Can Relative Performance Compensation Explain Analysts’ Forecasts of Earnings?¹

Edward Kutsoati²    Dan Bernhardt³

First version: May 1999
This version August 3, 2000

¹We are grateful to the Institutional Brokers Estimate System (I/B/E/S), a service of I/B/E/S International Inc. for providing analysts’ forecasts of earnings per share. We thank Lucy Ackert, George Deltas, Ted Juhl, Gregor Smith, Brett Trueman, Mike Waldman, Ján Zábojnik and an anonymous referee for very helpful suggestions. We would also like to thank Vaclav Polasek for his help with programming. The usual caveat applies.

²Department of Economics, Tufts University, Medford, MA, 02155
³Department of Economics, 463 Commerce West, University of Illinois, Champaign, IL. 61820
Abstract

This paper derives and tests the implications of relative performance incentives for the forecasts of financial analysts. If, in addition to compensation for absolute forecast accuracy, analysts are compensated on the basis of how their forecasts compare with those of other analysts, then earlier forecasts affect later announcements. The theoretical predictions regarding the direction of bias are tested using data on individual analysts' forecasts of earnings per share (EPS). We find very strong evidence that the last analyst to report a forecast strategically selects a contrarian forecast that overshoots the consensus (mean) forecast in the direction of his private information. In particular, we provide evidence that the bias in the last analyst's forecast is chosen strategically, and is not due to the analyst myopically ignoring information in the forecasts of others. We also find evidence that investors do not unravel biases in the forecasts of firms followed by fewer analysts.
1 Introduction

The forecasts of firms’ earnings made by financial analysts serve as a major source of information for brokerage houses and money management funds, who rely on the accuracy of these forecasts in providing services for their clients. This paper explores how a security analyst’s compensation affects his reported forecasts. In particular, we posit that an analyst’s compensation may be tied both directly to absolute forecast accuracy and to how the accuracy of his forecast compares with those of other analysts; that is, the analyst’s relative forecast accuracy. We derive the associated theoretical implications for forecasts, and provide extensive empirical evidence that relative performance compensation is crucial for understanding analyst forecasts.

Relative performance compensation emerges naturally in environments where potential employers and clients are trying to infer analysts’ abilities from their forecasts. Firms with better analysts win more clients, generating more brokerage commissions and profits. Because the relative accuracy of an analyst’s forecast helps firms and investors unravel the analyst’s forecasting ability, competition among firms for better analysts leads to relative performance compensation. Indeed, brokerage houses might want to design performance compensation contracts that induce analysts to over-emphasize their private information and hence reveal their ability more quickly. In fact, there is extensive evidence suggesting that analysts and their employers do care about the relative accuracy of analysts’ forecasts.

We allow for general functional forms of relative performance compensation and derive the implications of the different curvatures for the forecast of the last analyst to report earnings. If analysts are compensated on the basis of relative performance, then when reporting their forecasts, analysts will balance their aims of minimizing forecast errors against looking good relative to other forecasters. For example, the fear of being fired due to bad relative under-performance might make the analyst averse to the possibility of substantially under-performing other analysts, implying concave relative performance compensation on this range. Alternatively, being “right”, when everyone else is “wrong”, might lead to a sig-

---

1Dugar and Nathan (1995) discuss other incentives that may impact on analysts’ forecasts. Lin and McNichols (1998) document evidence suggesting that analysts working for investment banks that underwrite equity offerings tend to issue more favorable recommendations.

2In a labor market context, Holmstrom (1982) explored how performance contracts distort the actions of workers when the market is trying to learn workers’ abilities. See, also Scharfstein and Stein (1990), and Zweibel (1995). Chevalier and Ellison (1997) document empirically how mutual fund manager’s investments evolve over time, presumably reflecting learning by investors who direct investment funds according to performance.

3The annual All-Star Analysts Survey (conducted by the Wall Street Journal and Zacks Investment Research) and the All-American Research Team, published by the Institutional Investor magazine rank security analysts on the basis of their earnings forecasting skills. Analysts’ “skills” are measured by comparing the accuracy of their earnings estimates against those of other analysts following the same firm. Stickel (1990, 1992) documented that analysts’ salaries are often tied to their rank in the II poll. In the 1970s, Merrill Lynch hired top-ranked analysts using the II poll as a guide (see “Who Really Moves The Market,” Fortune, Oct. 27, 1997).

4Mikhail et al. (1999) finds that analysts who are relatively less accurate than their peers are more likely to
nificant increase in the number of clients desiring the analyst’s services, implying a convex compensation relationship over some range of relative forecasts (see e.g., Henry 1989, Laster et al. 1999).

We consider an environment in which forecasts are released sequentially and later analysts observe earlier forecasts before reporting their forecasts. We show that if relative performance compensation concave or anti-symmetric concave (so that the marginal gain (loss) to being better (worse) than the consensus is small), then the last reporting analyst strategically biases his report toward the consensus. This would cause forecasts to be “clustered” more closely around the consensus than analysts’ information would suggest. If, instead, relative performance compensation is a convex or anti-symmetric convex (i.e., reflecting both large bonuses for being “right” when others are “wrong”, and the possibility of being fired if one’s performance is significantly worse than others) function of relative performance, then the last analyst strategically biases his forecast in the direction of his private information, away from the consensus. We then show that this strategic bias is greater for firms followed by fewer analysts, so that the analyst is more uncertain about true earnings given his information.

Data on individual analysts’ forecasts of earnings per share (EPS) permit us to identify the strategic bias in the last forecast. We find overwhelming empirical evidence that the last analyst strategically exaggerates his private information by announcing a contrarian forecast, overshooting true EPS away from the consensus (mean) forecast. If the last forecast is unbiased, then the last forecast should be as likely to exceed actual EPS as fall short, independently of whether the last forecast exceeded or fell short of the consensus. However, we find the last forecast overshot EPS away from the consensus 62.5% of the time. If the last forecast fell short of the consensus, it also fell short of true earnings 68% of the time; and if the last forecast overshot the consensus, it also overshot true EPS 56% of the time.

A referee observed that these results are also consistent with a myopic last analyst who fails to incorporate the information contained in the consensus forecast into his own forecast. In such a scenario, the last forecast would tend to overshoot earnings when it exceeded the consensus, as the analyst ignores the more negative information in the consensus, and would tend to fall short when his forecast was less than the consensus. We test this alternative hypothesis, exploiting the observation that the consensus is more accurate when there are more analysts. The evidence is not consistent with a myopia explanation: the average bias in the last forecast, given the consensus, falls with the number of forecasts. In short, the data are consistent only with the hypothesis of convex relative performance compensation and strategic analyst behavior, and sharply inconsistent with concave or linear relative compensation or analyst myopia.

---

5The consensus forecast and the outstanding mean forecast are used interchangeably in this paper.
We then test to see whether investors correctly unravel the bias in the last analyst's forecast. In particular, we investigate whether the difference between the last forecast and the consensus can predict the market's response to the firm's earnings announcement. If investors do not unravel the bias, then if the last forecast exceeded (lagged) the consensus, they will systematically over-predict (under-predict) earnings and will be disappointed (excited) by the earnings announcement, leading to a negative (positive) excess returns. We find evidence that investors in fact do not unravel the bias for firms that are followed by relatively few analysts. That is, investors appear to be fooled where our theory predicts the bias should be greatest, and where investors have the fewest sources of information.

In our environment, forecasts are biased, not because analysts are irrational as De Bondt and Thaler (1990) and Abarbanell and Bernard (1992) argue, but rather because the market provides them incentives to rationally distort their forecasts. Keane and Runkle (1998) argue that “Since financial analysts’ livelihoods depend on the accuracy of their forecasts . . . we can plausibly argue that these numbers accurately measure the analysts’ expectations.” But this argument follows only when financial analysts are compensated solely on the basis of absolute accuracy — when the market uses relative forecast accuracy to glean information about an analyst’s ability, analysts will also be rewarded for relatively more accurate forecasts.

Recent studies that explored how economic environments can lead to biased financial forecasts include Trueman (1994). In a model based on heterogeneous forecasting abilities, Trueman (1994) showed that when investors attempt to infer ability from forecasts, less able analysts will tend to mimic the forecasts of others, leading them to herd toward the consensus forecast. In a similar model, Trueman (1990) argues that an analyst will be reluctant to revise her forecast upon receiving new information since such action will mean the initial forecast is less precise and the analyst will be perceived as incompetent. Ehrbeck and Waldmann (1996) investigated this phenomenon using agency models in which less able forecasters balance their aims of minimizing forecast errors and making small revisions in their forecasts like more able forecasters. However, they find that forecasters of interest rates on 91-day U.S. Treasury bills revise their forecasts by too much and that large revisions tend to be correlated with large forecast errors6. This finding can be rationalized by convex relative performance compensation for forecasters.

Laster, Bennett and Geoum (1999) assume that analysts are compensated according to mean squared error plus a bonus for the most accurate forecast among all forecasters. They find that independent forecasters of GDP growth are more likely to report outlying forecasts. Lamont (1995) and Hong et al. (2000) find evidence that such incentives may vary over time: analysts who have been in the business longer tend to make bolder and less accurate forecasts.

---

6Kandel and Pearson (1995) also found evidence of over-revision by security analysts and showed that analysts' differential interpretation of public signals could explain the relation between volume of trade and stock returns around public announcements.
Our empirical results appear inconsistent with the theoretical clustering predictions of Trueman (1994) and contrast sharply with the empirical findings of Keane and Runkle (1998). Keane and Runkle, after accounting for the fact that analysts receive information from common sources, argue that “The evidence strongly supports the view that professional stock market analysts make rational (unbiased) forecasts of earnings per share.” The difference between our findings and Keane and Runkle’s is startling. Importantly, we design a frequency test that is robust to arbitrary correlation in information across analysts, so that correlation cannot explain the differences in our findings. There are many plausible reasons for why Keane and Runkle fail to uncover our relationship. Importantly, Keane and Runkle do not distinguish among analysts. With convex compensation, robust predictions about the nature of bias in analyst’s forecasts only obtain for the last forecaster, because earlier forecasters must take into account how their forecasts affect subsequent forecasts. Also, self-selection may mean that there is something special about the last forecaster — for example, the last forecaster may be the one with the most to gain from differentiating himself from others. We only document the bias in the last forecast and are unconcerned with identifying the properties of who is last. Finally, Keane and Runkle consider forecasts for only 21 firms each of which is very heavily followed, whereas our sample has 237 firms with varying degrees of analyst interest. A likely reason for the different findings is that Keane and Runkle consider only very heavily followed firms, where our theoretical analysis predicts the bias should be the smallest, and which our empirical documents.

The paper is organized as follows. After presenting a simple model of sequential forecasting in section 2, we analyze the effect of relative performance compensation for announced forecasts in section 3. In section 4, we describe the data, develop testable hypotheses and present our empirical findings. Section 5 concludes. All proofs are in an appendix.

2 A Model of Sequential Forecasting

Consider \( I \) security analysts who sequentially report forecasts of the earnings per share, \( \theta \), of a firm. We take the sequence of release of forecasts as given and index analysts by order in which they report their forecast\(^a\): \( j = \{1, 2, \ldots, I\} \). The prior distribution of earnings per share, \( \theta \), is known, but the actual value is realized after the last forecast. Denote the \( j \)th analyst’s forecast by \( f_j \). Later analysts observe the existing forecasts before reporting their own forecasts. In practice, forecasts are produced and become publicly available in an almost continuous fashion. For example, subscribers (and contributors) to the Institutional Brokers

\(^7\)In an otherwise unrelated study, Cooper et al. (1999) develop and test procedures for identifying lead analysts.

\(^a\)We do not model the strategic choice of when to release a forecast. Our analysis focuses on the last forecaster and is consistent with the possibility that analysts may differ in the form of relative compensation, and that the forecaster who goes last does so because he gains the most from distinguishing his forecast from others.
Estimate System (I/B/E/S) EXPRESS receive the latest forecasts of earnings made by a panel of security analysts. This product is delivered electronically and is generally accompanied by the mean (or “consensus”) and other summary statistics, so that analysts have a good idea of the market consensus, even if they do not know what particular analysts are reporting. Forecasts are also disseminated through the media, newsletters, webpages (e.g. Yahoo), etc.

