The Spread of Innovations through Social Learning

H. Peyton Young

The Brookings Institution, Washington, DC
Johns Hopkins University
University of Oxford

CSED Working Paper No. 43

December 2005

The author is indebted to Pierangelo de Pace, Kirk Moore, and Jon Parker for help in gathering and analyzing the data. The paper also benefited from comments by Robert Axtell, Andrew Felton, Robert Moffitt, Suresh Naidu, Kislaya Prasad, Matthew Shum, Thomas Valente, and Tiemen Woutersen. An earlier version of this work was presented to the Working Group on the Formation and Decay of Economic Networks at the Russell Sage Foundation in December 2003.

Abstract

Innovations often spread by the communication of information among potential adopters. In the marketing literature, the standard model of new product diffusion is generated by information contagion: agents adopt once they hear about the existence of the product from someone else. In social learning models, by contrast, an agent adopts only when the perceived advantage of the innovation -- as revealed by the actions and experience of prior adopters -- exceeds a threshold determined by the agent's prior beliefs. We demonstrate that learning with heterogeneous priors generates adoption curves that have an analytically tractable, closed-form solution. Moreover there is a simple statistical test that discriminates between this type of process and a contagion model. Applied to Griliches' classic results on the adoption of hybrid corn, this test shows that learning with heterogeneous priors does a considerably better job of explaining the data than does the contagion model.

1. Overview

The adoption of new technologies and practices frequently follows an S-curve: at first a few innovators adopt, then others hear about the idea and they adopt, and the process takes off -- first accelerating and later decelerating as the saturation level is reached. In the marketing literature this phenomenon is usually modeled as an information contagion process: people hear about the innovation from prior adopters, and they adopt -- possibly with a lag -- once they have heard about it. This is analogous to a model of disease transmission in which previously uninfected individuals become infected with some probability when they interact with people who are already infected. The contagion model generates S-shaped curves that have been fitted to a wide variety of data, particularly the adoption of new products (Bass, 1969, 1980).

In the economics literature, by contrast, the standard explanation for adoption is that individuals learn about an innovation either by directly observing its outcomes for others, or by inferring positive outcomes given the fact that others have adopted. Processes of this type are called *social learning models*.[1] Although there is a sizable literature on social learning -- both theoretical and empirical – its implications for the *shape* of the adoption curve have not previously been studied in any generality.

The purpose of this paper is to analyze adoption curves generated by social learning when agents have heterogeneous priors, and to draw a contrast with the curves generated by contagion models. Learning is a more complex process than contagion, because it involves two separate effects: as more people adopt,

[1] Some authors restrict the term "social learning" to situations in which agents make *indirect* inferences about the outcomes for other agents instead of observing the outcomes themselves. Here we focus on the case where agents learn by directly observing the outcomes for other agents or are told what the outcomes are.

more information accumulates that helps persuade the remaining people to adopt; in addition, however, the remaining people are inherently more pessimistic and hence harder to persuade.[2] The main theoretical contribution of the paper is to show that such a process can be modeled in a very general form by a family of differential equations with closed-form solutions. We formulate a simple nonparametric test that can differentiate between this class of curves and those generated by contagion. In particular, if the agents' prior beliefs are unimodally distributed, the relative acceleration of the adoption process will be nonlinear: rising in the early phases of adoption and declining in the later phases. This is quite different from a contagion model, which predicts that the relative acceleration decreases at a uniform rate throughout the process. In the second part of the paper we apply this test to Griliches' adoption curves for hybrid corn and show that they are more consistent with a learning than with a contagion model. The aggregate nature of the data means that we cannot *identify* learning as the cause of adoption. Nevertheless we can decisively reject the contagion model and demonstrate that the curves have nonlinear properties that are predicted by the learning model.[3]

2. Prior literature

The literature on innovation is very extensive, and includes work in economics sociology, and marketing. Here we shall touch on the relationship between our approach and the prior literature without attempting to provide a comprehensive view.[4]

---

[2] An epidemiological analog would be a disease such that the number of exposures must rise above an individual's exposure threshold before he can contract it. Most diseases have just the opposite property: multiple exposures usually *decrease* the likelihood of getting sick because immunity builds up.

[3] Of course we do not claim that learning from prior adopters is always the most appropriate model; Valente (1993) discusses a number of cases where other factors, such as the mass media, played a major role in the adoption dynamics.

In sociology there is a long tradition of studying the diffusion of innovations through the interactions among agents.[5] In this literature it is common to assume that each agent has an *adoption threshold,* that is, a minimum proportion of adopters that will induce him to adopt also.[6] Sociologists usually think of these thresholds as reflecting differences in responsiveness to social pressure or a desire to conform rather than as differences in beliefs, although they certainly could arise for the latter reason. While it has been argued informally that threshold heterogeneity could account for the S-shaped pattern of adoption curves (see especially Valente, 1995, 1996), the analytical implications for the shape of the adoption curve have not been worked out in any generality, as we shall do here.

In the economics literature, some form of learning is usually assumed to be the driving force behind adoption, but not necessarily with heterogeneity in beliefs (several exceptions will be discussed below). Some authors posit that agents know enough about the payoff structure of the situation to be able to make sophisticated inferences about the innovation's payoffs given that others have already adopted, as in the herding literature.[7] Other authors assume some form of boundedly rational learning in which people act on word-of-mouth information about the payoffs to prior adopters or the relative popularity of different choices among prior adopters.[8] In these models heterogeneity in beliefs does not drive the adoption process; moreover the focus of analysis is on the long-run stochastic dynamics and the conditions under which the long-run outcome is efficient, rather than on the shape of the adoption curve *per se.*

---

[4] For reviews of the various literatures see Mahajan and Peterson, 1985; Geroski, 2002; and Valente, 2005.

[5] The seminal paper is Ryan and Gross (1943). Rogers (2003) provides many empirical examples; for a recent review of the literature see Valente (2005).

[6] See among others Granovetter, 1978; Granovetter and Soong, 1983, 1986, 1988; Macy, 1990, 1991; Valente, 1995, 1996, 2005.

[7] See in particular Banerjee, 1992; Bikhchandani, Hirshleifer, and Welch, 1992; Smith and Sorensen, 2000; Gale and Kariv, 2003; Banerjee and Fudenberg, 2004; Munshi, 2004. Strategic issues are considered by Kapur, 1995.

There is, however, another branch of the economics literature in which heterogeneity does play an important role. In these models, potential adopters differ in some characteristic that establishes different *thresholds* at which they will adopt. For example, they may differ in their degree of risk aversion, or their pessimism about the prospects of the new technology, or (in the case of firms) their size. As time runs on, some key parameter -- such as the cost of production or amount of information -- changes and causes an increasing fraction of the population to adopt.[9]

Within this literature Jensen (1982) is the closest to the present paper. He posits that differences in beliefs are the source of heterogeneity, and then derives the consequences for the shape of the adoption curve explicitly. Specifically, he assumes that agents directly observe the realized payoffs of two competing technologies, one of which is better on average than the other. An agent irreversibly adopts one or the other technology as soon as his posterior estimate of the payoff difference reaches a critical level, which is determined by his prior belief. Assuming that the priors are uniformly distributed, Jensen shows that the resulting adoption curve is, in expectation, either S-shaped or concave. The principal difference between Jensen's framework and ours is that we derive results for any distribution of priors -- in fact for any distribution of adoption thresholds, whether generated by priors or otherwise. We also formulate a simple statistical test that can distinguish between this model and standard contagion models.

