An Empirical Investigation of Analysts’ Objective Function

Youfei Xiao*

November 17, 2015

Abstract

Assumptions about sell-side analysts’ objective function are critical to empirical researchers’ understanding of their incentives and the resulting behavior. This paper provides empirical evidence about the objective function underlying analysts’ choice of forecasts. In contrast to approaches used in previous papers which rely exclusively on statistical properties of forecasts, I compare theoretical models with alternate objective functions based on their ability to explain observed forecasts. A linear loss objective function which incorporates the effect future analysts’ actions on analysts’ deviation from peer forecasts is best rationalized by the data. I find that assumptions about the objective function has a substantial impact on the conclusions from empirical tests about analysts’ incentives and resulting behavior.

*Doctoral Candidate in Accounting, Graduate School of Business, Stanford University. E-mail: youfei.xiao@stanford.edu. I am deeply indebted to Dave Larcker and Peter Reiss for their invaluable advice. I also thank Mary Barth, Anne Beyer, Ed deHaan, Ron Dye, Mike Klausner and Stefan Reichelstein for helpful comments and discussions. Financial support from the Chartered Accountants Education Foundation and the Social Sciences and Humanities Council of Canada is gratefully acknowledged. All errors are my own.
1. Introduction

This paper examines empirically which objective function analysts use in their forecasting decisions. I consider four models of analysts’ payoffs which capture similar incentives but differ slightly in functional form as well as the measurement of the incentives and analyze the differences in the resulting forecast strategies. I then compare the four models based on the extent to which the corresponding forecast strategies explain observed forecast data. My analysis shows that these four models differ substantially in their explanatory power and give rise to distinct inferences about the nature of analysts’ incentives.

The empirical properties of analyst forecasts is studied extensively in accounting and finance, with the evidence pointing to conflicting conclusions about analysts’ rationality and incentives. Often absent from this literature is a discussion of the assumptions about analysts’ objective function that underly the tests of rationality and the resulting inferences about incentives. In the few studies where assumptions about analysts’ objective function are discussed, researchers commonly assume that this objective function depends on squared forecast error. Other studies, motivated by the consideration that analysts are concerned about how their forecasts compare to forecasts from other analysts, assume that their objective function depend on both squared forecast error and squared deviation from average peer forecast. Prior research has noted that if an objective function based on quadratic terms is a plausible characterization of analysts’ incentives, then perhaps one based on absolute values could be equally plausible.

An additional assumption common to previous theoretical and empirical analysis about the role of peer forecast in analysts’ objective function is that peer forecast is necessarily measured as the consensus forecast. That is, prior research assumes that analysts are concerned about the average over forecasts from analysts who have already announced and deviation from analysts who have yet to announce does not enter into analysts’ objective function. Although such an assumption provides analytical tractability in solving the model, it is quite plausible that in the underlying
objective function which analysts use to determine their forecasts, peer forecast is measured over already-issued forecasts and those that have yet been announced. Thus, it remains an open question whether equilibrium forecasts arising from models with alternate assumptions about the functional form of analysts’ objective function and the measurement of peer forecasts are a better fit for the data than models considered in previous research.

This study considers four models which assume that analysts’ predictions about earnings are informed by public information observed by all analysts, private information they obtain through their research activities and inferences about previous analysts’ private information based on observing their forecasts. In all four models, analysts are concerned about their forecast accuracy and their deviation from peer forecasts. In each model, the analyst’s objective function is based on either (1) squared forecast error and squared deviation from a static measure of peer forecast which only accounts for analysts who have already announced, (2) squared forecast error and the squared deviation from a forward-looking measure of peer forecasts which incorporates forecasts from both past and future analysts, (3) the absolute value of forecast error and the absolute value of deviation from the static measure of peer forecast and (4) the absolute value of forecast error and the absolute value of deviation from the forward-looking measure of peer forecast. I assume that all analysts have identical preferences for deviating from peer forecast.

I show that although all four models predict bias in the same direction, relative to consensus, as that hypothesized in previous research, an objective function comprised of squared terms induces behavior that is economically distinct from behavior induced by one comprised of absolute value terms. Under the former objective function, forecast strategies are linear in their earnings prediction and consensus, where concerns about deviating from peer forecast impacts the sign and magnitude of the weight placed on each item. In contrast, when analysts are concerned with the absolute value of deviation from peer forecast, they behave similarly to players in a location game, albeit one with incomplete information. In such a setting, the analyst’s decision is driven by the distance between his forecast and his earnings prediction as well as the sign of the
difference between the prediction and consensus. The resulting forecast strategies are non-linear, with discontinuities or with regions over which forecasts are completely insensitive to the earnings prediction.

Additionally, the model assumption about whether the peer forecast is a static or forward-looking measure affects either the weighting in the forecast strategies (under an objective function based on squared terms) or the curvature in the shape of forecast strategies (in the case of an objective function based on absolute value terms). In particular, when peer forecast is measured over both past and future forecasts, analysts rationally anticipate future analysts’ response function to their choice of location (relative to the earnings prediction and consensus). Both this anticipation and the best response function itself vary depending on the analyst’s order. The forward-looking measure of peer forecasts also creates a mechanical effect in the sense that consensus forecasts matter less to earlier analysts relative to later analysts.

I assess the empirical performance of each model based on fitting the associated forecast strategy to the data. For the two models which assume an objective function based on squared terms, I estimate a linear model and find that the estimated weights are not consistent with the model-implied weights under the assumption that peer forecast is forward-looking. A formal comparison using the log-likelihood from estimating both models also indicates that a model based on static peer forecast is more plausible. For the two models which assume an objective function based on absolute value terms, I estimate non-linear models, where the precise form of non-linearity is restricted by the model-implied forecast strategy. The log-likelihood from fitting the strategy associated with a model which assumes a forward-looking measure of peer forecast is superior to that from fitting the strategy associated with a model which assumes a static measure of peer forecast. Overall, a model which assumes analysts’ objective functions are based on the absolute forecast error and the absolute deviation from the forward-looking measure of peer forecast performs the best at explaining the data.

Although my empirical approach makes no formal attempt to recover the structural parameters, the coefficients from my estimation of forecast strategies associated with
an objective function based on squared terms are loosely consistent with an objective function which penalizes deviations from peer forecasts while those from my estimation of forecast strategies associated with an objective function based on absolute value terms are loosely consistent with an objective function which rewards analysts for the exact same action. My findings suggest that inferences about analyst behavior, in this case their preference for deviating from peer forecast as well as the extent to which they are concerned about future forecasts, are sensitive to a seemingly innocuous functional form assumption about the underlying theoretical model. This sensitivity highlights the importance of formulating an explicit model in developing empirical tests of analysts’ behavior.

The remainder of the paper is organized as follows. Section 2 summarizes the related literature. Section 3 describes four possible objective functions as well as other model assumptions. In Section 4, I discuss the equilibrium for each of the models. Sections 5 and 6 develops the empirical estimation and presents the corresponding results. Section 7 concludes.

2. Related Research

There are a number of previous studies highlighting the importance of assumptions about analysts’ objective functions for the empirical tests of forecast rationality. Basu and Markov (2004) finds that using the Least Absolute Deviation (LAD) estimator in place of the Ordinary Least Squares (OLS) estimator in regressions of actual earnings on forecasts produces coefficients consistent with analyst rationality. The LAD and OLS estimators are based on distinct econometric criterion functions, which differ from the assumptions I am making about analysts’ objective functions. Since the former is derived from the researcher minimizing the sum of the absolute value of residuals and the latter is derived from the researcher minimizing the sum of squared residuals, the paper conclude that analysts must also be using an absolute linear loss function rather than a quadratic loss function. In similar work, Gu and Wu (2003) shows that if analysts’ payoffs depend exclusively on the absolute value of their forecast error, then the optimal forecast is the median of their posterior beliefs. In contrast, they note that
analysts with payoffs that depend on the squared loss of their forecast error should forecast the mean of their beliefs. So the difference between the optimal forecasts under these two objective functions should be increasing in the skewness of analysts’ beliefs about earnings. They document that in the cross-section, forecast errors are positively associated with the skewness of a firm’s time-series of earnings. These results lead Gu and Wu (2003) to conclude that analysts must be using an absolute linear loss function. Related work in Abarbanell and Lehavy (2003) acknowledges that tests of analyst forecast rationality are sensitive to assumptions about the objective function.

My study is distinct from these previous papers along two dimensions. First, the previous comparisons of the absolute value and quadratic objective functions do not consider the possibility that analysts have strategic incentives to be close or far from peer forecasts in addition to incentives to be accurate. In contrast, I model both of these incentives explicitly. Second, while the previously used approaches to distinguishing between the two objective functions are based exclusively on the statistical properties of forecasts and forecast errors, I use the analytical solutions from the economic model to motivate statistical tests to apply in the data.

More broadly, my work is related to prior evidence on the extent to which peer forecasts matter to analysts’ forecasting decisions. Like my study, these papers are usually motivated by theoretical models of herding (Banerjee, 1992; Scharfstein and Stein, 1990) as well as evidence that differentiation from peers plays a role in analysts’ career outcomes (Hong and Kubik, 2003; Groysberg et al., 2011). Bernhardt et al. (2006) classifies earnings forecasts based on their location relative to actual and consensus. The paper computes frequency statistics resulting from these classifications and concludes that analysts prefer to be far from consensus. There is no explicit model of analysts’ objectives in the paper. Chen and Jiang (2006) and Zitzewitz (2001) develop models in which analysts are concerned about squared forecast error and squared deviation from peer forecast, where peer forecast is measured as consensus.

\footnote{For example, Gu and Wu (2003) acknowledges in footnote 1 that deliberate departures from beliefs about earnings are absent from the model.}
The process through which later analysts infer earlier analysts’ private information in a potential sequential setting is not explicitly modeled. Chen and Jiang (2006) concludes from its empirical test that analysts overweight their private information while Zitzewitz (2001) concludes that there is exaggeration behavior within the specific model considered in the paper.

3. Model

This section describes a stylized model of a sequential forecasting setting and the alternate assumptions about analysts’ objective function. It also outlines the statistical properties of analysts’ beliefs.

3.1 Assumptions

The model considers $J + 1$ analysts who issue forecasts on earnings sequentially according to an exogenously determined order. I assume that there is a common prior belief about actual earnings $E$, where $E$ is normally distributed with mean $\mu_0$ and variance $\sigma_0^2$. Each forecaster $j$ observes the private signal $s_j = E + \xi_j$. Conditional on actual earnings $E$, signals $s_j$ are independently normally distributed with mean $E$ and variance $\tau_j^2$. The analyst forecasting in the $j$-th order issues his forecast $f_j$, after observing his own signal $s_j$, the $j - 1$ preceding forecasts $f_1, \ldots, f_{j-1}$ and the inferred signal from preceding analysts’ forecast $\hat{s}_1(f_1), \ldots, \hat{s}_{j-1}(f_{j-1})$ and forms beliefs about earnings. Clearly in a sequential forecasting setting, analyst $j$ does not observe $s_{j+1}, \ldots, s_{J+1}$ but has some conjectures about subsequent analysts’ information.

