

Herding among Investment Newsletters: Theory and Evidence

JOHN R. GRAHAM*

ABSTRACT

A model is developed which implies that if an analyst has high reputation or low ability, or if there is strong public information that is inconsistent with the analyst's private information, she is likely to herd. Herding is also common when informative private signals are positively correlated across analysts. The model is tested using data from analysts who publish investment newsletters. Consistent with the model's implications, the empirical results indicate that a newsletter analyst is likely to herd on *Value Line's* recommendation if her reputation is high, if her ability is low, or if signal correlation is high.

HERD BEHAVIOR IS OFTEN SAID TO OCCUR when many people take the same action, perhaps because some mimic the actions of others. Herding has been theoretically linked to many economic activities, such as investment recommendations (Scharfstein and Stein (1990)), price behavior of IPOs (Welch (1992)), fads and customs (Bikhchandani, Hirshleifer, and Welch (1992)), earnings forecasts (Trueman (1994)), corporate conservatism (Zwiebel (1995)), and delegated portfolio management (Maug and Naik (1995)). This paper adds to the herding literature by developing and empirically testing a model that examines the incentives investment advisors face when deciding whether to herd. In particular, the paper tests whether economic conditions and agents' individual characteristics affect their likelihood of herding. The results are interpreted as a test of the predictions of the general class of cascade and herding models.¹

* Fuqua School of Business, Duke University. I am grateful to David Hirshleifer and Jaime Zender for comments that helped to substantially improve the paper. I would also like to thank Pete Kyle, Alon Brav, Doug Foster, Dan Graham, Rita Graham, Paul Harrison, Eric Hughson, Ron Lease, Mike Lemmon, Ernst Maug, Susan Monaco, Carl Moody, Barb Ostdiek, Drew Roper, Steve Slezak, René Stulz, Tom Smith, Brett Trueman, Vish Viswanathan, anonymous referees, and seminar participants at Duke, Tulane, and the University of Utah for helpful comments. I am grateful to Mark Hulbert and *The Hulbert Financial Digest* for providing the newsletter data, to David Hsieh for providing the daily S&P 500 index volatility estimates, and to Yunqi Han and the Federal Reserve Bank of Philadelphia for providing the data on Treasury bill forecasts. I am responsible for all remaining errors. The theoretical part of the paper was a chapter of my doctoral dissertation at Duke University. The empirical work was started while I was at the University of Utah.

¹ Welch (1996) also tests implications from the general class of herding models. He finds that brokerage recommendations are influenced by the consensus opinion of many brokers, especially in bullish market conditions or when the consensus proves to be wrong. He interprets the latter condition as being consistent with the implications from models that show that herding is sometimes based on little or no information (e.g., Scharfstein and Stein (1990) or Bikhchandani et al. (1992)).

We investigate the herding phenomenon using a simple model of stock analysts, patterned after the model in Scharfstein and Stein (1990). Each analyst in our model is one of two types, smart or dumb, although the type is unobservable to all. Smart analysts receive informative private signals about the stock market's expected return, dumb analysts receive uninformative signals. The smart analysts' signals are positively cross-correlated, implying that smart analysts following their private information have a tendency to act similarly. Consequently, in certain circumstances, an analyst can "look smart" by herding.

The analysts in the model act sequentially. The theoretical part of the paper investigates several factors that provide incentives for the second-mover to discard her private information and instead mimic the action of the first-mover. The analysts use Bayes' rule to determine their optimal actions and so prior public information is an important input in their decision-making processes, as is the precision of their private information (which we interpret as ability). The amount of correlation across informative private signals is also instrumental because it affects the degree to which analysts can look smart by herding. Finally, given that analysts maximize expected posterior reputation, their prior reputations also influence their optimal decisions.

After documenting the existence of parameter regions associated with "herding" and "deviating" equilibria, comparative statics are used to show that the incentive for the second-mover to discard her private information and instead mimic the market leader

1. increases with her initial reputation
2. decreases with her ability
3. increases in the strength of prior public information that is consistent with the leader's action
4. increases with the level of correlation across informative signals.

Though these factors are obviously interrelated, it is instructive to isolate the individual contribution of each to herding behavior, rather than blurring the distinction among them, as is often done.² The intuition behind the reputation implication is that analysts with high reputation (and salary) herd to protect their current status and level of pay.³

² For example, *Institutional Investor's* All-American Research Team is made up of high reputation analysts. Stickel (1990, 1992) shows that All-Americans give more accurate earnings forecasts and "follow the crowd" less often than non-All-Americans. Based on these findings, it appears that having a high reputation reduces the incentive to herd. In contrast, our model indicates that, to preserve status and salary, high reputation All-Americans have *greater* incentive to herd than non-All-Americans of equal ability. This implication may seem to contradict Stickel's (1990) finding that All-Americans "follow the crowd" less; however, his results reflect the *net* effect of reputation, ability, and other factors. We can isolate the effect of reputation on herding only by controlling for the other factors.

³ This is consistent with the implication in Prendergast and Stole (1996) that "youngsters" exaggerate private information to look knowledgeable, while "old-timers" make more conservative decisions. However, their prediction arises because old-timers do not want to deviate too far from their *own* past decisions, while our model predicts that agents herd on a *leader's* current decision to remain part of the crowd.

We test the implications of the theoretical model with a sample of investment newsletter asset allocation recommendations. A typical newsletter contains four to eight pages of analysis of current economic trends, combined with the newsletter editor's interpretation of how the trends affect various investment strategies. Though the mode and frequency of information transfer varies widely, the typical newsletter is published monthly and mailed to subscribers for an annual fee of approximately \$200; some letters also have a telephone, Internet, or fax updating service. The best known investment newsletter is the *Value Line Investment Survey*. Our sample consists of the market timing advice (i.e., recommendations about what portion of an investor's wealth should be invested in the stock market, cash, etc.) offered by 237 newsletter strategies over the period 1980 to 1992. Using these data, we identify the attributes of newsletters that herd on the advice of *Value Line*. Our strongest empirical finding is that herding decreases with the precision of private information, which lends support to the broad class of cascade and herding models. We also find evidence supporting the predictions that the incidence of mimicking *Value Line* increases with newsletter reputation, when a proxy for private information is highly correlated across analysts, and when prior information is strong.

The herding literature can be subdivided in the following manner, although these categories are neither exhaustive nor mutually exclusive: (1) informational cascades, (2) reputational herding, (3) investigative herding, and (4) empirical herding. (For a general review of the herding literature, see Devenow and Welch (1996).) The first two types of herding occur when individuals choose to ignore or downplay their private information and instead jump on the bandwagon by mimicking the actions of individuals who acted previously. Informational cascades occur when the existing aggregate information becomes so overwhelming that an individual's single piece of private information is not strong enough to reverse the decision of the crowd. Therefore, the individual chooses to mimic the action of the crowd, rather than act on his private information. If this scenario holds for one individual, then it likely also holds for anyone acting after this person. This domino-like effect is often referred to as a cascade. Research by Welch (1992), Bikhchandani et al. (1992), Banerjee (1992), Lee (1993), Smith and Sorensen (1994), Khanna and Slezak (1998), Banerjee and Fudenberg (1995), and Brandenburger and Polak (1996) investigates cascades.

Like cascades, reputational herding takes place when an agent chooses to ignore her private information and mimic the action of another agent who has acted previously. However, reputational herding models have an additional layer of mimicking resulting from positive reputational externalities that can be obtained by acting as part of a group or choosing a certain project. Our theoretical model falls in the reputational herding category. Other reputational herding models include Scharfstein and Stein (1990), Trueman (1994), Zwiebel (1995), Huddart (1996), and Prendergast and Stole (1996). Because these papers deal with issues similar to those investigated by our paper, they are discussed in detail in later sections.

Investigative herding occurs when an analyst chooses to investigate a piece of information she believes others also will examine. The analyst would like to be the first to discover the information but can only profit from an investment if other investors follow suit and push the price of the asset in the direction anticipated by the first analyst. Otherwise, the first analyst may be stuck holding an asset that she cannot profitably sell. Papers by Brennan (1990), Froot, Scharfstein, and Stein (1992), Dow and Gorton (1994), Hirshleifer, Subrahmanyam, and Titman (1994), and Golec (1997) fall into this group.

To an outsider, it can be difficult to differentiate whether an observed "herd" occurs for reasons put forth by models in any of the above categories. Indeed, there is a group of papers that investigate empirical clustering without directly testing the implications of the herding models. Clustering has been observed by Lakonishok et al. (1991), Peles (1993), Grinblatt, Titman, and Wermers (1995), Wermers (1999), Falkenstein (1996), Nofsinger and Sias (1996), and Wylie (1996) among pension funds, mutual funds, and institutional investors when a disproportionate share of investors engage in buying, or at other times selling, the same stock. Among other things, these papers suggest that clustering can result from momentum-following (also called "positive feedback investment," e.g., buying past winners) or perhaps from repeating the predominant buy or sell pattern from the previous period. We control for momentum-following but do not find that it contributes to clustering in our sample. The existing empirical literature largely tests whether "too many" investors appear to make the same choice; our paper attempts to more directly test the implications of the theoretical herding models.⁴ Another distinction is that our paper tests whether individual analysts take the same action as a "market leader" (*Value Line*) who sequentially precedes them, rather than examining "clusters" of analysts as is done in most other empirical herding papers.

Papers by Lamont (1995) and Ehrbeck and Waldmann (1996) are similar in spirit to ours. Lamont finds that a forecaster's age is positively related to the absolute first difference between his forecast and the group mean. Lamont interprets this as evidence that as a forecaster ages, evaluators develop "tighter priors" about the forecaster's ability, and hence the forecaster has less incentive to herd with the group. We investigate how the log of age affects herding but do not find a statistically significant relation. Ehrbeck and Waldmann find that the empirical patterns across T-bill forecasts are not supportive of simple reputational herding models, but instead seem to support behavioral hypotheses. In contrast, our results are consistent with explanations of herd behavior put forth by theoretical models.

The rest of the paper proceeds as follows. Section I develops the reputational herding model. Section II contrasts the model with cascades and other herding models, derives empirical implications, and tests the implications with investment newsletter data. Section III concludes and offers some thoughts on testing theoretical herding models.

⁴ Golec (1997) provides empirical evidence consistent with investigative herding.

I. The Reputational Herding Model

Consider an economy in which two risk neutral agents, A and B, evaluate an investment. The investment can have either a high (X_H) or low (X_L) payoff, with the prior probability of the high payoff being α . All of the analysts receive information about the investment payoff in the form of a privately observed signal, which they use to update α . A high signal (s_H) provides information in favor of the high investment payoff, a low signal (s_L) does the same for a low payoff. A or B superscripts are used to indicate which agent receives the signal.