---

[8] Kirman, 1993; Arthur, 1989; Ellison and Fudenberg, 1993, 1995; Bala and Goyal, 1998; Chatterjee and Hu, 2002.

[9] See David, 1975; Davies, 1979; Stoneman, 1981; Jensen, 1982, 1983; Balcer and Lippman, 1984. There is also an empirical literature on social learning that has been largely concerned with the problem of disentangling learning from other factors, such as spatial correlation of outcomes, similarity among neighbors, and so forth (Besley and Case, 1994; Foster and Rosenzweig, 1995; Conley and Udry, 2003; Munshi and Myaux, 2005). This approach requires micro-level data, whereas ours uses macro-level data on the shape of the adoption curve.

3. Contagion models

We begin by recalling the basic features of contagion models, after which we shall contrast them with social learning models. Consider a group of n individuals who are exposed to a new idea, technology, or practice beginning at time $t = 0$. Let $p(t)$ denote the *proportion* of the group who have adopted the idea by time t. Suppose that at time $t = 0$ a nonempty subgroup has heard about the innovation through external sources, that is, $p(0) > 0$. In its simplest form, the contagion model posits that in each period the probability that a given individual adopts for the first time is proportional to the number who have already adopted up to that time. In expectation this leads to a discrete-time process of form

$$p(t+1) - p(t) = \alpha p(t)(1 - p(t)). \tag{1}$$

where $\alpha \in (0, 1)$. This process can be motivated as follows: in the $(t+1)^{st}$ period each individual i who has not yet adopted meets someone at random. Assuming uniform mixing in the population, the person he meets is a prior adopter with probability $p(t)$. This leads agent i to adopt in the current period with probability α, that is, $1 - \alpha$ is the *degree of inertia* in decision making. Since the current proportion of non-adopters is $1 - p(t)$, the expected number of converts in the $(t+1)^{st}$ period is given by (1). The continuous-time analog is

$$\dot{p}(t) = \lambda p(t)(1 - p(t)), \tag{2}$$

where $\lambda \in (0, 1]$ is the instantaneous conversion rate. The solution is the logistic function

$$p(t) = 1/[1 + ce^{-\lambda t}], \quad c = -1 + 1/p(0). \tag{3}$$

We shall call this the *simple contagion model*. The resulting adoption curve is decidedly S-shaped, with the particular feature that it is symmetric about p = .5 for all λ, a fact that can be used to test its empirical plausibility.

A useful generalization is the following model due to Bass (1969, 1980). Suppose that people hear about the innovation partly from internal sources and partly from external sources (see also Lekvall and Wahlbin, 1973). Let λ be the instantaneous conversion rate when the information comes from other members of the group, and let γ be the instantaneous conversion rate when the information comes from outside the group. We then obtain the differential equation

$$\dot{p}(t) = \lambda\, p(t)(1 - p(t)) + \gamma(1 - p(t)). \tag{4}$$

Assuming that λ and γ are both positive, the solution is

$$p(t) = [1 - \beta\gamma e^{-(\lambda+\gamma)t}]/[1 + \beta\gamma e^{-(\lambda+\gamma)t}], \tag{5}$$

where β is a positive constant. If $p(0) = 0$, then $\beta = 1/\gamma$ and we obtain

$$p(t) = [1 - e^{-(\lambda+\gamma)t}]/[1 + (\lambda/\gamma)\, e^{-(\lambda+\gamma)t}]. \tag{6}$$

This is known as the *Bass model* of product diffusion (Bass, 1969, 1980). Note that when $\gamma = 0$ and $\lambda > 0$, we obtain the simple contagion model of equation (3).

When $\lambda = 0$ and $\gamma > 0$, the solution is the negative exponential distribution

$$p(t) = 1 - e^{-\gamma t}. \tag{7}$$

This situation arises if the non-adopters hear about the innovation with fixed probability γ in each period. It also arises if everyone has already heard about the innovation but they act with a probabilistic delay. Note, however, that this *pure inertia model* generates an adoption curve that is concave throughout, not S-shaped. This provides a straightforward test of the model's plausibility when empirical data on adoption rates is available.

4. Social learning with heterogeneous thresholds

The difficulty with the contagion model is the assumption that people adopt an innovation simply because they have heard about its existence. A more plausible hypothesis is that they weigh its benefits before adopting, where information about benefits comes from the experience of prior adopters. In a social learning model, a potential adopter looks at the innovation's realized performance among prior adopters, and combines this revealed payoff information with any prior information he may have in order to reach a decision. Although this is similar in spirit to other social learning models in economics, our approach differs in two important respects from much of the literature (see the discussion in section 2 above). First, we assume that agents can (and typically do) have different prior information, and hence require different thresholds of evidence before they adopt. Second, we do not assume that agents know the distribution of prior beliefs or have any other basis on which they can reasonably *infer* payoff information from the decision of other agents to adopt. Instead, we shall assume that agents can directly observe the realized payoffs of prior adopters.[10]

[10] The theory discussed in sections 4-6 is developed more extensively in Young (2004).

Specifically, we posit that each individual i has a critical value such that if the observed payoffs of the innovation are high enough among enough other people, then agent i will adopt also. Let $h(\alpha, m)$ be a real-valued function that is strictly increasing in both α and m, where α is the number of people who have already adopted and m is the mean payoff advantage of the innovation (relative to the status quo) among the adopters. For analytical convenience we shall assume that α can be any nonnegative real number (representing the "size" of the adopting population).

For each agent i, assume that a real number θ_i exists such that i is "ready to adopt" if and only if $h(\alpha, m) \geq \theta_i$. θ_i is i's *critical value*. We say "ready to adopt" because we shall eventually want to build inertia into the process: an individual who is ready to adopt may nevertheless fail to adopt immediately. Fix a population size n and a payoff advantage m. For each agent i define

$$r_i(m, n) = \inf \{p \in [0, 1]: h(pn, m) \geq \theta_i\}$$
$$r_i(m, n) = \infty \text{ if there exists no such p.} \qquad (8)$$

The number $r_i(m, n)$ is i's *resistance* to adoption, also known as i's *adoption threshold*. It is the minimum proportion of the population such that i is ready to adopt once that many other people have adopted.

Let us illustrate these ideas with a concrete example. Suppose that the innovation in question has a random payoff X_i to agent i, where X_i is normally distributed $N(m, \sigma^2)$. Assume that the status quo has zero payoff, hence m represents the *expected payoff advantage* of the innovation, which we assume to be positive. Individuals have heterogeneous prior beliefs about the value of m. Suppose, in particular, that agent i thinks that m is normally distributed in the population with mean μ_i^0 and standard deviation σ_i. Here μ_i^0 may be positive or negative: in the former case i is *optimistic* whereas in the latter case he is *pessimistic*. Let us assume temporarily that there is no inertia: individuals adopt as soon as their updated beliefs lead them to think that the innovation is at least as good as the status quo. Thus, in the first period, only the optimists adopt. In the second period, those who were slightly pessimistic to begin with see the outcomes (which are positive in expectation) among the initial adopters, in which case they too adopt.

In period t, let m_t be the realized mean payoff among all those who have adopted by period t, and let p(t) be the proportion who have adopted by t. Under suitable assumptions on i's beliefs about the variance of m, the posterior estimate of m is a convex combination of the observed mean and the prior mean of form

$$\frac{np(t)m_t + \tau_i\mu_i^0}{np(t) + \tau_i}, \qquad (9)$$

where n is the number of people in the group and τ_i is a positive number that depends on i's beliefs.[11] As we would expect, agent i gives increasing weight to the observed mean as the number of observations np(t) becomes large.

