ELSEVIER
Journal of Financial Economics 58 (2000) 369–396
JOURNAL OF Financial ECONOMICS
www.elsevier.com/locate/econbase

Herding among security analysts☆

Ivo Welch*

School of Management at Yale University, 46 Hillhouse Avenue, New Haven CT 06520, USA

Received 14 September 1999; received in revised form 2 November 1999

Abstract

The paper shows that the buy or sell recommendations of security analysts have a significant positive influence on the recommendations of the next two analysts. This influence can be traced to short-lived information in the most recent revisions. In contrast, the influence of the prevailing consensus is *not* stronger if the consensus accurately forecasts subsequent stock price movements. This indicates consensus herding consistent with models in which analysts herd based on little information. The consensus also has a stronger influence when market conditions are favorable. The resulting poorer information aggregation could cause bull markets to be intrinsically more "fragile" (e.g., Bikhchandani et al., J. Political Economy 100(5) (1992) 992–1026). © 2000 Elsevier Science S.A. All rights reserved.

JEL classification: G11; G14; G24

Keywords: Herding; Imitation; Informational cascades; Analysts

1. Introduction

Although much progress has been made at identifying a number of theoretical mechanisms that could cause herding (mutual imitation), it is remarkable that

☆This research has been supported by the Anderson School of Management at UCLA and by a generous grant from the *Milken Foundation.* I thank the editor (Bill Schwert), Antonio Bernardo, Mark Britten-Jones, David Hirshleifer, Olivier Ledoit, Mary Ellen Nichols, and Russ Vulkanov for helpful insights. I especially thank the referee (Campbell Harvey) for an exceptional job in helping me improve the paper. All remaining errors are my own.

* Tel.: + 1-203-4360 777; fax: + 1-203-4360 779.

E-mail address: ivo.welch@yale.edu (I. Welch).

PII: S0304-405X(00)00076-3

most theories of herding have relied primarily on readers' faith in the general phenomenon and on anecdotal evidence. Herding in financial markets, in particular, is often presumed to be pervasive, even though the extant empirical evidence is surprisingly sparse. My paper provides an empirical test for the presence of herding and its determinants in the context in which herding is perhaps most frequently mentioned: the purchase (buy, hold, or sell) recommendations for individual stocks by security analysts. These data are available to both market participants and academic researchers and are of daily but irregular frequency. It is often presumed that it is these public recommendations that help aggregate information, and hence herding could be a prominent mechanism through which information is (or is not) incorporated into security prices.

Unfortunately, working with analysts' purchase recommendations is not easy. One complication is that herding theories suggest that prior analysts' choices are an important influence on the next recommendation. However, from the theory's perspective, the *prevailing consensus*, the *recent revisions*, or simply the *most recent revision*, could all potentially be sufficient statistics for the decision to be made by an analyst. This is because many herding theories are designed to explain a steady state in which all analysts herd perfectly, not to explain an ever-varying time-series of recommendations or a residual heterogeneity in opinion across analysts. In other words, the theories are static, not dynamic, while in reality analysts frequently differ in their recommendations. Therefore, to give "herding" empirical meaning, my paper investigates the influence[1] of the *prevailing consensus* and the *most recent revisions* by other analysts. A second complication is the lack of a parsimonious methodology to handle estimation of herding in *discrete choice* scenarios. In my context, analysts' recommendations come in one of five flavors: strong buy, buy, hold, sell, and strong sell. Theoretically, herding is likely to be intrinsically more relevant if decisions are discrete, i.e., if there is little room for individual decision-makers to tilt their decision using their private information, and to experiment with small changes. Indeed, tests of *informational cascades* (Bikhchandani et al., 1992) *should* focus on discrete rather than on continuous action choice scenarios. Ordinary continuous choice statistical methods, such as naive procedures based on OLS transformations, are therefore largely inapplicable. To handle the discrete choice setting, my paper develops a new statistical methodology to measure the influence of earlier analysts' recommendations on current analysts' choices. In addition to its intuitive ease, the methodology allows me to examine if herding is stronger in particular circumstances.

In brief, I find empirical evidence of a positive influence of the most recent two revisions on the next analyst's revision. This influence is stronger when the

[1] The term "influence" is not meant to imply causality, insofar as no empirical work can prove causality. To make reading of the paper easier, I use the term "influence" rather than "consistent with the presence of an influence" when describing intertemporal correlations.

recent revisions are more recent, and when they turn out to be more accurate predictors of security returns ex-post. Consequently, I argue that this influence could be related to analysts wishing to exploit *fundamental and short-lived information* in these revisions.

I also find that the prevailing consensus has influence on analysts' choices. However, this influence is not significantly stronger when the consensus turns out to be *correct* in its prediction of subsequent stock price movements. This finding suggests that herding towards the consensus is less likely caused by fundamental information, which is consistent with models in which analysts herd based on little or no information (e.g., Scharfstein and Stein, 1990). Furthermore, using the aforementioned ability of my technique to investigate situations in which imitation is stronger, I find that herding towards the consensus is significantly stronger in "good times" (when recent returns were positive and when the consensus is optimistic). Although one should be cautious in drawing conclusions, my paper will argue that up-markets may aggregate less information, and therefore that up-markets could be more "fragile" than down-markets (Bikhchandani et al., 1992).

Although I am able to detect behavior consistent with mutual imitation, my paper does not attempt to discriminate between different theories of herding. For example, there are herding theories based on [1] utility interactions (e.g., Becker (1991), Jones (1984)); [2] sanctions on deviants (e.g., Akerlof, 1980); [3] direct payoff externalities (e.g., Diamond and Dybvig (1983), Arthur (1989)); [4] principal-agent payoff externalities (e.g., Zwiebel (1995), Brennan (1990), Froot et al. (1992), Hirshleifer et al. (1994), Scharfstein and Stein (1990) and Trueman (1994)); [5] informational externalities (e.g., Banerjee (1992); Bikhchandani et al., 1992; Rogers, 1983; Shiller, 1995; Welch, 1992); and [6] irrational agent behavior (e.g., DeLong et al., 1991).[2] Typically these theories show that the incentive to adopt a behavior increases in the number of previous adopters. A discriminating (and more powerful) test would need more detailed information on analysts' decision processes (such as their utility functions, private information, information transmission linkages and communication networks, payoff functions, etc.) than is currently available to me. Moreover, the papers referenced above generally develop "conceptual theories" designed less for direct empirical application than for illustration of the possibility of herding. Consequently, my paper is motivated by these theories in an ad-hoc fashion and does not attempt to test a particular specification.

There are a number of papers that, like my own paper, attempt to empirically identify herding among analysts. In one of the first empirical papers, Graham

[2] Devenow and Welch (1996) survey the literature on rational herding in financial markets. Bikhchandani et al. (1998) provide an intuitive overview, and Bikhchandani et al. (1999) provide an ongoing annotated Web bibliography of the *informational cascades* literature.

(1999) tests if other analysts follow *Value Line* in their market-timing recommendations (the allocation decision between stocks, bonds, and cash). He finds that newsletters he classifies to have both high reputation and low ability are more likely to follow *Value Line*. Analysts seem to be willing to sacrifice some prediction accuracy in order to protect their reputation. Hong et al. (1998) show that older analysts are less likely to herd, both in choice of firms followed and deviation from the earnings consensus. Stickel (1990) finds that past changes in the earnings consensus estimate and the deviation of an analyst's standing recommendation from this consensus are good predictors of revisions in analysts' *earnings* forecasts. In related research, Lakonishok et al. (1992); Wermers (1994); Grinblatt et al. (1995), and Oehler (1995) examine whether mutual funds herd in their purchasing decisions. While the first paper finds little correlated trading activity, the latter three employ different methodologies and find significant correlated trading activity.

The paper now proceeds as follows. Section 2 describes the data set. Section 3 discusses the econometric methodology to detect herding. Section 4 provides evidence on whether various moving "targets" (the prevailing consensus, the most recent revision [by another analyst], and the second-most recent revision) help to explain an analyst's revision, above and beyond an estimated unconditional (constant) transition probability. Section 5 discusses whether the "pulling power" of some of these targets (the prevailing consensus and the most recent two revisions) changes systematically with a number of independent determinants, such as the predictive accuracy of the target and recent stock market conditions. Section 6 summarizes the findings.

