Investor learning about analyst predictive ability ☆

Qi Chen[a], Jennifer Francis[a], Wei Jiang[b]

[a] *Fuqua School of Business, Duke University, Durham, NC 27708, USA*
[b] *Graduate School of Business, Columbia University, New York, NY 10027, USA*

Received 8 January 2003; received in revised form 18 December 2003; accepted 22 January 2004

Abstract

Bayesian learning implies decreasing weights on prior beliefs and increasing weights on the accuracy of the analyst's past forecast record, as the number of forecast errors comprising her forecast record (its length) increases. Consistent with this model of investor learning, empirical tests show that investors' reactions to forecast news are increasing in the product of the accuracy and length of analysts' forecast records. Moreover, the Bayesian learning predicted by our model is more descriptive of investor reactions than is a static model which predicts that investors' responses condition only on the prior accuracy of the analyst.

1. Introduction

We develop and test a model of investor learning about the predictive ability of security analysts. The model shows that when investors rely on an analyst's record of

☆We appreciate the comments of Kin Lo (the referee) and Thomas Lys (the editor). We have also benefited from discussions with Alon Brav, David Hsieh, Gur Huberman, Cade Massey, Katherine Schipper, and Vish Viswanathan. We also thank seminar participants at Columbia University, Duke University, Humboldt University at Berlin, INSEAD, University of Minnesota, University of Technology, Sydney, Tulane University, the University of Washington and the Midwestern Finance Association Annual Meeting. This research was supported by the Fuqua School of Business, Duke University and Columbia Business School, Columbia University.

Published in: ***Journal of Accounting and Economics***

past forecasting performance to update their beliefs about the analyst's true predictive ability, they [investors] will increase their weights on the analyst's past performance and decrease their weights on prior perceived ability. Intuitively, as an analyst accumulates a record of forecasting performance—that is, as more forecast errors are revealed—estimating her true ability from her past performance becomes more precise. We develop a model in which investors learn about an analyst's predictive ability in this Bayesian manner. In our model, investors update their beliefs about the predictive accuracy of the analyst from forecast errors which become known when earnings are announced; this learning about predictive ability is then manifest in investors' reactions to subsequent forecasts made by the analyst. Specifically, we show that Bayesian learning predicts that the forecast response coefficient (relating the stock price response to the earnings news in the analyst's forecast) is increasing in the product of the accuracy and number of forecast errors comprising her past performance. Hereafter, we refer to the analyst's series of past forecast errors as her forecast record.

We test the model's predictions using a sample of over 73,000 quarterly earnings forecasts made over January 1, 1990–June 30, 2000. Our results are consistent with investors placing greater weight on the accuracy of the analyst's forecast record as the length of that record increases. We note that this result is distinct from prior researchers' finding that investors attach higher weights to forecasts or stock recommendations issued by superior analysts (e.g., Stickel, 1992; Park and Stice, 2000; Mikhail et al., 2004). We refer to this type of learning as ''static'' because it assumes that investors do not consider the length of the forecast record, only the average accuracy of the forecast errors comprising that record. Consistent with static learning, we find significantly larger weights on forecast revisions made by analysts with more accurate prior forecasts. This result disappears, however, when we include our measure of Bayesian learning in the regression. Nested model tests reject static learning in favor of Bayesian learning as more descriptive of market reactions to analysts' forecasts. Together, our results suggest that investors consider not only the accuracy of an analyst measured at a point, but also the history of her forecast errors up to that time.

Our findings contribute to several literatures. First, the results support an important assumption in theoretical studies of analyst behavior; notably, that long-term career concerns, such as reputation, are important incentive mechanisms affecting analysts' behaviors (Trueman, 1994; Ehrbeck and Waldmann, 1996; Avery and Chevalier, 1999; Holmstrom, 1999).[1] By showing that investors rationally process information about analysts' revealed ability, our results provide empirical support for why analysts care about their reputations and how they build them. More generally, the model and tests can be generalized to any setting where market participants learn about an agent's true ability from repeated realizations of

[1] For example, research shows that analysts' earnings forecast accuracy affects their job turnover rates (Mikhail et al., 1999) and their career opportunities (Hong and Kubik, 2003). Forecast accuracy is also an important criteria used by *Institutional Investor* to award *All-American Analyst* honors (Stickel, 1992).

indicators which are informative about the agent's type.[2] For example, it is applicable to a setting where investors form assessments about the credibility of management from the accuracy of managers' disclosures, or where investment dollars flow to/from mutual funds based on the past performance. We focus on the analyst setting because it provides both a large cross-section (i.e., many analysts) and a long time-series of observations (i.e., many forecasts and many forecast errors), and therefore, offers a more powerful setting to detect learning than one characterized by fewer agents and less frequent realizations.

Second, our results contribute to the debate on investor rationality by documenting, in an empirical-archival setting, aggregate-market behavior that is consistent with learning. The fact that we find the Bayesian model is more descriptive of investors' learning behavior than a static model provides additional evidence that investors are able to process information rationally and in a sophisticated manner. Our results, therefore, support the view that investors are able to use historical data to infer underlying parameters in a rational way. This assumption underlies the argument that investor learning about parameter uncertainty explains some of the predictability of stock returns (Timmermann, 1993, 1996; Brennan, 1998; Lewellen and Shanken, 2002).

The rest of the paper is organized as follows. In Section 2, we present a model of how investors learn about analysts' predictive ability and how the learning process is manifest in investors' reactions to analysts' earnings forecasts. Section 3 tests the predictions of the model; here we describe the sample selection procedures, detail the measurement of the variables, and present the results. Section 4 assesses the sensitivity of the results and considers alternative explanations. Section 5 summarizes the results and concludes.

2. Model

In our model, investors rationally update their beliefs about an analyst's predictive ability according to Bayes rule.[3] (Fig. 1 depicts the events in our model, with

[2] The learning mechanism we document is similar to Lang (1991), Antonji and Pierret (2001), and Krueger and Fortson (2003). Lang argues that, as more earnings observations are revealed, the level of uncertainty about the persistence parameter underlying the earnings process decreases; hence, investors place less reliance on new earning surprises in drawing inferences about the persistence parameter. Antonji and Pierret show that as employers learn about an employee's productivity, they attach decreasing weights to prior beliefs about that productivity (such as those based on easily observable characteristics, like education) and attach increasing weights to less observable measures of productivity revealed by the employee over time. Krueger and Fortson document a significant increase in the sensitivity of market interest rates to unemployment rate estimates released by the Bureau of Labor Statistics (BLS) after the BLS significantly increased the size of the sample used to determine those estimates.

[3] Although we model investors as Bayesian, some non-Bayesian models of learning also predict decreasing weights on prior beliefs and increasing weights on forecast records. For example, conservative learning (where investors place too much weight on their priors) also generates these comparative static results. (See Brav and Heaton, 2002, for a discussion on the implications of conservative learning in financial markets.) Our results, therefore, can be viewed as providing evidence about any learning model which predicts that, over time, investors shift weight to the forecast record as the latter becomes a more precise estimate of the analyst's true ability.

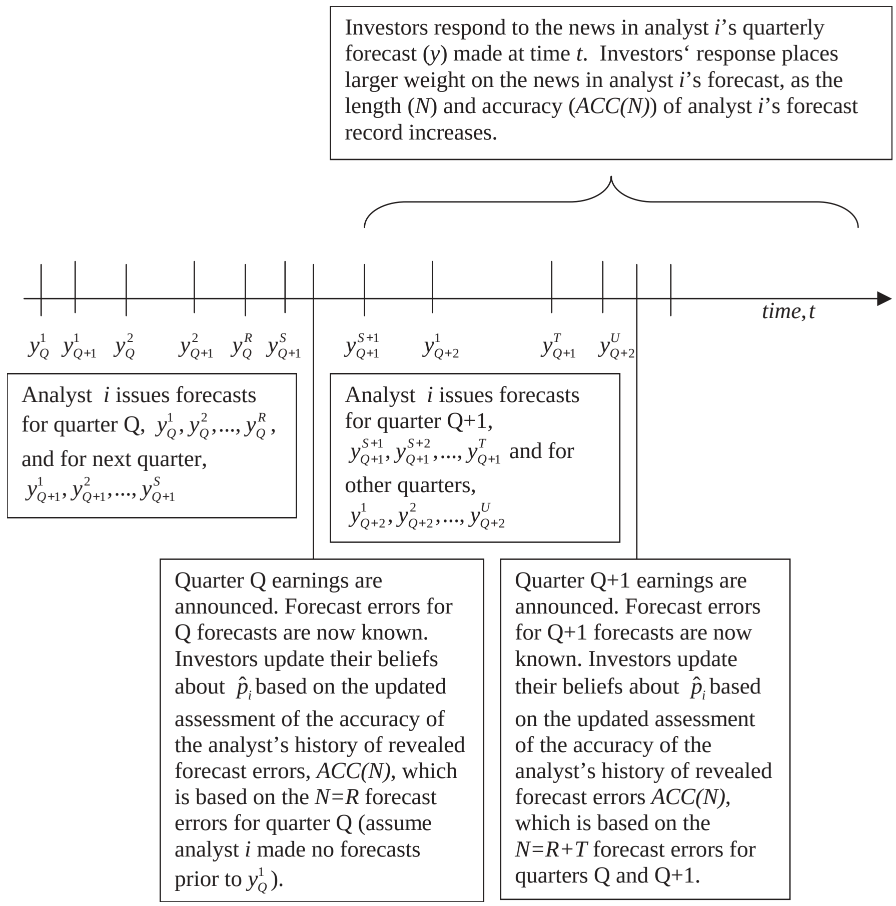

Fig. 1. Timeline of events in the model.

