5-2006

Loss Function Asymmetry and Forecast Optimality: Evidence from Individual Analysts' Forecasts

Stanimir Markov
Emory University

Min Yen TAN
Singapore Management University, minyen@aslouken.com

Loss function asymmetry and forecast optimality:
Evidence from individual analysts' forecasts

Stanimir Markov
(Stanimir_Markov@bus.emory.edu)

and

Min-Yen Tan
(Minyen_Tan @bus.emory.edu)

Goizueta Business School
Emory University
1300 Clifton Rd.
Atlanta, GA 30322

This version: May 9, 2006

Abstract:

We examine the optimality of quarterly earning forecasts issued by individual analysts. When we conduct Ordinary Least Squares (OLS) and Least Absolute Deviations (LAD) analyses, which assume loss function symmetry, we reject the null of forecast optimality at 5% significance level more than 5% of the time. Relaxing the symmetry assumption reduces the frequency of rejections below 5%. We demonstrate that the cross-sectional variation in the asymmetry parameter of the loss function is related to analyst employment. Overall, our evidence is consistent with the joint hypothesis of asymmetric loss and forecast optimality rather than the alternative of symmetric loss and lack of optimality.

We thank Orrie Barron, Sudipta Basu, Murgie Krishnan, Larry Brown, Marty Butler, Andy Leone, Jing Liu, Bruce Miller, Mike Parzen, Grace Pownall, Ane Tamayo, and seminar participants at Emory University, Georgia State University, and Penn State University. We also thank seminar participants at Baruch College and UCLA for their comments on a related paper. Finally, we thank Ron Harris and Christopher Baum for programming assistance and IBES for analysts' data.

Introduction

Many studies have documented that consensus forecast errors are predictable based on easily accessible public information, and have then concluded that financial analysts' forecasts are not optimal (Mendenhall, 1991; Abarbanell and Bernard, 1992).[1] These conclusions have been questioned recently from two perspectives. The first one questions the evidence of forecast error predictability based on the findings' lack of robustness (Abarbanell and Lehavy, 2003; Cohen and Lys, 2003). The second one questions the inference about lack of optimality. For example, Basu and Markov (2004) argue that this inference is driven by an inappropriate assumption of quadratic loss, and they document almost no predictability in consensus forecast errors when conducting the tests under the assumption of linear loss. They conclude that analysts' expectations are consistent with rationality and that suggestions that investors make the same cognitive mistakes are premature.[2]

We extend prior research on forecast rationality in two ways. Our first innovation is that we explore loss function-based explanations for the evidence of forecast predictability at the individual analyst level (Lys and Sohn, 1990; Jacobs and Lys, 1999; Mikhail et al., 2003, among others). While most of the debate about analyst rationality takes place at the consensus level, there are good reasons to study forecast optimality and information use at the individual analyst level. First, heterogeneity in loss functions makes ambiguous the interpretation of the tests of

[1] While different studies use different terms in interpreting the evidence of forecast error predictability, we think that the common theme in these interpretations is that the forecasts are not optimal. For example, serial correlation in forecast errors is explained with analysts' underreaction to information, or inefficient information use, due to the existence of cognitive biases (Mikhail et al., 2003) and analysts' underestimation of the persistence of their forecast errors (Mendenhall, 1991). See Kothari (2001) and Ramnath et al. (2006) for additional references to studies that document forecast error predictability, and infer that forecasts lack optimality.
[2] A third perspective is taken by Markov and Tamayo (2006), who explore Bayesian learning as an explanation for the evidence of forecast error predictability.

rationality at the consensus level.[3] When we average the forecasts to construct the consensus, we average not only over individual analysts' mistakes, but also over heterogeneous loss functions. Since no optimizing individual constructed the mean forecast to minimize her expected loss, it is not clear what rejection of the null, or failure to reject the null at the consensus level, means for how analysts use information. Second, forecast users make decisions based on access to specific analysts. Thus, knowledge of individual analysts' forecasting objectives and forecast inefficiencies is of enormous practical importance to forecast users.

Our second innovation is that we conduct bootstrap inferences to address Abarbanell and Lehavy's (2003) criticism about prior studies' lack of attention to the severe non-normality of the forecast errors. Bootstrap does not require assumptions about error distribution or infinitely large samples, and, in most samples, it typically provides a more accurate approximation to the distribution of an estimator than asymptotic theory (Horowitz, 1997; Efron and Tibshirani, 1998).

We analyze a comprehensive sample of forecasts issued late in the current quarter over the period of 1985 to 2004 by 2,489 analysts employed by 303 investment firms. Assuming that financial analysts' loss function is symmetric, we find that analysts do not efficiently use information in past forecast errors and earnings. In particular, our OLS (LAD)-based tests reject the null of forecast optimality at a 5% significance level between 11% and 16% (6 % and 8%) of the time. These rejection rates are based on the bootstrap approach, which does not assume normality or large sample sizes. We note that the asymptotic theory rejection rates can be twice

[3] Heterogeneity in loss functions likely arises from differences in the mix and relative importance of various activities funding investment research (Cohen et al., 2006). It seems unlikely that the set of implicit and explicit contracts faced by an analyst employed by a firm that sells research for a fee are the same as the ones faced by an analyst working for a full-service investment firm where research is bundled with the sale of other products and services. See Krigman et al. (2001) and Irvine (2000) for evidence of investment research's contribution to underwriting and brokerage revenues.

as high as the bootstrap rejection rates. Thus, we are confident that the severe non-normality of the forecast errors discussed in Abarbanell and Lehavy (2003), while an important cause for concern in prior research on earnings forecast errors, cannot be an explanation for these findings.

We next explore whether the rejections of the hypothesis of forecast optimality are due to the fact that individual analysts' forecasting objectives are (i) heterogeneous and (ii) poorly approximated by a symmetric loss function implied in the OLS and LAD analyses. In particular, we conduct tests of forecast optimality under more general loss functions that nest the traditional quadratic and linear loss functions. The lin-lin and quad-quad loss functions generalize the traditional linear and quadratic loss functions by introducing a single asymmetry parameter, α, to allow for the possibility that the cost of a forecast error depends on its sign. A parameter value of 0.75 (0.25) means that the penalty for a positive forecast error is three times as high (low) as the penalty for a negative forecast error. The quadratic and linear loss function obtain when α is equal to 0.50.

We have several reasons for using the lin-lin and quad-quad loss functions. First, since the arguments for why financial analysts' objectives are best captured by a symmetric loss function are not very strong (Lambert, 2004), it seems justified to use an approach that allows rather than assumes loss function symmetry. Second, many practitioners and researchers believe that, late in the quarter, financial analysts' forecasting objectives are, in fact, to issue forecasts that are easy to beat (Matsumoto, 2002; Bartov et al., 2002; Richardson et al., 2004). Such incentives are inconsistent with a symmetric loss function, α=0.50, but consistent with α being less than 0.50. If at the end of the quarter the cost of underpredicting earnings is lower than the cost of overpredicting earnings, then analysts will systematically underpredict earnings by issuing forecasts that are easy to beat. Finally, these loss functions are conventional in the economics literature, with recent advances in econometric theory enabling the estimation of the asymmetry parameter α (Granger and Newbold, 1986; Elliott et al., 2004; Elliott et al., 2005; Rodriguez, 2005; Markov and Tan, 2005).

When we use the lin-lin function, which nests the linear loss function, we reject the null of optimality less than 5% of the time, which is consistent with pure chance. When we use the quad-quad function, which nests the quadratic function, we reject between 6% and 12% of the time. The asymmetry parameter of the loss function, α, is generally lower than 0.5, which means that the cost of underpredicting earnings is generally lower than the cost of overpredicting earnings. We conclude that the traditional OLS and LAD tests reject the hypothesis that all analysts have the same symmetric loss function rather than the hypothesis that their forecasts deviate from optimality.