2. Data set

This paper uses the Zacks' Historical Recommendations Database, from which *The Wall Street Journal* compiles its regular performance reviews of major brokerage houses. The database covers about 302 thousand individual buy/sell recommendations issued by 226 brokers during the 1989–1994 period. The database itself is very reliable in its revision quotes, but because Zacks relies on printed reports, a revision could be quoted on a date later than that on which the public had access to it. This adds some noise to my empirical results.[3]

[3] Specifically it increases the probability of finding spurious short-term herding in the "last revision" target considered later. According to detailed conversations with Zacks, the majority of recommendations arrive either by fax or electronically within 1–2 days of the issued report. A small percentage arrive by regular mail. Zacks records the recommendation dates printed on the analysts' press release or electronic statement. It is possible that analysts' recommendations are released slightly before or after the date on the report (e.g., the analyst types and dates his/her report the day

Table 1
The 20 brokers with the highest number of recommendations in the Zacks Historical Recommendations Database

Description: The Zacks database is a commercially compiled database of analysts' recommendations. The database covers the period from 1989 to 1994. The total number of analyst recommendations in the database is 302,458. The total number of brokers in the sample is 226.

Interpretation: The number of recommendations seems to correlate with an intuitive notion of issuer quality.

	Firm Identification		Recommendations	
Rank	Abbrev.	Name	Number	Percent
1	VL	Value Line	24,667	8.16
2	SB	Smith Barney	11,496	3.80
3	MS	Morgan Stanley	11,161	3.69
4	SH	Lehman Brothers	10,413	3.44
5	ML	Merrill Lynch	10,140	3.35
6	DL	Donaldson Lufkin	8,960	2.96
7	MH	Paine Webber	8,797	2.91
8	PB	Prudential Sec.	7,795	2.58
9	SA	Salomon Bros	7,472	2.47
10	DW	Dean Witter	7,263	2.40
11	AB	Alex Brown	7,229	2.39
12	FB	CS First Boston	6,170	2.04
13	PS	Kemper Sec	5,949	1.97
14	BR	Bear Sterns	5,367	1.77
15	AG	A.G. Edwards	5,168	1.71
16	MY	Montgomery Sec.	5,036	1.67
17	WH	Wheat First/But	4,849	1.60
18	CR	NatWest Securities	4,471	1.48
19	OP	Oppenheimer	4,279	1.41
20	BD	Baird R.W.	3,996	1.32
		20 Brokers	160,678	53.12

Table 1 lists the 20 most prominent brokerage houses in the database. To avoid legal conflict with individual brokerage houses, Zacks does not permit researchers to publish more detail on individual brokers' or analysts' performance. *Value Line* is the most frequent provider of recommendations, with about 8% of all recommendations. Other frequent contributors are roughly

before or after the actual release). Zacks conducts extensive error checking on their database to assure consistency and accuracy. A competitor of Zacks, First-Call, offers recommendations through an electronic system, which assures dating. However, fewer brokers participate in the First-Call system (about 100 vs. 200 brokers) and buy/sell recommendations are only entered when earnings estimates are changed, making it even more difficult to ascertain proper dating. In addition, electronic communication was probably less important in the early 1990s than it is today.

as-expected, with retail brokerage firms *Smith Barney*, *Morgan Stanley*, *Lehman Brothers*, and *Merrill Lynch* heading the list. The top 20 brokers account for about half of all recommendations. Table 3 computes various analysts' consensus measures. The "weighted" consensus measures weights recommendations according to the frequency of recommendations by the broker in the database, e.g., a *Value Line* recommendation receives about 2.15 times more weight than a recommendation by *Smith Barney*.

The recommendations themselves range from 1 through 5, with 1 denoting a "strong buy," 3 a "hold," and 5 a "strong sell." This scheme is rather unfortunate, because writing about an "upgrade" could be interpreted to mean either a (more bearish) increase in the number or a (more bullish) optimistic upgrade. In my paper, an upgrade shall mean a bullish revision. Each record in the database contains both the previous recommendation and the current revision. When both are cited, the transition from x to y at time/index o is denoted as $\{x \to y\}_o$. A 6 denotes a "no-longer-followed," "not-followed," or "unknown." Revisions from or to a 6 are excluded from the sample. The sample also contains many "affirmation" records in which an analyst (or brokerage house) does not change their recommendation (e.g., $3 \to 3$). These are included in the sample. Finally, it is not meaningful to look for herding in stocks followed only by a single analyst. Therefore, I "bias" my study in favor of finding herding by considering only stocks which have received at least 16 recommendations.[4] The final sample contains about 50 thousand analyst recommendations. The top 10 followed firms in my sample are General Motors (492 observations), Apple (464), Compaq (441), Microsoft (438), Wal Mart (385), Merck (378), Pfizer (374), Intel (368), The Limited (362), and Bristol-Myers (360).

3. Herding estimation

My intent is to estimate the propensity of analysts to follow some moving "target," such as the herd's consensus. To make the discussion simpler, this section refers to the propensity of analysts to follow the consensus (one specific target), but the rest of the study considers multiple possible targets.

[4] An earlier draft excluded affirmations and found substantially similar results. (Ex ante, the inclusion or omission of these observations does not bias the estimated coefficients using the proposed methodology in later sections; see footnote 16.) Furthermore, this draft included only firms that were in the top 30 followed firms in 1989, and ran the estimation over 1990–1994. Again, the results were similar. Finally, the results were also robust when we computed firm-idiosyncratic transition matrices under the null hypothesis (discussed below). For example, analysts could have been more inclined to downgrade *IBM* during the sample period and more inclined to upgrade *Microsoft*, resulting in different transition matrices under the null hypothesis. Because this requires the estimation of a null matrix with 25 parameters for each firm, this is not sensible when firms with as few as 16 recommendations are included.

3.1. Naive estimation of herding

Estimating propensities to herd is not a simple task. With an average consensus recommendation of about 2.5 (between a hold and a buy), it is not too surprising that most "strong sell" (5) recommendations move towards the consensus. After all, recommendations are truncated at 1 (strong buy) and 5 (strong sell). This truncation also induces strong state dependence in the analyst revision process.

Worse, Table 2 shows that the overall transition matrix is highly irregular. One cannot simply assume that revisions, conditional on analysts' previous standing recommendation, are distributed normal, symmetric, or that the five conditional probability vectors (conditional on the previous recommendation) are identical and simply mean shifted. Consequently, one cannot simply use a linear procedure to predict a revised recommendation based on a starting point, such as

$$\textit{revised recommendation}_t = \alpha_0 + \sum_{i=1}^{5} \alpha_i \cdot (\textit{previous own recommendation} = i) + \alpha_6 \cdot (\textit{Prevailing Consensus}_t) + \cdots + \varepsilon_t, \quad (1)$$

where *revised recommendation* is the (discrete) revision to be explained and *previous own recommendation* is the analyst's standing recommendation.

3.2. The influence of targets on transition probabilities

I now develop a statistical procedure to address these problems. This procedure considers herding to be an external force that itself dynamically changes the

Table 2
The transition matrix

Description: See Table 1 for a description of the database. 53,745 observations are used in my study. This sample excludes revisions from or to the "not followed" category (6 in the Zacks database) and firms with fewer than 16 analyst recommendations.

Interpretation: The transition matrix is highly irregular. Transition probabilities are state-dependent.

From ↓	To →	1	2	3	4	5	Total
1	Strong buy	8,190	2,234	4,012	92	154	14,682
2	Buy	2,323	4,539	3,918	262	60	11,102
3	Hold	3,622	3,510	13,043	1,816	749	22,740
4	Sell	115	279	1,826	772	375	3,367
5	Strong sell	115	39	678	345	407	1,584
		14,365	10,601	23,477	3,287	1,745	53,475

transition probability matrix. Thus, a high frequency of revisions from 1 to 4 when the consensus is around 4 is not defined as "herding," while a *higher* probability of moving from 1 to 4 when the consensus is 4, rather than when it is 3 or 5, is defined as "herding." *It is only the time-series variability of the consensus that allows detection of what I call "herding."* In a sense, such estimation of herding is a conservative procedure, equivalent to the conservativeness of running regressions in differences instead of levels. Of course, it is quite possible that the unconditional transition matrix itself reflects consensus-following. But because I cannot decompose movements closer to the consensus into herding and into a mechanistic influence of the observed particular transition matrix, lacking a theory of transitions under the null hypothesis of independent choice (no herding), I cannot attribute simple movement towards the consensus to be herding.

Because there is no clear theory to guide the researcher either about the structure of (the restrictions on) the transition matrix or the precise form of the force of the consensus under the null or alternative hypothesis (i.e., that variation in the consensus influences the transition probabilities), I must define a reasonable form myself. My approach is conservative. I do not constrain the form of the unconditional overall transition matrix under the null hypothesis. It is assumed that the overall empirically observed transitional matrix is generated without the influence of the consensus (the null), and I test whether the alternative hypothesis of an influence of the prevailing consensus on the dynamics of the transition matrix can overcome this assumption.