Analysts are of two types, either smart or dumb. Smart analysts receive informative private signals regarding the investment payoff, dumb analysts receive purely random signals. The information structure is symmetric in that

$$\frac{1}{2} < \Pr(s_H|X_H, \text{smart}) = \Pr(s_L|X_L, \text{smart}) = p < 1.0$$

and

$$\Pr(s_H|X_H, \text{dumb}) = \Pr(s_L|X_L, \text{dumb}) = \frac{1}{2}, \quad (1)$$

where p (or $\frac{1}{2}$) measures the precision of the agent's signal. As in Scharfstein and Stein (1990), we assume that the type of a given analyst is unobservable to all. The common knowledge prior probability that an agent is smart is expressed as $\theta \in (0,1)$. We refer to θ as an analyst's initial reputation. Not knowing their own type, analysts expect to obtain a high signal given X_H with probability $p\theta + 0.5(1 - \theta)$.⁵

We implicitly assume the existence of private investors who "employ" the analysts. For example, if analysts' reports are in the form of investment newsletters, investors employ the analysts by subscribing to their newsletters. The investors do not know if a given analyst is smart or dumb (i.e., his type), but instead form an opinion as to whether an analyst is smart after observing his action and the investment outcome. Analyst wages (e.g., revenues from subscriptions) are assumed to increase linearly with the reputation of the analyst. Therefore, analysts choose a strategy that maximizes the probability that investors will think they are smart; that is, analysts maximize expected posterior reputation.

⁵ That is, the analyst's perception that her private information is accurate, which we refer to as her perception of her ability, is $p\theta + 0.5(1 - \theta)$. Notice that this quantity increases in p for all analysts. When analyzing the effect of signal accuracy on the incentive to herd, we sometimes loosely refer to p as the "ability" of all analysts, when to be more technically correct we should refer to "the effect of p on the analyst's perception of her ability." We do this to emphasize the separate effect of p as distinct from the effect of reputation θ on the analyst's perception of her ability.

We assume that the private signals of smart analysts are positively correlated and that the signals of dumb analysts are uncorrelated, with the level of correlation being measured by $\rho \in (0,1]$. This assumption is similar to that in Scharfstein and Stein (1990) (they assume that ρ is equal to one) and is critical for the reputational results that follow. If smart analysts' private information is positively correlated, they have a tendency to choose the same investment projects; that is, smart analysts often act as part of a group. In contrast, dumb analysts following their private information would appear to act independently. Analysts therefore deduce that by acting as part of a group they can "look smart," which provides an incentive to discard private information and "herd" to be part of a group. In a more general setting, reputational herding will occur as long as the positive correlation among smart analysts' signals is greater than that for dumb analysts' signals.⁶

The positive correlation assumption implies that the likelihood of two smart analysts both observing the "correct" signal is $(1 - \rho)p^2 + \rho p$. This expression represents a linear combination of the extreme cases of conditionally independent signals (p^2) and perfectly correlated signals (p).⁷ Analogously, the probability is $2p(1 - p)(1 - \rho)$ that two smart analysts receive different signals.

At the start of the game, the parameters α, p, θ , and ρ are common knowledge and each agent privately observes $\hat{s}^A$ or $\hat{s}^B$. In period 1, the leader, A, announces his investment decision publicly: $\hat{s}_H^A$ (invest) or $\hat{s}_L^A$ (do not invest). The follower, B, announces her investment decision in the second period. A primary focus of the paper is to determine the parameter values for which the announcements (with hat) are the same as private information (no hat) (e.g., $\hat{s}_H^B = s_H^B$) versus those that result in herding.

The investment outcome, X_H or X_L , is realized in period three. At that time, the investors use the agents' announcements, along with the investment outcome, to determine the revised probability an agent is smart: $\hat{\theta}^i(\hat{s}^A, \hat{s}^B, \alpha, p, \theta, \rho, X_k)$, $i = A, B$ and $k \in \{H, L\}$.⁸ Investors use Bayes' rule to calculate $\hat{\theta}^i(\cdot)$. For example, if the values for α, p, θ , and ρ are such that A is known to announce his private information in equilibrium, and B is known to herd, and investors observe $\hat{s}_H^A$, the following calculations determine the posterior probability that A is smart for each outcome:

$$\hat{\theta}^A(\hat{s}_H^A, X_L) \equiv \Pr(A \text{ smart} | \hat{s}_H^A, X_L) = \frac{\theta(1 - p)}{\theta(1 - p) + (1 - \theta)\frac{1}{2}}; \quad (2)$$

$$\hat{\theta}^A(\hat{s}_H^A, X_H) \equiv \Pr(A \text{ smart} | \hat{s}_H^A, X_H) = \frac{\theta p}{\theta p + (1 - \theta)\frac{1}{2}}. \quad (3)$$

⁶ Although not the situation modeled here, Hirshleifer (1993) points out that in certain situations it may be reasonable to think of dumb agents' information as being more positively correlated than smart agents' information, which can lead to anti-herding.

⁷ This relation is first identified in Graham (1996).

⁸ Henceforth, the functional dependence of $\hat{\theta}(\cdot)$ on α, p, θ , and ρ is suppressed for notational convenience.

In this situation, B's posterior reputation is equal to her initial reputation because she is known to herd. Notice that because B is known to herd in the situation depicted in equations (2) and (3), her action is informationless and is not used to update A's reputation. In contrast, if both A and B are known to truthfully announce their private information, their actions are evaluated relative to each other's action. The relative $\hat{\theta}(\cdot)$ updating rules are shown in the Appendix.

Analysts use Bayes' rule to update their beliefs. They calculate the probability of investment outcome given their signal and the common knowledge values of α , p , θ , and ρ . For example, the conditional probability of the low state given a low signal, from A's perspective, is

$$\Pr(X_L|s_L^A) = \frac{[\theta p + (1 - \theta)\frac{1}{2}](1 - \alpha)}{[\theta p + (1 - \theta)\frac{1}{2}](1 - \alpha) + [\theta(1 - p) + (1 - \theta)\frac{1}{2}]\alpha}. \quad (4)$$

Analyst B can condition on A's announcement as well as her own private information. If s^B differs from $\hat{s}^A$, and $\hat{s}^A$ reflects A's private information, then their information values cancel; for example, $\Pr(X_L|\hat{s}_L^A, s_H^B)$ equals the prior probability of the low investment outcome: $1 - \alpha$. In contrast, when $\hat{s}^A$ and s^B are identical, the probability is updated with the information from both signals and, for example, $\Pr(X_L|\hat{s}_L^A, s_L^B) > \Pr(X_L|\hat{s}_L^A, s_H^B)$.

This paper examines pure strategy Bayesian Nash equilibria. Given rational conjectures about the strategies of other analysts, each analyst chooses an action that maximizes his or her expected reputation. For example, if the leader's private information indicates that the investment outcome will be high, but the follower believes that it will be low, and if $(\alpha, p, \theta, \rho) = (0.30, 0.75, 0.60, 1.0)$, then the leader announces his private information and the follower herds in equilibrium. This behavior is optimal because the following inequalities hold:

$$\begin{aligned} & \hat{\theta}^A(\hat{s}_H^A, X_H)\Pr(X_H|s_H^A) + \hat{\theta}^A(\hat{s}_H^A, X_L)\Pr(X_L|s_H^A) \\ & > \hat{\theta}^A(\hat{s}_L^A, X_H)\Pr(X_H|s_H^A) + \hat{\theta}^A(\hat{s}_L^A, X_L)\Pr(X_L|s_H^A); \end{aligned} \quad (5)$$

$$\hat{\theta}^B(\hat{s}_H^A, \hat{s}_L^B, X_L)\Pr(X_L|\hat{s}_H^A, s_L^B) + \hat{\theta}^B(\hat{s}_H^A, \hat{s}_L^B, X_H)\Pr(X_H|\hat{s}_H^A, s_L^B) < \theta. \quad (6)$$

The left-hand side of equation (5) is the leader's expected reputation when he truthfully announces his private information (s_H^A), given that equation (6) holds. The right-hand side of equation (5) is the leader's expected reputation when he announces opposite his private information. Analogously, the left-hand side of equation (6) is the follower's expected reputation if she truthfully announces her private information (s_L^B), given that equation (5) holds, while the right-hand side of equation (6) is her reputation if she herds. The left-hand side of equation (6) is consistent with an out-of-equilibrium conjecture: analyst B deviates (i.e., does not announce the same signal as A)

only when telling the truth. This conjecture follows from the same refinement strategy used by Scharfstein and Stein (1990), and is “reasonable” in the sense that deviating to tell the truth is more attractive than deviating to lie for “herd equilibrium” parameter values.⁹

The following propositions describe the main theoretical results of the paper.

PROPOSITION 1:

- A. *If A and B have the same private information, they always make the same investment decision.*
- B. *If A and B do not have the same private information, then there exist separate nonempty $(\alpha, p, \theta, \rho)$ equilibrium parameter sets for which one of the following holds:*
 - (i) *The leader A and follower B both announce their private information*
 - (ii) *The leader A announces his private information but the follower B herds (i.e., B announces opposite her private information by mimicking the leader)*
 - (iii) *The leader A announces opposite his private information (e.g., announces $\hat{s}_L^A$ even though he observed s_H^A) and the follower B truthfully announces her private information.*

Proof: All proofs are in the Appendix.

The essence of Proposition 1 is that the leader will sometimes, but not always, announce his private information, and the follower will sometimes, but not always, herd. By expanding the Scharfstein and Stein (1990) parameter set, we are able to generalize their conclusion that the leader always announces truthfully and the follower always herds.

The parameter sets described in the first proposition are separated by hyperplanes that are defined by “indifference functions” (see the Appendix for details). By taking partial derivatives of the indifference functions with respect to model parameters, we gain insight into the economic conditions and incentives that lead to herding behavior, as described in the second proposition. Proposition 2 is stated for the situation in which the leader and follower receive different private signals.¹⁰

⁹ As noted by Scharfstein and Stein (1990), “perverse” equilibria can exist with different out-of-equilibrium beliefs. For example, an anti-herding equilibrium can be supported if followers are believed to deviate from the leader only to announce *opposite* their private information.

¹⁰ If both analysts receive the same private signal, we know from Proposition 1 that they both *always* take the same action; therefore, we do not explicitly calculate comparative statics for the “corner solution” in which the analysts receive the same private information. This has an important empirical implication in that, for example, high reputation analysts will take the same action as the leader when they receive the same private information, but may take the opposite action when they receive differing private information.