emphasis on the points in time when investor learning occurs and is manifest.) We assume that investors do not observe the analyst's true predictive ability, p, but have a common prior that p is unconditionally normally distributed with mean p_0 and variance $1/h_p$ (throughout we use h_r to denote the precision, or inverse variance, of a random variable r; $\hat{r}$ denotes the estimate of r based on all available information). Investors update those prior beliefs as information about that analyst's forecasting performance is revealed. We assume that the performance signal (such as the realized absolute forecast error), x_n, is related to p by $x_n = p + \delta_n$, with $n = 1, \ldots, N$, and δ_n representing random noise which is distributed i.i.d. normal with mean zero and variance $1/h_\delta$. After N observations of x_n, investors' best estimate (i.e., the estimate with the smallest variance among the class of unbiased estimators) of the analyst's ability, $\hat{p}(N)$, is

$$\hat{p}(N) = \frac{h_p}{h_p + Nh_\delta} p_0 + \frac{Nh_\delta}{h_p + Nh_\delta}\left(\frac{1}{N}\sum_{n=1}^{N} x_n\right)$$
$$= (1 - w(N))p_0 + w(N) \cdot ACC(N), \qquad (1)$$

where $ACC(N) \equiv (1/N)\sum_{n=1}^{N} x_n =$ the accuracy of the analyst's N prior performance signals,

$w(N) \equiv Nh_\delta/(h_p + Nh_\delta) =$ the weight investors place on $ACC(N)$,

$1 - w(N) =$ the weight investors place on their prior perceived ability for the analyst.

Eq. (1) introduces the variable $ACC(N)$, which captures the analyst's revealed predictive ability—that is, the accuracy of her past performance—as measured by the mean of the N absolute forecast errors available for this analyst.[4] By available, we mean that the forecast errors are known to investors at the time they calculate $ACC(N)$.

An important feature of (1) is that investor learning from the accuracy of the analyst's past performance, captured by $w(N) = Nh_\delta/(h_p + Nh_\delta)$, is an increasing and concave function of N. Intuitively, this means that as more signals about the analyst's predictive ability are available (that is, as N increases), investors shift weight off their prior beliefs and onto the revealed predictive ability of the analyst, $ACC(N)$. While this specification assumes that analyst ability is constant over time, Holmstrom's (1999) analyses imply that our predictions hold if we assume that analyst ability follows an autoregressive process or if it is a function of both natural ability and effort. As long as more observations lead to more precise estimation of the analyst's true ability, rational investors will put increasingly higher weights on longer forecast records when forming their posterior inference of the analyst's ability.

To test whether investor learning about analyst ability is consistent with (1), we link perceived ability to observed market data. Specifically, let z denote the firm-specific earnings that the analyst is forecasting; we assume that z is unconditionally normally distributed with mean z_0 and variance $\sigma_z^2 = 1/h_z$. The analyst's forecast, y, reflects her private signal about z; y is generated according to $y = z + \varepsilon_y$ where ε_y is normally distributed with mean zero and variance $1/h_y$, and is uncorrelated with z (i.e., $corr(\varepsilon_y, z) = 0$). Analysts differ in their predictive ability, as indexed by the precision of their private signal, h_y, with higher values of h_y denoting higher predictive ability. Let $h_y/(h_z + h_y)$ be the relative precision of the analyst's signal. We assume that the analyst's true predictive ability, p, is a monotonically increasing function of her relative precision: $p = g(h_y/(h_z + h_y))$ where $g(\cdot) \in R$ is a monotonically increasing function of both $h_y/(h_z + h_y)$ and h_y. That is, p can be viewed as

[4] Because the performance measures are conditionally independent and homoscedastic, the mean is a sufficient statistic for the entire history. Relaxing these assumptions does not change the inferences of the model.

an index for analyst precision, with larger values of p indicating higher ability analysts. For simplicity, we set $p = h_y/(h_z + h_y)$.[5]

Upon observing the analyst's forecast y, investors update their beliefs about earnings (z) using Bayes rule. Because all variables are normally distributed (by assumption), investors' updated assessment of z, $\hat{z} = E(z \mid z_0, y)$, is given by $\hat{z} = (1 - \hat{p})z_0 + \hat{p}y$, where $\hat{p} = E(p \mid I)$ is investors' updated belief about p, the analyst's true ability, conditional on the information set I available at the time of the forecast (Subramanyam, 1996). Note that $\hat{p}$ is the weight that investors place on the analyst's forecast y. The higher the analyst's perceived ability, the greater the weight investors place on her forecasts and therefore, the more influential are the analyst's forecasts in determining share price. Accordingly, we measure the market impact (*MI*) of the analyst's forecast as the stock price response to the news in y:

$$MI = \hat{z} - z_0 = \hat{p}(y - z_0). \tag{2}$$

The motivation for (2) comes from the fact that a company's expected future earnings are important determinants of current share price: the larger the change in the expectation of future earnings, the larger the price revision is to reflect the new information. For simplicity, Eq. (2) suppresses firm-specific factors (such as risk, growth opportunities, and size) that might affect the intensity of market reactions to earnings news. Subsequent tests (discussed in Section 4.2) show that our empirical results are not sensitive to the inclusion or exclusion of firm-specific multipliers on news.

An implicit assumption in (2) is that the analyst's forecast y is, on average, an unbiased estimate of earnings.[6] This is not a crucial assumption as long as, in equilibrium, investors can undo any systematic bias in analysts' forecasts. A feature of our empirical design that further mitigates this assumption is that we relate *MI* not to the forecast itself, but to the news component of the analyst's forecast (*NEWS*), measured as the difference between the analyst's earnings forecast and the prevailing consensus forecast. (As discussed in Section 3.1, we use the consensus to proxy for investors' expectation of earnings, z_0.) Hence, even if y is biased, as long as the consensus contains the same expected bias as y, *NEWS* will be an unbiased estimate of the information conveyed by the analyst's forecast. Further, even if the biases in the forecast and the consensus do not offset, any differences likely create

[5] Since $h_y/(h_z + h_y) \in [0, 1]$, setting p equal to $h_y/(h_z + h_y)$ violates the assumption that p is normally distributed. We make this simplification purely for notational ease: as long as the analyst's true predictive ability is a monotonic transformation of the relative precision of the analyst's private signals, p can be assumed to be normally distributed. More formally, we could decompose investors' updating of earnings expectations into two steps: (1) investors update their beliefs about p (which is a weighted average of investors' prior beliefs and the analyst's forecast record, both are assumed to be normally distributed); and (2) investors recover the relative precision of the analyst, $h_y/(h_z + h_y)$, from p to update their estimate of earnings. Because this two-step process adds no insights incremental to those identified in our simpler setting, we focus on the latter.

[6] Another implicit assumption is that analysts seek to minimize mean squared forecast errors, not mean absolute forecast errors. See Gu and Wu (2003) and Basu and Markov (2003). Since both squared and absolute forecast errors measure forecast accuracy, provided they are positively correlated (they are for our sample), the choice does not affect our analysis.

measurement error (assuming uncorrelated errors), biasing tests of the relation between MI to $NEWS$ toward zero.