Analyst $j$ is compensated according to both his absolute performance, measured by individual error; and his relative performance, measured by the difference between own error and the error in the consensus (mean) forecast:

$$w_j = \bar{w} - \lambda |f_j - \theta| + R(|f_m - \theta| - |f_j - \theta|),$$

where $\bar{w} > 0$ is a fixed reward, $f_m$ is the mean of all other forecasts, excluding $f_j$, and $\lambda > 0$ is the weight on absolute performance, $|f_j - \theta|$, in an analyst’s compensation. The relative compensation component, $R(|f_m - \theta| - |f_j - \theta|)$, is assumed to be increasing in relative performance: i.e., $R'(\cdot) > 0$. All of our theoretical findings, save one, would extend were we to relax the linear structure on absolute performance compensation and assume only that absolute performance compensation is strictly monotonically increasing in accuracy. Without loss of generality, we normalize $R(0)$ to be zero.

Analyst $j$ observes all preceding forecasts of other analysts in addition to a private signal, $s_j$ of the firm’s earnings per share: $j$’s information set, $\Omega_j$, includes at least $\{f_1, f_2, \ldots, f_{j-1}, s_j\}$. Analyst $j$’s signal is given by

$$s_j = \theta + \eta_j,$$

where $\eta_j$ is a symmetrically and independently distributed random variable with mean zero. Note that we assume that signal errors are independent; later we discuss how correlated signals affect our analysis. Given this information set, analyst $j$’s posterior beliefs about $\theta$ are captured by a density $g(\theta|\Omega_j)$ that is symmetrically distributed about its posterior mean, $\hat{\theta}_j$, and strictly positive on its connected support. Let $G(\theta|\Omega_j) = \int_{x<\theta} g(x|\Omega_j) \, dx$.

Except for the case where compensation functions are linear, we characterize only the forecast of the last analyst. Our analysis of the last analyst permits a variety of possible informational structures, including the possibility that new informational arrives over time. For example, we could let $\theta = \sum_{i=0}^{L+1} \theta_i$, where innovations $\theta_i$, are mean zero and independently and symmetrically distributed so that analyst $j$’s posterior is symmetrically distributed around $\hat{\theta}_j = \sum_{i=0}^{j} \theta_i$. This formulation allows later forecasters to see more innovations to earnings than earlier forecasters.

$Larger games that give rise to absolute and relative performance compensation include the learning about analyst ability discussed in the introduction, tournaments to overcome moral hazard among analysts, etc. We only derive and test the theoretical implications of different induced preferences, which, in turn, impose restrictions on possible larger games/primitive preferences.

\textsuperscript{10}Proposition 1(ii), which we do not test, requires more structure.
The information environment would have to be specified more carefully in order to characterize the forecasts of analysts other than the last. In particular, were inferences about $\theta$ by later forecasters affected by the forecasts of earlier analysts then we would have to worry about the incentives of earlier analysts to try to “fool” subsequent forecasters into believing that they observed a different private signal, thereby influencing their posteriors and thus their announcements. This caveat is irrelevant for the last analyst, as he has no other forecaster to “fool.”

3 Compensation Structure and the Pattern of Forecasts

To analyze how relative performance compensation affects forecast patterns, we consider, as a benchmark, a linear relative performance compensation function. We show that if $R(\cdot)$ is linear, then an analyst’s forecast, including the last, is unaffected by those made by others. Even though the reports of other analysts still affect an analyst’s compensation — more accurate forecasts by others reduce an analyst’s compensation — each analyst maximizes expected compensation by reporting his mean squared error (MSE) minimizing forecast: $f_j = \hat{\theta}_j$, $\forall j = 1, 2, \ldots, n$.

**Lemma 1** If an analyst’s relative compensation function is linear, then each analyst’s equilibrium forecast equals the posterior mean of his beliefs about $\theta$: $f_j = \hat{\theta}_j$, $\forall j = 1, 2, \ldots, n$.

We now show that when $R''(\cdot) \neq 0$, the outstanding forecasts (or the consensus of earlier forecasts) of other analysts will affect the last agent’s forecast: Analysts balance their aims of minimizing forecast errors against looking good after the outcome is realized. We first detail the set of feasible reports by the last analyst given his information and the consensus forecast.

**Lemma 2** If the consensus forecast is less than the last analyst’s expectation of earnings, i.e., $f_m < \hat{\theta}_L$, then his report will exceed the consensus, $f_L > f_m$. Conversely, if $\hat{\theta}_L < f_m$ then $f_L < f_m$.

Lemma 2 ensures that if the last analyst’s expectation of $\theta$ exceeds $f_m$, then so will his forecast. For every $f'_L < f_m$, since his posterior is symmetrically distributed about $\hat{\theta}_L$, there exists some forecast exceeding $f_m$ that dominates $f'_L$ in terms of expected reward and offers the same expected relative performance. Conversely, if he expects earnings to be less than $f_m$, then he reports a forecast less than $f_m$.

Having eliminated dominated strategies, we can now characterize the pattern of forecasts by the last analyst. Depending on the nature of the relative compensation function, the last analyst will choose either to “under-emphasize” or to “exaggerate” his information.
If $R''(|f_m - \theta| - |f_L - \theta|) < 0$, then analyst’s compensation is concave in relative performance. In this case, the last analyst will be averse to locating too far from the consensus, lest his forecast be inaccurate. He is willing to forego some of the compensation that derives from smaller absolute errors in order to be closer to the consensus: If he is wrong, then there is comfort in ‘numbers’ in having everyone else be wrong, too. The consequence is that the last analyst puts less weight on his information (or private signal) and biases his forecast, $f_L$, towards the consensus forecast, $f_m$:

**Proposition 1** Suppose $R(\cdot)$ is concave. Then
(i) If $f_m < \hat{\theta}_L$ then $f_m < f_L < \hat{\theta}_L$. Conversely, if $f_m > \hat{\theta}_L$ then $f_m > f_L > \hat{\theta}_L$.
(ii) $0 < \frac{\partial \hat{\theta}_L - f_L}{\partial (\hat{\theta}_L - f_m)} < 1$: that is, the bias in $f_L$ increases in $(\hat{\theta}_L - f_m)$.

Proposition 1 states that the last analyst will bias his forecast in the direction of the consensus report: he gets a higher expected payoff by reporting such a forecast since the reward from relative performance more than compensates for losses in payoff due to a worse absolute performance. Hence, if analysts’ payoffs are a concave function of relative performance, then forecasts can become “clustered” even if individual private information suggests otherwise. Indeed, as $f_m$ diverges further from the last analyst’s expectation, $\hat{\theta}_L$, his forecast, $f_L$, diverges further from $\hat{\theta}_L$ but at a rate less than one.11 Equivalently, if (say) the consensus forecast is revised, this will cause the last analyst to revise his forecast in the direction of revision even though his beliefs about earnings are unaltered.

Proposition 1 forms the basis of our predictions about a concave $R(\cdot)$. The conservative bias introduced in the forecast implies that if $f_L$ exceeds $f_m$, it is more likely that the last forecast will be less than $\theta$. Conversely, if $f_L$ is less than $f_m$, then $f_L$ is more likely to overestimate actual earnings. This generates a relationship between the pattern of forecasts and the direction of error in the last forecast.

Alternatively, compensation may be convex, $R''(|f_m - \theta| - |f_L - \theta|) > 0$, perhaps because one or two accurate forecasts that depart from the consensus could draw a substantial clientele. The next proposition details that convex relative performance compensation causes the last analyst to “exaggerate” his information by reporting an $f_L$ that overshoots $\hat{\theta}_L$ away from $f_m$.

**Proposition 2** Suppose that $R(\cdot)$ is convex. Then if $f_m < \hat{\theta}_L$, then $f_m < \hat{\theta}_L < f_L$. Conversely, if $f_m > \hat{\theta}_L$, then $f_m > \hat{\theta}_L > f_L$.

If the relative compensation function is convex, the last analyst has an incentive to bias his forecast to “locate” away from $f_m$: a relatively accurate forecast yields a much higher payoff.

---

11Absent specific parameterizations, similar predictions do not obtain for convex relative performance compensation.
than the losses incurred if his forecast is less accurate. This induces the last analyst to take forecasting risks by reporting contrarian forecasts. In particular, the last analyst will report an over-optimistic forecast of the firm’s EPS if \( \hat{\theta}_L > f_m \), and an over-pessimistic forecast if \( \hat{\theta}_L < f_m \). That is, \( f_L \) is more likely to be greater (less) than the actual earnings if \( f_L \) is greater (less) than \( f_m \).

As discussed earlier, characterizations of forecasts by analyst \( j \), \( 1 < j < L \) are more difficult because even when \( j \)'s forecast does not affect their inferences, it still affects the relevant consensus forecast and hence their announcement.\(^{12}\)

**Anti-symmetric Relative Compensation Functions**

The relative compensation function is said to be anti-symmetric about zero if:

\[
R(|f_m - \theta| - |f_j - \theta|) = \begin{cases} 
R(|f_m - \theta| - |f_j - \theta|), & \text{if } |f_m - \theta| > |f_j - \theta| \\
-R(|f_j - \theta| - |f_m - \theta|), & \text{if } |f_m - \theta| < |f_j - \theta|,
\end{cases}
\]

with \( R(0) = 0 \) and \( R'(\cdot) > 0 \). That is, analysts are penalized for bad relative performance just as much as they are rewarded for a good relative performance. The relative performance compensation function is anti-symmetric-concave if \( R(\cdot) \) is a concave function of positive relative performances, and anti-symmetric-convex if \( R(\cdot) \) is a convex function of positive relative performances.

One can show that the last analyst has an incentive to issue an exaggerated forecast if the relative performance function is “anti-symmetric convex”; and an incentive to bias his forecast toward the consensus if the relative performance function is “anti-symmetric concave.” That is, if relative performance compensation is anti-symmetric-concave, the last analyst “locates” closer to \( f_m \); and he “locates” further away from \( f_m \) if \( R \) is anti-symmetric-convex. The corresponding lemmas and propositions are proved in the same way. The key is to exploit the symmetry in the analysts’ beliefs about \( \theta \) and the fact that \( R'(z) = R'(-z) \) for all \( z \). The intuition is that relative to the consensus, the last analyst anticipates being right more often than being wrong, so that his forecast reflects more heavily the curvature assumptions on the payoffs for good relative performances.

**Example:** Let \( \theta \) be uniformly distributed on \([-4 + \hat{\theta}_L, 4 + \hat{\theta}_L] \), \( \lambda = \frac{1}{4} \) and \( f_m = 0 \). Figure 1(a) shows the equilibrium forecast of the last analyst as a function of \( \hat{\theta}_L \) when \( R \) is given by the

\(^{12}\)For example, consider 3 analysts. Suppose that \( f_1 < \hat{\theta}_2 = \hat{\theta}_3 \), and that relative compensation is convex. Analyst 2 wants to bias his forecast away from \( f_1 \) only if that bias raises \( |f_2 - 0.5(f_1 + f_3)\). But analyst 3 will bias his forecast away from \( 0.5(f_1 + f_2) \), so that if analyst 2 biases his forecast further away from \( f_1 \), it may cause analyst 3 to further bias his forecast away from \( f_1 \). But this may reduce \( |f_2 - 0.5(f_1 + f_3(\hat{\theta}_3))| \). Clearly, any predictions will depend finely on the particular parameterizations.
Figure 1: Anti-symmetric concave $R$ vs. anti-symmetric convex $R$

Optimal $f_L$ plotted against $\theta_L$, for (a) anti-symmetric concave $R$ and (b) anti-symmetric convex $R$, when $f_m = 0$, $\lambda = \frac{1}{2}$, and $\theta$ is uniformly distributed on $[-4 + \hat{\theta}_L, 4 + \hat{\theta}_L]$. The 45° line represents $f_L = \hat{\theta}_L$.

anti-symmetric-concave function

\[
R(z) = \begin{cases} 
\sqrt{z}, & \text{if } z > 0; \\
-\sqrt{-z}, & \text{if } z < 0.
\end{cases}
\]

where $z = |f_m - \theta| - |f_L - \theta|$ is the relative forecast error. The last analyst biases his forecast to “locate” closer to $f_m = 0$. Figure 1(b) illustrates the last analyst’s incentives to “over-emphasize” his information when $R(z)$ is given by the anti-symmetric-convex function,

\[
R(z) = \begin{cases} 
z^2, & \text{if } z > 0; \\
-z^2, & \text{if } z < 0.
\end{cases}
\]

The equilibrium forecast of the last analyst is biased away from $f_m$.