3.3. A parsimonious specification of changing transition probabilities

First, define a parameter θ that measures whether the analyst recommendation transition matrix is changing with the observed consensus. When θ is 0, the Markov probability transition matrix $\boldsymbol{P}$ should be independent of the prevailing consensus $\boldsymbol{C}$ (the null hypothesis). When θ is positive, it should indicate a tendency of the revision to follow the consensus. When θ is negative, it should indicate a tendency to avoid the consensus. Let $\boldsymbol{P}(\theta, \boldsymbol{C})$ be the 5-by-5 matrix of analysts' revision probabilities, with rows representing the previous recommendation of an analyst and columns representing the newly revised recommendation. Denote by $\boldsymbol{P}(0, \boldsymbol{C}) \equiv \boldsymbol{P}(0)$ the unconditional probability transition matrix under the null. The matrix is normalized so that the sum in each row is 1, i.e., the probability that an analyst moves from his previous recommendation, row i, to a new recommendation, column j, is $p_{i,j}$. To define the transition probability matrix $\boldsymbol{P}$, $p_{i,j}$ is allowed to be a function of θ and the target T (a scalar, e.g., the prevailing consensus $\boldsymbol{C}$):

$$p_{i,j}(\theta, T) \equiv p_{i,j}(0)\left\{\frac{[1 + (j - T)^2]^{-\theta}}{D_i}\right\}, \tag{2}$$

$$D_i = \sum_{j=1}^{5} p_{i,j}(0)[1 + (j - T)^2]^{-\theta}, \tag{3}$$

where the denominator D_i ensures that each row sums to 1. If θ is 0, $\boldsymbol{P}(\theta, T)$ simplifies to $\boldsymbol{P}(0)$ (the unconditional transition matrix, not influenced by the target). If θ is positive, for analyst revisions that are *farther* from the target ($[j - T]^2$ is greater), $p_{i,j}(0)$ is *more deflated.* If θ is infinity, the element in the transition matrix closest to the target is 1 and all other elements in that row are 0. In sum, a positive θ shifts probability mass from the unconditional transition matrix towards a transition matrix with more weight on revised recommendations closer to the target (e.g., the prevailing consensus).[5] Note that each $p_{i,j}(0)$ can be any probability. The only restriction under the null is that each $p_{i,j}(0)$ is constant and does not covary with the consensus target.

To provide intuition on this specification, Fig. 1 provides a graphical illustration of the influence of various θ's on the probability of revisions towards the target. In this figure, the target T is the prevailing consensus, $\boldsymbol{C}$, which happens to be a sell (4) and, given the analysts previously standing own recommendation, the unconditional probabilities under the null of a transition to $\{1, 2, 3, 4, 5\}$ are $\{0.05, 0.20, 0.30, 0.05, 0.40\}$. When θ is zero, the fact that the consensus is "4" is irrelevant, and the probabilities of observing a particular revision are just those under the null. A positive herding parameter θ causes the probability of observing a revision towards the target (of 4) to be higher than that of observing revisions farther away from the target. The figure shows that as θ increases, not only does the probability of observing a revision towards 4 increase, but the probabilities of observing a revision towards 3 or 5 (which are close to 4) also increases. Moreover, the positive herding parameter θ implies that, given the presumed target of 4, we should expect a revision towards a 1 to be less likely.

The figure also illustrates another interesting aspect of the specification: as θ increases beyond 1, the probability of a 3 or a 5 can decrease again. With a sufficiently strong herding parameter θ, the probability that the next revision is

[5] The chosen functional specification is the simplest that occurred to the author. The numerator has to contain a measure of the distance between the target and the chosen destination, thus $(j - T)^2$. θ must not add multiplicatively, or it would not enter the first-order condition in the optimization problem for θ. Adding 1 ensures that the distance is between 1 and 16, rather than 0 and 15. Thus, the factor multiplying $p_{i,j}(0)$ can be exponentiated by $-\theta$, so that a larger θ increases the weight on those row entries that are closer to T. Finally, the denominator normalizes the row probabilities to add back to 1. It should reassure the reader that simple variations in the functional specification (such as taking absolute values rather than squares) seem to have no impact on the empirical results (to three digits after the decimal points). Although the functional specification is arbitrary – it assumes a particular form of "pulling probabilities towards the "target" – this specification could still pick up other forms of "pulling towards the target," but of course less efficiently so.

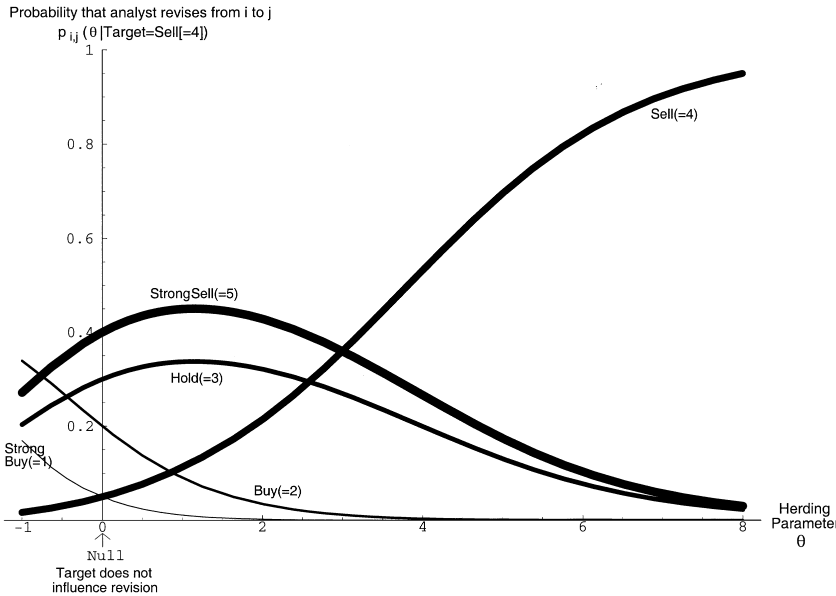

Fig. 1. Probability transformations. The figure plots $p_{i,j}$, i.e., the probability that the analyst will revise his own (given) earlier recommendation (i) towards the recommendation j indicated on each line, as a function of θ (the coefficient of herding). Under the null hypothesis of no influence ($\theta = 0$), these transition probabilities are 0.05 to a "strong buy," 0.20 to a "buy," 0.30 to a "hold," 0.05 to a "sell," and 0.40 to a "strong sell". The target itself is assumed to stand at "sell" = 4. **Interpretation**: When θ is zero, the target has no influence and the probabilities under the null hypothesis apply. When θ is positive, it becomes increasingly likely that the analyst chooses the target itself ("sell" = 4) and increasingly less likely that she chooses the farthest-away recommendation ("strong buy" = 1). The effect of the herding parameter θ on individual transition probabilities is not always monotonic. For example, the probability of a transition to a "strong sell" (5) or "hold" (3) increases until θ is about 1, because a "strong sell" (5) is close to the target of "sell" (4). The probability then decreases again for larger θ's, because more and more weight shifts towards the probability hitting the target *exactly*.

exactly on target becomes so large that it is less likely to just be *near* target. As θ approaches infinity, it becomes a certainty that the target is hit, and the probability of observing any other revisions vanishes.

3.4. The penalty/likelihood function

The specification of θ can apply to an underlying process for analysts' revisions, or can be used to derive an estimator, $\hat{\theta}$, of an underlying true herding parameter, θ. To compute this estimate, I minimize some penalty function defined over $\hat{\theta}$, given the empirically observed transitions $i_o \rightarrow j_o$ (each facing their own target T_o) over all observations $o \in [1, O]$. My chosen penalty function is the negative of the log-likelihood function. The probability of observing the o-th transition to new recommendation j_o, given an analyst's own previous recommendation i_o, given parameter θ and target T_o, is $p_{i,j}$ is

$$p_{i,j}(\theta, T_o)\,|_{\{i_o, j_o\}}, \tag{4}$$

with $p_{i,j}$ specified in (2). The likelihood function over O observations is

$$\prod_{\substack{\text{revisions } o\\ \text{from } i \text{ to } j\\ \{i \to j\}_o}}^{O} p_{i,j}(\theta, T_o)\,|_{i_o,\ j_o}, \tag{5}$$

which is the product of the individual revision probabilities evaluated at their realizations (o). Under the assumption that draws are independent, the product of the realized probabilities is the probability of empirically observing the full sequence of transitions.

