Boom and Bust Patterns in the Adoption of Financial Innovations

John C. Persons
Ohio State University

Vincent A. Warther
University of Michigan

We develop a dynamic model of the adoption of financial innovations. Each period, firms decide whether or not to adopt an innovation of uncertain value, and the profitability of each period's adoptions reveals information about the innovation's value. We show that characteristics of financial innovation waves cited by critics as evidence of irrational excess are, in fact, consistent with fully rational behavior. We also show that social welfare is enhanced when more firms adopt innovations of questionable value and that financial intermediaries have an incentive to encourage such adoption.

Financial markets seem to be afflicted with a boom and bust syndrome, and it is often asserted that these episodes are manifestations of irrational excess. In this article, we examine the dynamics that result from the introduction of a financial innovation, and find that fully rational decision making by well-informed agents produces many aspects of the boom and bust

---

We would like to thank Harry DeAngelo, Ming Dong, David Hirshleifer (the editor), John Matsusaka, René Stulz, two anonymous referees, and seminar participants at UBC, Carnegie-Mellon, Chicago, Northwestern, Ohio State, USC, Washington University, and the AFA and WFA annual meetings for their helpful comments. Persons gratefully acknowledges financial support from the Charles A. Dice Center for Research in Financial Economics and from a Dean's Summer Fellowship at the Fisher College of Business. Address correspondence to Vincent Warther, University of Michigan Business School, 701 Tappan, Ann Arbor MI 48109-1234; e-mail: vwarther@umich.edu.

The Review of Financial Studies Winter 1997 Vol. 10, No. 4, pp. 939–967

patterns that we observe. That an innovation creates a boom of activity followed by a sudden crash is not cause for alarm, nor is it a solid basis for questioning the rationality of market participants. It should rather be viewed as the natural result of learning in a well-functioning market.[1]

The leveraged buyout (LBO) wave of the 1980s is perhaps the most familiar example of the type of innovation-driven boom and bust cycle that we study. Sparked by innovations in the junk-bond market, the dollar volume of LBO deals went from $1 billion in 1980 to a peak of $60 billion in 1988 and then crashed to less than $1 billion in 1991. Many commentators have suggested that the LBO wave exhibited signs of irrationality such as "overpricing" and "overshooting."[2] Indeed, the very terms "boom" and "bust" are synonymous for many with irrationality. Another spectacular bust occurred recently in the market for collateralized mortgage obligations, after a dramatic boom in the early 1990s. Following serious losses by many CMO investors, the rate of issuance fell from more than $300 billion in 1993 to about $20 billion in 1995 [Jereski (1997)].

We construct a parsimonious model which incorporates the salient features of a financial innovation wave: (1) there is both systematic and firm-specific uncertainty about the value of a newly introduced innovation; (2) the expected gains from adoption vary across firms; (3) information about the innovation's value comes from the experience of adopters; and (4) adoption by one firm does not affect the profitability of adoption by others.

The first two features are common to virtually all innovations, but we consider the last two much more typical of financial innovations than other innovations, making this model particularly useful for understanding financial innovations. With respect to (3), it is clear a priori that some innovations are profitable (e.g., higher-speed computer chips or more efficient manufacturing equipment), and small-scale experiments can be performed to learn about the value of other innovations (e.g., new oil-refining techniques or the Dvorak keyboard). However, the only way to learn about the value of many innovations (such as new takeover defenses or new forms of corporate organization) is to observe what happens to those who adopt. Because experiments are either impossible or prohibitively expensive for most

[1] For different perspectives on the forces behind the recent surge in financial innovations, see Merton (1992), Miller (1986), and Ross (1989). Van Horne (1985) discusses the explosion of innovations and attributes it partly to irrationality.

[2] See Jensen (1991) for this point of view. Kaplan and Stein (1993) offer evidence to support this assessment and bluntly state that "the story is . . . one that does not fit comfortably with traditional notions of efficient markets."

financial innovations, they are characterized by learning from the experience of previous adopters. With respect to (4), the use of many nonfinancial innovations can be restricted by property rights, secrecy, competitive advantage, or the need for specialized inputs. Financial innovations are rarely patentable, and often must be understood by a wide range of users to be successful. Hence one firm's adoption or development of an innovation does not exclude or impede others from also adopting it. In Section 7, we discuss how much of the related literature analyzes cases where one of these two features is absent.

In our model, an innovation of uncertain value is introduced at date 0, and the expected payoff from adopting the innovation varies across firms. Although the true expected payoff is uncertain, differences in expected payoffs across firms are common knowledge, so everyone understands which firms are best suited to the innovation. Each period, every firm that has not adopted previously must decide whether to adopt the innovation; adopting is a one-time event for each firm. When a firm adopts the innovation, the uncertainty about its payoff is resolved in the following period. The realized payoff is the true expected payoff plus a random shock. Because a component of the payoff is common across firms, the remaining firms learn about the value of the innovation by observing the payoffs of those who adopted last period. A lot of information is generated if many firms adopt the innovation, and little information is generated if few firms adopt. Specifically, the precision of the signal that firms receive about the value of the innovation is proportional to the volume of adoption. Each period the remaining firms update their priors, and the process repeats. If the realized profits are good enough, some more firms adopt the next period, but if the realized profits are bad enough, no additional firms adopt. If no firms adopt, no new information about the innovation is generated, so the system comes to a halt.

When deciding whether to adopt, a firm faces a simple trade-off. Even though the expected payoff from adopting is positive, it might be better to wait and gain a more precise estimate of the value of the innovation by observing the payoffs of other firms. Waiting allows the firm to avoid some mistakes — the firm might never adopt the innovation if the realizations turn out to be poor. On the other hand, the cost of waiting an additional period is the time value of money, since payoffs that are expected to be positive are delayed for a period.

Because information is generated endogenously by firms' adoption decisions, rational innovation adoption naturally progresses in a boom and bust pattern — hence our use of the term "wave." The feedback between the volume of adoption and the amount of information generated produces interesting asymmetric dynamics. When initial adoptions are very profitable, subsequent adoption volume surges, and information revelation surges. When initial adoptions are disappointing, adoption volume drops, and information revelation falls off.

The model's structure is straightforward in that all participants are fully rational and all information is public. Nevertheless, the model generates dynamics similar to those observed in many innovation waves, so rational firm behavior may seem irrational to the casual observer. Below are examples of (rational) behavior that occurs in our setting, which may appear on the surface to be evidence of fads.

- Some firms will pass up an innovation today, when its estimated value is (say) v, then adopt the innovation next period when its estimated value is *less than* v (because bad news arrived in the interim). This behavior is puzzling because firms appear to blindly follow the lead of previous adopters while ignoring negative information. It is perfectly rational.
- Innovation waves always end badly because the last firms to adopt in a wave lose money on average. Critics may view this as proof of "overheating" or "overshooting" and mistakenly conclude that participants were caught up in an irrational fad.
- The end of an innovation wave is completely unpredictable. The probability that a wave continues for n additional periods is constant over time, regardless of the length of time since the innovation was introduced; the number of firms that adopted previously; the profits realized by previous adopters; the current estimated value of the innovation; and the precision of the current estimate of value. Thus, waves appear to be fragile because they often end abruptly rather than winding down in an "orderly" manner, and observers may misconstrue this as irrational excess.
- Even though the distribution of quality across innovations is symmetric, *most* innovations perform worse than expected. This could erroneously be cited as proof that adopters are too optimistic. In fact, the skewed distribution arises naturally when successful adoptions lead to an increase in future adoption volume.

The model also shows that, from a social perspective, there is too little adoption of financial innovations. A particular firm's decision to adopt generates information that other firms use to improve their estimate of the innovation's value, and this helps them make better adoption decisions. Adoption therefore produces a positive informational externality, and more adoption improves welfare. Regulators should encourage experimentation with new innovations, though their inclination may be to restrict such experimentation. Further, we show that a private financial intermediary who can capture some of the profits of future adopters may produce some of the welfare improvements that a benevolent social planner would achieve. Financial intermediaries who entice firms into adopting untried innovations are not socially destructive hucksters, but are, instead, performing a useful social function.

Finally, we show that when either the initial uncertainty or the rate of learning from adoptions is great enough, the average "diffusion curve" — the plot of cumulative adoptions against time — has an S shape. This is consistent with the results of empirical studies of major technological innovations [e.g., Mansfield (1961, 1968)]. This result does not rely on a thin-tailed distribution of firm types — rather we get this result assuming a *uniform* distribution of firm types.

The article proceeds as follows. Section 1 discusses applications of the model in more detail. Section 2 describes the model. Section 3 analyzes the model and derives the privately optimal adoption decision. Section 4 illustrates several facets of rational adoption behavior that could be mistaken for irrationality or fads. Section 5 shows that the average "diffusion curve" generated by the model matches the stylized facts from empirical studies of major technological innovations. Section 6 shows that individual firms choose too little innovation, so policymakers may be able to increase social welfare by encouraging adoption of innovations. Section 7 discusses the related literature, and Section 8 concludes. Some proofs are in the Appendix and the rest are available from the authors.