Our primary interest is in the path investors take to form $\hat{p}$; we predict that investors form $\hat{p}$ over time in a manner consistent with Bayesian inference. That is, when a new analyst makes her first forecast, investors assign p_0 as her prior perceived ability. Each time the firm announces earnings, investors acquire signals about the analyst's predictive ability from the realized forecast errors. Specifically, investors form $\hat{p}_t = E(p \mid p_0, N_t, ACC(N_t))$ based on: their prior, p_0; the length of the analyst's forecast record up to time t, N_t; and the accuracy of the analyst's forecast record at time t, $ACC(N_t)$. Forecasts whose errors are not known at time t, such as the forecast error associated with y_t, are excluded in calculating N_t and $ACC(N_t)$. Given the assumptions of our model, investors' updated estimate of the analyst's precision at time t can be expressed as $\hat{p}(N_t) = (1 - w(N_t))p_0 + w(N_t)p(ACC(N_t))$, where $p(ACC(N_t))$ is the inferred ability and $\partial p / \partial ACC(N) > 0$.

Substituting (1) into (2) and re-arranging terms yields the following equation relating the market impact of analyst i's forecast at time t to: (i) the news in this forecast, measured as its deviation from the consensus forecast ($NEWS$); (ii) investors' prior assessment of the analyst's type, $p_{i,0}$; (iii) the accuracy of the analyst's forecast record at time t and investors' updating rule, $w(N_{i,t}) \cdot ACC(N_{i,t})$:

$$\begin{aligned} MI_{i,t} &= \hat{p}(N_{i,t}) \cdot NEWS_{i,t} \\ &= [(1 - w(N_{i,t}))p_{i,0} + w(N_{i,t})p(ACC(N_{i,t}))] \cdot NEWS_{i,t} \\ &= p_{i,0}NEWS_{i,t} - p_{i,0}w(N_{i,t}) \cdot NEWS_{i,t} + w(N_{i,t}) \cdot p(ACC(N_{i,t})) \cdot NEWS_{i,t}. \end{aligned} \tag{3}$$

Our empirical tests focus on the updating rule, $w(N_{i,t})$. We require a learning specification that is globally increasing and concave, and has the desirable weighting function properties of being bounded by zero and one. One specification, implied by Eq. (1) of our model, is $w(N_{i,t}) = kN_{i,t}/(1 + kN_{i,t})$, where $k = h_\delta / h_p > 0$ captures the speed that investors shift weight off their prior and onto the accuracy of the analyst's forecast record. (Other weighting specifications are considered in Section 4.1.)

Since we do not observe investors' prior assessment ($p_{i,0}$) about an analyst's ability, we assume it depends on observable characteristics, such as the brokerage house that the analyst works for and the earnings predictability of the firm the analyst covers. Specifically, suppose $p_{i,0} = \vec{\beta}\vec{X}_{i,0}$ where $\vec{X}_{i,0}$ is a vector of characteristic variables that determines investors' priors about an analyst's ability and $\vec{\beta}$ is a conformable vector of coefficients describing the correlation between the hypothesized factors affecting prior beliefs and the true priors. Then we derive from (3) the following regression equation:

$$\begin{aligned} MI_{i,t} = \alpha_0 &+ \vec{\alpha_1}\vec{X}_{i,0} \cdot NEWS_{i,t} + \vec{\alpha_2}\vec{X}_{i,0} \cdot w(N_{i,t}) \cdot NEWS_{i,t} \\ &+ \alpha_3 w(N_{i,t}) \cdot ACC(N_{i,t}) \cdot NEWS_{i,t} + \varepsilon_{i,t}. \end{aligned} \tag{4}$$

Bayesian learning predicts that investors place increasing weights on the accuracy of the analyst's forecast record as the length of the forecast record increases. Therefore, our main hypothesis, stated in alternative form, is $\alpha_3 > 0$. Bayesian learning also predicts that investors place decreasing weights on their prior beliefs as the length of the forecast record grows, implying $\alpha_{2j} < 0$ ($\alpha_{2j} > 0$) (for $\beta_j > 0$ ($\beta_j < 0$)) where α_{2j} is the jth element in the vector $\vec{\alpha_2}$. However, because we do not observe investors' prior perceptions of analysts' abilities, estimating the weight on priors is conditional on our ability to correctly specify $\vec{\beta}$. Importantly, regardless of our ability to consistently estimate $\vec{\beta}$, the estimate of α_3 is consistent, and is the primary focus of our tests.

A special case of (4) is when investors assign a common prior, p_0, to all new analysts:

$$\begin{aligned} MI_{i,t} = {} & \alpha_0 + \alpha_1\, NEWS_{i,t} + \alpha_2 w(N_{i,t}) \cdot NEWS_{i,t} \\ & + \alpha_3 w(N_{i,t}) \cdot ACC(N_{i,t}) \cdot NEWS_{i,t} + \varepsilon_{i,t}. \end{aligned} \tag{5}$$

In (5), we expect $\alpha_1 > 0$, $\alpha_2 < 0$ and $\alpha_3 > 0$ because $p_0 > 0$. A positive value of α_1 corresponds to the well-documented and intuitive result that more news evokes larger market responses. As N increases, we expect a negative value of α_2 (reflecting decreasing reliance on the prior perceived ability of the analyst) and a positive value of α_3 (reflecting increasing weight on the accuracy of the analyst's forecast record).

Eq. (5) shows that the analyst's past forecast accuracy, $ACC(N)$, affects the intensity of the market's reaction to the forecast news through its interaction with $w(N)$. Importantly, the model does not show a role for the accuracy of the analyst's forecast record in the absence of information about the length of that record; that is, our model shows that $ACC(N) \cdot NEWS$ is not a separate independent variable in Eq. (4) or Eq. (5). This difference is subtle but important, as it highlights the distinction between a Bayesian model (which contains information about both the accuracy and the length of the forecast record observed at time t) and a static model (which conditions only on the accuracy of that forecast record).

The following example illustrates the difference in predictions from a Bayesian model versus a static model. Assume investors hold the same prior perceived ability for analysts A, B and C, with this prior equal to 0.10. At time t, A has accumulated a forecast record of length $N = 500$ with accuracy $ACC(N) = 0.01$, B has a forecast record of length $N = 2$ and accuracy $ACC(N) = 0.01$, and C has a forecast record of length $N = 100$ and accuracy $ACC(N) = 0.04$. A static model predicts equally large market reactions to the most accurate analysts' forecasts (A and B, each with accuracy of 0.01 at time t), and a smaller reaction to C's least accurate forecasts (C's accuracy is 0.04 at time t). In contrast, a Bayesian model predicts the following ordering of market reactions at time t: A > C > B. Investors react the least intensely to B's forecasts because, while B's forecast record is as accurate as A's, its short length means that investors will not weight the forecast record heavily in forming their posterior beliefs; hence investors' posterior beliefs will be close to their prior beliefs (0.10) for B. Investors react most intensely to A's forecasts because A's large N means that investors will weight A's forecast record heavily in forming their posterior; hence, investors' posterior beliefs are close to A's forecast record of 0.01. The intensity of investor reaction to C's forecasts lies between B and A, as investors' posterior belief for C's forecast accuracy will be close to 0.04 (as C's forecast record is relatively long), which is larger than that A's posterior (close to 0.01) but smaller than B's posterior (close to 0.10).

In summary, our model shows no role for the accuracy of the analyst's forecast record alone in explaining market reactions to the news in analysts' forecasts. In our setting, accuracy is important in explaining market reactions to forecast news only through its interaction with the length of the forecast record. This result is the basis of our main prediction of a positive coefficient on the term $w(N) \cdot ACC(N) \cdot NEWS$ in Eq. (5). In Section 3.3, we investigate whether this Bayesian specification is more (or less) descriptive of market reactions than the ad hoc static specification.

3. Empirical analysis

3.1. Sample and variable definitions

Our sample analysts are drawn from the Zacks Investment Research database. We begin by identifying analysts whose first forecasts (for any firm) recorded by Zacks were made on or after January 1, 1990. We then identify all quarterly earnings forecasts by these analysts for their first-covered firms, through June 30, 2000. We chose January 1, 1990 as our starting point because Zacks coverage is reasonably complete in the 1990s; this constraint increases the likelihood that the sample analysts have no forecast record prior to the first date in which they appear in our tests. We focus on analyst-firm pairings, as opposed to all forecasts made by an analyst, based on Park and Stice's (2000) finding that learning about analysts' forecasting ability is firm-specific (i.e., investors do not believe an analyst is good at predicting firm A simply because she is good at predicting firm B). These selection criteria produce 73,187 quarterly forecasts, representing 2,938 analyst-firm pairs and 1,969 firms.