An alternative explanation for our findings is that we are overfitting the data by introducing a parameter devoid of any economic substance. To preclude this explanation, we examine whether lin-lin and quad-quad loss functions are reasonable representations of analysts' forecasting objectives. We find that analysts with similar α's tend to work for the same employer, and that an analyst whose α deviates from that of her current employer is more likely to experience turnover. These results are especially strong for the asymmetry parameter of the lin-lin loss function, which validates the lin-lin function as a more reasonable representation of analysts' incentives than the quadratic loss function. We conclude that the variation in our estimates of the shape of the loss function is related to variation in incentives rather than noise, and that overfitting does not explain our findings of forecast optimality.

Next, we briefly discuss prior literature with an emphasis on studies that depart from the tradition of assuming symmetric loss. Section 3 describes our sample and reports the results from the traditional OLS and LAD analyses. Section 4 presents our main analyses based on Elliott et al.'s (2005) framework that allows for loss function asymmetry. Section 5 concludes.

2. Prior literature on analysts' loss functions and forecast optimality

The traditional approach to examining whether forecasts are optimal involves estimating with OLS the model

$$FE_{t+1} = \beta_0 + \beta_1 X_t + \delta_{t+1}, \tag{1}$$

where FE_{t+1} is analyst forecast error at time $t+1$ and X_t is a vector of variables known at the time of the forecast, t. Rejecting the hypothesis that β_0=0 and β_1=0 is viewed as evidence of lack of optimality as one can use publicly available information, X_t, to further reduce the mean squared forecast error. Positive values of β_1 are typically interpreted as evidence of underreaction to information or inefficient information use.

Basu and Markov (2004) point out that implicit in this approach is the assumption of a quadratic loss. In other words, the analyst is viewed as trying to minimize her mean squared error. Following Gu and Wu (2003), Basu and Markov (2004) argue that analysts' incentives are better represented by a linear loss function; analysts attempt to minimize their mean absolute error. They estimate a version of equation (1) using the LAD method rather than OLS and find that the coefficients are indistinguishable from the predicted values under the null of optimality. They still, however, often reject the null hypothesis that consensus forecasts are consistent with forecast optimality.

A number of recent studies on analyst forecasts question the appropriateness of the quadratic and linear loss functions (Lambert, 2004; Rodriguez, 2005; Clatworthy et al., 2005; Markov and Tan, 2005), and in particular, the assumption that the consequences of overpredicting earnings are the same as the consequences of underpredicting earnings. The symmetry of these loss functions is inconsistent with the view that analysts have incentives to issue optimistic reports early in the quarter and pessimistic forecasts late in the quarter, and the evidence of generally negative mean and median forecast errors for long-term forecasts and positive mean and median forecast errors for short-term forecasts (Richardson et al., 2004).[4]

Clatworty et al. (2005) examine the properties of financial analysts' forecast errors under an asymmetric loss function introduced first by Varian (1974), the linex loss function vis-à-vis the properties of the forecast errors under the linear loss function. A relation between forecast error bias and variance of the forecast error distribution is predicted under the asymmetric loss function, but not under the linear loss function (Christoffersen and Diebold, 1996, 1997). The evidence in Clatworty et al. (2005) rejects the linear loss function in favor of the asymmetric loss function. This evidence, however, does not shed light on the issue of forecast optimality, which is at the center of the other two empirical studies, Rodriguez (2005) and Markov and Tan (2005).

Both Markov and Tan (2005) and Rodriguez (2005) rely on the econometric framework of Elliott et al. (2005), also used in this study, to describe the asymmetry of the financial analysts' loss function and examine forecast optimality. Rodriguez (2005) analyzes a sample of 107 analysts and finds evidence consistent with forecast optimality. He also shows that risk aversion can be important for understanding the properties of analysts' forecast errors. An important difference between his study and our study is that our sample is much more comprehensive (2,489 analysts). This allows us to examine whether low power explains his failure to reject the null of optimality. In addition, we show that the individual analysts' loss function asymmetry varies systematically across firms, which alleviates the concern that the tests introduce an extra parameter devoid of economic content.

[4] The quadratic (linear) loss function predicts a zero mean (median) forecast error.

Markov and Tan (2005) examine the optimality of the consensus forecasts and find that consensus forecasts are more consistent with rationality in the period after Regulation FD became effective. We view empirical analyses of individual and consensus forecast errors as complementary. Researchers examine the properties of the consensus forecast, defined as the mean or median forecast, when their primary focus is on studying the beliefs of the marginal investor. They use the consensus forecast as a proxy for the unobservable marginal investor's beliefs.

However, there are good reasons to study individual analysts' errors. First, without knowledge of the forecasting objectives and rationality of individual analysts, we cannot really understand how the market for investment research functions. Second, since forecast users rely on investment research by individual analysts, it is important for them to know the forecasting objectives and biases of the analysts whose research they have purchased or consider for purchase. Third, if different analysts have different loss functions, then evidence about mean or median forecast becomes hard to interpret. For example, individual analysts may issue forecasts that are optimal under their own loss functions, but the average of their forecasts may be sub-optimal as each analyst pays attention *only* to her own loss function. The opposite is also possible. Individual analysts issue sub-optimal forecasts, but the average of their forecasts is optimal. Evidence of forecast optimality at the consensus level does not distinguish between the hypotheses that (i) all analysts have the same loss function and they make errors that cancel out, or (ii) analysts have different loss functions, make systematic mistakes, but there exists a loss function parameter value that rationalizes the consensus.[5]

[5] If analysts have different information sets, then tests of rationality at the consensus level can reject the null of rationality even if individual analysts are rational (Figlewski and Wachtel, 1983).

3. Tests of forecast optimality under loss function symmetry

In this section we describe our sample and report evidence about forecast optimality based on OLS and LAD regressions, which assume loss function symmetry.

3.1. Sample construction

Our primary data come from the Institutional Brokers Estimate System (I/B/E/S) database. We use the I/B/E/S Detail Earnings Estimate History File, which contains 1,564,054 individual analysts' forecasts of current quarter earnings of U.S. companies for the period from 1985 to 2004. We eliminate forecasts of quarter *t* earnings that are dated after the announcement of quarter t earnings (160,656 observations). We focus on forecasts that are issued in the second half of the quarter because they are the ones most likely to determine the earnings surprise on announcement dates (530,541 observations). These forecasts are more likely to incorporate information available at the beginning of the quarter (Soffer and Lys, 1999), and more likely to be systematically biased downward relative to forecasts issued at the beginning of the quarter (Richardson et al., 2004). We eliminate forecasts issued in the first half of the quarter (52,793), observations without two prior earnings announcements (41,928), observations where share price from two quarters ago is less than $1 (2,108), and observations without prior forecast errors (120,049). These filters ensure that we use the most up-to-date analyst forecasts, that we can examine analysts' use of publicly available information such as earnings and past forecast errors, and that we do not inflate forecast errors when we divide them by low stock price to alleviate heteroscedasticity concerns. The number of forecasts satisfying these criteria is 313,663. These forecasts were issued by 7,439 analysts. To ensure precision in our analyst-specific estimates, we require at least 30 observations for an analyst, which eliminates 42,942 forecasts (4,949 analysts). Finally, we drop forecasts issued by an analyst with I/B/E/S code of 0000000, as this code is assigned to all analysts who wish to stay anonymous (2,070). Our sample has 268,651 forecasts issued by 2,489 analysts.

3.2. Descriptive statistics

Panel A of Table 1 provides descriptive statistics about stock coverage for our sample of 2,489 analysts. These analysts were employed by 303 investment firms. They issued a total of 268,651 forecasts on 7,379 companies. Over her tenure, an analyst in our sample issued on average about 108 quarterly earnings forecasts for 22 firms over 27 quarters. The median number of forecasts issued, firms covered, and quarters on I/B/E/S tends to be lower, which suggests that there is a relatively high proportion of analysts who were very experienced, followed many stocks, and issued multiple forecasts. Likewise, some investment firms were much larger than others as the mean (median) number of analysts employed and stocks covered are 16 (4) and 178 (39), respectively. The reason for the relatively low number of analysts employed and stocks covered by an investment firm is that we include only analysts with at least 30 forecasts issued in the second half of the current quarter. As we noted above, there are about 5,000 analysts who issued fewer than 30 forecasts during our sample period.