Analysts issue forecasts by optimizing over one of two potential objective functions. An analyst $j$ is said to have a squared loss if his payoffs are determined by the following function:

$$L_j = (f_j - E)^2 + \lambda (f_j - \tilde{f}_{-j})^2.$$

Analyst $j$ is said to have absolute value loss if his payoffs are determined by the following function:

$$L_j = |f_j - E| + \lambda |f_j - \tilde{f}_{-j}|.$$
In both loss functions, the second term captures the impact of peer forecast on analysts’ payoffs. The inclusion of this term is commonly motivated by the thought that in a performance measurement framework, peer forecasts could be a performance measure that, used together with forecast error, exposes the analyst to less systematic risk (Hong et al., 2000; Hong and Kubik, 2003). There are two possibilities for how \( f_{-j} \) is defined. One assumption is that analysts are concerned with a measure of peer forecast in a static sense, incorporating only forecasts that have already been announced. Formally, I define the static measure of peer forecast as 
\[
\bar{c}_j = \sum_{k=1}^{j-1} f_k,
\]
where \( c_j \) is the variable conventionally referred to as consensus in this literature. An alternative assumption is that analysts’ payoff depend on the average over all forecasts, including those that have already been issued and those who will forecast later in the sequence. Accordingly, I define the forward-looking measure of peer forecast as
\[
\bar{f}_{-j} = \left( \sum_{k=1}^{j-1} f_k + \sum_{k=j+1}^{J+1} f_k \right) J^{-1}.
\]

The assumption that peer forecast is a static measure is used frequently in existing literature (Zitzewitz, 2001; Chen and Jiang, 2006) and gives rise to analytically simple model solutions. However, a forward-looking measure of peer forecast could be equally descriptive. If deviation from peer forecast enters into analysts’ objective functions as a performance measure, it is likely that all peer forecasts, irrespective of their order relative to the analyst, would be impounded in the variable. It is important to note that under both assumptions, analysts only observe consensus \( c_j \). The distinction is that, under the assumption of a forward-looking measure of peer forecast, analysts solve their optimization problem based on both \( c_j \) and on their expectations over peers who will forecast in the future. Thus, the salient question is whether this expectation over how later analysts’ forecast changes the analyst’s behavior in such a way that is distinguishable in the data.

To summarize, I consider the empirical implications of four potential models where analysts either follow a quadratic objective function with a static measure of peer forecast, quadratic objective function with a forward looking measure of peer forecast, absolute value objective function with a static measure of peer forecast or an absolute value objective function with a forward looking measure of peer forecast.
A final consideration in the model setup is that whenever the objective function is based on the static measure of peer forecast, \( c_j \) is undefined when \( j = 1 \), as such, the model does not provide predictions for how the first analyst behaves. In order to facilitate the comparison across the four models, I assume that the first analyst is unbiased. In my empirical tests, I restrict my analysis to analysts forecasting second or later. As such, the assumption about the first analyst only matters to the extent that it enters into subsequent analysts’ inferences about the first analyst’s private information.

3.2 Updating of Beliefs

With normal priors and a signal with an additive noise term that is also normal, each analyst’s posterior beliefs after observing his own signal and other forecasts is also normal. I use \( \mu_j \) and \( \sigma_j^2 \) to denote the mean and variance of this posterior distribution. For example, \( \mu_1 = \mathbb{E}(E|s_1)\), \( \mu_2 = \mathbb{E}(E|s_2, \hat{s}_1(f_1)) \) and \( \sigma_2 = Var(E|s_2, \hat{s}_1(f_1)) \). The posterior mean \( \mu_j \) can be computed using the standard formula with matrix partitions and inverses:

\[
\mu_j = \begin{bmatrix} \mu_0 \\ \mu_0 \\ \vdots \\ \mu_0 \end{bmatrix} + \begin{bmatrix} \sigma_0^2 \\ \sigma_0^2 \\ \vdots \\ \sigma_0^2 \end{bmatrix} \begin{bmatrix} \tau_1^2 + \sigma_0^2 & \sigma_0^2 & \cdots & \sigma_0^2 \\ \sigma_0^2 & \tau_2 + \sigma_0^2 & \cdots & \sigma_0^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_0^2 & \sigma_0^2 & \cdots & \tau_j^2 + \sigma_0^2 \end{bmatrix}^{-1} \begin{bmatrix} \hat{s}_1 \\ \hat{s}_2 \\ \vdots \\ s_j \end{bmatrix} - \begin{bmatrix} \mu_0 \\ \mu_0 \\ \vdots \\ \mu_0 \end{bmatrix}
\]

where \( \hat{s}_1, \hat{s}_2, \ldots, \hat{s}_{j-1} \) denote \( j \)‘s inferences about the previous analysts’ beliefs. These inferences, of course, are functions of \( f_1, \ldots, f_{j-1} \) together with the equilibrium forecasting strategies which are analyzed in the next section.

The expression for \( \mu_j \) can be expanded into a term that includes only \( s_j \) and \( \tau_j \) as well as a term made up of \( \hat{s}_1, \ldots, \hat{s}_{j-1} \) and \( \tau_1, \ldots, \tau_{j-1} \). This second term is identical to

\[\text{In the case where peer forecast is a forward-looking measure, this assumption coincides with the equilibrium outcome as well.}\]
how \( j - 1 \) updates his beliefs to form \( \mu_{j-1} \), except \( j - 1 \) uses the signal he actually observes \( s_j \) instead of \( \tilde{s}_{j-1} \). Since analyst \( j \) is rational about \( \tilde{s}_{j-1} \) and both \( j \) and \( j - 1 \) have the same beliefs about the preceding analysts’ \( \tilde{s} \), \( \mu_j \) can also be written as:

\[
\mu_j = W_j^p(\sigma_0, \bar{\tau})\hat{\mu}_{j-1} + (1 - W_j^p(\sigma_0, \bar{\tau}))s_j. 
\]  

(1)

where \( W_j^p(\sigma_0, \bar{\tau}) \equiv \tau_j^2 + \sigma_0^2 - \sigma_0^4 \)

\[
\begin{bmatrix}
1 \\
\vdots \\
1 \\
\end{bmatrix} \begin{bmatrix}
\tau_1^2 + \sigma_0^2 & \sigma_0^2 & \cdots & \sigma_0^2 \\
\sigma_0^2 & \tau_2^2 + \sigma_0^2 & \cdots & \sigma_0^2 \\
\vdots & \vdots & \ddots & \vdots \\
\sigma_0^2 & \sigma_0^2 & \cdots & \tau_{j-1}^2 + \sigma_0^2 \\
\end{bmatrix}^{-1} \begin{bmatrix}
1 \\
\vdots \\
1 \\
\end{bmatrix}^{-1}
\]

and where \( \hat{\mu}_{j-1} \) is \( j \)'s inferences about \( j - 1 \)'s beliefs. Note the \( p \) superscript in \( W_j^p(\cdot) \) is included to denote weights in forming analysts’ prediction about earnings. Similar to the case with \( \tilde{s}_1, \ldots, \tilde{s}_{j-1}, \hat{\mu}_{j-1} \) depends on what \( j - 1 \) actually forecasts through the equilibrium forecasting strategies that follow from each of the four objective functions I consider.

In this general case where every analyst has his own \( \tau_j \), then \( W_j^p(\cdot) \) can take on any value between 0 and 1, irrespective of the \( W(\cdot) \) for any other analyst. Even placing restrictions on the progression of \( \tau_j \) does not necessarily result in restrictions on \( W_j^p(\cdot) \). In the case where all analysts’ signals have the same precision i.e., \( \tau_j = \tau \), the expression for beliefs simplify to \( W_j^p(\sigma_0, \tau) = \frac{(j-1)\sigma_0^2 + \tau}{j\sigma_0^2 + \tau} \). If all analysts have the same precision or if a later analyst is less precise than an earlier analyst, the later analyst always places a lower weight on his private information than the earlier one, with \( W_{j-1}^p(\cdot) > W_j^p(\cdot) \). However, if \( \tau_j \) is decreasing in \( j \) (i.e., every analyst’s signal is more precise than the previous analyst’s signal), the weights in Equation (1) can be either increasing or decreasing with every subsequent analyst.

4. **Equilibrium Strategies**

Given each analyst’s beliefs \( \mu_j \) and observed consensus \( c_j \), the analyst solves his optimization problem, which consists of the quadratic loss objective function or
absolute value objective functions. Each objective function has different implications for how \( f_j^* \) varies with \( \mu_j \) and \( c_j \). These implications are discussed in this section, with the formal derivations included in Appendix A. Within each model being considered, I first discuss general features of the forecast strategies. Then, I examine specifically how the strategies differ depending on whether the model assumes that peer forecast is static or forward-looking.

### 4.1 Quadratic Loss Function

#### A. Properties of Forecast Strategies

When the objective function is quadratic, analysts’ strategies are linear in \( \mu_j \) and observed consensus \( c_j \). For \( j \geq 2 \), the strategy is:

\[
f_j^* = W_j^s(\lambda) c_j + \left(1 - W_j^s(\lambda)\right) \mu_j.
\]  

The weight in the forecasting strategy, \( W_j^s(\cdot) \), varies depending on the value of \( \lambda \). The superscript \( s \) is used to denote strategy (whereas the \( p \) superscript in Equation (1) corresponds to prediction about earnings). The exact functional form of is described in the Appendix.\(^3\)

If analysts’ objective functions includes a penalty for deviating from peer forecasts (i.e., \( \lambda > 0 \)), \( W_j^s(\lambda) \) is positive and the analyst issues a forecast between \( \mu_j \) and \( c_j \). If analysts are rewarded for deviations from consensus, I impose an additional assumption that \( \lambda \) is not so negative where there is no interior solution to the analyst’s optimization problem. For the admissible values of \( \lambda \), \( W_j^s(\lambda) \) is negative and the analyst biases from \( \mu_j \) in the opposite direction as \( c_j \). To summarize, \( W_j^s(\lambda) \) always has the same sign has \( \lambda \) and in the special case where \( \lambda = 0 \) the analysts honestly forecast their beliefs with \( f_j^* = \mu_j \).