We first examine Eq. (5), which assumes that the market assigns a common prior precision to all new analysts. We require measures for market impact, the news in the forecast, and the accuracy and length of analyst's forecast record at each forecast date t. Because there is usually a zero- to one-day lag between an analyst issuing her forecast and its reporting on Zacks, we measure the market impact of analyst i's forecast at time t, $MI_{i,t}$, as the cumulative abnormal return on the stock over a 2-day period ending on the publication date of the analyst's report. Abnormal returns are calculated as the firm's CRSP raw return on day t less the value-weighted market return on the same day. The news in analyst i's forecast made at time t, $NEWS_{i,t}$, is measured as the difference between the analyst's earnings forecast (from Zacks) and the consensus analyst forecast for this firm-quarter prevailing at time t. Observations with $NEWS_{i,t} = 0$ are retained in the sample, however, their inclusion or exclusion has no effect on the results.

We calculate the consensus forecast by weighting all outstanding forecasts made by all analysts for this firm-quarter.[7] Since more recent forecasts should contain more updated information, we use a rank inverse-weighting scheme to assign higher

[7] If there are no outstanding forecasts for the firm-quarter, we use the firm's actual earnings from four quarters ago as the consensus estimate. The results are not sensitive to the exclusion of these observations.

weights to recent forecasts. If there are $q = 1, \ldots, Q$ prevailing forecasts, F_q, that are issued d_q days before t, with $d_Q > d_{Q-1} > \cdots > d_1$, then rank inverse-weighting assigns weight $v_q = 1/q/\sum_{s=1}^{Q} 1/s$. We obtain similar results (not reported) using the raw values (rather than the rank values), $v_q = 1/d_q/\sum_{s=1}^{Q} 1/d_s$. We also tried three other weighting schemes: two based on other weightings of the timeliness of outstanding forecasts (equal-weighting and linear-weighting) and one based on the accuracy of analysts at the dates of their outstanding forecasts (accuracy-weighting). Equal-weighting assigns weight $v_q = 1/Q$ to each F_q, and linear-weighting assigns $v_q = (d_Q - d_q + 1)/\sum_{s=1}^{Q}(d_Q - d_s + 1)$. For accuracy-weighting, we calculate the accuracy of each analyst at the time of her forecast, and use the ranked values of these accuracy measures to weight outstanding forecasts, with more accurate analysts' forecasts assigned higher weights. The correlations between *NEWS* based on these weighting methods are 0.95 or higher. Because results using these other measures (not reported) are similar in all respects to those reported for the rank inverse-weighted measure, we report only results for consensus measures based on the rank inverse-weighting scheme.

We order the forecasts for each analyst-firm pair chronologically by the forecast report dates and assign ordinal values to each forecast. We define $N_{i,t}$ as the number of forecasts made by analyst i (for a given firm) up to time t where the forecasted earnings have been realized before t; that is, $N_{i,t}$ captures the length of analyst i's *proven* forecast record at time t. The accuracy of analyst i's forecast record at time t, $ACC(N_{i,t})$, is measured as the negative of the average absolute forecast errors of her $N_{i,t}$ forecasts. Absolute forecast error equals the absolute value of the difference between forecasted earnings and actual earnings, scaled by share price 5 days prior to the forecast date. We take the negative of the mean absolute forecast error so that larger values of $ACC(N)$ correspond to more accurate forecasts. As with the measure of the consensus forecast, we use three weighting schemes to combine prior absolute forecast errors in calculating the accuracy of the analyst's forecast record: equal-weighting, linear-weighting and inverse-weighting. If the noise in each normalized forecast error is of similar magnitude, equal weighting of prior absolute forecast errors should be optimal. The inverse-weighting and linear-weighting schemes account for two possible effects: a recency effect (investors place greater weight on recent forecast errors even though they contain no more information than distant forecast errors) or changing analyst ability over time such that recent forecast errors are more indicative of her current ability level. Correlations among the test variables (Table 1, panel B) show pair-wise correlations of 0.952 or higher between the equal-weighted and inverse-weighted *ACC*. We report results using equal-weighting for $ACC(N)$ construction, but results are similar using the other weighting schemes (not reported).

Finally, we include in (5) the 1-day lagged abnormal return (*LAGRET*) of firm j. The inclusion of this variable controls for any non-synchronous trading effects that might lead to serial correlation in daily stock returns. Its inclusion or exclusion has no substantive effect on the results.

Table 1, panel A reports descriptive information about the test variables. On average, the sample analysts' forecast records contain a mean (median) of $N = 24.8$

Table 1
Descriptive statistics and pairwise correlations

Panel (A) Summary statistics for test variables

	Mean	Standard deviation	10%	25%	Median	75%	90%
MI	−0.0020	0.0686	−0.0621	−0.0255	−0.0016	0.0238	0.0620
NEWS							
Rank inverse weighting	−0.0016	0.0101	−0.0077	−0.0026	−0.0003	0.0005	0.0024
Raw inverse weighting	−0.0025	0.0114	−0.0102	−0.0037	−0.0006	0.0005	0.0025
N	24.8	30.4	1	4	13	34	64
(−) *ACC*(*N*)							
Equal weighting	−0.0083	0.0150	−0.0007	−0.0018	−0.0039	−0.0090	−0.0185
Inverse weighting	−0.0081	0.0166	−0.0007	−0.0015	−0.0035	−0.0084	−0.0181
LAGRET	−0.0009	0.0387	−0.0351	−0.0153	−0.0003	0.0157	0.0379

Panel (B) Pearson (above diagonal) and Spearman (below diagonal) correlations (p-value in parentheses)

Spearman \ Pearson	*MI*	*NEWS* (*rank inverse*)	*NEWS* (*raw inverse*)	*N*	*ACC*(*N*) (*equal*)	*ACC*(*N*) (*inverse*)	*LAGRET*
MI	1	0.059	0.053	0.018	0.019	0.016	0.016
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
NEWS	0.092	1	0.980	−0.010	−0.065	−0.077	0.015
(*rank inverse*)	(0.00)		(0.00)	(0.01)	(0.00)	(0.00)	(0.00)
NEWS	0.084	0.958	1	−0.016	−0.095	−0.110	0.013
(*raw inverse*)	(0.00)	(0.00)		(0.00)	(0.00)	(0.00)	(0.00)
N	0.008	−0.051	−0.064	1	0.050	0.042	0.004
	(0.01)	(0.00)	(0.00)		(0.00)	(0.00)	(0.25)
ACC(*N*)	−0.001	−0.103	−0.121	0.294	1	0.952	−0.004
(*equal*)	(0.76)	(0.00)	(0.00)	(0.00)		(0.00)	(0.27)
ACC(*N*)	−0.001	−0.114	−0.133	0.264	0.976	1	−0.005
(*inverse*)	(0.82)	(0.00)	(0.00)	(0.00)	(0.00)		(0.16)
LAGRET	0.013	0.046	0.040	0.004	−0.011	−0.011	1
	(0.00)	(0.00)	(0.00)	(0.20)	(0.00)	(0.00)	

Sample description and variable definitions: the sample consists of 73,187 quarterly earnings forecasts made by 2,938 analysts following 1,969 firms between January 1, 1990 and June 30, 2000. *MI* is the cumulative abnormal return over the 2-day period ending on the forecast date, where abnormal return is calculated as the difference between the CRSP stock return and the value weighted market return. *NEWS* is the difference between sample analysts' forecasts and the consensus forecast, where the consensus forecast is a weighted average of all prevailing forecasts made prior to time t. With *rank* (*raw*) *inverse weighting*, we assign a weight to each observation that is inversely proportional to the ranked value (raw value) of the distance between the date of that observation and the current forecast date. With *equal weighting*, we assign an equal weight to each observation. N is the number of the analyst's prior forecasts with observed forecast errors at time t. $ACC(N)$ *is* the negative of the analyst's average absolute forecast error for her N prior forecasts. *LAGRET* is the lagged abnormal 1-day return 2 days before the announcement date.

(13) forecast errors. The mean accuracy of the forecast records, $ACC(N)$, is -0.0083, with a standard deviation of 0.0150. The mean $NEWS$ conveyed by the sample forecasts is -0.0016 (or -0.16% of stock price), with a standard deviation of 0.0101. The mean market reaction to the sample forecasts is -0.0020 (or -0.20% of stock price), with a standard deviation of 0.0686.