Assume that each agent is adopts in the first period t such that his posterior estimate is nonnegative. Our interest is in the expected motion of the process, hence we shall ignore variability in the realizations of m_t and work with its expectation m. Thus, in expectation, agent i adopts at time t + 1 if and only if

$$\frac{np(t)m + \tau_i\mu_i^0}{np(t) + \tau_i} \geq 0. \qquad (10)$$

[11] For example, (9) holds in a normal-normal updating framework (de Groot, 1970, Chapter 9). We assume that multiple observations of the same individual do not convey additional information to potential adopters, that is, the uncertainty lies in the distribution of payoffs among individuals, not in their distribution among repeated trials by a *given* individual.

In this case the function h takes the especially simple form $h(\alpha, m) = \alpha m$, where $\alpha = p(t)n$ is the number of adopters at time t and i's critical value is $\theta_i = -\tau_i \mu_i^0$. Hence i's resistance $r_i = r_i(m, n)$ is

$$\begin{array}{ll} r_i = 0 & \text{if } \mu_i^0 > 0 \\ r_i = -\tau_i \mu_i^0/nm & \text{if } 0 \leq -\tau_i \mu_i^0/nm \leq 1 \\ r_i = \infty & \text{if } -\tau_i \mu_i^0/nm > 1 \end{array} \qquad (11)$$

In other words, i is ready to adopt if and only if either i is an optimist (i's resistance is zero), or i is initially a pessimist but changes his mind once the proportion of adopters is at least $r_i = -\tau_i \mu_i^0/nm$.

Let us return now to the general situation. Fix the population size n and the payoff advantage of the innovation m, which we assume to be positive. Denote agent i's resistance by $r_i = r_i(m, n)$. It follows from the definition of r_i that i first adopts at time t + 1 if and only if

$$p(t-1) < r_i \leq p(t). \qquad (12)$$

Let F(r) be the cumulative distribution function of the resistance parameter r in the population. We then obtain the discrete-time process

$$\forall\, t \geq 0, \qquad p(t+1) = F(p(t)). \qquad (13)$$

We can easily extend the model by introducing an inertia parameter. Let F(r) be the distribution function of r. At time t + 1, F(p(t)) – p(t) is the proportion of individuals who are prepared to adopt (because their resistance has been overcome), but they may not do so immediately out of inertia. These are the *susceptible* individuals at time t + 1. Suppose that each susceptible individual adopts with probability α in the current time period. This leads to the discrete-time process

$$p(t+1) = \alpha[F(p(t)) - p(t)] + p(t). \tag{14}$$

The continuous-time analog is a differential equation of form

$$\dot{p}(t) = \lambda\, [F(p(t)) - p(t)]. \tag{15}$$

where $\lambda > 0$. Assume that $F(0) > 0$ and $p(0) = 0$. Let

$$b = \min\{r: F(r) \leq r\}. \tag{16}$$

Then (15) is a separable ODE with the solution

$$\forall\, x \in [0, b),\ \ t = p^{-1}(x) = 1/\lambda \int_0^x dr/(F(r) - r). \tag{17}$$

We assume here that $p(0) = 0$, hence the constant of integration is zero. Note that if $F(r)$ is any c.d.f. satisfying $F(r) > r$ everywhere on the interval $[0, b)$, then (17) can be integrated to obtain the function $p^{-1}(x)$, which uniquely determines $p(t)$ provided that $p^{-1}(x)$ is strictly increasing. The differential equation (15) and its solution (17) will be called a *heterogeneous resistance model with distribution function* F *and inertia parameter* $1/\lambda$. It applies to any process that is driven by heterogeneous adoption thresholds, whether or not they arise from *ex ante* differences in beliefs.

5. Examples

To illustrate, consider the normal-normal learning model and suppose that the numbers $\theta_i = -\tau_i\mu_i^0$ are normally distributed. Then the resistances have a truncated normal distribution with point masses at 0 and 1, corresponding to the optimists and ultra-pessimists respectively. The adoption curve takes the form

$$\forall x \in [0, b),\ t = p^{-1}(x) = 1/\lambda \int_0^x dr/(N((r-\mu)/\sigma) - r), \tag{18}$$

where N is the cumulative standard normal distribution, μ is the mean and σ^2 the variance. Figure 1 illustrates the case where $\mu = .20$ and $\sigma = .10$. The point mass at r = 0 represents the optimists who propel the adoption process forward initially; there is also a tiny point mass at r = 1 (not shown here) that represents the individuals who are so pessimistic they will never adopt. The middle panel shows the cumulative distribution function. The bottom panel shows the adoption curve generated by this distribution when $\lambda = 1$. (Notice that λ determines the time scale, but does not alter the shape of the trajectory, so there is no real loss of generality in fixing its value.)

6. Structural characteristics of adoption curves generated by learning

We have seen how to derive the adoption curve from the distribution of resistances and the degree of inertia. In practice, however, we often want to go in the other direction and test whether a given adoption curve is consistent with a given class of distribution functions and inertia levels. There is a straightforward way to do so.