1. Applications

Good applications of the model are innovations that are characterized by assumptions (3) and (4) discussed in the introduction. That is, an ideal application is one in which all learning about the value of the innovation comes from observing the experience of adopters, and one in which there are no interactions among adopters — one firm's decision to adopt does not affect another firm's ability to adopt nor its payoff from adopting. There are many nice applications in the area of corporate control beyond the LBO wave mentioned earlier. These include the waves of horizontal mergers in the 1890s, vertical mergers in the 1920s, conglomerate mergers in the 1960s, and hostile takeovers and the resulting defensive innovations in the 1980s. In the case of defensive tactics, we saw innovations such as poison pills that proved to be quite effective and sparked a sustained boom in usage [Malatesta and Walkling (1988), Ryngaert (1988)], and we saw other innovations, such as greenmail, that faded in popularity as they proved ineffective at deterring takeovers [Bradley and Wakeman (1983), Eckbo (1990)]. For each of these, the only way to find out how well the innovation worked was through experiential learning — some firms had to try it out, and everyone else learned about its value from the success or failure of the early adopters. These innovations could not be patented or kept secret, so other firms were free to adopt after learning from the early outcomes. Furthermore, payoff interactions among adopters are minimal — the benefits to one firm of conglomerate mergers or poison pills do not depend in important ways on other firms' decisions to adopt.

Our model is very well suited to new corporate financing tactics; recent examples include the relatively successful Dutch auction repurchase method [Bagwell (1992), Persons (1994)] and the unsuccessful STAC (simultaneous tender and call) strategy for retiring nonrefundable debt [Dillon, Noe and Ramirez (1995)]. Again, these tactics become "public domain" once they are used, the only way to learn about their efficacy is by observing firms that try them, and there are no significant payoff interactions among adopters.

Our model also applies well to new forms of legal organization, which share the important characteristics of observational learning and free imitability. When the public corporation was new, its future dominance was anything but a foregone conclusion — its usefulness had to be demonstrated by experience. A more recent example is the exploding popularity of limited liability partnerships [Geyelin (1994)].[3]

Other applications include new strategies for hedging with derivatives, which has experienced a recent boom, and innovative investment strategies such as portfolio insurance, which experienced a boom in the mid-1980s and a bust following its failure to perform adequately during the 1987 crash. To some extent, one can analyze the value of these strategies by running pseudoexperiments with historical data, but there remains significant residual uncertainty that can be resolved only by executing the strategies in real time.[4]

Although the characteristics we analyze are sometimes important for new types of securities, the model is less suited to most of these innovations. Because liquidity is a critical concern in securities trading, the value of these innovations to a particular user often depends a great deal on how many others choose to use the same security.

While the model was motivated by thinking about financial innovations, it can also shed some light on boom and bust patterns in other

[3] One factor that limits application of the model to these examples is that payoff interactions among adopters might be important because of "network" externalities. For example, capital markets for corporate securities are very liquid because there are lots of corporations.

[4] The caveats that make these applications less than ideal are that it may be possible to keep the strategy's implementation details confidential, and it might be difficult to accurately observe a firm's use of derivatives, attenuating other firms' ability to learn from its experience. In addition, portfolio insurance might be another case where the innovation's value depends on the number of other adopters. Here the externality is negative because a portfolio insurance strategy is harder to execute successfully when many other investors are following the same strategy, as during the crash of 1987 [Brady Commission (1988), Jacklin, Kleidon and Pfleiderer (1992)].

areas. One particularly nice example is management strategies — total quality management, reengineering, downsizing, etc. These innovations are easily imitated, learning about their effectiveness comes from observing firms that try the strategies, and payoff interactions among adopters are minimal. Other examples where observational learning is important include oil exploration and real estate development. One firm's experience drilling in an area provides valuable information to other firms that hold leases in the area, and one firm's experience developing real estate of a particular type or in a particular area provides information about demand to other potential developers.

2. The Model

Consider an innovation that has just become available to a set of risk-neutral firms. The innovation is provided to firms at cost by financial intermediaries in a competitive market. The value of the innovation to a particular firm is the present value of the expected cash flows that it generates, and we let β denote the discount factor. Each firm chooses its adoption policy to maximize the present value of its future cash flows.[5] The innovation is better suited to some firms than others, so we assume that the innovation's expected value varies across firms. Let θ denote the value of the innovation to the "best" firm — the firm with the largest expected gains from adopting. For other firms, the innovation's expected value is lower than θ, and we refer to a firm that has expected payoff $\theta - x$ as "firm x." One can think of θ as the value of the innovation and think of x as the firm's cost of adopting the innovation, though the cross-sectional differences in value may result from many factors other than differences in adoption expenses. Adoption costs x are distributed on $[0, \infty)$, and this distribution is assumed uniform with unit density.[6]

The value of the innovation, θ, is uncertain. All information is public, and the (common) prior distribution of θ is normal with mean $\bar{\theta}_0$ and precision h_0 (i.e., variance $1/h_0$). The decision of firm x about whether to adopt immediately can be viewed as a call option exer-

[5] We assume that each firm acts as a value maximizer, but do not address how the gain from a firm's adoption might be shared. For instance, in a leveraged buyout, some of the gain accrues to the buyer of the firm and some to the firm's existing shareholders. We leave questions concerning the division of surplus to models tailored to specific applications, and simply note that the dynamics we analyze follow from any bargaining process that produces firm-surplus maximizing adoption decisions.

[6] Note that this density is not a *probability* density, but simply a convenient way of saying that, for any set S, the measure of firms with adoption costs $x \in S$ is the Lebesgue measure of S. We assume a density of 1 to simplify the exposition. Assuming some arbitrary measure of ω firms on any interval $[x, x+1)$ would not change the results in any substantive way, but would merely require a reinterpretation of the model's variables.

cise decision, where x is the exercise price and the payoff is stochastic with mean $\bar{\theta}_0$. Intuition suggests that if any firms adopt the innovation in a particular period, it will be the low-cost firms that do so, since they have the most to gain from adopting. This is confirmed in the first lemma.

Lemma 1. *Suppose firms x and $\hat{x}$ have not adopted prior to date t, and $x < \hat{x}$. If firm $\hat{x}$ adopts at date t, so does firm x.*

Proof. Suppose x does not adopt immediately, and let τ denote the (random) future adoption date and u denote the expected payoff, so $u = E[\beta^{\tilde{\tau}-t}(\tilde{\bar{\theta}}_\tau - x)] \geq \bar{\theta}_t - x$. One feasible strategy for $\hat{x}$ is to do what is optimal for x. This would yield expected payoff

$$\begin{aligned} E[\beta^{\tilde{\tau}-t}(\tilde{\bar{\theta}}_\tau - \hat{x})] &= u - (\hat{x} - x)E[\beta^{\tilde{\tau}-t}] \\ &> u - (\hat{x} - x) \geq \bar{\theta}_t - x - (\hat{x} - x) = \bar{\theta}_t - \hat{x}, \end{aligned}$$

producing the contradiction that immediate adoption is suboptimal for $\hat{x}$. ■

Given Lemma 1, at any given date t, all the firms on some interval $0 \leq x < x_t$ have already adopted the innovation, and firm x_t is the low-cost remaining firm. (By the definition of θ, $x_0 = 0$.) If the innovation's current estimated value is low enough ($\bar{\theta}_t \leq x_t$), none of the remaining firms adopt, because their expected profit is negative — the option is "out of the money." Since no additional firms adopt, no new information is generated, and the situation next period is unchanged — we are in the "stopping region" and the system has halted permanently.

If $\bar{\theta}_t$ is greater than x_t, on the other hand, some new firms $x \in [x_t, x_{t+1})$ will exercise their option to adopt the innovation, and we let $q_t = x_{t+1} - x_t$ denote the number of new adopters.[7] Between t and $t+1$, these firms realize their gains or, a bit more realistically, the uncertainty about their future cash flows from adopting is resolved. The average realized profit for the adopting firms is denoted

$$\pi_t = q_t^{-1} \int_{x_t}^{x_{t+1}} (\theta - x)\ dx + \gamma_t,$$

where γ_t is a normally distributed mean-zero shock. We assume that

[7] It must be the case that some firms adopt, for suppose the optimal decision was that no additional firms adopted at date t. Then no new information is generated, so the situation is identical at $t + 1$, and no adoption must again be optimal. By induction, no adoption occurs at any future date, so all the remaining firms get a payoff of zero. For any firm $x \in [x_t, \bar{\theta}_t)$, this is strictly less than the payoff from immediate adoption, so firm x should adopt — a contradiction.

the amount of new information generated is proportional to the number of firms adopting the innovation — the precision of γ_t is $q_t h$, where h is a parameter that measures the rate of learning about θ from the experience of adopting firms. This assumption ensures that the information generated is the same as in a world where q_t discrete firms produce signals with identical precision. To avoid integer complications, we assume a continuum of firms rather than discrete firms.[8]

Let ϵ_t denote the difference between the realized profit and the expected profit:

$$\epsilon_t = \pi_t - q_t^{-1} \int_{x_t}^{x_{t+1}} (\bar{\theta}_t - x)\, dx = \theta - \bar{\theta}_t + \gamma_t. \tag{1}$$

Firms update their opinions of the innovation in standard Bayesian fashion, so the posterior distribution of θ is normal, with mean $\bar{\theta}_{t+1} = \bar{\theta}_t + \frac{q_t h}{h_t + q_t h}\epsilon_t$ and precision $h_{t+1} = h_t + q_t h$. The process then repeats at date $t+1$. If the realized profit π_t is low enough, $\bar{\theta}_{t+1}$ drops below x_{t+1}, so no more firms adopt at $t+1$. When this happens, the system halts, because no new information is generated when no firms adopt. But if π_t is high enough that $\bar{\theta}_{t+1} > x_{t+1}$, additional firms adopt at date $t+1$, again generating new information that the remaining firms use to revise their beliefs about θ.

