Innovation Diffusion in Heterogeneous Populations

H. Peyton Young

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

CSED Working Paper No. 45

December 28, 2006

The author thanks Robert Axtell, Paul David, Steven Durlauf, Joshua Epstein, James Heckman, Josef Hofbauer, Thomas Norman, Thomas Valente, and Tiemen Woutersen for constructive comments on earlier drafts. Jon Parker assisted with the simulations.

Abstract

New products and practices take time to diffuse, a fact that is often attributed to some form of *heterogeneity* among potential adopters. People may realize different benefits and costs from the innovation, or have different beliefs about its benefits and costs, hear about it at different times, or delay in acting on their information. This paper analyzes the dynamics arising from different sources of heterogeneity in a completely general setting without placing parametric restrictions on the distribution of the relevant characteristics. The structure of the dynamics, especially the pattern of acceleration, depends importantly on which type of heterogeneity is driving the process. These differences are sufficiently marked that they provide a potential tool for discriminating empirically among diffusion mechanisms. The results have potential application to marketing, technological change, fads, and epidemics.

JEL Classification: O33, D8, M3. Keywords: *diffusion, innovation, learning*

1. Introduction

The diffusion of new products and practices usually takes time, and the proportion of people who have adopted at each point in time frequently, though not invariably, traces out an S-shaped curve. There is an extensive theoretical and empirical literature on this phenomenon and the mechanisms that might give rise to it.[1] Different lines of explanation have been pursued in the various disciplines -- marketing, sociology, and economics – where innovation diffusion has been most intensively studied. A crucial feature of some of these explanations is that *heterogeneity* among the agents is the reason they adopt at different times. Nevertheless, most of the extant models incorporate heterogeneity in a very restricted fashion, say by considering two homogeneous populations of agents, or by assuming that the heterogeneity is described by a particular family of distributions. [2]

In this paper we shall show how to incorporate heterogeneity into some of the benchmark models in marketing, sociology, and economics without imposing *any* parametric restrictions on the distribution of parameters. The resulting dynamical systems turn out to be surprisingly tractable analytically; indeed, some of them can be solved explicitly for any distribution of the parameter values. We then demonstrate that different models leave distinctive 'footprints'; in particular, they exhibit noticeably different patterns of acceleration, especially in the start-up phase, with few or no assumptions on the distribution of the

[1] For reviews of the literature see Mahajan and Peterson (1985), Mahajan, Muller, and Bass (1990), Geroski (2000), Stoneman (2002), and Valente (1995, 1996, 2005).

[2] See among others Jeuland, 1981; Jensen, 1982; Karshenas and Stoneman, 1992; Geroski, 2000.

parameters. The reason is that the models themselves have fundamentally different structures that even large differences in the distributions cannot overcome. It follows that, given sufficient data on the aggregate dynamics of a diffusion process, one could assess the relative plausibility of different mechanisms that might be driving it with little or no prior knowledge about the distribution of parameters. While this type of analysis is certainly no substitute for having good micro-level data, it could be useful in situations where such data are unavailable. [3]

We shall consider five general approaches to explaining innovation diffusion.

1. *Inertia*. People delay adopting out of inertia or because they need to wait for a revision opportunity to come along.

2. *Contagion.* People adopt the innovation when they hear about it from someone who has already adopted.

3. *Conformity*. People adopt when enough other people in the group have adopted.

4. *Social learning*. People adopt once they see enough evidence among prior adopters to convince them that the innovation is worth adopting.

5. *Moving equilibrium*. As external conditions change, say as the cost of the innovation decreases or information about it increases, more and more people adopt as their reservation thresholds are crossed.

---

[3] An exploratory study of this type using Griliches' data on hybrid corn can be found in Young (2006).

Contagion (or epidemic) models are common in the marketing literature. Conformity (or threshold) models are the standard explanation in sociology. Learning and moving equilibrium models are the preferred approaches in economics. The paper is structured as follows. For each type of explanation we begin with a benchmark model and show how to incorporate heterogeneity of the parameters in complete generality. (In the case of social learning we develop the benchmark model from scratch, since surprisingly little has been done on this approach in the prior literature.) We then show how to solve the resulting dynamical systems and investigate their dynamic properties, particularly the pattern of acceleration. In all of these cases I adopt a mean-field approach, in which agents are assumed to interact at random and the population is large. This allows the expected motion of the process to be analyzed using systems of differential equations. The analysis can be extended to small population settings and to situations where agents interact through a social network. These extensions require substantially different techniques, however, and will be treated separately.

2. Inertia

To fix ideas it will be helpful to begin with one of the simplest explanations of diffusion, namely, that people sometimes delay in acting on their information. Such delays might be caused by pure procrastination, for example, or the need to wait until a replacement opportunity arises, e.g., a person adopts a new product only when his current model wears out. Suppose first that there is no heterogeneity among agents, and and let λ be the instantaneous rate at which any given non-adopter first adopts.

We treat adoption as an irreversible process in the short run. Let $p(t)$ be the proportion of adopters by time t, where we set the clock so that $p(0)=0$. The function $p(t)$ is called the *adoption curve*. When the population is homogeneous with instantaneous adoption rate λ, the expected motion is described by the ordinary differential equation $\dot{p}(t)=\lambda(1-p(t))$, and the unique solution is $p(t)=1-e^{-\lambda t}$.

Notice that this curve is *concave* throughout; in particular, it is not S-shaped. We claim that this remains true when any amount of heterogeneity is introduced. Specifically, let $\nu(\lambda)$ be a distribution of λ in the population whose support lies in some bounded interval $0 \le \lambda \le b$. Then the expected trajectory of the process is given by

$$p(t)=1-\int e^{-\lambda t} d\nu . \tag{1}$$

Differentiating (1) twice over, we see that $\ddot{p}(t)<0$ *irrespective of the distribution* $\nu(\lambda)$. The intuition is straightforward: agents with low inertia (high λ) tend to adopt earlier than those with high inertia (low λ). Hence the rate of adoption falls for two reasons: first, because the average degree of inertia in the remaining population of non-adopters is *increasing* over time, and second, because the number of non-adopters is *decreasing* over time.

This simple example illustrates the kinds of results that hold in more complex diffusion processes: the logic of the model has implications for the shape of the curve that remain true even when an arbitrary degree of heterogeneity is introduced.

3. Contagion

The next example we shall consider is the benchmark model in the marketing literature, which is variously known as the *Bass model* of new product diffusion (Bass, 1969, 1980) or the *mixed-influence diffusion model* (Mahajan and Peterson, 1985). The basic idea is that people adopt an innovation once they hear about it, and they can hear about it in one of two ways: from sources within the group or from sources external to the group (or both). The relative strength of these two information channels determines the shape of the curve.

Specifically, let λ be the instantaneous rate at which a current non-adopter hears about the innovation from a previous adopter within the group, and let γ be the instantaneous rate at which he hears about it from sources outside of the group. We shall assume that λ and γ are nonnegative, and that not both are zero. In the absence of heterogeneity, such a process is described by the ordinary differential equation $\dot{p}(t) = (\lambda p(t) + \gamma)(1 - p(t))$, and the solution is

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

When contagion is generated purely from internal sources ($\gamma = 0$) this boils down to the ordinary logistic function, which is of course S-shaped.[4] When innovation is driven solely by an external source ($\gamma > 0$ and $\lambda = 0$), the result is the negative exponential distribution, just as in the case of pure inertia. When both γ and λ are positive, we can choose β in expression (2) so that $p(0) = 0$; namely, with $\beta = 1/\gamma$ we obtain

[4] The logistic model was common in the early work on innovation diffusion; see for example Griliches (1957) and Mansfield (1961). Dixon (1980) showed that Griliches' data are better modeled by Gompertz functions.

$$p(t) = [1 - e^{-(\lambda+\gamma)t}]/[1 + (\lambda/\gamma)e^{-(\lambda+\gamma)t}]. \tag{3}$$

This basic model has spawned many variants, some of which assume a degree of heterogeneity, such as two groups with different contagion parameters (Karshenas and Stoneman , 1992; Geroski, 2000) or employ a specific parametric form such as the gamma distribution (Jeuland, 1981).