4 Empirical Analysis

4.1 Data and Sample Selection

We next test the theoretical predictions of the model and investigate the market’s recognition of any bias in the last forecast. Individual analysts forecasts, forecast publication dates and the actual EPS reported by the firms are provided by the Institutional Brokers Estimate System (I/B/E/S). Data on daily returns, share prices and shares outstanding are taken from the Center for Research in Security Prices (CRSP).

We focus on one-quarter ahead forecasts in the I/B/E/S file between 1988-I and 1995-IV for all firms with fiscal year-end in December that are followed over that period by at least 2
analysts, have actual earnings available and have daily returns on the CRSP tapes. We choose one-quarter ahead forecasts because there are more observations per firm and fewer forecast revisions than longer-term forecasts. We do not consider forecasts prior to 1988, because of the lag between the date of an analyst's forecast and the date the forecast was entered in the I/B/E/S database during this period. O'Brien (1988) reported an average publication lag of 34 trading days over the period 1975–1982. After 1988, improvements in technology reduced this turn-around time to less than 24 hours (see I/B/E/S Research Bibliography). Our sample selection ensures that the publication dates are close to the actual dates at which analysts released their forecasts and, more importantly, the last analyst observes the earlier forecasts.

This leaves us with a sample size of 237 firms with 7584 firm-quarters. Analysts normally revise their initial forecasts for a firm-quarter before the actual earnings are announced by the firm. Hence, the total number of forecasts in any given firm quarter will include both old forecasts and their revisions. For the most part, we focus on only the most recent forecast reported by all analysts following the firm in any given firm-quarter.

Our theory imposes predictions only on the last analyst. If multiple analysts provide forecasts on the last day, we both do not know which one was 'last', nor what their information is. For firms followed by many analysts, the forecasts of analysts on the last date are unlikely to be affected significantly by beliefs about what other analysts are reporting on that date. Still, to avoid taking stands, we restrict our focus to firm-quarters where a single forecast was reported on the last date. This reduces the number of observations (firm-quarters) by 1702. We note, however, that our empirical findings are qualitatively unaffected if we include observations with multiple forecasts on the last day, and interpret the last forecast as the mean of those forecasts.

For the purpose of analyzing market reaction to the last forecast, we require all dates of announcements be known and in the required order — publication dates of forecasts must precede the date of announcement of actual EPS. For some firm-quarters, the date of announcement of actual EPS is either missing or the publication date of the last forecast is later than date of release of actual EPS. In either case, we delete that firm-quarter. Finally, we deleted observations for which the error in the outstanding mean forecast exceeds $10 as these are more likely to have resulted from data-entry error. Table 1 summarizes the final sample and shows the type of industries/sectors represented in the sample.

Table 2 gives the distribution of analyst coverage, total forecasts reported and the average number of forecasts per analyst across all firms-quarters. An average of 9 analysts follow a firm each quarter, and the average number of forecasts per analyst is about 1.31, suggesting

13For example, in the third quarter of 1988, Pan Energy Corporation reported an EPS of $1.98 whilst the mean forecast from the I/B/E/S file for the same firm-quarter was $50.35. O'Brien (1988), Kandel and Pearson (1995) and Lim (2000) deleted similar suspected data-entry errors.
Table 1: Sample Selection

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
<th>Number of Firms</th>
<th>Firm-quarters</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Firms with at least 2 analysts following, fiscal year end in December and actual EPS available for all quarters in 1988-1995</td>
<td>279</td>
<td>8928</td>
</tr>
<tr>
<td>2</td>
<td>Firms in (1) with daily returns on CRSP tapes</td>
<td>237</td>
<td>7584</td>
</tr>
<tr>
<td>3</td>
<td>Firms-quarters in (2) with at least 2 analysts reporting their last forecast at different dates</td>
<td>237</td>
<td>7571</td>
</tr>
<tr>
<td>4</td>
<td>Sample size after deleting observations with multiple forecasts on last forecast date</td>
<td>237</td>
<td>5869</td>
</tr>
<tr>
<td>5</td>
<td>Firm-quarters with publication date of last forecast in correct order (i.e., before date of actual EPS)</td>
<td>237</td>
<td>5842</td>
</tr>
<tr>
<td>6</td>
<td>Final sample after deleting observations with absolute mean forecast error exceeding $10</td>
<td>237</td>
<td>5832</td>
</tr>
</tbody>
</table>

Sample of firms by Sectors

<table>
<thead>
<tr>
<th>Sector</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Finance</td>
<td>34</td>
</tr>
<tr>
<td>Health</td>
<td>13</td>
</tr>
<tr>
<td>Consumer Non-Durables</td>
<td>20</td>
</tr>
<tr>
<td>Consumer Services</td>
<td>19</td>
</tr>
<tr>
<td>Consumer Durables</td>
<td>20</td>
</tr>
<tr>
<td>Energy</td>
<td>22</td>
</tr>
<tr>
<td>Transportation</td>
<td>10</td>
</tr>
<tr>
<td>Technology</td>
<td>13</td>
</tr>
<tr>
<td>Basic</td>
<td>41</td>
</tr>
<tr>
<td>Capital/Machinery</td>
<td>22</td>
</tr>
<tr>
<td>Utility</td>
<td>31</td>
</tr>
</tbody>
</table>

Table 2: Distribution of analysts and forecasts

<table>
<thead>
<tr>
<th></th>
<th>min</th>
<th>max</th>
<th>25</th>
<th>50</th>
<th>75</th>
<th>Mean</th>
<th>Std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total sample:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>analysts</td>
<td>2</td>
<td>27</td>
<td>6</td>
<td>9</td>
<td>12</td>
<td>9</td>
<td>4.50</td>
</tr>
<tr>
<td>total forecasts</td>
<td>2</td>
<td>67</td>
<td>7</td>
<td>11</td>
<td>17</td>
<td>13</td>
<td>7.33</td>
</tr>
<tr>
<td>fcts. per analyst</td>
<td>1</td>
<td>3.3</td>
<td>1.11</td>
<td>1.25</td>
<td>1.44</td>
<td>1.31</td>
<td>0.27</td>
</tr>
<tr>
<td>1st Qtr:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>analysts</td>
<td>2</td>
<td>30</td>
<td>5</td>
<td>8</td>
<td>11</td>
<td>8</td>
<td>4.37</td>
</tr>
<tr>
<td>total forecasts</td>
<td>2</td>
<td>50</td>
<td>6</td>
<td>9</td>
<td>14</td>
<td>11</td>
<td>6.93</td>
</tr>
<tr>
<td>fcts. per analyst</td>
<td>1</td>
<td>2.3</td>
<td>1</td>
<td>1.21</td>
<td>1.40</td>
<td>1.26</td>
<td>0.25</td>
</tr>
<tr>
<td>2nd Qtr:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>analysts</td>
<td>2</td>
<td>27</td>
<td>6</td>
<td>8</td>
<td>11</td>
<td>9</td>
<td>4.36</td>
</tr>
<tr>
<td>total forecasts</td>
<td>2</td>
<td>47</td>
<td>7</td>
<td>11</td>
<td>15</td>
<td>12</td>
<td>6.88</td>
</tr>
<tr>
<td>fcts. per analyst</td>
<td>1</td>
<td>3.3</td>
<td>1.08</td>
<td>1.25</td>
<td>1.43</td>
<td>1.29</td>
<td>0.27</td>
</tr>
<tr>
<td>3rd Qtr:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>analysts</td>
<td>2</td>
<td>28</td>
<td>7</td>
<td>9</td>
<td>13</td>
<td>10</td>
<td>4.51</td>
</tr>
<tr>
<td>total forecasts</td>
<td>2</td>
<td>67</td>
<td>8</td>
<td>12</td>
<td>17</td>
<td>13</td>
<td>7.51</td>
</tr>
<tr>
<td>fcts. per analyst</td>
<td>1</td>
<td>2.9</td>
<td>1.11</td>
<td>1.28</td>
<td>1.50</td>
<td>1.33</td>
<td>0.28</td>
</tr>
<tr>
<td>4th Qtr:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>analysts</td>
<td>2</td>
<td>32</td>
<td>7</td>
<td>10</td>
<td>14</td>
<td>10</td>
<td>4.47</td>
</tr>
<tr>
<td>total forecasts</td>
<td>2</td>
<td>47</td>
<td>9</td>
<td>13</td>
<td>19</td>
<td>14</td>
<td>7.52</td>
</tr>
<tr>
<td>fcts. per analyst</td>
<td>1</td>
<td>2.8</td>
<td>1.14</td>
<td>1.28</td>
<td>1.50</td>
<td>1.35</td>
<td>0.28</td>
</tr>
</tbody>
</table>

NOTES:

a. N represents the number of observations (firm-quarters). All firm-quarters have at least two forecasts reported by distinct analysts on different dates.
b. Total forecasts is the total number of forecasts (including both old forecasts and their revisions) reported by all analysts in a firm-quarter.
that analysts infrequently revise their one-quarter ahead forecasts. Although the distributions of analysts and forecasts between quarters are similar, on average the number of analysts and forecasts per firm-quarter increases slightly from the first to the fourth quarter — more forecasts are reported as the fiscal year-end approaches. Forecasts are also more likely to be revised as the fiscal year-end approaches.

4.2 Evaluating Bias in the Last Forecast

4.2.1 Frequency tests

We now develop and test simple hypotheses regarding the direction of bias in last analyst's forecast. The objective of our empirical work is not to estimate the relative compensation function, but rather to determine which form of compensation better describes analysts' behavior. Once the nature of the relative compensation is determined, several empirical implications of the effect of relative compensation on analysts' forecasts can be explored.

The empirical analysis uses only the last forecast reported by each analyst in a firm-quarter. In practice, an analyst may receive new information at any time and revise his forecast upon the arrival of new information. Hence, we give more attention to an analyst's last forecast, because more recent forecasts contain more information and are more precise than older forecasts (O'Brien 1988). In any event, quarterly forecasts are rarely revised so this procedure should not alter the results. The empirical issue is what is the “appropriate” consensus forecast with which the last analyst's forecast is being compared. Choosing the wrong measure for “consensus” only reduces our chances of finding that the (very) last forecast is systematically affected by the outstanding consensus. Using the mean of the forecasts of other analysts to construct the consensus seems appropriate because the mean forecast is highlighted in the I/B/E/S database and by other researchers.

We use the following notation: For firm quarter $\tau$, let $\theta_\tau$ be the true earnings reported by the firm. Denote the last reported forecast of the firm's earnings by $f_{L\tau}$ and the outstanding mean forecast (the mean of the forecasts excluding $f_{L\tau}$), by $f_{m\tau}$.