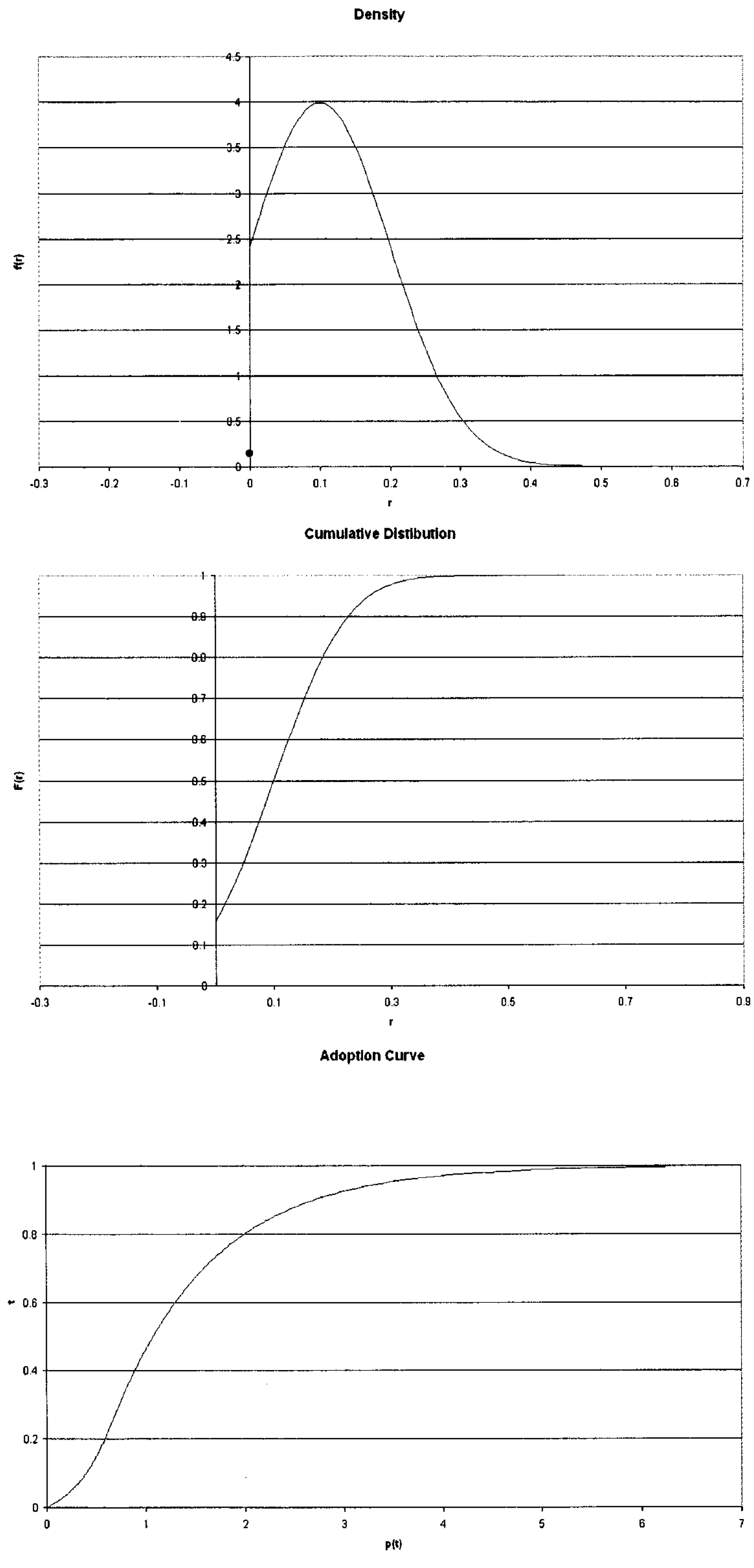

Figure 1. Normal distribution N(.10, .01): truncated density, cumulative distribution and adoption curve with $\lambda = 1$.

Let p(t) be an empirically observed adoption curve over the time interval $0 \le t \le T$. For analytical convenience assume that p(t) is strictly increasing and twice differentiable. If p(t) is generated by a social learning model, then it satisfies equation (15) for some unknown cumulative distribution function F(r). Differentiating (15) we obtain

$$\forall t \in [0, T], \qquad \ddot{p}(t) = \lambda\ [f(p(t))\ \dot{p}(t)\ - \dot{p}(t)]. \qquad (19)$$

By assumption p(t) is strictly increasing, so $\dot{p}(t) > 0$ and we obtain

$$\forall t \in [0, T], \qquad \ddot{p}(t)/\dot{p}(t) = \lambda[f((p(t))\ - 1]. \qquad (20)$$

Assume that p(0) = 0 and let b = sup $\{p(t): 0 \le t \le T\}$. For every r in the interval [0, b), let g(r) be the relative acceleration rate when r is the proportion of the population that has already adopted, that is,

$$\forall\ r \in [0, b), \qquad g(r) = \ddot{p}(t_r)/\dot{p}(t_r) \text{ where } p(t_r) = r. \qquad (21)$$

By assumption p(t) is strictly increasing, so t_r is uniquely defined. Equations (20) and (21) imply that *the relative acceleration function* g(r) *is a positive linear transformation of the unobserved density function* f(r), that is, for all r in some subinterval [0, b),

$$g(r) = \lambda[f((r)\ - 1]. \qquad (22)$$

Note that g(r) is an *observable*: it can be computed directly from the slope of the adoption curve. Note also that g(r) does not depend on time as such; rather it is the acceleration rate at the time when the adoption level is r. Equation (22) shows that g(·) can be used to estimate the density function generating the data. We claim further that g(·) can be used to evaluate the likelihood of the alternative, namely, the contagion (or Bass) model. To see why, recall from (4) that the Bass model is defined by the differential equation

$$\dot{p}(t) = (\lambda\, p(t) + \gamma)(1 - p(t)). \tag{23}$$

Taking logarithms of both sides and differentiating with respect to t, we obtain

$$g(r) = \lambda - \gamma - 2\lambda r. \tag{24}$$

This is a linear decreasing function of r. *Hence we can reject the contagion model if we can reject the linearity of the observed function* g(r).[12]

If the data are generated by social learning with a normal distribution of critical values θ_i, then g(r) will be nonlinear. In this case there are three possible shapes for g(r), depending on where the neutral value $\theta = 0$ falls in the distribution. Recall that, in general, the tails of a normal density are strictly convex, while the part between the tails is strictly concave. (These properties also hold for a variety of other unimodal distributions, including the lognormal, the t, and the Cauchy distribution.) Thus we have the following:

---

[12] A comparison of (22) and (24) reveals that the Bass model is a special case of a social learning model in which the density of resistances is a straight line with negative slope. While conceivable, this would be a most unusual distribution; in any event it is rejected outright for the data on hybrid corn, as we shall show in section 7.

1. If $\theta = 0$ lies in the right tail of the normal density, then g(r) is convex for all r.

2. If $\theta = 0$ lies between the two tails, then g(r) is concave for smaller values of r and convex for larger values of r.

3. If $\theta = 0$ lies in the left tail, then g(r) is convex for smaller values of r, concave for intermediate values of r, and convex for larger values of r.

7. The diffusion of hybrid corn

One of the most carefully documented examples of innovation diffusion is Griliches' study of hybrid corn (Griliches, 1957). Using extensive unpublished data collected by the Field Crop Statistics Branch of the Agricultural Marketing Service, he was able to show that regional differences in the rate of diffusion were related to differences in hybrid corn's potential profitability relative to traditional varieties. He also drew attention to the strongly S-shaped pattern of diffusion in almost all of the regions he studied.

Griliches' study is similar to ours in that he deduced certain qualitative features of the individual learning process from the overall *shape* of the adoption curves: in particular, he showed that in regions where the expected payoff from adopting hybrid corn was high, the rate of adoption by individuals tended to be high also. However, Griliches' approach differs from ours in several important respects. First, he fitted logistic functions to the data, whereas we will use a nonparametric estimation procedure.[13] Second, he did not provide a decision-theoretic model of the individual adoption process, whereas we are able to compare the implications of two alternative models -- learning with

[13] Subsequently Dixon (1980) re-analyzed Griliches' data by fitting them to Gompertz functions instead of logistic functions; he also used a longer data series. Although this produced a better fit, the conclusion remained intact that the speed of adoption is positively correlated with the expected payoff advantage of hybrid over traditional corn.

heterogeneous thresholds and contagion. Of course, such an approach does not *identify* learning as the driving force behind adoption; to go this far we would need micro-level data on individual adoption behavior conditional on others' behavior (controlling for common effects). Griliches' data do not permit such an analysis. Nevertheless, they are sufficient to distinguish statistically between contagion and learning at the macro-level. [14]

Figure 2 shows the percentage of corn acreage planted in hybrid corn, by crop reporting district, for a group of corn belt states in each year during the period 1933-1952. They were retrieved from Griliches' papers in the Harvard University Archives.[15] Unfortunately not all of the original data could be located, but we do have results for seven of the twelve states in the corn belt: Ohio, Indiana, Michigan, Illinois, Wisconsin, Iowa, and Kansas.[16] Each of these states contains nine reporting districts, which yields 63 distinct adoption curves for the analysis.

Some qualitative differences among the states are worth noting. Kansas and Michigan are distinguished by the fact that in several districts the rate of acceleration was significantly slower and choppier than the norm. (In Michigan, these were the districts in the northern part of the states and in Kansas they were in the southwestern part of the state; in both cases these are the least desirable areas for growing corn.) Ohio and Illinois present another interesting contract: In Ohio the district curves rise rapidly and nearly in tandem, whereas in Illinois they all rise rapidly but some lag behind others by as much as 4-5 years.