The fully heterogeneous version can be formulated as follows. Let μ be the joint distribution of the contagion parameters λ and γ. For convenience we shall assume that μ has bounded support, which we may take to be $\Omega = [0,1]^2$. (Rescaling λ and γ by a common factor is equivalent to changing the time scale, so this involves no real loss of generality.) In what follows we shall always assume that $\bar{\gamma} = \int_\Omega \gamma d\mu > 0$, for otherwise the process cannot get out of the initial state $p(0) = 0$.

Let $p_{\lambda,\gamma}(t)$ be the *proportion* of all type-(λ,γ) individuals who have adopted by time t. Then the proportion of all individuals who have adopted by time t is

$$p(t) = \int p_{\lambda,\gamma}(t)d\mu . \tag{4}$$

(Hereafter integration over Ω is understood.) Each subpopulation of adopters $p_{\lambda,\gamma}(t)$ evolves according to the differential equation

$$\dot{p}_{\lambda,\gamma}(t) = (\lambda p(t) + \gamma)(1 - p_{\lambda,\gamma}(t)). \tag{5}$$

This defines a system of first-order differential equations coupled through the common term $p(t)$. We can reduce it to an ordinary differential equation by the following device. Let $x_{\lambda,\gamma}(t) = \ln(1 - p_{\lambda,\gamma}(t))$ and observe that (5) is equivalent to the system $\dot{x}_{\lambda,\gamma}(t) = -(\lambda p(t) + \gamma)$ for all (λ, γ). From this and the initial condition $x_{\lambda,\gamma}(0) = 0$ we obtain

$$x_{\lambda,\gamma}(t) = -\int_0^t (\lambda p(s) + \gamma) ds = -\lambda \int_0^t p(s) ds - \gamma t . \tag{6}$$

From the definition of $x_{\lambda,\gamma}(t)$ it follows that

$$p(t) = 1 - \int e^{x_{\lambda,\gamma}(t)} d\mu ; \tag{7}$$

that is, $p(t)$ satisfies the integral equation

$$p(t) = 1 - \int e^{-\gamma t - \lambda \int_0^t p(s) ds} d\mu . \tag{8}$$

Differentiating we obtain

$$\dot{p}(t) = \int (\lambda p(t) + \gamma) e^{-\gamma t - \lambda \int_0^t p(s) ds} d\mu . \tag{9}$$

Expression (9) can be put in more standard form by defining $y(t) = \int_0^t p(s) ds$. Then $\dot{y}(t) = p(t)$, $\ddot{y}(t) = \dot{p}(t)$, and (9) becomes a second-order differential equation in y. Note that the right-hand side is Lipschitz continuous in $t, y,$ and $\dot{y}$, hence on any finite interval $0 \le t \le T$ there exists a unique continuous solution satisfying the initial condition $p(0) = 0$. By the Picard-Lindelöf theorem, such a solution can be constructed by successive approximation (Coddington and Levinson, 1955). It turns out, however, that we can deduce some key dynamic properties of the process without solving it explicitly: in particular, we will show that $\dot{p}(t)/p(t)$ is strictly decreasing *irrespective of the joint distribution of* λ *and* μ. These and other properties of the model will be derived in section 7.

4. Conformity

The sociological literature on innovation stresses the idea that people have different 'thresholds' that determine when they will adopt as a function of the number (or proportion) of others who have adopted. The dynamics of these models were first studied by Schelling (1971, 1978), Granovetter (1978), and Granovetter and Soong (1988); for more recent work in this vein see Macy (1991), Valente (1995, 1996, 2005), Centola (2006), and Lopez-Pintado and Watts (2006).

For each agent i, suppose that there exists a minimum proportion $r_i \geq 0$ such that i adopts as soon as r_i or more of the group has adopted. (If $r_i > 1$ the agent never adopts.) This is called the *threshold* or *resistance* of agent i. Let $F(r)$ be the cumulative distribution function of resistance in the population. One can then define the discrete-time version of the process as follows (Granovetter, 1978). Let $p(t)$ be the proportion of adopters at period $t = 0,1,2,...$. The clock starts in period 0 when no one has yet adopted. In period 1, everyone adopts whose thresholds are zero. These are the *innovators*. By definition of F the innovators constitute the fraction $F(0)$ of the population, which we shall assume henceforth is strictly positive. In period 2, everyone adopts whose thresholds are at most $F(0)$. Thus at the end of the second period the fraction $F(F(0))$ have adopted.

Proceeding in this way, we obtain $p(t) = F^{[t]}(0)$, where $F^{[t]}$ is the t-fold composition of F with itself.

A useful generalization is to allow for some inertia in the adoption decision. Specifically, let us assume that in each period only a fraction $\alpha \in (0,1)$ of those who are prepared to adopt actually do so. In other words, among those people whose thresholds have been crossed but have not yet adopted by the end of period t, only α will adopt by the end of the next period. This yields the discrete-time process

$$p(t+1) - p(t) = \alpha[F(p(t)) - p(t)]. \tag{10}$$

The continuous-time analog is

$$\dot{p}(t) = \lambda[F(p(t)) - p(t)], \lambda > 0. \tag{11}$$[5]

Assume now that $F(0) > 0$ and let b be the smallest number in (0, 1] such that $F(b) = b$, if any such exists; otherwise let $b = 1$. We then have $F(p) > p$ for all $p \in [0,b)$. Since (11) is a separable ordinary differential equation, we obtain the following explicit solution for the inverse function $t = p^{-1}(x)$:

$$\forall x \in [0,b), \qquad t = p^{-1}(x) = (1/\lambda)\int_0^x dr/(F(r) - r). \tag{12}$$

[5] Lopez-Pintado and Watts (2006) derive the same continuous-time generalization and study its fixed points under various assumptions about F.

Observe that the right-hand side is integrable because $F(r)$ is monotone nondecreasing and $F(r)-r$ is bounded away from zero for all r in the interval $[0,x]$ whenever $x<b$. (The constant of integration is zero because of the initial

condition $p(0)=0$.) The fact that this kind of process has an *explicit* analytic solution for any distribution seems not to have been recognized before.

Suppose now that the parameters $0 \le \lambda, r \le 1$ are *jointly distributed* in the population. Assume that the joint distribution can be expressed as a conditional cumulative distribution $F_\lambda(r)$ for each λ together with an unconditional density $\nu(\lambda)$.[6] Then the cumulative joint distribution function can be written $G(\bar{\lambda}, \bar{r}) = \int_0^{\bar{\lambda}} F_\lambda(r)\nu(\lambda)d\lambda$. Let $p_\lambda(t)$ be the proportion of adopters in the λ–population at time t, and let $p(t) = \int p_\lambda(t)\nu(\lambda)d\lambda$ be the proportion of adopters in the total population at time t. Then

$$\dot{p}_\lambda(t) = \lambda[F_\lambda(p(t)) - p_\lambda(t)], \tag{13}$$

and

$$\dot{p}(t) = \int \lambda F_\lambda(p(t))\nu(\lambda)d\lambda - \int \lambda p_\lambda(t)\nu(\lambda)d\lambda. \tag{14}$$

Unlike the previous case this is not necessarily separable. However, the right-hand side is Lipschitz continuous in a suitably-defined Banach space (namely, the space of continuous functions endowed with the max-norm), hence the system has a unique continuous solution (see for example Lang, 1999). Under quite mild conditions on the distribution it also has certain distinctive acceleration properties, as we shall show in section 7.