PROPOSITION 2: *The leader's incentive to truthfully announce private information*

- (i) *increases in ability (p)*
- (ii) *increases in informative signal correlation (ρ)*
- (iii) *increases in initial reputation (θ)*
- (iv) *increases (decreases) in the strength of prior information (α) when it is consistent (inconsistent) with his private information.*

*When the leader announces truthfully in equilibrium, the follower's incentive to truthfully announce private information*¹¹

- (i) *increases in ability (p)*
- (ii') *decreases in informative signal correlation (ρ)*
- (iii') *decreases in initial reputation (θ)*
- (iv) *increases (decreases) in the strength of prior information (α) when it is consistent (inconsistent) with her private information.*

We now discuss the intuition behind the propositions. When we analyze the follower's incentives, we assume that the leader is known to truthfully announce his private information in equilibrium. If A does not announce his private information, B's incentives are identical to those discussed for A (because A's announcement is effectively ignored, and B becomes a "leader"), so we do not repeat those implications. For the following discussion, assume that the analysts privately observe s_H^A and s_L^B . (Because of symmetry in the model, the arguments also hold if the analysts observe s_L^A and s_H^B .)

Prior information (α): Without knowledge of the analysts' private information, the effect of prior public information is ambiguous empirically; consequently, to keep this section streamlined, the effects are discussed at the end of the Appendix. Perhaps the most important empirical implication is the following: When a group of analysts make similar predictions in the face of strong prior information it often is because the analysts effectively herd on the prior, not on other analysts.

Informative signal correlation (ρ): The assumption that smart analysts' signals are positively correlated, but dumb analysts' signals are not, is a key input into the reputational incentive to herd. As informative signals become more positively correlated, it becomes more likely that smart analysts will act as part of a group, and the incentive to herd increases. Conversely, as correlation declines, follower analysts are more likely to announce their private information. With respect to A, the leader is more likely to announce his private information when he anticipates that the follower will not contradict him. Therefore, as correlation declines, and followers become less likely to herd, the leader has less incentive to announce his private information.

¹¹ When the leader A does not announce his private information in equilibrium, investors ignore his advice and the follower's incentives to announce her private information are shown in (i)–(iv) because the follower essentially assumes the role of "market leader."

Ability (p): As an analyst's private information becomes more accurate, it becomes more likely that he will announce his private information because it becomes more likely that his revised belief about the investment outcome will be consistent with his private information. For example, if $\alpha < \frac{1}{2}$ and A observes s_H^A , the strength of his posterior belief that X_H will occur (in which case he announces $\hat{s}_H^A$) increases with p .

Initial reputation (θ): The incentives associated with initial reputation are opposite for the two analysts. Recall that, because they do not know if they are smart or dumb, the analysts expect to obtain a high signal given X_H with probability $p\theta + 0.5(1 - \theta)$. For the leader, this probability increases in θ , and so higher initial reputation makes it more likely that the leader will truthfully announce his private information.

With respect to the follower, if her private information is exactly offset by A's announced private information, her posterior belief that X_L will occur is $1 - \alpha$, regardless of the value of θ . That is, B's belief about the quality of her private information (relative to the quality of A's information) is invariant with respect to θ . Consequently, her belief about the likelihood of X_L occurring is not affected by θ , and a "positive" effect of θ analogous to that described for the leader is absent. Instead, θ affects B's incentives because it represents her salary if she herds in equilibrium (because investors cannot draw inference on her type when she herds and therefore her reputation remains $\hat{\theta}^B = \theta$). When θ is low, B receives a small salary unless she distinguishes herself from the pack and improves her reputation. When θ is high, she has strong incentive to herd to preserve her reputation and high salary.¹² Therefore, the follower's incentive to announce her private information decreases with her initial reputation θ . Notice that the "salary preservation" incentive does not apply to A because he acts first and does not have the opportunity to herd on another agent's action.

If age and reputation are positively correlated, the notion that high-reputation analysts are more likely to herd has the same empirical implication as a result from Prendergast and Stole (1996), who find that experienced agents herd on their own past actions to look like they are (and have always been) knowledgeable. Though the empirical implication is similar, in our model followers herd on another analyst, without any direct reference to their own past actions. Lamont (1995) compares the implication that high-reputation analysts herd because they have "farther to fall" to the counter-argument that investors may hold "tighter priors" on experienced analysts' underlying abilities, thus providing less benefit to herding. He finds that

¹² In our model, low reputation agents herd less because they have "nothing to lose"; instead of herding, they act according to their private information, thereby improving informational efficiency. Diamond (1989), in his analysis of reputation acquisition in debt markets, also concludes that those of low standing are likely to have a "nothing to lose" mentality. However, he argues that this promotes *inefficient* behavior because these borrowers act "too risky" in a gamble to achieve high reputation and the benefits it provides.

older agents herd less than young agents, counter to the implications from Prendergast and Stole (1996) and our model. In the empirical section we explore whether age is related to herding behavior.

Our model is also related to work by Trueman (1994) and Zwiebel (1995). Although the general topic area is similar, these reputational herding models are fairly dissimilar to ours. In a model with independently distributed signals, Trueman derives an incentive to “herd on the prior.” By assuming that the *ex ante* likelihood of the “low payoff” occurring is relatively high, Trueman gives relatively uninformed analysts the incentive to sometimes ignore their private information and announce “low.” Clustering on the low announcement occurs because the low payoff is *ex ante* favored and relatively informed analysts are therefore likely to announce “low.” Thus, analysts can look informed and enhance their reputation by “herding” on the low announcement.¹³

In Zwiebel (1995), managers sometimes herd on projects, rather than on the actions of others as they do in our model.¹⁴ Zwiebel’s managers choose between a project based on existing technology and an innovative project, the latter having a higher expected return but also leading to less precise reputation evaluation. Though low- and high-ability managers both choose the innovative project, medium-ability managers opt to “herd” on the existing technology. For a medium-ability manager, the increased risk of being misevaluated (and terminated) due to less precise evaluation more than offsets the higher expected return associated with the innovative project. In Zwiebel’s model, reputation depends on absolute action (i.e., the choice between innovative and existing technology) and relative performance (relative to a market benchmark, which is based on the aggregate outcome of the managers’ choices). That is, the actual actions taken by other managers are not used by the labor market when forming an opinion of an individual manager’s reputation; the other managers’ actions are only used to form the benchmark. In our model, relative action (i.e., relative to other analysts’ actions) is very important.

II. Empirical Tests for Herding among Investment Newsletters

In this section we use regressions to test the model. Because the reputational model is closely related to cascade and other herding models, we interpret our empirical analysis as a step toward testing implications from the

¹³ Brandenburger and Polak (1996) also discuss the “herd on the prior” intuition. Trueman’s model also incorporates a second form of herding: if the first-mover’s announcement is inconsistent (consistent) with the second-mover’s private information, the second-mover becomes relatively less (more) likely to announce her private information. This cascading effect occurs because of Bayesian updating (i.e., the second-mover updates her prior after observing the first-mover’s announcement, which alters her belief about the probability that the low payoff will occur), and is also captured in our model and in cascades models.

¹⁴ Holmstrom (1982) presents a model in which reputational concerns cause managers to never invest.

general class of herding models, as called for by Hirshleifer (1993, p. 158). We begin by stating the empirical implications.

A. Empirical Implications from the Reputational Herding Model

The best possible test of a herding model would compare analysts' announcements to their private information signals. This comparison would allow one to know whether analysts discard their private information when they take the same actions as others. Unfortunately, we are not aware of a data source containing analysts' private information. Instead, we test the model's implications by observing whether an analyst "acts the same as the leader."

Proposition 2 derives the follower's incentives for the case in which her private information differs from the leader's. However, we know from Proposition 1 that the follower will *always* take the same action as the leader when her private information is the same as the leader's (also see footnote 10). Thus, the model implies, for example, that analysts making announcements that differ from the leader's will *ceteris paribus* have low reputation, but that followers taking the same action as the leader will be comprised of high reputation analysts who discard their private information to mimic the leader *and* low reputation followers who receive the same private signal as the leader. By examining whether the follower "acts the same as the leader," we should find that the *average* reputation of deviators is lower than the average reputation of those acting the same as the leader, but the power of our tests is reduced relative to tests based on analysts' actual private information.

Finally, we perform our tests on data points for which the comparative static implications about the reputation and correlation variables are unambiguous (i.e., we consider observations with parameter values for which the model implies that (i), (ii'), (iii'), and (iv) of Proposition 2 hold). For this sample, the conditions that lead to followers acting the same as the leader (which we refer to as herding) can be summarized as follows:

EMPIRICAL IMPLICATIONS: *All else equal, herding should be observed when*

1. *follower analyst ability (p) is low*
2. *private informative signals are highly correlated (ρ)*
3. *follower analyst initial reputation (θ) is high*
4. *prior information (α) is strong and consistent with the leader's announcement.*

B. Relation between the Reputational Herding Model, Cascades Models, and Other Herding Models

The implication that analysts herd more often when the precision of their private information (p) is low is not unique to the reputational model. For example, Bikhchandani et al. (1992, p. 1002) state that a "low-precision individual imitates a high-precision predecessor" and Trueman (1994) finds that the incidence of herding by low-precision analysts is relatively high.

Given that the models are unanimous in their predictions, we expect that this should be a strong effect empirically.

The result that herding increases with informative signal correlation is specific to the reputational model (and, by extension, to Scharfstein and Stein's (1990) model) in the sense that most other herding investigations treat information as being independently distributed. However, we conjecture that introducing positive information correlation into the other models would lead to increased herding because the agents would become relatively more likely to receive the same private information, and hence act similarly. Thus, we interpret a positive relation between herding and signal correlation as being consistent with the general class of models.

The result that the incentive to herd increases with initial reputation (θ) is unique to the reputational herding model, and is crucially tied to the assumptions of positive signal correlation and relative performance evaluation. When analysts are evaluated individually (or based on absolute performance), as in other cascade and herding models, they cannot "protect their reputation" by herding. To the extent that reputation improves an individually evaluated analyst's perception of the precision of her private information (e.g., if an analyst does not condition on a leader's action), reputation will be negatively related to the propensity of herding. Or, if an individually evaluated follower's private information exactly offsets the leader's, the follower always announces consistent with the prior, regardless of the value of θ . It is only when the follower is evaluated relative to the leader that initial reputation provides a positive incentive to herd. Thus, the implication that reputation and herding on the leader are *ceteris paribus* positively correlated distinguishes the reputational model from the general class of herding models.

