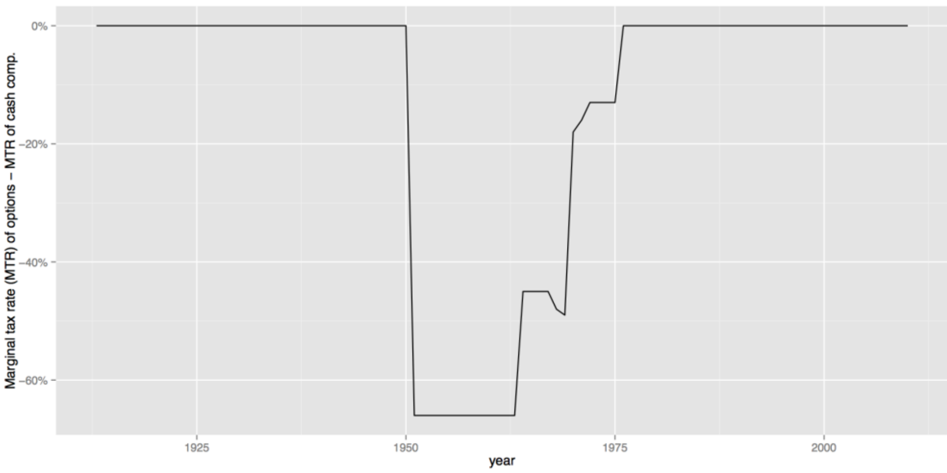

Figure 6. Highest marginal tax rate on options compensation minus the highest marginal tax rate on cash compensation, 1913–2012.

7.2 Market volatility. A second exogenous factor that affects the value of options compensation is the volatility of stock prices. Lower volatility makes stock options more appealing for both the firm and the recipient (Holderness et al. 1999, Stulz 1996). For the recipient, options are more appealing due to decreased risk. For the firm, reduced volatility makes stock market value a more accurate measure of managerial performance. A standard measure of volatility is 5% Value-at-Risk (VaR) (Linsmeier and Pearson 2000, Gourieroux and Jasiak 2010). The value of the statistic indicates that there is a 5% chance of portfolio losses *at least* as large as the statistic on any given day. Using daily S&P 500 market returns data obtained through the CRSP database, we calculated the 5% VaR for each year since 1936.[13] To smooth out idiosyncratic effects we used the average 5% VaR of the S&P 500 over five-year intervals as the variable of interest.

The results are shown in Table 5. Higher marginal tax rates and higher volatility have negative effects on the subsequent adoption rate, just as theory would predict.

	(1)	(2)	(3)
CDF of integral of prev. adoptions (%)	7.11*** (1.69)	3.0228*** (0.493)	3.6665*** (0.6103)
Level of previous adoptions (%)	-7.47** (3.42)	-5.1269*** (1.8394)	-4.2604** (1.777)
Year (scaled)	0.25 (2.22)	-0.1941 (2.2403)	-0.7588 (2.2098)
Net benefit (5 yr avg), adopters vs non-adopters	--	1.9576*** (0.7643)	1.6823** (0.768)
Δ Relative Marginal Tax on Options	--	--	-1.5179* (0.8764)
Value-at-Risk (5-yr avg)	--	--	-70.5028** (32.4707)

Table 5: Estimation of the learning model with all variables included. Standard errors are given in parentheses. Model (3) includes the relative marginal tax on options and market volatility. Models (1) and (2) are reprinted from earlier tables for easy reference.

8. Learning from peers

Our final test of the learning hypothesis is whether firms are responding to the prior experience of *similar* firms. To examine this possibility we grouped the 101 firms in our sample by NAICS sectors and sub-sectors (see Appendix A.4 for a list of these sectors). In particular we studied the number of adoptions per year in each manufacturing sub-

[13] For example the 5% VaR statistic equals 2% in a given year if there is a 5% chance that the S&P 500 will lose at least 2% of its value on any given day in that year.

sector and each non-manufacturing sector that contained at least four firms in our sample. The adoption curves are shown in Figure 7.