In other words, any chosen θ parameter implies, for each observation o, a probability vector p_i, which in turn allows us to compute the "probability loss" attributable to the actual observed revision. Intuitively, the penalty is higher when an actual transition is observed for which $p_{i,j}$ is lower. For example, if the o-th transition is from 1 to 2, the penalty is higher when $p_{1,2}(\theta)$ is 0.5 than when $p_{1,2}(\theta)$ is 1. Taking the log of the likelihood function, the estimator $\hat{\theta}$ is

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}} \sum_{o} \log[p_{i,j}(\theta, T_o)\,|_{\{i_o, j_o\}}]. \tag{6}$$

Statistical inferences can be drawn from a likelihood ratio test (LRT), which computes the likelihood ratio of the probability of the data given a constant transition probability versus a transition probability that varies with T according to θ. Standard asymptotic theory prescribes that

$$-2\left\{\sum_{o} \log[p_{i,j}(0)\,|_{\{i_o, j_o\}}] - \sum_{o} \log[p_{i,j}(\theta, T_o)\,|_{\{i_o, j_o\}}]\right\} \sim \chi^2_1. \tag{7}$$

Unfortunately, as noted in Section 3.2, $p_{i,j}(0)$ is not observable. Therefore, I bias my test against the herding hypothesis by assuming that the $\boldsymbol{P}(0)$ matrix under the null hypothesis is the empirically observed transition matrix. It is of course both correct under the null and it is conservative to assume that the empirically observed probabilities are the probabilities under the null hypothesis.

The maximum-likelihood estimation typically converges quickly and produces asymptotically unbiased estimates (see the appendix). An earlier version of this paper showed (via Monte-Carlo simulations) that the maximum-likelihood ratio has excellent small-sample characteristics with as few as 3,200 observations (used in an earlier draft).[6] With 54,000 observations, not only is there no distinguishable difference between asymptotics and small-sample inference, almost all estimated variables also tend to be either statistically significant at below the 1% level or not significant at all. The reader is thus advised to consider the estimated parameters close to the population parameters and focus on their economic rather than their statistical significance.

3.5. Extension: multiple targets and independent determinants

The main advantage of the proposed procedure is its simplicity, both conceptually and in terms of implementation. The parsimony of my procedure permits easy extensions of the methodology to examine if extra variables (e.g., the optimism of the consensus) influence the tendency to follow the herd, or if there are multiple targets toward which analysts herd (e.g., the consensus and the most recent revision). Extend Eq. (2) to an equivalent specification that offers K targets,[7] and allows each target to have L_k independent determinants. Thus, estimate $\sum_k L_k$ parameters ($\theta_{k,l}$) using realizations of $\sum_k L_k$ independent variables ($\{\boldsymbol{x}_k\}_k$; $\boldsymbol{x}_k$ has length L_k) and realizations of the length-K target vector ($\boldsymbol{T}$) for each revision in the data set. The herding function is now

$$p_{i,j}(\{\theta_{k,l}\}; \boldsymbol{T}, \boldsymbol{x}_k) \equiv p_{i,j}(0) \left\{ \frac{\prod_{k=1}^{K} [1 + (j - T_k)^2]^{-\theta_{T_k}(\boldsymbol{x}_k)}}{D_i} \right\}. \tag{8}$$

$$\theta_{T_k}(\boldsymbol{x}_k) \equiv \sum_{l=1}^{L} \theta_{k,l} x_{k,l}, \tag{9}$$

and D_i normalizes $p_{i,j}(\{\theta\}_{k,l}; \boldsymbol{T}, \boldsymbol{x}_k)$ to ensure that the probabilities add to one for each row. Note that each target's independent determinants could be different. For example, the independent determinant of the "prevailing consensus" target could be a variable that measures how optimistic the prevailing target is, while the independent determinant of the "last revision" target could be a variable that measures the optimism of the last revision. Finally, there is a change in the asymptotic likelihood ratio statistic: it is now computed against an optimally chosen parameter vector that restricts one parameter to zero at a time.

My paper considers two types of targets: the prevailing consensus, and the most recent revisions by other analysts. Unfortunately, it is difficult to make the notation in the tables and text both concise and mnemonic. I continue to denote

[6] An additional advantage of having conducted Monte-Carlo simulations reported in the first version of this paper is that they unambiguously lay to rest the concern that any documented herding tendency could be a mechanistic artifact of the discrete/limited number of options available to analysts or the estimation procedure or the fact that the target itself (the independent variable) is the outcome of prior transitions.

[7] Note that multiple targets do not have a linear influence. In other words, having the first target be a 1 and the second target be a 5 implies a lower – not higher – probability of observing a 3.

the prevailing consensus as $\boldsymbol{C}$. Although it would make sense to denote the most recent revision of another analyst as $j_{t=-1}$ (remember that j is used as index for the revised recommendation) or Revision(− 1), to save subscripts I denote the most recent revision by another analyst as $\boldsymbol{R}(-1)$, the second-most recent revision as $\boldsymbol{R}(-2)$. Further, I refer to the coefficient on the variable x applicable to target k as $\theta_{\underset{x}{k}}$. If x is the constant, I sometimes write θ_k instead of $\theta_{\underset{1}{k}}$.

For example, assume the use of two targets, the prevailing consensus ($\boldsymbol{C}$) and the most recent revision ($\boldsymbol{R}(-1)$). Further assume that the objective is to examine if the tendency to herd towards the consensus is stronger when the consensus is more optimistic and if the tendency to herd towards the last revision is stronger when the last revision is more optimistic. Optimism can be measured as "3 minus the target" (remember that 1 denotes a "strong buy"). The procedure involves estimating four θ parameters on the objective function:

$$p_{i,j}\left(\theta_{\underset{1}{C}}\theta_{\underset{3-C}{C}},\ \theta_{\underset{1}{R(-1)}},\ \theta_{\underset{3-R(-1)}{R(-1)}};\ \boldsymbol{C},\ \boldsymbol{R}(-1),\ (3-\boldsymbol{C}),\ (3-\boldsymbol{R}(-1))\right)$$

$$\propto p_{i,j}\left\{[1+(j-\boldsymbol{C}^2]^{-[\theta_{\underset{1}{C}}+\theta_{\underset{3-C}{C}}\cdot 3-C]}\right.$$

$$\left.[1+(j-\boldsymbol{R}(-1))^2]^{[\theta_{\underset{1}{R(-1)}}+\theta_{\underset{3-R(-1)}{R(-1)}}\cdot(3-\boldsymbol{R}(-1))]}\right\}. \tag{10}$$

where the proportionality factor (D_i) ensures that $\sum_{j=1}^{5} p_{i,j} = 1$ for each originally standing recommendation i.

4. Is there herding? Univariate measures of herding

4.1. *Uni-target estimation*: *Do analysts herd towards the consensus*?

I begin with an examination of the tendency of analysts to herd towards the prevailing consensus. I compute the empirical estimate $\hat{\theta}_C$ according to the specification in Eq. (2). The first specification of the target uses the ordinary prevailing consensus. The second specification weights recommendations in the prevailing consensus by the broker quality[8] (as in Table 1). The third specification uses "time-decayed consensus," in which later observations receive relatively more weight in the consensus computation.[9]

---

[8] Although my broker-quality metric is really a measure of the *quantity* (not *quality*) of recommendations of broker, a casual look at Table 1 confirms intuitively that the two measures are highly correlated.

[9] The current specification decays the weight given to each subsequent recommendation by a factor of $\frac{1}{2}$. If one uses the number of days between revisions to compute the weight factor (instead of an index), the results are again similar.

Table 3
Herding towards the consensus

Description: This table shows $\hat{\theta}_C$, the empirical estimates of the tendency of analysts to herd towards various measures of the prevailing consensus, using a data set of 44,781 analysts revisions from the 1989–1994 period with sufficient data to compute a consensus. A positive $\hat{\theta}_C$ means that analysts have a tendency to move towards the consensus, above and beyond the tendency to move towards the consensus implied by the empirically observed unconditional transition probabilities (the null hypothesis).

Interpretation: The prevailing consensus has a statistically significant positive influence on analysts' revisions. The economic meaning of these coefficients is discussed in Table 4.

Consensus is	Estimate	χ^2_1 significance
Ordinary prevailing	0.130	$< 0.01\%$
Broker-quality weighted	0.140	$< 0.01\%$
Time-decayed	0.164	$< 0.01\%$

Table 3 shows that the estimated herding coefficient,0 $\hat{\theta}_C$, is 0.13 for the ordinary consensus, 0.14 for the broker-quality-weighted consensus, and 0.16 for the time-weighted consensus. With 53,475 observations, the estimated parameters are statistically significant at the 0.01% level.

4.2. Assessing economic significance

To assess the economic significance of these herding coefficients, Table 4 adopts as an "intuitive herding criterion" the probability of hitting the target. The target is as described in the first column, and probabilities are calibrated using the unconditional transition matrix from Table 2. If $\theta = -\infty$, the target will always be avoided. If $\theta = 0$, the probability of hitting the target is equal to the unconditional probability of hitting the target. If $\theta = +\infty$, the target will always be hit.