We denote firm *j*'s quarterly I/B/E/S earnings per share (EPS) for quarter *t+1* as ${}_jA_{t+1}$, and analyst *i*'s forecast of firm *j*'s EPS for quarter *t+1* as ${}_jF^i_{t+1}$. The forecast error, denoted as ${}_jFE^i_{t+1}$, is defined as ${}_jA_{t+1} - {}_jF^i_{t+1}$. All variables are scaled by share price recorded for the earnings-announcement month of the quarter *t-1* obtained from I/B/E/S to alleviate heteroscedasticity concerns and are winsorized at the 1% level on both tails to eliminate outliers. The mean and median forecast errors of analyst *i* are calculated based on all forecasts, ${}_jF^i_{t+1}$

issued by analyst *i*. Thus, they are analyst-specific rather than analyst-firm-specific. There are both advantages and disadvantages to analyzing the properties of FE^i rather than ${}_jFE^i$. An obvious advantage is that we are less prone to survival bias; we do not require that analyst *i* issued at least 30 forecasts on firm *j*, which is the approach taken by Mikhail et al. (2003), but only that analyst *i* issued 30 forecasts. In addition, our choice to combine the distributions ${}_jFE^i$ and ${}_{j+1}FE^i$ is justified by Jacobs and Lys's (1999) findings about the existence of common components in the properties of the distribution of ${}_jFE^i$ and ${}_{j+1}FE^i$. The disadvantage is that analyst *i*'s forecast errors in quarter *t* are cross-correlated, a problem that we address later on by clustering observations or conducting a block bootstrap. We report mean, median, standard deviation, and the 25th and 75th percentile of the cross-sectional distribution of 2,489 mean and median forecast errors in Panel B of Table 1.

We find strong evidence that the distribution of median forecast errors is centered at a positive number. Both the mean and the median of the distribution of median forecast errors are positive, and we reject the null of zero median forecast errors at 5% level in favor of "greater than zero" 44% of the time. The frequency of rejecting the null in favor of "smaller than zero" is only 2% and consistent with chance. The evidence on mean forecast errors is mixed. The mean of the distribution of mean forecast errors is negative, -0.0002, but the median is zero. In addition, we obtain high rejections of the null in favor of "greater than zero" 22% of the time, and in favor of "less than zero", 14% of the time. The higher frequency of rejections of the null in favor of "greater than zero" is not necessarily inconsistent with the negative mean of the distribution of mean forecast errors. If statistical precision is higher in the sub-sample of analysts with positive forecast errors, then we may reject more often. Overall, the general tendency is toward issuing low, beatable forecast errors that result in positive earnings "surprises."

3.3 OLS and LAD evidence about forecast optimality

The OLS (LAD) method is appropriate if analysts have quadratic (linear) loss function. Basu and Markov (2004) document that the estimated coefficients deviate less from the predicted values when they use the LAD method than when they use the OLS method and argue that consensus annual forecasts are generally consistent with optimality under the linear loss function. However, even with the LAD method, Basu and Markov (2004) often reject the null hypothesis of forecast optimality.

To provide a baseline for our empirical analyses, we first conduct tests of forecast optimality under the traditional assumption of symmetric linear or quadratic loss function. In particular, for every analyst *i*, we regress the forecast errors ($_{j}FE^{i}_{t+1}$) on intercept and information variables known to the analyst at the time of the forecast; past forecast errors ($_{j}FE^{i}_{t}$) and intercept (Model 1), earnings at lags 1 and 2 ($_{j}A_{t}$ and $_{j}A_{t-1}$) and intercept, (Model 2), and past forecast errors and past earnings at lags 1 and 2 ($_{j}FE^{i}_{t}$, $_{j}A_{t}$, and $_{j}A_{t-1}$) and intercept (Model 3).[6] We estimate these models using OLS and LAD methods. In Table 2 we report the mean parameter estimates from our 2,489 analyst-specific regressions. The last column summarizes the results from our 2,489 tests of optimality; we report the percentage of times that we reject the null of optimality at 5%. In view of the severe non-normality of forecast error distribution (Abarbanell and Lehavy, 2003), we conduct both asymptotic-theory and bootstrap inferences.

[6] Consistent with prior literature, we assume that the cost of accessing and processing publicly available information for a sell-side analyst is zero. We think that this assumption is more appropriate for analyst's own forecast errors and past earnings than for other variables such as accruals and stock returns, hence our focus on forecast errors and past earnings. Our analysis, however, can be extended to other information variables.

The mean coefficients on past forecast errors and past earnings are positive, which is consistent with prior evidence. The common interpretation is that analysts do not understand the properties of the quarterly earnings process, or underreact to earnings information. We note that the coefficient on past forecast errors from our analyst-specific regressions is 0.0823, which is lower than the corresponding coefficient in Mikhail et al. (2003), 0.1400. Given that our sample covers a more recent period, 1985 to 2004, vis-à-vis 1980 to 1995 in Mikhail et al. (2003), and that analysts issue more accurate forecasts in recent years (Brown and Caylor, 2005), we view our findings as being in the same range.

When we base our inferences on asymptotic theory, we get very high rejections of the null hypothesis that all coefficients are jointly equal to zero. The OLS regressions reject the null hypothesis at 5% level between 34% and 56% of the time. When we use the LAD method, we reject the null between 49% and 74% of the time. These results are not directly comparable, however. The LAD regressions assume independent and identically distributed errors, while the OLS regressions assume only time independence. With heteroscedasticity and cross-correlation likely present in the data, the LAD frequency of rejections likely overstates the true frequency of rejections. Due to the lack of asymptotic results about the distribution of LAD parameters in the presence of heteroscedasticity or cross-correlation that is analogous to the results about OLS parameters, using the bootstrap in LAD analysis should be the preferred choice.[7]

[7] Rogers (1992) provides Monte Carlo evidence that LAD asymptotic standard errors in the presence of heteroscedasticity significantly understate the true standard errors; he recommends that inferences be based on bootstrap standard errors instead.

The Jarque-Bera test (results not tabulated) rejects the null hypothesis that the residuals come from a normal distribution more than 90% of the time.[8] The distribution of residuals has both high skewness and kurtosis. The lack of normality of the residuals combined with the limited sample sizes used in the analyst-specific estimations makes asymptotic inferences suspect and provides additional motivation for conducting bootstrap inferences.

We view observations from different quarters as independent and observations from the same quarter as dependent. Therefore, we sample blocks of observations rather than individual observations; each block consists of observations for the same quarter. Drawing observations by quarters ensures that each draw is independent of the other draws. We draw 1,000 samples with replacement from our original sample. We estimate our model 1,000 times to obtain the parameters' bootstrap distribution. This distribution is the basis for calculating standard errors and for conducting statistical tests.

We document a substantial drop in rejection rates when we conduct bootstrap inferences. In the case of the OLS regression, the rejection rates are now between 12% and 20%, which are still too high to be explained by chance. In the case of LAD regressions, we obtain rejection rates between 6% and 8%. Given the substantial drop in rejection rates, we recommend that researchers studying the properties of analyst forecast errors conduct bootstrap inferences in addition to asymptotic theory inferences.[9]

---

[8]The Jarque-Bera test determines whether the sample skewness and kurtosis are unusually different from their expected values under the normality assumption of 0 and 3. The test statistic is $\frac{n}{6}\left[S^2 + \frac{(K-3)^2}{4}\right] \sim \chi^2_2$, where n is number of observations, S and K are sample skewness and kurtosis. It is possible that non-normality of the forecast errors is at least partly driven by pooling quarter *t*'s forecast errors of analyst *i*.

[9] Our recommendation is most pertinent to studies on individual analysts where we have a combination of small sample size and highly non-normal residuals.

3.4. Optimality tests when the true loss function differs from the one assumed by the researcher

The evidence so far suggests lack of forecast optimality at the individual analyst level assuming, of course, that analysts have symmetric loss function. The objective of the analysis in this sub-section is to demonstrate that an earnings forecast constructed under a slightly asymmetric loss function would be found sub-optimal in OLS and LAD tests, which assume symmetric loss.