---

[6] In fact all the arguments go through when λ has a discrete distribution; the assumption of a density is purely for notational convenience.

5. Social learning

Next we consider processes in which agents adopt only when they see enough positive evidence from the outcomes among previous adopters. These are called *social learning* models, or more precisely, *social learning models based on direct observation*.[7] We shall first outline the general approach and then work out a specific example using normal-normal updating.

Consider a large population of individuals and suppose (for the moment) that the adoption process operates in discrete time $t = 1, 2, 3, \ldots$. Each adopter i generates a payoff X_i that can be observed by those who have not yet adopted. Assume that the realizations X_i are i.i.d. with finite mean μ and variance σ^2. We shall interpret μ to be the expected payoff difference between the innovation and the status quo, and we shall assume that in expectation the innovation is superior to the status quo, that is, $\mu > 0$. Ex ante people do not know that $\mu > 0$; rather, they start with different beliefs (based on their private information) about the value of μ. These beliefs are updated as they see the realized payoffs among prior adopters.

For simplicity let us assume that all agents are risk neutral, and they adopt once they *believe* the mean is positive (given their posteriors). For the process to get started there must exist a group that needs no persuading; these are the optimists. Once they adopt (possibly with some lag), their outcomes are observed by others who were initially pessimistic. Since $\mu > 0$ by assumption,

[7] Another form of social learning occurs in *herding models*, where agents decide whether to adopt based on the fact that others have adopted, not on the realized outcomes, which are private information (Bikchandani, Hirshleifer, and Welsh, 1992; Banerjee, 1992).

the mean outcome among these prior adopters will be positive (in expectation), which will tip some of the pessimists into the optimists' camp. In sum, as more and more people adopt, a larger base of information is created, this information is on average positive, which causes the next group of agents to become optimistic and adopt, which further enlarges the information base, and so forth. Whether this *snowball effect* reaches saturation or fizzles out depends on the distribution of prior beliefs in the population, as we shall see in a moment. In any event, the expected dynamics of the process can be expressed in a surprisingly simple way as a function of the distribution of prior beliefs.

We shall first walk through the argument assuming a discrete-time process and a large but finite population, then pass to the continuous limit. For the sake of concreteness let us temporarily assume a specific parametric structure for the updating process, namely, normal-normal updating; it will soon become apparent that the argument does not depend on this particular parametric structure. Suppose then that the random variables X_i are i.i.d. normal with mean $\mu > 0$ and variance σ^2. At the start of the process, before any outcomes have been observed, each agent i has a prior belief about the value of μ, where the belief is normal with mean $m_i(0)$ and variance τ_i^2. These beliefs are based on agents' private information and may differ among agents. Let there be n agents in the population, where n is large. By period t the proportion $p(t)$ will have adopted, and they will have generated $np(t)$ outcomes with a realized mean

$\mu(t)$. Assume for simplicity that all agents can view $\mu(t)$.[8] Then $i's$ Bayesian posterior $m_i(t)$ is a convex combination of $m_i(0)$ and $\mu(t)$, namely,[9]

$$m_i(t) = \frac{\mu(t) + m_i(0)/\tau_i^2 np(t)}{1 + 1/\tau_i^2 np(t)} \quad . \tag{15}$$

By assumption i is risk-neutral, so she is prepared to adopt once $m_i(t) > 0$. In particular, agents such that $m_i(0) > 0$ need no convincing and are prepared to adopt right away. These *optimists* propel the process forward initially. By contrast, an agent i such that $m_i(0) < 0$ is initially *pessimistic;* she only changes her mind once she sees enough positive outcomes among prior adopters. Specifically, expression (15) shows that she will change her mind provided that $\mu(t) > -m_i(0)/\tau_i^2 np(t)$, which is equivalent to

$$p(t) > -m_i(0)/\tau_i^2 n\mu(t) . \tag{16}$$

By assumption, $\mu(t)$ is $N(\mu, \sigma/\sqrt{np(t)})$. Assuming that n is large, the realized mean $\mu(t)$ is close to the actual mean with high probability unless $p(t)$ is very small. Thus, except possibly when $p(t)$ is near zero, we can say that i adopts with high probability once the proportion $p(t)$ passes the threshold

$$r_i = -m_i(0)/\tau_i^2 n\mu . \tag{17}$$

---

[8] A more realistic assumption would be that each agent observes a *subset* of outcomes, say those generated among his acquaintances. This modification does not change the aggregate dynamics in any fundamental way, since idiosyncratic variations among individual realizations are smoothed out when aggregated over the whole population.

[9] See, for example, DeGroot (1970).

(We shall consider the situation when $p(t)$ is small in a moment.) Observe that the number r_i an *ex ante* property of agent i: it depends on her initial beliefs, as well as on n and μ, all of which are fixed.[10] Thus, r_i functions like a *resistance* in the model considered in the previous section, and the aggregate dynamics are determined by the distribution of the resistances in the population. Specifically, let $F(r)$ be the c.d.f of r in the population, where r is derived from the initial beliefs as in (17). Then the expected motion of the discrete-time process is well-approximated by the difference equation (10), that is,

$$p(t+1) - p(t) = \alpha[F(p(t)) - p(t)],$$

where $0 < \alpha < 1$ and $F(0) > 0$. We claim that this remains a good approximation even in the start-up phase when $p(t)$ is small. The reason is that the process is initially driven forward by the optimists, who by assumption represent a positive fraction $F(0)$ of the population. In other words, adoption is initially driven by a population of $n(F(p(t)) - p(t))$ individuals, which is large even in the start-up phase when $p(t)$ is small; hence the preceding approximations remain valid.

The continuous-time analog is analogous to (11), namely,

$$\dot{p}(t) = \lambda[F(p(t)) - p(t)] \text{ for some } \lambda > 0.$$

While we derived this expression using normal-normal updating, this was not crucial to the argument: the essential point is that, as the proportion of adopters grows, the information they generate gradually overcomes the skepticism of

[10] If agent i obtains information only from a circle of s_i acquaintances, her threshold will instead be $r_i = -m_i(0) / \tau_i^2 s_i \mu$.

those who remain. This property holds in expectation for a variety of Bayesian updating procedures; for example, it holds when the outcome variable X is binomially distributed and agents have different priors about its mean (Lopez-Pintado and Watts, 2006). [11]

We now consider a variant of the preceding model in which the information generated by each prior adopter *accumulates over time*. Suppose, for example, that the innovation is a new medication whose efficacy can only be determined when taken over a substantial period of time. An agent who is deciding whether to adopt the medication will therefore be interested, not only in *how many* prior adopters there are, but *how long* each of them has been using it. In this and other situations, each adopter's outcome needs to be weighted by the length of time since he first adopted. If all adopters are weighted equally and there is no discounting, the total amount of information generated up to time t is found by *integrating* the adoption curve up to t, namely,

$$r(t) = \int_0^t p(s)ds \,.^{12} \tag{18}$$

Following the previous line of argument, we may suppose that each agent i has an information threshold or resistance $r_i \geq 0$, determined by his initial beliefs, such that he is adopts with high probability once the amount of information exceeds his threshold: $r(t) > r_i$. Letting $F(r)$ be the cumulative distribution function of r, we obtain the dynamical equation

[11] In earlier work, Jensen (1982) studied the special case in which X is binomial and the prior beliefs are uniformly distributed.