3. Privately Optimal Adoption Decisions

Firm x chooses an adoption policy that maximizes its expected discounted value, taking the actions of other firms as given. First consider the initial decision. If $\bar{\theta}_0 \leq 0$, no firm has positive expected profits from adopting the innovation, so no firms adopt and the system never starts. If $\bar{\theta}_0$ is positive, then all firms $x \in [0, \bar{\theta}_0)$ have positive expected payoffs from immediate adoption. Some of these firms must innovate — otherwise, the system would stop and they would all get zero payoffs. This would be irrational since immediate adoption yields a positive expected payoff. However, given that some of these firms adopt today, each firm also has a positive expected payoff from waiting a period to see what it learns about θ. Because of this, not all of them will adopt immediately — x_1 is less than $\bar{\theta}_0$. Although it is easy to determine the expected payoff for firm x from immediate adoption (just $\bar{\theta}_0 - x$), the expected payoff from waiting is harder

[8] The Appendix explains in more detail the connection between this model and a model that has discrete firms with equally spaced adoption costs. When firms are not atomistic, there is also the possibility of multiple equilibria due to strategic considerations.

to determine because information revelation is endogenous. The expected payoff from waiting depends on how many firms adopt today, since this determines how much new information is generated, and it also depends on the adoption policies that firms will follow in the future. However, the problem simplifies if one considers the situation of the new marginal firm, x_1.

Consider what happens when date 1 rolls around and $\bar{\theta}_1$ is realized. If $\bar{\theta}_1 \leq x_1$, the system stops, and all remaining firms get a payoff of zero. But if $\bar{\theta}_1$ is larger than x_1, some firms will innovate, and firm x_1, being the lowest-cost firm, will certainly innovate. The expected payoff from waiting is therefore very simple for the marginal firm; it is just

$$\beta E[(\tilde{\bar{\theta}}_1 - x_1)^+ \mid x_1] = \beta \int_{x_1}^{\infty} (\bar{\theta}_1 - x_1) f(\bar{\theta}_1 \mid x_1)\ d\bar{\theta}_1, \tag{2}$$

where f is the density of $\bar{\theta}_1$. By waiting, the firm has an *option* to adopt next period, and the option turns out to be valuable if it is "in the money." Since the marginal firm must be indifferent between immediate adoption and waiting, this expression must equal the expected payoff from adopting, $\bar{\theta}_0 - x_1$. Equating these expressions determines x_1 and, hence, the volume of adoption at date 0. Notice that in Equation (2), the density $f(\bar{\theta}_1)$ is conditioned on x_1; determining the marginal firm is not as simple as it might appear because the distribution of $\tilde{\bar{\theta}}_1$ changes as the volume of adoption changes, since more adoption generates more information, making $\tilde{\bar{\theta}}_1$ more variable.

This logic is equally valid at any other date t. Suppose the current state of the system is $(\bar{\theta}_t, x_t, h_t)$, and $\bar{\theta}_t$ is greater than x_t, so the system has not stopped. The new marginal adopter, x_{t+1}, has an expected payoff from immediate adoption of $\bar{\theta}_t - x_{t+1}$. If this firm waits, it will adopt at date $t+1$ if and only if $\bar{\theta}_{t+1} > x_{t+1}$, in which case its expected payoff is $\bar{\theta}_{t+1} - x_{t+1}$. The marginal adopter is determined by equating these payoffs:

$$\begin{aligned} \bar{\theta}_t - x_{t+1} &= \beta E[(\tilde{\bar{\theta}}_{t+1} - x_{t+1})^+ \mid x_{t+1}] \\ &= \beta \int_{x_{t+1}}^{\infty} (\bar{\theta}_{t+1} - x_{t+1}) f(\bar{\theta}_{t+1} \mid x_{t+1})\ d\bar{\theta}_{t+1}. \end{aligned} \tag{3}$$

(The function f is the density of $\bar{\theta}_{t+1}$ (at date t), which is normal with mean $\bar{\theta}_t$ and variance $\frac{q_t h}{h_t(h_t+q_t h)}$.) Equation (3) says that, for the marginal adopter, the option to adopt is worth exactly the same amount dead or alive. One can show that x_{t+1} is well defined since there is a unique firm that satisfies Equation (3).

Equation (3) can be written in a way that nicely illustrates the trade-off firms face, by rewriting the left-hand side as $\int_{-\infty}^{\infty}(\bar{\theta}_{t+1} - x_{t+1}) f(\bar{\theta}_{t+1} \mid x_{t+1})\, d\bar{\theta}_{t+1}$ and rearranging to get

$$\int_{-\infty}^{x_{t+1}} (x_{t+1} - \bar{\theta}_{t+1}) f(\bar{\theta}_{t+1} \mid x_{t+1})\ d\bar{\theta}_{t+1} = (1-\beta) \int_{x_{t+1}}^{\infty} (\bar{\theta}_{t+1} - x_{t+1}) f(\bar{\theta}_{t+1} \mid x_{t+1})\ d\bar{\theta}_{t+1}, \tag{4}$$

or

$$E[(\tilde{\bar{\theta}}_{t+1} - x_{t+1})^{-} \mid x_{t+1}] = (1-\beta) E[(\tilde{\bar{\theta}}_{t+1} - x_{t+1})^{+} \mid x_{t+1}].$$

The left-hand side of Equation (4) is the expected benefit for the marginal firm of avoiding date-t adoptions when the subsequent information indicates that an adoption is probably unprofitable ($\bar{\theta}_{t+1} < x_{t+1}$). The right-hand side is the expected loss in time value from delaying an innovation that the firm ends up adopting at date $t+1$ anyway, because the innovation still looks profitable ($\bar{\theta}_{t+1} > x_{t+1}$). When the subsequent information turns out to be (relatively) good, the marginal firm regrets having waited, since waiting delayed the cash flows from the innovation. The decision to wait pays off when the subsequent information is bad, because the firm avoids an innovation that looks unprofitable ex post — Bernanke (1983) calls this the "bad news principle."

The adoption volume at date t is determined by the solution to Equation (3) or Equation (4). Figure 1 illustrates what happens for a specific set of parameters ($h_0 = 0.1$, $h = 0.2$, and $\beta = 0.95$) by plotting adoption volume q_t as a function of $\bar{\theta}_t$ and x_t.[9] The system begins on the left side of the graph at the point $(\bar{\theta}_0, x_0 = 0)$ and moves from left to right over time. The height of the surface tells us the speed at which the system moves. Movement continues until a shock drives the system into the stopping region where $\bar{\theta}_t \leq x_t$. Since expected profits $\bar{\theta}_t - x$ are now negative for the remaining firms, they reject the innovation, and $q_t = 0$. Because there are no new adopters, information revelation stops, and $\bar{\theta}$ stops evolving. Throughout the article, Figure 1 will be helpful for developing intuition.

The following comparative statics results are intuitively clear, and easy to prove using Equation (3) or Equation (4): for $\bar{\theta}_t > x_t$, q_t is increasing in $\bar{\theta}_t - x_t$ and h_t, and decreasing in h and β. The expected profit for the existing marginal firm is $\bar{\theta}_t - x_t$, and more firms adopt

[9] The plot omits h_t as a state variable because, with h_0 and h fixed, h_t and x_t are linked by $h_t = h_0 + x_t h$. Either one of the two can therefore be omitted as redundant.

