DRAFT

The Spread of Innovations by Social Learning

H. Peyton Young

Johns Hopkins University

University of Oxford

December, 2005

This version: February 17, 2006

The author is indebted to Pierangelo de Pace, Kirk Moore, and Jon Parker for assistance in gathering and analyzing the data. He also thanks Robert Axtell, Samuel Bowles, Paul David, Joshua Epstein, Andrew Felton, Robert Moffitt, Suresh Naidu, Kislaya Prasad, Matthew Shum, Thomas Valente, and Tiemen Woutersen for helpful comments.

Abstract

One common explanation for the diffusion of innovations is information contagion: agents adopt once they hear about the existence of the innovation from someone else. An alternative explanation is learning: agents adopt once the perceived gain from adoption – as revealed by the outcomes among prior adopters – is high enough to overcome their initial reservations (as embodied in their prior beliefs). We show that learning with heterogeneous beliefs generates an adoption dynamic that has a simple and completely general closed-form solution. Moreover, the pattern of acceleration generated by a learning model often differs qualitatively from the pattern generated by contagion, even when both give rise to S-curves. Applied to Griliches' classic study of hybrid corn, the theory suggests that learning does a better job of explaining the shape of the adoption curves 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 models, their implications for the *shape* of the adoption curve have not previously been studied in any generality.

In this paper we analyze the shape of adoption curves generated by social learning when agents have heterogeneous priors, and show that they differ in important respects from curves generated by contagion models. Learning is a more complex process than contagion because it involves two separate effects: as more people adopt, more information accumulates that helps persuade the

[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 shall use the term in a more comprehensive sense.

remaining people to adopt; in addition, however, the remaining people are inherently more pessimistic and hence harder to persuade. The main theoretical contribution of this paper is to show that such a process has a completely general closed-form solution with *no* restrictions on the ex ante distribution of beliefs. In addition, we formulate a simple nonparametric test that can differentiate between this class of adoption curves and those generated by contagion under mild nonparametric assumptions about the underlying distribution. In particular, the relative acceleration rate of the process will frequently be nonlinear: rising in the early phases of adoption and declining in the later phases. Curves generated by contagion never have this property; instead, the relative acceleration rate decreases at a uniform rate from beginning to end.

In the second part of the paper we apply this framework 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 Griliches' data means that we cannot *identify* learning as the cause of adoption (this exercise would require micro-level panel data). Nevertheless, the shapes of the adoption curves are such that we can reject the contagion model at a high level of significance. The curves also display, at a high level of significance, the particular nonlinear properties that would be predicted by a learning model with heterogeneous priors.

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 a comprehensive overview.[2]

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

In sociology there is a long tradition of studying the diffusion of innovations. In this literature it is common to assume that the diffusion process is driven by differences in agents' *adoption thresholds*, where the threshold is the minimum proportion of adopters in the agent's reference group that induces him to adopt also. [3] Sociologists usually think of these thresholds as reflecting differences in responsiveness to social pressure or the 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. Indeed the theoretical solution we shall derive applies equally well to any process that is driven by heterogeneous adoption thresholds, whether or not they arise from differences in beliefs.[4]

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 (some notable 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.[5] Other authors assume some

[3] See for example Ryan and Gross, 1943; Coleman, Katz, and Menzel, 1966; Granovetter, 1978; Granovetter and Soong, 1983; Macy, 1991; Rogers, 2003; Valente, 1995, 1996, 2005.

[4] Our results are derived under the assumption that interactions among agents are completely random (a *mean-field* model). The situation is more complex when agents interact through a given social network; we shall not treat that case here.

[5] See in particular Banerjee, 1992; Bikhchandani, Hirshleifer, and Welch, 1992; Kapur, 1995; Smith and Sorensen, 2000; Gale and Kariv, 2003; Banerjee and Fudenberg, 2004; Munshi, 2004.

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.[6] In these models it is not heterogeneity in beliefs that drives the adoption process (although heterogeneity could be introduced as an additional element). 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*.

There are, however, several branches of the diffusion literature in which heterogeneity plays a more central role. One line of work assumes that agents differ in certain characteristics (not necessarily beliefs) that establish different thresholds at which they will adopt. For example, they may differ in their degree of risk aversion, cost of adoption, scale of operations, and so forth. As time runs on, some key *exogenous* parameter -- such as the cost of a new product—changes and causes an increasing fraction of the population to adopt as their thresholds are crossed. This approach is sometimes referred to as a *moving equilibrium diffusion model*.[7] A key difference between this approach and the present one is that diffusion is driven *externally* by changes in an exogenous parameter, whereas in our approach diffusion is driven *internally* by information generated by the experience of prior adopters. Externally driven models have fundamentally different dynamical properties than internally driven models, and typically do not exhibit the same acceleration patterns (Young, 2004).

---

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

[7] See David, 1969, 1975, 2003; Davies, 1979; Stoneman, 1981; Jensen, 1982, 1983; Balcer and Lippman, 1984; David and Olsen, 1986.

The line of work that is closest to the present paper is due to Jensen (1982, 1983). He posits that differences in agents' beliefs are the source of heterogeneity, and that belief updating is governed by the outcomes among prior adopters. He then works out the implications for the shape of the adoption curve in particular cases. Specifically, he assumes that agents directly observe the realized payoffs of two competing technologies, which can take on two values, high or low. The two technologies differ in the distribution of high-low outcomes, and one is superior in expectation to the other. An agent irreversibly adopts one of the technologies 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 expected motion of the adoption curve is 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.