Analysts’ inferences about previous analysts’ earnings predictions must be rational given the equilibrium forecast strategies, with \( \hat{\mu}_k\left(f_k^* (\mu_k)\right) = \mu_k \). For \( j = 2 \), \( \hat{\mu}_1 = f_1 \)

---

\(^3\)In particular, under the assumption that peer forecast is a forward-looking measure, Equations (A.4.2) and (A.4.3) characterize the strategy. Under the assumption that peer forecast is a static measure, Equation (A.1.1) characterizes \( W_j^s(\cdot) \).
because I have assumed that the first analyst honestly forecasts $\mu_1$. For $j > 2$, the expression for the analyst’s prediction about earnings in Equation (1) can be completed using the following relation between analyst $j$’s inferences about analyst $j - 1$’s prediction and $f_{j-1}$ as well as $c_{j-1}$.

$$\hat{\mu}_{j-1} = \frac{f_{j-1} - W_{j-1}^s(\lambda)c_{j-1}}{1 - W_{j-1}^s(\lambda)}$$

Note that $c_{j-1}$ itself is mechanically related to $c_j$ and $f_j$, with $c_{j-1} = \frac{c_j(j-1) - f_{j-1}}{j - 2}$. Thus, a substitution can be made so that in Equation (1), $\mu_j$ can be expressed in terms of only $c_j, s_j$ and $f_{j-1}$:

$$\mu_j = W_j^P(\sigma_0, \bar{\tau}) \frac{f_{j-1} - W_{j-1}^s(\lambda)\frac{c_j(j-1) - f_{j-1}}{j - 2}}{1 - W_{j-1}^s(\lambda)} + (1 - W_j^P(\sigma_0, \bar{\tau})) s_j.$$  

Finally, with substitution of the previous expression and the assumption that $s_j = E + \xi_j$, the analyst’s strategy in Equation (2) can be written entirely in terms of consensus $c_j$, the previous forecast $f_{j-1}$, actual earnings $E$ and the noise in analyst’s private signal $\xi_j$.

$$f_j^* = \begin{cases} 
W_j^s(\lambda)c_j + (1 - W_j^s(\lambda))(W_j^P(\sigma_0, \bar{\tau})c_j + (1 - W_j^P(\sigma_0, \bar{\tau}))(E + \xi_j)) & \text{if } j = 2 \\
W_j^s(\lambda)c_j + (1 - W_j^s(\lambda)) \\
\cdot \left( W_j^P(\sigma_0, \bar{\tau})\frac{f_{j-1} - W_{j-1}^s(\lambda)\frac{c_j(j-1) - f_{j-1}}{j - 2}}{1 - W_{j-1}^s(\lambda)} + (1 - W_j^P(\sigma_0, \bar{\tau}))(E + \xi_j) \right) & \text{if } j > 2 
\end{cases}$$

Note that Equation (3) forms the basis for the econometric model developed in the next section.

**B. Static vs. Forward-Looking Measure of Peer Forecast**

If the objective function is based on a forward-looking measure of peer forecast, analysts’ incentives differ depending on whether they are early or late in the timing of their forecasts, where the terminology is used somewhat unconventionally in this setting. Specifically, I consider an analyst to be “early” (“late”) if there are many (few)
forecasts that will follow his forecast. That is, fixing the total number of analysts $J, j$ reflects whether an analyst is early or late. However, if the comparison in question is based on firms with different $J$’s, then it is $J – j$ (i.e., the number of analysts yet to forecast) which captures the notion of early or late\(^4\).

Whereas later analysts know with almost certainty what consensus will be after all analysts have issued forecasts, earlier analysts face uncertainty over future analysts’ beliefs about earnings. I show in the appendix that this distribution has the same mean as his beliefs over earnings. Earlier analysts have a first-movers advantage in the sense that when he adds bias to his forecast, a later analyst’s best response is to introduce bias in the same direction (relative to the later analyst’s beliefs)\(^5\).

As such, the early mover is in a position to bias less than the late mover (therefore incurring a lesser forecast accuracy loss) to effect the same expected deviation from average peer forecasts. There is also mechanical difference in analysts’ incentives due to how the average peer forecast is defined in the objective function. For all analysts, the peer forecast component of their objective functions is averaged over $J$. The consensus forecast that the second analyst faces only impacts average peer forecast by $\frac{1}{J}$ while the consensus that the second-to-last analyst faces impacts his average peer forecast by $\frac{J - 1}{J}$. This difference in how prevailing consensus enters into the objective function increases the weight placed on consensus for later analysts. This latter effect dominates the former such that overall, the magnitude of $W_j(\lambda)$ increases with every subsequent analyst.

If peer forecast is a static measure, then the weight used by the last analyst $W_{J+1}(\lambda)$ corresponds to that predicted for the objective function under forward looking consensus. Since every analyst faces the same objective function for a given pair of $\mu_j$ and $c_j$, consensus has the same impact on each analysts’ incentives irrespective of the number of previous analysts that $c_j$ is averaged over. Additionally, earlier analysts do not consider how their actions enter into later analysts’ actions which makes every analyst behave effectively like they are the last analysts. It follows that fixing the

\(^4\)For example, an analyst who forecasts third out of eight analysts is considered early and one who forecasts second out of three analysts is considered late.

\(^5\)Critical to this result is the assumption that $\lambda$ is the same across all analysts.
total number of analyst at $J$, all analysts place the same weight on consensus and on the earnings prediction, with $W^s_j(\lambda) = W^s_{j+1}(\lambda), \forall j$.

### 4.2 Absolute Value Objective Function

If analysts’ have an absolute value objective function, the resulting strategies are substantially different from those that result from a quadratic objective function. In particular, with the latter objective function, $\lambda$ changes how analysts weight their prediction relative to consensus while with the former, $\lambda$ affects the distance of the analyst’s optimal forecast relative to his earnings prediction.

#### A. Properties of Forecast Strategies

Under an absolute value objective function, the forecast strategies cannot necessarily be solved in closed form. The shape of his deviation from his earnings prediction is highly non-linear in the difference between the analyst’s prediction and consensus, which I denote as $D_j \equiv \mu_j - c_j$. In addition, the sign of $\lambda$ affects the shape of this non-linear strategy. I first assume that $\lambda > 0$. When $D_j$ is small and positive, analysts will forecast exactly $c_j$. When $D_j$ is positive and moderately large, $f^*_j$ increases gradually. As $D_j$ grows large, the shape of $f^*_j$ becomes almost linear, where the analyst’s departure from his earnings prediction does not exceed a constant function of $\lambda$ and the posterior variance of his earnings prediction. Formally, the following list includes characteristics about the shape of $f^*_j$ over the range $D_j > 0$ case (where the $D_j < 0$ analysis is exactly symmetrical). The details of the derivations are included in Appendix A.

1. $f^*_j > \mu_j$.
2. $f^*_j$ is bounded below by $\mu_j - \hat{\sigma}_j \Phi^{-1} \left( \frac{1+\lambda}{2} \right)$, where $\Phi(\cdot)$ is the standard normal cumulative distribution function.
3. $\frac{df^*_j}{dD_j} \in [0, 1]$.
4. $\frac{d^2f^*_j}{dD^2_j} \geq 0$. 

13
v) There is a threshold \( D' \) such that when \( D_j < D' \), \( f_j^* = c_j \).

vi) In the special case of the last analyst \((j = J + 1)\), \( D' = \hat{\sigma}_{J+1} \Phi\left(\frac{1+\lambda}{2}\right) \) and \( f_{J+1}^* = \mu_{J+1} - \hat{\sigma}_{J+1} \Phi^{-1}\left(\frac{1+\lambda}{2}\right) \).

When \( \lambda < 0 \) and \( D_j \) is positive, the slope of \( f_j^* \) is very large for values of \( D_j \) close to zero. As \( D_j \) increases, this slope decreases and at the extreme values of \( D_j \), \( f_j^* \) will converge approximately a line with a slope of 1. For all positive values of \( D_j \), the difference between \( f_j^* \) and \( \mu_j \) is bounded above by a constant function of \( \lambda \) and \( \hat{\sigma}_j \). Formally, some characteristics of \( f_j^* \) can be summarized in the list below (with the derivations included in Appendix A.2).

i) \( f_j^* < \mu_j \).

ii) \( f_j^* \) is bounded above by \( \mu_j + \hat{\sigma}_j \Phi^{-1}\left(\frac{1-\lambda}{2}\right) \).

iii) \( \frac{df_j^*}{dD_j} \in (1, \infty) \).

iv) \( \frac{d^2f_j^*}{dD_j^2} < 0 \).

v) In the case of the last analyst \((j = J + 1)\), \( f_{J+1}^* = \mu_{J+1} + \hat{\sigma}_{J+1} \Phi^{-1}\left(\frac{1-\lambda}{2}\right) \).

Similar to the insight from the discussion of a quadratic objective function in Section 4.1, each analyst's inference about the preceding analysts' prediction about earnings must be consistent with the equilibrium described above. Since the equilibrium strategy is not necessarily solvable in closed form, the function associated with this inference may also lack analytical expressions. With the absolute value objective function, \( f_j^* \) is non-monotonic in \( \mu_j \) if \( \lambda > 0 \) and \( \mu_j < c_j + D' \). In such a case, analyst \( j + 1 \) does not infer \( \mu_j \) exactly but rather his beliefs about \( \mu_j \) follows a normal distribution truncated above by \( c_j + D' \) and below by \( c_j - D' \). His inferences about analyst \( j \)'s signal is therefore the mean of this distribution, with \( \hat{\mu}_{j-1} = c_j \).

B. Forward-looking vs. Static Peer Measure Assumption

In contrast to the quadratic loss setting where models with different assumptions about how peer forecast is measured have distinct implications for analysts' weighting
of their earnings prediction and consensus, in an absolute value loss setting the two assumptions give rise to different predictions about the shape of the forecasting strategy (see Appendix [A.2] for the formal derivations). For illustrating these differences, I maintain the assumption from the previous subsection that $D_j > 0$.

Under the forward-looking measure of peer forecast assumption, the shape of $f_j^*$ varies depending on $j$. When $\lambda < 0$, the shape of $f_j^*$ is steeper (i.e., $\frac{d^2 f_j^*}{dD_j^2}$ is more negative) for later analysts than earlier analysts in the region close to $D_j = 0$. For the very last analyst, $f_j^*$ jumps from consensus when $D_{J+1} = 0$ to $\mu_{J+1} + \hat{\sigma}_j \Phi^{-1}\left(\frac{1-\lambda}{2}\right)$. The left plot in Figure I further illustrates the shape of $f_j^*$. As analysts' order increases, the region of the plot close to $c_j$ becomes steeper and converges to the shape in the right plot. When $\lambda > 0$, $D'$, which is the range of $D_j$ over which the analyst will imitate consensus, is larger for later analysts than for earlier analysts. In the region where $D_j > D'$, $f_j^*$ will be steeper for later analysts than for earlier analysts, but the slope never exceeds 1. The late analysts versus early analyst comparison can also be seen in Figure I. In both cases, for a fixed large value of $D_j$, the magnitude of the bias is increasing in the order of the analysts.

Like the case with the quadratic objective function, when peer forecast is a forward-looking measure, earlier analysts’ strategies reflect the fact that later analysts’ bias will respond endogenously to their own choice of bias. Unique to this setting is that both forecast strategies and corresponding response functions reflect a cross-partial effect between $\mu_j$ and $c_j$. That is, later analysts’ response to earlier analysts’ bias is the most pronounced when $\mu_j$ and $c_j$ are close together. For this reason, earlier analysts bias less than later ones for small values of $D_j$ while later analysts are less biased than earlier analysts when $D_j$ is large.

When peer forecast is a static measure, the shape of each analyst’s strategy is the same no matter his order. That is, the strategy always has the shape in the right plot of Figure I. The only difference between earlier and later analysts is the magnitude of the bias. Specifically, when $\lambda > 0$, the analyst’s forecast is the greater of $\mu_j - \hat{\sigma}_j \Phi^{-1}\left(\frac{1+\lambda}{2}\right)$
and $c_j$. For the $\lambda < 0$ case, the analyst forecasts $\mu_j + \hat{\sigma}_j \Phi^{-1} \left( \frac{1-\lambda}{2} \right)$ whenever $D_j > 0$.