The table shows that the probability of hitting a hypothetical hold (3) target is about 42% under the null hypothesis. The estimated coefficient of 0.10 to 0.16 raises this probability to about 47%. Similar probability increases are observed for other target levels. Although these are respectable increases in probability, the reader must recognize that one could not have expected to find "overwhelming" evidence of herding: it should come as no surprise that analysts' recommendations are "all over the place," i.e., one rarely finds that all analysts share the same view. Analysts are of course also in the business of disagreeing with other analysts, not just following one another. In addition, with the consensus itself being a function of lagged revisions, there are limits to the standard deviation of the target and thus to my ability

Table 4
Economic significance of θ parameters: probability of hitting the target

Description: This table computes the probability (in percent) that a revision hits the target, varying the herding coefficients (θ). If $\theta = -\infty$, the target will always be avoided. If $\theta = 0$, the probability of hitting the target is equal to the unconditional probability of hitting the target. If $\theta = +\infty$, the target will always be hit. This table was produced using the unconditional transition matrix (from Table 2).

Interpretation: An empirically estimated coefficient of around 0.10 provides a small, but meaningful change in the probability of hitting the target. This probability change is a very conservative measure, because herding could also influence the unconditional transition probability matrix (see discussion in Section 3.2).

	Herding coefficient θ_T									
Target (T)	− 25.00	− 1.00	0.00	0.05	0.10	0.15	0.20	0.50	1.00	25.00
1	0.0	7.0	27.5	28.9	30.4	31.9	33.5	43.0	58.2	100.0
2	0.0	9.4	20.8	21.4	22.1	22.8	23.6	28.1	36.2	100.0
3	0.0	17.1	42.5	44.0	45.4	46.8	48.3	56.5	68.4	100.0
4	0.0	1.3	6.3	6.7	7.1	7.6	8.0	11.1	17.5	100.0
5	0.0	0.3	2.9	3.3	3.6	4.0	4.4	7.5	16.2	100.0

to detect herding. Without variance in the target, no herding could be detected by my procedure in the first place. Second, it must be reemphasized that I conduct the equivalent of a conservative estimation *in differences* (see Section 3.2). I can only detect the influence of *changes* in the target, not the influence of the target itself. Herding is likely to influence the transition matrix under the null hypothesis, but I am unable to measure this absent a theory of transitions under the null. In other words, the unconditional probability of moving to a "hold" could be as high as 42% *primarily because* the consensus itself tends to hover around an optimistic "buy" or "hold" recommendation.

4.3. *Tri-target estimation: Do analysts herd towards the consensus and/or recent revisions?*

I now expand the examination of potential targets towards which an analyst can herd from just the prevailing consensus to include two additional targets: the most recently changed recommendation by another analyst (***R*(− 1)**), and the second-most recently changed recommendation (***R*(− 2)**). To ensure that I do not predict transitions with recent revisions that are not available to market participants at the time, I exclude the (rare) observations in which the most recent revision occurred on the same day. The three herding

Table 5
Univariate tri-target ($\boldsymbol{C}$, $\boldsymbol{R}(-1)$, $\boldsymbol{R}(-2)$) herding

Description: This table examines the tendency of analysts to herd towards three targets: [a] the prevailing consensus ($\mathscr{C}$), [b] the last revision ($\boldsymbol{R}(-1)$), and [c] the second-to-last revision ($\boldsymbol{R}(-2)$). The sample are 44,781 analysts revisions from the 1989–1994 period with sufficient data to compute a consensus. A positive $\hat{\theta}$ means that analysts have a tendency to move towards the target, above and beyond the tendency to move towards the target implied by the empirically observed constant transition probabilities under the null hypothesis.

Interpretation: The consensus and the most recent two analyst revisions have a significant influence on the next analyst revision. (Not reported: the third-most through fifth-most recent revisions are not statistically significant.)

Consensus is	Consensus ($\boldsymbol{C}$) θ_C	Last revision ($\boldsymbol{R}(-1)$) $\theta_{\boldsymbol{R}(-1)}$	2nd-to-last revision ($\boldsymbol{R}(-2)$) $\theta_{\boldsymbol{R}(-2)}$
Ordinary prevailing	0.045	0.087	0.054
Broker-quality weighted	0.066	0.081	0.045
Time-decayed	0.079	0.064	0.041*

*All χ^2_1 significance levels for these estimates are $< 0.01\%$.

parameters ($\hat{\theta}_C$, $\hat{\theta}_{\boldsymbol{R}(-1)}$, $\hat{\theta}_{\boldsymbol{R}(-2)}$) to be estimated are now the three target coefficients in Eq. (8), i.e.:

$$p_{i,j}(\theta_{\mathbf{C}}, \theta_{\boldsymbol{R}(-1)}, \theta_{\boldsymbol{R}(-2)}; \boldsymbol{C}, \boldsymbol{R}(-1), \boldsymbol{R}(-2)) = p_{i,j}(0) \times \left\{ \frac{[1+j-\boldsymbol{C})^2]^{-\theta_{\mathrm{C}}} \cdot [1+(j-\boldsymbol{R}(-1))^2]^{-\theta_{\boldsymbol{R}(-1)}} \cdot [1+(j-\boldsymbol{R}(-2))^2]^{-\theta_{\boldsymbol{R}(-2)}}}{D_i} \right\}. \tag{11}$$

The results of this tri-target herding estimation are in Table 5. The most important target is the most recent revision. The coefficient estimate for $\theta_{\boldsymbol{R}(-1)}$ ranges from 8.7% to 6.4%, depending on the definition of the consensus. In addition, both the second-to-last revision and the consensus continue to have economically and statistically significant magnitudes ranging from about 4% to 8%. (Not reported in the table, any further revisions beyond the two most recent revisions are neither economically nor statistically significant.) Thus, I conclude that analysts are influenced by both the prevailing consensus and the two most recent revisions, and henceforth concentrate on these three targets.

5. When is herding stronger? Multivariate tri-target estimation

The most interesting analysis in this paper concerns systematic variations in the inclinations of analysts to herd. This section examines the determinants thereof.

5.1. Is there more herding when the herd is correct?

Herding could be either irrational or rational. Analysts could either irrationally follow other analysts or information, or they could rationally tend to herd towards a target – such as the consensus $\boldsymbol{C}$ – if this target contained information useful for return prediction. Naturally, such herding could either be the result of analysts independently following the same fundamental information, or the result of direct observation and mimicry of earlier analysts' decision. I am unable to distinguish between these two causes. However, I can measure if information in the target contributes to analysts' decision to herd. If I observe analysts tending to follow the consensus or recent revisions even when this target does not provide useful information, I would relatively discount a rational scenario and upgrade some version of an irrational or little-information scenario. As a proxy for whether following the target would be helpful to an analyst in increasing the accuracy of her recommendation, I introduce a new variable, $\mathscr{T}_t = (3 - T) \cdot r_{0,t}$, where T is the target and $r_{0,t}$ is the stock return[10] from zero to t days after the *to-be-explained* revision. For the consensus, I label this variable $\mathscr{T}\mathscr{C}_t = (3 - \boldsymbol{C}) \cdot r_{0,t}$; for the last two revisions, $\mathscr{T}\mathscr{R}(-1)_t = [3 - \boldsymbol{R}(-1)] \cdot r_{0,t}$ and $\mathscr{T}\mathscr{R}(-2)_t = [3 - \boldsymbol{R}(-2)] \cdot r_{0,t}$.

A positive $\mathscr{T}$ indicates that both the target is optimistic (a strong buy or a buy) and the ex-post return is positive; or that both the target is pessimistic (a strong sell or a sell) and the ex-post return is negative. Conversely, a negative $\mathscr{T}$ indicates that the target mispredicts the subsequent realization. A good mnemonic for $\mathscr{T}$ would be "herding is useful," "good herding," or "target-is-correct."" If it were information that simultaneously led analysts and the herd to buy or sell, then I would expect that *on average* the tendency to move towards a buy (sell) by both the analyst and the herd is higher when the return later on indeed turns out to be higher (lower). That is, I use the ex-post return as a proxy for fundamental ex-ante information, and ask if analysts are more likely to move in the same direction because of this fundamental information.

In sum, my empirical question is "Is there a greater propensity to herd toward a target when this target (the herd) later proves to be correct?" Increased herding when the target later proves to be correct could happen either if fundamental

[10] The results remain the same if the market adjusted return is used instead of the raw return.

information drove many or the most recent analysts to be bullish/bearish (i.e., fundamental news has set the target), or if the target was informative in the sense that following a more optimistic target today led to subsequently higher returns and analysts rationally followed the target, or if the target itself were to drive subsequent stock returns.[11]

Our estimated equation is isomorphic to Eq. (10) estimated over all revisions, each with individual targets and individual ex-post "target-is-correct" data realizations. Intuitively, a positive $\hat{\theta}_{\underset{\mathcal{TC}}{\boldsymbol{C}}}$ indicates a stronger tendency to follow the consensus when the consensus later turns out to be correct.