[14] Moreover, we could not find any data sets that are more suitable for this purpose, because statistical significance will require an adoption curve with a great many observations, or many adoption curves running in parallel (as in Griliches' case).

[15] I am indebted to Diane Asseo Griliches for giving me access to her late husband's papers.

[16] The "corn belt," consists of the twelve states Ohio, Indiana, Michigan, Illinois, Wisconsin, Minnesota, North Dakota, South Dakota, Iowa, Missouri, Kansas, and Nebraska (United States Department of Agriculture, 1952).

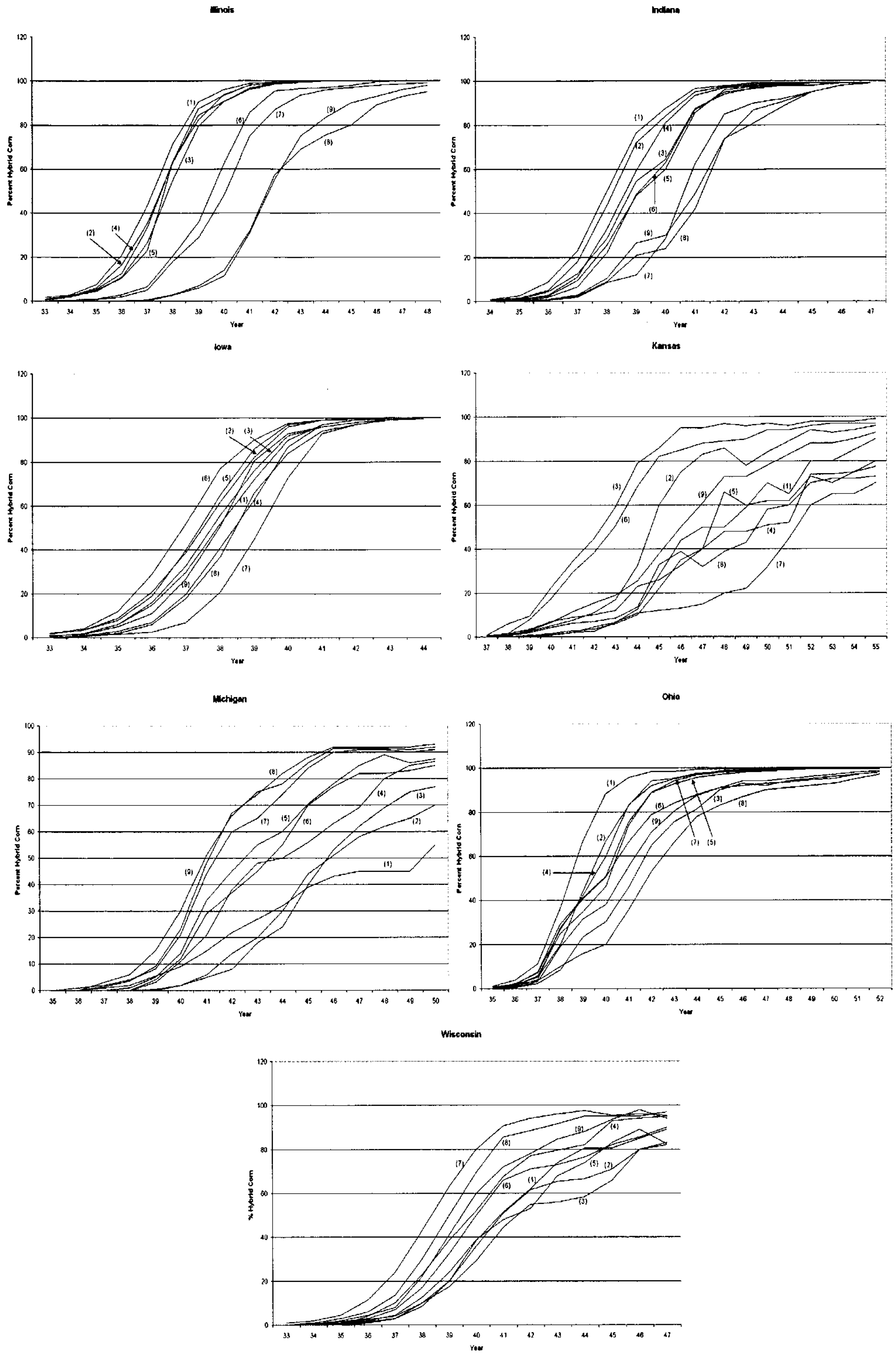

Figure 2. Adoption curves for hybrid corn in seven states, by crop reporting district: 1933-1952.

Overall, however, the great majority of the curves have the characteristic S-shape, with a sharp acceleration in the early phases of adoption.

We now turn to the statistical estimations. Let x_{it} be the percentage of hybrid corn planted in state i in year t. The absolute rate of change from t to t + 1 is given by $\Delta_{it} = x_{it+1} - x_{it}$. Let

$$y_{it} = \Delta_{it}/\Delta_{it-1} \text{ if } \Delta_{it-1} > 0,$$

(If $\Delta_{it-1} = 0$, y_{it} is undefined; we do not include these points in the analysis.) Then y_{it} - 1 estimates the relative rate of acceleration at time t. The contagion model implies that for each i, $y_{it} - 1$ is a decreasing linear function x_{it}, hence the same holds for y_{it}. The social learning model, by contrast, implies that $y_{it} - 1$ (and hence also y_{it}) is a positive linear transformation of some unknown density $f(x_{it})$.

To test the plausibility of the contagion model, we first pool the data from all regions and estimate y as a quadratic function of x. Note that the quadratic term is different from zero at an extremely high level of significance. Performing the same analysis on each state separately we find that linearity is rejected at the 1% level for five out of the seven states (see Table 3 in the Appendix). Overall, we can therefore reject the contagion model with a high degree of confidence.

Table 1. Pooled data: OLS estimation of $g(x) = c(1) + c(2)x + c(3)x^2$.

	Coefficient	Std. Error	t-Statistic
C(1)	2.413073	0.062948	38.33456
C(2)	-0.048165	0.004232	-11.38173
C(3)	0.000313	4.89E-05	6.397485

Next let us consider the social learning model. We begin by estimating the shape of the acceleration function y = g(x) nonparametrically using the k nearest neighbors method (Härdle, 1990, Chapter 3). To control for outliers we use the median value of the nearest k neighbors instead of the mean. The results for the pooled data are shown in Figure 3. (We used k = 77, which is approximately a 10% span.)[17]

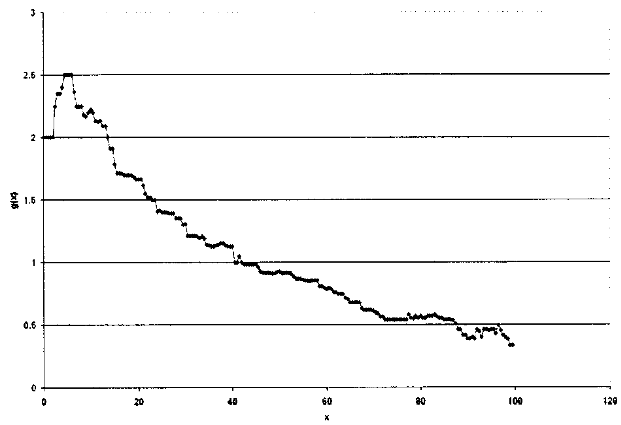

Figure 3. Relative acceleration rate g(x) as a function of adoption level x. Median of 77 nearest neighbors (span = 10%). Pooled data.