3.4.1. Generating optimal forecasts

We simulate forecasts of A_{t+1} that incorporate all information in A_t and A_{t-1} to minimize the sum of absolute forecast errors, where positive and negative forecast errors are weighted by α and $(1-\alpha)$ respectively; $\alpha \in \{0.40, 045, 0.50, 0.55, 0.60\}$. In particular, we estimate the regression

$$A_{t+1} = \chi_{0,\alpha} + \chi_{1,\alpha} A_t + \chi_{2,\alpha} A_{t-1} + \varepsilon_{t+1,\alpha}\ , \qquad (3)$$

at various quantiles of the earnings distribution at time t+1 for the same values of $\alpha; \alpha \in \{0.40, 045, 0.50, 0.55, 0.60\}$. The estimated coefficients, by construction, describe the conditional quantile of A_{t+1} as a linear function of A_{t+1} and A_{t+1}. They minimize the sum of absolute errors where positive and negative errors are weighted α and $(1-\alpha)$ respectively.[10]

Coefficients and standard errors are reported in Panel A of Table 3. If a forecaster has a loss function in which the cost of underpredicting earnings is 2/3 of the cost of over-predicting earnings (α=0.40), then her optimal forecast of A_{t+1} would be equal to $0.006 + 0.7034 * A_t + 0.2086 * A_{t-1}$. A forecaster with an α of 0.60 would construct her forecast

[10] In contrast, OLS coefficients describe the conditional mean of the dependent variable as a function of the dependent variable, and minimize the sum of squared residuals.

differently; her forecast would be equal to $0.0038 + 0.6702*A_t + 0.1724*A_{t-1}$. In other words, each regression decomposes A_{t+1} into forecast $\hat{A}_{t+1,\alpha}$ and forecast error, $\hat{\varepsilon}_{t+1,\alpha}$. The forecast and the forecast error are indexed by α because they depend on how errors are penalized.

3.4.2. Are generated forecasts optimal in OLS and LAD analyses?

In the second step of our analysis, we examine whether the forecast errors $\hat{\varepsilon}_{t+1,\alpha}$, can be predicted by A_t and A_{t-1} using OLS and LAD regression methods. We estimate the model

$$\hat{\varepsilon}_{t+1,\alpha} = \beta_{0,\alpha} + \beta_{1,\alpha} A_t + \beta_{2,\alpha} A_{t-1} + \delta_{t+1} \tag{4}$$

using OLS and LAD regression methods. In Panel B of Table 3, we report the OLS and LAD coefficients from the estimation of equation (4) as well as F-statistic from the joint test that all coefficients equal zero. We reject the null of forecast optimality in all specifications with the exception of α=0.5 in the LAD regression.[11] These rejection rates are based on the more conservative bootstrap approach. It is important to note that not only the intercept but also the slope coefficients significantly deviate from the predicted values of zero. We conclude that forecast that are constructed to be optimally under asymmetric loss, appear sub-optimal in OLS and LAD tests that assume symmetric loss.[12]

In practice, however, we do not know the value of the asymmetry parameter of the analyst's loss function. Elliott et al. (2005) develop a method to estimate this parameter and

[11] This is not surprising since we generate the forecasts and tests for optimality under linearity and for the same α.
[12] The bootstrap was implemented by drawing pairs of observations. Thus, the bootstrap rejection rates are probably too high given the lack of independence in the cross-section. In our OLS analysis we also used heteroscedasticity – consistent and robust to intra-quarter cross-correlation standard errors, which rely on asymptotic theory, and similarly rejected the null of forecast optimality for all values of α.

examine the extent to which the forecasts are consistent with forecast optimality.[13] The next section provides a brief overview of their framework, used in our study, and reports our main findings.

4. Tests of forecast optimality that do not assume symmetry

4.1 Econometric method

The consequences of making an inaccurate forecast are represented by the loss function

$$L(p,\alpha,\theta) \equiv \left[\alpha + (1-2\alpha)\cdot 1\left(A_{t+1} - f_{t+1}(\theta) < 0\right)\right]\cdot \left|A_{t+1} - f_{t+1}(\theta)\right|^{p}. \quad (5)$$

The second term, $\left|A_{t+1} - f_{t+1}(\theta)\right|^{p}$ is the analyst's forecast error defined as the difference between earnings, A_{t+1} and the earnings forecast, $f_{t+1}(\theta)$. The latter is a linear function of variables W_t observed by the analyst at time t, $f_{t+1}(\theta) = \theta \cdot W_t$. Different values of θ represent different forecasting rules, which in turn result in different forecast errors. The first term in equation (5), $\left[\alpha + (1-2\alpha)\cdot 1\left(A_{t+1} - f_{t+1}(\theta) < 0\right)\right]$ makes the cost of a forecast error conditional on its sign. If α is equal to 0.5, then positive and negative forecast errors are equally costly. In fact, when α=0.50 and p=1 or p=2, the loss function reduces to that of the familiar cases of a linear or a quadratic loss function, widely used in prior research on financial analysts. If α>0.5, however, then overpredictions are less costly to the analyst. In other words, the analyst has incentives to overpredict earnings. In sum, our ignorance about the analysts' objectives consists only of not knowing the value of the single parameter α, $\alpha \in (0,1)$.

[13] The general idea of recovering a parameter from the data that is most consistent with optimizing behavior and assessing the extent to which optimality restrictions are satisfied in the data appears first in Hansen and Singleton's seminal (1982) study.

As an optimizing agent, the analyst chooses a forecasting rule $f_{t+1}(\theta) = \theta \cdot W_t$ to minimize her expected loss

$$\min_{\theta} E\left[L(p, \alpha, \theta)\right].^{14} \tag{6}$$

If θ is chosen optimally, then the forecast errors must satisfy the first-order conditions

$$E\left[W_t \cdot \left(1\left(\varepsilon_{t+1}^{*} < 0\right) - \alpha\right) \cdot \left|\varepsilon_{t+1}^{*}\right|^{p-1}\right] = 0\,, \tag{7}$$

where $\varepsilon_{t+1}^{*} = A_{t+1} - \theta^{*} W_t$.[15] Having access only to a subset of the information available to the analyst at time *t*, which we denote as V_t, does not prevent us from estimating α. Since an optimizing analyst exploits any information available to her at time *t*, we can substitute V_t for W_t in the moment conditions and use the corresponding sample moments to back out the asymmetry parameter α.

As long as we have more moment conditions than parameters to estimate, we are able to recover the asymmetry parameter without ad hoc rationalizing the forecasts. The reason for this is that the same α has to set two or more sample moments simultaneously to zero. Our estimator of α minimizes a quadratic form

$$q = g_T(\alpha)' S g_T(\alpha) \tag{8}$$

where $g_T(\alpha)$ is the sample equivalent of equation (7), and S is a weighting matrix. Our weighting matrix is the inverse of the covariance matrix of the moment conditions, which

[14] In other words, we view the forecast as a choice that analysts make in trying to enhance their welfare—a departure from the literature's tradition of viewing forecasts as exogenously given (Demski, 2004). In a survey of the use of expectations in accounting research, Demski forcefully argues that reliance on exogenous expectations structures limits the depth and boundaries of teaching and research (p. 519).
[15] This is proposition 1 in Elliot et al. (2005).

minimizes the asymptotic variance of the GMM estimator.[16,17] In sum, GMM picks the value of α that minimizes the squared distance between zero and the moment conditions divided by the covariance of the moment conditions.