Table 6 shows some interesting results. The herding coefficient on my proxy for whether the target is "correct" is positive for both recent revisions ($\hat{\theta}_{\underset{\mathcal{TR}(-1)}{\boldsymbol{R}(-1)}}$ and $\hat{\theta}_{\underset{\mathcal{TR}(-2)}{\boldsymbol{R}(-2)}}$) regardless of the ex-post return horizon. These positive coefficients suggest that analysts have a greater propensity to follow the most recent revisions (**R(− 1)** and **R**(− **2**)) if these revisions later on prove to be correct. This supports information-based theories of rational herding. In contrast, the negative coefficient on the consensus target-is-correct variable ($\hat{\theta}_{\underset{\mathcal{TC}}{\boldsymbol{C}}}$) suggests that analysts are more inclined to follow the prevailing consensus ***C*** when it later on turns out to be wrong. The negative coefficient indicates that the prevailing consensus has a "pulling power" that is not easily explained by theories arguing that herding is the outcome of analysts using information contained in the consensus to improve their predictions.[12] Because I do not find informational advantage to following the consensus, I am inclined to interpret this evidence as supportive of theories in which analysts follow the consensus based on very little or no information (e.g., Bikhchandani et al., 1992; DeLong et al., 1991; Scharfstein and Stein, 1990).

[11] At this point, it is worthwhile to mention two peripheral issues. First, on theoretical grounds, it is unclear why information arrival would influence analysts' recommendations sequentially in a systematic way. Security prices could react faster to information than analysts (issued on average about once every one or two weeks), or the analyst's recommendation itself could have a price influence. Second, any information sequentiality is more likely to be picked up by the most recent revision. When news comes out that unidirectionally changes the prefered recommendation of many analysts, the first analyst would alter the forecast, then the second analyst would adjust towards this revision, etc. In a sense, the consensus "discriminates" against simple sequential information releases.

[12] Barber et al. (1998) find that the consensus has significant stock return forecasting power in our Zacks data set. An earlier draft of my paper found that this finding did not extend to the data in an earlier sample of the 30 most followed firms, using a much simpler methodology than that in Barber et al. (1998). Because return performance is not the focus of this paper, this evidence has been omitted.

As to the economic magnitude of these effects, the variables $\mathscr{TR}(-1)$ and $\mathscr{TR}(-2)$ have a standard deviation of about 6% on the 5-day horizon, 10.7% on the 20-day horizon, 15.0% on the 40-day horizon, 18.5% on the 60-day horizon, 26.5% on the 120-day horizon, and 37.0% on the 250-day horizon; the $\mathscr{TC}$ measure has about 60% of the standard deviation of the target-is-correct variables on the same horizon. Thus, a one-standard deviation difference in $\mathscr{TR}(-1)$ and $\mathscr{TR}(-2)$ accounts for an implied difference in the tendency of herding equivalent to a θ of 0.05 to 0.10 – similar in magnitude to the overall tendency to herd. The influence is particularly strong on a short horizon time-span. In contrast, the sign of the $\mathscr{TC}$ coefficient indicates that there is no extra tendency to herd if the consensus later on turns out to be correct, so economic magnitude computations are not called for.

In sum, I conclude that herding based on the two last revisions is consistent with information contained in these revisions that aids analysts in making better predictions, while herding based on the consensus is less likely due to rational informational considerations.

5.2. *Other determinants of herding* (*bivariate estimation*)

There are a number of other possible determinants of herding, and I now examine the potential influence of proxies for these determinants. For the sake of brevity, I limit the discussion to only the most interesting findings:

Time Is the tendency of analysts to herd higher when the target is more up to date? Following the same format as Table 6, I now use the age of each target, ΔT,[13] as independent variables.

Determinants of targets' influence: Age of Target (ΔT)

Target:	Consensus		Last revision		2nd-to-last tevision	
Determinant:	Constant	Age of Target (ΔT)	Constant	Age of Target (ΔT)	Constant	Age of Target (ΔT)
$\hat{\theta}$	– 0.033	0.000242	0.112	– 0.0029	0.065	– 0.00059
χ^2_1 Significance level	0.00%	0.10%	0.00%	0.07%	0.00%	30.26%

[13] This notation is lazy, because there are three different ΔT's (one for each target). To avoid lengthy definitions, which would mostly serve to distract the reader, the in-text discussions in this section clarify whether the independent variables are identical or specific to each target.

Table 6
Multivariate tri-target (**C**, **R(− 1)**,**R(− 2)**) herding – fundamental information?

Description: This table examines the tendency of analysts to herd towards three targets (the prevailing consensus (**C**), and the most recent two revisions (**R(− 1)** and **R(− 2)**) by other analysts), and whether these tendencies are stronger when the herding target later turns out to have been correct. $\mathcal{TC}$ is $(3 - \boldsymbol{C})\cdot r$; $\mathcal{TR}(-1)$ is $(3 - \boldsymbol{R}(-1))\cdot r$; $\mathcal{TR}(-2)$ is $(3 - \boldsymbol{R}(-2))\cdot r$, where r is the ex-post return over the horizon indicated in the left-most column. $3 - T$ equals 2 if T is a "strong buy"; and equals -2 if T is a "strong sell." These variables ($\mathcal{TC}$, $\mathcal{TR}(-1)$, $\mathcal{TR}(-2)$) are positive either if the ex-post return is positive and the target recommendation was a buy or a strong buy, or if the ex-post return is negative and the recommendation was a sell or strong sell). The sample are 44,656 analysts revisions from 1989–1994 with sufficient data to compute both a consensus and appropriate ex-post 5-day returns. (The sample sizes are similar for longer return horizons.)

Interpretation: Information could drive the tendency of analysts to herd towards the most recent revisions: There is evidence that analysts tend to follow the recent revisions more often when they later on turn out to be correct. Thus, herding towards recent revisions could be based on analysts' desire to use fundamental information in these revisions to improve their own recommendations. But there is also evidence (especially on shorter return horizons) that analysts do not tend to follow the prevailing consensus more often when it later on turns out to be correct. This finding suggests herding towards the consensus could be based on little information or irrational behavior.

Target:	Consensus (C)		Last revision (**R(− 1)**)		2nd-to-last revision (**R(− 2)**)	
Ex-post return window	Constant $\hat{\theta}^{C}_{1}$	target-is-correct $\hat{\theta}^{C}_{\mathcal{TC}}$	Constant $\hat{\theta}^{R(-1)}_{1}$	target-is-correct $\hat{\theta}^{R(-1)}_{\mathcal{TR}(-1)}$	Constant $\hat{\theta}^{R(-2)}_{1}$	target-is-correct $\hat{\theta}^{R(-2)}_{\mathcal{TR}(-2)}$
5	0.057	− 0.0079	0.0795	0.0132	0.050	0.0142
	(0.00)	(1.61)	(0.00)	(0.00)	(0.00)	(0.00)
20	0.054	− 0.0039	0.081	0.0056	0.050	0.0061
	(0.00)	(3.73)	(0.00)	(0.00)	(0.00)	(0.00)
40	0.052	− 0.0018	0.082	0.0030	0.050	0.0034
	(0.00)	(19.09)	(0.00)	(0.00)	(0.00)	(0.00)
60	0.052	− 0.0022	0.082	0.0024	0.051	0.0019
	(0.00)	(4.55)	(0.00)	(0.00)	(0.00)	(0.00)
120	0.048	0.0005	0.083	0.0009	0.050	0.0010
	(0.00)	(53.20)	(0.00)	(0.26)	(0.00)	(0.07)
250	0.049	0.0005	0.080	0.0007	0.049	0.0005
	(0.00)	(41.79)	(0.00)	(0.07)	(0.00)	(2.23)

The estimated negative coefficient on the age of the most recent revision ($\hat{\theta}_{\underset{\Delta T}{R(-1)}}$) suggests that herding towards the most recent revision is stronger when the most recent revision is "fresh." This complements my earlier finding that $\mathcal{TR}(-1)$ has the strongest influence at the short horizon – the information that induces herding is short-lived. There is an average of 8.6 days to the previous revision (standard deviation of 9.1 days) and an average of 16.4 days (standard deviation of 13.7 days) to the second-to-last revision. (Table 7 lists the mean and standard deviations of all variables discussed in this section in the right hand columns.) A one-standard deviation difference in the most recent revision's age can thus account for a difference of about 0.025 in the herding tendency (θ). In contrast, $\hat{\theta}_{\underset{\Delta T}{R(-2)}}$ fails to indicate that herding towards the second-to-last revision is stronger when it is fresher. Furthermore, $\hat{\theta}_{\underset{\Delta T}{C}}$ indicates that herding towards the consensus is stronger when the recommendation-weighted consensus is older, not younger.

Market Conditions Is the tendency of analysts to herd greater when analysts are more optimistic? As before, I measure optimism as $3 - T$, where T is the target.