[12] More generally, $r(t) = \gamma \int_0^t p(s)ds$ where γ is the rate at which information is generated by prior adopters. For notational simplicity we shall take $\gamma = 1$.

$$\dot{p}(t) = \lambda[F(\int_0^t p(s)ds) - p(t)], \lambda > 0\,. \tag{19}$$

A process of this form will be called a *cumulative learning model,* in contrast to the previous class of processes, which will be called *non-cumulative learning models* or simply *threshold models*.

When λ and r are jointly distributed, we obtain a system analogous to (13), namely,

$$\forall \lambda \in [0,1], \quad \dot{p}_{\lambda}(t) = \lambda[F_{\lambda}(r(t)) - p_{\lambda}(t)], \text{ where } r(t) = \int_0^t \int_0^1 p_{\lambda}(s) d\nu ds . \qquad (20)$$

As before, a unique continuous solution is guaranteed on any finite interval $t \in [0,T]$, assuming that $F_{\lambda}(0) > 0$ for each value of λ. This system is typically more difficult to solve than the non-cumulative version, even when λ does not vary. However, when r is uniformly distributed (and λ is constant) we can obtain explicit solutions in both cases.

First, consider the threshold model generated by the uniform distribution $F(r) = ar + b$, where $a > 0$, $a \neq 1$, $0 < b < 1$, and $0 \leq r \leq (1-b)/a$. Integrating as in (12), it follows that on some interval $t \in [0,T]$, the unique solution is

$$p(t) = (b/(a-1))(e^{\lambda(a-1)t} - 1) . \qquad (21)$$

If $a > 1$ the path is exponentially increasing until the first time T such that $p(T) = 1$, after which $p(t)$ remains constant. If $a < 1$, the path is increasing, concave, and approaches the value $p = b/(1-a)$ asymptotically as $t \to \infty$.

Consider now the cumulative learning model generated by the same distribution. Differentiating (19) with respect to t, and using the fact that $p(t) = \dot{r}(t)$, we obtain

$$(1/\lambda)\ddot{p}(t) + \dot{p}(t) - ap(t) = 0. \qquad (22)$$

This is a second-order equation with constant coefficients. The initial condition is $p(0) = 0$; moreover from (19) we know that $\dot{p}(0) = \lambda b$. It follows that the unique solution is[13]

$$p(t) = [b/\sqrt{1+4a/\lambda}][e^{(\lambda/2)[-1+\sqrt{1+4a/\lambda}]t} - e^{-(\lambda/2)[1+\sqrt{1+4a/\lambda}]t}]. \qquad (23)$$

The shape of this curve is illustrated in Figure 1. Notice that it is concave initially, then convex; in other words it is *inverse-S-shaped*. Curves like this are probably quite rare in practice; in particular, empirical studies suggest that adoption curves usually *decelerate* as they approach their upper bound (Valente, 1995; Rogers, 2003).

[13] Substitute $p(t) = e^{\alpha t+\beta}$ into (22) and solve the resulting quadratic equation for α and β. Both roots appear in the solution (23) in order to satisfy the initial conditions.

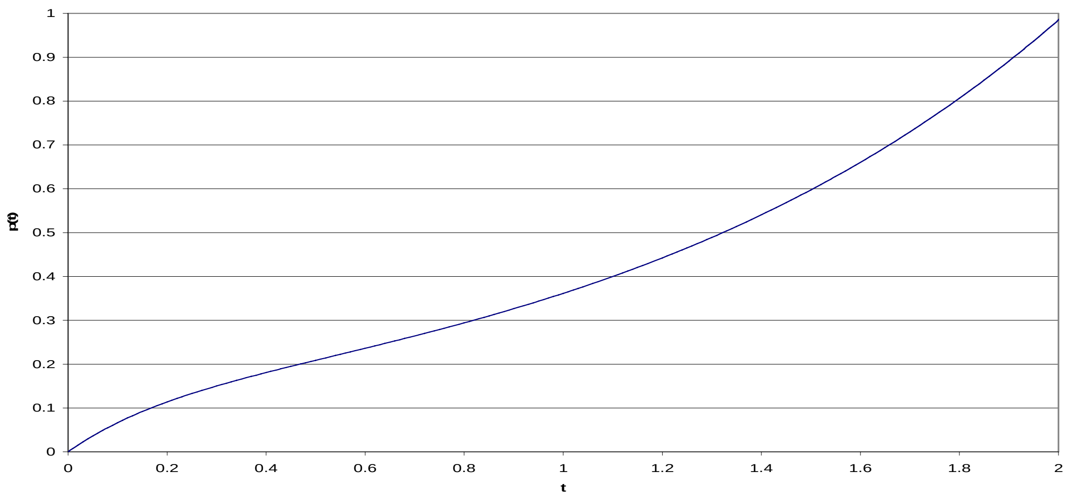

Figure 1. Cumulative learning dynamics generated by the uniform distribution $F(r) = 5r/4 + 1/5$ and $\lambda = 4$.

The peculiar behavior of this curve arises from the assumption that resistances are uniformly distributed. If the distribution of resistances has a sufficiently thin right tail, then deceleration eventually sets in, as one would expect. This effect is illustrated in Figure 2 for a normal distribution. Note, however, that the cumulative learning curve still *decelerates initially*, just as it did under the uniform distribution. It turns out that initial deceleration is a feature of the cumulative learning model *no matter what the distribution of resistances*, as we shall show in section 7. The reason is that cumulative learning attaches a lot of weight to information generated by very early adopters, of which there are very few, which creates an initial drag on the process.