Hansen's J-statistic, which is equal to T times the minimized value of the quadratic form, measures the distance between zero and the moment conditions, or how well the first order conditions from the analyst's optimization problem, are satisfied in the data. It follows a chi-square distribution with degrees of freedom equal to number of moments minus 1, the number of parameters estimated. Large values of the J-statistic mean that the distance between zero and the moment conditions is too large to be explained by chance, and that we should reject the joint hypothesis that (i) analyst's loss function is well approximated by equation (5) and (ii) the forecasts are optimal.[18]

4.2. Main findings

We use as instruments the regressors in the OLS and LAD regressions of Table 2; past forecast errors and an intercept; past earnings and an intercept; forecast errors, past earnings, and an intercept. Panel A of Table 4 provides evidence about the cross-sectional distribution of the asymmetry parameter and the frequency with which we reject the symmetry null of α=0.50 in favor of α>0.50 or α<0.50. In the case of the lin-lin function, we reject the null of symmetry at the 5% level in favor of α<0.50 about 80% of the time. We conclude that analysts have incentives to systematically underpredict earnings (Brown, 2001, Bartov et al., 2002, among

[16] The weighting matrix determines the relative importance of setting a particular moment condition to zero when estimating α.
[17] We used Stata's **ivreg2** command and its options **cluster** and **robust** to produce heteroscedasticity-consistent and robust to intra-quarter cross-correlation standard errors. Sample code is available upon request.
[18] It is standard to refer to the test as a test of over-identifying restrictions. In our setting, the restrictions hold if the forecast solves the minimization problem (equation (6)), and, thus, we refer to the test as an optimality test. Cochrane (2001) provides an insightful discussion of the J-test and its applications to tests of asset pricing models.

others). We find no evidence that any analysts have incentives to overpredict earnings since the frequency of rejections, which occur 2% of the time, is low enough to be due to chance. The cross-sectional mean and median are about 0.31, which means that the cost of underpredicting earnings by 1 cent is about three times lower than the cost of overpredicting earnings by 1 cent. In other words, analysts have incentives to consistently underpredict earnings.

The asymmetry parameter of the quad-quad loss function is 0.48. This should not be surprising given our sample evidence that the mean forecast error is very close to zero. As we have argued above, the assumption that analysts' true loss function is quadratic (linear) leads to the prediction that the mean (median) forecast error is equal to zero. However, we observe a significant amount of variation in the cross-sectional distribution of the asymmetry parameter, as the 25th and 75th percentile are about 0.32 and 0.63 respectively. Thus, we get high frequency of rejections of the null not only in favor of $\alpha<0.50$, between 27% and 32%, but also in favor of $\alpha>0.50$, between 24% and 30% of the time. The difference between $\alpha_{quad\text{-}quad}$ and $\alpha_{lin\text{-}lin}$ is due at least partially to the fact that the quad-quad loss function estimations use squared forecast errors, which increases the sensitivity of our estimates to extreme observations in either tail of the distribution.

In Panel B of Table 4 we present the frequency with which we reject the null of forecast optimality for the lin-lin and quad-quad functions. To stress the significance of relaxing the symmetry assumption, we also present the rejection frequencies when we assume symmetry (OLS and LAD analysis). In the case of the lin-lin loss function, the rejection frequency is about 5%, which is consistent with pure chance. We conclude that relaxing the symmetry assumption implicit in the LAD estimations significantly changes our inferences about forecasts' apparent lack of optimality.

In the case of the quad-quad loss function, we reject the null of forecast optimality less often than in the case of OLS estimations, but more than 5% of the time: between 8% and 12% of the time under the quad-quad loss function, and between 11% and 16% of the time under the quadratic loss function. The evidence about the forecasts' lack of optimality is weakened, but still existent.

We conclude that relaxing the symmetry assumption is generally important for making correct inferences about forecast optimality. Under the lin-lin loss function, which nests the traditional linear loss function, the evidence is consistent with forecast optimality. Under the quad-quad loss function, which nests the traditional quadratic loss function, the evidence is inconsistent with forecast optimality. The frequency of rejections of the null hypothesis of forecast optimality is slightly higher than what we would expect by pure chance.

A maintained assumption of this study is that the asymmetric loss function is a meaningful representation of the forecaster's incentives, and that variation in α is not just statistical noise, but captures variation in incentives. We document slightly higher rejection rates of the null under the quad-quad loss function, which raises the question of which specification is more appropriate. To validate the lin-lin and quad-quad loss functions as reasonable representations of analysts' forecasting objectives and to help determine which specification is more reasonable, we next explore the link between the estimated asymmetry parameters of the lin-lin and quad-quad loss functions and analyst employment.

4.3. Validating the asymmetry parameter

We view analysts employed by the same firm as conducting investment research in the same institutional setting, which means that they should face similar sets of explicit and implicit compensation contracts. If α is informative about analyst incentives, not just noise, we should observe that analysts employed by the same firm have similar α's. We test this prediction by regressing individual analysts' α's on 210 investment-firm indicator variables which are equal to 1 when an individual analyst is employed by a particular firm and are 0 otherwise. The difference between 210 and 303 (number of investment firms in Table 1) is due to the fact an analyst who works for more than one employer is paired up with her first employer. If employment does not influence α, then these indicator variables will not help explain the cross-sectional variation in α.

We estimate $\alpha_{lin\text{-}lin}$ and $\alpha_{lquad\text{-}quadn}$ using all instruments (reported in the first column of Panel A of Table 5).[19] The second and third columns report adjusted R-squared and F-stats from the test that the coefficients on the employment indicator variables are jointly equal to 0.

The employment indicator variables explain over 12% (about 8%) of the variation in $\alpha_{lin\text{-}lin}$ ($\alpha_{quad\text{-}quad}$). In all specifications we strongly reject the null that α is unrelated to employment. We conclude that variation in $\alpha_{lin\text{-}lin}$ and $\alpha_{quad\text{-}quad}$ is related to variation in incentives rather than driven by statistical noise. The higher R-squared and stronger rejections in the case of $\alpha_{lin\text{-}lin}$ suggest that variation in $\alpha_{lin\text{-}lin}$ captures better variation in the circumstances in which forecasting performances are evaluated. Thus, we have more confidence in our results from the tests under the lin-lin loss function.

We also examine whether the discrepancy between an analyst's α and the current employer's α, estimated by pooling observations of all analysts employed by the same firm, has an effect on analyst turnover. The higher the discrepancy between an analyst's and an employer's α, the more divergent the analyst's behavior from the behavior of the other analysts

[19] The results do not change when we use subsets of the available instruments.

employed at the same firm and the more likely that she will be separated from her current employer.

There are 22 employers with not enough observations to estimate employer's α. This reduces our sample from 2,489 analysts to 2,457 analysts. The number of turnover observations is reduced from 1,388 to 1,356.[20] We document a statistically significant positive effect of the discrepancy in $\alpha_{lin\text{-}lin}$ on the probability of a turnover (Panel B of Table 5). To help interpret the logit coefficient of 1.44, we calculate the probability of a turnover at the mean value of the independent variable and at the mean-plus-one standard deviation of the independent variable. We find that increasing the independent variable by one standard deviation increases the probability of a turnover from 54% to 57%. The fairly small effect is not surprising, however, given the parsimony of the loss function specification and the estimation error in $\alpha_{lin\text{-}lin}$. The coefficient on $\alpha_{quad\text{-}quad}$ is positive, but not statistically significant.

Finally, we examine whether the turnover event matches analysts and employers with similar α's. We predict that, when an analyst with high α (greater than 0.50) changes jobs, he is more likely to pair up with a high α employer. We examine this prediction conditional on the analyst being currently employed by a high α or low α firm. The need to estimate new employer's α reduced our sample from 1,356 turnover observations to 1,304 observations.

The results from our estimations are reported in Panel B of Table 5 (Model 2). In three out of four specifications, we document that an analyst with high α is more likely to pair up with an employer with high α.[21] In sum, we find that our parameter estimate, α, can explain turnover

[20] There are 681 analysts who have more than one turnover observation. The analyzed sample includes only the first turnover observations.

[21] In the fourth specification, we do not have enough variation in the independent variable to estimate the model.

outcomes. This evidence makes it less likely that our findings are due to overfitting the data by introducing a parameter devoid of any economic content.

5. Conclusions

Our study makes several contributions. First, we document that rejections of the null hypothesis of forecast optimality at the individual analyst level are driven by the invalid assumption of loss function symmetry. This assumption is (i) inconsistent with the argument that analysts have incentives to issue beatable forecasts and (ii) rejected in the data. After allowing for loss function asymmetry, we find evidence consistent with forecast optimality.