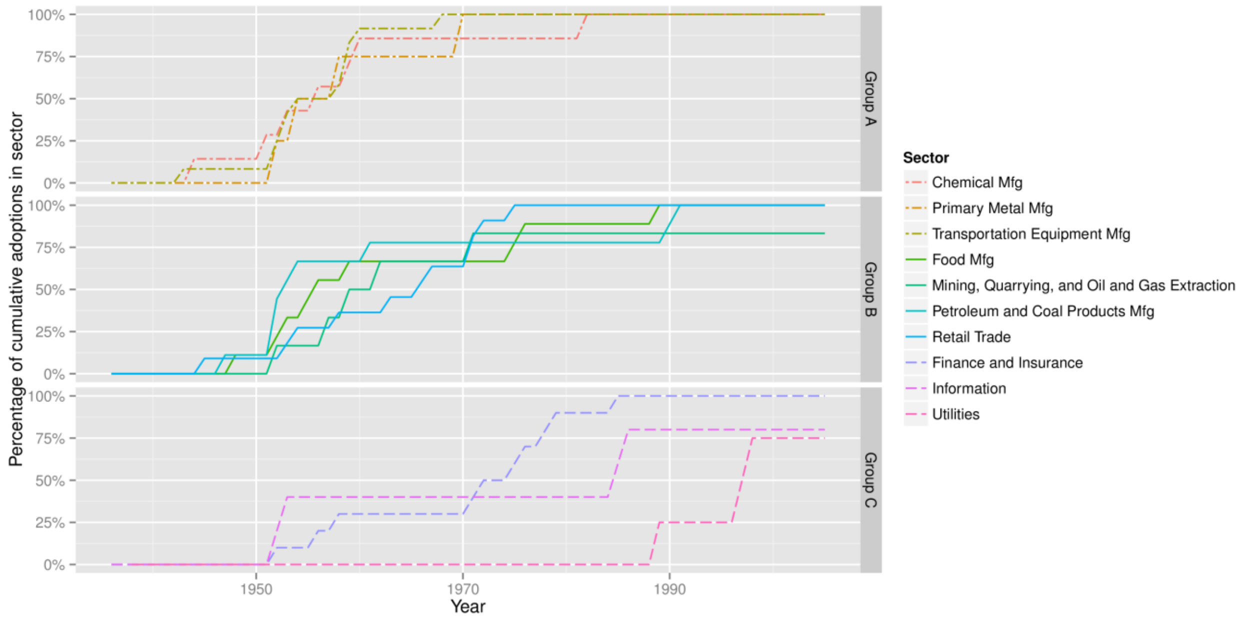

Figure 7: Cumulative adoptions per sector, for all sectors with over four firms. Firms are split into three arbitrary groups, which are clustered to show the differences in adoption curves. Groups are arbitrary and are purely for visual convenience.

Within each sector we estimated the learning model using the same techniques discussed in Section 5. Due to the small number of observations within these groups we omitted several variables including marginal tax rates and volatility. However we included fixed effects for each of the 10 groups. The results are shown in Table 6. The key finding is that the integral of previous within-group adoptions and the perceived benefits from adoption are both highly significant. This lends support to the idea that firms are learning from the experiences of prior adopters in a similar line of business.

	Coefficient	S.E.	*t*-value	*p*-value

CDF of integral of prev. adoptions (%)	2.5921***	0.7404	3.50	0.0005
Level of previous adoptions (%)	-4.8104***	1.0833	-4.44	0.0000
Year (scaled)	1.5665	1.1868	1.32	0.1868
Net benefit (5 yr avg), adopters vs non-adopters	1.6956***	0.4849	3.50	0.0005
Sector Fixed Effects:	10			

Table 6. Estimation of the learning model conditional on prior adoptions by firms in the same economic sector. Firms are grouped into 10 sectors.