3.2. Tests of Bayesian learning

Our first set of tests estimates the following regression:

$$\begin{aligned} MI_{i,t} = {} & \alpha_0 + \alpha_1\, NEWS_{i,t} + \alpha_2 w(N_{i,t}) \cdot NEWS_{i,t} \\ & + \alpha_3 w(N_{i,t}) \cdot ACC(N_{i,t}) \cdot NEWS_{i,t} + \alpha_4\, LAGRET_{i,t} + \varepsilon_{i,t}. \end{aligned} \quad (6)$$

Recall that $w(N_{i,t}) = kN_{i,t}/(1 + kN_{i,t})$. Because $k = h_\delta/h_p$ is not directly estimable, we solve for k and all the α coefficients simultaneously by estimating (6) using non-linear least-squares procedures. This estimation produces a value of $k = 0.0989$, which implies that investors resolve about half of their uncertainty about analyst ability after observing 10 forecast errors. Throughout our analyses, we use $k = 0.10$; hence the learning function is given by $w(N_{i,t}) = 0.1N_{i,t}/(1 + 0.1N_{i,t})$.

The first column of Table 2 reports the OLS coefficient estimates and White (1980) heteroscedasticity consistent t-statistics for Eq. (6). In Table 2 and throughout the paper, we report results based on the full sample; we find similar results (not reported) for samples that trim the 1% extremes based on absolute studentized residuals or based on the absolute values of MI or $NEWS$. As expected given prior studies' findings concerning market reactions to the news in analysts' forecasts (e.g., Lys and Sohn, 1990), the coefficient relating $NEWS$ to MI, α_1, is significantly positive (t-statistic $= 4.61$). We do not find $\alpha_2 < 0$ at conventional levels of significance. This may be because of the fact that the assumption of common priors is needed to predict $\alpha_2 < 0$. The key prediction of the model, $\alpha_3 > 0$, is supported by the results: the coefficient estimate for $w(N) \cdot ACC(N) \cdot NEWS$, $\hat{\alpha}_3$, is equal to 7.890 (t-statistic $= 3.34$).

We examine the sensitivity of these results to different estimation procedures. First, we assess the sensitivity of the results to the pooled regression assumption of a constant intercept across all analyst-firm pairs by estimating fixed effects models which allow for analyst-specific intercepts and, separately, for firm-specific intercepts. The coefficient estimates and t-statistics for the analyst fixed effects model and for the firm fixed effects model are reported in columns (2) and (3) of Table 2; the 2,938 analyst-specific intercepts and the 1,969 firm-specific intercepts are not tabulated. In both fixed effects models, the coefficient on $NEWS$ is positive (t-statistics on $\hat{\alpha}_1$ are about 3), and the coefficient estimate on $w(N) \cdot NEWS$ is insignificant, similar to the results in column (1). Both fixed effects specifications show a significant positive coefficient on $w(N) \cdot ACC(N) \cdot NEWS$, with the t-statistic on $\hat{\alpha}_3$ equal to 2.80 for the analyst fixed effects model and 3.21 for the firm fixed effects model.

Based on prior research documenting differential market responses to unexpected earnings information conditional on firm-specific factors (e.g., Easton and

Table 2
Tests of Bayesian learning with common priors about analyst predictive ability[a]

Independent variable	(1) Pooled OLS	(2) Analyst fixed effects	(3) Firm fixed effects	(4) Firm slope effects	(5) Year effects	(6) Quantile regression	(7) Cluster regression
Intercept	−0.001	—	—	−0.000	—	−0.004	−0.001
	(−4.93)			(−0.86)		(−27.27)	(−1.83)
NEWS	0.428	0.324	0.330	0.572	0.393	0.182	0.428
	(4.61)	(3.24)	(3.34)	(1.52)	(2.66)	(5.33)	(3.11)
$w(N) \cdot NEWS$	0.189	0.140	0.196	−0.130	0.138	0.151	0.189
	(1.22)	(0.96)	(1.30)	(−0.53)	(0.89)	(2.45)	(0.85)
$w(N) \cdot ACC(N) \cdot NEWS$	7.890	6.114	7.738	8.114	9.025	3.634	7.890
	(3.34)	(2.80)	(3.21)	(2.64)	(3.99)	(4.56)	(2.83)
LAGRET	0.026	0.005	0.005	0.017	0.024	−0.025	0.026
	(2.41)	(0.46)	(0.56)	(1.64)	(2.27)	(−6.32)	(0.94)
Adj. R^2	0.004	0.142	0.115	0.063	0.007	0.023	0.004

Sample and variable definitions: See Table 1.

[a] We report the coefficient estimates (t-statistics) associated with regressions of Eq. (6):

$$MI_{i,t} = \alpha_0 + \alpha_1\, NEWS_{i,t} + \alpha_2 w(N_{i,t}) \cdot NEWS_{i,t} + \alpha_3 w(N_{i,t}) \cdot ACC(N_{i,t}) \cdot NEWS_{i,t} + \alpha_4\, LAGRET_{i,t} + \varepsilon_{i,t}$$

Column (1) shows the OLS pooled regression results. Analyst- and firm-fixed effect results are reported in columns (2) and (3), respectively; we do not tabulate the analyst-specific and firm-specific intercepts. Column (4) reports results with firm-specific slope coefficients for *NEWS*; we report the mean value of the $J = 1{,}969$ firm-specific *NEWS* coefficients. Column (5) reports results with year-specific intercepts and year-interaction slope coefficients for *NEWS*; we report the mean value of the 11 year-specific *NEWS* coefficients (year intercepts are not tabulated). Quantile regressions are reported in column (6). In column (7), we report the coefficient estimates for the OLS cluster regression, where the t-statistics are based on standard errors adjusted for both heteroscedasticity and intra-cluster error correlation. Each cluster consists of all observations within the same forecast week and same industry (as measured by 2-digit SIC code).

Zmijewski, 1989), we examine whether the results in Table 2 are driven by firm-specific responses to forecast news. Specifically, we repeat our tests allowing for firm-specific response coefficients on *NEWS*. Column (4) in Table 2 reports the results. The key coefficient of interest, $\hat{\alpha}_3$, remains significantly positive, with a t-statistic of 2.64. As an alternative test (not tabulated), we interact *NEWS* with measures of systematic risk (beta, estimated using the market model and all trading days in year $t-1$), growth opportunities (proxied by the book value of equity to the market value of equity at the end of year $t-1$), and firm size (measured as the market value of equity in the end of year $t-1$). For the sub-sample of 55,317 quarterly forecasts with data on these variables, we find their inclusion in Eq. (6), if anything, enhances the significance of the main test variable: $\hat{\alpha}_3 = 12.62$, t-statistic $= 3.45$.

We also examine whether our results are affected by over time changes in market responses to analysts' forecasts, potentially related to over-time changes in properties of those forecasts. Such patterns will affect our results to the extent that $w(N)$ is correlated with calendar time; for our sample forecasts, this pair-wise correlation is 0.19. We re-estimate Eq. (6) with year intercepts and year dummies interacted with *NEWS*; results for the main test variables are reported in column (5) of Table 2. Importantly, $\hat{\alpha}_3$ remains significantly positive, with a *t*-statistic of 3.99.

The last two columns of Table 2 report the results of two robustness checks. In column (6) we report the results of quantile regressions which constrain the median residual to be zero.[8] Similar to the other estimation procedures, the quantile regressions show a significantly positive coefficient estimate on $w(N) \cdot ACC(N) \cdot NEWS$, $\hat{\alpha}_3 = 3.634$ (*t*-statistic = 4.56). Column (7) reports cluster-adjusted standard errors which allow for error correlations induced by common shocks. All observations within the same forecast week and same industry at two-digit SIC code are clustered, and the sample variance–covariance matrix of each cluster is used to form the weighting matrix for estimating standard errors; the point estimates of the coefficients are unaffected by this procedure.[9] The results show that while clustering affects the standard errors of the constants, inferences about $\hat{\alpha}_3$ (*t*-statistic = 2.83) are not altered.

Overall, the results in Table 2 suggest that market reactions to analyst reports are consistent with investors placing greater weight on an analyst's displayed ability (as captured by $ACC(N)$) as her forecast record becomes a more precise estimate of her true ability (that is, as N increases). This Bayesian learning result is robust to analyst-, firm-, and year-specific effects as well as to deviations from distributional assumptions on the error term underlying ordinary least squares estimation.