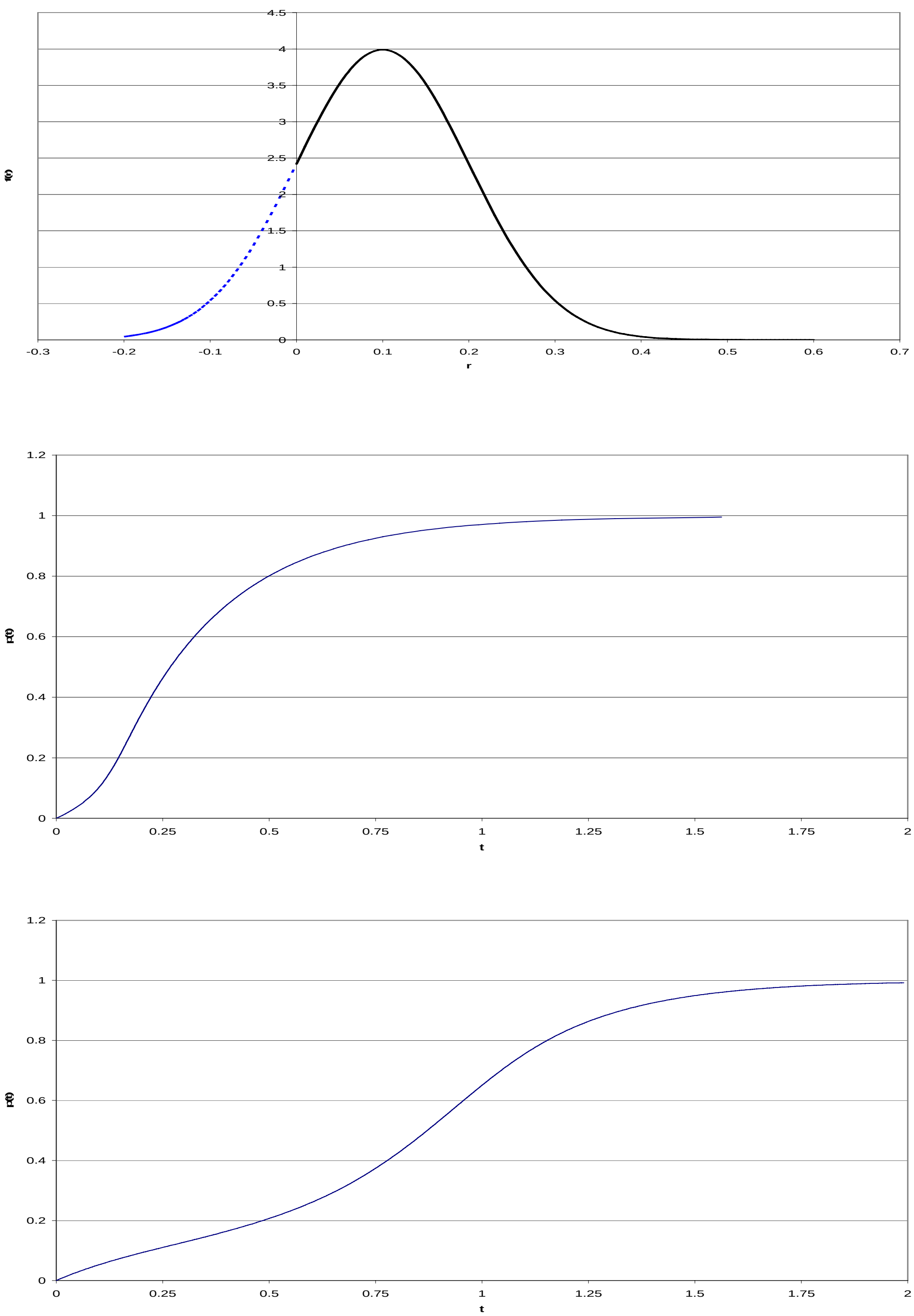

Figure 2. Top: normal distribution of resistances N(.10, .10); Middle: noncumulative learning curve $(\lambda = 4)$; Bottom: cumulative learning curve $(\lambda = 4)$.

6. Exogeneously driven moving equilibrium models

In contagion, conformity, and social learning models the process is propelled by a feedback loop between prior adopters and future adopters. In this section we consider models in which the dynamics are driven *solely* by changes in an exogenous variable and there is no internal feedback.

As an example, consider a new product whose price declines over time. If agents have different costs of adoption, those with the lowest costs will adopt first, then the next lowest, and so forth (David, 1966, 1969, 1975, 2003; David and Olsen, 1984, 1986; Stoneman, 2002). Another example would be increasing information about the new product that is generated from outside sources. As more becomes known about it, agents who were initially skeptical change their minds. Note that this is very similar to the learning model, except that in this case the information is generated exogenously. In general, let $\theta(t)$ be the value of an exogenous scalar parameter at time t, which is assumed to be monotonically increasing. Each agent adopts when $\theta(t)$ is large enough, and the crucial value of θ represents that agent's threshold. Heterogeneity is described by a cumulative distribution function $F(\theta)$ in the population of potential adopters. Thus the proportion of adopters at time t is

$$p(t) = F(\theta(t)). \tag{24}$$

Such a process is sometimes called a "moving equilibrium model" (David, 1969). Here I shall call it an *externally driven moving equilibrium model* to emphasize the importance of the external driving force. Indeed, the learning dynamics discussed earlier have an equal claim to being called moving equilibrium models, because at each point in time agents make optimal choices given their information. The crucial difference between the two approaches is that, in learning models, the key parameter that propels the process forward is information generated *internally* by prior adopters, whereas in the models discussed here the process is propelled by changes in an *exogenous* parameter. For example, if $\theta(t)$ represents the mean realized payoff from a series of trials external to the group, and θ is the resistance of a given individual within the group as determined by his prior beliefs, then the exogenous aspect is the only essential difference between this and the previous class of social learning models.

Unlike social learning models and the other models discussed above, there is no differential equation to solve in this case: $p(t)$ is simply the composition of two monotone increasing functions $F(\theta)$ and $\theta(t)$. Since any monotone increasing function $p(t)$ can be expressed in many ways as the composition of two monotone functions, the model has no predictive power without more knowledge about the nature of $F(\theta)$ and $\theta(t)$. Nevertheless, if we assume a definite form for $\theta(t)$, for example that it is linear, then something interesting can be said about the acceleration of the process without placing undue restrictions on F, as we shall see towards the end of the next section.

7. Acceleration analysis

In this section we shall show that each of the models introduced above leads to predictions about the *acceleration pattern* of the diffusion curve that require few, if any, restrictions on the underlying distribution of heterogeneous characteristics. The key measures that we shall study are the rate of acceleration $\ddot{p}(t)$ and the relative rate of acceleration $\ddot{p}(t)/\dot{p}(t)$.

To fix ideas, let us first consider the threshold model when there is no heterogeneity. As we have already seen, the differential equation describing such a process is given by

$$\dot{p}(t) = \lambda[F(p(t)) - p(t)], \tag{25}$$

where $F(r)$ is the distribution function of "resistance" in the population. As before, we shall assume that $F(r) > r$ on some initial interval $0 \leq r < b$, for otherwise the process cannot get started. In this case $\dot{p}(t)$ is positive over some initial interval $[0, T]$. Assume that $F(r)$ has a continuous density $f(r)$ on the open interval $r \in (0,1)$. (Note that, since $F(0) > 0$, the density f is not defined at $r = 0$.) Differentiating (25) with respect to t and dividing through by $\dot{p}(t)$, which by assumption is positive, we obtain

$$\forall t \in [0, T], \qquad \ddot{p}(t) / \dot{p}(t) = \lambda[f(p(t)) - 1]. \tag{26}$$

It follows that the relative acceleration rate traces out a positive linear transformation of the underlying density.

Notice that the process has positive acceleration if and only if the density is large enough, namely, if and only if $f(r) > 1$. Suppose further that $f(r)$ is unimodal -- first strictly increasing and then strictly decreasing. Then the relative acceleration rate exhibits the same pattern. In particular, the start-up phase of the process exhibits *superexponential growth*. (When $\ddot{p}(t) / \dot{p}(t)$ is constant $p(t)$ grows exponentially, and when $\ddot{p}(t) / \dot{p}(t)$ is strictly increasing $p(t)$ grows superexponentially.)

This phenomenon results from the compounding of two effects. First, as more and more people adopt, the amount of information available to the remainder of the population increases. Second, the number of people persuaded by each additional bit of information increases as the process moves up the left tail of the distribution of resistances. These conclusions continue to hold when λ is heterogeneously distributed, under quite weak non-parametric assumptions on the joint distribution of resistances and inertia levels. $F_{\lambda}(r)$.

Proposition 1. *Suppose that diffusion is driven by a heterogeneous threshold model, where for each level of inertia* $\lambda \in (0,1]$ *the conditional distribution of resistance has a density* $f_{\lambda}(r)$ *that is continuous and bounded, and* $F_{\lambda}(0) > 0$. *Let* $\overline{f}_{\lambda}(0) = \lim_{r \to 0^{+}} f_{\lambda}(r)$. *Then:*

i) if $\overline{f}_{\lambda}(0) > 1$ *for every* λ, *the process initially accelerates;*

ii) if the conditional densities $f_{\lambda}(r)$ *are strictly increasing on an open interval* $(0,\overline{r})$, *the relative acceleration rate* $\ddot{p}(t)/\dot{p}(t)$ *is strictly increasing on the same interval.*

Proof. The equations of motion are

$$\dot{p}_{\lambda}(t) = \lambda[F_{\lambda}(p(t)) - p_{\lambda}(t)]. \tag{27}$$

Hence for $t > 0$,

$$\ddot{p}_{\lambda}(t) = \lambda[f_{\lambda}(p(t)) - 1]\dot{p}_{\lambda}(t) \tag{28}$$

and

$$\ddot{p}(t) = \int \lambda[f_{\lambda}(p(t)) - 1]\dot{p}_{\lambda}(t)d\nu . \tag{29}$$

Letting $t \to 0^{+}$ and using the continuity of $p(t)$ and $\ddot{p}(t)$ we have

$$\ddot{p}(0) = \int \lambda^2 [\overline{f}_\lambda(0) - 1] F_\lambda(0) d\nu \,, \tag{30}$$

from which i) follows immediately.