Second, we address Abarbanell and Lehavy's (2003) concern about the sensitivity of prior findings to distributional assumptions by conducting bootstrap inferences. While the use of bootstrap does not reverse the OLS findings of lack of forecast optimality, we find a significant drop in rejection rates. We believe that in some circumstances the use of bootstrap can change our inferences, and recommend its use.

Third, we further establish the role of incentives in investigations of the time-series properties of analyst forecast errors. While we are able to reject the quadratic and linear loss function in favor of the quad-quad and lin-lin loss functions, we acknowledge that there could be alternative loss functions that better approximate the analysts' forecasting problem. We think that proposing and estimating loss functions that better approximate analysts' forecasting problem should be an exciting area of future research.

Our evidence should not be considered to be a general indictment of the quadratic and linear loss functions. There could be circumstances in which symmetry is a valid assumption. Assuming symmetry because it is convenient, however, not only leads to invalid inferences about forecast optimality, but also prevent us from learning about the nature of the forecasters' incentives from the data.

There are other potential avenues for future research. In this study, we consider only a few information variables suggested by prior research as being inefficiently used by financial analysts. An immediate extension of our study is to examine additional information variables such as extreme past earnings (Easterwood and Nutt, 1999), accruals (Bradshaw et al., 2001), and past returns (Lys and Sohn, 1990). Another potential avenue for future research would be to examine the shape of the manager's loss function implicit in her earnings forecasts and the optimality of her earnings forecasts.

References

Abarbanell, J. S., and V. L. Bernard, 1992, Tests of Analysts' Overreaction/ Underreaction to Earnings Information as an Explanation for Anomalous Stock Price Behavior, The Journal of Finance 47 (3), 1181-1207.

Abarbanell, J. S., and R. Lehavy, 2003, Biased forecasts or biased earnings? The role of reported earnings in explaining apparent bias and over/underreaction in analysts' earnings forecasts, Journal of Accounting and Economics 36 (1-3), 105-146.

Bartov, E., D. Givoly, and C. Hayn, 2002, The rewards to meeting or beating earnings expectations, Journal of Accounting and Economics 33, 173-204.

Basu, S., and S. Markov, 2004, Loss function assumptions in rational expectations tests on financial analysts' earnings forecasts, Journal of Accounting and Economics 38, 171-203.

Bradshaw, M., S. Richardson, and R. Sloan, 2001, Do analysts and auditors use information in accruals? Journal of Accounting Research 39 (1), 45-74.

Brown, L., 2001, A temporal analysis of earnings surprises: Profits vs. losses, Journal of Accounting 39, 221-241.

Brown, L., and M. Caylor, 2005, A temporal analysis of quarterly earnings thresholds: Propensities and valuation consequences, The Accounting Review 80 (2), 423-440.

Christofferson, P., and F. Diebold, 1996, Further results on forecasting and model selection under asymmetric loss, Journal of Applied Econometrics 11, 561-572.

Christofferson, P., and F. Diebold, 1997, Optimal prediction under asymmetric loss, Econometric Theory 13, 808-817.

Clatworthy M., Peel, P., and P. Pope, 2005, Are analysts' loss functions asymmetric?, Working Paper, Lancaster University Management School.

Cochrane, J., 2001, Asset Pricing, Princeton University Press.

Cohen, A., Groysberg, B., and P. Healy, 2006, Which types of analyst firms are more optimistic? Journal of Accounting and Economics 41, 119-146.

Cohen, D. A., and T. Z. Lys, 2003, A note on analysts' earnings forecast errors distribution, Journal of Accounting and Economics 36 (1-3), 147-164.

Demski, J., 2004. Endogenous expectations, The Accounting Review 70 (2), 519-539.

Easterwood, J., and S. Nutt, 1999, Inefficiency in analysts' earnings forecasts: systematic misreaction or systematic optimism? The Journal of Finance 54 (5), 1777-1797.

Efron, B., and R.J. Tibshirani, 1998, An Introduction to the bootstrap, Chapman and Hall, New York.

Elliot, G., I. Komunjer, and A. Timmerman, 2005, Estimation and testing of forecast rationality under flexible loss, Review of Economic Studies 72, 1107-1125.

Elliot, G., I. Komunjer, and A. Timmerman, 2004, Biases in macroeconomic forecasts: irrationality, or asymmetric loss? Working Paper, University of California at San Diego.

Figlewski, S., and P. Wachtel, (1983), Rational expectations, information efficiency, and tests using survey data, Review of Economics and Statistics 65, 529-531.

Gu, Z., and J. Wu, 2003, Earnings skewness and analyst forecast bias, Journal of Accounting and Economics 35 (1), 5-29.

Granger, C., and P. Newbold, 1986, Forecasting Economic Series, 2nd Edition, Academic Press.

Hansen, L., and K. Singleton, 1982, Generalized instrumental variables estimation of nonlinear rational expectations models, Econometrica 50, 714-735.

Horowitz, J., 1997, Advances in econometrics: theory and applications, Vol. 3, Ed. by D. Kreps, and K. Wallis, Cambridge University Press.

Irvine, P., 2000. Do analysts generate trades for their firms? Evidence from the Toronto stock exchange, *Journal of Accounting and Economics* 30 (2), 209-226.

Jacobs, J., and T. Lys, (1999), Determinants and implications of the serial correlation in analysts' earnings forecast errors, Working Paper, University of Colorado and Northwestern University.

Kothari, S.P., 2001, Capital markets research in accounting, Journal of Accounting and Economics 31, 105-231.

Krigman, L., Shaw, W., and K. Womack, 2001, Why do firms switch underwriters? Journal of Financial Economics 60, 245-284.

Lambert, R., 2004, Discussion of analysts' treatment of non-recurring items in street earnings and loss function assumptions in rational expectations tests on financial analysts' earnings forecasts, Journal of Accounting and Economics 38, 205-222.

Lys, T., and S. Sohn, 1990, The association between revisions of financial analysts' earnings forecasts and security-price changes, Journal of Accounting and Economics 13 (4), 341-363.

Markov, S., and A. Tamayo, 2006, Predictability in financial analyst forecast errors: Learning or irrationality, Forthcoming in Journal of Accounting Research.

Markov, S., and Tan, Min-Yen, 2005, Separating the Effects of Asymmetric Incentives and Inefficient Use of Information on Financial Analysts' Consensus Earnings Forecast Errors, Available at SSRN: http://ssrn.com/abstract=768324.

Matsumoto, D., 2002, Management's incentives to guide analysts' forecasts, The Accounting Review 77, 483-514.

Mendenhall, R., 1991, Evidence of Possible Underweighting of Earnings-Related Information, Journal of Accounting Research 29, 170-180.

Mikhail, M., Walther, B., and R. Willis, 2003, The effect of experience on security analyst underreaction, Journal of Accounting and Economics 35, 101-116.

Ramnath, S., Rock S., and P. Shane, 2006, A review of research related to financial analysts' forecasts and stock recommendations, Available at SSRN: http://ssrn.com/abstract=848248.

Richardson, S., Teoh, S., and P. Wyscoki, 2004, The walk-down to beatable analyst forecasts: The role of equity issuance and insider trading incentives. Contemporary Accounting Research 21 (4), 885-924.

Rodriguez, M., 2005, Financial analysts' incentives and forecast biases, Working Paper University of California, San Diego.

Rogers, 1992, Quantile regression standard errors, Stata Technical Bulletin 9, 133-139.

Soffer, L., and T. Lys, 1999, Post-Earnings Announcement Drift and the Dissemination of Predictable Information, Contemporary Accounting Research, 16 (2) 305-331.

Varian, H., 1974, A Bayesian approach to real estate assessment, Studies in Bayesian Economics in honor of L.J. Savage, Edited by S. Feinberg and A. Zellner, Amsterdam: North Holland: 195-208.