3.3. Static versus Bayesian learning

In this section, we examine whether prior studies' evidence of static learning exists for our sample, and compare the ability of the static model and the Bayesian model to explain market reactions to forecast news. Mikhail et al. (1997) and Park and Stice (2000) find larger market reactions to forecasts issued by superior analysts, defined as analysts with more precise forecasts (based on prior forecast errors). The precision of the analyst's prior forecasts at time t is captured by $ACC(N_{i,t})$, and is calculated as previously defined. We begin by replicating prior studies' evidence that the market reaction to forecast news is positively related to $ACC(N_{i,t})$, i.e., $\beta_2 > 0$ in Eq. (7):

$$MI_{i,t} = \beta_0 + \beta_1\, NEWS_{i,t} + \beta_2\, ACC(N_{i,t}) \cdot NEWS_{i,t} + \beta_3\, LAGRET_{i,t} + \varepsilon_{i,t}. \quad (7)$$

The results of estimating Eq. (7) are shown in column (1) of Table 3. Consistent with prior studies' evidence of larger market reactions to reports made by analysts who have more accurate forecast records, we document a significant positive coefficient on $ACC(N) \cdot NEWS(\hat{\beta}_2 = 3.005,\ t\text{-statistic} = 2.92)$.

[8] Koenker and Bassett (1978) show that when the distribution of the residuals has fat tails, quantile regression estimators are more efficient than conventional least squares estimators and are more robust to outliers.

[9] As long as the number of clusters approaches infinity as the sample size goes to infinity, the resulting standard error estimators are robust to arbitrary forms of heteroskedasticity and intra-cluster correlation (Wooldridge, 2002).

Table 3
Tests of static learning and tests comparing Bayesian and static learning[a]

Independent variable	(1)[a]	(2)[b]
Intercept	−0.001	−0.001
	(−5.00)	(−4.93)
NEWS	0.504	0.448
	(8.51)	(3.78)
$w(N) \cdot NEWS$	—	0.152
		(0.80)
$ACC(N) \cdot NEWS$	3.005	0.545
	(2.92)	(0.37)
$w(N) \cdot ACC(N) \cdot NEWS$	—	6.758
		(2.16)
LAGRET	0.026	0.026
	(2.43)	(2.42)
Adj. R^2	0.0041	0.0042
F-statistic, excluding $ACC(N) \cdot NEWS$	—	0.222
(p-value)		(*0.64*)
F-statistic, excluding $w(N) \cdot ACC(N) \cdot NEWS$	—	6.094
(p-value)		(*0.01*)

Sample and variable definitions: See Table 1.

[a] Column (1) reports the coefficient estimates (t-statistics) from estimating Eq. (7):

$$MI_{i,t} = \beta_0 + \beta_1\, NEWS_{i,t} + \beta_2\, ACC(N_{i,t}) \cdot NEWS_{i,t} + \beta_3\, LAGRET_{i,t} + \varepsilon_{i,t}.$$

[b] Column (2) reports the coefficient estimates (t-statistics) from estimating Eq. (8):

$$MI_{i,t} = \rho_0 + \rho_1\, NEWS_{i,t} + \rho_2 w(N_{i,t}) \cdot NEWS_{i,t} + \rho_3\, ACC(N_{i,t}) \cdot NEWS_{i,t} + \rho_4 w(N_{i,t}) \cdot ACC(N_{i,t}) \cdot NEWS_{i,t} + \rho_5\, LAGRET_{i,t} + \varepsilon_{i,t}.$$

The positive value for β_2 is consistent with a static learning model in which investors condition their response to *NEWS* on the accuracy of the analyst's forecast record at time t, but not (necessarily) on the path that produced that forecast record. In contrast, a Bayesian learning model requires that investors condition their response to *NEWS* on both the accuracy and length of the analyst's past forecast record. Bayesian learning not only implies static learning, it requires that when forming perceptions of an analyst's ability, investors put more weight on proven accuracy (as well as proven inaccuracy) as an analyst builds her forecast record. Note that Bayesian learning subsumes static learning as a Bayesian learner would, other things equal, react more intensely to analysts with better forecast records.

To test whether the static model or the Bayesian model is more descriptive of investor learning, we conduct F-tests of the following nested model (see Judge et al., 1985):

$$MI_{i,t} = \rho_0 + \rho_1\, NEWS_{i,t} + \rho_2 w(N_{i,t}) \cdot NEWS_{i,t} + \rho_3\, ACC(N_{i,t}) \cdot NEWS_{i,t} + \rho_4 w(N_{i,t}) \cdot ACC(N_{i,t}) \cdot NEWS_{i,t} + \rho_5\, LAGRET_{i,t} + \varepsilon_{i,t}. \quad (8)$$

If static learning better explains investor behavior than Bayesian learning, we expect to observe a significant increase in residual variances from excluding $ACC(N) \cdot$

$NEWS$ from (8); if, however, Bayesian learning better explains investor behavior, the increase in residual variances will be greater when $w(N) \cdot ACC(N) \cdot NEWS$ is excluded. Coefficient estimates and White t-statistics for (8), along with F-statistics and p-values, are reported in column (2), Table 3. These results show that removing the variable, $ACC(N) \cdot NEWS$, from the regression has no discernible effect on residual variances (F-statistic $= 0.222$, p-value $= 0.64$). In contrast, removing the dynamic regressor, $w(N) \cdot ACC(N) \cdot NEWS$, leads to a significant increase in residual variances (F-statistic $= 6.094$, p-value $= 0.01$), indicating that the Bayesian model provides significantly greater explanatory power than the static model.

Overall, the evidence in Table 3 is more consistent with market reactions to analyst reports being described by a Bayesian model of learning about analyst predictive ability than by a static model of learning. These results are consistent with arguments that investors use historical data to rationally infer underlying parameters (Timmermann, 1993, 1996; Brennan, 1998; Lewellen and Shanken, 2002), and, on the whole, suggest a high level of rationality and sophistication in information processing.

4. Extensions and alternative explanations

We consider several extensions and three alternative explanations for our findings. First, we revisit the assumption that investor learning is described by $w(N) = kN_{i,t}/(1 + kN_{i,t})$. Second, we examine the sensitivity of the results to the assumption of common prior beliefs. Third, we explore the implicit assumption that all investors are equally capable of learning. Finally, we examine whether analyst learning, researcher learning or attrition bias explains the results. Because none of these additional tests alters inferences about the main results, we summarize (but do not tabulate) their results in this section.

4.1. Alternative learning functions

We report two tests of the sensitivity of our results to the assumed learning function, $w(N) = kN_{i,t}/(1 + kN_{i,t})$, with $k = 0.10$. First, we repeat the analyses using other values for k. Recall that k captures the speed that investors shift their weight (off the prior and) onto the forecast record: the higher the value of k, the faster investors shift weight toward the forecast record when forming their posterior belief about an analyst's ability. Results based on $k = 0.05$, 0.075, 0.3, 0.5 and 1.0 show that the Bayesian learning coefficient is uniformly positive, with t-statistics ranging from 2.75 to 3.81. Second, we examine the sensitivity of the results to two other learning specifications: (i) a family of exponential functions, $w(N) = (\exp(kN) - 1)/\exp(kN)$; and (ii) a quadratic learning function, $w(N) = aN - bN^2$, where we predict $a > 0$ and $b > 0$. The results of both specifications show that the Bayesian learning coefficients have the predicted signs, and are significant at the 0.01 level. We conclude that our results are not sensitive to the specification of the learning function.

4.2. Tests of Bayesian learning with non-common priors

Our main tests assume that investors have a common prior, p_0, about the predictive ability of an analyst making her maiden forecast. Intuitively, this assumption means that investors are not able to distinguish among new analysts who do not yet have forecast records. We extend our analysis to consider different priors, $p_{0,i}$, formed by investors based on observable characteristics of the analyst at the time she issues her first forecast: $p_{0,i} = \vec{X}_i\vec{\beta}$, where $\vec{X}_i$ is a vector of characteristics. Based on prior researchers' findings that analysts employed by larger or more prestigious brokerage firms have more accurate forecasts (Clement, 1999; Jacob et al., 1999; Hong and Kubik, 2003), we examine whether investors attach a higher prior to forecasts issued by analysts employed by brokerage houses with large research departments (*SIZE*, equal to the number of analysts with forecasts on Zacks in analyst *i*'s first year) or elite banks (*ELITE*, equal to one if the analyst's employer is one of *Institutional Investor's* top 10 brokerage houses that year, and zero otherwise). Investors may also consider the predictability of the earnings being forecasted (*PREDICT*, measured as the median absolute forecast error for this firm, calculated across all forecasts made over the four quarters prior to analyst *i*'s first forecast).[10] We assign a higher prior to forecasts made by analysts who cover companies with historically more predictable earnings. Substituting these variables for $X_{i,0}$ in Eq. (4) yields the following regression equation:

$$\begin{aligned} MI_{i,t} = \lambda_0 &+ \lambda_1\, NEWS_{i,t} + \lambda_2\, SIZE_i \cdot NEWS_{i,t} + \lambda_3\, ELITE_i \cdot NEWS_{i,t} \\ &+ \lambda_4\, PREDICT_i \cdot NEWS_{i,t} \\ &+ \lambda_5 w(N_{i,t}) \cdot NEWS_{i,t} + \lambda_6 w(N_{i,t}) \cdot SIZE_i \cdot NEWS_{i,t} \\ &+ \lambda_7 w(N_{i,t}) \cdot ELITE_i \cdot NEWS_{i,t} \\ &+ \lambda_8 w(N_{i,t}) \cdot PREDICT_i \cdot NEWS_{i,t} \\ &+ \lambda_9 w(N_{i,t}) \cdot ACC(N_{i,t}) \cdot NEWS_{i,t} + \lambda_{10}\, LAGRET_{i,t} + \varepsilon_{i,t} \end{aligned}$$