The remainder of the paper is organized as follows. In the next section we recall the structure of the standard contagion model, in which hearing about the existence of an innovation is enough to cause people to adopt it (possibly with a lag). Then we develop an alternative model in which an agent adopts only if prior adopters generate enough favorable information to overcome his initial reservations (i.e., to tip him over a threshold determined by his prior belief). The latter model has a simple, closed-form solution for any distribution of thresholds. We then show that, although both the contagion and learning models can produce S-curves, the pattern of acceleration is typically quite different in the two cases, especially in the early phases of adoption. In a learning model, the relative rate of acceleration will often rise initially, whereas in a contagion model it invariably falls from start to finish. In the final section we apply these findings to Griliches' data on hybrid corn.

3. Contagion models

We begin by recalling the basic features of contagion 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

$$p(t + 1) - p(t) = \alpha p(t)(1 - p(t)). \quad (1)$$

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: smaller values of α correspond to greater levels of inertia. 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

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

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

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

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. 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

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

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 $\lambda = 0$ 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, and adopt as soon as they have heard about it. The same model 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. 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. In particular, let us suppose that each 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 framework 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 directly observe the realized payoffs of prior adopters.

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 $h(\mu, \alpha)$ be a real-valued function that is strictly increasing in both μ and α, where α is the number of people who have already adopted and μ is the mean payoff advantage of the innovation (relative to the status quo) among the adopters. In principle μ may be positive, negative, or zero.

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 $h(\mu, \alpha) \geq \theta_i$. The interpretation is that θ_i is the critical level such that the evidence embodied in $h(\mu, \alpha)$ is just sufficient to convince agent i that the innovation is better than the status quo. We shall refer to θ_i as i's *evidentiary threshold*, which may be positive, negative, or zero. A positive value means that the agent is initially pessimistic and requires a sufficiently positive average outcome among a sufficiently large number of people in order to adopt. A negative value means that the agent is initially optimistic and requires a sufficiently negative average outcome among a sufficiently large number of people in order *not* to adopt.

Given these evidentiary thresholds, we may now define a related concept called a "population threshold" or "resistance level." Specifically, let us fix the population size n and a payoff difference μ. For each agent i define

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

In other words, $r_i(\mu, n)$ is the minimum proportion of adopters that is required for i to want to adopt also. We shall refer to this as i's *resistance* level. (In sociology it would be called an *adoption threshold*.) Notice that, unlike an evidentiary threshold, a resistance (if it is finite) must always lie between zero and one because it represents a population proportion. Furthermore it is infinite if a given agent remains unconvinced even when everyone else adopts.

We can illustrate these ideas with a simple example. Suppose that the innovation in question has a random payoff X_i to agent i, where X_i is normally distributed $N(\mu, \sigma^2)$. Assume that the status quo has zero payoff, hence μ represents the *expected payoff gain* from the innovation. For concreteness let us assume that μ is positive. Individuals do not necessarily know this, however, and have heterogeneous beliefs about its actual value. Suppose, in particular, that agent i thinks that μ 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 average outcome among the initial adopters, which is positive in expectation, in which case they too adopt.

In period t, let μ^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, his posterior estimate of μ is a convex combination of the observed mean and his prior:

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

Here n is the number of people in the group and τ_i is a positive number that depends on i's beliefs about μ. For example, (9) holds in a normal-normal updating framework (de Groot, 1970, Chapter 9). As we would expect, agent i gives increasing weight to the observed mean as the number of observations np(t) becomes large.[8]

Assume that each agent 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 μ^t and work with its expectation μ. (This is justified if the size of the population is sufficiently large.) Thus, in expectation, agent i adopts at time t + 1 if and only if

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

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

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

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 $r_i = -\tau_i \mu_i^0/n\mu$.

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

---

[8] We are assuming here that the uncertainty lies in the distribution of payoffs among individuals, not in their distribution among repeated trials by a *given* individual. An alternative assumption

$$p(t-1) < r_i \leq p(t). \tag{12}$$

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

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

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 δ in the current time period. This leads to the discrete-time process

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

The continuous-time analog is a differential equation of form

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

where λ > 0. Assume that F(0) > 0 and p(0) = 0. Let

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

Then (15) is a separable ODE with the solution

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

---

would be that an agent updates based on outcomes in all prior periods, whether by the same or different agents.

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 is very general and applies to any process that is driven by heterogeneous adoption thresholds, whether or not they arise from *ex ante* differences in beliefs. [9]

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

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

where N is the cumulative standard normal distribution, μ is the mean and σ^2 the variance. Figure 1 illustrates the case where $\mu = .10$ and $\sigma^2 = .10$.

[9] Lopez-Pintado and Watts (2005) extend this mean-field technique to study a variety of collective dynamics, including situations where agents' choices have negative as well as positive feedback effects on the choices of other agents.