9. Conclusion

In this paper we have analyzed various factors that led to the increasing use of stock options to compensate U.S. corporate executives over the past 80 years. The pattern of diffusion is very similar to that of many technological innovations: initially the rate of diffusion is rather slow and erratic, then it starts to accelerate rapidly as more and more people become convinced that the innovation is worthwhile, and finally it tails off as saturation is reached. Our basic premise is that the accumulation of information from prior adopters drove the process forward. Although this idea has been suggested by various authors in other contexts, it is difficult to test without extensive longitudinal data on individual adoption decisions and the resulting payoffs. The dataset assembled by Frydman and Saks (2010), together with publicly available market data, provides the type of information one needs to analyze this issue. We have a detailed account of the dates at which corporations first adopted, and we can observe the subsequent changes in market values for adopters and non-adopters alike. The major factors that affect the value of options compensation – changes in tax treatment and underlying market volatility – are well-documented and can be introduced as control variables. Unlike many technological inovations, there was no organized campaign promoting their usage, nor is there any reason to think that the innovation itself was substantially improved and refined over time. Both of these factors pose serious identification problems in the spread of most technological innovations.

Our findings can be summarized as follows. First, the overall pattern of diffusion is consistent with a learning model, whereas it is not consistent with standard contagion models. In fact, using the relative hazard rate as a statistic, one can reject the hypothesis that the process was driven by a contagion model with heterogeneous parameters (see section 3). The key problem is that contagion models do not predict such a sharp period of acceleration in the middle phase of the adoption process, whereas this is a natural feature of a learning model. Second, we showed that the learning model is validated at a high level of statistical significance under the assumption that resistance in the population follows a normal distribution. In particular, the process appears to be driven forward by the amount of cumulative information generated by prior adopters, and not the current level of adoption, as would be the case with a contagion or bandwagon model. Third, we showed that the average net benefit to prior adopters, as measured by the increase in stock values compared to non-adopters, has a significant effect on the rate of subsequent adoption. Finally, we showed that the same results hold when one restricts attention to information generated by firms within a given economic sector. In other words, it appears that firms pay attention to information generated by firms who are similar to themselves. This idea – that people learn by observing the outcomes of prior adopters --- is one of the central propositions in the early work by Ryan and Gross (1943, 1950).

Of course, these results do not rule out the possibility that some firms may have imitated their peers without paying attention to the associated payoffs. Nor can we rule out the idea that as options became widespread some firms felt compelled to go along with the trend, i.e., options compensation took on aspects of a social norm. What we can say is that learning from the favorable experiences of others appears to have played a substantial role in the spread of options, and it thus provides a validation of one of the central propositions of the early literature on innovation diffusion.

References

Abowd, J. M. (1990). Does performance-based managerial compensation affect corporate performance? *Industrial and Labor Relations Review*, 52S–73S.

Banerjee, A. (1992). A simple model of herd behavior. *The Quarterly Journal of Economics 107*(3), 797–817.

Banerjee, A. A.G. Chandrasekhar, E. Duflo, M. Jackson (2013). The diffusion of microfinance. *Science 341*, 363-370.

Bass, F. (1969). A new product growth for model consumer durables. *Management Science 15*(5), 215–227.

Bass, F., T. Krishnan, and D. Jain (1994). Why the Bass model fits without decision variables. *Marketing Science 13*(3), 203–223.

Bass, F. and J. Wind (1995). Introduction to the special issue: empirical generalizations in marketing. *Marketing Science 14*(3), 2.

Batten, D. (1987). The balanced path of economic development: a fable for growth merchants. In *Economic Evolution and Structural Adjustment*, pp. 64. Springer.

Bebchuk, L. A. and J. M. Fried (2003). Executive compensation as an agency problem. *Journal of Economic Perspectives 71*, 92.

Berle, A. A. and G. G. C. Means (1932). *The modern corporation and private property*. Transaction Publishers.

Bikhchandani, S., D. Hirshleifer, and I. Welch (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy 100*(5), 992–1026.

Blume, L. E. (1993). The statistical mechanis of strategic interaction. *Games and Economic Behavior 5*, 387-424.

Blume, L. E. (1995). The statistical mechanics of best-response strategy revision. *Games and Economic Behavior 11*, 111-145.

Brickley, J. A., S. Bhagat, and R. C. Lease (1985). The impact of long-range managerial compensation plans on shareholder wealth. *Journal of Accounting and Economics 7*(1), 115–129.

Cameron, C. A. and P. K. Trivedi (2005). *Microeconometrics: Methods and applications*. New York, NY: Cambridge University Press.