C. The Newsletter Data Sample

Implications from the theoretical model are tested with a collection of asset allocation advice (i.e., recommendations about how investment holdings should be split between cash and equities). The sample of investment newsletter recommendations is first used by Graham and Harvey (1996, 1997) and ranges over the thirteen years ending in December 1992.¹⁵ We examine whether a follower's first recommendation in a given month mimics the market leader's last recommendation in the previous month. If a newsletter does not make a recommendation during a month, then the last position from the previous month is maintained as the recommended asset allocation. We assume that newsletters maximize expected reputation in each month.

As discussed below, our primary specifications use two different definitions of newsletter reputation and consequently also examine two different subsets of the data sample. We initially use a static definition of reputation based on

¹⁵ See Graham and Harvey (1996) for a complete description of the institutional details of the data. See Metrick (1999) for an analysis of the selectivity advice from the newsletter sample. See Womack (1996) for an analysis of the timing and selectivity advice from a different sample of analyst recommendations.

whether the newsletter exists in the first observation of the sample. This analysis uses the “short-horizon” sample, which consists of 279 observations spanning 1980 to 1981 and represents advice from twenty-three investment newsletter strategies. Our second measure of reputation is dynamic in the sense that it is updated monthly based on economic conditions, the newsletter’s action, the leader’s action, and the market return. The dynamic measure of reputation is defined over the “long-horizon” sample period, 1980 to 1992, and contains 5,293 recommendations made by 237 newsletter strategies.¹⁶

The *Value Line Investment Survey* is the best known investment newsletter. *Value Line*’s advice is freely observable to all market participants because it is available without charge in most public libraries. Furthermore, *Value Line* is well-respected and its recommendations are known to be studied extensively (see, e.g., Black (1973), Copeland and Mayers (1982), Stickel (1985), and Lewis, Rogalski, and Seward (1997)). Given its stature and wide availability, we argue that investors use the recommendations made by *Value Line* as a benchmark against which they compare the advice of other newsletters. Knowing that they will be judged relative to *Value Line*, the other newsletters can be thought of as followers in the context of the theoretical model, and *Value Line* can be thought of as the market leader.¹⁷ *Value Line* makes a recommendation in each year-month of the sample.

We represent $X_H(X_L)$ by an upward (downward) movement in the equity market, as measured by a positive (negative) excess return on the S&P 500 index. Analogously, newsletter announcements are represented by changes in the recommended long position: a high (low) announcement is conveyed by increasing (decreasing) the recommended position in the equity market. So the basic idea is: Does a follower increase equity weights after observing that *Value Line* increased its market position?¹⁸

¹⁶ There are 330 total observations available for the short-horizon analysis; however, 51 observations are deleted from the sample because they represent parameter values that are inconsistent with parts (ii') and (iii') of Proposition 2 (i.e., they have different implications about the sign of the reputation and correlation variables). There are 12,426 total observations available for the “long-horizon” analysis, 7,132 of which are deleted because they represent parameter values that are inconsistent with the implications in parts (ii') and (iii') of Proposition 2.

¹⁷ Our approach differs from that in many empirical herding papers in that our main specification does not analyze whether the follower mimics the consensus opinion. For our sample, it is not clear how a newsletter would know what the consensus opinion is, short of subscribing to more than 200 newsletter forecasts. It seems highly unlikely that all newsletters subscribe to every other newsletter to determine the consensus. In contrast, any newsletter can observe *Value Line*’s announcement simply by going to the library. Nonetheless, we perform an unreported analysis that provides weak evidence that followers mimic the group mean.

¹⁸ The model derived in Section I is a one-period model in which agents all have the same initial reputation. In Section II we test the model’s implications using cross-sectional time-series data, implicitly assuming that the model’s comparative statics carry over to a more general setting. We can show that the comparative statics are qualitatively unchanged if the leader and follower have different initial reputations. In particular, with one exception, the propositions remain unaltered. The one exception involves the effect of initial reputation on the follower’s incentive to herd. In Section I, if the leader’s action conflicts with the follower’s private information, θ affects B only because it represents her initial salary; a large θ provides an

D. Empirical Specification and Technique

The dependent variable for the empirical tests is a dummy variable equal to one if a follower newsletter herds on *Value Line* and equal to zero otherwise. We use a logistic transformation of the explanatory variables, which are described next, because of the binary nature of the dependent variable. (The qualitative results do not change if we do not transform the variables.) The model is estimated using Parzen's (1957) weights (see Andrews (1991)) in Hansen's (1982) generalized method of moments. This approach estimates consistent coefficients in the presence of conditionally heteroskedastic, serially correlated errors. The explanatory variables are described next. At the end of this section, we discuss regressions that remove all of the economy-wide effects (i.e., correlation, prior information, and ability) and thereby verify that our proxies for these effects do not distort the inference we draw on the other explanatory variables.

Ability (p): The model predicts that analysts are more likely to deviate when private information is accurate, as measured by the economy-wide value of p . We measure p as the proportion of analysts who make the "correct" recommendation, which is consistent with the model's definition of "for a given realization X_H or X_L , p measures the accuracy of informative signals." A forecast is "correct" if a newsletter recommends increasing equity weights in period t before the monthly market excess return is positive in $t + 1$, or decreasing equity weights preceding a market decline. All else equal, there should be less herding when p is high.

The models in Bikhchandani et al. (1992, p. 1002) and Trueman (1994) imply that *individual* analysts with high ability are more likely to deviate. In our model, ability is held constant across analysts in each month-year. Thus, our primary empirical analysis does not examine cross-sectional variation in ability but rather tests whether the economy-wide measure of ability varies with the incidence of herding in the time-series. (However, in alternative regressions, we find that the implications about ability are unchanged when we use a measure of ability that varies across newsletters.)

Magnitude of the Prior Probability of Market Movements (α): There is mounting evidence that stock returns are predictable using publicly available information (e.g., see Fama and French (1989) or Harvey (1989)). For example, Harvey's (1989) statistical model predicts the one-month-ahead excess market return using a January dummy and four lagged instruments: the excess return on the CRSP equally-weighted index, the excess return on a three-month Treasury bill, Moody's Baa minus the Aaa bond yield, and the excess dividend yield on the S&P 500 index. We use Harvey's model to measure the strength of publicly available information, dividing the absolute value of the

incentive to herd. However, if the follower has a unique reputation, a relatively large value for θ^B gives evidence that her private information is accurate and provides a disincentive to herd. This "information accuracy" incentive offsets the "salary preservation" incentive; however, numerical simulation indicates that the salary preservation effect dominates for most parameter values. On average, high initial reputation provides a positive incentive to herd. The general comparative statics also carry over to a dynamic setting.

forecast from Harvey's model by the largest absolute value over the sample period. Dividing by the largest absolute value bounds the variable between zero and one, so it can be interpreted as a probability.

The effect of prior information is somewhat ambiguous for the entire class of herding models. The general notion from the theory is that strong prior (or public) information can lead to clustering because it may induce a cascade beginning with the leader that will also affect the follower. However, if the leader happens to make an announcement that is contrary to the prior, the follower's incentive to herd with the leader is decreasing in the strength of the prior information. Our primary specification calls for a positive relation between the magnitude of prior information and the incidence of herding. We also perform an analysis in which the prior information variable is interacted with a dummy variable that indicates whether *Value Line's* announcement is consistent with the prior information.

Signal Correlation (ρ): The model predicts a positive relation between herding behavior and the degree to which informative signals are positively correlated. Given that private information signals are not observable, we propose an alternative measure of signal correlation. First, recall that the excess return on T-bills is one of the instruments that is useful in predicting (via Harvey's (1989) model) the one-month-ahead market return. We conjecture that analysts incorporate private forecasts of the T-bill rate as part of their private information about the anticipated market movement. Consequently, we use the scaled cross-sectional standard deviation of private forecasts of the three-month T-bill rate as our measure of signal correlation ρ ; ¹⁹ the scaling is accomplished by dividing each observation by the maximum sample standard deviation so the variable can be interpreted as a correlation coefficient. This variable is based on quarterly forecasts collected by the Federal Reserve Bank of Philadelphia in its Survey of Professional Forecasters. The survey of this variable started in the third quarter of 1981, so the sample is truncated slightly when correlation is included in the specification.

Our variable is not a perfect proxy for the private information of investment newsletters. The identities of the economists who make the T-bill forecasts are unknown but are unlikely to be the same as the editors of the newsletters in our sample. Further, the T-bill forecast is at best one element of the information set used by the newsletter editors to form their predictions about future market movements, and the T-bill forecasts are updated quarterly, rather than monthly like the rest of our data. Finally, it is possible that the T-bill forecasts themselves are influenced by herding behavior and hence a positive relation with newsletter herding may not be causal. While recognizing these concerns, we feel that the dispersion in T-bill forecasts may still be a reasonable proxy for ρ . It is unlikely that the economists in the Federal Reserve Bank's survey are influenced by herding behavior because their identities are not released to the public (i.e., their potential

¹⁹ We would prefer to measure correlation based on (private) forecasts of the market return, rather than based on one element of analysts' information sets; however, we are unable to locate a sample of market return forecasts.

clients). Consequently, to the extent that T-bill forecasts are important inputs into the information sets of the newsletter editors, and to the extent that the professional economists and newsletter editors draw their private information from the same sources, we feel that our variable is a reasonable measure of ρ .

Reputation (θ): Mark Hulbert produces *The Hulbert Financial Digest* to evaluate the investment newsletters in our sample. Hulbert starts his analysis in June 1980 by collecting information on a sample of eleven newsletters that had high reputations at that time. He then adds information on other newsletters through the end of 1981, and each year since. Given that Hulbert starts his sample with the highest reputation letters, the letters added in July 1980 and beyond have relatively lower reputations than those in the initial sample (or possibly did not exist in June 1980). Our first definition of reputation uses a dummy variable with a value of one (zero) if the letter was added to the sample in (after) June 1980 to identify a letter with a high (low) reputation. Given that this definition of reputation is static, our analysis using this variable focuses on the short-horizon sample, with observations in the 1980 to 1981 period.

We also specify a dynamic measure of reputation that updates a newsletter's prior reputation each period using the $\hat{\theta}(\cdot)$ function. In particular, we assign an initial reputation of 0.65 to those newsletters in Hulbert's initial sample (including *Value Line*) and an initial reputation of 0.35 to the other letters. If the $(\alpha, p, \theta, \rho)$ parameters are such that the follower newsletter's equilibrium behavior is to herd, regardless of its private information, its reputation is not updated in that month. If the letter deviates, or if the parameter values are such that it is common knowledge that the follower does not discard private information in equilibrium but instead announces its private signal, the letter's reputation is revised using $\hat{\theta}(\cdot)$ functions found in the text and Appendix.