To establish ii), notice that (28) implies $\ddot{p}_\lambda(t)/\dot{p}_\lambda(t) = \lambda[f_\lambda(p(t)) - 1]$ for every λ. By the hypothesis in ii), the functions $f_\lambda(p(t))$ are strictly increasing on some interval $0 < p(t) < \overline{r}$. It follows that, for every λ and all t in a suitable interval $(0, T]$,

$$\dddot{p}_\lambda(t)/\dot{p}_\lambda(t) - (\ddot{p}_\lambda(t)/\dot{p}_\lambda(t))^2 > 0 \,, \tag{31}$$

that is,

$$\sqrt{\dddot{p}_\lambda(t)\dot{p}_\lambda(t)} > \ddot{p}_\lambda(t) \,. \tag{32}$$

Hence

$$\int \sqrt{\dddot{p}_\lambda(t)\dot{p}_\lambda(t)} d\nu > \int \ddot{p}_\lambda(t) d\nu = \ddot{p}(t) \,. \tag{33}$$

By Hölder's Inequality,

$$\sqrt{\dddot{p}(t)\dot{p}(t)} = \left(\int \dddot{p}_\lambda(t) d\nu \int \dot{p}_\lambda(t) d\nu\right)^{1/2} \geq \int \sqrt{\dddot{p}_\lambda(t)\dot{p}_\lambda(t)} d\nu \,. \tag{34}$$

Combining this with (33) we conclude that $\sqrt{\dddot{p}(t)\dot{p}(t)} > \ddot{p}(t)$, which implies that $\ddot{p}(t)/\dot{p}(t)$ is strictly increasing on $(0, T]$. This concludes the proof of the proposition.

We now show that the cumulative learning model always decelerates initially. We shall first run through the argument assuming *constant* λ; it will then be clear how to generalize it to the heterogeneous case. Let $F(r)$ be the distribution function of resistance r and let $r(t) = \int_0^t p(s) ds$. Assume that $F(r)$

has a differentiable density $f(r)$ such that $f'(r)$ is continuous and bounded on $0 < r \le 1$. Differentiating the defining equation (19) with respect to t, and recalling that $p(t) = \dot{r}(t)$, we obtain the following expression

$$(1/\lambda)\ddot{p}(t) = p(t) f(r(t)) - \dot{p}(t)\,. \qquad (35)$$

The solution $p(t)$ is continuous and therefore $\lim_{t \to 0^+} p(t) = p(0) = 0$. Hence $f(r(t))p(t)$ approaches zero as $t \to 0^+$. We also know from (19) that $\dot{p}(0) = \lambda F(0) > 0$. It follows from this and (35) that

$$\lim_{t \to 0^+} \ddot{p}(t) = -\lambda^2 F(0) < 0\,. \qquad (36)$$

In short, the adoption curve must be *decelerating* in a neighborhood of the origin. (Figures 1 and 2 illustrate this phenomenon.) The reason is that the initial block of optimists $F(0)$ exerts a decelerative drag on the process: they contribute at a decreasing rate as their numbers diminish, while the information generated by the new adopters gathers steam fairly slowly because at first there are so few of them. These arguments continue to hold when there is heterogeneity in λ, as the reader may verify.

Next we shall show that the relative acceleration rate is strictly increasing in a neighborhood of the origin provided that $\overline{f}(0) = \lim_{r \to 0^+} f(r) > 0$. Let $\phi(t) = (1/\lambda)\ddot{p}(t)/\dot{p}(t)$. From (35) we deduce that

$$\phi(t) = f(r(t))p(t)/\dot{p}(t) - 1\,. \qquad (37)$$

Differentiating (37) we obtain

$$\dot{\phi}(t) = f'(r(t))p^2(t)/\dot{p}(t) + f(r(t)) - f(r(t))\ddot{p}(t)p(t)/\dot{p}^2(t). \qquad (38)$$

As $t \to 0^+$ the first term goes to zero, because by assumption f' is bounded, $p(t) \to 0$, and $\dot{p}(0) > 0$. The third term also goes to zero. However, $f(r(t)) \to \overline{f}(0) > 0$, so the second term is positive in the limit. It follows from continuity that $\dot{\phi}(t)$ is strictly positive on some initial interval $0 \le t \le T$.

The reader may verify that a similar argument holds in the heterogeneous case provided that all of the conditional distributions satisfy $\overline{f}_\lambda(0) = \lim_{r \to 0^+} f_\lambda(r) > 0$, and the derivatives are bounded and continuous in a neighborhood of zero. These findings are summarized in the following.

Proposition 2. *Suppose that diffusion is driven by cumulative social learning, where for each level of inertia $\lambda \in (0,1]$ the conditional distributions of resistance satisfy $F_\lambda(0) > 0$ and $\overline{f}_\lambda(0) = \lim_{r \to 0^+} f_\lambda(r) > 0$, and the derivatives $f'_\lambda(r)$ are continuous and bounded on $0 < r \le 1$. Then initially the process strictly decelerates whereas the relative acceleration rate strictly increases.*

Next we shall study the shape of the curves generated by heterogeneous contagion. It turns out that a key statistic in this case is the *hazard rate* $h(t) = \dot{p}(t)/(1 - p(t))$. Consider a heterogeneous contagion model where μ is the joint distribution of the internal and external contagion parameters λ and γ. As before we shall assume that the support of μ lies in the unit square $\Omega = [0,1]^2$. To assure that the process gets started, we shall also suppose that $\overline{\gamma} = \int_\Omega \gamma d\mu > 0$.

Proposition 3. *Suppose that diffusion is driven by heterogeneous contagion with joint distribution* μ *on the parameters* $(\lambda,\gamma)\in[0,1]^2$ *such that* $\bar{\gamma}>0$. *For all* $t>0$, $h(t)/p(t)$ *is strictly decreasing in* t, *equivalently,*

$$\forall t>0, \qquad \ddot{p}(t)/\dot{p}(t)<(1-2p(t))h(t)/p(t)\,. \tag{39}$$

Furthermore,

$$\bar{\lambda}<\bar{\gamma}\Rightarrow \ddot{p}(0)<0\,. \tag{40}$$

Corollary 3.1. *If* $p(t)$ *is generated by heterogeneous contagion, then* $\dot{p}(t)/p(t)$ *is strictly decreasing.*

The corollary follows immediately from the fact that $h(t)/p(t)$ is strictly decreasing: namely, if $h(t)/p(t)<h(t')/p(t')$ for some $t<t'$, then $\dot{p}(t)/p(t)<\dot{p}(t')/p(t')$. Notice, however, that this does not necessarily imply that the relative acceleration rate $\ddot{p}(t)/\dot{p}(t)$ is strictly decreasing. Rather, the fact that $h(t)/p(t)$ is decreasing implies that $\ddot{p}(t)/\dot{p}(t)$ is bounded above (see (39)), where the bound goes to zero as $p(t)$ approaches one-half. These predictions should be straightforward to check given sufficient empirical data.

Proof of proposition 3. Define the function $H(t)=h(t)/p(t)$: this is well-defined for all $t>0$ because $\dot{p}(0)=\bar{\gamma}>0$ and hence $p(t)>0$. To establish the first claim of the proposition we shall show that $\dot{H}(t)<0$.