Our analysis shows that if $R$ is linear in relative performance, then analysts report their true expectation. That is, since the last forecast is not affected by the consensus, the probability that the last forecast exceeds actual earnings should not vary with whether or not the last forecast exceeded the public information consensus forecast. To test this hypothesis, and the alternative hypotheses developed below, we first employ a simple frequency test. To construct the test statistic, we group the data into mutually exclusive events according to the realization of the true earnings relative to $f_{L\tau}$ and $f_{m\tau}$. Define $\delta^+_{\tau}$ and $\delta^-_{\tau}$ as

$^{14}$It is worth noting that not all analysts contribute their forecasts to the I/B/E/S database, so that in actuality there will be more analysts per firm than documented here.
The events $\delta^+_r$ and $\delta^-_r$ are illustrated in Figure 2: the top panel shows the event $\delta^+_r$ and the lower line segment depicts $\delta^-_r$. While earnings per share are continuously distributed, in practice, both earnings per share and forecasts by analysts are reported in cents, so that some unobserved rounding occurs. This creates a problem with inference whenever the last forecast equals the consensus or reported EPS. Without rounding, both of these events should happen with probability zero. With rounding, we cannot discern whether the last analyst’s expectation exceeds the consensus or not; and whether it exceeds true EPS or not. Consequently, for our frequency tests, we must drop an observation when either $f_{Lr} = \theta_r$ or $f_{Lr} = f_{mr}$. Accordingly, define $\rho_r$ as $\rho_r = 1$ if $f_{Lr} \neq \theta_r$ and $f_{Lr} \neq f_{mr}$; and $\rho_r = 0$ otherwise (i.e., $\rho_{rl} = 1$ if and only if the observation is counted in our sample). Our test statistic is

$$T = \frac{\sum_r \delta^+_r + \sum_r \delta^-_r}{\sum_r \rho_r}.$$  

The test statistic, $T$, is equal to one-half under the null hypothesis that $f_{Lr}$ is unbiased.\(^{15}\)

Under the alternative hypothesis that $R(\cdot)$ is a convex, or anti-symmetric convex, function of relative performance, $T < 0.5$. That is, were $R(\cdot)$ convex, the last forecaster would bias his forecast past $\theta_r$ away from $f_{mr}$. Hence, if $f_{Lr} > f_{mr}$, then $f_{Lr}$ would exceed the last forecaster’s expectation of $\theta_r$, so that more often than not, $\theta_r$ should fall short of $f_{Lr}$. Conversely, if $f_{Lr} < f_{mr}$, more often than not, $\theta_r$ should exceed $f_{Lr}$. Therefore, convex relative compensation implies $T < 0.5$.

Finally, under the alternative hypothesis that $R(\cdot)$ is a concave, or anti-symmetric concave function of relative performance, $T > 0.5$. That is, concave relative compensation would cause the last forecaster to bias his forecast toward $f_{mr}$. Hence, if $f_{Lr}$ is greater (less) than $f_{mr}$, then $f_{Lr}$ is more likely to be less (greater) than realized earnings, so $T > 0.5$.

Our frequency tests have several advantages:

- They non-parametric and unrelated to the scale of errors across firms: no assumption is

\(^{15}\)If the frequency with which the last forecast exceeded the consensus differed significantly from one-half, then the test statistic would have to be modified in order for it to be robust against arbitrary correlation in information amongst analysts,

$$T' = \frac{\sum_j \gamma^+_j \sum_r \delta^+_r + \sum_j \gamma^-_j \sum_r \delta^-_r}{\sum_r \rho_r}.$$  

Here, $\gamma^+_j = 1$ if $f_{Ljr} > f_{mr}$ and $f_{Ljr} \neq \theta_r$, and $\gamma^-_j = 0$ otherwise; and $\gamma^-_j = 1$ if $f_{Ljr} < f_{mr}$ and $f_{Ljr} \neq \theta_r$, and $\gamma^-_j = 0$ otherwise.
made about the structure of the relationship between the error in the last forecast and the difference in $f_{Lr}$ and $f_{mr}$.

- Outliers do not have disproportionate effects on frequencies. Consequently, the tests are robust to failing to exclude unusual “one–time” events such as large discretionary write–downs of assets that analysts do not seek to predict (see, for example, the concerns detailed in Keane and Runkle (1998)).

- Most importantly, our test statistic is robust to cross–sectional correlation in forecast errors. To see this, suppose that $\theta_{lt} = \bar{\theta}_{lt} + e_{lt}$, where $\bar{\theta}_{lt}$ is independently distributed across time and firms, and $e_{lt}$ is a shifter that analysts do not forecast: Analysts forecast $\bar{\theta}_{lt}$, missing the common component, $e_{lt}$. This unforecasted component $e_{lt}$ could be an earnings shock realized after the last forecast; or it could be a common component of earnings that is hidden by the firm from forecasters. If $e_{lt} = e_{jt}, \forall i, j$, then it could represent a unforecasted market shock to earnings common to all firms. If $E[e_{lt}] = 0$, then if analysts issue unbiased forecasts of $\bar{\theta}_{lt}$, they issue unbiased forecasts of $\theta_{lt}$. To see that $T$ is robust to cross–sectional correlation, suppose that $e_{lt} > 0$. Recall that $\theta_{lt}$ is distributed symmetrically about $\bar{\theta}_{lt}$. While $e_{lt}$ does not affect the frequency with which $f_{Lr} > f_{mr}$, it reduces the likelihood with which event $\delta_{T}^+$ occurs (i.e., the likelihood that the last forecast exceed true earnings), but raises by an identical amount the likelihood with which event $\delta_{T}^-$ occurs (i.e., the likelihood that the last forecast falls short of true earnings). Thus, our frequency test statistic is unaffected by correlated information across forecasters, or cross–sectional correlation in earnings across firms.

The results, reported in Table 3, show that $T = 0.375$. That is, the analyst overshoots earnings in the direction away from the consensus 62.5% of the time. This qualitative pattern of overshooting in the direction away from the consensus continues to hold when we condition
on analyst coverage and financial year.\textsuperscript{16} The table also highlights the robustness of our test statistic to common information shocks. To see this observe that there is almost no variation in $T$ over different fiscal years, even though there is substantial variation from year to year in the frequency with which the last forecast exceeds true EPS. $T$ also does not vary systematically with the number of analysts following the firm, even though the mean probability that the last forecast exceeds true earnings is shifted down by greater analyst coverage. That is, while variations across different sub-samples shift up or down the mean probability that the consensus exceeds actual EPS, and hence the frequency with which event $\delta^+_r$ occurs, there is a corresponding offsetting shift in the frequency with which event $\delta^-_r$ occurs.

Were the last forecast unbiased, and unaffected by the consensus, then $T$ should be one-half. The fact that event $\delta_r = \delta^+_r \cup \delta^-_r$ occurs so infrequently ($T = 0.375$) is overwhelming evidence that the last analyst “over-emphasizes” his own information in his forecast and/or ignores information in earlier forecasts to which they have access. Indeed, despite the possibly countervailing effects of common shocks, the last analyst’s forecast exhibits a contrarian bias both when he has positive relative to the consensus, and when he has negative private information. That is, if his forecast exceeds the consensus, then it tends to exceed actual earnings, $\Pr(f_{\text{LR}} > \theta_r | f_{\text{LR}} > f_{\text{MR}}) = 0.56$; and if his forecast is less than the consensus then it generally falls short of actual earnings, $\Pr(f_{\text{LR}} < \theta_r | f_{\text{LR}} < f_{\text{MR}}) = 0.68$.

Note that on average, the last forecaster does not appear to make systematically overly-optimistic forecasts: $\Pr(f_{\text{LR}} > \theta_r) = 0.43$. Further, the probability that the last analyst’s forecast exceeds the actual EPS is greater for firms that are followed by few analysts. There are many papers documenting that not only does the last forecast tend, unconditionally, to fall short of true earnings per share, but so does the consensus forecast (see e.g. Burgstahler and Dichev (1997), Burgstahler and Eames (1999), Degeorge et al. (1999) and Bernhardt and Campello (2000)). These papers interpret this as evidence that firms seek to manage earnings in order to beat analysts’ consensus earnings forecast by a small amount; firms seek to generate a small positive earnings surprise. An alternative possibility that is consistent with these data, however, is that analysts and investors were (pleasantly) surprised by the high earnings growth experienced over this time period. These consistently pleasant surprises are reflected by the correspondingly sharp rise in stock prices.

A possible concern with our frequency analysis is that while it robustly identifies that the last analyst’s forecast is biased, it does not identify why the forecast is biased. Our explanation for the findings is that the last analyst has convex relative performance compensation of some form, and that he strategically biases his forecast in the direction of his private information. A referee observed that an alternative explanation for our findings is that the last

\textsuperscript{16}Similar patterns emerge if the data is segmented according to the total number of forecasts reported.

\textsuperscript{17}Note that Table 3 only reports the conditional probabilities for $f_{\text{LR}} > \theta_r$.  

15
Table 3: Bias in Last Forecast

| Samples | Test | Pr($f_{t,r} > \theta_r$) | Pr($f_{t,r} > \theta_r | f_{t,r} < \theta_r$) | Pr($f_{t,r} > \theta_r | f_{t,r} > \theta_r$) |
|---------|------|--------------------------|---------------------------------|---------------------------------|
| All     | 0.375| 0.43                     | 0.32                            | 0.56                            |
|         | [0.36, 0.39] | [0.42, 0.44] | [0.30, 0.34] | [0.54, 0.58] |
| Panel B: Segmented by Number of Analysts^b |
| N.A ∈ [2, 5] | 0.377 | 0.47                     | 0.35                            | 0.60                            |
|         | [0.35, 0.41] | [0.44, 0.50] | [0.31, 0.39] | [0.55, 0.64] |
| N.A ∈ [6, 7] | 0.377 | 0.45                     | 0.34                            | 0.58                            |
|         | [0.35, 0.41] | [0.42, 0.48] | [0.30, 0.38] | [0.53, 0.62] |
| N.A ∈ [8, 9] | 0.402 | 0.45                     | 0.37                            | 0.55                            |
|         | [0.37, 0.43] | [0.42, 0.48] | [0.32, 0.41] | [0.51, 0.60] |
| N.A ∈ [10, 12] | 0.358 | 0.42                     | 0.30                            | 0.57                            |
|         | [0.33, 0.39] | [0.39, 0.45] | [0.26, 0.34] | [0.53, 0.62] |
| N.A ≥ 13 | 0.368 | 0.37                     | 0.27                            | 0.51                            |
|         | [0.34, 0.39] | [0.35, 0.40] | [0.24, 0.30] | [0.47, 0.55] |
| Panel C: Segmented by Fiscal year |
| 1988 | 0.378 | 0.43                     | 0.31                            | 0.55                            |
|         | [0.34, 0.42] | [0.39, 0.47] | [0.26, 0.36] | [0.50, 0.61] |
| 1989 | 0.381 | 0.46                     | 0.36                            | 0.59                            |
|         | [0.34, 0.42] | [0.41, 0.50] | [0.30, 0.41] | [0.52, 0.65] |
| 1990 | 0.380 | 0.50                     | 0.37                            | 0.67                            |
|         | [0.31, 0.39] | [0.46, 0.54] | [0.31, 0.42] | [0.60, 0.72] |
| 1991 | 0.361 | 0.46                     | 0.35                            | 0.62                            |
|         | [0.32, 0.40] | [0.42, 0.50] | [0.30, 0.40] | [0.56, 0.68] |
| 1992 | 0.397 | 0.47                     | 0.38                            | 0.58                            |
|         | [0.36, 0.43] | [0.43, 0.50] | [0.34, 0.43] | [0.52, 0.64] |
| 1993 | 0.378 | 0.44                     | 0.32                            | 0.56                            |
|         | [0.34, 0.41] | [0.40, 0.47] | [0.27, 0.37] | [0.51, 0.61] |
| 1994 | 0.366 | 0.36                     | 0.23                            | 0.49                            |
|         | [0.33, 0.40] | [0.32, 0.40] | [0.19, 0.28] | [0.44, 0.54] |
| 1995 | 0.383 | 0.36                     | 0.25                            | 0.48                            |
|         | [0.35, 0.42] | [0.33, 0.40] | [0.21, 0.30] | [0.42, 0.53] |

**Notes:**
- This table provides values for our test statistic, $T$, as described in the text, as well as conditional and unconditional relative frequencies of the direction of error in the last forecast. 95% confidence intervals are reported in square brackets.
- The sample used in this frequency tests excludes observations for which that last forecast equals the outstanding consensus, or equals the actual EPS (see text).
- N.A denotes that number of analysts following the firm. For example, N.A ∈ [2, 6] refers to the sub-sample in which forecasts for a firm’s quarterly EPS were provided by 2 to 6 analysts, and so on.
analyst is systematically myopic and ignores the information contained in earlier forecasts. If each analyst receives an independent, identically normally distributed signal and issues an unbiased forecast based only on his private information, then the unbiased forecast given the last forecast and the outstanding consensus is \( \frac{(L-1)f_{L-1} + f_L}{L} \). More generally, if the last analyst ignores the information contained in earlier forecasts, and myopically issues an unbiased forecast based only on his private information, then, true earnings will tend to be between the consensus and the last forecast (see Ehrbeck and Waldmann 1996). In this story, the last analyst does not strategically bias his forecast in the direction of his private information, but rather his forecast myopically reflects only his information. Hence, both our strategic model with convex relative performance compensation and this analyst myopia model predict that the analyst should overshoot true earnings in the direction away from the consensus with a probability exceeding 0.5.