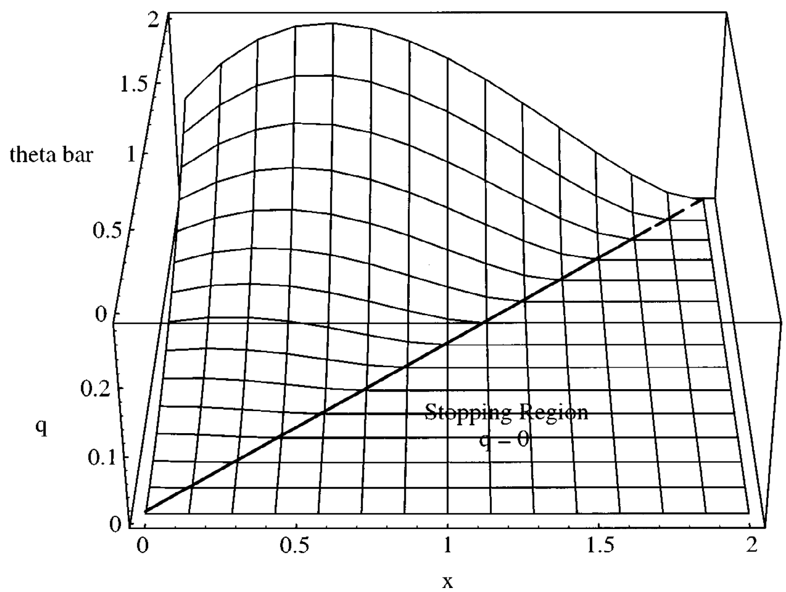

Figure 1
Adoption volume versus adoption cost and estimated value
The number of firms that adopt the innovation at any date (q) is plotted on the vertical axis as a function of the marginal firm, x, and the estimated value of the innovation, $\bar{\theta}$. The system moves from left to right over time as firms adopt. The region in the lower right-hand corner of the plot is the stopping region where no firms adopt and information revelation ceases. The plot is constructed using the following parameters: prior precision $h_0 = 0.1$, rate of learning $h = 0.2$, and discount factor $\beta = 0.95$.

when this is higher. More firms also adopt when h_t is higher because the expected benefits of waiting for more information are smaller when current information is more precise. Using the same logic, fewer firms adopt when h is larger because more information is generated when the rate of learning is higher, so the expected benefits of waiting are greater. Fewer firms adopt when β is higher because it is less costly to wait for more information.

Proposition 1. *When $\bar{\theta}_t \leq x_t$, $q_t = 0$. When $\bar{\theta}_t > x_t$, there is a unique value x_{t+1} that satisfies Equation (3), and $0 < q_t < \bar{\theta}_t - x_t$. As long as $\bar{\theta}_t > x_t$, q_t is increasing in $\bar{\theta}_t$ and h_t, and is decreasing in h and β.*

The results of Proposition 1 are much more general than our specific distributional assumptions. The assumption that adoption costs are distributed uniformly is not used at all — the proposition holds for any continuous density of adoption costs, or even if all firms have identical adoption costs. The Gaussian randomness that we have assumed is not crucial either. What is important for the results is that an increase in q_t or h or a decrease in h_t makes the distribution of $\tilde{\bar{\theta}}_{t+1}$ riskier in the sense of Rothschild and Stiglitz (1970). An example with completely different distributions [based on Raiffa and Schlaifer (1961, Section 10.2)] is available from the authors. The analysis shows explicitly that the results go through under these different distributional assumptions.

4. Rational Innovation versus Irrational Fads

In this section we further characterize the behavior of the system. Under the assumption that participants act rationally, we derive results that are sometimes mistakenly cited as evidence that an innovation wave is an irrational fad.

4.1 Firms adopt following bad news

Our first result shows that it is often rational for firms to adopt in one period even though their most recent information is negative and they chose not to adopt in the previous period. At first blush, it appears that these firms are acting like lemmings — they are adopting just to be part of the crowd, ignoring the warning signs of recent history. Critics of the innovation could interpret this as evidence of an irrational fad, but in the next proposition, we show that this reasoning is faulty — adopting in the face of recent bad news is not evidence of irrational excess.

Proposition 2. *Firms sometimes adopt at date t even though* $\bar{\theta}_t < \bar{\theta}_{t-1}$ *and they chose not to adopt at date* $t-1$.

Additional firms adopt at time t whenever $\bar{\theta}_t > x_t$, because some remaining firms have positive expected profits. Recall from Equation (4) that x_t is less than $\bar{\theta}_{t-1}$, that is, some firms with positive expected profits at $t-1$ choose to wait for more information. Therefore, moderately bad news — news that reduces $\bar{\theta}$, but not below x_t — will reduce the rate of adoption, but will not kill the wave.

In fact, efficiency is often enhanced by behavior that seems even more extreme. We shall see later that a social planner not only encourages adoption after moderately unfavorable information, but sometimes encourages adoption by firms that have *negative* expected profits from adoption. It is even possible that a social planner will encourage adoption when *all* of the adopting firms have negative expected profits from adoption, because the social value of the information generated exceeds the expected losses.

4.2 Innovation waves always end badly

We now look at the fate of the last firms to adopt the innovation. Define T to be the (random) period in which the system stops. At date T, the system moves into the stopping region where $\bar{\theta}_T \leq x_T$. Equation (3) implies that $x_T < \bar{\theta}_{T-1}$. Combining these two inequalities, we get $\bar{\theta}_T < \bar{\theta}_{T-1}$. The innovation wave, therefore, always ends on a bad note; the realization of $\bar{\theta}_T$ is always a disappointment because it is less than expected.[10] Furthermore, average profits are *negative* for the firms that were least eager to adopt (firms with $x = x_T$). This is so despite the fact that these firms expected positive profits when they decided to adopt.

Proposition 3. *Firms that adopt in the last period are disappointed on average: $\bar{\theta}_T < \bar{\theta}_{T-1}$. Furthermore, average profits are negative for the marginal adopters.*

Because an innovation wave always ends in disappointment, critics of the wave point to this disappointment as proof that the wave grew into an irrational fad. For instance, in popular discussions of LBOs, critics often point to the lamentable fate of late adopters as proof that the wave had gotten out of control and investors were overly optimistic about the profits to be made. When adoptions are profitable, they lead to further adoption, so adoption stops only when recent information is negative. Since adoption volume falls to zero at the end of the wave, the bad taste left by the losses of the last adopters remains the last word on the subject. A disappointing ending is inevitable even though all adopters rationally expect positive profits when they adopt.

4.3 The end of an innovation wave is unpredictable

One feature of irrational fads is that it is hard to predict when the "bubble will burst," but it would seem that one could forecast the end of a wave of rational innovation with some success. To the contrary, we find that it is impossible to predict when an innovation wave will end based on any information available before the system stops. At any date before the system stops (i.e., when $\bar{\theta}_t > x_t$), the "hazard rate" is the probability that the system will stop the next period (i.e., $\Pr[\tilde{\bar{\theta}}_{t+1} \leq x_{t+1}]$). The next proposition states that the hazard rate never changes, regardless of how the system evolves — the hazard rate depends only on β. For β in the range 0.80 to 0.99, the hazard rate ranges from 26% ($\beta = 0.80$) to 4% ($\beta = 0.99$). Taking LBOs as an

[10] The fact that $\bar{\theta}_T$ is always less than $\bar{\theta}_{T-1}$ does not contradict the martingale property of conditional expectations. Suppose date t turns out to be the stopping date. At date $t-1$, we do not yet know that next period is the stopping date, and $E[\tilde{\bar{\theta}}_t] = \bar{\theta}_{t-1}$, as it must.

example, a period in the model would be about 2 years because it appears to take about that long to determine the valuation effects of the restructuring. If the interest rate is 6%, the discount factor is then 0.8836, and the probability that firms stop doing LBOs in a given period is about 20%.

Proposition 4. *Assume that $\bar{\theta}_t > x_t$, so the system has not stopped. The hazard rate is independent of the date t, the rate of learning h, the current state of the system $(\bar{\theta}_t, x_t, h_t)$, and the history of the system $(\{\bar{\theta}_\tau\}_{\tau=0}^{t-1}, \{x_\tau\}_{\tau=0}^{t-1}, \{h_\tau\}_{\tau=0}^{t-1})$. Hence, the hazard rate is also independent of $\{q_\tau\}_{\tau=0}^{t-1}$, $\{\gamma_\tau\}_{\tau=0}^{t-1}$, and $\{\pi_\tau\}_{\tau=0}^{t-1}$.*

The end of a wave is, therefore, sudden and unpredictable, so waves appear to be fragile. Innovation waves naturally proceed in a boom and bust pattern. Before the system stops, the shocks to the system are positive, on average, so estimates of the innovation's value become increasingly optimistic — a boom occurs. Then a large negative shock hits the system "without warning" and adoption stops completely. As noted above, the last adopters are disappointed with their profits and regret their adoption ex post, so the wave ends in a sudden bust. Boom and bust patterns are often considered socially undesirable and signs of excess, greed, irrationality, or speculative mania. (Simply applying the pejorative labels boom and bust suggests irrationality and social impropriety.) Many seem to feel that innovation must be orderly and smooth to be rational. However, this result shows that unpredictability and fragility result from endogenous information revelation and are consistent with rationality — *predictable* stopping would require irrationality.

One might think that the hazard rate would be higher when $\bar{\theta}_t - x_t$ is small, because we are then close to the stopping region, and it would seem that the chances of moving into the stopping region are higher when we are close to the boundary. Figure 1 helps us see why this is not so. Holding x_t constant and decreasing $\bar{\theta}_t$ moves us closer to the stopping region but, as the figure shows, it also decreases the volume of adoption. Since less new information is generated, the variance of $\tilde{\bar{\theta}}_{t+1}$ decreases, and this tends to reduce the probability of stopping next period. The effects of moving closer to the stopping region and decreasing the variance of $\tilde{\bar{\theta}}_{t+1}$ exactly offset, leaving the hazard rate unchanged. If, instead, we hold $\bar{\theta}_t$ fixed and increase x_t (move left to right in the figure), we again move closer to the stopping region, and Figure 1 shows that the effect on the volume of adoption is ambiguous. However, an increase in x_t means that more firms have adopted previously, so more information has been incorporated into the current estimate $\bar{\theta}_t$. This makes $\tilde{\bar{\theta}}_{t+1}$ less volatile because less weight is put on the new information and more on the current estimate, $\bar{\theta}_t$. This decrease in volatility, combined with the change in adoption volume, exactly offsets the effect of being closer to the stopping region.