### 5. Empirical Estimation

This section develops the empirical approach for fitting the forecast strategies derived in Section 4 to the data. The forecast strategies are functions of consensus, earnings, the previous analyst’s forecast and an unobservable term $\xi_j$. The estimation requires a number of additional assumptions. I assume that $\sigma_0$, the variance of analysts’ prior beliefs about earnings, is constant in the cross-section of firms. A second assumption is that $\tau_j$ is homogeneous in the cross-section. That is, the variance of analysts’ private signal is permitted to vary depending on the order but is homogeneous for analysts who forecast the same order. Also, consistent with the previous section, I assume that all analysts have the same $\lambda$. The reason such assumptions are necessary is that distinguishing between the static and forward-looking measure of peer forecast requires comparison of either the weighting or shape of analysts’ forecasts across analysts who forecast in a different order (both relative to the start and to the total number of analysts). Empirical identification of this type of heterogeneity requires the researcher to assume that there is no heterogeneity along other dimensions (i.e., across observations with the same $j$).

My sample is comprised of forecasts issued within 30 days of annual earnings announcements between 1990 and 2012 in the I/B/E/S Unadjusted Details File. In order to ensure that the data comes from a setting in which analysts have sufficient time to observe preceding analysts’ forecasts, I use two-year ahead forecasts where the average forecast occurs three days apart. Since my model is intended to explain analysts’ concern about peer forecasts, I exclude the first analyst from the sample (including firm-years covered by only a single analyst). Because the size of the parameter space, along with the associated computational burden, increases with $J$ in my model, firm-years covered by more than 10 analysts (approximately 10% of the sample) are excluded from my analysis. Following evidence from Cheong and Thomas (2011) that analysts’ think in terms of raw earnings per share, I use EPS forecasts without any scaling by share price and or assets.
5.1 Estimation of Strategies under Quadratic Objective Function

As noted in Section 4.1, strategies resulting from the quadratic objective function are linear in consensus $c_j$, the previous analyst’s forecast $f_{j-1}$ and actual earnings $E$ as well as an unobservable term. Thus, if observed forecasts are determined by the quadratic loss function, then they can be related to each of $c_j$, $f_{j-1}$ and $E$ with a linear econometric model.

$$f_{it,j} = \omega_c c_{it,j} + \omega_E E_{it} + \omega_l f_{it,j-1} + \epsilon_{it,j}$$

s.t. $\omega_c + \omega_E + \omega_l = 1$ (4)

In Equation (4), $\omega_c$, $\omega_l$ and $\omega_E$ are reduced form coefficients in the sense that they correspond to the coefficients on $c_j$, $f_{j-1}$ and $E$ in (3), which are determined by the model parameters $\sigma$, $\bar{\tau}$ and $\lambda$. Note that the restriction on the model weights follow from the observation that in Equation (3), the weighting on $f_{j-1}$, $c_j$ and $E$ mechanically sum to 1.

Since $\epsilon_{it,j}$ is homoscedastic by assumption and enters linearly into Equation (4), I estimate $\omega_c$ and $\omega_E$ using the standard Ordinary Least Squares (OLS) approach. The choice is motivated by the well-known property that OLS estimates have the lowest variance amongst all unbiased estimators. Because $\xi_j \sim N(0, \tau_j)$ by assumption and because $\epsilon_{it,j}$ is equal to a constant times $\xi_j$, the OLS estimate coincides with the maximum likelihood estimate. This equivalence allows me to compare the goodness of fit from estimating other models where OLS cannot be implemented. To emphasize, imposing the economic assumption that analysts’ follow a least squares objective function has no direct bearing on my econometric criterion function. That I select an estimator which minimizes the mean squared error between the data and the model prediction, is solely due to the special statistical properties of OLS estimator.

Under the static peer measure assumption, the analyst’s order does not matter in how analysts weight their earnings prediction and consensus, with $W_j^s(\lambda) = W^s(\lambda)$.

---

7Although I assume $\epsilon_{it,j}$ is normally distributed, the assumption is not needed for this property to hold.
However, the order does matter for the weights placed on consensus, the previous forecast and the private signal in determining the earnings prediction (i.e., $W^p_j(\sigma_0, \tau)$). As such, the overall weights in Equation (3) vary depending on order and, for the static peer measure assumption, I estimate (4) with a specific $\omega_c$ and $\omega_E$ parameters for each $j$. That is, I allow the two parameters to vary depending on whether the observation corresponds to the 2nd, 3rd, . . . , 10th forecast.

In contrast, the forward-looking peer measure assumption implies that even fixing the analyst’s order, $W^s_j(\lambda)$ varies with the number of analysts remaining. Specifically, $W^s_j(\lambda)$ is greater for later analysts. Thus, an examination of the weights in (3) reveal that for a fixed $j$, the estimated value of $\omega_c$ ($\omega_E$) should be decreasing (increasing) in $J$.

To summarize the above discussion, the equilibrium forecast strategies arising from the quadratic loss function can be estimated in the data using Equation (4). Holding fixed the assumption that the loss function is quadratic, the static-peer measure assumption implies that (4) should be estimated with $\omega_c$ and $\omega_E$ which varies depending on the order $j$ but with the restriction that the coefficients should be the same irrespective of the total number of analysts $J$. Under the forward-looking peer measure assumption, estimation of (4) should allow for variation in $\omega_c$ and $\omega_E$ depending on $j$ and also variation in $J$, but only to the extent that this second variation is consistent with the inequality relations predicted by the model.

To implement the estimation, I follow the standard procedure for OLS estimation with linear restrictions. I first estimate an unrestricted version of (4) where $\omega_c$ and $\omega_E$ are permitted to vary based on both $j$ and $J$. I compute the log-likelihood of the model using the natural logarithm of the mean squared residuals where, under the equality restrictions predicted by the assumption that peer forecast is a static measure, the residuals for all observations are determined using the estimated weights for the last analyst (but specific to the order $j$). For example, if I find that $\omega_E$ for $j = 4$ and $J = 4$ is 0.2, then for the purposes of calculating the log-likelihood I assume that

\[ \omega_c + \omega_E + \omega_l = 1 \]  

Note that no extra estimation steps are needed to ensure that $\omega_c + \omega_E + \omega_l = 1$ and $\omega_c + \omega_E = 1$. The restrictions can be implemented by subtracting $f_{j-1}$ from the other regression variables (and $c_j$ from $E$ and $f_j$ for $j = 2$).
\( \omega_E = 0.2 \) for all observations where \( j = 4 \), even though the estimates of \( \omega_E \) specific to \( j = 4 \) and \( J = 8 \) may be different from 0.2. To compute the log-likelihood with the inequality restrictions predicted by the model where I assume a forward-looking measure of peer forecast, I use residuals based on weights where \( \omega_c (\omega_E) \) is set to the greater (lesser) of the actual estimated weights specific to the \( j \) and \( J \) for each observation and the weights used for the next highest \( J \). For example, if find that \( \omega_c = 0.2 \) for \( j = 4 \) and \( J = 4 \) and \( \omega_c = 0.3 \) for \( j = 4 \) and \( J = 5 \), I replace the estimated weight with 0.2 for observations where \( j = 4 \) and \( J = 5 \).

### 5.2 Estimation of Strategies under Absolute Value Objective Function

Recall from Section [5.2](#) that forecast strategies cannot be characterized in closed form under the an absolute value objective function. Fitting the forecast strategy exactly for just a single set of parameters requires the use of a numerical solver to predict either \( f_j^* \) or \( \xi_j \) from the inverse of \( f_j^* \) for every observation. Of course, this routine would have to nested in another numerical solver procedure that searches for parameters which best explain the data. Because such an approach would be impractical to implement, I use an approximation of \( f_j^* \) which reduces the computational burden but captures the essential features in the shape of \( f_j^* \). Since this shape varies depending on the sign of \( \lambda \), I use separate approximations for the \( \lambda > 0 \) and \( \lambda < 0 \) case.

For the \( \lambda < 0 \) case, I use a piece-wise error function approximation:\(^{[10]}\)

\[
 f_j = \begin{cases} 
 \mu_j - \text{erf}(\alpha_0 \alpha_1) & \text{if } \mu_j < c_j - \alpha_0 \\
 \text{erf}(\alpha_1 \cdot (\mu_j - c_j)) + c_j & \text{if } c_j - \alpha_0 \leq \mu_j \leq c_j + \alpha_0 \\
 \mu_j + \text{erf}(\alpha_0 \alpha_1) & \text{if } \mu_j > c_j + \alpha_0. 
\end{cases}
\]

\(^{[5]}\)

The left plot in Figure II overlays this approximation function (solid curve) against the exact forecast strategy obtained from a numerical solver (dashed curve), where I

---

\(^{[9]}\)To emphasize, this approach is not the same as computing the mean squared residual from estimating a model where \( \omega_E \) and \( \omega_c \) varies by \( j \) but does not vary by \( J \).

\(^{[10]}\)The error function is the same as the normal cumulative density function up to a constant transformation (i.e., the two functions have the same shape), with \( \text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} \, dt. \)
only include the strategy over the region where \( \mu_j \geq c_j \) because the strategy over the \( \mu_j < c_j \) region is symmetrical. In the approximation, the parameter \( \alpha_1 \) determines the sensitivity of \( f_j^* \) (i.e., \( \frac{\partial^2 f_j^*}{\partial D_j} \)) in the region where \( \mu_j \) and \( \mu_j \) are close together while \( \alpha_0 \) determines the threshold, relative to \( c_j \), beyond which \( f_j^* \) becomes approximately linear in \( \mu_j \).

For the \( \lambda > 0 \) case, I use a piece-quadratic error function approximation:

\[
f_j = \begin{cases} 
-\alpha_2(\mu_j - c_j + \alpha_0)^2 + \alpha_1(\mu_j - c_j + \alpha_0) + c_j & \text{if } \mu_j < c_j - \alpha_0 \\
c_j & \text{if } c_j - \alpha_0 \leq \mu_j \leq c_j + \alpha_0 \\
\alpha_2(\mu_j - c_j - \alpha_0)^2 + \alpha_1(\mu_j - c_j - \alpha_0) + c_j & \text{if } \mu_j > c_j + \alpha_0 \end{cases}
\]  

(6)

In this approximation, \( \alpha_0 \) determines the range of \( \mu_j \)'s relative to \( c_j \) over which the analyst imitates \( c_j \). This parameter is not an approximation and is exactly equal to the \( D' \) variable discussed in Section 4.2. The parameters \( \alpha_1 \) and \( \alpha_2 \) approximate the convex shape of \( f_j^* \) in the region where \( \mu_j \) is outside this interval. The right plot in Figure II illustrates both the approximation function and the exact forecast strategy computed using a numerical solver.

As discussed in Section 4.2 and illustrated in Figure I, the model with a forward-looking measure of peer forecast predicts that when \( \lambda > 0 \), the interval over which \( f_j^* \) is the same as \( c_j \) should be smaller and the convexity in \( f_j^* \) should be larger for earlier analysts. Thus, conditional on the order \( j \), \( \alpha_0 \) should decrease and \( \alpha_1 \) should increase with \( J \). When \( \lambda < 0 \), the model predicts that \( f_j^* \) is steeper in the region where \( \mu_j \) is close to \( c_j \) for later analysts than for earlier analysts, with an infinite slope in the case of the last analyst. Accordingly, \( \alpha_1 \) has to be decreasing in \( J \) for a fixed \( j \). The model also predicts that for \( \lambda < 0 \) and large values of \( \mu_j \), the upper bound in the magnitude of the bias is larger for earlier analysts than earlier analysts. As such, \( \alpha_0 \) has to be increasing in \( J \) conditional on \( j \).