For each parameter pair (λ,γ) let $q_{\lambda,\gamma}(t)=(1-p_{\lambda,\gamma}(t))$ denote the *proportion* of the (λ,γ)-population *that has not yet adopted by* time t. The proportion of the total population that has not adopted by t is therefore

$$q(t) = \int q_{\lambda,\gamma}(t) d\mu \ . \tag{41}$$

For each (λ,γ) we have

$$\dot{p}_{\lambda,\gamma}(t) = (\lambda p(t) + \gamma) q_{\lambda,\gamma}(t) \ . \tag{42}$$

Integration with respect to μ yields

$$\dot{p}(t) = [\lambda(t) p(t) + \gamma(t)] q(t) , \tag{43}$$

where

$$\lambda(t) = q^{-1}(t) \int \lambda q_{\lambda,\gamma}(t) d\mu \text{ and } \gamma(t) = q^{-1}(t) \int \gamma q_{\lambda,\gamma}(t) d\mu \ . \tag{44}$$

Note that $\lambda(t)$ and $\gamma(t)$ are the *expected* values of λ and γ in the population of non-adopters at time t. It follows that

$$H(t) = \dot{p}(t) / [p(t) q(t)] = \lambda(t) + \gamma(t) / p(t) \ . \tag{45}$$

Claim: *For every* $t > 0$, $\dot{\lambda}(t) p(t) + \dot{\gamma}(t) \leq 0$. (46)

Proof of claim. For every $t > 0$ we have

$$\dot{\lambda}(t) = \frac{\int \lambda \dot{q}_{\lambda,\gamma}(t) d\mu}{\int q_{\lambda,\gamma}(t) d\mu} - \frac{\int \lambda q_{\lambda,\gamma}(t) d\mu \int \dot{q}_{\lambda,\gamma}(t) d\mu}{[\int q_{\lambda,\gamma}(t) d\mu]^2} , \tag{47}$$

and

$$\dot{\gamma}(t) = \frac{\int \gamma \dot{q}_{\lambda,\gamma}(t) d\mu}{\int q_{\lambda,\gamma}(t) d\mu} - \frac{\int \gamma q_{\lambda,\gamma}(t) d\mu \int \dot{q}_{\lambda,\gamma}(t) d\mu}{[\int q_{\lambda,\gamma}(t) d\mu]^2}. \tag{48}$$

To show that $\dot{\lambda}(t) p(t) + \dot{\gamma}(t) \leq 0$, multiply (47) by $p(t)$ and add it to (48); after simplifying we obtain the equivalent condition

$$\int (\lambda p(t) + \gamma) \dot{q}_{\lambda,\gamma}(t) d\mu \int q_{\lambda,\gamma}(t) d\mu - \int (\lambda p(t) + \gamma) q_{\lambda,\gamma}(t) d\mu \int \dot{q}_{\lambda,\gamma}(t) d\mu \leq 0. \tag{49}$$

(Notice that t does not vary in this expression; t is fixed and integration is taken with respect to λ and γ.)

We know from (42) that $\dot{q}_{\lambda,\gamma}(t) = -(\lambda p(t) + \gamma) q_{\lambda,\gamma}(t)$ for every λ, γ, and t. Substituting this into (49) we obtain

$$\int (\lambda p(t) + \gamma)^2 q_{\lambda,\gamma}(t) d\mu \int q_{\lambda,\gamma}(t) d\mu \geq [\int (\lambda p(t) + \gamma) q_{\lambda,\gamma}(t) d\mu]^2. \tag{50}$$

Fix $t > 0$ and define the random variables

$$X = (\lambda p(t) + \gamma) \sqrt{q_{\lambda,\gamma}(t)} \text{ and } Y = \sqrt{q_{\lambda,\gamma}(t)}. \tag{51}$$

The realizations of X and Y are determined by random draws from μ. Then (50) follows directly from Schwarz's inequality: $E[X^2]E[Y^2] \geq (E[XY])^2$. This establishes the claim.

We now use this result to show that $H(t)$ is strictly decreasing in t for all $t > 0$. Direct differentiation leads to

$$\dot{H}(t) = \dot{\lambda}(t) + \dot{\gamma}(t) / p(t) - \gamma(t)\dot{p}(t) / p^2(t) . \tag{52}$$

By the above claim, $\dot{\lambda}(t)p(t) + \dot{\gamma}(t) \le 0$, so division by $p(t) > 0$ yields $\dot{\lambda}(t) + \dot{\gamma}(t)/p(t) \le 0$. Thus the sum of the first two terms on the right-hand side of (52) is nonpositive. But the last term is strictly negative, because $\gamma(t) > 0$ for all $t > 0$ given the initial condition $\overline{\gamma} = \gamma(0) > 0$. Hence $H(t)$ is strictly decreasing in t, which establishes the first claim of the proposition. Expression (39) is an immediate consequence of the fact that $\dot{H}(t) < 0$.

To prove (40), recall that

$$\dot{p}(t) = \int \dot{p}_{\lambda,\gamma}(t) d\mu = \int (\lambda p(t) + \gamma) q_{\lambda,\gamma}(t) d\mu . \tag{53}$$

Differentiate (53) and evaluate it at $t = 0$ to obtain

$$\ddot{p}(0) = \dot{p}(0) \int \lambda q_{\lambda,\gamma}(0) d\mu + p(0) \int \lambda \dot{q}_{\lambda,\gamma}(0) d\mu + \int \gamma \dot{q}_{\lambda,\gamma}(0) d\mu . \tag{54}$$

Now use the fact that $p(0) = 0$, $\dot{p}(0) = \overline{\gamma}$, and $\dot{q}_{\lambda,\gamma}(0) = -\gamma$ to deduce that

$$\begin{aligned} \ddot{p}(0) &= \overline{\gamma}\overline{\lambda} - \int \gamma^2 q_{\lambda,\gamma}(0) d\mu \\ &= \overline{\gamma}\overline{\lambda} - \overline{\gamma}^2 - \sigma_\gamma^2 \le \overline{\gamma}(\overline{\lambda} - \overline{\gamma}) . \end{aligned} \tag{55}$$

Hence $\ddot{p}(0) < 0$ if $\overline{\lambda} < \overline{\gamma}$. This concludes the proof of Proposition 3.

There exist perfectly reasonable S-shaped curves for which $h(t)/p(t)$ is *strictly monotone increasing*, and which are therefore inconsistent with a heterogeneous contagion model for *any* distribution of the contagion parameters. Consider, for example, curves of form $\dot{p}(t) = p^a(t)(1-p(t))$, which were first proposed by Easingwood, Mahajan, and Muller (1981, 1983). When $a > 1$, $h(t)/p(t) = p^{a-1}(t)$ is strictly increasing, hence the process cannot arise from a heterogeneous contagion model. Yet it generates adoption curves that look superficially very much like other S-shaped curves, including some, like the Bass model, that do arise from contagion (see Figure 3). The differences are only revealed by studying the behavior of the first and second order derivatives.