We expect that $\lambda_2 > 0,\ \lambda_3 > 0,\ \lambda_4 > 0$ and $\lambda_6 < 0,\ \lambda_7 < 0,\ \lambda_8 < 0$ if investors believe *SIZE*, *ELITE* and *PREDICT* proxy (positively) for the analyst's prior ability. While the results show some evidence of $\lambda_3 > 0$ and $\lambda_4 > 0$, we find no evidence of the predicted shift in weights away from these measures of prior beliefs as the length of the forecast record increases. That is, none of the coefficients interacting *SIZE*, *ELITE* or *PREDICT* with $w(N) \cdot NEWS$ is distinguishable from zero. However, the coefficient on the Bayesian learning variable, λ_9, remains reliably positive (t-statistic = 2.66). Further, repeating the F-tests of the nested model for the case of non-common prior beliefs show, if anything, stronger evidence that Bayesian learning dominates static learning in explaining investor responses to forecast news. In particular, removing the static variable, $ACC(N) \cdot NEWS$, has no significant

[10] If an insufficient number of forecasts exists to calculate *PREDICT* (for example, if the analyst is among the first to cover a firm that just went public), we set *PREDICT* equal to the 90th percentile value of the *PREDICT* distribution under the view that sparsely covered firms have low predictability. Results are not affected if we exclude these observations.

effect on residual variances (F-statistic $= 0.691$, p-value $= 0.41$); however, excluding $w(N) \cdot ACC(N) \cdot NEWS$, leads to a significant increase in residual variances (F-statistic $= 8.358$, p-value $= 0.004$).

We also estimate (4) by allowing investors to have employer-specific or analyst-specific priors, as operationalized by permitting the slope coefficients on $NEWS$ and on $w(N) \cdot NEWS$ to vary across the 257 employers and 2,938 analysts in our sample. In both the employer-specific case and the analyst-specific case, we find that the coefficient of main interest, α_3, is significantly positive (at the 0.07 level or better). These results indicate that Bayesian learning is significant under the least restrictive specification about investors' prior beliefs about new analysts' predictive abilities. Further, although we motivate the employer- and analyst-specific tests as controlling for differing prior beliefs, the results are more broadly consistent with the view that Bayesian learning is not driven by firm- or analyst-specific factors relating market responses to analysts' forecasts.

4.3. Differential investor learning ability

Our main tests assume that firm l's investors are as capable of learning about analyst predictive ability as firm m's investors. We probe this assumption by examining whether the Bayesian learning effects documented for our sample are concentrated in firms with high levels of institutional ownership. This analysis is motivated by prior research which frequently characterizes institutional investors as more sophisticated than individual investors (see, for example, Hand, 1990; Walther, 1997; Sias and Starks, 1997). If sophistication is a pre-condition for learning, we expect to observe more evidence of learning in sub-samples of firms with high institutional holdings relative to sub-samples of firms with low institutional holdings.

Using institutional holding data from Spectrum, we rank firms based on their average percent common shares held by institutions over 1990–2000, and divide the ranked distribution into thirds. For our sample, the 1/3 and 2/3 percentile cutoff points are 39.2% and 57.1%, respectively; firms in the top third (the High institutional holding sample) have mean institutional holdings of 67.4% compared to 24.5% for the bottom third (the Low institutional holding sample). We then estimate a variant of Eq. (6) which includes the interaction terms, $High \cdot w(N) \cdot NEWS$ and $High \cdot w(N) \cdot ACC(N) \cdot NEWS$, where $High = 1(0)$ if the forecast is for a firm in the High (Low) Institutional Holding sample. The results show that the coefficient estimate on $High \cdot w(N) \cdot ACC(N) \cdot NEWS$ is positive (t-statistic $= 2.02$), suggesting that, relative to investors of low institutionally held firms, investors of high institutionally held firms shift significantly greater weight onto the analyst's forecast record as the latter becomes a more precise estimate of the analyst's true ability. Further tests show that this result is not driven by the larger size of firms with high levels of institutional holdings.

Overall, these results are consistent with prior research which characterizes institutional investors as more sophisticated than retail investors. Specifically, we find that Bayesian learning is significantly more pronounced in firms with high levels of institutional holdings, even controlling for the larger size of these firms.

4.4. Alternative explanations: analyst learning, researcher learning and attrition bias

Our final extension considers three alternative explanations for the results: analyst learning, observational equivalence of investor learning and researcher learning, and an attrition bias related to the length of analysts' forecast records. The first explanation—analyst learning—is based on the premise that it is not investors who learn, but analysts.[11] Specifically, if analysts learn over time in a manner that results in improvements in their forecast accuracy that is positively related to N, we would expect to find that the coefficient on our main test variable, $w(N) \cdot ACC(N) \cdot NEWS$, would be positive, i.e., $\alpha_3 > 0$; this is the same prediction as generated by the scenario where investors learn. This result obtains because $w(N) \cdot ACC(N) \cdot NEWS$ will be positively correlated with $ACC(N) \cdot NEWS$; the latter captures the information in the analyst's forecast that a non-learning investor would respond to. Importantly, however, if investors are not learning, we would expect the static model to perform as well as the Bayesian model. That is, if investors are not learning, then at any time t, they do not consider the length of the forecast record that produced the period t measure of the analyst's forecast accuracy, rather they care only about the average forecast accuracy of that record; this is precisely the static model's prediction. Because the evidence in Table 2 rejects the static model in favor of the Bayesian model, we conclude that our results are consistent with investor learning. Note that our results do not speak to whether analysts themselves learn; they may or may not.[12] Importantly, analyst learning has no impact on our conclusion that investors learn.

The second explanation—researcher learning—is also based on the premise that investors do not learn about analysts' predictive ability (perhaps because investors, somehow, directly observe the analyst's true ability),[13] but rather the researcher (who does not observe the analyst's true ability and uses the forecast record as a proxy) is simply better able to measure ability as the length of the forecast record increases. Stated differently, there is an errors-in-variables problem which biases toward our results if the error is less severe for longer forecast records. The researcher is, therefore, more likely to detect a positive correlation between market responses and forecast records when forecast records are long. Although researcher learning is

[11] Mikhail et al. (1997) find a significant improvement in forecast accuracy as the analyst's firm-specific experience increases, consistent with the view that analysts "learn-by-doing." Jacob et al. (1999) argue, however, that this effect is eliminated if one controls for the fact that better forecasters are more likely to survive, and poorer forecasters are more likely to fail. Because we do not require our sample analysts, or the firms they follow, to have survived for any length of time, it is unlikely that analyst- or firm-selection bias explains our findings.

[12] For our sample analysts, unreported tests show no evidence that forecast accuracy improves as analysts become more experienced.

[13] We note that prior research is not consistent with the explanation that market participants know analysts' true ability (without learning). In particular, prior studies' finding that analysts' promotions and job terminations are related to analysts' past forecast accuracy (Mikhail et al., 1999; Hong et al., 2000; Hong and Kubik, 2003) suggests that analysts' employers do *not* know analysts' abilities, but rather learn about them over time. Assuming that investors have no more information about analysts than their employers, these results suggest that investors also learn about analyst ability.

inherently irrefutable (because the analyst's true ability can not be measured without error by the researcher), we provide evidence of its influence by adding as a control variable to Eq. (6), $AllACC_i \cdot NEWS_{i,t}$, where $AllACC_i$ = the accuracy of analyst i's forecast record measured at the time of her *last* sample forecast date. When the analyst has a long forecast record, $AllACC_i$ should be a very precise measure of her true ability. If investors know the analyst's true ability, and therefore do not learn about it over time, we expect $AllACC_i \cdot NEWS_{i,t}$ to be significant in explaining MI, while the Bayesian learning term will be insignificant in the presence of $AllACC \cdot NEWS$. Results show that the coefficient on $AllACC \cdot NEWS$ is negative, not positive as predicted by researcher learning. More importantly, we find no evidence that $w(N) \cdot ACC(N) \cdot NEWS$ loses significance in the presence of $AllACC \cdot NEWS$; across all specifications, the Bayesian learning variable is positive, with t-statistics ranging between 3.84 and 5.04.