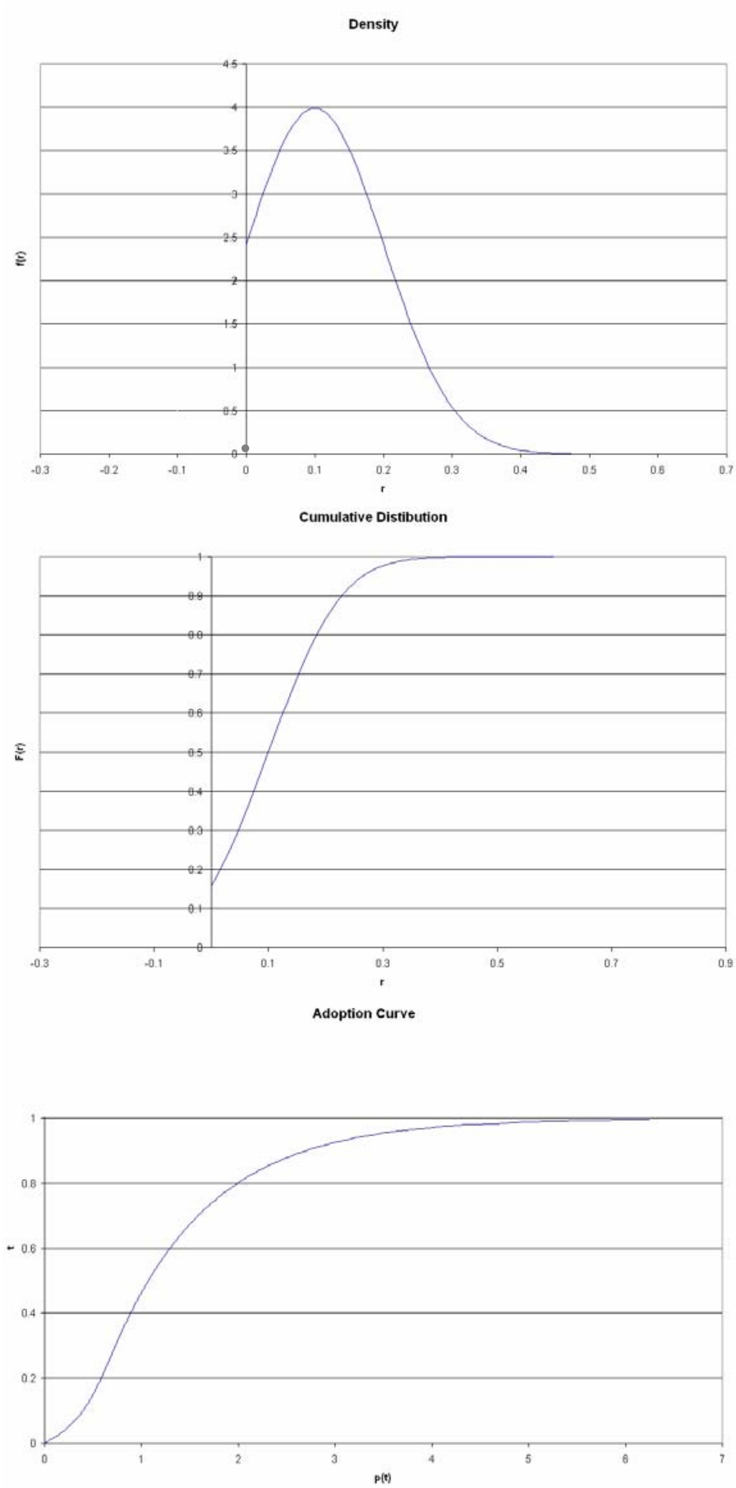

Figure 1. Normal distribution N(.10, .01): truncated density, truncated cumulative distribution function, and adoption curve with $\lambda = 1$.

The top panel shows the distribution of resistances, which is essentially a truncated normal density. There is a point mass at r = 0 representing the optimists (those with negative evidentiary thresholds) who propel the adoption process forward initially; there is also a tiny point mass at r = 1 that represents the individuals who are so pessimistic they will never adopt. The middle panel shows the cumulative distribution function of the resistances. 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.

Let p(t) be an empirically observed adoption curve over the time interval $0 \leq t \leq T$. 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 ≤ t ≤ 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 latter model is defined by the differential equation

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

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

$$g(r) = \lambda - \gamma - 2\lambda r. \qquad (24)$$

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).[10]

If the data are generated by learning with normally distributed evidentiary thresholds θ, then the resistances will have a truncated normal distribution (see (8)) and g(r) will be nonlinear. In this case there are three possible shapes for g(r), depending on where the value θ = 0 falls in the distribution.

1. If θ = 0 lies in the right tail of the normal density (i.e., to the right of the right-most inflection point), then g(r) is convex for all r.

2. If θ = 0 lies between the two tails, then g(r) is concave for smaller values of r and convex for larger values of r.[11]

3. If θ = 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.

Of course it would be going too far to assume that the evidentiary thresholds must be normally distributed. But it is not unreasonable to think that their distribution is unimodal and approximately normal. The reason is that an agent's prior belief will typically be derived from a variety of sources: news

---

[10] A comparison of (22) and (24) reveals that the Bass model is a special case of a 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.

[11] This is the situation depicted in Figure 1.

items, hearsay, experience with other innovations, and so forth. Some of these sources will suggest a positive outcome and others a negative one. Assuming that these outcomes are to some degree random, that the sources are more or less independent, and that there are many of them, the resulting distribution of beliefs will (under quite mild restrictions) be approximately normal due to the central limit theorem. In what follows we shall not in fact assume normality, but we shall look for certain qualitative features of the adoption curve that would arise under a unimodal distribution.

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 adoption process from the overall *shape* of the adoption curves. In particular, he showed that in regions where there was a large expected gain in profitability from switching to hybrid corn, the rate of adoption tended to be high. His focus on the shape of the curves was dictated by the nature of the data, which did not provide dates of adoption by individual farmers (or any other characteristics of the farmers). Instead, he had regional data on the percentage of corn acreage planted in hybrid corn over a period stretching from the early 1930s to the early 1950s. Each of these regions typically consists of several adjacent counties, and is defined by the U.S. Department of Agriculture as a "crop reporting district" for statistical purposes. This is very far from the kind of micro-level data that would be needed to identify in detail the factors that go into individual farmers' decisions to adopt.[12] Nevertheless we can carry out some exploratory data analysis on the shape of the adoption curves to see whether some of the features discussed in the preceding sections might be present. In particular, we can check whether the relative acceleration is decreasing and linear, as would be implied by the contagion model, or has the concave-then-convex shape that could easily arise in a heterogeneous thresholds model.