Table 1. Descriptive statistics

Panel A. Number of analysts

We denote firm *j*'s quarterly I/B/E/S earnings per share (EPS) for quarter *t+1* as ${}_{j}A_{t+1}$, and the forecast for quarter *t+1* made by analyst *i* as ${}_{j}F^{i}_{t+1}$. ${}_{j}FE^{i}_{t+1}$ is the forecast error for the quarter *t+1* and equal to ${}_{j}A_{t+1} - {}_{j}F^{i}_{t+1}$. All variables are scaled by the share price recorded for the earnings-announcement month of the quarter *t-1* obtained from I/B/E/S to alleviate heteroscedasticity concerns and are winsorized at the 1% level on both tails to eliminate outliers. We include only observations with available past quarterly earnings, ${}_{j}A_{t}$ and, ${}_{j}A_{t-1}$, and past forecast, ${}_{j}F^{i}_{t}$, and $Price_{t-1}$ greater than \$1. We also drop analysts coded as 0000000 in I/B/E/S and analysts that do not have more than 30 forecasts in the sample. The final sample has 268,651 individual analyst forecast observations for 2,489 analysts.

Number of analysts	2,489	
Number of analyst investment firms	303	
Number of covered companies (unique IBES tickers)	7,379	
Number of quarterly earnings forecasts	268,651	
Mean (median) number of forecasts per analyst	107.94	(74.00)
Mean (median) number of quarters that an analyst is on IBES	27.13	(24.00)
Mean (median) number of analysts employed by an investment firm	16.22	(4.00)
Mean (median) number of companies covered by an analyst	22.43	(19.00)
Mean (median) number of companies covered by an investment firm	177.52	(39.00)

Panel B. Cross-sectional distribution of mean and median forecast errors

We calculate the mean and median forecast errors for each analyst. We report the mean, median, standard deviation, 25th percentile and 75th percentile of these analyst-specific mean and median forecast errors. For every analyst we test the null hypotheses of zero mean and zero median forecast error. We report the percentage of cases where we reject the null at 5% level in favor of "greater than 0" and in favor of "less than 0."

	Descriptive Statistics					Rejections of hypothesis of location parameter = 0 in favor of	
	Mean	Median	STD	25th perc	75th perc	Greater than 0	Less than 0
Mean FE	-0.0002	0.0000	0.0015	-0.0008	0.0006	21.53%	13.70%
Median FE	0.0003	0.0002	0.0006	0.0000	0.0005	43.75%	2.01%

Table 2. Tests of forecast optimality under loss symmetry: OLS and LAD tests

For each analyst, using OLS and LAD methods, we estimate the models

$$\text{Model 1: } {}_jFE^i_{t+1} = \beta_0 + \beta_1\, {}_jFE^i_t + {}_j\varepsilon_{t+1}$$
$$\text{Model 2: } {}_jFE^i_{t+1} = \beta_0 + \beta_2\, {}_jA_t + \beta_3\, {}_jA_{t-1} + {}_j\varepsilon_{t+1}$$
$$\text{Model 3: } {}_jFE^i_{t+1} = \beta_0 + \beta_1\, {}_jFE^i_t + \beta_2\, {}_jA_t + \beta_3\, {}_jA_{t-1} + {}_j\varepsilon_{t+1}$$

where ${}_jA_{t+1}$ is firm *j*'s quarter *t+1* earnings per share; ${}_jA_t$ and ${}_jA_{t-1}$ are defined analogously; ${}_jFE^i_{t+1}$ is analyst's forecast error in forecasting firm *j*'s quarter *t+1* earnings per share; ${}_jFE^i_t$ is defined analogously. We report cross-sectional means of 2,489 parameter estimates and the frequency with which the null hypothesis of forecast optimality (coefficients are jointly equal to zero) is rejected at the 5% significance level. We make the following assumptions in the estimation of the variance-covariance matrix. The combination of asymptotic theory inferences and OLS estimation method assumes that errors from different quarters are independent (STATA's **cluster** and **robust** options). The case of asymptotic theory inferences and LAD estimation method assumes that errors are independent and identically distributed (no easy fix available in STATA to deal with deviations from the i.i.d. ideal). The two cases of bootstrap inferences assume that observations from different quarters are independent. We draw 1,000 samples by drawing clusters of observations, with each quarter representing a separate cluster, rather than pairs of observations (STATA's **bootstrap** command and **cluster** option).

					Rejections of forecast optimality	
	β_0	β_1	β_2	β_3	Asymptotic theory	Bootstrap
Model 1, OLS	-0.0002	0.0823			33.82%	12.20%
Model 2, OLS	-0.0008		0.0499	-0.0090	48.43%	17.78%
Model 3, OLS	-0.0007	0.0649	0.0471	-0.0094	56.31%	19.87%
Model 1, LAD	0.0002	0.0656			48.55%	5.97%
Model 2, LAD	-0.0001		0.0285	-0.0079	62.37%	7.71%
Model 3, LAD	-0.0001	0.0549	0.0274	-0.0077	73.53%	7.46%

Table 3. OLS and LAD tests when the forecasts are generated under asymmetric linear loss function

In Panel A we report the results from the estimation of (3)

$$ {}_jA_{t+1} = \chi_{0,\alpha} + \chi_{1,\alpha}\, {}_jA_t + \chi_{2,\alpha}\, {}_jA_{t-1} + {}_j\varepsilon_{t+1,\alpha}\,, $$

where α is the forecasted quantile of the earnings distribution. α takes the values of 0.5, 0.45, 0.40, 0.55 or 0.6. We report the estimated coefficients and bootstrap standard errors in parentheses. We draw 1,000 samples by drawing pairs of observations (not clusters).

All variables are scaled by the share price recorded for the earnings-announcement month of the quarter t-1 and are winsorized at the 1% level on both tails to eliminate outliers. We have 180,065 firm-quarter observations. *** indicate significance at 1%.

Panel A. Forecasting future earnings under asymmetric loss function

Asymmetry parameter	β_0	β_1	β_2
α=0.40	0.0006	0.7034	0.2086
	(0.0000)	(0.0007)	(0.0007)
α=0.45	0.0013	0.7033	0.2004
	(0.0000)	(0.0007)	(0.0006)
α=0.50	0.0020	0.6988	0.1913
	(0.0000)	(0.0006)	(0.0006)
α=0.55	0.0028	0.6879	0.1819
	(0.0000)	(0.0006)	(0.0006)
α=0.60	0.0038	0.6702	0.1724
	(0.0000)	(0.0008)	(0.0008)

Panel B. OLS and LAD tests of forecast optimality when forecast is optimal under asymmetric linear loss function

In panel B we report the results from the estimation of model (4)

$$\hat{\varepsilon}_{t+1,\alpha} = \beta_{0,\alpha} + \beta_{1,\alpha\ j} A_t + \beta_{2,\alpha\ j} A_{t-1} + {}_j\delta_{t+1}$$

where $\hat{\varepsilon}_{t+1,\alpha}$, is the residual from the estimation of equation (3), and other all other variables are as defined earlier. We estimate this model using OLS and LAD estimation methods. We report parameter estimates, bootstrap standard errors (in parentheses) and F-stats from the joint test that all coefficients are equal to zero. We draw 1,000 samples by drawing pairs of observations. *** (**) [*] indicate significance at 1% (5%) [10%].