To test between our strategic explanation for forecast bias and the alternative analyst myopia explanation, we now tease out additional theoretical predictions of our model when analyst compensation is convex:

**Proposition 3** Suppose that \( R(\cdot) \) is convex, and consider two posterior belief distributions for the last analyst, \( g_L(\theta|\Omega_L) \) and \( \tilde{g}_L(\theta|\Omega_L) \), where \( \tilde{g}_L(\cdot) \) is a mean preserving spread of \( g_L(\cdot) \). Then for any given difference between the consensus and the last analyst’s expectation, \( |f_m - \hat{\theta}_L| \), the last analyst chooses a more severely biased report when his posterior beliefs are less precise:

\[
|f_L(\tilde{g}_L) - \hat{\theta}_L| > |f_L(g) - \hat{\theta}_L|.
\]

Proposition 3 has no frequency analog. Although a less-dispersed posterior reduces the bias in the last analyst’s forecast, there is more probability mass close to the median, so that the probability of overshooting may be higher or smaller if more analysts follow a firm.

Assuming that the uncertainty that the last analyst faces is reduced by the number of existing forecasts, we can implement Proposition 3’s content empirically as follows:

**Corollary 1** Suppose that \( R(\cdot) \) is convex and that the last analyst’s posterior about \( \theta \) is tighter if he observes more forecasts about a firm’s EPS. Then for any given difference between the consensus and the last analyst’s forecast, \( f_m - f_L \), the expected bias in the last analyst’s forecast is less when there are more forecasts:

\[
E[(f_L \tau - \hat{\theta}_L)|f_L \tau - f_m\tau] \text{ is a decreasing function of the number of forecasts at the last analyst’s disposal.}
\]
That is, the more analysts following the firm (or the more total forecasts), the smaller is the strategic bias in the last analyst’s forecast. Consequently, for any given deviation in the last forecast away from the consensus, the expected error in his forecast falls with the number of forecasts/analysts.

The myopic analyst model makes the opposite prediction:

**Corollary 2** Suppose that the more forecasts reported by analysts following a firm, the more information that is contained in the outstanding consensus forecast, i.e. $E[|f_{Lr} - \hat{\theta}_r|]$ falls with the number of forecasts. Then for any given difference between the consensus and the last analyst’s forecast, $f_m - f_L$, the expected bias in the last analyst’s forecast is greater when there are more forecasts:

$$E[(f_{Lr} - \hat{\theta}_r)|f_{Lr} - f_m]$$

is an increasing function of the number of forecasts.

### 4.2.2 Convexity of $R$ versus analysts’ myopia

To test the compensation hypothesis against the alternative explanation of analyst myopia, we see how the bias is related to the number of forecasts. Since earnings are expressed on a per share basis, to eliminate scale effects we must compute the bias as a percent of stock price:\(^{18}\)

$$BIAS_{Lr} = \frac{f_{Lr} - \theta_r}{P_r},$$

where $P_r$ is the share price at the time of publication of the last forecast in the firm quarter. For similar reasons, we express the difference between the last forecast and the consensus as a percentage of share price,

$$SFD_r = \frac{f_{Lr} - f_m}{P_r}.$$

We then investigate the relationship between these two variables by running regressions of $BIAS$ on $SFD$, using different proxies to capture the amount of information available to the last analyst. The first model is estimated using dummy variables defined on analyst coverage:

$$BIAS_{Lr} = \alpha_0 + \alpha_1 NA_{[6,7]} + \alpha_2 NA_{[8,9]} + \alpha_3 NA_{[10,12]} + \alpha_4 NA_{[13]} + \alpha_5 \left(NA_{[2,5]}SFD_r\right)$$

$$+ \alpha_6 \left(NA_{[6,7]}SFD_r\right) + \alpha_7 \left(NA_{[8,9]}SFD_r\right) + \alpha_8 \left(NA_{[10,12]}SFD_r\right) + \alpha_9 \left(NA_{[13]}SFD_r\right) + \epsilon_r$$

where $NA_{[2,5]} = 1$ if 2 to 5 analysts reported forecasts in a firm-quarter, and zero otherwise; $NA_{[6,7]} = 1$ if 6 to 7 analysts reported forecasts in a firm-quarter, and zero otherwise, and so on, with $NA_{[13]} = 1$ if more than 13 analysts report forecasts, and equal zero otherwise.\(^{18}\)

\(^{18}\)This method of standardizing forecasts error and forecast differences is standard (see e.g., Lys and Sohn (1990), Alexander and Ang (1997) and Lim (2000). Following Lim we also delete firm-quarters (16 observations) with stock prices below five dollars, in order to reduce the sensitivity of our results to noise in prices.
The second model uses the total number of forecasts reported as a proxy for information:

\[
\text{BIAS}_{Lr} = \beta_0 + \beta_1 TF_{[7,9]} + \beta_2 TF_{[10,12]} + \beta_3 TF_{[13,17]} + \beta_4 TF_{[17]} + \beta_5 \left( TF_{[2,6]} SFD_r \right) \\
+ \beta_6 \left( TF_{[7,9]} SFD_r \right) + \beta_7 \left( TF_{[10,12]} SFD_r \right) + \beta_8 \left( TF_{[13,17]} SFD_r \right) + \beta_9 \left( TF_{[18]} SFD_r \right) + \epsilon_r
\]

where \( TF_{[2,6]} = 1 \) if 2 to 6 total forecasts are reported in a firm–quarter, and zero otherwise; \( TF_{[7,9]} = 1 \) if 7 to 9 forecasts reported, and zero otherwise, and so on, with \( TF_{[18]} = 1 \) if more than 18 are reported, and equal zero otherwise.\(^{19}\) We also consider models with tighter functional forms:

\[
\text{BIAS}_{Lr} = \zeta_0 + \zeta_1 \ln(NA_r) + \zeta_2 SFD_r + \zeta_3 \ln(NA_r) \ast SFD_r + \epsilon_r
\]

and

\[
\text{BIAS}_{Lr} = \nu_0 + \nu_1 \ln(TF_r) + \nu_2 SFD_r + \nu_3 \ln(TF_r) \ast SFD_r + \epsilon_r\].

If the last forecast is unbiased, then \( \text{BIAS}_{Lr} \) should not be systematically related to \( SFD_r \). That is, none of the coefficients on the interaction terms between \( SFD_r \) and measures of the number of analysts or forecasts in each of the regressions should be significantly different from zero, nor should their sum.

If, instead, convex relative compensation of some form drives the last analyst’s forecast so that the last analyst strategically biases his forecast away from the consensus, then the sum of the coefficients on the interaction terms should be positive, and the bias should fall as the last analyst’s uncertainty is resolved. That is, the coefficient on interaction terms between \( SFD_r \) and the number of analysts (forecasts) should fall as the number of analysts (forecasts) increases, reflecting that the last analyst’s uncertainty falls with the number of analysts (forecasts), inducing him to issue a less biased forecast.

Finally, if analyst myopia drives the last analyst’s forecast, i.e. if the last analyst ignores the information contained in the outstanding forecasts, then the sum of the coefficients on the interaction terms should also be positive, but the coefficients on the interaction terms between \( SFD_r \) and the number of analysts (forecasts) should rise with the number of analysts (forecasts), reflecting that as the number of analysts (forecasts) rises, the consensus forecast becomes more accurate so that fixing the difference between the myopic analyst’s forecast and the consensus, a myopic last analyst, on average, will overshoot by more.

Table 4 presents the results from these regression analyses. To account for possible correlation in the forecast errors, the standard errors are computed using the Hubert–White (robust cluster) method.

The results strongly support the hypothesis that the bias in the last analyst’s forecast is due to some form of convex relative performance compensation and is strategically chosen.\(^{19}\)

\(^{19}\)The dummies with number of forecasts are constructed differently from those with analysts coverage to ensure similar sub-sample sizes. Recall that the average number of forecasts per analyst is about 1.31 (see Table 2).
Table 4: Convexity of $R$ versus myopic analyst

This table reports the results from firm-fixed effects regressions of BIAS (i.e., the error in last forecast expressed as a percentage of stock price at date of release of last forecast, $P^t$) on the scaled difference in the last and outstanding mean forecasts, $SFD = \frac{\bar{P}_{-1} - \bar{P}_{-2}}{\bar{P}_{-1}}$. Various proxies are used to capture the information available to the last analyst: Panel A uses dummies defined on analyst coverage, Panel B uses dummies defined on total forecasts reported (including older ones), and Panels C and D uses the logarithm of these measures. Robust standard errors (in parentheses) were computed using the Hubert–White method to account for possible correlation of forecast errors within firms. (*) (**) and (***)) indicate significant at 90%, 95% and 99% levels.

Panel A: Dummies, as proxy for information, defined on analysts following

<table>
<thead>
<tr>
<th>const.</th>
<th>$NA_{15,7}$</th>
<th>$NA_{19,9}$</th>
<th>$NA_{19,12}$</th>
<th>$NA_{21,3}$</th>
<th>$NA_{21,7}$SFD</th>
<th>$NA_{21,9}$SFD</th>
<th>$NA_{21,12}$SFD</th>
<th>$NA_{21,13}$SFD</th>
<th>$NA_{21,15}$SFD</th>
<th>$R^2$</th>
<th>$\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.130</td>
<td>-0.083</td>
<td>-0.112</td>
<td>-0.082</td>
<td>-0.155</td>
<td>1.560</td>
<td>0.894</td>
<td>0.769</td>
<td>0.624</td>
<td>0.462</td>
<td>0.176</td>
<td>10.3***</td>
</tr>
<tr>
<td>(0.104)</td>
<td>(0.110)</td>
<td>(0.121)</td>
<td>(0.106)</td>
<td>(0.103)</td>
<td>(0.300)***</td>
<td>(0.164)***</td>
<td>(0.154)***</td>
<td>(0.110)***</td>
<td>(0.335)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Panel B: Dummies, as proxy for information, defined on total number of forecasts

<table>
<thead>
<tr>
<th>const.</th>
<th>$TF_{15,9}$</th>
<th>$TF_{19,12}$</th>
<th>$TF_{19,15}$</th>
<th>$TF_{19,17}$</th>
<th>$TF_{19,26}$SFD</th>
<th>$TF_{19,27}$SFD</th>
<th>$TF_{19,7,12}$SFD</th>
<th>$TF_{19,10,12}$SFD</th>
<th>$TF_{19,13,15}$SFD</th>
<th>$R^2$</th>
<th>$\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.110</td>
<td>-0.088</td>
<td>-0.069</td>
<td>-0.062</td>
<td>-0.086</td>
<td>1.352</td>
<td>1.125</td>
<td>0.491</td>
<td>0.100</td>
<td>0.195</td>
<td>0.187</td>
<td>28.59***</td>
</tr>
<tr>
<td>(0.102)</td>
<td>(0.108)</td>
<td>(0.118)</td>
<td>(0.101)</td>
<td>(0.101)</td>
<td>(0.318)***</td>
<td>(0.135)***</td>
<td>(0.253)***</td>
<td>(0.228)***</td>
<td>(0.137)***</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Panel C: $\ln(N.A)$ as proxy for information