The nonparametric fit supports our previous finding that g(x) is not a linear function of x. Specifically, the curve is concave for low values of x: first it increases, then it decreases at a decreasing rate. To check this pattern parametrically, we divided the data into two disjoint subsamples: A, $0.5 \le x \le 20$ and B, $20 < x \le 90$.[18] We then estimated a quadratic model separately on each subsample by OLS after removing outliers.[19] The estimated quadratic coefficient on subsample A is negative at the 1.1% level of significance, while on subsample

[17] Shorter spans produce rather choppy fits, whereas longer spans lose some of the resolution needed to estimate the early part of the curve.

[18] Adoption levels of less than 0.5% were discarded because the estimation errors are very large. Adoption levels above 90% were not included because this censors outcomes in regions that had not reached full penetration by the end of the study period. Splitting the subsamples at 20% has no particular significance; other values between 15% and 25% could have been used as well.

[19] Outliers were defined as points lying more than two standard deviations from the mean as determined by a locally estimated model.

B it is positive at the 3.7% level of significance (see Table 2). This lends further support to the hypothesis that the curve is concave for low values of x and convex for high values of x.

Table 2. Pooled data: quadratic estimation on two subsamples.

Subsample A: $0.5 \le x \le 20$

	Coefficient	Std. Error	t-statistic	1-sided p-value
C(1)	2.166858	0.136397	15.88636	1
C(2)	0.056931	0.041629	1.367586	0.9136
C(3)	-0.005204	0.002247	-2.315373	0.01075

Subsample B: $20 < x \le 90$

	Coefficient	Std. Error	t-Statistic	1-sided p-value
C(1)	1.778686	0.225581	7.884897	0
C(2)	-0.02594	0.008685	-2.98668	0.9985
C(3)	0.000136	7.54E-05	1.798211	0.03655

A similar pattern emerges when the seven states are analyzed separately. Figure 4 shows the nonparametric curves estimated by the nearest neighbors method, again using the median instead of the mean to control for outliers. (Here k = 11, which represents a span of about 10%.) From the figure we see that five of the seven states -- Ohio, Indiana, Illinois, Iowa, and Wisconsin -- exhibit an increasing then decreasing pattern for low values of x. Indiana, Illinois, and Wisconsin have a convex appearance for x > 20, while Iowa appears to be almost linear. Ohio peaks unusually early, and decreases very sharply thereafter. Kansas and Michigan do not have a marked concave shape initially, and decrease in more nearly linear fashion.

Next we test for concavity and convexity separately in each of the seven states. Namely, for each state s let us choose a dividing point x_s^* and test for concavity below x_s^* and for convexity above x_s^*. (In the following we let $x_s^* = 20\%$ in all states except Ohio, where x_s^* = 10%. Even better results can be obtained by tailoring the dividing points more finely on a state-by-state basis.) Table 4 in the Appendix shows that below the dividing points the curves are concave -- the estimated quadratic coefficients are negative -- in six out of seven states. They are significantly negative at the 5% level in two states (Illinois and Iowa), and at the 7% level or better in three states (Illinois, Iowa, and Ohio). Above the dividing points the curves are convex -- the estimated quadratic coefficients are positive -- in six out of the seven states, and they are significantly positive at the 5% level or better in three states (Kansas, Illinois, and Wisconsin).

Although not all of the values for individual states are statistically significant, they are highly significant as a group.[20] To see why, let us compute the *average* t-value of the quadratic coefficients on each of the subsamples: below the dividing points the average t-value is -1.2004, while above the dividing points the average t-value is 1.0347. For each of the state estimates the t-statistic has over 30 degrees of freedom, hence they are very close to being normal with mean 0 and variance 1. Assuming independent realizations, it follows that the average t-values are very close to being normal with mean 0 and variance 1/7. Hence the average t-value below the dividing points represents about $1.2004\sqrt{7} = 3.176$ standard deviations, which is significantly negative (in a one-sided test) at the 1% level. Similarly, the average t-value above the dividing points is about $1.0347\sqrt{7} = 2.738$ standard deviations, which is significantly positive at the 1% level.

[20] In general, given n realizations of independent t-statistics, we cannot expect all of them to have an equally high level of significance, but we can apply a test of significance to their average.

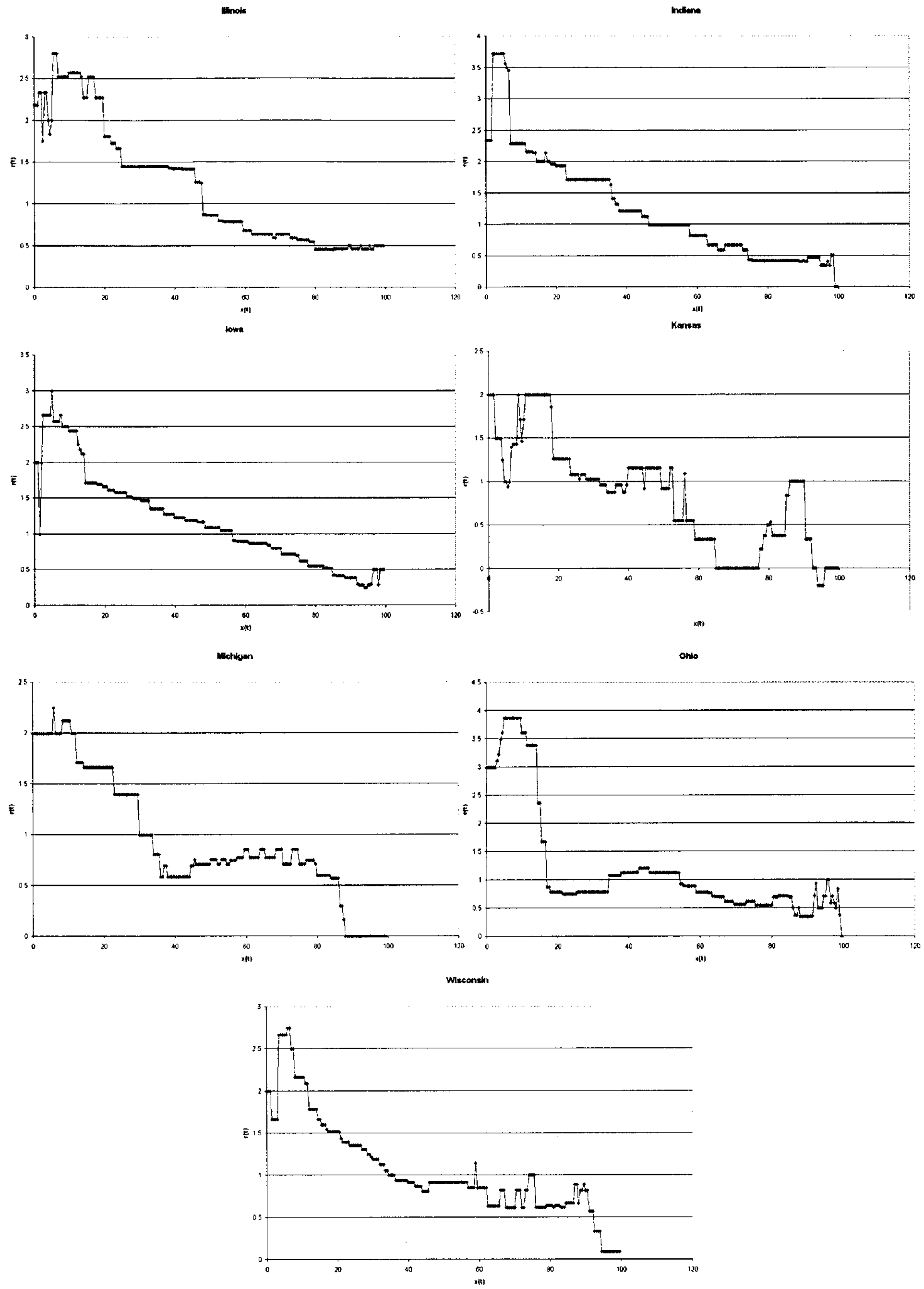

Figure 4. Relative acceleration rate g(x) as a function of adoption level x. Median of 11 nearest neighbors (span = 10%). State-level data.