Determinants of targets' influence: optimism

Target:	Consensus		Last revision		2nd-to-last revision	
Determinant:	Constant	Optimism	Constant	Optimism	Constant	Optimism
$\hat{\theta}$	– 0.027	0.140	0.085	0.0031	0.054	0.0014
χ^2_1 Significance level	0.00%	0.00%	0.00%	70.0%	0.00%	90.05%

The positive coefficient on optimism of the consensus ($\hat{\theta}_{\underset{3-C}{C}}$) suggests that a more optimistic consensus has significantly more pulling power. With optimism having a standard deviation of about 0.34, the 0.14 coefficient indicates a differential herding coefficient of about 0.047 for a one-standard deviation difference in attitude. In contrast, the influence of the most recent revisions do not change when $\boldsymbol{R}(-1)$ or $\boldsymbol{R}(-2)$ are more or less optimistic.

It is also interesting to ask if the tendency of analysts to herd is greater when recent past returns have been higher. For the 60-day return horizon, the implied coefficients are:

Determinants of targets' influence: past 60-day returns

Target:	Consensus		Last revision		2nd-to-last revision	
Determinant:	Constant	Past 60-day returns	Constant	Past 60-day returns	Constant	Past 60-day returns
$\hat{\theta}$	0.034	0.0043	0.092	− 0.0019	0.058	− 0.0012
χ^2_1 Significance level	0.00%	0.00%	0.00%	0.07%	0.00%	2.71%

The past return coefficient for the consensus ($\hat{\theta}^{C}_{\gamma-60,-1,}$) of 0.0043 suggests that recent positive returns (like general optimism in the consensus) increase the pulling power of the consensus. The standard deviation of the past return is about 15.6%, indicating an implied influence of a one-standard deviation difference in 60-day return on the consensus' influence of about 0.067. In contrast, the past return coefficient for the most recent revision ($\hat{\theta}^{R(-1)}_{\gamma-60,-1}$) suggests that recent positive returns reduce the pulling power of the most recent revision. The coefficient suggests that analysts tend to follow the consensus in up-markets and recent revisions in down-markets. In sum, the picture that emerges suggests that consensus herding among analysts occurs primarily when market conditions are favorable. Clearly, analysts are a source of information for the market as a whole. If the same anomaly were to apply to market participants in general, i.e., that consensus herding is more prevalent in up-markets than in down-markets, it would indicate that there is less independent information in the market (poor information aggregation) when conditions are bullish. "News" could thus have a more dramatic impact in optimistic markets than in pessimistic markets. Such a consequence of poorer information aggregation is called "fragility" in Bikhchandani et al. (1992), and it could cause more fickle markets in times when markets have been bullish. This is broadly consistent with a notion of markets – not only for individual stocks but also in the aggregate – that crashes are more frequent than frenzies. The herding theories themselves, however, offer no explanation as to why there is more herding in bull markets than in bear markets.[14]

[14] The related hypothesis of whether there is a tendency for analysts to herd more when *subsequent* returns are positive – and the question of whether this could have driven the $\mathscr{T}$ variables, of which subsequent returns are a component – can be dismissed. Subsequent returns have no significant influence on the tendency of analysts to herd.

Disagreement Is the tendency to herd greater when the cross-sectional standard deviation in the consensus ($\sigma_{\mathbf{C}}$) is lower?

Determinants of targets' influence: consensus std. dev.

Target:	Consensus		Last revision		2nd-to-last revision	
Determinant:	Constant	Consensus std.dev.	Constant	Consensus std.dev.	Constant	Consensus std.dev.
$\hat{\theta}$	0.413	– 0.367	0.027	0.056	– 0.032	0.083
χ_1^2 Significance level	0.00%	0.00%	47.61%	7.38%	38.90%	1.58%

The significantly negative coefficient on the consensus standard deviation for the consensus ($\hat{\theta}_{\mathbf{C}_{\sigma\mathbf{C}}}$) suggests that the prevailing consensus has more pulling power when there is more agreement among analysts. As expected, the heterogeneity in the prevailing consensus does not appear to influence the pulling power of the most recent revisions.

Broker Quality Do high-quality brokers tend to herd less? The independent variable is now the measure of broker quality (BQ) for the revision to be predicted, with BQ measured as the percentage tabulated in Table 1.

Determinant's of targets' influence: Broker quality

Target:	Consensus		Last revision		2nd-to-last revision	
Determinant:	Constant	Broker quality	Constant	Broker quality	Constant	Broker quality
$\hat{\theta}$	0.057	– 0.0058	0.094	– 0.003	0.045	0.0043
χ_1^2 Significance level	0.00%	35.17%	0.00%	51.21%	0.05%	34.32%

I conclude that the quality of the revising broker is not an important determinant of the tendency to herd. However, this result is not robust. An earlier draft found a positive coefficient on the tendency to follow the consensus when broker quality was high for the 30 most-followed stocks.

5.3. Multivariate tri-target estimation

Because it is unclear whether some of the independent variables only proxy the influence of other variables, I also employ a grand simultaneous estimation with all three targets (**C**, **R**(− **1**), and **R**(− **2**)), each with all the aforementioned independent determinants, plus the differences between the targets. The results are in Table 7. The table is easy to summarize: none of the significant variables

Table 7
Multivariate tri-target (**C**, **R**(− **1**), **R**(− **2**)) herding (all variables)

Description: This table examines the determinants of the tendencies of analysts to herd towards three targets (the prevailing consensus (**C**), and the two most recent revisions (**R**(− **1**) and **R**(− **2**)) by other analysts). The determinants are identical to those discussed in the text for the bivariate cases. The maximum-likelihood optimization is conducted over 44,641 analysts revisions from 1989 to 1994 with sufficient data to compute all variables. Boldfaced variables are statistically significant at the 1% level.

Interpretation: There are *no* important differences between the tri-target bivariate-determinant estimations discussed earlier and the multivariate estimations displayed here. The coefficients are remarkably independent of one another.

Independent variable	$\hat{\theta}$	χ^2_1 Significance	μ	σ
Target is consensus **C**				
Constant	0.237	< 0.01%	1	0
Target-is-correct ($\mathscr{T}$) (60-day ex-post return times (3-Target))	− 0.0009	60.10%	1.16	10.28
60-day ex-ante return	0.0038	< 0.01%	2.30	15.58
Optimism (3-target)	0.108	0.05%	0.61	0.33
60-day ex-post return	0.0006	55.77%	1.91	15.18
Broker quality	− 0.0041	52.39%	2.19	1.91
Target age	0.00027	0.03%	326.05	137.59
Consensus variance	− 0.385	< 0.01%	0.96	0.25
Absolute value of consensus - last rev.	0.054	1.19%	0.82	0.55
Target is last revision (**R**(− **1**))				
Constant	0.083	4.70%	1	0
60-day ex-post return	− 0.0019	0.38%	1.91	15.18
Target-is-correct ($\mathscr{T}$) (60-day ex-post return times (3-Target))	0.0027	< 0.01%	1.41	18.69
60-day ex-ante return	− 0.0018	0.10%	2.30	15.58
Optimism (3-target)	− 0.0014	87.83%	0.63	1.05
Broker quality	− 0.0021	63.41%	2.19	1.91
Target age	− 0.0026	0.20%	8.59	9.14
Consensus variance	0.055	9.49%	0.96	0.25
Absolute value of consensus - last rev.	− 0.0089	62.83%	0.82	0.55

Table 7 (*continued*)

Independent variable	$\hat{\theta}$	χ^2_1 Significance	μ	σ
Target is 2nd-to-last revision (***R*(− 2)**)				
Constant	− 0.0227	55.77%	1	0
60-day ex-post return	− 0.0020	0.33%	1.91	15.18
Target-is-correct ($\mathscr{T}$)				
(60-day ex-post return times (3-target))	0.0023	< 0.01%	1.31	18.45
60-day ex-ante return	− 0.0011	4.28%	2.30	15.58
Optimism (3-target)	0.0032	72.37%	0.63	1.04
Broker quality	0.0038	41.51%	2.19	1.91
Target age	− 0.0006	32.89%	16.42	13.75
Consensus variance	0.0717	3.01%	0.96	0.25
Absolute value of consensus - last rev.	− 0.0039	84.33%	0.82	0.54
Absolute value of last rev - 2nd-to-last rev.	0.0056	67.84%	1.03	0.93

switch sign when competing with other determinants. The table also introduces two additional explanatory variables: the difference between the target and the consensus, and (for the second-to-last revision) the difference between the two previous recommendations. Neither is significant.