The actual values of α , p , θ , and ρ (as assigned to the explanatory variables), and the actual occurrence of either X_L or X_H , are used as inputs into $\hat{\theta}(\cdot)$. For example, consider an observation in period t for which ability p is 0.60, the strength of prior information α is 0.40, signal correlation ρ is 0.33, initial reputation θ^1 for a particular newsletter is 0.65, and X_L occurs (i.e., the excess return on the market is negative). The numerical values of the explanatory variables enter the period t regression directly, and they are also the inputs used to update reputation for this particular newsletter using $\hat{\theta}(\cdot)$ functions such as equation (A1) in the Appendix. The updated reputation value becomes the "prior reputation" in period $t + 1$. This calculation is done separately for each newsletter, in each month of the sample. Given that the dynamic specification of reputation is updated each month, it is used in the long-horizon regressions over the entire 1980 to 1992 sample.

For our primary specification, we initially assign *Value Line* a reputation of 0.65 because it is in Hulbert's initial sample; however, the qualitative results do not change if *Value Line*'s reputation is held fixed at 0.50 for the entire sample period. Moreover, the conclusions are unchanged if we assign all newsletters an initial reputation of 0.50 and allow the variation in rep-

utation to occur entirely from the monthly $\hat{\theta}(\cdot)$ updating. This latter check indicates that the reputation results hold when the dichotomization related to appearing in Hulbert's initial sample is completely ignored; in contrast, the short-horizon reputation results are based entirely on the Hulbert dichotomization.²⁰

Control variables: About one-third of the newsletter strategies have a stated policy of never recommending a short position in the equity market. To ensure that this restriction does not influence our findings, we include a dummy variable equal to one if a letter restricts its advice so as to never short the market.

Grinblatt et al. (1995) and Wermers (1999) conclude that a large portion of herding behavior occurs when analysts "momentum-follow," that is, they buy recent winners or sell recent losers.²¹ If both the leader and follower momentum-follow, then it can look like the latter is herding on the former, when in fact both are simply mimicking the market movement. To make sure that momentum-following does not drive our results, we include a dummy variable equal to one if a newsletter's recommended change in equity weight has the same sign as the lagged S&P 500 return and equal to zero otherwise. We expect the coefficient of this variable to be positive if momentum-following contributes to observed empirical clustering.

Annual dummy variables are included in the long-horizon estimation to account for trends not captured by the other variables. A measure of market uncertainty is also included in a version of the long-horizon regression to control for the effects of market volatility on the incentive to herd (alternative regression 7 in Table I). We use the standard deviation of five-minute returns on the S&P 500 futures contract to measure daily market volatility. The volatility series begins in April 1982, and, therefore, when this variable is included the long-horizon regression is estimated with a truncated sample. The annual dummies and volatility variable are not included in the short-horizon analysis.

Finally, our two most problematic explanatory variables are the measures for prior information and private signal correlation. Ideal specifications for these variables should vary by year-month but be identical across newsletters within each year-month. To ensure that these proxies in no way influence the inference that we draw on the other variables, we run a long-

²⁰ Note that, given values of α , p , θ , and ρ , the dynamic reputation variable is calculated directly from the model. In this sense, the regressions in this paper are closer to being direct tests of a herding model than are the tests in most empirical herding papers. Note also that if the follower herds in equilibrium, the $\hat{\theta}(\cdot)$ function is not updated; thus, by construction, herding on *Value Line* will not lead to a newsletter obtaining a high reputation, eliminating concerns that the reputation variable is "hardwired" to be positively correlated with herding behavior. Finally, notice that (to be consistent with the model) the ability variable only varies by month, thus capturing an absolute and economy-wide "accuracy" effect, which is empirically distinct from the "relative performance" effect measured by the reputation variable.

²¹ The Grinblatt et al. (1995) and Wermers (1999) analyses look at herding on individual stocks; our analysis examines the market.

horizon regression that is blanketed with dummy variables to remove all year-month effects (alternative regression 6). In particular, we include a separate dummy for each year-month in the sample (except the first), 149 total variables. This specification allows us to focus on the marginal effects of momentum-following and reputation, after purging the direct influence of signal correlation, public information, and ability.²²

E. Empirical Results

Short-horizon results: The estimation results appear in Table I. We refer to the specification that includes all the explanatory variables as the “base case” (see Panel A). The adjusted R^2 of 9.5 percent for the short-horizon base case indicates that the model explains the data reasonably well. The negative coefficient of the ability variable (significant with a p -value of 0.013) is consistent with the hypothesis that when private information is accurate, newsletters are relatively unlikely to herd. This finding is consistent with Welch’s (1996) conclusion that herds are often based on little or no information. The positive coefficient of the strength of prior information is consistent with the notion that herding occurs in the presence of strong prior information, although this variable is not statistically significant.

The positive coefficient of the reputation variable implies that high reputation newsletters are relatively likely to herd on *Value Line*’s recommendation. This lends support for the reputational model beyond that for the general class of herding models. The control for newsletters that have a restriction against shorting the market is not related to the dependent variable in a significant manner, nor does the evidence indicate that momentum-following contributes to clustering. Finally, in separate specifications, we drop the prior information, reputation, and ability variables, respectively (see alternative regressions 2a, 3a, and 4a in Table I). The inference on the remaining variables does not change qualitatively, indicating that any misspecification in any one of the proxies (or pairwise collinearity) is not driving our results.

Long-horizon results: The long-horizon regression results are shown in Panel B of Table I. The overall fit of the base case regression is measured by an adjusted R^2 of 15.26 percent. The base case findings provide statistically significant support for the hypotheses that newsletters herd on *Value Line* when their reputation is high, average ability is low, prior information is strong, and signal correlation is high. In contrast, the momentum-following and short-sales-restriction variables are not significantly related to the incidence of herding. The estimated coefficients of the annual dummies are not reported to save space, although most are statistically significant.

²² In a final specification, we define herding by the correlation coefficient between an individual newsletter’s recommendations and those from *Value Line* (rather than using a binary measure to define the dependent variable); this analysis contains one observation per newsletter, purging the effects of ability, signal correlation, and public prior information, but it still allows us to test for reputational effects.

Table I
Regression Results

Table I contains estimated coefficients from regressing the propensity to herd, as measured by a dummy variable equal to 1 if an investment newsletter recommends asset allocation weights that change in the same direction as *Value Line's* recommendation and equal to 0 otherwise, on a logistic transformation of explanatory variables. The explanatory variables include the scaled expected one-period-ahead excess return on the S&P 500 (prior information), ex post accuracy of the newsletters' investment recommendations (ability), the cross-sectional standard deviation in forecasts of the three-month T-bill divided by the maximum standard deviation (signal correlation); a no-short dummy variable equal to one if a newsletter strategy has an explicit restriction against recommending short equity positions; an indicator variable $I(\cdot)$ that is assigned a value of 1 if the recommended change in investment weights has the same sign as the S&P 500 excess return from the previous period and is equal to 0 otherwise (momentum-following); the standard deviation of the S&P 500 excess return as measured with five-minute returns on the day the newsletter makes its recommendation (market uncertainty), and annual dummy variables or a dummy variable for all but one year-month in the sample. The measure of reputation in the short-horizon regression is a dummy variable equal to 1 if the letter was among the first group analyzed by the *Hulbert Financial Digest*. The reputation variable in the long-horizon analysis has an initial value of 0.65 (0.35) if the variable is (is not) included in Hulbert's initial sample, and is updated each month using the $\hat{\theta}(\cdot)$ function. The p -values (based on two-sided t -tests and shown below the estimated coefficients in parentheses) are consistent with respect to both conditional heteroskedasticity and serial correlation.

Panel A: Short Horizon (1980–1981)					
Explanatory Variable	Expected Sign	Base Case 1a	Alternative Regression 2a	Alternative Regression 3a	Alternative Regression 4a
Intercept		0.495 (0.276)	0.549 (0.110)	1.202 (0.003)	0.532 (0.307)
Abs[$E_{t-1}(r_{mkt,t})/\text{Max}(E_{t-1}(r_{mkt,t}))]$ Prior information (α)	+	0.287 (0.842)		0.406 (0.770)	1.815 (0.138)
Hulbert's initial sample Reputation (θ)	+	0.999 (0.002)	1.002 (0.002)		1.027 (0.002)
Ex post accuracy Ability (p)	–	–7.372 (0.013)	–6.969 (0.002)	–7.603 (0.009)	
No shorting		0.444 (0.263)	0.448 (0.260)	0.670 (0.092)	0.426 (0.260)
$I(\text{sign}(\Delta w_t) = \text{sign}(r_{mkt,t-1}))$ Momentum-following	+	0.206 (0.566)	0.206 (0.564)	0.219 (0.536)	0.234 (0.508)
Adjusted R^2		9.54%	9.87%	6.08%	6.08%
Number of observations		279	279	279	279

Panel B: Long Horizon (1980–1992)								
Explanatory variable	Expected Sign	Base Case 1b	Alternative Regression 2b	Alternative Regression 3b	Alternative Regression 4b	Alternative Regression 5	Alternative Regression 6	Alternative Regression 7
Intercept		−1.422 (0.003)	−0.828 (0.518)	−1.080 (0.025)	−0.941 (0.001)	−0.032 (0.906)	2.489 (0.999)	3.854 (0.001)
Abs[$E_{t-1}(r_{mkt,t})/\text{Max}(E_{t-1}(r_{mkt,t}))]$ Prior information (α)	+	1.400 (0.001)		1.492 (0.001)	2.804 (0.001)	1.371 (0.001)		1.740 (0.077)
Updated $\hat{\theta}(\cdot)$ Reputation (θ)	+	0.696 (0.001)	0.707 (0.001)		0.616 (0.001)	0.721 (0.001)	0.354 (0.004)	0.610 (0.001)
Ave. ex post accuracy Ability (p)	−	−7.463 (0.001)	−8.101 (0.001)	−7.771 (0.001)		−7.149 (0.001)		−7.401 (0.001)
σ (T-bill forecast) Signal correlation (ρ)	+	2.134 (0.001)	2.067 (0.002)	2.400 (0.001)	0.927 (0.067)			4.439 (0.001)
No shorting		−0.094 (0.406)	−0.083 (0.460)	−0.031 (0.796)	−0.058 (0.613)	−0.092 (0.409)	−0.133 (0.233)	−0.121 (0.299)
I(sign(Δw_t) = sign($r_{mkt,t-1}$)) Momentum-following	+	−0.108 (0.156)	−0.138 (0.063)	−0.110 (0.001)	−0.074 (0.314)	−0.127 (0.091)	−0.407 (0.006)	0.015 (0.853)
$\sigma(r_{mkt,t})$ Market uncertainty								−0.132 (0.885)
Annual dummies		✓	✓	✓	✓	✓		✓
149 year-month dummies							✓	
Adjusted R^2		15.26%	14.92%	13.24%	10.77%	15.27%	27.87%	12.49%
Number of observations		5,293	5,293	5,293	5,293	5,293	5,422	5,076

With one exception, separately dropping the explanatory variables in a series of regressions does not change the inference on the remaining variables (see alternative regressions 2b, 3b, 4b, and 5 in Table I). This implies that misspecification in any one of the proxies (or pairwise collinearity) does not substantially affect the other variables. The one exception is that the momentum-following variable is significant in a few of the specifications, but has a sign opposite that hypothesized.