<table>
<thead>
<tr>
<th>const.</th>
<th>$\ln(N.A)$</th>
<th>SFD</th>
<th>$\ln(N.A) + SFD$</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.168</td>
<td>-0.056</td>
<td>1.506</td>
<td>-0.345</td>
<td>0.165</td>
</tr>
<tr>
<td>(0.105)</td>
<td>(0.042)</td>
<td>(0.257)***</td>
<td>(0.105)**</td>
<td></td>
</tr>
</tbody>
</table>

Panel D: $\ln(TF)$ as proxy for information

<table>
<thead>
<tr>
<th>const.</th>
<th>$\ln(TF)$</th>
<th>SFD</th>
<th>$\ln(TF) + SFD$</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.131</td>
<td>-0.035</td>
<td>1.609</td>
<td>-0.334</td>
<td>0.168</td>
</tr>
<tr>
<td>(0.112)</td>
<td>(0.042)</td>
<td>(0.362)***</td>
<td>(0.151)**</td>
<td></td>
</tr>
</tbody>
</table>
That is, if the last analyst’s forecast overshoots the consensus, he tends to overshoot earnings, but the amount by which he overshoots falls with the amount of information (as captured either by the number of analysts or the number of forecasts) at his disposal. For example, the first regression in panel A (i.e., with dummies defined on numbers of analysts reporting forecasts) indicate that if the last analyst overshoots the consensus by 1% of the stock’s price and there are less than five analysts following the firm, then on average he will exceed true EPS by 1.4%. However, if there are ten to twelve analysts, then he will only exceed true EPS by 0.68%. With more than 13 analysts reporting forecasts of the firm’s EPS, this bias is only about 0.45% of the stock price. A test of the differences in these estimates returned a $\chi^2$ of 10.3, which is significant at the 5% level. The results are inconsistent with the myopic analyst model, which predicts that the bias should rise with the number of analysts. This is reassuring, as it indicates that analysts who have a lot at stake do respond to economic incentives when choosing their forecasts.

A similar pattern emerges when the dummies are defined on the total number of forecasts (panel B). The first regression in panel B reveals that if the last analyst overshoots the consensus by 1% of the stock’s price and there are less than six forecasts at his disposal, then he will exceed true EPS by 1.46%, but he exceeds true EPS by only 0.52% if he has more that eighteen forecasts at his disposal. Again, the $\chi^2$ = 28.59 shows that interaction terms are not jointly equal.20

Finally, panels C and D of Table 4 reveal that the qualitative results are unchanged even when we impose logarithmic functional forms on our measures of information (logarithms fit less well for large numbers of analysts) to test the hypotheses of convexity of relative performance compensation against that of analysts’ myopia. That is, the bias falls with both the numbers of analysts and the numbers of forecasts at the last analyst’s disposal.

4.2.3 Does the market unravel this bias?

Our analysis has revealed that the last analyst’s reported forecast over-emphasizes his private information: his forecast tends to over-shoot the actual EPS away from the outstanding consensus. This raises the following question: Are investors systematically fooled by the bias in the last analyst’s forecast? For example, a large positive deviation, $(f_{Lr} - f_{mr})$, in the last forecast away from the consensus, may mislead investors who do not account for the bias in this forecast into over-estimating by how much earnings will go up. Investors who treat the last report as unbiased will over-react to their forecasts of earnings. Then, if the ‘correct’ information does not leak out until near the earnings announcement date, a large positive $(f_{Lr} - f_{mr})$ should be associated with subsequent disappointment and a downward revision

20The qualitative findings are robust to other partitions of the data according to number of analysts or total forecasts, as well as to trimming the data by deleting observations in the tails of $SFD_r$ (as in Lim 2000).
in prices around the earnings announcement, and hence negative excess returns. Conversely, a large negative $f_{Lt} - f_{m,t}$ should be associated with positive excess returns around the earnings announcement. In this section, to see whether investors are systematically fooled by the forecasts, we see if there is a systematic relationship between the degree to which the last forecast overshoots the consensus and excess returns.

To obtain excess returns, we first estimate a market model for each firm-quarter:

$$r_{it} = \alpha_0 + \alpha_1 r_{mt} + \epsilon_t,$$

(7)

where $r_{it}$ is the daily return on security $i$ and $r_{mt}$ is the return on the CRSP equal weighted market index. For each firm-quarter, the market model regression (equation (7)) uses returns for the period $t = -170$ through to $t = -21$, where time $t = 0$ denotes the day of actual earnings announcement by the firm. Hence, each market model is estimated with 150 data points. The least squares estimates, $\hat{\alpha}_0$ and $\hat{\alpha}_1$, are then used to calculate the cumulative abnormal returns (CAR) for different event windows around the earnings announcement date for each firm-quarter $\tau$:

$$CAR_{\tau} = \sum_{t=L}^{T} \left[ r_{it} - (\hat{\alpha}_0 + \hat{\alpha}_1 r_{mt}) \right],$$

(8)

where $L$ and $T$ are starting and end points of the period over which abnormal returns are cumulated.

In our analysis, it is important to exclude the market’s reaction to the information in the last forecast. For instance, if the last forecast exceeds the consensus, then this good news will be reflected in the price and lead to immediate positive excess returns; cumulating those immediate returns may cause us to fail to find evidence of bias, since the market’s failure to incorporate the bias leads to predicted negative excess returns. To disentangle the impact of new information from the reaction to the last forecast, we drop all observations for which the last forecast was not at least two days from the beginning of our event window. Our findings are not systematically affected by other choices.

Note that the empirical regularity for earnings surprise is that, on average, daily abnormal returns evolve monotonically throughout a long event window (see e.g. Foster, Olsen and Shevlin (1984)). Since we are searching for evidence that investors are systematically fooled — which implies a non-monotonic relationship in abnormal returns as investors over-react to the last forecast — we are “battling” the standard empirical regularity.

We then estimated separate regressions of $CAR_{[-2,0]}$ and $CAR_{[-2,1]}$ on SFD using analyst coverage and total forecasts reported to control for information available to the last analyst. Table 5 presents the findings. Qualitatively, the following empirical regularities emerge:\footnote{Similar empirical regularities regarding percentage mean cumulative abnormal returns emerge for other windows.}
Table 5: Regressions of mean CARs on SFD: Dummies as proxy for information

This table reports results from OLS regressions of mean cumulative abnormal returns (CAR), on scaled forecast difference, SFD = \( \frac{P - P_{0}}{P_{0}} \), where \( P \) is the stock price at the date of release of the last forecast. Abnormal returns are cumulated from time \( t \) to \( t \). Date \( t = 0 \) denote the date of actual earnings announcement. Dummy variables defined on analyst coverage (Panel A) and total number of forecasts (Panel B), as described in the text. Estimated standard errors are reported in parentheses: (*), (**) and (***) indicate significant at 90%, 95% and 99% levels. All regressions have \( R^2 \) of about 1%.

### Panel A: Dummies defined on number of analysts

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>const.</th>
<th>( N_{A,0} )</th>
<th>( N_{A,1} )</th>
<th>( N_{A,1} )</th>
<th>( N_{A,2} )</th>
<th>( N_{A,3} )</th>
<th>( SFD )</th>
<th>( SFD )</th>
<th>( SFD )</th>
<th>( SFD )</th>
<th>( SFD )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( CAR_{-2,0} )</td>
<td>0.082</td>
<td>-0.007</td>
<td>0.075</td>
<td>-0.149</td>
<td>0.129</td>
<td>-0.373</td>
<td>-0.344</td>
<td>0.165</td>
<td>-0.061</td>
<td>-0.229</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.002)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.002)*</td>
<td>(0.001)**</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td></td>
</tr>
<tr>
<td>( CAR_{-2,1} )</td>
<td>0.120</td>
<td>-0.054</td>
<td>-0.101</td>
<td>-0.157</td>
<td>-3.08E-06</td>
<td>-0.591</td>
<td>-0.858</td>
<td>0.008</td>
<td>0.163</td>
<td>-0.352</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)**</td>
<td>(0.001)**</td>
<td>(0.002)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Panel B: Dummies defined on total number of forecasts

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>const.</th>
<th>( TF_{1,0} )</th>
<th>( TF_{1,1} )</th>
<th>( TF_{1,2} )</th>
<th>( TF_{1,3} )</th>
<th>( SFD )</th>
<th>( SFD )</th>
<th>( SFD )</th>
<th>( SFD )</th>
<th>( SFD )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( CAR_{-2,0} )</td>
<td>0.033</td>
<td>0.115</td>
<td>0.113</td>
<td>-0.084</td>
<td>0.174</td>
<td>-0.487</td>
<td>-0.438</td>
<td>0.024</td>
<td>-0.007</td>
<td>-0.172</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.002)</td>
<td>(0.001)</td>
<td>(0.002)*</td>
<td>(0.001)**</td>
<td>(0.001)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td></td>
</tr>
<tr>
<td>( CAR_{-2,1} )</td>
<td>0.063</td>
<td>0.083</td>
<td>0.032</td>
<td>-0.208</td>
<td>0.090</td>
<td>-0.717</td>
<td>-0.902</td>
<td>-0.005</td>
<td>0.009</td>
<td>-0.249</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.003)**</td>
<td>(0.001)**</td>
<td>(0.002)</td>
<td>(0.003)</td>
<td>(0.002)</td>
<td></td>
</tr>
</tbody>
</table>

### Panel C: \( \ln(NA) \) as proxy for information

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>const.</th>
<th>( \ln(NA) )</th>
<th>( SFD )</th>
<th>( \ln(NA) ) * ( SFD )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( CAR_{-2,0} )</td>
<td>0.008</td>
<td>0.011</td>
<td>-0.706</td>
<td>0.236</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.001)</td>
<td>(0.004)*</td>
<td>(0.002)</td>
</tr>
<tr>
<td>( CAR_{-2,1} )</td>
<td>0.103</td>
<td>-0.020</td>
<td>-1.352</td>
<td>0.448</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.001)</td>
<td>(0.004)**</td>
<td>(0.002)**</td>
</tr>
</tbody>
</table>

### Panel D: \( \ln(TF) \) as proxy for information

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>const.</th>
<th>( \ln(TF) )</th>
<th>( SFD )</th>
<th>( \ln(TF) ) * ( SFD )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( CAR_{-2,0} )</td>
<td>0.020</td>
<td>0.032</td>
<td>-0.580</td>
<td>0.145</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.006)</td>
<td>(0.003)*</td>
<td>(0.001)</td>
</tr>
<tr>
<td>( CAR_{-2,1} )</td>
<td>0.120</td>
<td>-0.025</td>
<td>-1.320</td>
<td>0.363</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(-0.001)</td>
<td>(0.004)**</td>
<td>(0.002)**</td>
</tr>
</tbody>
</table>
Mean cumulative abnormal returns, from two days prior to earnings announcement till the day of, or day after, the announcement are negatively correlated with SFD.

The negative impact of SFD on mean cumulative abnormal returns is greater when fewer analysts follow the firm (and hence, fewer forecasts of earnings are reported).

These relationships are stronger if returns are cumulated up to one day after actual earnings have been announced.

The data indicate that if the last forecast overshoots the consensus, then the stock market price response to the actual earnings announcement is in the opposite direction; and the greater is the difference between the last forecast and the consensus forecast, the greater is the expected change in stock price. Also, the magnitude of the market’s response to the actual earnings announcement is greater for firms followed by fewer analysts.

The data therefore suggest that investors are fooled by the last analyst’s forecast, underestimating the magnitude of the bias in his report and attaching an undue weight to the last forecast relative to the consensus. Further, investors do not appear to unravel the fact that the incentive for the last analyst to issue a biased forecast is greater for firms followed by fewer analysts, for which there is more uncertainty. Consequently, when there are fewer forecasts, they are misled by the last analyst’s forecast by a greater amount.