Cameron, A. C. and P. Trivedi (2013). *Regression Analysis of Count Data*. Cambridge University Press.

Centola, D. and Baronchelli (2015). The spontaneous emergence of conventions: an experimental study of cultural evolution. *Proceedings of the National Academy of Sciences 112 (7)*.

Chatterjee, K. and S. H. Xu (2004). Technology diffusion by learning from neighbors. *Advances in Applied Probability 36*, 355-376.

Conley, T.G. and C. R. Udry (2010). Learning about a new technology: pineapple in Ghana. *American Economic Review 100,* 35-69.

Cowan, R. and N. Jonard (2004). Network structure and the diffusion of knowledge. *Journal of Economic Dynamics and Control 28*, 1557-1575.

David, P. (1975). *Technical Change, Innovation, and Economic Growth*. Cambridge: Cambridge University Press.

De Groot, M. (1970). *Optimal Statistical Decisions*. New York: McGraw-Hill.

Dodds, P. S. and D. J. Watts (2004). Universal behavior in a generalized model of contagion. *Physical Review Letters 92 (21)*, 218701-4.

Dodds, P. S. and D. J. Watts (2005). A generalized model of social and biological contagion. *Journal of Theoretical Biology 232*, 587-604.

Frydman, C. and R. S. Molloy (2011). Does tax policy affect executive compensation? Evidence from postwar tax reforms. *Journal of Public Economics 95*(11), 1425–1437.

Frydman, C. and R. E. Saks (2010). Executive compensation: A new view from a long-term perspective, 1936–2005. *Review of Financial Studies 23*(5), 2099–2138.

Golub, B. and M. O. Jackson (2010). Naïve learning in social networks and the wisdom of crowds. *American Economic Journal: Microeconomics 2*, 112-149.

Goolsbee, A. (2000). What happens when you tax the rich? Evidence from executive compensation. *Journal of Political Economy 108*, 352-378.

Gourieroux, C. and J. Jasiak (2010). Value-at-risk. *Handbook of Financial Econometrics*, 553–609.

Granovetter, M. (1978). Threshold models of collective behavior. *American Journal of Sociology 83*, 1420-1443.

Granovetter, M. and R. Soong (1988). Threshold models of diversity:Chinese restaurants, residential segregation, and the spiral of silence. *Sociological Methodology* 18, 69-104.

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

Hall, B. J. and J. B. Liebman (2000). The taxation of executive compensation. In *Tax Policy and the Economy, Volume 14*, pp. 1–44. MIT Press.

Hall, B. J. and K. J. Murphy (2003). The trouble with stock options. *Journal of Economic Perspectives 17*(3), 49–70.

Holderness, C. G., R. S. Kroszner, and D. P. Sheehan (1999). Were the good old days that good? Changes in managerial stock ownership since the great depressionchanges in managerial stock ownership since the Great Depression. *The Journal of Finance 54*(2), 435–469.

Jackson, M. O. (2010). *Social and Economic Networks*. Princeton NJ: Princeton University Press.

Jackson, M. O. and B. W. Rogers (2001). Relating network structure to diffusion properties through stochastic dominance. *BE Journal of Theoretical Economics 7*, 1-13.

Jackson, M. O. and L. Yariv (2007). Diffusion of behavior and equilibrium properties in network games. *American Economic Review 97*, 92-98.

Kohler, H.-P. (1997). Learning in social networks and contraceptive choice. *Demography 34*, 369-383.

Kohler, H.-P., J.R. Bereman, and S.C. Watkins (2001). The density of social networks and fertility decisions: evidence from South Nyanza District, Kenya. *Demography 38*, 43-58.

Kreindler, G. E. and H. P. Young (2014). Rapid innovation diffusion in social networks. *Proceedings of the National Academy of Sciences 111, Suppl 3*, 10881-10888.

Lawless, J. F. (1987). Negative binomial and mixed poisson regression. *Canadian Journal of Statistics 15*(3), 209–225.

Levy, F. and P. Temin (2007). Inequality and institutions in 20th century America. Technical report, National Bureau of Economic Research.