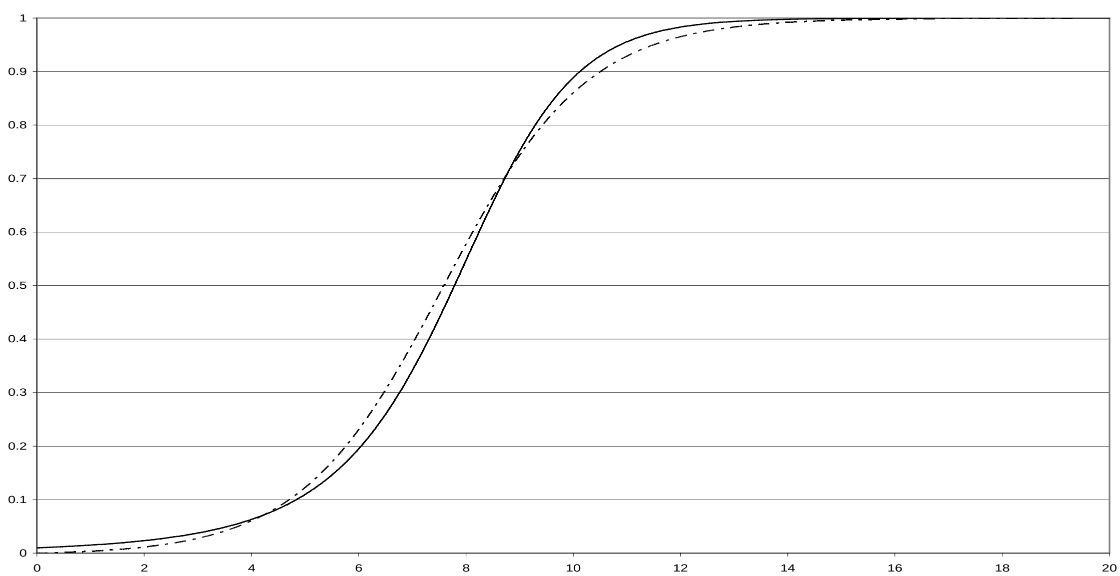

Figure 3. Two adoption curves: the solid line is generated by $\dot{p}(t) = p^{1.2}(1-p)$ and $p(0) = 0.01$, the dashed line by the Bass model with $\lambda = .75$ and $\gamma = .0025$.

Finally, let us turn to the externally driven moving equilibrium models. Somewhat less definitive statements can be made in this case due to the extremely general nature of the model; nevertheless something can be said.

Recall from (24) that such a process takes the form $p(t) = F(\theta(t))$, where $F(\theta)$ is the distribution function of a scale parameter θ, which increases over time. As with the previous models, we are interested in analyzing the dynamics without making restrictive assumptions about the distribution of heterogeneity. Notice that in this case the heterogeneity is embodied entirely in $F(\theta)$; to say anything about the dynamics we must first specify the behavior of $\theta(t)$. The simplest assumption is that θ increases at a constant rate, that is, $\dot{\theta} > 0$ and $\ddot{\theta} = 0$. In this case $p(t)$ traces out a portion of the cumulative distribution curve starting at $p(0) = F(\theta(0))$. If the density $f(\theta)$ is unimodal, this will be an S-shaped curve. To make comparisons with the previous models, note that

$$\ddot{p}(t)/\dot{p}(t) = af'(\theta(t))/f(\theta(t)), \text{ where } a = \dot{\theta}(t) > 0. \tag{56}$$

For many common distributions $f'(\theta(t))/f(\theta(t))$ is nonincreasing, that is, the density $f(\theta)$ is logconcave. These include the normal, lognormal, exponential, and uniform distributions. For all of these distributions the relative acceleration rate $\ddot{p}(t)/\dot{p}(t)$ will be nonincreasing. This stands in marked contrast with heterogeneous threshold models, where the relative acceleration rate is increasing whenever the density $f(\theta)$ is increasing.

Proposition 4. *Suppose that* $p(t)$ *arises from an externally driven moving equilibrium model* $p(t) = F(\theta(t))$, where $\dot{\theta}(t) > 0$, $\ddot{\theta}(t) = 0$, *and the density* $f(\theta)$ *is continuously* *differentiable and logconcave. Then the relative acceleration rate* $\ddot{p}(t)/\dot{p}(t)$ *is nonincreasing.*

8. Summary

In this paper we have studied five families of diffusion models, and shown how to solve them for completely general distributions of the underlying heterogeneous characteristics. Each family of models has a distinctive pattern of acceleration, as shown in Table 1. In situations where good micro-level adoption data are not available, this framework has the potential for assessing the relative plausibility of different diffusion models based on the behavior of the aggregate dynamics. Of course, actual tests of significance would require a detailed analysis of the error structure in a finite-population setting, which we have side-stepped in order to study the mean-field dynamics. The extension of the results to a fully stochastic framework, and their application to empirical adoption curves, will be treated separately.

Model	Footprint	Restrictions on distribution
1. Inertia	$\ddot{p}(t) < 0$	none
2. Threshold	$\ddot{p}(0) > 0$ $\ddot{p}(t) / \dot{p}(t) \uparrow$ initially	density initially increasing and greater than unity
3. Cumulative learning	$\ddot{p}(0) < 0$ $\ddot{p}(t) / \dot{p}(t) \uparrow$ initially	none density initially increasing
4. External moving equilibrium $F(\theta(t))$	$\ddot{p}(0) > 0$ $\ddot{p}(t) / \dot{p}(t) \downarrow$ initially	density initially increasing and logconcave; $\theta(t)$ linear
5. Contagion	$\dot{p}(t) / p(t) \downarrow$ $\ddot{p}(t) / \dot{p}(t) < (1-2p(t))h(t) / p(t)$	none

Table 1. Diffusion models and their acceleration 'footprints.' Non-cumulative learning and conformity fall under the heading of threshold models.

References

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

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.

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.

Centola, Damon (2006), "Free riding and the weakness of strong incentives," Working Paper, Department of Sociology, Cornell University.

Coddington, E.A., and N. Levinson (1955), *Theory of Ordinary Differential Equations*. New York: McGraw-Hill.

David, Paul A. (1966), "The mechanization of reaping in the ante-bellum Midwest," In Henry Rosovksy, ed., *Industrialization in Two Systems: Essays in Honor of Alexander Gershenkron*. New York: Wiley.

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 (1984), "Anticipated automation: a rational expectations model of technological diffusion," Technological Innovation Project Working Papers No. 2, Center for Economic Policy Research, Stanford University.

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.

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

Dixon, Robert (1980), "Hybrid corn revisited," *Econometrica* 48, 1451-1461.

Easingwood, Christopher J., Vijay Mahajan, and Eitan Muller (1981), "A non-symmetric responding logistic model for technological substitution," *Technological Forecasting and Social Change*, 20, 199-213.

Easingwood, Christopher J., Vijay Mahajan, and Eitan Muller (1983), "A nonuniform influence innovation diffusion model of new product acceptance," *Marketing Science*, 2, 273-296.

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

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

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

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

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

Jeuland, Abel P. (1981). "Incorporating heterogeneity into parsimonious models of diffusion of innovation," University of Chicago Center for Research on Marketing, Working Paper No. 45.

Karshenas, M., and Paul Stoneman (1992), "A flexible model of technological diffusion incorporating economic factors with an application to the spread of colour television ownership in the UK," *Journal of Forecasting*, 11, 577-601.

Lang, Serge (1999). *Fundamentals of Differential Geometry*. New York: Springer-Verlag.

Lopez-Pintado, Dunia, and Duncan J. Watts (2006), "Social Influence, Binary Decisions, and Collective Dynamics," Working Paper, Institute for Social and Economic Research and Policy, Columbia University.

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

Mahajan, Vijay and Robert A. Peterson (1985), *Models for Innovation Diffusion*. Beverly Hills: Sage Publications.

Mahajan, Vijay, Eitan Muller, and Frank M. Bass (1990), "New product diffusion models in marketing: a review and directions for further research," *Journal of Marketing*, 54, 1-26.

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

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

Schelling, Thomas C. (1971), "Dynamic models of segregation," Journal of Mathematical Sociology, 1, 143-186.

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

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

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, *Models and Methods in Social Network Analysis*. Cambridge UK: Cambridge University Press.

Young, H. Peyton (2006), “The spread of innovations through social learning,” Johns Hopkins University, Mimeo.