First let us examine the adoption curves themselves. Figure 2 shows the percentage of corn acreage planted in hybrid corn, by crop reporting district, for seven "corn belt" states in each year during the period 1933-1952. Each state has nine reporting districts, so there are 63 adoption curves in all. The data in this form are not in Griliches' 1957 article; they had to be retrieved from his papers in the Harvard University Archives.[13]

The great majority of the curves have the characteristic S-shape, with a sharp acceleration in the early phases of adoption. Nevertheless certain qualitative differences among the states can be discerned. 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

[12] For empirical work of this type see Besley and Case, 1994; Foster and Rosenzweig, 1995; Conley and Udry, 2003; Munshi and Myaux, 2005

[13] I am indebted to Diane Asseo Griliches for giving me access to her late husband's papers. The original data set was considerably larger and included all twelve states comprising the corn belt; only data for these seven states in the corn belt could be located.

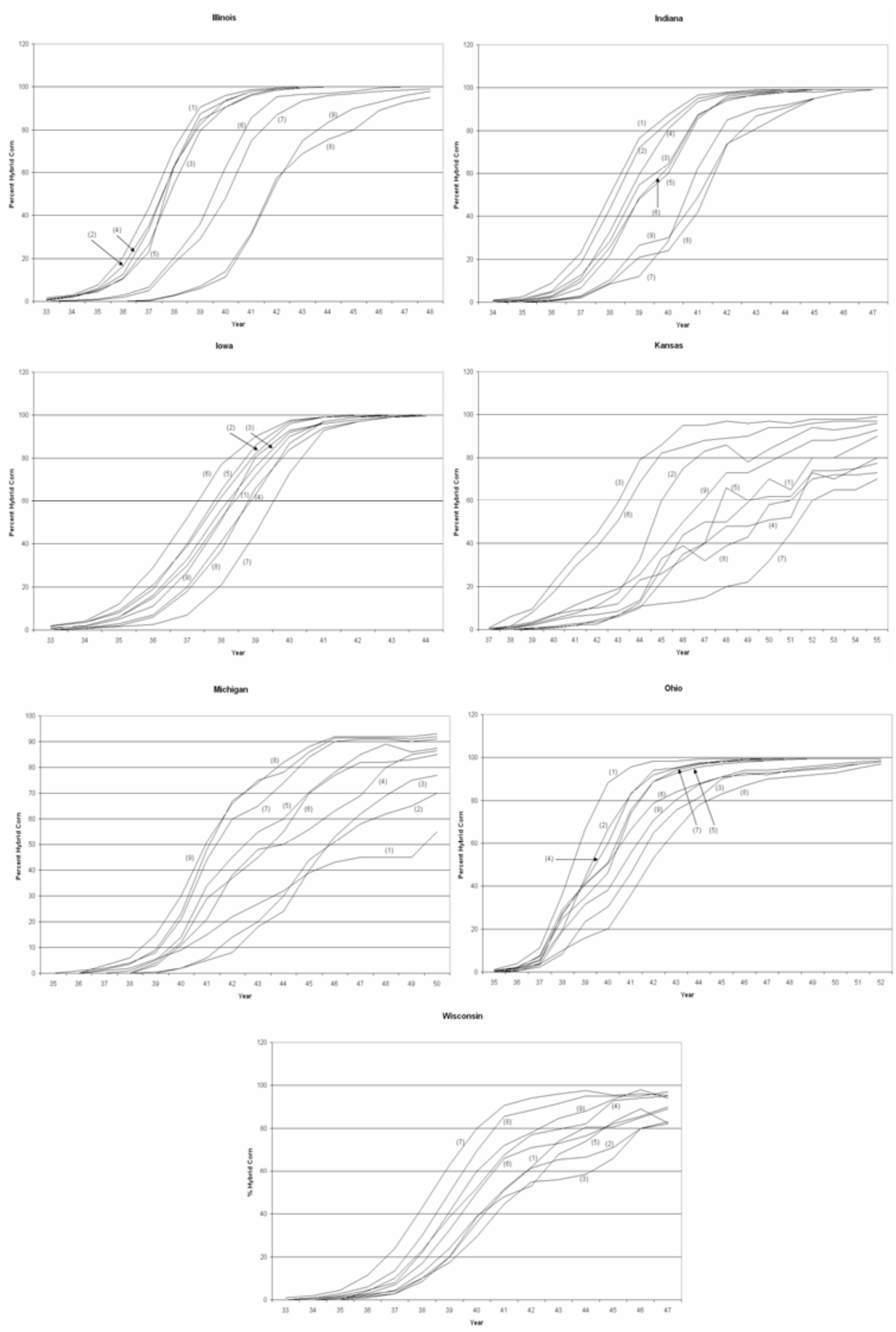

Figure 2. Adoption curves for hybrid corn in seven corn-belt states, by crop reporting district: 1933-1952.

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. These lags are due in part to the fact that hybrid seed needs to be tailored to growing conditions in specific areas, and the appropriate seed was developed for some areas earlier than for others (Griliches, 1957). They may also be due in part to differences among local agricultural extension agents in promoting the new technology. (We shall control for differences in start-up times in the statistical estimations to follow.)

We now turn to the statistical estimations. We remind the reader that this is an exploratory data analysis of the shape of the curves; we cannot delve into the micro-level data that generated the curves because it is not available (nor was it available to Griliches as far as we know). First we shall estimate the acceleration function $g_i(\cdot)$ for each crop reporting district i. In particular, let x_{it} be the *percentage* of corn acreage planted in hybrid corn in district 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 is an estimate of the relative rate of acceleration $g_i(r_{it})$ at the resistance level $r_{it} = x_{it}/100$. (Note that x_{it} is a percentage, whereas resistance is a proportion.)