Overall, therefore, the state-level data are consistent with a social learning model that is generated by unimodal distributions of resistances -- possibly differing across states -- where the modal levels of resistance lie somewhere between 0 and 20%. The state-level data are not consistent with the contagion model, which is rejected at the 1% level in five of the seven states.

8. Social learning or social conformity?

The preceding analysis is predicated on a model of social learning with heterogeneous beliefs. However, it could also describe an adoption process that is driven by heterogeneity in other characteristics of the agents, such as a desire to conform, or responsiveness to social pressure. As mentioned earlier, these types of explanations are standard in the sociology literature (see among others Ryan and Gross, 1943; Rogers, 1983; Granovetter, 1978; Granovetter and Soong, 1983, 1986, 1988; Macy, 1990, 1991; Valente, 1993, 1995, 1996). In the present case the data allow us to distinguish between these explanations and social learning based on realized payoffs. The reason is that, *if the adoption dynamics were driven purely by conformity, the payoff advantage from adopting hybrid corn should not matter for the shape of the adoption curve.* But Griliches' analysis, as well as Dixon's subsequent re-analysis, shows that it does matter (Griliches, 1957; Dixon, 1980). In particular, their analysis shows that the speed with which adoption occurs within a given region increases with the expected gain in profitability from adopting hybrid corn in that region. (This follows from their estimation of the regional coefficients in the logistic and Gompertz functions respectively as a function of regional profitability.) [21]

---

[21] Under the logistic function (3), the elapsed time t that the process takes to go from an adoption level x to some higher level y is $t = (1/\lambda)\ln[(1/x - 1)/(1/y - 1)]$. Griliches showed that λ tends to be higher in those regions where farmers could expect large gains in profitability by planting hybrid corn instead of traditional open pollinated varieties. Subsequent studies of technology adoption in agriculture have found, using micro-level data, that high realized payoffs in a given farmer's reference group increases his propensity to adopt (Foster and Rosenzweig, 1995; Conley and Udry, 2003; Munshi, 2004).

The same conclusion follows readily from our model with virtually no assumptions about the parametric form of the distribution of beliefs. Suppose that there is a common distribution of prior beliefs about the value of the mean across regions, but that the *actual* mean differs among regions. Recall from (8) that agent i's resistance $r_i(m)$ is a function of i's prior beliefs and the true mean m, where by assumption $r_i(m)$ is strictly decreasing in m. (In the normal-normal case $r_i(m) = (1/m)r_i(1)$.) Consider two different regions, 1 and 2, with means $m_1 < m_2$. Let $F_i(r)$ be the resulting distribution of resistances in region i, and assume that both are strictly increasing in r. Then F_2 represents an upward shift relative to F_1, that is, $F_2(r) > F_1(r)$ for all r. (In the normal-normal case, $F_2(r) = F_1((m_2/m_1)r) > F_1(r)$.) Now consider any two realized levels of adoption $x < y$, and let $T_i(x, y)$ be the time it takes in region i to go from x to y. From (17) it follows that

$$T_i(x, y) = 1/\lambda \int_x^y dr/(F_i(r) - r). \tag{25}$$

From the preceding discussion we know that the denominator of (25) is larger for i = 2 than for i = 1 for all $r \in [x, y]$, hence $T_1(x, y) > T_2(x, y)$. In other words, it takes less time in expectation to go from adoption level x to a higher level y when the payoff from the innovation is larger. This is effectively what Griliches (and Dixon) showed: the rate of acceleration of the adoption curve is positively correlated with the payoff advantage of the innovation. While we cannot rule out the possibility that social pressure or conformity played some role, it seems unlikely that they were the sole explanation for the S-shaped adoption pattern of hybrid corn.

References

Arthur, W. Brian (1989), "Competing technologies, increasing returns, and lock-in by historical events," *Economic Journal*, 99, 116-131.

Balcer, Y., and S. Lippman (1984). "Technological expectations and adoption of improved technology," *Journal of Economic Theory*, 34, 292-318.

Banerjee, Abhijit, and Drew Fudenberg (2004), "Word-of-mouth learning," *Games and Economic Behavior* 46, 1-22.

Bass, Frank (1969), "A new product growth model for consumer durables," *Management Science*, 15, 215-227.

Bass, Frank (1980). "The relationship between diffusion rates, experience curves and demand elasticities for consumer durables and technological innovations," *Journal of Business*, 53, 551-567.

Besley, Timothy, and Anne Case (1994). "Diffusion as a learning process: evidence from HYV cotton. Princeton University, RPDS Discussion Paper #174.

Chatterjee, Kalyan, and Susan Hu (2004). "Technology diffusion by learning from neighbors," *Advances in Applied Probability*, 36, 355-376.

Conley, Timothy G., and Christopher R. Udry (2003), "Learning about a new technology: pineapple in Ghana," Working Paper, University of Chicago and Yale University.

Davies, S. (1979). *The Diffusion of Process Innovations*. Cambridge UK: Cambridge University Press.

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

DeGroot, Morris (1970), *Optimal Statistical Decisions.* New York: McGraw Hill.

Dixon, Robert (1980), "Hybrid corn revisited," *Econometrica* 48, 14511461.

Ellison, Glenn, and Drew Fudenberg (1993), "Rules of thumb for social learning," *Journal of Political Economy*, 101, 612-643.

Ellison, Glenn, and Drew Fudenberg (1995), "Word of mouth communication and social learning," *Quarterly Journal of Economics*, 110, 93-125.

Foster, Andrew and Mark Rosenzweig (1995), "Learning by doing and learning from others: human capital and technical change in agriculture," *Journal of Political Economy*, 103, 1176-1209.

Gale, Douglas, and Shachar Kariv (2003). "Bayesian learning in social networks," *Games and Economic Behavior*, 45, 329-346.

Geroski, P. A. (2000), "Models of technology diffusion," *Research Policy*, 29, 603-625.

Griliches, Zvi (1957), "Hybrid corn: an exploration of the economics of technological change," *Econometrica*, 25, 501-522.

Härdle, W. (1990), *Applied Nonparametric Regression.* Cambridge UK and New York: Cambridge University Press.

Jensen, Richard (1982), "Adoption and diffusion of an innovation of uncertain profitability," *Journal of Economic Theory*, 27, 182-193.

Jensen, Richard (1983), "Innovation adoption and diffusion when there are competing innovations," *Journal of Economic Theory*, 29, 161-171.

Kapur, Sandeep (1995), "Technological diffusion with social learning," *Journal of Industrial Economics*, 43, 173-195.

Kirman, Alan (1993), "Ants, rationality, and recruitment," *Quarterly Journal of Economics*, 108, 137-156.