Under the assumption that peer forecast is a static measure, \( f_j^* \) is piece-wise linear but, unlike the case of a quadratic objective function, the location of the linear
strategy varies depending on whether the analyst is early or late. In the approximation functions, $\alpha_1$ determines the curvature in $f_j^*$ when $\lambda < 0$ and when $\lambda > 0$ the curvature in $f_j^*$ is jointly determined by $\alpha_1$ and $\alpha_2$. For the $\lambda > 0$ case, a model with a static measure of peer forecast has the restrictions $\alpha_1 = 1$ and $\alpha_2 = 0$. For the $\lambda < 0$ case, the model restricts $\alpha_1$ to be the same for all analysts as the last analyst conditional in the order $j$ \[1\] In both approximations, I maintain the assumption from the model with a forward-looking measure of peer forecast that $\alpha_0$ varies across $j$ and $J$.

Equations (5) and (6) is expressed in terms of $\mu_j$ and $c_j$ but $\mu_j$ itself is linked to $s_j = E + \xi_j$ and $c_j$ through the expression in (1). Whereas in the strategy from a quadratic objective function, the parameters in (1) are subsumed into the weights on actual, consensus and lagged forecast, in this setting the weights in Equation (1) can be estimated separately due to the non-linearity of $f_j^*$. The additional parameters I estimate are $\sigma_0$ and $\tau_j$ for $j = 2, \ldots, 10$. Note that in (1), $\hat{\mu}_{j-1}$ depends on analyst $j$’s inferences about $j - 1$’s strategy, i.e., $\alpha_0, \alpha_1$ and $\alpha_1$ in the approximation. I use the estimated parameters for $j - 1$ in the data to compute $\hat{\mu}_{j-1}$.

I estimate both Equations (5) and (6) with a maximum likelihood approach (see Appendix B for derivation of the likelihood function). Like the approach used for the strategy from the quadratic objective function, I first estimate the models without any restriction on how the $\alpha$’s vary with $j$ and $J$. Then I impose the inequality restrictions from the model where peer forecast is a forward-looking measure and compute the corresponding log-likelihood statistic. I also compute a log-likelihood statistic from the equality restrictions on $\alpha_1$ and $\alpha_2$ implied by the model where peer forecast is a static-measure.

6. Results

Prior to discussing the estimation results for each of the four models, I present in Table 1 some summary statistics about forecast error (i.e., $f_j - A$) and deviation from consensus (i.e., $f_j - c_j$). Panel A reports summary statistics after sorting the

\[1\] Even though the model predicts that $\alpha_1 = \infty$, such a restriction, of course, cannot be implemented empirically.
sample based on the order of forecasts $j$. Both the mean and median forecast error decreases as each analyst forecasts. This finding is loosely consistent with results from Richardson et al. (2004), although the previous study considers a different event window. The mean (median) forecast error for analysts announcing second is $0.198$ ($0.149$) while the mean (median) forecast error for analysts announcing tenth is $0.148$ ($0.078$). The average deviation from consensus also appears to decline as each analyst announces. However, this decline appears to be less pronounced than that for forecast error. The mean deviation from consensus for analysts announcing second is $0.001$ and -$0.007$ for analysts announcing tenth.

Panel B of Table 1 reports the summary statistics after sorting firms by coverage. Similar to Panel A, the mean and median forecast error are both decreasing in coverage. The mean (median) forecast error for firms covered by two analysts is 0.224 (0.119) while the mean (median) forecast error for those covered by 10 analysts is 0.158 (0.085). Regarding the deviation from consensus, the summary statistics do not appear to vary substantially with analyst coverage.

Table 2 presents the weights from estimating (4) without any restrictions. Panel A reports the estimated weights on actual earnings $\omega_E$. To illustrate the interpretation of estimates, I find that $\omega_E$ is 0.26 (standard error=0.01) for the second out of two forecasts while for the second out of ten forecasts, the estimate of $\omega_E$ is 0.20 (standard error=0.01). A comparison of successive rows in the table indicates that, conditional on the analyst’s order $j$, $\omega_E$ generally declines with the number of analysts remaining (although for the larger $j$’s, such as $j = 7$ and $j = 9$, there is weak evidence of an increase). This pattern is not consistent with the inequality relation predicted by the model where peer forecast is a forward-looking measure. The inconsistency becomes more pronounced when I focus on the estimates of the weight on consensus $\omega_c$ in Panel B. A model with a forward-looking peer measure predicts a smaller $\omega_c$ for earlier analysts relative to later analysts. In Panel B, almost all the weights increase with $J$ for any given $j$. For example, the estimate of $\omega_c$ for the fifth out of five forecasts is 0.74 (standard error=0.02) while the estimate is 0.98 (standard error=0.02) for the fifth out of eight forecasts.
The log-likelihood statistics in Panel C provide a formal test of the fit of the strategy from the model with a static measure of peer forecast against that from the model with a forward-looking measure of peer forecast. Under the inequality restrictions implied by the latter model, the log-likelihood is -25,883 while under the equality restrictions implied by the former model the log-likelihood is -26,010. The difference is a mere 127. Although there is no statistical result on the limiting distribution of this difference, based on a qualitative comparison and based on the comparisons in Panels A and B, I do not find evidence that a model with forward-looking peer forecast explains the data better than one with a static-looking peer forecast.

It is important to emphasize that it is likely not possible to infer the sign and magnitude of $\lambda$ based on estimates of $\omega_E$ and $\omega_c$ because each is a confluence of a number of model parameters. In the case of $\omega_E$, it incorporates both the weight the analyst places on his signal when he forms beliefs (i.e. $W^p_j(\sigma_0, \tau)$ from Equation (1)) and the weight he places on beliefs as part of his equilibrium strategy (i.e., $W^s_j(\lambda)$ from Equation (2)). Similarly, $\omega_c$ impounds the effect of both $W^s_{j-1}(\lambda)$ to the extent that consensus matters in analysts’ objective function as well as both $W^b_{j-1}(\sigma_0, \tau)$ and $W^s_{j-1}(\lambda)$ through $j$’s inferences about $j-1$’s beliefs. Absent the ability to observe $\sigma_0$ and $\tau$, it is not possible to disentangle $\lambda$ from those two parameters. However, I note that in Equation (3), the weight on consensus is positive if and only if $\lambda$ itself is positive. I interpret the generally large positive estimates of $\omega_c$ as evidence that in a model which assumes a quadratic objective function, the data can only be explained by the assumption that $\lambda > 0$.

In Table 3, I present the results from fitting the approximation of a strategy from an Absolute Value objective function with $\lambda < 0$ (i.e., Equation (5)). Panel A includes my estimates of $\alpha_0$ pertaining to the $j$-th out of $J$ total forecasts. The estimates of $\alpha_0$ generally increases with $J$ across all the columns of $j$. For example, $\alpha_0$ for the third out of three forecasts is 0.004 while the estimate is 0.377 for the third out of ten analysts. This trend is consistent with the inequality restriction implied by the model

\footnote{Because $\alpha_0$ and $\alpha_1$ are the two parameters of interest, particularly for distinguishing between static versus forward-looking measure of peer forecast, the estimates of $\sigma$ and $\tau_j$ from Equation (1) are not reported.}
with a forward-looking measure of peer forecast. The estimates of $\alpha_1$ in Panel B are more mixed. For some values of $j$, $\alpha_1$ increases with $J$. For example, $\alpha_1$ for the sixth out of six forecasts is 1.24 while $\alpha_1$ for the sixth out of ten forecasts is 0.31. However, for other values of $j$, for example $j = 3$, there is evidence that $\alpha_1$ is larger for earlier analysts than for earlier analysts.

For formal comparison of the models with static vs. forward-looking measure of peer forecast, I use the log-likelihood statistics in Panel C. With the inequality restrictions implied by the model with a forward-looking measure, I find that the log-likelihood is -16,877. This is substantially larger than the log-likelihood from imposing the equality restrictions implied by the model with a static measure. Although there is no exact statistical test for this setting, I use a likelihood ratio test as a loose guideline. The likelihood ratio test statistic is $2 \times (22,940 - 16,877) = 12,126$ and follows a chi-squared distribution. Even after accounting for the difference in the number of free parameters used in each model, it is possible to reject the model with equality restrictions in favor of the one with inequality restrictions. Thus, under the assumption that analysts have an absolute value objective function, a model in which peer forecast is a forward-looking measure explains the data better than one in which peer forecast is a static-measure.

Comparing Panel C across Tables 2 and 3, I conclude that the log-likelihood from fitting the strategy from a model with an absolute value objective function is higher than the log-likelihood from fitting the strategy from a model with a quadratic objective function. Although I also estimate the approximation in Equation (6) corresponding to the strategy from an Absolute Value objective function with $\lambda > 0$, I am unable to find admissible values of the approximation parameters which explain the data better than fitting (4) and (5). That is, for all parameters where $\alpha_0 > 0$ and $\alpha_2 > 0$, the likelihood corresponding to Equation (5) is strictly worse than the statistics reported in Panel C of Tables 2 and 3. Taken together, I conclude that a model based on the absolute value objective function with a forward-looking measure of peer forecast and $\lambda < 0$ is most plausible given the data.
7. Conclusion

In this paper, I investigated the empirical implications of theoretical assumptions about analysts’ objective function. Prior empirical research on analyst behavior typically do not articulate such a theory. A select number of studies which focus on behavior induced by incentives related to peer forecasts develop their empirical tests using one specific choice of assumption about the functional form the objective function and the measurement of peer incentives. I show that minor changes in those assumptions produce distinct analytical predictions about forecasting behavior and that the data is much more consistent with certain assumptions rather than others.

Specifically, I find that an objective function based on squared forecast error and squared deviation from consensus performs almost equally well at explaining the data as one based on squared forecast error and squared deviation from both past and future forecasts. I also find that an objective function based on the absolute value of forecast error and the absolute value of deviation from both past and future forecasts explains the data better than one based on the absolute value of forecast error and the absolute value of deviation from consensus. Across all four models, the one where the analysts are concerned with absolute forecast error and the absolute value of deviation from past and future forecasts is the most plausible empirically. In addition, the data is consistent with both a squared objective function in which analysts prefer to be close to peers as well as an absolute value objective function in which analysts prefer to be far from peers.

A multitude of additional assumptions were used in deriving the analytical results of interest and the corresponding empirical estimates. Some critical assumptions include the exogenous timing of forecast announcements, the absence of any conflicts-of-interest as well as homogeneity in incentives and information precision across different firms and analysts. The statistical tests I use to distinguish between various objective functions are not necessarily valid to the extent that these additional assumptions are not valid. Thus, my empirical comparison should be qualified in the sense that alternate assumptions along these additional dimensions may produce
different inferences about the most plausible objective function and I do not interpret my evidence as definitive support for a certain kind of peer-induced behavior. Overall, my study underscores the need for an explicit model in empirical analysis to draw economically meaningful inferences about analysts’ behavior.
Appendix A: Derivation of Model Solutions

This appendix contains the derivation of the model solutions. Each section corresponds to one specific assumption about analysts’ objective functions. I first consider the two models where the objective function is based on a static measure of peer forecast. The solutions for the model where the objective function is based on a forward-looking measure of peer forecast builds on these results. For brevity, I omit functional arguments where they are not ambiguous. In the case the the two models with a quadratic objective function, although in the model I assume that all analysts have the same $\lambda$, the derivations below allow for heterogeneity and, as such, $\lambda$ is indexed by $j$.