Recall that the contagion model implies that for each i, $y_{it} - 1$ should be a decreasing linear function of x_{it}, hence the same should hold for y_{it}. The learning model, by contrast, implies that $y_{it} - 1$ (and hence also y_{it}) is a positive linear transformation of some unknown density $f(x_{it})$.

The analysis consists of four parts. First we shall pool the pairs (x_{it}, y_{it}) from all 63 districts and study the shape of y as a function of x using a nonparametric, nearest neighbors estimation procedure. Next we confirm the nonparametrically estimated shape by breaking the data into disjoint subsamples and testing for linearity, concavity, and convexity on the subsamples. (Of course we do these parametric estimations with the original data, not the nearest neighbors smoothed data.) In the third and fourth steps we apply the same procedure to each state separately. All of these estimates automatically control for differences in start-up times among the districts. This is because we are estimating the acceleration rate in each district as a function of the penetration level in each district; it is irrelevant how long it took to reach that level.

Let us first pool all of the pairs (x_{it}, y_{it}) and estimate the shape of the 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. (We chose $k = 77$, which is approximately a 10% span.)[14] The results for the pooled data are shown in Figure 3. This curve appears to be concave for low values of x (below about 20%) and convex for higher values of x. It does not appear to be linear.

Next we verify these features parametrically by fitting a quadratic function to the data using OLS. First let us check for nonlinearity by fitting a quadratic to the entire range of x-values. As Table 1 shows, the quadratic term is different from zero at an extremely high level of significance, so we can confidently reject linearity. This means, in particular, that the data are very likely inconsistent with a contagion process whose parameters are common across states and districts. It remains to be seen whether this remains true when we allow district-specific contagion parameters; we shall return to this issue in a moment.

---

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

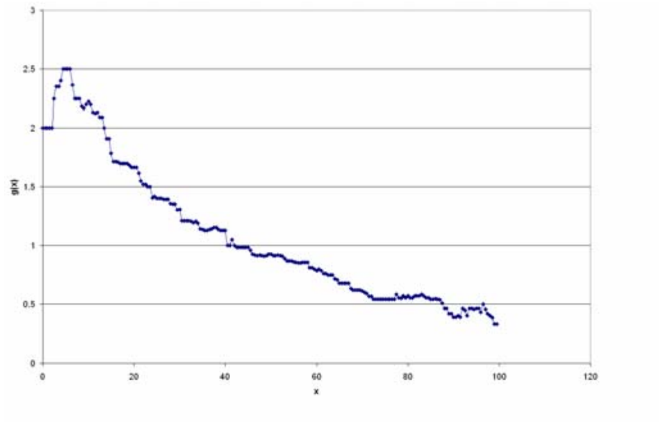

Figure 3. Relative acceleration rate g(x) as a function of adoption level x. Median of 77 nearest neighbors (span = 10%). Pooled data.

Table 1. Pooled data: OLS estimation of $g(x) = c_0 + c_1x + c_2x^2$.

	Coefficient	Std. Error	t-Statistic
c_0	2.413073	0.062948	38.33456
c_1	-0.048165	0.004232	-11.38173
c_2	0.000313	4.89E-05	6.397485

Next let us check the concave-then-convex shape suggested by the nonparametric analysis using a parametric estimation. Specifically, let us divide the data into two disjoint subsamples: A, $0.5 \le x \le 20$ and B, $20 < x \le 90$.[15] We then estimate a quadratic model separately on each subsample by OLS, after removing outliers.[16] The estimated quadratic coefficient on subsample A is negative at the 1.1% level of significance, while on subsample B it is positive at the 3.7% level of significance (see Table 2). Thus the parametric and nonparametric methods complement each other, and indicate that the function $y = g(x)$ 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_0	2.166858	0.136397	15.88636	1
c_1	0.056931	0.041629	1.367586	0.9136
c_2	-0.005204	0.002247	-2.315373	0.01075

Subsample B: $20 < x \le 90$

	Coefficient	Std. Error	t-Statistic	1-sided p-value
c_0	1.778686	0.225581	7.884897	0
c_1	-0.02594	0.008685	-2.98668	0.9985
c_2	0.000136	7.54E-05	1.798211	0.03655

We now refine the analysis by allowing for coefficients that vary among states and districts. This involves a considerable loss of power in conducting tests of significance, but something of interest can still be said.

[15] Adoption levels of less than 0.5% were excluded because the estimation errors are very large. Adoption levels above 90% were excluded because many districts did not reach this level of 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.

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

For each of the 63 districts, let $y_i = g_i(x_i)$ be the acceleration function for district i, where x_i is the level of penetration in the district. If the contagion model is correct, then $g_i(x_i)$ should be linear and decreasing in x_i with coefficients that may be specific to district i. To test this hypothesis we introduce district-specific slopes and intercepts and estimate a model of form

$$y_i = c_{0i} + c_{1i}\, x_i + c_2 x_i^2. \qquad (24)$$

Note that the quadratic coefficient is not specific to i; if the contagion model (which is our null hypothesis) is correct, then it should be zero. In fact, it turns out to be significantly different from zero with a t-statistic of 4.87.[17] Thus the observed acceleration pattern is not consistent with the contagion model even allowing for contagion parameters that vary across districts.

In principle we would also like to test whether the district acceleration functions g_i are concave-then-convex with coefficients that are specific to each district, but the district-level data are too sparse to carry out such a test. The data are rich enough, however, to test the concave-then-convex hypothesis on a state-by-state basis. This amounts to assuming that there is a common process generating the data among the districts within each state, but possibly different processes across states.[18] As before we employ both nonparametric and parametric estimation techniques to analyze the shape of the acceleration functions.

---

[17] Specifically, $c_2 = 0.000265$ with a standard error of 0.0000544 and t-statistic equal to 4.874255. The units are small because x is measured in percent, hence x^2 can be as high as $100^2 = 10{,}000$.