6. Conclusions

This paper has produced the following primary findings about analyst herding:

1. An analyst's recommendation revision has a positive influence on the next two analysts' revisions.
2. The influence of these most recent revisions can be traced to *short-lived information*. The influence is stronger when short-run ex-post stock returns are accurately predicted by the revision and when the most recent revision has occurred more recently. Lacking access to the underlying information flow, I cannot discern if the influence of recent revisions is either a similar response by multiple analysts to the same underlying information or is caused by direct mutual imitation.
3. The prevailing consensus has a positive influence on the recommendation revisions of analysts.
4. The influence of the prevailing consensus is not stronger when it is a good predictor of subsequent stock returns. Thus, the influence of the consensus is probably less related to attempts by analysts to uncover fundamental

information which can help improve their own recommendations. This favors theories in which analysts' consensus herding is not the result of rational and efficient information aggregation.
5. The influence of the prevailing consensus is stronger when recent market conditions have been bullish. This finding suggests that information aggregation is relatively poorer in up-markets, perhaps causing more fragility and thus a higher incidence of "crashes" in up-markets than of "frenzies" in down-markets.

My paper is merely a first step in this line of research and two possible future improvements immediately come to mind: First, if there was a theory of the decision processes under the null hypothesis of "no herding," it would be much easier to detect the alternative. Second, if a researcher had access to more "micro" information, such as the specific information transmission mechanisms and information sets available to individual analysts when making decisions, one could discriminate between true herding and information that is simply received simultaneously and interpreted likewise by two analysts. These much more demanding tasks are left to future research.

Appendix A

A.1. Asymptotic unbiasedness

This appendix outlines why the log penalty function leads to asymptotically unbiased estimates of θ.[15] Consider estimating a vector of N transition probabilities, $\hat{p}_j$, where the transition is to a state j ($j \in [1, N]$) from an arbitrary given state i. Assume transitions to state j truly occur with probability p_j. Consequently, asymptotically, the penalty function will minimize the fraction p_j of incidences of $\log(\hat{p}_{\hat{j}})$, and thus the procedure would solve

$$\min_{\{\hat{p}_1, \hat{p}_2, \dots, \hat{p}_N\}} \sum_{j=1}^{N} p_j \log(\hat{p}_j), \tag{A.1}$$

$$\text{subject to} \sum_{j=1}^{N} p_j = 1, \quad \sum_{j=1}^{N} \hat{p}_j = 1. \tag{A.2}$$

After substituting the two conditions for the N-th probabilities, the j-th partial derivative of the objective function with respect to $\hat{p}_j$ is

$$\frac{\partial}{\partial \hat{p}_j}: \frac{p_j}{\hat{p}_j} - \frac{1 - \sum_{i=1}^{N-1} p_i}{1 - \sum_{i=1}^{N-1} \hat{p}_i}. \tag{A.3}$$

[15] It is straightforward to show that a sum-squared penalty function on $(1 - p_{i,j})$ leads to asymptotic coefficient bias.

Setting this to zero, and simplifying, I obtain

$$p_j\left(1-\sum_{i=1}^{N-1}\hat{p}_i\right)=\hat{p}_j\left(1-\sum_{i=1}^{N-1}p_i\right). \tag{A.4}$$

Symmetry among $N-1$ such conditions reveals that $\hat{p}_j = p_j$, for all j, is the solution to this system of equations.

Now allow only one parameter θ to be estimated by $\hat{\theta}$, but substitute the correct probabilities, p_j, for $\hat{p}_j$ in such a way that when $\theta = 0$ the first-best solution obtains:

$$\min_{\hat{\theta}} \sum_{j=1}^{N} p_j \log[p_j g_j(\hat{\theta}, T)], \tag{A.5}$$

$$\text{subject to} \sum_{j=1}^{N} p_j = 1, \quad \sum_{j=1}^{N} \hat{p}_j = 1, \tag{A.6}$$

and $g_j(\theta, T)$ is a function that depends on the state j, the target T, and the parameter θ ($g_j(\theta, T)$ is identically 1 iff $\theta = 0$). If the transition vector is truly constant, so that the target has no influence, the minimization procedure can achieve the first-best minimum by choosing only one parameter $\hat{\theta} = 0$ instead of choosing all $N-1$ probabilities. A similar argument (and unreported Monte-Carlo simulations) confirm that $\hat{\theta}$ is not only unbiased under the null hypothesis, but also unbiased under the alternative hypothesis (p_j needs to be replaced by $p_j g_j(\theta, T)$).[16]

References

Akerlof, G.A., 1980. A theory of social custom, of which unemployment may be one consequence. Quarterly Journal of Economics 94, 749–775.

Arthur, B.W., 1989. Competing technologies, increasing returns, and lockin by historical events. Economic Journal 99, 116–131.

Banerjee, A., 1992. A simple model of herd behavior. Quarterly Journal of Economics 107 (3), 797–818.

Barber, B., Lehavy, R., McNichols, M., Trueman, B., 1998. Can investors profit from the prophets? Consensus analyst recommendations and stock returns. Unpublished working paper. Stanford University.

[16] The omission of the diagonal elements would not bias the estimated coefficients. By omitting a row's diagonal probability d, the optimization is on

$$\min_{\{\hat{p}_1,\hat{p}_2,\ldots,\hat{p}_M\}} \sum_{j=1}^{M} \frac{p_j}{d} \log\left(\frac{\hat{p}_j}{d}\right) = \frac{1}{d} \cdot \sum_{j=1}^{M} [p_j \log(\hat{p}_j) - \log(d)], \tag{A.7}$$

where $M = N-1$, and the estimated coefficients are still unbiased. An earlier draft did indeed run an analysis without recommendation affirmations and found substantially similar results.

Becker, G.S., 1991. A note on restaurant pricing and other examples of social influences on price. Journal of Political Economy 99, 1109–1116.

Bikhchandani, S., Hirshleifer, D., Welch, I., 1992. A theory of fads, fashion, custom, and cultural change as informational cascades. Journal of Political Economy 100 (5), 992–1026.

Bikhchandani, S., Hirshleifer, D., Welch, I., 1998. Learning from the behavior of others: conformity, fads, and informational cascades. Journal of Economic Perspectives 12 (3), 151–170.

Bikhchandani, S., Hirshleifer, D., Welch, I., 1999. Informational cascades and rational herding: an annotated bibliography and resource reference. http://linux.agsm.ucla.edu/cascades/p.n/a.

Brennan, M.J., 1990. Latent assets. Journal of Finance 45 (3), 709–730.

DeLong, B.J., Shleifer, A., Summers, L.H., Waldmann, R.J., 1991. The survival of noise traders in financial markets. Journal of Business 64 (1), 1–20.

Devenow, A., Welch, I., 1996. Rational herding in financial economics. European Economic Review 40 (3–5), 603–616.

Diamond, D.W., Dybvig, P.H., 1983. Bank runs, deposit insurance, and liquidity. Journal of Political Economy 91, 401–419.

Froot, K.A., Scharfstein, D.S., Stein, J.C., 1992. Herd on the street: informational inefficiencies in a market with short-term speculation. Journal of Finance 47 (4), 1461–1484.

Graham, J.R., 1999. Herding among investment newsletters: theory and evidence. Journal of Finance 54 (1), 237–268.

Grinblatt, M., Titman, S., Wermers, R., 1995. Momentum investment strategies, portfolio performance, and herding: a study of mutual fund behavior. American Economic Review 85 (5), 1088–1105.

Hirshleifer, D., Subrahmanyam, A., Titman, S., 1994. Security analysis and trading patterns when some investors receive information before others. Journal of Finance 49 (5), 1665–1698.

Hong, H., Kubik, J.D., Salomon, A., 1998. Security analysts' career concerns and herding of earnings forecasts. Unpublished working paper, Stanford University, Syracuse University, Salomon-Smith-Barney.

Jones, S.R.G., 1984. The Economics of Conformism. Blackwell, Oxford.

Lakonishok, J., Shleifer, A., Vishny, R.W., 1992. The impact of institutional trading on stock prices. Journal of Financial Economics 32 (1), 23–44.

Oehler, A., 1995. Do institutional investors herd? Unpublished working paper. Universität Bamberg.

Oehler A (1998): "Do mutual funds specializing in German stocks herd?" Finanzmarkt und Portfolio Management, no. 4, pp 452–465.

Rogers, E.M., 1983. Diffusion of Innovations. Free Press, New York.

Scharfstein, D.S., Stein, J.C., 1990. Herd behavior and investment. American Economic Review 80 (3), 465–479.

Shiller, R.J., 1995. Conversation, information, and herd behavior. American Economic Review 85, 181–185.

Stickel, S.E., 1990. Predicting individual analyst earnings forecasts. Journal of Accounting Research 28 (2), 409–417.

Trueman, B., 1994. Analyst forecasts and herding behavior. Review of Financial Studies 7 (1), 97–124.

Welch, I., 1992. Sequential sales, learning, and cascades. Journal of Finance 47 (2), 695–732.

Wermers, R., 1994. Mutual fund herding and the impact on stock prices. Journal of Finance 2, 581–622.

Zwiebel, J., 1995. Corporate conservatism and relative compensation. Journal of Political Economy 103 (1), 1–25.