A.1 Quadratic objective function, static measure of peer forecast

When peer forecast is based on a static measure, every analyst’s behavior is similar to the last analyst. Specifically, the optimization problem is:

$$\min_{f_j} \mathbb{E}(f_j - E)^2 + \lambda_j (f_j - c_j)^2.$$ 

The first order condition from solving this optimization problem is:

$$f_j - \mathbb{E}(E|s_j, \hat{s}_1, \ldots, \hat{s}_{j+1}) + \lambda_j (f_j - c_j) = 0.$$ 

Thus each analyst forecasts:

$$f_j^* = \frac{\lambda_j c_j + \frac{1}{\lambda_j + 1} \mu_j}{1 + \lambda_j}.$$ 

(A.1.1)

A.2 Absolute Value objective function, static measure of peer forecast

For the absolute Value loss function, consider first the $\lambda = 0$ case as a benchmark. When the analyst’s payoffs only depend on the expected unsigned forecast error, the optimal forecast is the median of his beliefs about earnings. As noted in Gu and Wu (2003), normally distributed posteriors have the well-known property that the median coincides with the mean, $\mu_j$. In general, the strategies have to be solved backwards like those for the quadratic loss function. I use $\Phi(\cdot)$ and $\phi(\cdot)$ to denote the standard normal distribution and density functions, respectively.
Each analyst solves the optimization problem:

$$\min_{f_{j+1}} \mathbb{E} ( |f_j - E| ) + \lambda_j |f_j - c_j| .$$

Since the objective function is not differentiable at $f_j = c_j$, the optimization problem has to be solved piece-wise over two intervals. Over the range $f_j \geq c_j$, the first order condition describing the optimal forecast is, denoted by $f_j^0$ is:

$$2 \left( \frac{1}{\hat{\sigma}_j} \right) \phi \left( \frac{f_j^0 - \mu_j}{\hat{\sigma}_j} \right) (f_j^0 - \mu_j) + \left[ 2 \Phi \left( \frac{f_j - \mu_j}{\hat{\sigma}_j} \right) - 1 \right]$$

$$- 2 \left( \frac{f_j - \mu_j}{\hat{\sigma}_j} \right) \phi \left( \frac{f_j^0 - \mu_j}{\hat{\sigma}_j} \right) + \lambda = 0,$$

which implies $f_j^0 = \mu_j + \hat{\sigma}_j \cdot \Phi^{-1} \left( \frac{\lambda}{2} \right)$. Similarly, there is another first order condition characterizing the optimal forecast over the interval where $f_j < c_j$:

$$2 \left( \frac{1}{\hat{\sigma}_j} \right) \phi \left( \frac{f_j^1 - \mu_j}{\hat{\sigma}_j} \right) (f_j^1 - \mu_j) + \left[ 2 \Phi \left( \frac{f_j - \mu_j}{\hat{\sigma}_j} \right) - 1 \right]$$

$$- 2 \left( \frac{f_j^1 - \mu_j}{\hat{\sigma}_j} \right) \phi \left( \frac{f_j^1 - \mu_j}{\hat{\sigma}_j} \right) - \lambda = 0.$$

So the corresponding solution is $f_j^1 = \mu_j + \hat{\sigma}_j \cdot \Phi^{-1} \left( \frac{\lambda}{2} \right)$. And if neither $f_j^0$ or $f_j^1$ are feasible, the analyst forecasts $c_j$.

There are certain solutions that can be ruled out based on knowledge that $\lambda < 0$. For illustration, suppose that $\mu_j > c_j$, then any forecast less than $\mu_j$ is suboptimal because there is another forecast symmetric around $\mu_j$ where he in expectation, he incurs the same forecast error but with certainty he will be better off due to being further away from consensus. Forecasting $c_j$ is never optimal because the benefit of deviating is constant and the analyst always benefits from this deviation in the direction that would reduce his forecast error. Furthermore since $\lambda < 0$, the objective function is always increasing at $f_j = c_j$, so $f_j^0$ is also the global optimum as long as $\mu_j \neq c_j$. Similar reasoning can be applied when $\mu_j < c_j$. Accordingly, when $\lambda < 0$, the
optimal strategy is:

\[
 f^*_j = \begin{cases} 
 \mu_j + \hat{\sigma}_j \cdot \Phi^{-1} \left( \frac{1 + \lambda}{2} \right) & \text{if } \mu_j + \hat{\sigma}_j \cdot \Phi^{-1} \left( \frac{1 + \lambda}{2} \right) < c_j \\
 c_j & \text{if } \mu_j \in \left[ c_j + \hat{\sigma}_j \cdot \Phi^{-1} \left( \frac{1 - \lambda}{2} \right) , c_j + \hat{\sigma}_j \cdot \Phi^{-1} \left( \frac{1 + \lambda}{2} \right) \right] \\
 \mu_j + \hat{\sigma}_j \cdot \Phi^{-1} \left( \frac{1 - \lambda}{2} \right) & \text{if } \mu_j + \hat{\sigma}_j \cdot \Phi^{-1} \left( \frac{1 - \lambda}{2} \right) > c_j.
\end{cases}
\]  

(A.2.1)

In contrast, when \( \lambda < 0 \), it is never optimal for analysts to deviate from beliefs in the opposite direction as consensus. As such over analysts always prefer to forecast \( f^0_j \) when \( \mu_j > c_j \) over \( f^1_j \) and they prefer \( f^1_j \) over \( f^0_j \) if \( \mu_j < c_j \). However with \( \lambda > 0 \), the corner solution of not deviating from consensus may be optimal. Formally: as long as \( \mu_j \neq c_j \). Similar reasoning can be applied when \( \mu_j < c_j \). Accordingly, when \( \lambda < 0 \), the optimal strategy is:

\[
 f^*_j = \begin{cases} 
 \mu_j + \hat{\sigma}_j \cdot \Phi^{-1} \left( \frac{1 + \lambda}{2} \right) & \text{if } \mu_j < c_j \\
 \mu_j & \text{if } \mu_j = c_j \\
 \mu_j + \hat{\sigma}_j \cdot \Phi^{-1} \left( \frac{1 - \lambda}{2} \right) & \text{if } \mu_j > c_j.
\end{cases}
\]  

(A.2.2)

A.3 Prediction about subsequent analysts’ beliefs

In order to derive strategies in the models where peer forecast is forward-looking, I first establish an expression for analysts’ expectations about subsequent analysts’ earnings predictions. I assume (for now) that analysts’ strategies are strictly monotone in their signals. Under such an assumption, analyst \( k \)'s prediction about \( j \)'s signal \( \hat{s}_{k,j}(f^*_{it,j}) = s_j \ \forall k > j \). As such, analyst \( j \)'s beliefs about analyst \( k \)'s beliefs can be computed using a general version of the law of iterated expectations.

\[
 E(\mu_k | s_1, \ldots, s_j) = \mathbb{E}_{s_{j+1}, \ldots, s_k} \left[ E_E \left( E | s_1, \ldots, s_k \right) | s_1, \ldots, s_j \right] \\
 = \int \cdots \int_{\mathbb{R}^{k-j+1}} E g(E | s_1 \ldots s_k) g(s_{j+1} \ldots s_k | s_1 \ldots s_j) dE \ ds_{j+1} \cdots ds_k \\
 = \int \cdots \int_{\mathbb{R}^{k-j+1}} E g(E, s_1, \ldots, s_j, s_{j+1}, \ldots, s_k) g(s_1, \ldots, s_k) g(s_{j+1} \ldots s_k) dE \ ds_{j+1} \cdots ds_k \\
 = \int_E E g(E, s_1, \ldots, s_j) \ g(s_1, \ldots, s_j) \ dE \\
 = \mu_j
\] 

(A.3.1)
A.4 Quadratic objective function, forward-looking measure of peer forecasts

For the last analyst, the forward-looking measure of peer forecast is the same as the static measure. Since the strategies for the model with a static measure of peer forecast has already solved, I substitute $J + 1$ into Equation (A.1.1):

$$f_{J+1}^* = \frac{\lambda_{J+1}}{1 + \lambda_{J+1}} c_{J+1} + \frac{1}{1 + \lambda_{J+1}} \mu_{J+1}. \quad (A.4.1)$$

Equation (A.4.1) shows that analyst $J + 1$’s strategy is a convex combination of consensus and beliefs and, as such, is fully monotonic in beliefs. Suppose that for analysts $k \geq j + 1$ forecast strategies are convex combinations of $\mu_k$ and $c_k$. Analyst $j$’s first order condition is:

$$2(f_j - \mu_j) + 2\lambda_j \left( f_j - \frac{c_j(j - 1)}{J} - \frac{\sum_{k=j+1}^{J+1} \mathbb{E}(f_k^*)}{J} \right) \left( 1 - \frac{d}{df_j} \sum_{k=j+1}^{J+1} J^{-1} f_k^* \right) = 0. \quad (A.4.1)$$

Since $f_{j+1}^*$ through $f_{J+1}^*$ are linear in beliefs and consensus by assumption, then they are also linear in $f_j$ and $c_j$. I define the constant $v_j = \frac{d}{df_j} \sum_{k=j+1}^{J+1} f_k^*$ so that the optimality condition can be re-expressed as:

$$2(f_j - \mu_j) + 2\lambda_j \left( f_j - \frac{c_j(j - 1)}{J} - \frac{(1 - v_j) \sum_{k=j+1}^{J+1} \mathbb{E}(\mu_k) + v_j f_j + (j - 1)v_j c_j}{J} \right) \left( 1 - \frac{v_j}{J} \right) = 0. \quad (A.4.1)$$

Using the result from Equation (A.3.1), $\sum_{k=j+1}^{J+1} \mathbb{E}(\mu_k) = (J - j) \mu_j$. After simplifying and solving this expression, the optimal forecast satisfies:

$$f_j^* = \frac{\lambda_j(j - 1)(1 + v_j)(1 - \frac{v_j}{J})}{J \left( 1 + \lambda_j \left( 1 - \frac{v_j}{J} \right)^2 \right)} c_j + \left[ 1 - \frac{\lambda_j(j - 1)(1 + v_j)(1 - \frac{v_j}{J})}{J \left( 1 + \lambda_j \left( 1 - \frac{v_j}{J} \right)^2 \right)} \right] \mu_j. \quad (A.4.2)$$

Equation (A.4.2) shows that $f_j^*$ can be expressed as a convex combination of $c_j$ and $\mu_j$ if $f_k^*$ can be expressed as a convex combination of $\mu_k$ and $c_k$ for $k \geq j + 1$.