5 Conclusion

This paper presents a compensation-based model to explain the strategic bias in individual analysts’ forecast of earnings per share (EPS). When analysts are compensated both on the basis of the accuracy of their forecasts and how the accuracy of their forecasts compare with other analysts, then earlier forecasts affect later announcements. For different forms of relative performance compensation, we derive the implications for the pattern of forecasts.

We find overwhelming evidence that the last forecast is not an unbiased reflection of that analyst’s information: The last analyst to report a forecast of a firm’s EPS strategically reports a strongly contrarian forecast, biasing his forecast away from the consensus forecast — 62% of the time, the last forecast overshoots actual EPS in the direction away from the outstanding mean forecast. This is consistent with some form of convex relative performance and implies that a better estimate of EPS than either the last forecast or the average of all forecasts could be obtained by accounting for the bias in the last forecast (even just by averaging the last and consensus forecasts). We also document that the bias in the last forecast falls with the number of forecasts, indicating that the bias is chosen strategically, rather than due to the last forecaster myopically ignoring the information in the consensus.
Finally, we test to see whether investors correctly unravel the strategic bias in the last forecast. We show that the difference between the last analyst’s forecast and the consensus scaled by price predicts excess returns cumulated around the earnings announcement date. In particular, the data indicate that investors appear to be systematically fooled by the final forecast, attaching too much weight to the last forecast, especially for firms followed by few analysts.

Appendix

Proof of Lemma 1:
If compensation is linear in relative performance, then an analyst’s compensation can be written as

\[ w_j = \overline{w} - (1 + \lambda)|f_j - \theta| + |f_m - \theta|. \]

Consider the last analyst to report his forecast. Since the \((L - 1)\) preceding forecasts have been announced, \(f_L\) will not affect \(f_m\), the mean of all other forecasts. Hence, maximizing \(E(w_L|\hat{\theta}_L)\) is equivalent to minimizing

\[ E[|f_L - \theta|\mid \hat{\theta}_L] = \int_\theta (|f_L - \theta|)dG(\theta|\hat{\theta}_L) \]

To show that the optimal forecast is \(f_L = \hat{\theta}_L\), we prove that the expected forecast error is greater if the last analyst reports any other forecast.

Suppose he reports \(f_L = \tilde{\theta}_L - \xi\) (a lesser forecast), where \(\xi > 0\). Then the difference in expected errors is

\[
\int_\theta (|\tilde{\theta}_L - \theta| - |\hat{\theta}_L - \xi - \theta|)dG(\theta|\hat{\theta}_L) = \int_{\theta \leq \hat{\theta}_L - \xi} \xi dG(\theta|\hat{\theta}_L) + \int_{\hat{\theta}_L - \xi}^{\hat{\theta}_L} (2\tilde{\theta}_L - 2\theta - \xi) dG(\theta|\hat{\theta}_L) - \int_{\theta > \hat{\theta}_L} \xi dG(\theta|\hat{\theta}_L)
\]

\[ \leq \int_{\theta \leq \hat{\theta}_L} \xi dG(\theta|\hat{\theta}_L) - \int_{\theta > \hat{\theta}_L} \xi dG(\theta|\hat{\theta}_L) \tag{9} \]

since \(\theta > \hat{\theta}_L - \xi\) in the second term on the RHS of (9), which implies that \(2\tilde{\theta}_L - 2\theta - \xi \leq \xi\). By the symmetry of \(g(\theta|\hat{\theta}_L)\) about \(\hat{\theta}_L\), the last term in (9) vanishes. Hence,

\[ \int_\theta (|\tilde{\theta}_L - \theta|)dG(\theta|\hat{\theta}_L) \leq \int_{\theta \leq \hat{\theta}_L} \xi dG(\theta|\hat{\theta}_L) \]

Similarly, \(\forall \xi > 0\), we get \(\int_\theta (|\tilde{\theta}_L - \theta|)dG(\theta|\hat{\theta}_L) < \int_{\theta \leq \hat{\theta}_L + \xi - \theta} \xi dG(\theta|\hat{\theta}_L)\). The last analyst has a larger expected forecast error if he announces a forecast greater than his posterior expectation. Hence the optimal forecast is \(f_L = \hat{\theta}_L\).

Now consider the \((L - 1)\)th analyst. Since the \((L - 2)\) preceding forecasts have been announced and the last analyst will report his true signal anyway\(^{22}\), the forecast of the \((L - 1)\)th analyst will not affect the mean of the other forecasts: that is, \(\eta_{(L-1)} = 0\). Hence, maximizing \(E(w_{L-1}|\hat{\theta}_{L-1})\) is equivalent to minimizing

\[ E(|F_{L-1} - \theta|\mid \hat{\theta}_{L-1}) \]

Repeating the analysis above, it is straightforward to show that the optimal forecast of the \((L - 1)\)th analyst will then be \(f_{L-1} = \hat{\theta}_{L-1}\). By backward induction, we get \(F_j^* = \hat{\theta}_j, \forall j = 1, 2, \ldots, L\).

Proof of Lemma 2:
\(^{22}\)By assumption, \(f_{L-1}\) will not affect \(\hat{\theta}_L\) since analysts cannot mislead others.
Given his posterior beliefs, the last analyst’s expected payoff will be given by

\[ E \left[ w_L(f_L) \mid \hat{\theta}_L \right] = \pi - \lambda \int_0^{f_L} (f_L - \theta) dG(\theta \mid \hat{\theta}_L) + \int_0^{f_L} R(\theta) dG(\theta \mid \hat{\theta}_L). \]

Suppose \( f_m < \hat{\theta}_L \). We first show that \( f_L < f_m \). To do this, we prove that for every forecast \( f_L < f_m \), there exists some other \( f_L' > f_m \) that yields a higher payoff. Consider the following two reports by the last analyst: \( f_L = f_m + \xi \) and \( f_L = f_m - \xi \), where \( \xi > 0 \). The difference in payoffs will be given by

\[
E \left[ w_L(f_m + \xi) - w_L(f_m - \xi) \mid \hat{\theta}_L \right] = \int_{\theta < f_m - \xi} (-2 \lambda \xi + R(-\xi)) \] 
\[ + \int_{f_m - \xi < \theta < f_m + \xi} (2 \lambda (\theta - f_m) + R(-\xi) - R(2 \theta - 2f_m - \xi)) dG(\theta \mid \hat{\theta}_L) \] 
\[ + \int_{f_m + \xi < \theta} (2 \lambda (\theta - f_m) + R(2 \theta - 2f_m - \xi) - R(-\xi)) dG(\theta \mid \hat{\theta}_L) \] 
\[ + \int_{\theta > f_m + \xi} (-2 \lambda \xi + R(-\xi)) dG(\theta \mid \hat{\theta}_L). \]  

Clearly, the sum of the first and fourth terms on the RHS of (10) is positive by the symmetry of \( g(\theta \mid \hat{\theta}_L) \) about \( \hat{\theta}_L \). Secondly, the third term is at least equal to the second term since \( g(\theta \mid \hat{\theta}_L) \) is single–peaked at \( \hat{\theta}_L \).

Therefore,

\[ E \left[ w_L(f_m + \xi) - w_L(f_m - \xi) \mid \hat{\theta}_L \right] > 0 \]
and \( f_m + \xi \) dominates \( f_m - \xi, \forall \xi > 0 \).

It is also straightforward to show that \( f_L \neq f_m \); that is, it is not optimal for the last analyst to mimic the mean of earlier forecasts by reporting \( f_L = f_m \). Because,

\[
E \left[ w_L(f_m + \epsilon) - w_L(f_m) \mid \hat{\theta}_L \right] = \int_{\theta < f_m} (-\lambda \epsilon + R(-\epsilon)) dG(\theta \mid \hat{\theta}_L) 
+ \int_{f_m < \theta < f_m + \epsilon} (\lambda(2\theta - 2f_m - \epsilon) + R(2\theta - 2f_m - \epsilon)) dG(\theta \mid \hat{\theta}_L) 
+ \int_{\theta > f_m + \epsilon} (\lambda + R(\epsilon)) dG(\theta \mid \hat{\theta}_L). 
\]

Dividing through by \( \epsilon \) and taking the limit as \( \epsilon \to 0 \), we get the derivative of \( w_L(f_L) \) at \( f_L = f_m \) to be

\[
\lim_{\epsilon \to 0} E \left[ \frac{w_L(f_m + \epsilon) - w_L(f_m)}{\epsilon} \right] = -\int_{\theta < f_m} (\lambda + R'(\theta)) dG(\theta \mid \hat{\theta}_L) + \int_{\theta > f_m} (\lambda + R'(\theta)) dG(\theta \mid \hat{\theta}_L) \]

Using the symmetry of \( g(\theta \mid \hat{\theta}_L) \) about \( \hat{\theta}_L \), we get derivative of \( w_L(f_L) \) w.r.t \( f_L \) at \( f_L = f_m \) to be

\[
\lim_{\epsilon \to 0} E \left[ \frac{w_L(f_m + \epsilon) - w_L(f_m)}{\epsilon} \right] = 2 \int_{f_m}^{\hat{\theta}_L} (\lambda + R'(\theta)) dG(\theta \mid \hat{\theta}_L) > 0. \]

That is \( w_L(f_L) \) is increasing at \( f_L = f_m \). Hence, \( f_L > f_m \) if \( \hat{\theta}_L > f_m \). The proof that \( f_L < f_m \) if \( \hat{\theta}_L < f_m \) is analogous.

**Proof of Proposition 1:**

(i) Suppose \( f_m < \hat{\theta}_L \), then by Lemma 2, \( f_L > f_m \). So it remains to show that \( f_L \notin [\hat{\theta}_L, \infty) \). To do this, we show that \( w_L(f_L) \) is decreasing in \( f_L \) over \([\hat{\theta}_L, \infty) \). Consider \( f_L = \hat{\theta}_L + \xi \), where \( \xi \geq 0 \). The derivative of \( w_L(f_L) \) at this point is given by

\[
\lim_{\epsilon \to 0} E \left[ \frac{w_L(\hat{\theta}_L + \xi + \epsilon) - w_L(\hat{\theta}_L + \xi)}{\epsilon} \right] = -\int_{\theta < \hat{\theta}_L} (\lambda + R'(\hat{\theta}_L - \xi)) dG(\theta \mid \hat{\theta}_L) 
- \int_{\hat{\theta}_L < \theta \leq \hat{\theta}_L + \xi} (\lambda + R'(\hat{\theta}_L - \xi - f_m)) dG(\theta \mid \hat{\theta}_L) 
+ \int_{\theta \geq \hat{\theta}_L + \xi} (\lambda + R'(\hat{\theta}_L + \xi - f_m)) dG(\theta \mid \hat{\theta}_L). \]  

26
Consider the second term on the RHS of (11). Since \( \theta < \bar{\theta}_L + \xi \), we have \( R'(2\theta - \bar{\theta}_L - \xi - f_m) \geq R'(\bar{\theta}_L + \xi - f_m) \) by the concavity of \( R(\cdot) \). Also \( R'(f_m - \bar{\theta}_L - \xi) > R'(\bar{\theta}_L + \xi - f_m) \). Therefore,

\[
\lim_{\varepsilon \to 0} E \left[ \frac{w_L(\hat{\theta}_L + \xi + \varepsilon) - w_L(\hat{\theta}_L + \xi)}{\varepsilon} \right] < -\int_{\theta \leq \hat{\theta}_L + \xi} (\lambda + R'(\hat{\theta}_L + \xi - f_m)) dG(\theta) \hat{\theta}_L \nonumber \nonumber
\]  

\[
+ \int_{\theta > \hat{\theta}_L + \xi} (\lambda + R'(\hat{\theta}_L + \xi - f_m)) dG(\theta) \hat{\theta}_L. \nonumber
\]

Exploiting the symmetry of \( g(\theta) \) about \( \hat{\theta}_L \), we get

\[
\lim_{\varepsilon \to 0} E \left[ \frac{w_L(\hat{\theta}_L + \xi + \varepsilon) - w_L(\hat{\theta}_L + \xi)}{\varepsilon} \right] < -\int_{\hat{\theta}_L - \xi}^{\hat{\theta}_L + \xi} (\lambda + R'(\hat{\theta}_L + \xi - f_m)) dG(\theta) \hat{\theta}_L < 0, \nonumber
\]

\( \forall \varepsilon \geq 0 \). So the last analyst’s payoff is decreasing in \( f_L \in [\hat{\theta}_L, \infty) \). Combining this result with Lemma 2, it follows that the last analyst will “locate” between \( f_m \) and \( \hat{\theta}_L \).