In an untabulated regression, we allow ability to vary across newsletters in a given month, and find the same strong negative effect of ability on herding. (This supports the implication in Bikhchandani et al. (1992) and Trueman (1994) that analyst-specific ability is negatively related to herding.) At the other extreme, a separate unreported analysis shows that the inference drawn on the other variables does not change when ability is held constant at its sample mean in every observation.

Alternative regression 7 adds a measure of market uncertainty to the base case specification. The estimated coefficient indicates that the incidence of herding behavior is unrelated to market uncertainty. The signs of the estimated parameters for the other variables are unchanged in regression 7 relative to the base case.

Finally, in a separate untabulated regression, we include two variables to measure the strength of prior information. We interact the base case prior information variable with a dummy variable that conditions on whether the leader makes an announcement that is consistent with the prior information. We expect that when *Value Line* announces consistent with (opposite from) the prior, the incidence of herding will increase (decrease) with the strength of the prior, resulting in a positive (negative) coefficient. However, the coefficients of both of the prior information variables are estimated to be significantly positive. Finding that both coefficients are positive is inconsistent with implication (iv) in Proposition 2.

Robustness of the reputational results and link to the existing literature: A number of additional tests are performed to validate the reputational results in the context of the existing literature. First, we consider different definitions of reputation. Unfortunately, information for some candidate measures of newsletter reputation, such as the number of subscribers or the cost of a subscription, is not publicly available.

Following the lead of Lamont (1995), we test whether a newsletter's age is related to its incidence of herding. One could argue that older letters have higher reputations; we find evidence consistent with this hypothesis in the form of a small but significantly positive correlation between age and the dynamic measure of herding.²³ If age does measure reputation, Lamont's evidence that old analysts herd less often than young analysts runs counter to our finding that high reputation analysts herd more. However, there are

²³ There are 150 months in our sample. We take the natural logarithm of age (i.e., log of months in the sample) because we believe that any effect age has on reputation is not such that an age of 150 months leads to a reputation double that for an age of 75 months.

several possible disadvantages to using age to proxy for reputation: (1) “high” age could be associated with decreased reputational incentives because old newsletter editors are closer to retirement and therefore care less about reputation; (2) age can never decrease, even in the face of poor performance, unlike the dynamic measure of reputation; (3) a newsletter entering the sample midway will always have a reputation less than that of a newsletter that enters the sample early. In contrast, the dynamic measure sets a new entrant’s reputation equal to 0.35 but then updates based on relative performance, allowing a new letter’s reputation to rise above an old letter’s. Unreported regression results indicate that there is no statistically significant relation between age and herding in our sample.

Second, we focus on the reputation result by purging the data of all time-series effects. In particular, alternative regression 6 includes a separate dummy variable for all but one year-month in the sample. Blanketing the data with year-month dummies purges the direct effects of prior information, ability, and signal correlation (and any other unmodeled year-month effect) and the remaining coefficients therefore measure purely cross-sectional effects. The sign of the estimated reputation coefficient is the same as reported in the base case specification (see Panel B of Table I).

We also perform analysis using the correlation coefficient between a newsletter’s recommendations and those of *Value Line* as dependent variable (rather than the usual binary herd/deviate dependent variable). The advantage of this approach is that it eliminates all time-series effects from the data and focuses on which letters generally move in the same direction as the leader; the disadvantage is that it cannot distinguish whether a letter herds in some periods but not others. The untabulated results indicate that high reputation newsletters make recommendations that are highly correlated with those of *Value Line*. This provides evidence of the robustness of the positive reputation result found in the main analysis.

The evidence indicates that the incentive to herd increases with analyst reputation. This is consistent with the “farther to fall” hypothesis proposed by Scharfstein and Stein (1990) and Lamont (1995). Our evidence is not supportive of the “tighter priors” hypothesis (suggested in the same papers) that analysts herd less as time passes and their reputation becomes more precisely known.

Finally, to get a feel for the economic importance of the various causes of herding behavior, we compare the adjusted R^2 from the alternative regressions to that for the base case specification. For example, the R^2 in long-horizon regressions drops from 15.26 percent to 10.77 percent when the ability proxy is dropped from the specification. This indicates, in an informal way, that ability explains 29.4 percent of the observed incidence of herding $((15.26 - 10.77)/15.26 = 0.294)$. Using this informal metric, the economic importance of the various causes of herding behavior are ranked as follows, with the percent of explained variation in observed herding behavior shown in parentheses: ability (29.4 percent), reputation (13.2 percent), prior information (2.2 percent), and signal correlation (0 percent). Thus, the precision

of private information signals (i.e., ability) is the most important factor affecting whether a newsletter herds on the advice of the market leader. The importance of ability is consistent with the unanimous predictions from cascades and herding models (e.g., the models in Bikhchandani et al. (1992), Trueman (1994), and the reputational herding model).

III. Conclusion and Discussion

In the past decade there has been an explosion of theoretical and empirical research into the causes and effects of herding behavior (see the introduction for citations). The theoretical papers offer well-crafted models that provide insight into why agents may *rationally* choose to mimic the actions of others. The empirical papers generally investigate whether “too many” agents appear to take the same action, or how tightly individuals’ forecasts are distributed around the consensus forecast. Although the empirical papers document interesting patterns in the data, few are close to the theory. An interesting area for future research is to use the theoretical models to help structure and interpret empirical tests of herding. This may lead to more unambiguous and powerful empirical tests of the causes and effects of herding, while at the same time put the models to the test. We view the main contribution of our paper as its use of relatively direct empirical tests to evaluate the implications from a theoretical herding model.

There are several issues to consider when formulating empirical tests of herding. First, do the theoretical hypotheses make conditional or unconditional predictions about the environment and incentives that cause herding? For example, our model makes a *conditional* prediction about the effect of reputation on herding, depending on the model parameters and the leader’s action. We use the structure of the model to focus on the conditions (i.e., data points and parameter values) for which the expected effect of reputation is unambiguous. Similarly, one may expect the incidence of herding to vary in response to certain events (e.g., increased market volatility or less precise information signals). If one does not pay careful attention to the conditional nature of many herding predictions, one may inadvertently “average away” the signs and symptoms of herding.

Researchers also need to consider the “all else equal” aspect of the predictions made by most herding models. Consequently, it makes sense to test for the causes of herding in a multivariate setting, controlling for the various incentives to herd. In our paper, we attempt to sort out several implications from the model, rather than focusing on just one. Given that it is difficult to empirically measure some of the variables, careful robustness checks are necessary to validate the results. (For example, at one point we purge the data of all time-series effects and focus exclusively on the cross-sectional incentives to herd.) Similarly, it is important to control for potential causes of herding that are outside of the candidate model. For example, if a group

of agents mimic an external event (such as “momentum-following” by investing according to past returns), it may appear that they are copying each other, when really they are not.

It is important to retain as much of the model’s structure as possible when empirically testing its hypotheses. To test our model, we use a dynamic measure of reputation that is constructed with Bayesian updating functions specified in the model. Although this makes our analysis a joint test of the causes of herding and the model, we feel that using the dynamic reputation variable helps keep the tests reasonably close to the model. As another example, we test whether followers mimic a market leader (as implied by the model), rather than substitute the “group mean” in place of the leader.

Finally, and perhaps most fundamentally, researchers must carefully define what they mean by “herding.” The definition of the dependent variable and general structure of the tests may be very different depending on the type of herding one is investigating. For example, “everyone observes and updates efficiently with identical private information” and “agents observe different private signals but some discard their private information and instead mimic the actions of others” are very different events and may require different empirical tests. Unfortunately, it is difficult to use off-the-shelf databases to test deep implications about herding, such as knowing how an agent uses his private information. In this paper, we restate the model’s implications in terms of what we can observe (i.e., act the same as the leader versus act differently). We suggest that researchers use models for guidance whenever possible to define herding, as well as to derive tests and variables, and that they design extensive robustness checks to address the shortcomings of their data and specifications.

Perhaps researchers using an experimental setting will be able to obtain the unique data necessary to sort out the many factors that contribute to herding behavior. An important caveat is in order with respect to experimental research, however. If one were to implement the exact mathematical model derived in this paper in an experimental setting, complete with the assumptions and objectives (such as reputation maximization and the ability to apply Bayes’ rule), one should by construction confirm the implications presented here. If one does not confirm the implications from the model, this would suggest that the precise modeling environment was not implemented. For example, it may be possible that experimental subjects are not capable of implementing Bayes’ rule in the correct fashion. If this is the case, observing or not observing clustering would not necessarily reveal whether agents respond to the hypothesized incentives put forth in the model (although it may document interesting cognitive results).

Much work remains to be done in the area of testing theoretical herding models. Consider these theoretical explanations for observing a group of agents clustered together:

- They all receive the same private information.
- All the agents study the same inputs (as in investigative herding).

- They all herd on some observable event but not explicitly on other agents (e.g., momentum-following investing).
- The aggregate prior information is so strong that even after updating with their private information, the agents all take the same action (as in cascades).
- They choose to mimic the actions of a leader (perhaps for reputational reasons), even though if they used their private information in a traditional application of Bayes' rule, they would deviate from the leader.
- The agents act similarly for behavioral reasons such as under- or over-confidence that, in contrast to the previous items, may not be considered "rational."
- The agents are all part of a speculative bubble.

With the exception of a few recent papers (e.g., Ehrbeck and Waldmann (1996), Golec (1997), Lamont (1995), Welch (1996), this paper), we know very little about which, if any, of these reasons contribute to observed empirical clustering. Nor do we know if "more efficient" decisions would be made if agents faced different incentives. It seems that designing tests that distinguish between the potential causes of herd behavior is a fertile area for future research.