Linsmeier, T. J. and N. D. Pearson (2000). Value at risk. *Financial Analysts Journal*, 47–67.

Lopez-Pintado, D. and D. J. Watts (2008). Social influence, binary decisions, and collective dynamics. *Rationality and Society 20*, 399-443.

Macy, M. (1991). Chains of cooperation: threshold effects in collective action. *American Sociological Review 56*, 730-747.

Madeo, S. and T. Omer (1994). The effect of taxes on switching stock option plans: evidence from the tax reform act of 1969. *Journal of the American Taxation Association 16*, 24-42.

Masson, R. T. (1971). Executive motivations, earnings, and consequent equity performance. *The Journal of Political Economy*, 1278–1292.

Murphy, K. J. (1999). Executive compensation. *Handbook of Labor Economics 3*, 2485–2563.

Noble, K. B. (June 28, 1981). Revival for stock options? *New York Times*.

Rogers, E. (2003). *Diffusion of innovations: Fifth Edition*. New York: Free Press.

Ryan, B. and N. Gross (1943). The diffusion of hybrid seed corn in two Iowa communities. *Rural Sociology 8*(1), 15–24.

Ryan B. and N. Gross (1950). Acceptance and diffusion of hybrid corn seed in two Iowa communities. Research Bulletin 372, Agricultural Experiment Station, Iowa State College of Agriculture and Mechanic Arts, Ames Iowa.

Schelling, T. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology 1*, 143-186.

Schelling, T. (1978). *Micromotives and Macrobehavior*. New York: Norton.

Stoneman, P. (2002). *The Economics of Technological Diffusion*. Oxford: Blackwell.

Stulz, R. M. (1996). Rethinking risk management. *Journal of Applied Corporate Finance 9*(3), 8–25.

Sultan, F., J. Farley, and D. Lehmann (1990). A meta-analysis of applications of diffusion models. *Journal of Marketing Research 27*(1), 70–77.

Valente, T. W. (1995). *Network Models of the Diffusion of Innovations*. New York: Hampton Press.

Valente, T. W. ((1996). Social network thresholds in the diffusion of innovations. *Social Networks 18*, 69-89.

Valente, T. W. (2005). Network models and methods for studying the diffusion of innovations. In *Models and Methods in Social Network Analysis*, ed. P. J. Carrington, J. Scott, and S. Wasserman, 98-116. Cambridge: Cambridge University Press.

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

Washington, G. T. and V. H. Rothschild (1962). *Compensating the Corporate Executive*, Volume 2. Ronald Press Co.

Yermack, D. (1997). Good timing: CEO stock option awards and company news announcements. *The Journal of Finance 52*(2), 449–476.

Young, H. P. (2003). The diffusion of innovations in social networks. In *The Economy as a Complex Evolving System, vol. III*. Edited by L. E. Blume and S. N. Durlauf. Oxford: Oxford University Press.

Young, H. P. (2009). Innovation Diffusion in Heterogeneous Populations: Contagion, Social Influence, and Social Learning. *American Economic Review 99*(5), 1899–1924.

APPENDIX

A.1. Executive compensation data for 1936–1991 from Frydman and Saks

The cornerstone of the analysis is based on Frydman and Saks's excellent and comprehensive database (2010) of executive compensation from 1936 until 1991, which they have made publicly available online.[14] For a sample of firms, the authors hand-collect compensation information from the 10-K reports and proxy statements submitted to the Securities and Exchange Commission (SEC).

There are 101 firms in the unbalanced panel tracked by Frydman and Saks. These firms are chosen primarily because during 1940, 1960, or 1990, they were one of the 50 largest American firms, as measured by total sales (for 1960 and 1990) or market value (for 1940). However, the authors do not include firms for which the data are sparse or sporadic, as explained below:

> "[W]e only use firms for which we can find information for at least 20 years in a 30-year window. These 30-year windows are 1936–1966 for the 1940 sample, 1943–1973 for the 1960 sample, and 1970–2000 for the 1990 sample. In addition, we also require that annual data must be available for at least three blocks of five consecutive years within this 30-year period. This requirement is necessary because only consecutive data on stock option grants and exercises can allow us to reliably estimate an individual's holdings of unexercised stock options. If a firm does not meet these criteria, we replace it with the next largest firm on the list. In this manner, we move down the rankings until we have a total of 50 firms for each list of rankings." (Frydman and Saks 2010, p. ??)