The proof that \( \hat{\theta}_L < f_L < f_m \) if \( \bar{\theta}_L < f_m \) is analogous.

(ii) Let \( f_m < \hat{\theta}_L \) and \( f_L \) be the optimal forecast of the second analyst given \( \hat{\theta}_L \) and \( f_m \). Then it will suffice to show that if the consensus forecast were to be revised to \( \bar{f}_m = f_m - \delta, \) \( \delta > 0 \), then the last analyst will report \( f_L^* \in (f_L - \delta, f_L) \).

By Proposition 1, we know that \( f_L \in (f_m, \hat{\theta}_L) \), and by the optimality of \( f_L(\hat{\theta}_L) \), \( \frac{\partial w}{\partial f} = 0 \) at \( f_L(\hat{\theta}_L) \).

But

\[
\lim_{\varepsilon \to 0} E \left[ \frac{w_L(f_L + \varepsilon) - w_L(f_L)}{\varepsilon} \right] = -\int_{\theta \leq f_m} (\lambda + R'(f_m - f_L)) dG(\theta) \hat{\theta}_L - \int_{f_m}^{f_L} (\lambda + R'(2\theta - f_m - f_L)) dG(\theta) \hat{\theta}_L - \int_{\theta > f_L} (\lambda + R'(f_L - f_m)) dG(\theta) \hat{\theta}_L = 0. \nonumber\]

Let \( H(f_L) \) be defined as

\[
H(f_L) = -\int_{\theta \leq f_m} (\lambda + R'(f_m - f_L)) dG(\theta) \hat{\theta}_L - \int_{f_m}^{f_L} (\lambda + R'(2\theta - f_m - f_L)) dG(\theta) \hat{\theta}_L + \int_{\theta > f_L} (\lambda + R'(f_L - f_m)) dG(\theta) \hat{\theta}_L = 0. \nonumber\]  

(12)

Now, suppose the consensus forecast were \( \bar{f}_m = f_m - \delta \) \( \delta > 0 \). Then the derivative of \( w_L \) at \( f_L \) will be

\[
\lim_{\varepsilon \to 0} E \left[ \frac{w_L(f_L + \varepsilon) - w_L(f_L)}{\varepsilon} \right] = -\int_{\theta \leq f_m - \delta} (\lambda + R'(f_m - f_L - \delta)) dG(\theta) \hat{\theta}_L - \int_{f_m - \delta}^{f_L} (\lambda + R'(2\theta - f_m + \delta - f_L)) dG(\theta) \hat{\theta}_L + \int_{\theta > f_L} (\lambda + R'(f_L - f_m + \delta)) dG(\theta) \hat{\theta}_L. \nonumber\]  

(13)

By a simple change of variable, we can write the RHS of equation (13) as

\[
-\int_{\theta \leq f_m - \delta} (\lambda + R'(f_m - f_L - \delta)) dG(\theta) \hat{\theta}_L - \int_{f_m - \delta}^{f_L + \delta} (\lambda + R'(2\theta - f_m - f_L - \delta)) dG(\theta) \hat{\theta}_L + \int_{\theta > f_L + \delta} (\lambda + R'(f_L + \delta - f_m)) dG(\theta) \hat{\theta}_L \nonumber
\]

which is \( H(f_L + \delta) \). By the optimality of \( f_L \),

\[
\lim_{\delta \to 0} E \left[ \frac{H(f_L + \delta) - H(f_L)}{\delta} \right] < 0 \nonumber
\]

27
which implies that \( H(f_L + \delta) < H(f_L) = 0 \). Hence,

\[
\lim_{\varepsilon \to 0} E \left[ \frac{w_L(f_L + \varepsilon) - w_L(f_L)}{\varepsilon} \right]_{\hat{\theta}_L, f_m^*} < 0
\]

and the last analyst will shade his forecast in the direction of revision in the outstanding consensus.

Finally, it is straightforward to show that

\[
\lim_{\varepsilon \to 0} E \left[ \frac{w_L(f_L - \delta + \varepsilon) - w_L(f_L - \delta)}{\varepsilon} \right]_{\hat{\theta}_L, f_m^*} > 0
\]

which implies that the last analyst will bias his forecast by less than \( \delta \) in response to the new mean. That is,

\[
0 < \frac{\partial \hat{\theta}_L - f_m}{\partial \hat{\theta}_L - f_m} < 1.
\]

**Proof of Proposition 2:**

(i) Suppose \( f_m < \hat{\theta}_L \). Then by Lemma 2, \( f_L > f_m \). So we need to show that \( w_L(f_L) \) is increasing in \( f_L \) over \( (f_m, \hat{\theta}_L) \). Consider two reports by the last analyst: \( f_L = \hat{\theta}_L - \xi + \varepsilon \) and \( f_L = \hat{\theta}_L - \xi \), where \( 0 \leq \xi < \hat{\theta}_L - f_m \). The payoffs from these two reports can be used to compute the derivative of \( w_L(f_L) \) at \( f_L = \hat{\theta}_L - \xi \). This is given by

\[
\lim_{\varepsilon \to 0} E \left[ \frac{w_L(\hat{\theta}_L - \xi + \varepsilon) - w_L(\hat{\theta}_L - \xi)}{\varepsilon} \right]_{\hat{\theta}_L} = -\int_{\theta \leq f_m} (\lambda + R'(f_m - \hat{\theta}_L + \xi)) dg(\theta|\hat{\theta}_L) \\
- \int_{f_m}^{\theta \leq \hat{\theta}_L - \xi} (\lambda + R'(2\theta - \hat{\theta}_L + \xi - f_m)) dg(\theta|\hat{\theta}_L) \\
+ \int_{\theta > \hat{\theta}_L - \xi} (\lambda + R'(\hat{\theta}_L - \xi - f_m)) dg(\theta|\hat{\theta}_L). \tag{14}
\]

Consider the second term on the RHS of (14). Since \( \theta < (\hat{\theta}_L - \xi) \), we have \( R'(2\theta - \hat{\theta}_L + \xi - f_m) \leq R'(\hat{\theta}_L - \xi - f_m) \) by convexity of \( R(\cdot) \). Also \( R'(f_m - \hat{\theta}_L + \xi) \leq R'(\hat{\theta}_L - \xi - f_m) \). Hence,

\[
\lim_{\varepsilon \to 0} E \left[ \frac{w_L(\hat{\theta}_L - \xi + \varepsilon) - w_L(\hat{\theta}_L - \xi)}{\varepsilon} \right]_{\hat{\theta}_L} > -\int_{\theta \leq \hat{\theta}_L - \xi} (\lambda + R'(\hat{\theta}_L - \xi - f_m)) dg(\theta|\hat{\theta}_L) \\
+ \int_{\theta > \hat{\theta}_L - \xi} (\lambda + R'(\hat{\theta}_L - \xi - f_m)) dg(\theta|\hat{\theta}_L).
\]

Again, by the symmetry of \( g(\theta|\hat{\theta}_L) \) about \( \hat{\theta}_L \), we get

\[
\lim_{\varepsilon \to 0} E \left[ \frac{w_L(\hat{\theta}_L - \xi + \varepsilon) - w_L(\hat{\theta}_L - \xi)}{\varepsilon} \right]_{\hat{\theta}_L} > \int_{\hat{\theta}_L - \xi}^{\hat{\theta}_L + \xi} (\lambda + R'(\hat{\theta}_L - \xi - f_m)) dg(\theta|\hat{\theta}_L) \geq 0,
\]

\( \forall 0 \leq \xi < \hat{\theta}_L - f_m \). The last analyst’s payoff is therefore increasing in \( f_L \) over \( (f_m, \hat{\theta}_L) \). Hence, the optimal \( f_L \) will be such that \( f_m < \hat{\theta}_L < f_L \).

The proof that \( f_L < \hat{\theta}_L < f_m \) if \( \hat{\theta}_L < f_m \) is analogous.

**Proof of Proposition 3:**

Without loss of generality, normalize \( \hat{\theta} \) to zero, and suppose that \( f_m < 0 \). Consider two feasible forecasts \( f_L, f_L = f_L + \varepsilon, \varepsilon > 0 \) and where since we have convex relative performance compensation, \( f_L \) must be positive. Then for \( \varepsilon \) sufficiently small, \( f_L \) is a relatively more profitable forecast than \( f_L \) when \( g_L \) than when \( g_L \):

\[
E \left[ -\lambda |f_L - \theta| + R(|f_m - \theta| - |f_L - \theta|) \right]_{g_L} = \left[ -\lambda |f_L - \theta| + R(|f_m - \theta| - |f_L - \theta|) \right]_{g_L}
\]

\[
- E \left[ -\lambda |f_L - \theta| + R(|f_m - \theta| - |f_L - \theta|) \right]_{g_L} = \left[ -\lambda |f_L - \theta| + R(|f_m - \theta| - |f_L - \theta|) \right]_{g_L}
\]

\[
= \int_0 \left[ -\lambda |f_L - \theta| - |f_L - \theta| + R(|f_m - \theta| - |f_L - \theta|) - R(|f_m - \theta| - |f_L - \theta|) \right] (g_L(\theta) - g_L(\theta)) d\theta > 0.
\]

To see this last inequality, note for any transfer of probability mass from \((-b, b)\) to \((-a, a)\), \( b > a > 0 \) associated with a mean and symmetry preserving reduction in uncertainty, that
(i) The relative absolute error from submitting a forecast closer to the mean does not increase:

\[ |f_E - a| + |f_E + a| - ([f_E - a] + [f_E + a]) - |f_E - b| + |f_E + b| - ([f_E - b] + [f_E + b]) \leq 0. \]

where the inequality is strict if and only if \( b > f_E > a \) or \( f_E > b > f_E \). To see this, note that for \( \theta \geq 0 \),

\[ |f_E - \theta| + |f_E + \theta| - ([f_E - \theta] + [f_E + \theta]) \]

is zero for \( \theta > f_E \) and then increases linearly with \( \theta \in [f_E, f_E] \), and is equal to \( f_E - f_E < 0 \) for \( 0 < \theta < f_E \). Hence, if \( a, b > f_E \) both terms are zero so the difference is zero, and if both \( a, b < f_E \), then both terms are \( f_E - f_E \), so the difference is zero. Otherwise the difference is strictly negative.

(ii) The strict convexity of \( R(\cdot) \) ensures that \( f_E \) has a marginally greater relative expected payoff than \( f_E \) when the gamble is \((-a, a)\), than \((-b, b)\):

\[
\lim_{\epsilon \to 0} \frac{1}{\epsilon} \left[ R(|f_m - a| - |f_E - a|) - R(|f_m - a| - |f_E - a|) + R(|f_m + a| - |f_E + a|) - R(|f_m + a| - |f_E + a|) \right.
\]

\[
- (R(|f_m - b| - |f_E - b|) - R(|f_m - b| - |f_E - b|) + R(|f_m + b| - |f_E + b|) - R(|f_m + b| - |f_E + b|)) \right]
\]

\[
= -R(|f_m - a| - |f_E - a|) - R(|f_m + a| - |f_E + a|) + R(|f_m - b| - |f_E - b|) + R(|f_m + b| - |f_E + b|) > 0
\]

References