Appendix

This appendix contains proofs of the propositions. In addition to the $\hat{\theta}(\cdot)$ revisions that evaluate an analyst without regard to the other agents' actions, such as those shown in equations (2) and (3) in the text, updating rules analogous to the following are sometimes used to evaluate an analyst relative to the actions of the other analyst:

$$\hat{\theta}^B(\hat{s}_H^A, \hat{s}_L^B, X_L) = \frac{\frac{1}{2}p(1-\theta)\theta + p(1-p)(1-\rho)\theta^2}{\frac{1}{2}p(1-\theta)\theta + \frac{1}{2}(1-p)\theta(1-\theta) + \frac{1}{4}(1-\theta)^2 + p(1-p)(1-\rho)\theta^2}; \quad (\text{A1})$$

$$\hat{\theta}^B(\hat{s}_H^A, \hat{s}_L^B, X_H) = \frac{\frac{1}{2}(1-p)(1-\theta)\theta + p(1-p)(1-\rho)\theta^2}{\frac{1}{2}p(1-\theta)\theta + \frac{1}{2}(1-p)\theta(1-\theta) + \frac{1}{4}(1-\theta)^2 + p(1-p)(1-\rho)\theta^2}. \quad (\text{A2})$$

Unless otherwise stated, assume in this appendix that B observes s_L^B .

Proof of Proposition 1A: If the prior information is so strong that A makes an announcement consistent with the prior, regardless of his private information, then his action is ignored. Given common values of α , p , ρ and θ , if

the follower receives the same private information as the leader, she finds herself in the exact same position and hence takes the same action as the leader.

If prior information is not strong enough to start a cascade with the leader, but instead the leader reliably announces his private information, the follower also announces her private information. This behavior is optimal because the following inequality holds for all parameter values:

$$\begin{aligned} & \hat{\theta}^B(\hat{s}_L^A, \hat{s}_L^B, X_H) \Pr(X_H | \hat{s}_L^A, s_L^B) + \hat{\theta}^B(\hat{s}_L^A, \hat{s}_L^B, X_L) \Pr(X_L | \hat{s}_L^A, s_L^B) \\ & > \hat{\theta}^B(\hat{s}_L^A, \hat{s}_H^B, X_H) \Pr(X_H | \hat{s}_L^A, s_L^B) + \hat{\theta}^B(\hat{s}_L^A, \hat{s}_H^B, X_L) \Pr(X_L | \hat{s}_L^A, s_L^B). \end{aligned} \quad (A3)$$

The left-hand (right-hand) side of this equation is expected reputation when B truthfully announces (announces opposite) her private information. An identical result holds for $\hat{s}_H^A, s_H^B$. This completes the proof of part A.

Proof of Proposition 1B: Assume that the leader observes s_H^A and that the follower observes s_L^B . If the following inequalities hold, the leader announces his private information, and the follower herds (for convenience, these are replications of equations (5) and (6) in the text):

$$\begin{aligned} & \hat{\theta}^A(\hat{s}_H^A, X_H) \Pr(X_H | s_H^A) + \hat{\theta}^A(\hat{s}_H^A, X_L) \Pr(X_L | s_H^A) \\ & > \hat{\theta}^A(\hat{s}_L^A, X_H) \Pr(X_H | s_H^A) + \hat{\theta}^A(\hat{s}_L^A, X_L) \Pr(X_L | s_H^A); \end{aligned} \quad (A4)$$

$$\hat{\theta}^B(\hat{s}_H^A, \hat{s}_L^B, X_L) \Pr(X_L | \hat{s}_H^A, s_L^B) + \hat{\theta}^B(\hat{s}_H^A, \hat{s}_L^B, X_H) \Pr(X_H | \hat{s}_H^A, s_L^B) < \theta. \quad (A5)$$

The left-hand (right-hand) sides of these expressions represent expected reputation when the analysts announce (announce opposite) their private information.

Consider first the optimal behavior of analyst A for $(\alpha, p, \theta, \rho)$ parameter values for which he rationally conjectures that B herds. In this case, A knows that he will be evaluated individually. We can numerically determine the parameter values for which equation (A4) holds, or, by holding one of the variables fixed, obtain analytic solutions. If $\rho = 1$, the α -function for which the left-hand side of equation (A4) equals the right-hand side is $\alpha^{A,1B}(p, \theta) = [1 - (2p - 1)\theta]/2$. Given values for p and θ , this function provides the value of prior information α for which A is indifferent between announcing his private information and announcing opposite his information. The “1B” superscript indicates that the function is associated with Proposition 1B. Agent A announces truthfully when $\alpha > \alpha^{A,1B}(p, \theta)$ (the prior “supports” s_H^A) because the inequality in (A4) holds; A announces opposite his private information when $\alpha < \alpha^{A,1B}(p, \theta)$.

Analogously, we can determine solutions to equation (A5) either numerically or (by holding one of the variables fixed) analytically. If $\rho = 1$, the α -function defining parameter values for which the left-hand side of equa-

tion (A5) equals the right-hand side is $\alpha^{B,1B}(p, \theta) = [2p - \theta - 1]/[4p - 2]$. Given a rational conjecture that analyst A announces truthfully, B herds when $\alpha > \alpha^{B,1B}(p, \theta)$ (the prior does not support s_L^B) because the inequality in (A5) holds. Given these indifference functions, the parameter set that applies to Proposition 1B(ii) can now be described. Proposition 1B(ii) holds for α 's greater than $\alpha^{A,1B}(\cdot)$ and $\alpha^{B,1B}(\cdot)$.

With respect to part (i) of Proposition 1B, if the leader observes s_H^A he becomes more conservative when he anticipates that the follower will not herd (if the follower observes s_L^B). This implies that the leader's optimal action will be reflected by a different α -function, one that anticipates the likelihood that the follower will deviate. Formally, given a rational conjecture that the other analyst announces his or her private information, each analyst will announce his or her private information when the following inequalities hold:

$$\begin{aligned}
 & (\hat{\theta}^A(\hat{s}_H^A, \hat{s}_H^B, X_H) \Pr(X_H | s_H^A) + \hat{\theta}^A(\hat{s}_H^A, \hat{s}_H^B, X_L) \Pr(X_L | s_H^A)) \Pr(s_H^B | s_H^A) \\
 & + (\hat{\theta}^A(\hat{s}_H^A, \hat{s}_L^B, X_H) \Pr(X_H | s_H^A) + \hat{\theta}^A(\hat{s}_H^A, \hat{s}_L^B, X_L) \Pr(X_L | s_H^A)) \Pr(s_L^B | s_H^A) \\
 & > (\hat{\theta}^A(\hat{s}_L^A, \hat{s}_H^B, X_H) \Pr(X_H | s_H^A) + \hat{\theta}^A(\hat{s}_L^A, \hat{s}_H^B, X_L) \Pr(X_L | s_H^A)) \Pr(s_H^B | s_H^A) \\
 & + (\hat{\theta}^A(\hat{s}_L^A, \hat{s}_L^B, X_H) \Pr(X_H | s_H^A) + \hat{\theta}^A(\hat{s}_L^A, \hat{s}_L^B, X_L) \Pr(X_L | s_H^A)) \Pr(s_L^B | s_H^A);
 \end{aligned} \tag{A6}$$

$$\begin{aligned}
 & \hat{\theta}^B(\hat{s}_H^A, \hat{s}_L^B, X_H) \Pr(X_H | \hat{s}_H^A, s_L^B) + \hat{\theta}^B(\hat{s}_H^A, \hat{s}_L^B, X_L) \Pr(X_L | \hat{s}_H^A, s_L^B) \\
 & > \hat{\theta}^B(\hat{s}_H^A, \hat{s}_H^B, X_H) \Pr(X_H | \hat{s}_H^A, s_L^B) + \hat{\theta}^B(\hat{s}_H^A, \hat{s}_H^B, X_L) \Pr(X_L | \hat{s}_H^A, s_L^B).
 \end{aligned} \tag{A7}$$

To account for B's possible actions, A weights between the conditional probabilities that B will announce s_H^B (the top line of either side of (A6)) and s_L^B (the lower line). To aid understanding of how this weighting is accomplished, $\Pr(s_H^B | s_H^A)$ is calculated (from A's perspective) next. Recall that if economic conditions are such that Proposition 1B(i) holds in equilibrium, all the players of the game rationally conjecture that A and B announce their respective private information.

The following two probabilities are weighted to obtain $\Pr(s_H^B | s_H^A)$:

$$\Pr(s_H^B | s_H^A, X_H) = \frac{[p^2(1 - \rho) + p\rho]\theta^2 + p\theta(1 - \theta) + \frac{1}{4}(1 - \theta)^2}{\frac{1}{2}(1 - \theta) + p\theta}; \tag{A8}$$

$$\Pr(s_H^B | s_H^A, X_L) = \frac{[(1 - p)^2(1 - \rho) + (1 - p)\rho]\theta^2 + (1 - p)\theta(1 - \theta) + \frac{1}{4}(1 - \theta)^2}{\frac{1}{2}(1 - \theta) + (1 - p)\theta}. \tag{A9}$$

These equations calculate the probability B will observe (and announce) s_H^B , given that A announced $\hat{s}_H^A$.

The probability of B announcing s_H^B , given $\hat{s}_H^A$ (again, from A's perspective), is obtained by weighting the previous two equations by conditional probabilities of state occurrence:

$$\Pr^A(s_H^B | \hat{s}_H^A) = \Pr(s_H^B | \hat{s}_H^A, X_H) \Pr(X_H | \hat{s}_H^A) + \Pr(s_H^B | \hat{s}_H^A, X_L) \Pr(X_L | \hat{s}_H^A). \quad (A10)$$

The probability that B announces s_L^B can be calculated in a similar manner or by noting that $\Pr(s_L^B | \hat{s}_H^A) = 1 - \Pr(s_H^B | \hat{s}_H^A)$. These probabilities are used to weight the parenthetical expressions on either side of equation (A6).

As before, an "indifference function" $\alpha^{A,1B^*}(\cdot)$ can be derived that defines values of α for which A is indifferent about announcing his private information. With respect to the follower, equation (A7) holds for all parameter values for which equation (A5) is violated. This means that $\alpha^{B,1B}(\cdot)$ still defines the "boundary" between herding and deviating. Analysts A and B both announce their private information when $\alpha^{A,1B^*}(\rho, p, \theta) < \alpha < \alpha^{B,1B}(\rho, p, \theta)$ because equations (A6) and (A7) hold.

Finally, Proposition 1B(iii) holds when α is less than values defined by $\alpha^{A,1B}(\cdot)$, $\alpha^{A,1B^*}(\cdot)$, and $1 - [1 - (2p - 1)\theta]/2$.²⁴ This last condition arises because the follower is evaluated individually if the leader announces opposite his private information in equilibrium. However, the last condition is redundant because indifference α 's determined by this function are always greater than those determined from $\alpha^{A,1B^*}(\cdot)$ and $\alpha^{A,1B}(\cdot)$. Therefore, Proposition 1B(iii) holds if $\alpha < \alpha^{A,1B^*}(\cdot)$ or $\alpha < \alpha^{A,1B}(\cdot)$.