[18] As we noted earlier, the analysis automatically controls for differences in start-up times among districts, not merely among states.

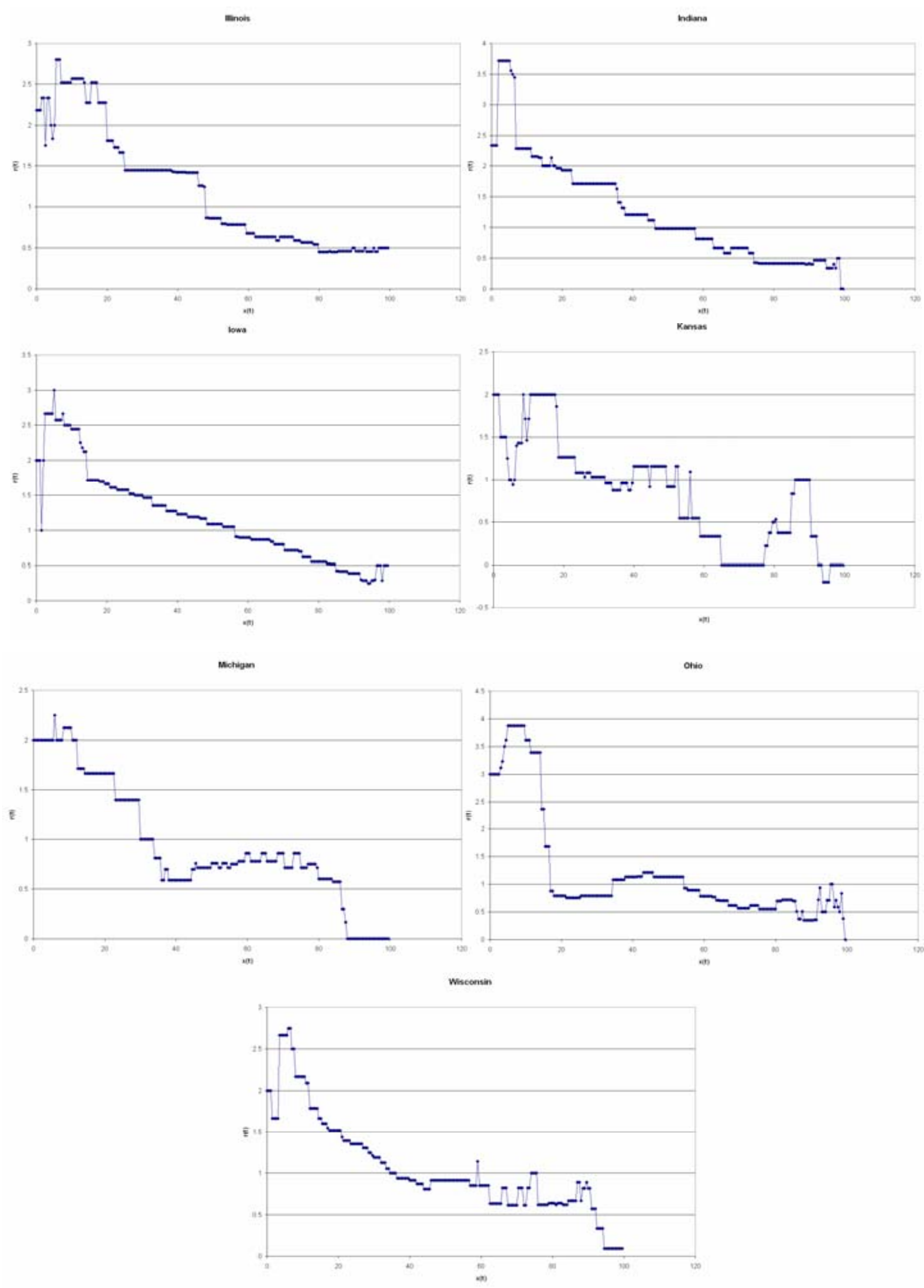

Figure 4. Relative acceleration rate g(x) as a function of adoption level x. Median of 11 nearest neighbors (span = 10%). State-level data.

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 $x \leq 20$. 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 a choppy but more nearly linear fashion.

These nonparametric estimations give an overall sense of the shape of the acceleration functions; we now estimate the shapes parametrically state-by-state. In state s let us choose a dividing point x_s^* and test for concavity below x_s^* and for convexity above x_s^*. For simplicity we choose $x_s^* = 20\%$ in all states except Ohio, where we let $x_s^* = 10\%$. These choices are loosely based on the nonparametric estimations; even better results can be obtained if we tailor the dividing points more finely on a state-by-state basis. The results of the parametric estimations are presented in Table 4 in the Appendix. On the subsamples $x \leq x_s^*$, we find that in six out of the seven states the estimated quadratic coefficients are negative (indicating concavity). 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). On the subsamples $x > x_s^*$, the estimated quadratic coefficients are positive (indicating convexity) 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 are statistically significant for individual states, they are highly significant as a group.[19] To see why, let us compute the *average* t-statistic of the quadratic coefficients for each of the subsamples. On the A-subsamples in Table 4, the average of the seven t-statistics is -1.2179. On the B-subsamples the average of the seven t-statistics is 1.0347. Each of these t-statistics 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* of seven t-statistics is close to being normal with mean 0 and variance 1/7. Hence the realized average t-statistic on the A-subsamples is $-1.2179\sqrt{7} = -3.222$ standard deviations below the mean. This is significantly negative (in a one-sided test) at the 1% level. Similarly, the realized average t-statistic on the B-subsamples is $1.0347\sqrt{7} = 2.738$ standard deviations above the mean, which is significantly positive at the 1% level.