The third explanation for our results is an attrition bias: the fact that an analyst has survived to produce a long forecast record makes investors more likely to believe that she has high forecasting ability.[14] In this case, α_2, the coefficient on $w(N) \cdot NEWS$ is biased upwards because this term now measures both the changing weight on prior beliefs and investors inferences about ability drawn from the length of the forecast record. Although analyst attrition bias offers a potential explanation for why our estimates of $\hat{\alpha}_2$ are higher than expected, it has no influence on $\hat{\alpha}_3$ which measures the marginal contribution of $w(N) \cdot ACC(N) \cdot NEWS$ to explaining market responses that is not correlated with $w(N) \cdot NEWS$.

There is, however, a more subtle form of attrition bias that may affect our results. If higher ability analysts survive not only because they are more accurate but because they are more persuasive (i.e., their forecasts have greater market impact), we expect to find a positive correlation between market reactions and long, accurate forecast records, even without investor learning. To investigate the influence of this form of attrition bias on our results, we include variables interacting $NEWS$ and $w(N) \cdot NEWS$ with a measure of the analyst's historical market impact, $HistMI_{i,t}$, measured as the average of the absolute value of the 2-day market responses to analyst i's forecasts made before the current forecast date t. If analyst persuasiveness explains our results, we expect that the coefficient on $HistMI \cdot NEWS$ is positive, and the coefficient on the Bayesian learning variable is insignificant. Results of this test show no evidence of reliably positive coefficients on the variables interacting $HistMI$ with $NEWS$ or with $w(N) \cdot NEWS$. Moreover, we note that the significance of $\hat{\alpha}_3$ is largely unaffected by the inclusion of these controls for analysts' persuasiveness (t-statistics for $\hat{\alpha}_3$ range between 2.68 and 3.25).

In summary, while analyst learning, researcher learning and analyst attrition bias are potential explanations for our findings, further analyses indicate that none of

[14] Because our sample includes forecasts made by analysts who survived to the last date of the sampling period *and* by those who did not, our results are not subject to the common sampling-survivor bias (i.e., the bias from making predictions about the population based on a sub-sample that contains survivors only).

these explanations drives our results. Overall, our results continue to support a model of Bayesian learning by investors.

5. Summary and conclusion

Bayesian learning predicts that investors shift weight from their prior perceived ability to the analyst's forecast record, as that forecast record accumulates evidence about the true (unobservable) predictive ability of the analyst. Based on this prediction, our model shows that the intensity of market reactions to analyst reports is increasing in the product of the length and accuracy of the analyst's past series of forecast errors (i.e., her forecast record). Results of empirical tests, based on a large sample of quarterly earnings forecasts over 1990–2000, are consistent with this prediction. The results are robust to the measurement of the test variables, to outliers, to alternative estimation procedures, to assumptions about prior beliefs, to different specifications of the learning function, and to alternative explanations.

We compare our results to those of prior empirical studies which document larger market responses to forecasts issued by more accurate analysts. We find that the process by which investors distinguish among analysts is richer than a static process that focuses on just the accuracy of the forecast record at a given point in time. Investors incorporate the history of an analyst's past performance in a rational and Bayesian manner by placing greater weight on the accuracy of the analyst's forecast record as the length of that forecast record increases. Once this Bayesian learning is accounted for, our results show that static learning does not provide extra information.

At the analyst level, our results provide evidence on why analysts care about their reputations and how they build them. At a broader level, our findings are consistent with the market as a whole processing information in a rational and sophisticated manner.

References

Antonji, J., Pierret, C., 2001. Employer learning and statistical discrimination. Quarterly Journal of Economics 116, 313–350.

Avery, C., Chevalier, J., 1999. Herding over the career. Economic Letters 63, 327–333.

Basu, M., Markov, S., 2003. Loss function assumptions in rational expectations tests on financial analysts' earnings forecasts. Working paper, Emory University.

Brav, A., Heaton, J., 2002. Competing theories of financial anomalies. Review of Financial Studies 15 (2), 575–606.

Brennan, M., 1998. The role of learning in Bayesian portfolio decisions. European Finance Review 1, 295–306.

Clement, M., 1999. Analyst forecast accuracy: do ability, resources, and portfolio complexity matter? Journal of Accounting and Economics 27 (3), 285–303.

Easton, P., Zmijewski, M., 1989. Cross-sectional differences in the market response to the announcement of accounting earnings. Journal of Accounting and Economics 11, 117–141.

Ehrbeck, T., Waldmann, R., 1996. Why are professional forecasters biased? Agency versus behavioral explanations. Quarterly Journal of Economics 111, 21–41.
Gu, Z., Wu, J., 2003. Earnings skewness and analyst forecast bias. Journal of Accounting and Economics 35, 5–29.
Hand, J., 1990. A test of the extended functional fixation hypothesis. The Accounting Review 65 (4), 764–780.
Holmstrom, B., 1999. Managerial incentive problems: a Bayesian perspective. Review of Economics Studies 66, 169–182.
Hong, H., Kubik, J., 2003. Analyzing the analysts: career concerns and biased earnings forecasts. Journal of Finance LVIII (1), 313–351.
Hong, H., Kubik, J., Solomon, A., 2000. Security analysts' career concerns and herding of earnings forecasts. Rand Journal of Economics 31, 121–144.
Jacob, J., Lys, T., Neale, M., 1999. Expertise in forecasting performance of security analysts. Journal of Accounting and Economics 28, 51–82.
Judge, G., Griffiths, W., Hill, C., Lutkepohl, H., Lee, T., 1985. The Theory and Practice of Econometrics, 2nd Edition. Wiley, New York.
Koenker, R., Bassett, G., 1978. Regression quantiles. Econometrica 46, 33–50.
Krueger, A.B., Fortson, K.N., 2003. Do markets respond more to more reliable labor market data? A test of market rationality. Journal of the European Economic Association 1 (4), 931–957.
Lang, M., 1991. Time-varying stock price response to earnings induced by uncertainty about the time-series properties of earnings. Journal of Accounting Research 29 (2), 229–257.
Lewellen, J., Shanken, J., 2002. Learning, asset-pricing tests, and market efficiency. Journal of Finance 57, 1113–1146.
Lys, T., Sohn, S., 1990. The association between revisions of financial analysts' earnings forecasts and security-price changes. Journal of Accounting and Economics 13, 341–363.
Mikhail, M., Walther, B., Willis, R., 1997. Do security analysts improve their performance with experience? Journal of Accounting Research 35, 131–157.
Mikhail, M., Walther, B., Willis, R., 1999. Does forecast accuracy matter to security analysts? The Accounting Review 74, 185–200.
Mikhail, M., Walther, B., Willis, R., 2004. Are there differences in stock recommendation profitability across individual analysts? Journal of Financial Economics, forthcoming.
Park, C., Stice, E., 2000. Analyst forecasting ability and the stock price reaction to forecast revisions. Review of Accounting Studies 5, 259–272.
Sias, R., Starks, L., 1997. Return autocorrelation and institutional investors. Journal of Financial Economics 46, 103–131.
Stickel, S., 1992. Reputation and performance among security analysts. Journal of Finance 47, 1136–1811.
Subramanyam, K.R., 1996. Uncertain precision and price reactions to information. The Accounting Review 71, 207–220.
Timmermann, A., 1993. How learning in financial markets generates excess volatility and predictability in stock prices. Quarterly Journal of Economics 108, 1135–1145.
Timmermann, A., 1996. Excess volatility and predictability of stock prices in autoregressive dividend models with learning. Review of Economic Studies 63, 523–557.
Trueman, B., 1994. Analyst forecast and herding behavior. Review of Financial Studies 7, 97–124.
Walther, B., 1997. Investor sophistication and market earnings expectations. Journal of Accounting Research 35 (2), 157–179.
White, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48, 817–838.
Wooldridge, J., 2002. Econometric Analysis of Cross Section and Panel Data. The MIT Press, Cambridge, MA.