By the induction principle, Equation (A.4.2) holds for all $j \geq 2$. To complete the description of the strategy, note that $\frac{df_{j+2}^*}{df_j} = \frac{df_{j+2}^*}{df_j} + \frac{df_{j+2}^*}{df_{j+1}} \frac{df_{j+1}^*}{df_j} = \frac{df_{j+2}^*}{df_{j+1}} \left( 1 + \frac{df_{j+1}^*}{df_j} \right)$. From
j’s perspective, his own forecast enters into j + 2’s strategy directly through the effect of $f_j$ on $c_{j+2}$. However, $f_j$ also enters into $f_{j+1}^*$, which in turns enters into $j + 2$’s strategy through $c_{j+2}$. With this identity, I perform the following expansion:

$$\sum_{k=j+1}^{J+1} \frac{df_k^*}{df_j} = \frac{df_{j+1}^*}{df_j} + \left( 1 + \frac{df_{j+1}^*}{df_j} \right) \sum_{k=j+2}^{J+1} \frac{df_k^*}{df_j}$$

Using the definition of $\frac{df_k^*}{df_j}$ and differentiating Equation (A.4.2) with respect to $c_j$, I derive the following recursive definition for $v_j$:

$$v_j = \left( 1 + v_{j+1} \right) \frac{\lambda_{j+1} \left( 1 + v_{j+1} \right) \left( 1 - \frac{v_{j+1}}{J} \right)}{J \left( 1 + \lambda_{j+1} \left( 1 - \frac{v_{j+1}}{J} \right)^2 \right)} + v_{j+1}, \quad (A.4.3)$$

where $v_{J+1} = 0$ by Equation (A.4.1).

First Analyst

The first analyst does not observe other analysts’ forecasts and instead only sees his own beliefs. In addition, his beliefs about all other analysts’ beliefs is also $\mu_1$ (by Equation (A.3.1)). His optimization problem is:

$$\min_{f_1} E \left( (f_1 - e)^2 \right) + \lambda_1 \left( f_1 - J^{-1} \sum_{k=2}^{J+1} f_k^* \right)^2.$$  

The first order condition for the first analyst is:

$$f_1 - \mu_1 + \lambda_1 \left( f_1 - J^{-1} \sum_{k=2}^{J+1} f_k^* \right) \left( 1 - J^{-1} \sum_{k=2}^{J+1} \frac{df_k^*}{df_1} \right) = 0. \quad (A.4.4)$$

If he forecasts $\mu_1$, then in expectation, the second analyst will forecast $f_2^* = \mu_1$ by Equation (A.4.2). Applying Equation (A.3.1) iteratively over subsequent analysts’ strategies and computing an expectation, it becomes clear that forecasting $\mu_1$ is a solution to the optimality condition in (A.4.4). Since $f_1^* = \mu_1$, the first analyst is unbiased.
A.5 Absolute Value objective function, forward-looking measure

Strategies for $\lambda < 0$ case

For the last analyst, peer forecast in this model is simply consensus, which is the same as the static measure of peer forecast. The solution for the model with a static measure of peer forecast was already derived in the previous section. Using the expression in (A.2.2) and substituting $j = J + 1$, I obtain:

\[
 f_j^* = \begin{cases} 
 \mu_{J+1} + \hat{\sigma}_{J+1} \cdot \Phi^{-1} \left( \frac{1+\lambda}{2} \right) & \text{if } \mu_{J+1} < c_{J+1} \\
 \mu_{J+1} & \text{if } \mu_{J+1} = c_{J+1} \\
 \mu_{J+1} + \hat{\sigma}_{J+1} \cdot \Phi^{-1} \left( \frac{1-\lambda}{2} \right) & \text{if } \mu_{J+1} > c_{J+1}.
\end{cases}
\]  

(A.5.1)

For convenience, I express strategies as the the optimal bias $b_j^* = f_j^* - \mu_j$. Consistent with the notation in 4.2 I denote $D_j = \mu_j - c_j$ as the difference between analyst $j$’s prediction and consensus forecast. For $J + 1$, the above expressions for $f_{J+1}^*$ shows that $D_j$ is a sufficient statistic for $b_j^*$. Specifically, when $\lambda < 0$, the optimal bias for the last analyst can be written as:

\[
b_{J+1}^*(D) = \hat{\sigma}_{J+1} \cdot \Phi^{-1} \left( \frac{1-\lambda}{2} \right) [1 (D > 0) - 1 (D < 0)].
\]  

(A.5.2)

By expressing the optimal strategy in terms of bias, then for analyst $j$ the deviation from peer performance can be simplified into a function of $b_j$ and $D_j$ using the following steps.

\[
f_j - \sum_{k \neq j} f_k = b_j + \mu_j - c_j - \frac{1}{J} \sum_{l=J+1}^{J+1} f_l^* \\
= b_j - \frac{(c_j - \mu_j)(j - 1)}{J} + \mu_j \left( 1 - \frac{j - 1}{J} \right) - \frac{1}{J} \left( \sum_{l=J+1}^{J+1} \mu_l + \sum_{l=J+1}^{J+1} b_l \right) \\
= b_j + D_j \frac{j - 1}{J} + \left( 1 - \frac{j - 1}{J} \right) \mu_j - \frac{1}{J} \sum_{l=J+1}^{J+1} \left( \mathbb{E}(\mu_l | \mu_j) + z \cdot \hat{\sigma}_{j,l} \right) - \frac{1}{J} \sum_{l=J+1}^{J+1} b_l \\
= b_j + D_j \frac{j - 1}{J} - \frac{1}{J} \sum_{l=J+1}^{J+1} b_l - \frac{z}{J} \sqrt{\sum_{l=J+1}^{J+1} \hat{\sigma}_{j,l}^2}
\]

where $z \sim N(0, 1)$, $\hat{\sigma}_{j,l}$ is the standard deviation of analyst $j$’s posterior beliefs about future analyst $l$’s earnings prediction and the last equality holds because of the result
in Equation (A.3.1).

For analysts $j = 2, \ldots, J$, the optimization problem can be re-characterized such that he chooses the level of bias based on $D_j$:

$$\min_{b_j} \int \left| b_j - y \cdot \hat{\sigma}_j \right| \phi(y) dy$$

$$+ \lambda \int b_j + D_j \frac{j-1}{J} - \frac{1}{J} B_{j+1} \left( \frac{b_j - D_j(j-1)}{j} + y \cdot \hat{\sigma}_{j,j+1} \right) - \frac{y}{J} \sum_{l=j+1}^{J+1} \hat{\sigma}_{j,l}^2 \phi(y) dy,$$

(A5.3)

where $B_j(D) = b_j^*(D) + \int B_{j+1} \left( \frac{b_j^*(D) - D(j-1)}{j} + y \cdot \hat{\sigma}_{j,j+1} \right) \phi(y) dy$ and $B_{J+1}(D) = b_{J+1}^*(D)$. In particular, the original optimization problem requires computing an expectation over the joint density of future analysts’ prediction because each of $f_{j+1}^*, \ldots, f_{J+1}$ are random from $j$’s perspective. Using this modified formulation of the optimization problem, the sum of future analysts’ forecasts can be expressed as a function, $B_j(\cdot)$. Since $B_j(\cdot)$ can is defined recursively, it is sufficient consider only the expectation over the density of the $j+1$-st analyst’s prediction. Equivalently, $B_j(\cdot)$ can be thought of a function that captures the extent to which future analysts’ best response to $j$’s choice of bias enters his payoffs.

Based on the revised optimization problem, the following properties hold when $D > 0$ (where the properties of the $D < 0$ are exactly symmetric):

i) **The bias positive, with $b^*(D) > 0$.**

In Equation (A5.2), $b_{J+1}^*(D)$ is symmetric around 0 and is positive (negative) whenever $D > 0$ ($D < 0$). The induction principle can be used to show that this property also holds for $B_j(\cdot)$ whenever it holds for $B_j(\cdot), \ldots, B_{J+1}(\cdot)$. If the analyst forecasts $b_j = 0$, then with greater than 50% probability $B_{J+1} \left( \frac{b_j}{J} - \frac{D_j(j-1)}{j} + y \cdot \hat{\sigma}_{j,j+1} \right) < 0$. For a fixed negative value of $b_j < 0$, the shape of the normal density implies that there is a relatively minor impact on the probability of negative vs. positive $B_{J+1}$’s. When the analyst chooses a positive $b_j$ of the same magnitude instead, the location of $B_{J+1} \left( \frac{b_j}{J} - \frac{D_j(j-1)}{j} + y \cdot \hat{\sigma}_{j,j+1} \right) < 0$ shifts closer to zero and, as such, there is a large increase in the probability of positive $B_{J+1}$’s relative to negative ones. A fixed value of $b_j$, whether positive or negative, has the same impact on expected forecast error.

ii) **The bias is bounded above by $\hat{\sigma}_j \Phi^{-1} \left( \frac{1-\lambda}{2} \right)$**.
In Equation (A.5.2), \( b_{J+1}^*(D) > 0 \) and \( \frac{\partial b_{J+1}^*(D)}{\partial D} \) is bounded between 0 and 1. Using the principle of induction, one can show that \( \frac{\partial b_{J+1}^*(D)}{\partial D} \) must be bounded between 0 and \( \frac{\ell - J + 1}{J + 1} \). The marginal benefit of deviating from consensus is greater than \( \lambda \). Therefore \( b_{j}^*(D) \) is always less than a setting where the marginal benefit is at least \( \lambda \) (i.e., Equation (A.2.2)).

iii) The bias is concave in \( D_j \), with \( \frac{\partial^2 b_{j}^*(D)}{\partial D^2} < 0 \)

The absolute value operator is a convex function. The first term in (A.5.3) is convex in \( b_j \) while the second term is concave (due to \( \lambda < 0 \)). The second term is always more concave than the first term is convex. Thus an increase in \( D_j \) always changes \( b_{j}^*(D_j) \) by more than the increase.

**Strategies assuming that \( \lambda > 0 \)**

In the case where \( \lambda > 0 \), I use a similar approach for the last analysts’ strategy as that used in the \( \lambda < 0 \) case. That is, I use the result in Equation (A.2.1) with the substitution that \( j = J + 1 \):

\[
\begin{align*}
  f_{j+1}^* = \begin{cases} 
    \mu_{j+1} + \hat{\sigma}_{j+1} \cdot \Phi^{-1} \left( \frac{1 + \lambda}{2} \right) & \text{if } \mu_{j+1} + \hat{\sigma}_{j+1} \cdot \Phi^{-1} \left( \frac{1 + \lambda}{2} \right) < c_{j+1} \\
    c_{j+1} & \text{if } \mu_{j+1} \in \left[ c_{j+1} + \hat{\sigma}_{j+1} \cdot \Phi^{-1} \left( \frac{1 - \lambda}{2} \right), c_{j+1} + \hat{\sigma}_{j+1} \cdot \Phi^{-1} \left( \frac{1 + \lambda}{2} \right) \right] \\
    \mu_{j+1} + \hat{\sigma}_{j+1} \cdot \Phi^{-1} \left( \frac{1 - \lambda}{2} \right) & \text{if } \mu_{j+1} + \hat{\sigma}_{j+1} \cdot \Phi^{-1} \left( \frac{1 - \lambda}{2} \right) > c_{j+1}.
  \end{cases}
\end{align*}
\]

Also following my analysis for the \( \lambda < 0 \) case, I write all analysts’ strategies in terms of the optimal bias \( b_{j}^* \) in terms of the difference between the earnings prediction and consensus \( D_j \). After subtracting \( \mu_j \) from the strategy in Equation (A.5.4), the optimal bias for the last analyst is:

\[
b_{J+1}^*(D) = \min \left\{ D_{J+1} \cdot \Phi^{-1} \left( \frac{1 + \lambda}{2} \right) \right\} \left[ 1 (D < 0) - 1 (D > 0) \right]. \tag{A.5.5}
\]

For middle analysts forecasting in the \( j \)-th order, they solve the optimization problem in (A.5.3). Based on an analysis of this revised objective function, \( f_{j}^* \) has the following properties whenever \( D_j > 0 \):

i) There is a threshold over which \( b_{j}^*(D_j) = -D_j \).