4.4 Innovations are usually disappointments

Because the amount of new information generated depends on whether previous information was favorable or unfavorable, a bit of luck (good or bad) in the early tests of an innovation can dramatically alter the adoption process and the terminal opinion of the innovation's value. The endogenous information flow produces a fundamental asymmetry between innovations that experience initial good luck and those that experience initial bad luck.

Suppose that, by chance, the initial adopters of an innovation experience low profits, leading all firms to reduce their estimate of its value. If these outcomes are bad enough, no firms adopt in the following period, and the flow of information stops. When poor profits push an innovation into the stopping region where $\bar{\theta}_t < x_t$, it is "stuck" there forever. Sometimes the bad outcomes occur because the innovation is truly inferior, but sometimes they result from randomness; in the latter case, the mistaken pessimism is terminal because no new information arrives to correct it. A corresponding amount of *good* initial luck has the opposite dynamics. Rather than permanently locking in opinions about the innovation, the high profits induce a lot of subsequent adoption, which generates new information about θ. When the early profits were due to randomness rather than a truly superior innovation, the subsequent information will usually correct the mistaken optimism.

Because innovations that wander into the stopping region due to bad luck cannot escape, a disproportionate number of innovations end up permanently labeled bad — or at least disappointing — ideas, with a low terminal estimate of θ (denoted $\bar{\theta}_T$). This produces a skewed distribution of $\tilde{\bar{\theta}}_T$, with the median and mode less than the prior mean, $\bar{\theta}_0$. This skewness also means that the right tail of the distribution of $\tilde{\bar{\theta}}_T$ is fatter than the left tail, since the average value of $\bar{\theta}_T$ must equal the prior mean, $\bar{\theta}_0$. One can interpret the skewed distribution of $\tilde{\bar{\theta}}_T$ as follows: although most innovations turn out worse than expected ($\bar{\theta}_T < \bar{\theta}_0$), the positive surprises ($\bar{\theta}_T > \bar{\theta}_0$) tend to be bigger surprises than the negative surprises. If information about the innovation were independent of the rate of adoption, there would be no feedback loop, and the distribution of terminal estimates would be symmetric rather than skewed.

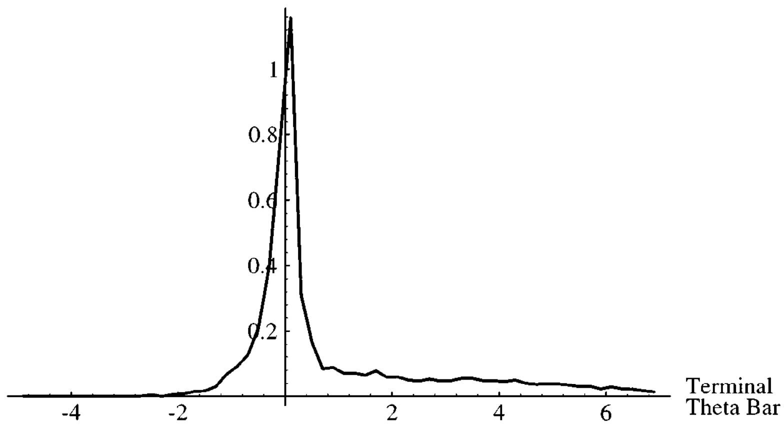

Figure 2
Distribution of terminal estimated value, $\bar{\theta}_T$
This figure plots the distribution of $\bar{\theta}_T$, the terminal estimate of the value of the innovation, for an example with the following parameters: prior precision $h_0 = 0.1$, rate of learning $h = 0.2$, and discount factor $\beta = 0.95$. The distribution resulted from a simulation with 5000 trials.

We illustrate this via simulation by setting $\bar{\theta}_0 = 1$, $h_0 = 0.1$, $h = 0.2$, and $\beta = 0.95$, and running 5000 simulations of the model. Figure 2 plots the frequency distribution of the values of $\bar{\theta}_T$, and clearly shows the skewness of the distribution. Even though the prior (and average posterior) mean is 1, the median $\bar{\theta}_T$ is only 0.081 and the mode is about 0.05. More than 71% of the simulated innovations have terminal estimates less than 1. The right tail of the distribution is noticeably thicker than the left tail.

Thus, because of the link between current profitability and future information flow, most innovations end up as disappointments. Failure to recognize these dynamics could lead a regulator or other observer to conclude that adopters tend to be overly optimistic or euphoric about an innovation, perhaps because of ambitious or fraudulent promoters. This conclusion, however, ignores the other side of the coin, that the successful innovations succeed more spectacularly — again, not because they have been promoted too aggressively, but because of the natural feedback that occurs when firms learn by observing the experiences of others. Casual empiricism can be deceiving in an innovation wave if one overlooks these inherently asymmetric dynamics.

4.5 Generality of the results

The results in this section hold under more general conditions than we have assumed. For example, Propositions 2 and 3 hold for any continuous density of adoption costs and any continuously distributed random variables. It is also clear that the asymmetric dynamics would cause most innovations to be disappointments for any symmetric distribution, not just the normal distribution we assumed. The result that the end of an innovation wave is completely unpredictable (Proposition 4) also holds for any continuous density of adoption costs. However, Proposition 4 will not hold for arbitrary distributions of the random variables. The probability of stopping hinges on the distribution of next period's mean, $\tilde{\bar{\theta}}_{t+1}$, which is normal under our assumptions. The property of the normal distribution that yields the unpredictability result is that the density $f(\bar{\theta}_{t+1})$ can be written as $(1/\sigma)\phi((\bar{\theta}_{t+1} - \bar{\theta}_t)/\sigma)$, where ϕ is a function independent of the parameters of the distribution — specifically, ϕ is the standard normal density. Thus the density at a particular $\bar{\theta}_{t+1}$ is determined if one knows the standard deviation and knows the distance from $\bar{\theta}_{t+1}$ to the mean, measured in standard deviations. This property is not unique to the normal distribution — it is shared by the uniform, exponential, double exponential, logistic, and Gumbel (or extreme value) distributions.

5. The Diffusion Curve

In this section we characterize the volume of adoptions over time. We frame the discussion in terms of the "diffusion curve," which is a plot of the cumulative number of adoptions as a function of time. The empirical literature has established the stylized fact that diffusion curves for technological innovations tend to be S-shaped.[11] The rate of adoption is low to begin with, it increases to a maximum (the inflection point of the S curve), and it decreases thereafter. In some cases, the innovation is adopted rapidly from the outset and the rate of adoption decreases, leading to a concave diffusion curve. We now show that the model generates diffusion curves consistent with the stylized findings.

The diffusion curve will vary depending on the profits that adopters realize, so we would like to characterize the "average" diffusion curve. We begin by showing what the diffusion curve looks like when there are no surprises, that is, when it turns out that $\bar{\theta}_t$ is constant because unexpected profits ϵ_t happen to be zero each period.

When there are no surprises, the state variable $\bar{\theta}_t$ is constant over time. Only the variables x_t and h_t change, and they are linked via $h_t = h_0 + x_t h$. We therefore investigate how the rate of adoption q_t varies with the cumulative volume of adoption x_t. In terms of Figure 1,

[11] For example, see Mansfield (1961, 1968) and the studies in Nasbeth and Ray (1974).

this corresponds to moving from left to right parallel to the x axis. For the parameters used in the figure, the rate of adoption starts low, increases to a maximum, and then decreases, yielding an S-shaped pattern of cumulative adoptions. The proposition below states that this is always true when the initial information is imprecise (h_0 is small), the rate of learning h is large, or the discount factor β is high. All these factors increase the benefits of waiting vis-à-vis immediate adoption, so adoption volume starts low and builds momentum. In the opposite circumstances, firms adopt aggressively from the outset, and the diffusion curve is concave.

Proposition 5. *Suppose that unexpected profits ϵ_t turn out to be zero each period. There exist critical values $h_0^* \in (0, \infty)$, $h^* \in (0, \infty)$, and $\beta^* \in (0, 1]$ such that if $h_0 < h_0^*$, $h > h^*$, or $\beta > \beta^*$, the diffusion curve is S-shaped ($\partial q/\partial x > 0$ for x less than some $\hat{x}$ and $\partial q/\partial x < 0$ for $\hat{x} < x < \bar{\theta}_0$). Otherwise the diffusion curve is concave ($\partial q/\partial x < 0$ for all x).*

We have also verified, via simulation, that the average diffusion curve behaves as expected when there *are* unexpected profits. Higher values of h and β and lower values of h_0 make firms less aggressive in their adoption, so the average diffusion curve is more S-shaped. Opposite changes in the parameters make firms adopt more aggressively, so the average diffusion curve is more concave.