Asymmetry parameter	OLS test of optimality				LAD test of optimality			
	β_0	β_1	β_2	F test	β_0	β_1	β_2	F test
α=0.40	0.0023	-0.1998	0.0151	161.92***	0.0014	-0.0047	-0.0173	6.94***
	(0.0002)	(0.0134)	(0.0087)		(0.0000)	(0.0006)	(0.0006)	
α=0.45	0.0016	-0.1997	0.0233	149.98***	0.0007	-0.0045	-0.0091	2.59*
	(0.0002)	(0.0134)	(0.0087)		(0.0000)	(0.0006)	(0.0006)	
α=0.50	0.0009	-0.1952	0.0324	131.81***				
	(0.0002)	(0.0134)	(0.0087)					
α=0.55	0.0000	-0.1843	0.0418	107.49***	-0.0008	0.0109	0.0094	4.21**
	(0.0002)	(0.0134)	(0.0087)		(0.0000)	(0.0006)	(0.0006)	
α=0.60	-0.0009	-0.1666	0.0513	80.46***	-0.0018	0.0286	0.0189	24.25***
	(0.0002)	(0.0134)	(0.0087)		(0.0000)	(0.0006)	(0.0006)	

Table 4. Estimation of asymmetry parameter, α, and test of forecast optimality

For every analyst, we estimate α from the FOC (equation: $E\left[V_t \cdot \left(1\left({}_jFE_{t+1}^i < 0\right) - \alpha\right) \cdot \left|{}_jFE_{t+1}^i\right|^{p-1}\right] = 0$) using the two-step efficient Generalized Method of Moments (GMM). V_t is a vector of instruments, ${}_jFE_{t+1}^i$ is the forecast error, p=1 (2) is represents the lin-lin (quad-quad) loss function (firm and analyst indexes are suppressed for brevity). Instruments used in the estimation are reported in the first column. We report mean, median, 25th and 75th percentile of the cross-sectional distribution of analyst's α. α=0.5 represents the case of loss function symmetry. α>0.5 represents the case of analysts' incentives to issue optimistic forecasts. α<0.5 represents the case of analysts' incentives to issue pessimistic forecasts. For every analyst we test the hypothesis of loss function symmetry, α=0.5, allowing for heteroscedasticity and intra-quarter cross-correlation. We report the percentage of cases where we reject symmetry at 5% level in favor of α<.50 and in favor of α>0.50. All variables are scaled by the share price recorded for the earnings-announcement month of the quarter *t-1*, and are winsorized at the 1% level on both tails to eliminate outliers. The sample has 268,651 individual analyst forecast observations for 2,489 analysts.

Panel A. Estimation of asymmetry parameter, α

Model	Cross-sectional Distribution of *α*				Rejections of symmetry (*α=0.5)*	
	Mean	Median	Q25	Q75	*In favor of α<0.5*	*In favor of α>0.5*
Lin-lin loss						
Interc., ${}_jFE_t^i$	0.3149	0.3089	0.2296	0.3940	78.83%	1.44%
Interc., ${}_jA_t$, ${}_jA_{t-1}$	0.3154	0.3077	0.2332	0.3956	78.50%	1.56%
All	0.3055	0.3011	0.2141	0.3882	80.39%	2.01%
Quad-quad loss						
Interc., ${}_jFE_t^i$	0.4797	0.4913	0.3269	0.6332	27.25%	23.88%
Interc., ${}_jA_t$, ${}_jA_{t-1}$	0.4781	0.4937	0.3233	0.6316	27.33%	22.74%
All	0.4834	0.4936	0.2937	0.6642	32.38%	29.97%

Panel B. Test of forecast optimality

The Hansen's J-statistic[22] has a χ^2 distribution with degrees of freedom equal to the difference between the number of moments and number of parameters estimated and large values rejecting the null of forecast optimality. We report the percentage of times the null hypothesis is rejected at 5% in column 2. In the last column we report the percentage of times we reject optimality at 5% with traditional OLS and LAD tests. Instruments used in the estimation are reported in the first column.

Model	Rejections of forecast optimality when	
	Symmetry not assumed	Symmetry assumed (Table 2)
Lin-lin loss		
Interc., ${}_jFE_t^i$	6.03%	5.97%
Interc., ${}_jA_t$, ${}_jA_{t-1}$	4.90%	7.71%
All	4.22%	7.46%
Quad-quad loss		
Interc., ${}_jFE_t^i$	11.53%	12.20%
Interc., ${}_jA_t$, ${}_jA_{t-1}$	8.19%	17.78%
All	8.19%	19.87%

[22] The Hansen's J-statistic is equal to the minimized value of $J=Ng_t(\alpha)'Sg_t(\alpha)$, where N is the sample size, $g_t(\alpha)$ is the first order condition, $E\left[V_t\cdot\left(1\left({}_jFE_{t+1}^i<0\right)-\alpha\right)\cdot\left|{}_jFE_{t+1}^i\right|^{p-1}\right]=0$, and S is the optimal weighting matrix (see Section 2.3). In STATA we use **ivreg2** with the **gmm** option which utilizes the two-step efficient GMM, and the optimal weighting matrix is the inverse of the covariance matrix of orthogonality conditions. Note that there should be at least two instruments to be able to estimate this statistic.

Table 5 Validating α

Panel A. Employment as a determinant of α

We regress 2,489 individual analysts' α's, from Panel A of Table 4, on 210 investment firm indicator variables. An investment firm indicator variable is equal to 1 when the firm is the analyst employer and 0 otherwise. In the case of an analyst who has worked for more than one investment firm, we include only the first employer.[23] We report adjusted R squared and F statistic with 209 and 2,279 degrees of freedom from the test that the investment firm dummies are jointly equal to zero. We use heteroscedasticity-consistent standard errors. Instruments used in the estimation are reported in the first column. *** indicate significance at 1%.

Loss function and instruments used in deriving α	Adj. R squared	F stat
Lin-lin loss		
Interc., ${}_jFE_t^i$	0.1120	2.471***
Interc., ${}_jA_t$, ${}_jA_{t-1}$	0.1182	2.562***
All	0.1219	2.617***
Quad-quad loss		
Interc., ${}_jFE_t^i$	0.0794	2.005***
Interc., ${}_jA_t$, ${}_jA_{t-1}$	0.0893	2.143***
All	0.0748	1.942***

[23] There are 1,388 analysts who have worked for more than one investment firm in the sample. Including the employer that employed such an analyst for the longest period of time resulted in higher adjusted R-squares and F statistics.

Panel B. Discrepancy between analyst's and employer's α as a predictor of analyst turnover

Model 1 is a logit regression that estimates the probability of a turnover as a function of the discrepancy between the analyst's α and current employer's α:

$$P\left[TURN_i\right] = \beta_o + \beta_1 DISC_i + \varepsilon_i ,$$

where $TURN_i$ is an indicator variable equal to 1 if analyst *i* changes jobs, and zero otherwise; $DISC_i$ is the absolute value of the difference between the analyst's α and the current employer's α. When an analyst experiences turnover more than once, we include only the first turnover observation. There are 1,388 analysts experiencing job turnover, or separation, from 210 employers. Employer's α is estimated from forecasts issued by all analysts employed by that firm. Because there are 22 employers that do not have enough observations for the estimation of employer's α, our sample includes 2,457 analysts (1,356 turnover observations) working for 189 employers.

Model 2 is a logit regression that estimates the probability that analyst's new employer will have high alpha (α>0.5) as a function of analyst's alpha:

$$P\left[HI\_NEW\_EMPL_i\right] = \beta_i + \beta_1 HI\_ANLST_i + \varepsilon_i ,$$

where HI_ANLST_i is equal to 1 if analyst's alpha is high (α>0.5), and zero otherwise, and $HI\_NEW\_EMPL_i$ is equal to 1 if new employer's alpha is high (α>0.5) and zero otherwise. The number of new employers is 201. We were not able to estimate an α for 27 employers. Thus, model 2 is estimated on a sample of 1,304 turnover observations. We estimate the model separately on two sub-samples: analysts currently employed at high-alpha firms (105 observations under linear loss and 1,090 number of observations under quadratic loss) and analysts currently employed at low-alpha firms (1,199 observations under linear loss and 214 number of observations under quadratic loss). We report parameter estimates, heteroscedasticity-consistent standard errors in parentheses. *** (**) [*] indicate significance at 1% (5%) [10%].

Loss function and instruments used to estimate α	Model 1	Model 2	
		$\alpha_{Old\ employer}<0.5$	$\alpha_{Old\ employer}>0.5$
Lin-lin, All instruments			
β_0	0.0217	-4.6303***	-3.157#
	(0.0679)	(0.3031)	(0.7295)
β_1	1.4422***	1.6859**	-
	(0.5218)	(0.6656)	-
Quad-quad, All instruments			
β_0	0.0848	0.8076***	-1.4803***
	(0.0688)	(0.2099)	(0.1087)
β_1	0.4341	0.8239*	0.4584***
	(0.2812)	(0.4216)	(0.1467)