Overall, therefore, the state-level data are broadly consistent with a heterogeneous thresholds model in which the underlying heterogeneity is unimodal. This does not identify the source of the heterogeneity, nor does it rule out the possibility that other models of adoption might predict the observed pattern of acceleration equally well. However, we are unaware of any other models in the literature that do make such predictions. Moreover, we have shown that the usual way of modeling adoption curves, namely the contagion model, is rejected at a high level of significance even when we allow for district-specific contagion parameters.

[19] 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.

8. Social learning or social conformity?

In this concluding section we ask whether there is any evidence that it is heterogeneous beliefs about *payoffs* that are driving the process, as opposed to other kinds of heterogeneity, such as responsiveness to social pressure or the desire to conform (as is posited in some sociological studies, including the influential paper by Ryan and Gross, 1943). In fact we can say something about this possibility. For if the adoption dynamics were driven *purely* by conformity or social pressure, the gain in payoff from adopting hybrid corn should not affect 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 analyses show that the speed with which adoption occurs within a given district tends to increase with the expected *gain in profitability* from adopting hybrid corn in that district. (This follows from their estimation of the coefficients in the logistic and Gompertz functions respectively as a function of district-level profitability.) [20]

The same conclusion follows readily from our model with virtually no assumptions about the parametric form of the distribution of beliefs. Suppose, for example, that there is a common distribution of prior beliefs about the value of the mean among districts, but that the *actual* mean differs among districts. Suppose further that the size, n^*, of an agent's reference group is the same in all districts. From (8) we know that agent i's resistance $r_i(\mu, n^*)$ is a function of i's prior beliefs and the true mean μ, where $r_i(\mu, n^*)$ is strictly decreasing in μ. (In the normal-normal case $r_i(\mu, n^*) = (1/\mu)r_i(1, n^*)$.)

---

[20] 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).

Now consider two different districts, say 1 and 2, with means $\mu_1 < \mu_2$. Let $F_i(r)$ be the resulting distribution of resistances in district 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, $F_2(r) = F_1((\mu_2/\mu_1)r) > F_1(r)$.) Now consider any two realized levels of penetration $x < y$, and let $T_i(x, y)$ be the length of time it takes in district i to go from level x to level y. From equation (17) it follows that

$$T_i(x, y) = 1/\lambda \int_x^y dr/(F_i(r) - r). \qquad (26)$$

The preceding discussion shows that the denominator of (26) 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 gain from the innovation. Thus, while we cannot rule out the possibility that social pressure or conformity played some role, it seems likely that beliefs about payoffs must be part of the explanation too.

9. Conclusion

We conclude with the following observations. First, we have shown that learning with heterogeneous beliefs generates a simple and general class of dynamical processes whose solution depends on the cumulative distribution function of the implied resistances. Second, we have shown that the relative acceleration rates, which are directly computable from a given adoption curve, can be used to recover the underlying density of resistances, or at least a positive linear translation thereof. If adoption were generated by information contagion, the relative acceleration rates should be linear and decreasing. In a learning model, the acceleration rate will often display a more complex pattern. In particular, if evidentiary thresholds are unimodally distributed in the population, the relative acceleration rate will first rise and then fall. This pattern is in fact seen in Griliches' data on hybrid corn, whereas the linearity implied by contagion is decisively rejected. Moreover, the acceleration rate increases with the actual gains in profitability, as the learning model would predict. If instead adoption were driven by a heterogeneous *conformity* parameter, there should be no payoff effect. While no claim is made that this identifies learning as the cause of adoption, the model we have proposed does lead to interesting and testable predictions about the shape of adoption curves that appear not to have been studied before.

References

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

Bala, Venkatesh, and Sanjeev Goyal (1998), "Learning from neighbors," *Review of Economic Studies*, 65, 595-621.

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

Banerjee, Abhijit (1992), " A simple model of herd behavior," *Quarterly Journal of Economics*, 107, 797-817.

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.

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch (1992), "A theory of fads, fashion, custom, and cultural change as informational cascades," *Journal of Political Economy*, 100, 992-1026.

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

Coleman, James S., Elihu Katz, and H. Menzel (1966). *Medical Innovation: A Diffusion Study*. New York: Bobbs Merrill.

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

David, Paul A. (1969), "A contribution to the theory of diffusion," Research Center in Economic Growth Memorandum No. 71, Stanford University. Mimeo.

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

David, Paul A. (2003), "Zvi Griliches on diffusion, lags, and productivity growth…connecting the dots," Department of Economics, Stanford University. Mimeo.

David, Paul A. and Trond E. Olsen (1986), " Equilibrium dynamics of diffusion when incremental technological innovations are foreseen," Ricerche Economiche, (Special Issue on Innovation Diffusion), 40, 738-770.

Davies, S. (1979). *The Diffusion of Process Innovations*. 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.

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

Granovetter, Michael S., and Roland Soong (1983), "Threshold models of diffusion and collective behavior," *Journal of Mathematical Sociology*, 9, 165-179.

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.

Lopez-Pintado, Dunia, and Duncan J. Watts (2005), " Social Influence, Binary Decisions, and Collective Dynamics," Working Paper, Department of Sociology, Columbia University.

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

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

Mansfield, Edwin (1961), "Technical change and the rate of innovation," *Econometrica*, 29, 741-766.

Munshi, Kaivan (2004), "Social learning in a heterogeneous population: technology diffusion in the Indian green revolution," *Journal of Development Economics*, 73, 185-215.

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.

Smith, Lones, and Peter Sorensen (2000), " Pathological outcomes of observational learning, " *Econometrica*, 68, 371-398.

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. (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 $g(x) = c_0 + c_1x + c_2x^2$ by state.