6. Cooperative Adoption Decisions and the Role of Financial Intermediaries

6.1 Discussion

This section explores the positive informational externality that arises because one firm's decision to adopt the innovation generates information that is valuable to other firms. Since individual firms ignore the externality when deciding whether to adopt, they often choose to wait even though immediate adoption would increase expected aggregate wealth. From a social perspective, there is too little adoption; adoption volume would be higher if firms could act cooperatively. The message for policymakers is that adoption should not be restricted but, if anything, encouraged.[12]

Even if regulators do not adequately encourage adoption of risky innovations, financial intermediaries with some market power have a natural incentive to encourage adoption. A dominant intermediary —

[12] We should add a caveat. We are assuming that θ captures the innovation's effect on social wealth. The policy implications would be different if an adopter's gains came at others' expense.

such as Drexel Burnham Lambert in junk bonds during much of the 1980s or Salomon Brothers when mortgage-backed securities were a recent innovation — can capture large rents from future adopters, providing a reason to subsidize early adopters. Such an intermediary internalizes the informational externality and effectively improves social welfare.[13] Market power is necessary, though, for intermediaries to play this role. If the market for the innovation is perfectly competitive, intermediaries earn no rents and have no incentive to spur adoption.

To see this, take the extreme case where the innovating intermediary has a monopoly on the use of the innovation, and suppose the innovator can make a credible take-it-or-leave-it offer to any firm at any time. Then the intermediary could decide which firms should adopt in the current period and could capture all the adoption profits by making the following offer to each of them: "adopt the innovation now and turn over all profits to me in exchange for a cash payment of one cent." In this case, the intermediary's maximization problem is identical to that of the social planner, that is, maximize the present value of all profits from adoption. In reality, intermediaries have less monopoly power and less bargaining power, but they do capture some of the profits from subsequent adoption. Consequently they have the incentive to encourage adoption by subsidizing marginal firms, thereby partially achieving the gains that a social planner would achieve.

6.2 Results

We now denote the privately chosen adoption volume q_t^p and the cooperative decision q_t^c. The determination of q_t^p was analyzed in Section 3; the determination of q_t^c is, unfortunately, less straightforward. We can show that the cooperative adoption volume is always at least as great as the private level ($q_t^c \geq q_t^p$), and is generally strictly greater. The task of a social planner is to choose an adoption policy that, at any date, maximizes the expected aggregate payoff from all current and future adoption decisions. The problem has three state variables: the estimated value of the innovation, $\bar{\theta}_t$; the existing marginal firm, x_t; and current precision, h_t.[14] The transition rules for the state variables are straightforward. If q_t firms adopt the innovation,

[13] Tufano (1989) documents that firms that introduce financial innovations earn their profits primarily from the large volume of subsequent business they capture.

[14] In this presentation, which seems to be the most straightforward, the problem has three state variables ($\bar{\theta}_t$, x_t, and h_t) and two parameters (h and β). As mentioned in note 9, one can also view the problem as having two state variables ($\bar{\theta}_t$ and x_t) and three parameters (h_0, h, and β). In fact, the problem can be simplified to have only two state variables ($\bar{\theta}_t - x_t$ and h_t) and two parameters (h and β) — given h and β, all that really matters about the current state is the marginal expected profitability of adoption, and the precision of this estimate.

the marginal adopter changes to $x_{t+1} = x_t + q_t$, and the precision changes to $h_{t+1} = h_t + q_t h$. The state variable $\bar{\theta}_t$ is stochastic, and the date-t distribution of $\bar{\theta}_{t+1}$ is normal with mean $\bar{\theta}_t$ and variance $\frac{q_t h}{h_t(h_t+q_t h)}$.

The social planner chooses a policy function that prescribes the number of new adopters, q_t, as a function of the current state. We can equivalently think about choosing the new marginal adopter $x_{t+1} = x(\bar{\theta}_t, x_t, h_t)$, since $x_{t+1} = x_t + q_t$. The goal is to solve

$$V(\bar{\theta}_0, x_0 = 0, h_0) = \sup_{x(\cdot)} E\left[\int_0^{x_1} (\widetilde{\theta} - x)\, dx + \sum_{t=1}^{\infty} \beta^t \int_{x(\tilde{\bar{\theta}}_{t-1}, \tilde{x}_{t-1}, \tilde{h}_{t-1})}^{x(\tilde{\bar{\theta}}_t, \tilde{x}_t, \tilde{h}_t)} (\widetilde{\theta} - x)\, dx\right],$$

subject to the transition rules given in the previous paragraph. The Bellman principle requires that, for any state, the policy function solve

$$V(\bar{\theta}_t, x_t, h_t) = \sup_{x_{t+1} \geq x_t} \left\{\int_{x_t}^{x_{t+1}} (\bar{\theta}_t - x)\, dx + \beta E[V(\widetilde{\bar{\theta}}_{t+1}, x_{t+1}, h_{t+1})]\right\}. \tag{5}$$

It is important to recognize that the distribution of $\bar{\theta}_{t+1}$ in the last term depends on the choice x_{t+1} — more adoption generates more information.[15]

By examining the first-order conditions of this maximization problem, one can show that, in any state, the social planner will choose at least as much adoption as firms will choose privately, and will often choose strictly more. The reason is straightforward. In addition to trading off the marginal firm's expected profits from immediate adoption against its expected profits from waiting (as independent firms would do), additional adoption generates information that is valuable in making future adoption decisions for the remaining firms. If the flow of information were independent of the volume of adoption, the social planner would set $q_t^c \equiv q_t^p$.

Proposition 6. *For any state* $(\bar{\theta}_t, x_t, h_t)$*,* $q_t^c \geq q_t^p$*, and the inequality is sometimes strict.*

Clearly the externality is more important when there is more uncertainty about the innovation, since additional information is more

[15] The social planner's problem bears some similarities to the problem analyzed by Tsur et al. (1990), in which a farmer chooses how many acres to convert to a new technology each period. Some of the important differences are that their farmer is risk averse (with constant absolute risk aversion), acreage can be costlessly switched back to the old technology, and the expected payoff from adopting the new technology is the same for each acre.

valuable when there is more uncertainty. The next proposition shows that the private and cooperative decisions converge as the precision of the current estimate increases, so the social benefits of encouraging adoption are small when there is little uncertainty.

Proposition 7. *As h_t approaches ∞, $q_t^c - q_t^p$ approaches 0.*

When a lot of firms have adopted previously, most of the uncertainty has been resolved (h_t is large), so intervention does not accomplish much. However, when h_t is small, intervention can produce large welfare improvements, so efforts to encourage adoption should concentrate on unproven innovations — innovations that are new or that died an early death.

The most apparent case is when $\bar{\theta}_t$ is less than x_t, so each firm expects to incur losses if it adopts. The system halts, but this may be grossly inefficient from a social perspective. Even though current adoptions are expected to generate losses, the value of the information to the remaining firms can far exceed the expected losses of current adopters (especially when h_t is small). In fact, the next proposition shows that for *any* $\bar{\theta}_t$ and x_t, it is optimal to choose positive adoptions if h_t is small enough. Similarly, for any h_t, it is optimal to choose positive adoptions even though $\bar{\theta}_t - x_t$ is negative, as long as $\bar{\theta}_t$ is close enough to x_t. Finally, for any state $(\bar{\theta}_t, x_t, h_t)$, positive adoptions are optimal if the rate of learning h is large enough.

Proposition 8. *If $\bar{\theta}_t \geq x_t$, then $q_t^c > 0$. If $\bar{\theta}_t < x_t$, then $q_t^c > 0$ if h_t is small enough, if h is large enough, or if $x_t - \bar{\theta}_t$ is small enough.*

7. Related Literature

This article extends the literature that studies the effect of informational externalities that spring from observational learning. Banerjee (1992), Bikhchandani, Hirshleifer, and Welch (1992), and Welch (1992) advance models along this line. They show that when agents cannot observe others' information but only their decisions, rational agents may choose to ignore their own private signals and follow the actions of previous decision makers. Such cascades would not develop if all agents shared the same information, which highlights the fundamental difference between these models and ours. In our model, all information is *public* and is incorporated in individual decisions. Nevertheless, rational decisions lead to boom and bust adoption patterns that may appear irrational. Another interesting difference is that those models behave symmetrically, with positive and negative information producing the same dynamics (but in opposite directions). In our model, the dynamics induced by positive and negative information are very different because of the endogenous information flow.[16]

In a similar vein, Bulow and Klemperer (1994) allow endogenous timing of transactions and show that frenzies of activity may have rational origins.[17] While their model differs fundamentally from ours, their concern with the volume of trading activity is similar to our concern with the volume of adoption in an innovation wave.