Further, the authors attempt to maintain a consistent identifier for firms even as the firms experience mergers or company name changes:

[14] Last accessed April 2013 at `http://people.bu.edu/cfrydman/Frydman_webpage/Data.html`.

> “Our intent is to keep a post-merger company in the sample if the new firm is similar to the original company. Therefore, we continue to follow a company for as long as the firm maintains the same permanent company identification number (PERMNO) in the CRSP database. We also include a post-merger firm with a different permanent number if either (1) all or part of the name of the old company is retained in the new company’s name, or (2) the 2-digit SIC code of the new and the old company are the same. Out of the 101 firms in our sample, there are seven cases where a firm’s identification number changes but it retains the name of the original firm, and 25 cases where the identification number and name changes but the industrial classification remains the same. There are 11 cases where we stop following a firm after a merger because the new firm takes on an entirely new name and operates in a different industry. There are also another 14 cases where we cease to follow a firm because it becomes foreign-owned (and therefore not subject to the Securities and Exchange Commission (SEC) reporting requirements) or because the firm has gone out of business.” (Frydman and Saks 2010, p. ??)

For each of these 101 firms, Frydman and Saks report compensation data for at least the top three executives in the firm. From 1936–1943, the SEC required compensation reports only for these top three executives at each firm. From 1943–1978, the SEC required reports for the top three executives *and* any employees receiving compensation exceeding a specified threshold in salary and bonuses (not including stock grants or options), which was increased over time. After 1978, the SEC required reports for the top five executives. For consistency, the analysis in this research is restricted to the top three executives at each firm.

A.2. Executive compensation data for 1992–2005 from ExecuComp

For the years from 1992 through 2005, Standard and Poor’s Computstat Executive Compensation (ExecuComp) database provides compensation data from the SEC’s 10-K reports and proxy statements similar to the pre-1992 data of Frydman and Saks (2010).[15]

[15] Available through Wharton Research Data Services, last accessed April 2013 at `http://wrds-web.wharton.upenn.edu/wrds/ds/comp/index.cfm`.

To maintain consistency with Frydman and Saks (2010), we only include firms from ExecuComp in the list of 101 firms selected as the top 50 largest firms in 1940, 1960, or 1990. Since this research concerns the *initial* adoption of options packages by firms and since this process of adoption is largely exhausted by 1990, including a different selection of top firms after 1990 does not alter the results.

A.3. Maximum likelihood estimation procedure

Here we describe the estimation procedures that were used to generate the results in Section 5. We explain the log-likelihood function, describe the grid search used in the optimization process, and conduct tests for GARCH errors.

The number of new adoptions in a given year t is a non-negative integer y_t. Poisson distributions are commonly used to model such count data but they require the mean to equal the variance, a requirement that is violated in this case: the mean of adoptions is 1.4, and the variance is 5.4. Using a likelihood ratio test for over-dispersion (Lawless 1987; Cameron and Trivedi 2013), we can reject the Poisson distribution with a p-value of 0.0002. Therefore, instead of the Poisson distribution, we used a negative binomial model, which allows a separate parameter (α^{-1}) to parameterize the variance. The intensity of the process is a scalar θ_t that depends on a normal CDF of prior adoptions with mean μ and variance σ^2, together with a vector x_t of other explanatory variables as explained in the text. Specifically we posit a model of the form

$$\theta_t = \beta \cdot x_t + \gamma \Phi\left(\int_0^t p(s)ds \mid \mu, \sigma\right). \qquad \text{(A1)}$$

Using maximum likelihood techniques we estimated the coefficients in this expression simultaneously with the variance and mean of Φ.