This completes the proof of Proposition 1. Q.E.D.

Proof of Proposition 2: The comparative statics in Proposition 2 can be derived numerically or by taking partial derivatives of the α -functions derived in the previous proof. For example, the partial derivative of $\alpha^{B,1B}(\cdot)$ with respect to θ is $-1/(4p - 2)$, which is always negative. This implies that, all else equal, as θ increases, the α -function produces a smaller indifference value of α . Given that B observes s_L^B and therefore herds when α is greater than the value produced by $\alpha^{B,1B}(\cdot)$, the follower is more likely to herd as θ increases.

Likewise, the partial derivatives of $\alpha^{A,1B}(\cdot)$ with respect to p and θ are both negative. Because the leader truthfully announces his private information when $\alpha > \alpha^{A,1B}(\cdot)$, these derivatives imply that, all else equal, the leader announces his private information more often as p and θ increase.

The effect of prior information on the incentive to herd. In the Scharfstein and Stein (1990) case, prior information is diffuse (because they assume $\alpha = \frac{1}{2}$). When the leader privately observes s_H^A , his revised belief of the prob-

²⁴ The function is derived assuming that $\rho = 1$. We are unable to solve analytically if ρ is allowed to vary, although we can obtain numerical solutions.

ability that the high investment outcome will occur is greater than $\frac{1}{2}$; therefore, he announces $\hat{s}_H^A$. This optimal behavior holds for all Scharfstein and Stein parameter values.

In our model, α can range between zero and one. If α is high (or not much below $\frac{1}{2}$) and the leader observes s_H^A , he believes that the high outcome is likely and truthfully announces his private information. However, if the prior probability of X_H is quite low, the leader's revised probability belief about the high investment outcome occurring will be less than $\frac{1}{2}$ and he will announce $\hat{s}_L^A$, opposite his private information. Consequently, the incentive for the leader to truthfully announce his private information increases with the strength of the prior when it is consistent with his private information, but decreases with the prior when it is inconsistent. (A similar effect of aggregate information can be found in Bikhchandani et al. (1992) and Trueman (1994).) Finally, for extremely high or low α , the leader will always make an announcement that is consistent with the prior, regardless of his private information, and his announcement is therefore ignored by the follower and by investors.²⁵

With respect to the effect of prior information on the follower, again consider the Scharfstein and Stein case. If the follower observes $(\hat{s}_H^A, s_L^B)$, her private information and the leader's announcement effectively cancel, and her revised probability belief about X_H occurring is $\frac{1}{2}$. Without positive correlation among smart analysts' signals, B is indifferent between announcing $\hat{s}_L^B$ and $\hat{s}_H^B$. However, as long as informative signals are positively correlated, the follower has incentive to mimic the leader by announcing $\hat{s}_H^B$, and herding occurs. In our model, because α is not fixed at $\frac{1}{2}$, if the prior information is "strong" and consistent with s_L^B (in this case, well below $\frac{1}{2}$), B may optimally announce her private information. That is, she will not always herd. The precise meaning of "strong" depends on the values of p , θ and ρ .

REFERENCES

- Andrews, Donald, 1991, Heteroscedastic and autocorrelation consistent covariance matrix estimation, *Econometrica* 59, 817–858.
- Banerjee, Abhijit, 1992, A simple model of herd behavior, *Quarterly Journal of Economics* 107, 797–817.
- Banerjee, Abhijit, and Drew Fudenberg, 1995, Word-of-mouth learning, Working paper, Department of Economics, Massachusetts Institute of Technology.
- 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.

²⁵ In our model, the leader does not always mimic prior information as he would in the proposed pure-strategy equilibrium of Brandenburger and Polak (1996). In their model, the actions of analysts are not evaluated relative to the investment outcome; instead, market beliefs are formed after the agent's announcement but before the outcome is realized, and the agent's pay is proportional to the strength of these beliefs. By Bayes' rule, the strength of these beliefs can always be maximized by making an announcement that is consistent with the prior public information, regardless of private information (if the market believes agents are announcing their private information). Thus, agents would always herd on the prior in Brandenburger and Polak's proposed pure strategy equilibrium.

- Black, Fischer, 1973, Yes, Virginia, there is hope: Tests of the Value Line ranking system, *Financial Analysts Journal* 29, 10–14.
- Brandenburger, Adam, and Ben Polak, 1996, When managers cover their posteriors: Making the decisions the market wants to see, *Rand Journal of Economics* 27, 523–541.
- Brennan, Michael, 1990, Latent assets, *Journal of Finance* 45, 709–730.
- Copeland, Thomas, and David Mayers, 1982, The Value Line enigma (1965–1978): A case study of performance evaluation issues, *Journal of Financial Economics* 10, 289–321.
- Devenow, Andrea, and Ivo Welch, 1996, Rational herding in financial markets, *European Economic Review* 40, 603–616.
- Diamond, Douglas, 1989, Reputation acquisition in debt markets, *Journal of Political Economy* 97, 828–862.
- Dow, James, and Gary Gorton, 1994, Arbitrage chains, *Journal of Finance* 49, 819–849.
- Ehrbeck, Tilman, and Robert Waldmann, 1996, Why are professional forecasters biased? Agency versus behavioral explanations, *Quarterly Journal of Economics* 111, 21–40.
- Falkenstein, Eric, 1996, Preferences for stock characteristics as revealed by mutual fund portfolio holdings, *Journal of Finance* 51, 111–135.
- Fama, Eugene, and Kenneth French, 1989, Business conditions and expected returns on stocks and bonds, *Journal of Financial Economics* 25, 23–50.
- Froot, Kenneth, David Scharfstein, and Jeremy Stein, 1992, Herd on the street: Informational inefficiencies in a market with short-term speculation, *Journal of Finance* 47, 1461–1484.
- Golec, Joseph, 1997, Herding on noise: The case of Johnson Redbook's weekly retail sales data, *Journal of Financial and Quantitative Analysis* 32, 367–381.
- Graham, John, 1996, Is a group of economists better than one? Than none?, *Journal of Business* 69, 193–232.
- Graham, John, and Campbell Harvey, 1996, Market timing ability and volatility implied in investment newsletters' asset allocation recommendations, *Journal of Financial Economics* 42, 397–421.
- Graham, John, and Campbell Harvey, 1997, Grading the performance of market timing investment newsletters, *Financial Analysts Journal* 53, 54–66.
- Grinblatt, Mark, Sheridan Titman, and Russ Wermers, 1995, Momentum investment strategies, portfolio performance, and herding: A study of mutual fund behavior, *American Economic Review* 85, 1088–1105.
- Hansen, Lars, 1982, Large sample properties of generalized method of moments estimators, *Econometrica* 50, 1029–1054.
- Harvey, Campbell, 1989, Time-varying conditional covariances in test of asset pricing models, *Journal of Financial Economics* 24, 289–317.
- Hirshleifer, David, 1993, Managerial reputation and corporate investment decisions, *Financial Management* 22, 145–160.
- Hirshleifer, David, Avanidhar Subrahmanyam, and Sheridan Titman, 1994, Security analysis and trading patterns when some investors receive private information before others, *Journal of Finance* 49, 1665–1698.
- Holmstrom, Bengt, 1982, Managerial incentive problems: A dynamic perspective; in *Essays in Economics and Management in Honor of Lars Wahlbeck* (Swedish School of Economics, Helsinki).
- Huddart, Steven, 1996, Reputation and performance fee effects on portfolio choice by investment advisors, Working paper, Duke University.
- Khanna, Naveen, and Steve Slezak, 1998, The effect of organizational form on information flow and decision making: Informational cascades in group decision making, Working paper, University of North Carolina.
- Lakonishok, Josef, Andrei Shleifer, Richard Thaler, and Robert Vishny, 1991, Window dressing by pension fund managers, *American Economic Review, Papers and Proceedings* 81, 226–231.
- Lamont, Owen, 1995, Macroeconomic forecasts and microeconomic forecasters, NBER working paper 5284, Cambridge, MA.
- Lee, In Ho, 1993, On the convergence of informational cascades, *Journal of Economic Theory* 61, 395–411.

- Lewis, Craig, Richard Rogalski, and James Seward, 1997, The information content of Value Line convertible bond rankings, *Journal of Portfolio Management* 24, 42–52.
- Maug, Ernst, and Narayan Naik, 1995, “Herding and delegated portfolio management: The impact of relative performance evaluation on asset allocation,” Working paper, Duke University.
- Metrick, Andrew, 1999, Performance evaluation with transactions data: The stock selection of investment newsletters, *Journal of Finance*, forthcoming.
- Nofsinger, John, and Richard Sias, 1996, Herding by institutional and individual investors, Working paper, Marquette University.
- Parzen, Emmanuel, 1957, On consistent estimates of the spectrum of a stationary time series, *Annals of Mathematical Statistics* 28, 329–348.
- Peles, Nadav, 1993, The determinants of institutional demand for equity: An empirical study, Working paper, Columbia University.
- Prendergast, Constance, and Lars Stole, 1996, Impetuous youngsters and jaded old-timers: Acquiring a reputation for learning, *Journal of Political Economy* 104, 1105–1134.
- Scharfstein, David, and Jeremy Stein, 1990, Herd behavior and investment, *American Economic Review* 80, 465–479.
- Smith, Lones, and Peter Sorensen, 1994, Pathological models of observational learning, Working paper 94-24, Department of Economics, Massachusetts Institute of Technology.
- Stickel, Scott, 1985, The effect of Value Line survey rank changes on common stock prices, *Journal of Financial Economics* 14, 121–144.
- Stickel, Scott, 1990, Predicting individual analyst earnings forecasts, *Journal of Accounting Research* 28, 409–417.
- Stickel, Scott, 1992, Reputation and performance among security analysts, *Journal of Finance* 47, 1811–1836.
- Trueman, Brett, 1994, Analyst forecasts and herding behavior, *The Review of Financial Studies* 7, 97–124.
- Welch, Ivo, 1992, Sequential sales, learning, and cascades, *Journal of Finance* 47, 695–732.
- Welch, Ivo, 1996, Herding among security analysts, Working paper #8-96, University of California, Los Angeles.
- Wermers, Russ, 1999, Mutual fund herding and the impact on stock prices, *Journal of Finance*, forthcoming.
- Womack, Kent, 1996, Do brokerage analysts’ recommendations have investment value?, *Journal of Finance* 51, 137–167.
- Wylie, Samuel, 1996, Tests of the accuracy of measures of herding using UK data, Working paper, London School of Business.
- Zwiebel, Jeffrey, 1995, Corporate conservatism and relative compensation, *Journal of Political Economy* 103, 1–25.