A number of articles study adoption of innovations where there is uncertainty about the value of the innovation and firms learn about its value through experience. Jensen (1982), Stoneman (1981), and Tsur, Sternberg, and Hochman (1990) are examples. The important difference between those models and ours is that, in those models, the adopting firm captures the future benefits of the information generated by adoption. By contrast, adoption in our model produces an informational externality because *other* firms benefit from the information generated when a firm chooses to adopt.[18]

This article is also related to the literature that studies decision making in the face of irreversible investment opportunities which was pioneered by Arrow and Fisher (1974) and is reviewed by Pindyck (1991). The basic trade-off faced by a firm in our model when deciding whether to adopt an innovation is the same as in those models. Delaying the adoption of a good innovation is costly but, because adoption has uncertain consequences and more information becomes available over time, a firm can sometimes avoid a mistake if it waits for more information.[19]

As in a subset of the contagion literature, we develop an information-based model where the actions of a large group of agents are correlated. Chari and Jagannathan (1988) show how an initial wave

---

[16] Caplin and Leahy (1993) develop a very different model that is based on observational learning by potential entrants about production costs in an industry, and the amount of information revealed is endogenously determined by the number of entrants. They derive results similar to some of ours, and use their model to explore the macroeconomic effects of sectoral shocks and the speed with which the economy adjusts to those shocks.

[17] Gul and Lundholm (1995) present another model with endogenous timing in which clustering occurs.

[18] The innovation model of Bhattacharya, Chatterjee and Samuelson (1986) incorporates informational externalities, but differs from ours in several important ways. First, firms in their model run experiments to gather information; such experiments are impossible in our model. Second, the results of a firm's experiments are not public information — in our model, all information is public. Third, they rely on asymptotic decision rules that are approximately correct when the uncertainty about the innovation is small. One must therefore treat their results with caution when there is substantial uncertainty about the innovation — precisely the situation that interests us.

[19] It is not crucial to our results that the decision to adopt the innovation is irreversible, only that it involves some irreversible consequences. For example, some unrecoverable costs are incurred in executing an LBO, and if the firm later decides to reverse the LBO, the losses suffered in the interim cannot be recovered.

of bank withdrawals can lead to a panic when other depositors make incorrect inferences about the reason for the withdrawals, and Gorton (1985) models banking panics caused by depositors acting on imperfect signals about the return on their deposits. In these models, some participants have information that others do not, whereas all information is public in our model. Hence our model shows that contagion-like dynamics can be generated even with symmetric information and it shows that these dynamics are not necessarily undesirable.

The network externality literature pioneered by Farrell and Saloner (1985, 1986) and Katz and Shapiro (1986) concentrates on the case where an individual firm's payoff from adoption is increasing in the total number of adopters, and shows how a "bandwagon" can develop under those circumstances. Because we also examine a positive externality, some of our results are similar: individual neglect of the externality can lead to under-adoption, and agents able to internalize the externality can mitigate the inefficiency. But our model generates booms and busts even though a firm's payoff from adopting is completely independent of other firms' decisions. The dynamics derive from learning with endogenous information generation — an effect absent from the network externality literature.

Although the modeling approach is very different, the literature on bubbles in security prices, represented by Allen and Gorton (1993), Samuelson (1958), and Tirole (1982), addresses the same issues of rationality during apparent manias that we confront in this article. Rather than focusing on security prices, though, we focus on volume. Because frenzied trading volume is a major reason bubbles are perceived to be irrational, our study of the volume dimension can be seen as a step toward understanding this neglected dimension of apparent manias.

8. Conclusion

We present a simple model of the adoption of financial innovations that produces the boom and bust patterns that we often observe. Even though all information is public and all participants are rational, apparently irrational behavior emerges in equilibrium. For example: (1) Firms will pass up an innovation today, then adopt next period after receiving *unfavorable* information about the innovation. (2) On average, the last firms to adopt an innovation lose money. (3) The end of an innovation wave is completely unpredictable. (4) Even though the distribution of quality across innovations is symmetric, *most* innovations perform worse than expected. We also show that private adoption decisions lead to too little adoption of innovations, and discuss how financial intermediaries can reduce this social loss.

We assumed that the innovation is supplied in a perfectly competitive market with constant costs. Hence, it is priced at cost, so each firm captures the full surplus from adoption and acts accordingly. This follows from the basic assumption that the adoption of financial innovations cannot be restricted by patents or other methods of exclusion. The results are unchanged if we assume, instead, that the surplus is split between the supplier and the firm in fixed proportions. We also discuss how the results change under the alternate assumption that the innovation is supplied by a monopolist who can make firm-specific take-it-or-leave-it offers, and who therefore capture the entire surplus. It would complicate the analysis if suppliers of the innovation interacted strategically, a situation addressed by many models in the literature. Our assumption that the costs of adoption are constant over time could be relaxed to allow suppliers to have increasing efficiency. The temporal decrease in adoption costs would yield a lower volume of adoption for each given level of productive efficiency.

For reasons of tractability, we assumed that a firm realizes its payoff in the period that it adopts the innovation. If the uncertainty about the firm's payoff were resolved over many subsequent periods, the same economic forces would be at work but even richer dynamics would result. For example, a wave of adoption might halt and then start up again later when favorable surprises hit firms that adopted several periods previous. Another interesting extension would be to have firms choose between two competing innovations. To take a simple case, suppose all firms rank the two innovations identically, so all adopters in a given period choose the same innovation. In such a setting, private decision making could lead to gross inefficiencies in the choice of innovation as well as in the total amount of adoption. For example, suppose there is a lot of uncertainty about innovation A, but firms have a very precise estimate of the value of innovation B, and innovation B appears to be the better of the two. Firms acting independently will adopt innovation B, but the social value of some adoption of innovation A could be enormous because of the great uncertainty — if only someone would try it, we might find out that innovation A is a spectacular idea.

Appendix: Proofs of Propositions

Parallel between continuous case and discrete case

One can think of the model presented in this article as a continuous version of the following discrete setting. Suppose there are countably many firms $j = 0, 1, 2, \ldots$ with adoption costs $x^j = j + 1/2$, and firms x_t through (but not including) x_{t+1} innovate in period t. The number of new adopters is then $q_t = x_{t+1} - x_t$. Assume that the payoff to firm j from adopting the innovation is $\theta - x^j + \delta^j$, where the random shocks δ^j are distributed independently as $N(0, 1/h)$. Realized profits from adoption are observed by all and used in updating the prior distribution over θ. If we define $\hat{\gamma}_t = q_t^{-1} \sum_{j=x_t}^{x_{t+1}-1} \delta^j$ to be the average shock, one can see that the discrete system behaves like the continuous system. First of all, expected profits are $\bar{\theta}_t q_t - q_t(x_t + q_t/2)$ in both cases:

$$\begin{aligned}
\sum_{j=x_t}^{x_{t+1}-1} (\bar{\theta}_t - x^j) &= \sum_{j=x_t}^{x_t+q_t-1} (\bar{\theta}_t - j - 1/2) \\
&= \bar{\theta}_t q_t - q_t(x_t + q_t/2) \\
&= \int_{x_t}^{x_t+q_t} (\bar{\theta}_t - x)\ dx \\
&= \int_{x_t}^{x_{t+1}} (\bar{\theta}_t - x)\, dx.
\end{aligned}$$

Furthermore, the variance of $\hat{\gamma}_t$ is $(q_t^{-1})^2 \sum_{j=x_t}^{x_t+q_t-1} h^{-1} = (q_t h)^{-1}$ (since the δ^j are independent), so $\hat{\gamma}_t$ and γ_t have identical distributions. Therefore, the two systems not only have identical expected profits, but realized profits have identical distributions in the two settings, and the Bayesian updating rules are exactly the same.

Decision making would be somewhat more complicated in the discrete setting for a couple of reasons. First, one would have to bother with integer constraints. Second, a firm's decision to adopt would affect the amount of information generated, thereby affecting the value of the firm's option to adopt — in the continuous setting, the value of this option is independent of whether a particular firm chooses to adopt. This creates a potential for multiple equilibria.

Proof of Proposition 3. If the system has stopped by date t, then $\bar{\theta}_t \leq x_t$, else some firms would adopt. If the system *just* stopped at date t (i.e., $T = t$), then q_{t-1} was positive, and Equation (4) requires $x_t < \bar{\theta}_{t-1}$. Combining these two inequalities yields $\bar{\theta}_t \leq x_t < \bar{\theta}_{t-1}$ whenever $T = t$, i.e., $\bar{\theta}_T < \bar{\theta}_{T-1}$.

To prove that the marginal adopters lose money at date T, first note that the average profit for all the firms that adopt at date $T - 1$ is

$$\begin{aligned}
\pi_{T-1} &= \frac{1}{q_{T-1}} \int_{x_{T-1}}^{x_T} (\theta - x)\ dx + \gamma_{T-1} \\
&= \theta - (x_T + x_{T+1})/2 + \gamma_{T-1}.
\end{aligned}$$

Because we know that $\bar{\theta}_T$ must be between $\bar{\theta}_{T-1}$ and $\theta + \gamma_{T-1}$, the fact that $\theta_T < \theta_{T-1}$ implies that $\theta + \gamma_{T-1} < \bar{\theta}_T$. This gives

$$\pi_{T-1} < \bar{\theta}_T - (x_T + x_{T+1})/2. \tag{6}$$

Since the average adoption cost in this period is $(x_T + x_{T-1})/2$, but the adoption cost for the marginal adopter is x_T, the average payoff for the marginal adopter is

$$\pi_{T-1} - [x_T - (x_T + x_{T-1})/2] < \bar{\theta}_T - x_T \leq 0,$$

the first inequality by Equation (6) and the second because the system stopped at date T.