Illinois		Coefficient	Std. Error	t-Statistic	p-value
	c_0	3.060617	0.188065	16.27422	0
	c_1	-0.075797	0.011472	-6.606876	0
	c_2	0.000559	0.000126	4.436975	0

Indiana		Coefficient	Std. Error	t-Statistic	p-value
	c_0	2.368583	0.131639	17.993	0
	c_1	-0.0494	0.009134	-5.408568	0
	c_2	0.000371	0.000108	3.42806	0.0009

Iowa		Coefficient	Std. Error	t-Statistic	p-value
	c_0	2.354163	0.15061	15.63088	0
	c_1	-0.024892	0.011515	-2.161701	0.0346
	c_2	0.0000191	0.000135	0.141644	0.8878

Kansas		Coefficient	Std. Error	t-Statistic	p-value
	c_0	1.973356	0.126018	15.65934	0
	c_1	-0.032913	0.0074	-4.447779	0
	c_2	0.00017	8.39E-05	2.03239	0.0455

Michigan		Coefficient	Std. Error	t-Statistic	p-value
	c_0	1.973356	0.126018	15.65934	0
	c_1	-0.032913	0.0074	-4.447779	0
	c_2	0.00017	8.39E-05	2.03239	0.0455

Ohio		Coefficient	Std. Error	t-Statistic	p-value
	c_0	2.185582	0.279818	7.810739	0
	c_1	-0.045885	0.013443	-3.413248	0.0011
	c_2	0.000349	0.000143	2.437297	0.0177

Wisconsin		Coefficient	Std. Error	t-Statistic	p-value
	c_0	2.429842	0.15327	15.85332	0
	c_1	-0.033255	0.011388	-2.920241	0.0049
	c_2	0.000128	0.000135	0.948789	0.3465

Table 4. OLS estimation of $g(x) = c_0 + c_1x + c_2x^2$ on subsamples by state. Subsample A is $0.5 \le x \le 20$ and subsample B is $20 < x \le 90$ except for Ohio, where they are $0.5 \le x \le 10$ and $10 < x \le 90$ respectively.

Subsample A

Illinois		Coefficient	Std. Error	t-Statistic	p-value
	c_0	1.455313	0.313819	4.637436	0.99995
	c_1	0.328757	0.105272	3.122942	0.998
	c_2	-0.018954	0.006221	-3.0466	0.00245

Indiana		Coefficient	Std. Error	t-Statistic	p-value
	c_0	2.925513	0.387747	7.544897	1
	c_1	-0.008494	0.141299	-0.060116	0.4763
	c_2	-0.004124	0.009119	-0.452209	0.3276

Iowa		Coefficient	Std. Error	t-Statistic	p-value
	c_0	1.939869	0.329981	5.878734	1
	c_1	0.172919	0.105004	1.64678	0.945
	c_2	-0.010328	0.005426	-1.903544	0.0333

Kansas		Coefficient	Std. Error	t-Statistic	p-value
	c_0	1.921343	0.338986	5.667905	1
	c_1	-0.050417	0.093235	-0.540754	0.29565
	c_2	0.00224	0.005059	0.442719	0.67

Michigan		Coefficient	Std. Error	t-Statistic	p-value
	c_0	1.791391	0.316375	5.662248	1
	c_1	0.044601	0.091428	0.487828	0.68495
	c_2	-0.004147	0.004781	-0.867481	0.19715

Ohio		Coefficient	Std. Error	t-Statistic	p-value
	c_0	1.822649	0.570476	3.194959	0.99675
	c_1	0.704467	0.355268	1.982918	0.9663
	c_2	-0.063179	0.039799	-1.587478	0.06735

Wisconsin		Coefficient	Std. Error	t-Statistic	p-value
	c_0	2.072809	0.2758	7.515611	1
	c_1	0.069304	0.086023	0.805642	0.787
	c_2	-0.005302	0.004489	-1.181058	0.1229

Subsample B

Illinois		Coefficient	Std. Error	t-Statistic	p-value
	c_0	2.735526	0.257174	10.63687	0
	c_1	-0.04563	0.009841	-4.636492	0.99995
	c_2	0.000221	8.47E-05	2.605663	0.00725

Indiana		Coefficient	Std. Error	t-Statistic	p-value
	c_0	1.101416	0.661786	1.664309	0.0529
	c_1	0.012206	0.026206	0.465777	0.32225
	c_2	-0.000221	0.000229	-0.965295	0.8292

Iowa		Coefficient	Std. Error	t-Statistic	p-value
	c_0	2.000526	0.16623	12.03468	0
	c_1	-0.019999	0.00648	-3.086379	0.9977
	c_2	0.0000322	5.66E-05	0.568914	0.28705

Kansas		Coefficient	Std. Error	t-Statistic	p-value
	c_0	2.756219	0.722111	3.816893	0.00015
	c_1	-0.078598	0.027938	-2.813338	0.9967
	c_2	0.000623	0.000244	2.555345	0.0066

Michigan		Coefficient	Std. Error	t-Statistic	p-value
	c_0	1.633805	0.431821	3.78352	0.0002
	c_1	-0.0204	0.016629	-1.22679	0.8873
	c_2	0.000067	0.000146	0.459487	0.3239

Ohio		Coefficient	Std. Error	t-Statistic	p-value
	c_0	1.403541	0.342773	4.094661	0.0001
	c_1	-0.013028	0.014735	-0.884144	0.80935
	c_2	4.10E-05	0.000135	0.30375	0.38135

Wisconsin		Coefficient	Std. Error	t-Statistic	p-value
	c_0	2.064414	0.537286	3.842297	0.00015
	c_1	-0.040594	0.020588	-1.971739	0.9729
	c_2	0.00031	0.000181	1.715367	0.04625