Mahajan, V., and R. Peterson, (1985), *Models of Innovation Diffusion*. Beverly Hills CA: Sage Publications.

Munshi, Kaivan (2004), "Social learning in a heterogeneous population: technology diffusion in the Indian green revolution," *Journal of Development Economics*.

Munshi, Kaivan, and Jacques Myaux (2005), "Social norms and the fertility transition, Working Paper, Department of Economics, Brown University. Forthcoming in the *Journal of Development Economics.*

Rogers, Everett (2003), *Diffusion of Innovations.*, 5th edition. New York: Free Press.

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

Stoneman, P. (1981). "Intra-firm diffusion, Bayesian learning, and profitability," *Economic Journal* 91, 375-388.

United States Department of Agriculture, (1952), *Handbook of Agriculture.* Washington, D.C.: U.S.Government Printing Office.

Valente, Thomas W. (1993), "Diffusion of innovations and policy decision-making," Journal of Communication 43, 30-45.

Valente, Thomas W. (1995), *Network Models of the Diffusion of Innovations*. Cresskill NJ: Hampton Press.

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

Valente, Thomas W. (2005), "Network models and methods for studying the diffusion of innovations," in Peter J. Carrington, John Scott, and Stanley Wasserman (eds.), *Recent Advances in Social Network Analysis*. Cambridge UK: Cambridge University Press.

Young, H. Peyton (2004), "The spread of innovations within groups," Mimeo, Department of Economics, Johns Hopkins University, Baltimore MD 21218.

Appendix

Table 3. OLS estimation of $g(x) = c(1) + c(2)x + c(3)x^2$ by state.

Illinois		Coefficient	Std. Error	t-Statistic	p-value
	C(1)	3.060617	0.188065	16.27422	0
	C(2)	-0.075797	0.011472	-6.606876	0
	C(3)	0.000559	0.000126	4.436975	0

Indiana		Coefficient	Std. Error	t-Statistic	p-value
	C(1)	2.368583	0.131639	17.993	0
	C(2)	-0.0494	0.009134	-5.408568	0
	C(3)	0.000371	0.000108	3.42806	0.0009

Iowa		Coefficient	Std. Error	t-Statistic	p-value
	C(1)	2.354163	0.15061	15.63088	0
	C(2)	-0.024892	0.011515	-2.161701	0.0346
	C(3)	0.0000191	0.000135	0.141644	0.8878

Kansas		Coefficient	Std. Error	t-Statistic	p-value
	C(1)	1.973356	0.126018	15.65934	0
	C(2)	-0.032913	0.0074	-4.447779	0
	C(3)	0.00017	8.39E-05	2.03239	0.0455

Michigan		Coefficient	Std. Error	t-Statistic	p-value
	C(1)	1.973356	0.126018	15.65934	0
	C(2)	-0.032913	0.0074	-4.447779	0
	C(3)	0.00017	8.39E-05	2.03239	0.0455

Ohio		Coefficient	Std. Error	t-Statistic	p-value
	C(1)	2.185582	0.279818	7.810739	0
	C(2)	-0.045885	0.013443	-3.413248	0.0011
	C(3)	0.000349	0.000143	2.437297	0.0177

Wisconsin		Coefficient	Std. Error	t-Statistic	p-value
	C(1)	2.429842	0.15327	15.85332	0
	C(2)	-0.033255	0.011388	-2.920241	0.0049
	C(3)	0.000128	0.000135	0.948789	0.3465

Table 4. OLS estimation of $g(x) = c(1) + c(2)x + c(3)x^2$ on subsamples by state. Subsample A is $0.5 \leq x \leq 20$ and subsample B is $20 < x \leq 90$ except for Ohio, where they are $0.5 \leq x \leq 10$ and $10 < x \leq 90$ respectively.

Subsample A

Illinois		Coefficient	Std. Error	t-Statistic	p-value
	C(1)	1.455313	0.313819	4.637436	0.99995
	C(2)	0.328757	0.105272	3.122942	0.998
	C(3)	-0.018954	0.006221	-3.0466	0.00245

Indiana		Coefficient	Std. Error	t-Statistic	p-value
	C(1)	2.925513	0.387747	7.544897	1
	C(2)	-0.008494	0.141299	-0.060116	0.4763
	C(3)	-0.004124	0.009119	-0.452209	0.3276

Iowa		Coefficient	Std. Error	t-Statistic	p-value
	C(1)	1.939869	0.329981	5.878734	1
	C(2)	0.172919	0.105004	1.64678	0.945
	C(3)	-0.010328	0.005426	-1.903544	0.0333

Kansas		Coefficient	Std. Error	t-Statistic	p-value
	C(1)	1.921343	0.338986	5.667905	1
	C(2)	-0.050417	0.093235	-0.540754	0.29565
	C(3)	0.00224	0.005059	0.442719	0.67

Michigan		Coefficient	Std. Error	t-Statistic	p-value
	C(1)	1.791391	0.316375	5.662248	1
	C(2)	0.044601	0.091428	0.487828	0.68495
	C(3)	-0.004147	0.004781	-0.867481	0.19715

Ohio		Coefficient	Std. Error	t-Statistic	p-value
	C(1)	1.822649	0.570476	3.194959	0.99675
	C(2)	0.704467	0.355268	1.982918	0.9663
	C(3)	-0.063179	0.039799	-1.587478	0.06735

Wisconsin		Coefficient	Std. Error	t-Statistic	p-value
	C(1)	2.072809	0.2758	7.515611	1
	C(2)	0.069304	0.086023	0.805642	0.787
	C(3)	-0.005302	0.004489	-1.181058	0.1229

Subsample B

Illinois		Coefficient	Std. Error	t-Statistic	p-value
	C(1)	2.735526	0.257174	10.63687	0
	C(2)	-0.04563	0.009841	-4.636492	0.99995
	C(3)	0.000221	8.47E-05	2.605663	0.00725

Indiana		Coefficient	Std. Error	t-Statistic	p-value
	C(1)	1.101416	0.661786	1.664309	0.0529
	C(2)	0.012206	0.026206	0.465777	0.32225
	C(3)	-0.000221	0.000229	-0.965295	0.8292

Iowa		Coefficient	Std. Error	t-Statistic	p-value
	C(1)	2.000526	0.16623	12.03468	0
	C(2)	-0.019999	0.00648	-3.086379	0.9977
	C(3)	0.0000322	5.66E-05	0.568914	0.28705

Kansas		Coefficient	Std. Error	t-Statistic	p-value
	C(1)	2.756219	0.722111	3.816893	0.00015
	C(2)	-0.078598	0.027938	-2.813338	0.9967
	C(3)	0.000623	0.000244	2.555345	0.0066

Michigan		Coefficient	Std. Error	t-Statistic	p-value
	C(1)	1.633805	0.431821	3.78352	0.0002
	C(2)	-0.0204	0.016629	-1.22679	0.8873
	C(3)	0.000067	0.000146	0.459487	0.3239

Ohio		Coefficient	Std. Error	t-Statistic	p-value
	C(1)	1.403541	0.342773	4.094661	0.0001
	C(2)	-0.013028	0.014735	-0.884144	0.80935
	C(3)	4.10E-05	0.000135	0.30375	0.38135

Wisconsin		Coefficient	Std. Error	t-Statistic	p-value
	C(1)	2.064414	0.537286	3.842297	0.00015
	C(2)	-0.040594	0.020588	-1.971739	0.9729
	C(3)	0.00031	0.000181	1.715367	0.04625