Proof of Proposition 4. Assume that, at date t, the estimated value of the innovation is $\bar{\theta}_t$ and the existing marginal firm is x. If q firms adopt at date t, the new marginal firm is $x + q$. From Equation (3), q must satisfy

$$\bar{\theta}_t - x - q - \beta \int_{x+q}^{\infty} (\bar{\theta}_{t+1} - x - q) f(\bar{\theta}_{t+1} \mid x, q)\ d\bar{\theta}_{t+1} = 0, \tag{7}$$

where $f(\cdot \mid x, q)$ is a normal density with mean $\bar{\theta}_t$ and variance $\sigma^2 = \frac{qh}{(h_0+xh)(h_0+(x+q)h)}$. Now define $u = (\bar{\theta}_t - x - q)/\sigma$; since there is a one-to-one correspondence between q and u, we can think of firms choosing u rather than q. If we divide Equation (7) by σ and introduce the change of variables $z = (\bar{\theta}_{t+1} - \bar{\theta}_t)/\sigma$, the chosen u satisfies

$$u - \beta \int_{-u}^{\infty} (u + z)(2\pi e^{z^2})^{-1/2}\ dz = 0.$$

Note that u depends only on β — it is independent of all other parameters. The hazard rate is

$$\begin{aligned} \Pr[\tilde{\bar{\theta}}_{t+1} \leq x + q] &= \Pr\left[\frac{\tilde{\bar{\theta}}_{t+1} - \bar{\theta}_t}{\sigma} \leq -\frac{\bar{\theta}_t - x - q}{\sigma}\right] \\ &= \Pr\left[\frac{\tilde{\bar{\theta}}_{t+1} - \bar{\theta}_t}{\sigma} \leq -u\right] = \Phi(-u), \end{aligned}$$

where $\Phi(\cdot)$ is the standard normal distribution function. Because u depends only on β, the hazard rate also depends only on β.

References

Allen, F., and G. Gorton, 1993, "Churning Bubbles," *Review of Economic Studies*, 60, 813–836.

Arrow, K. J., and A. C. Fisher, 1974, "Environmental Preservation, Uncertainty, and Irreversibility," *Quarterly Journal of Economics*, 88, 312–320.

Bagwell, L. S., 1992, "Dutch Auction Repurchases: An Analysis of Shareholder Heterogeneity," *Journal of Finance*, 47, 71–106.

Banerjee, A. V., 1992, "A Simple Model of Herd Behavior," *Quarterly Journal of Economics*, 107, 797–817.

Bernanke, B. S., 1983, "Irreversibility, Uncertainty, and Cyclical Investment," *Quarterly Journal of Economics*, 97, 85–106.

Bhattacharya, S., K. Chatterjee, and L. Samuelson, 1986, "Sequential Research and the Adoption of Innovations," *Oxford Economic Papers*, 38, 219–243.

Bikhchandani, S., D. Hirshleifer, and I. Welch, 1992, "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades," *Journal of Political Economy*, 100, 992–1026.

Bradley, M., and L. M. Wakeman, 1983, "The Wealth Effects of Targeted Share Repurchases," *Journal of Financial Economics*, 11, 301–328.

Brady Commission (Presidential Task Force on Market Mechanisms), 1988, *Report*, U.S. Government Printing Office, Washington.

Bulow, J., and P. Klemperer, 1994, "Rational Frenzies and Crashes," *Journal of Political Economy*, 102, 1–23.

Caplin, A., and J. Leahy, 1993, "Sectoral Shocks, Learning, and Aggregate Fluctuations," *Review of Economic Studies*, 60, 777–794.

Chari, V. V., and R. Jagannathan, 1988, "Banking Panics, Information, and Rational Expectations Equilibrium," *Journal of Finance*, 43, 749–761.

Dillon, U. S., T. Noe, and G. G. Ramirez, 1995, "Bond Calls, Credible Commitment and Equity Dilution: A Theoretical and Clinical Analysis of Simultaneous Tender and Call (STAC) Offers," working paper, Georgia State University.

Eckbo, B. E., 1990, "Valuation Effects of Greenmail Prohibition," *Journal of Financial and Quantitative Analysis*, 25, 491–506.

Farrell, J., and G. Saloner, 1985, "Standardization, Compatibility, and Innovation," *Rand Journal of Economics*, 16, 70–83.

Farrell, J., and G. Saloner, 1986, "Installed Base and Compatibility: Innovation, Product Preannouncements, and Predation," *American Economic Review*, 76, 940–955.

Geyelin, M., 1994, "The Business of Law: Lawyers and Clients," *Wall Street Journal*, Sept. 2, B3.

Gorton, G., 1985, "Bank Suspension of Convertibility," *Journal of Monetary Economics*, 15, 177–194.

Gul, F., and R. Lundholm, 1995, "Endogenous Timing and the Clustering of Agents' Decisions," *Journal of Political Economy*, 103, 1039–1066.

Jacklin, C. J., A. W. Kleidon, and P. Pfleiderer, 1992, "Underestimation of Portfolio Insurance and the Crash of October 1987," *Review of Financial Studies*, 5, 35–64.

Jensen, M. C., 1991, "Corporate Control and the Politics of Finance," *Journal of Applied Corporate Finance*, 4, 13–33.

Jensen, R., 1982, "Adoption and Diffusion of an Innovation of Uncertain Profitability," *Journal of Economic Theory*, 27, 182–193.

Jereski, L. 1997, "Investors Returning to CMOs: Mortgage Securities Have Yield as a Lure" *Wall Street Journal*, February 24, C1.

Kaplan, S. N., and J. C. Stein, 1993, "The Evolution of Buyout Pricing and Financial Structure in the 1980s," *Quarterly Journal of Economics*, 63, 313–357.

Katz, M. L., and C. Shapiro, 1986, "Technology Adoption in the Presence of Network Externalities," *Journal of Political Economy*, 94, 822–841.

Malatesta, P. H., and R. A. Walkling, 1988, "Poison Pill Securities: Stockholder Wealth, Profitability, and Ownership Structure," *Journal of Financial Economics*, 20, 347–376.

Mansfield, E., 1961, "Technical Change and the Rate of Imitation," *Econometrica*, 29, 741–766.

Mansfield, E., 1968, *Industrial Research and Technological Innovation: An Econometric Analysis*, Norton, New York.

Merton, R. C., 1992, "Financial Innovation and Economic Performance," *Journal of Applied Corporate Finance*, 4, 12–22.

Miller, M. H., 1986, "Financial Innovation: The Last Twenty Years and the Next," *Journal of Financial and Quantitative Analysis*, 21, 459–471.

Nasbeth, L., and G. F. Ray, eds., 1974, *The Diffusion of New Industrial Processes: An International Study*, Cambridge University Press, London.

Persons, J. C., 1994, "Signaling and Takeover Deterrence with Stock Repurchases: Dutch Auctions versus Fixed Price Tender Offers," *Journal of Finance*, 49, 1373–1402.

Pindyck, R. S., 1991, "Irreversibility, Uncertainty, and Investment," *Journal of Economic Literature*, 29, 1110–1148.

Raiffa, H., and R. Schlaifer, 1961, *Applied Statistical Decision Theory*, Harvard University Press, Boston.

Ross, S. A., 1989, "Institutional Markets, Financial Marketing, and Financial Innovation," *Journal of Finance*, 44, 541–556.

Rothschild, M., and J. E. Stiglitz, 1970, "Increasing Risk: I. A Definition," *Journal of Economic Theory*, 2, 225–243.

Ryngaert, M. H., 1988, "The Effect of Poison Pill Securities on Shareholder Wealth," *Journal of Financial Economics*, 20, 377–418.

Samuelson, P., 1958, "An Exact Consumption-Loan Model of Interest with or without the Social Contrivance of Money," *Journal of Political Economy*, 66, 467–482.

Stoneman, P., 1981, "Intra-Firm Diffusion, Bayesian Learning and Profitability," *Economic Journal*, 91, 375–388.

Tirole, J., 1982, "On the Possibility of Speculation under Rational Expectations," *Econometrica*, 50, 1163–1181.

Tsur, Y., M. Sternberg, and E. Hochman, 1990, "Dynamic Modeling of Innovation Process Adoption with Risk Aversion and Learning," *Oxford Economic Papers*, 42, 336–355.

Tufano, 1989, "Financial Innovation and First-Mover Advantages," *Journal of Financial Economics*, 25, 213–240.

Van Horne, J. C., 1985, "Of Financial Innovations and Excesses," *Journal of Finance*, 40, 621–631.

Welch, I., 1992, "Sequential Sales, Learning, and Cascades," *Journal of Finance*, 47, 695–732.