Out of the possible specifications of the negative binomial distribution, we used the 'NB2' model favored by Cameron and Trivedi (2005), in which the individual log-likelihood function is:[16]

$$\ln l_t(\alpha,\theta_t) = \ln\Gamma(\alpha^{-1}+\theta_t) - \ln\Gamma(\alpha^{-1}) - \ln\Gamma(\theta_t+1) + \alpha^{-1}\ln(\alpha^{-1}) \\ + y_t\ln(\theta_t) - (\alpha^{-1}+y_t)\ln(\alpha^{-1}+\theta_t)$$

To obtain the initial values for estimating the numeric maximum-likelihood estimation procedure, we first ran a grid search over the values displayed in Table A.1.[17] Given the initial values with the largest log-likelihood, a negative binomial model was estimated using maximum-likelihood procedures.

	Lower Bound	Upper Bound	Steps
α^{-1}	1	5	5
CDF of integral of previous adoptions (%): weight	-15	15	30
CDF of integral of previous adoptions (%): mean	-20	20	30
CDF of integral of previous adoptions (%): sd	0.05	10	15
Level of previous adoptions (%)	-10	10	10
Year (scaled)	-4	4	5

Table A.1. Bounds for the grid search. 'Steps' refers to the number of values tested between the lower and upper bounds.

We used the coefficients estimated by this procedure to model the probability distribution of the integer number of adoptions in each year, namely,

$$f(y_t = k \mid \alpha^{-1},\theta_t) = \frac{\Gamma(k+\alpha^{-1})}{\Gamma(k+1)\Gamma(\alpha^{-1})}\frac{(\alpha^{-1})^{\alpha^{-1}}}{(\alpha^{-1}+\theta_t)^{\alpha^{-1}}}\frac{\theta_t^k}{(\alpha^{-1}+\theta_t)^k}. \qquad \text{(A3)}$$

[16] See Cameron and Trivedi 2013, p. 70, Equation 3.26, or Cameron and Trivedi 2005, p. 675, Equation 20.12. Equivalent after simplification to Hilbe 2011, p. 199, Equation 8.39.

[17] Note that the initial values, identified by their log-likelihood, simply represent the numerical 'starting point' for maximum-likelihood estimation, which starts the estimation procedure close to its likely solution. Using randomly-selected initial values would give the same results, but would take more computing time.

Thus the expected total number of adoptions in year t is

$$\sum_{k=1}^{100} kf(y_t = k \mid \alpha^{-1}, \theta_t). \tag{A4}$$

This expression was used to derive the estimated adoption curve in Figure 5.

We also tested the residuals for GARCH errors. Using portmanteau white-noise tests on the squared residuals from the above model, we can accept the null hypothesis of uncorrelated heteroskedasticity for up to 5 lags . (The Box-Pierce and Ljung-Box portmanteau tests both give nearly identical results.) We therefore conclude that GARCH corrections are not needed in this case.

A.4. NAICS classification of firms in the sample

Using the North American Industrial Classification System (NAICS), we classify each of these 101 firms by sector (two-digit NAICS code) and sub-sector (three-digit NAICS code). Since these firms are overwhelmingly concentrated in manufacturing, we use the sub-sector to classify any firm in the manufacturing sector, and the sector to classify any firm outside the manufacturing sector. For analysis that specifically incorporates NAICS codes, only sectors with at least four firms in the manufacturing sub-sector or non-manufacturing sector are included (Table A3).

Sector	#
Transportation Equipment Mfg	12
Retail Trade	11
Finance and Insurance	10
Food Mfg	9
Petroleum and Coal Products Mfg	9
Chemical Mfg	7
Mining, Quarrying, and Oil and Gas Extraction	6
Information	5
Primary Metal Mfg	4
Utilities	4
Beverage and Tobacco Product Mfg	3
Paper Mfg	3
Plastics and Rubber Products Mfg	3
Computer and Electronic Product Mfg	2
Electrical Equipment Mfg	2
Professional and Scientific Services	2
Transportation and Warehousing	2
Wholesale Trade	2
Agriculture, Forestry, Fishing and Hunting	1
Fabricated Metal Product Mfg	1
Machinery Mfg	1
Nonmetallic Mineral Product Mfg	1
Non-classifiable Establishments	1

Table A3: Number of firms, out of the 101 in this sample, classified by NAICS sectors and manufacturing sub-sectors. For sector-based analysis, only sectors with at least four firms are included.