The variance of the analyst’s beliefs over future analysts’ beliefs is always smaller than those about earnings itself, with \( \hat{\sigma}_{j,l} \leq \hat{\sigma}_j, l = j + 1 \ldots J + 1 \). As a result, deviations from expected future forecasts are costlier than deviations from expected earnings. The result is that for small values of \( D_j \), the analyst prefers
to choose $b_j = -D_j$ so that the deviation from peer forecasts is equal to zero, in expectation, even if such a strategy results in a small amount of forecast error.

ii) **The bias is negative, with $b_j^*(D_j) < 0$.**

The reasoning is similar to the $\lambda < 0$ case. In Equation (A.5.5), the bias for the last analyst is weakly negative for $D > 0$. If the same property about $b_k(D)$ holds for all $k = j + 1, \ldots, J$, then $B_{j+1}(D) < 0$. By choosing $b_j < 0$, the analyst increases the expected value of $-\frac{1}{J} B_{j+1}(\frac{b_j}{J} - D_j \frac{j+1}{J} + y \hat{o}_{j,j+1})$ and therefore minimizes the second term in Equation (A.5.3). By the induction principle, $b_j(D) < 0$ holds for all $j$.

iii) **The bias is bounded below by $\hat{o}_j \Phi^{-1}(\frac{1-\lambda}{2})$.**

Similar to the reasoning for the $\lambda < 0$ case, Equation (A.5.5) shows that $b_{j+1}^*(D)$ is negative and $\frac{\partial b_{j+1}^*}{\partial D}$ is bounded between -1 and 0. Once again using an induction argument, if these two properties hold for $k = j + 1, \ldots, J$ then $\frac{\partial B_{j+1}^*(D)}{\partial D}$ must be bounded between $-J-j+1$ and 0. The marginal cost of deviating from consensus is always greater than $\lambda$. Therefore the analyst will never choose $b_j$ which is more negative than what they would choose when faced with certainty that their forecast will exceed consensus. From the derivation in Section A.2, we know that this latter quantity is exactly $\hat{o}_j \Phi^{-1}(\frac{1-\lambda}{2})$. To close the loop on the induction step, if $b_j^*(D)$ is bounded below by this quantity and since result (i) holds, then $\frac{\partial b_j^*(D)}{\partial D}$ must be greater than -1 for all $j$. 

Appendix B: Likelihood Function for Strategy Under Absolute Value Objective Function

This appendix derives the likelihood function for the approximation of the forecast strategies under an Absolute Value Objective Function. I derive the likelihood separately for the $\lambda < 0$ and $\lambda > 0$ case. Since I use a piece-wise approximation for both cases, the likelihood has to be derived piece-wise using the density function of $f_j^*$ for each region.

B.1 Approximation Function for the $\lambda < 0$ case

To reiterate the discussion in Section 5.2, the approximation I use for the strategy from the absolute value objective function when $\lambda < 0$ is:

$$f_j = \begin{cases} 
\mu_j - \text{erf}(\alpha_0 \alpha_1) & \text{if } \mu_j < c_j - \alpha_0 \\
\text{erf} \left( \alpha_1 \cdot (\mu_j - c_j) \right) + c_j & \text{if } c_j - \alpha_0 \leq \mu_j \leq c_j + \alpha_0 \\
\mu_j + \text{erf}(\alpha_0 \alpha_1) & \text{if } \mu_j > c_j + \alpha_0.
\end{cases}$$ (B.1.1)

I denote the cumulative distribution function of $f_j^*$ as $G(\cdot)$. Over the range where $|\mu_j - c_j| < \alpha_0$, the approximation is non-linear in $\mu_j$ and, as such, it is non-linear in the unobservable term $\xi_j$. Thus, the density cannot be determined simply by inverting $f_j^*$ in terms of $\xi_j$ and has to be derived using first principles, starting the distribution function, as follows:

$$G(r) = Pr(f_j^* \leq r) = Pr \left[ \text{erf} \left( \alpha_1 \left( \mu_j - c_j \right) \right) \leq r \right] = Pr \left( \mu_j - c_j \leq \frac{\text{erf}^{-1}(r - c_j)}{\alpha_1} \right)
$$

$$= Pr \left[ W_j^p(\sigma_0, \bar{\tau})\hat{\mu}_{j-1} + (1 - W_j^P(\sigma_0, \bar{\tau})) s_j \leq c_j + \frac{\text{erf}^{-1}(r - c_j)}{\alpha_1} \right]
$$

$$= Pr \left[ \tau_j \left( 1 - W_j^P(\sigma_0, \bar{\tau}) \right) \xi_j \leq c_j + \frac{\text{erf}^{-1}(r - c_j)}{\alpha_1} - \mathbb{E} \left( \mu_{it,j} | E_{it} \right) \right]
$$

$$= Pr \left[ \xi_j \leq \frac{1}{\tau_j \left( 1 - W_j^P(\sigma_0, \bar{\tau}) \right)} \left( \frac{\text{erf}^{-1}(r - c_j)}{\alpha_1} - \mathbb{E} \left( \mu_{it,j} | E_{it} \right) + c_j \right) \right].$$

In the above expression, $\mathbb{E} \left( \mu_{it,j} | E_{it} \right) = W_j^P(\sigma_0, \bar{\tau})\hat{\mu}_{j-1} + (1 - W_j^P(\sigma_0, \bar{\tau})) E$ is the de-
terministic portion of $\mu_j$ observable to the researcher. Using the assumption that $\xi_j \sim N(0, 1)$, the density function follows:

$$g(r) = \frac{d}{dr} \Phi \left[ \frac{1}{\tau_j (1 - W^p_j (\sigma_0, \bar{\tau}))} \left( \frac{\text{erf}^{-1} (r - c_j) - \mathbb{E} (\mu_{it,j} | E_{it}) + c_j}{\alpha_1} \right) \right]$$

$$= \Phi' \left[ \frac{1}{\tau_j (1 - W^p_j (\sigma_0, \bar{\tau}))} \left( \frac{\text{erf}^{-1} (r - c_j) - \mathbb{E} (\mu_{it,j} | E_{it}) + c_j}{\alpha_1} \right) \right] \frac{d}{dr} \frac{\text{erf}^{-1} (r - c_j)}{\tau_j (1 - W^p_j (\sigma_0, \bar{\tau}))}$$

$$= \frac{\sqrt{\pi} e^{\text{erf}^{-1} (f_j - c_j)^2}}{2\alpha_1 \tau_j (1 - W^p_j (\sigma_0, \bar{\tau}))} \Phi \left[ \frac{1}{(1 - W^p_j (\sigma_0, \bar{\tau}))} \tau_j \left( \frac{\text{erf}^{-1} (f_j - c_j)}{\alpha_1} - \frac{\mathbb{E} (\mu_{it,j} | E_{it}) - c_j}{\tau_j (1 - W^p_j (\sigma_0, \bar{\tau}))} \right) \right] \tag{B.1.2}$$

Note that in computing $g(r)$ from $G(r)$, the non-linear relation between $f_j^*$ and $\xi_j$ is accounted for through the inverse of the Jacobian of this transformation.

For the regions where $\mu_j - c_j > \alpha_0$, the approximation is linear. So $f_j^*$ can be inverted where $\mu_j = f_j - \text{erf}(\alpha_0 \alpha_1)$. Using the expression for $\mu_j$ in Equation (1), $\xi_j$ so that over this region, the density is:

$$g(r) = \phi \left( \frac{f_j - \mathbb{E} (\mu_{it,j} | E_{it}) - \text{erf} (\alpha_0 \alpha_1)}{\tau_j (1 - W^p_j (\sigma_0, \bar{\tau}))} \right). \tag{B.1.3}$$

The density over the region where $\mu_j - c_j < -\alpha_0$ can be derived similarly, with:

$$g(r) = \phi \left( \frac{f_j - \mathbb{E} (\mu_{it,j} | E_{it}) + \text{erf} (\alpha_0 \alpha_1)}{\tau_j (1 - W^p_j (\sigma_0, \bar{\tau}))} \right). \tag{B.1.4}$$

Combining Equations (B.1.1), (B.1.2) and (B.1.3) and taking the natural logarithm, the log-likelihood can be written as:
\[
\ln \mathcal{L} (\alpha_0, \alpha_1, \sigma_0, \tau) = \sum_{it} -\frac{1}{2} \ln \left( \tau_j^2 (1 - W_j^p (\sigma_0, \tau))^2 \right) \\
+ \ln \left( \left[ \left( \text{erf}^{-1}(f_{it,j} - c_{it,j}) \right) \right]^2 - \frac{\left( \mathbb{E} (\mu_{it,j} | E_{it}) - c_{it,j} - \alpha_1^{-1} \text{erf}^{-1}(f_{it,j} - c_{it,j}) \right)^2}{\tau_j^2 (1 - W_j^p (\sigma_0, \tau))^2} - \ln(\alpha_1) \right)
\]

Using the log-likelihood functions in Equation (B.1.5), I search over the parameters \( \lambda, \sigma_0 \) and \( \tau_j \) which maximizes the expression subject to the restrictions that all \( \sigma_0 \) and \( \tau_j \) have to be strictly positive and that \( \lambda \) must be strictly negative.

### B.2 Approximation Function for the \( \lambda > 0 \) case

When \( \lambda > 0 \), recall from Section 5.2 that the approximation of the strategy is is:

\[
f_j = \begin{cases} 
-\alpha_2(\mu_j - c_j + \alpha_0)^2 + \alpha_1(\mu_j - c_j + \alpha_0) + c_j & \text{if } \mu_j < c_j - \alpha_0 \\
\alpha_2(\mu_j - c_j - \alpha_0)^2 + \alpha_1(\mu_j - c_j - \alpha_0) + c_j & \text{if } \mu_j > c_j + \alpha_0,
\end{cases}
\]

Similar to the \( \lambda < 0 \) case, I use \( G(r) \) to denote the distribution function of \( f_j^* \). Over the region where \( \mu_j > c_j + \alpha_0 \), the mapping between \( f_j^* \) and \( \xi_j \) is non-linear (because \( \mu_j \) depends on \( \xi_j \) and \( f_j^* \) is non-linear in \( \mu_j \)). The cumulative distribution function, determined using first principles, follows:

\[
G(r) = Pr(\alpha_2(\mu_j - c_j - \alpha_0)^2 + \alpha_1(\mu_j - c_j - \alpha_0) + c_j \leq r) \\
= Pr(\mu_j \leq \alpha_0 + c_j + \frac{\sqrt{\alpha_1^2 + 4\alpha_2(r - c_j) - \alpha_1} \cdot 2\alpha_2}{2\alpha_2}) \\
= Pr(\xi_j \leq (1 - W_j^p (\sigma_0, \tau))^{-1} \frac{1}{\tau_j} \left( \alpha_0 + c_j - \mathbb{E} (\mu_{it,j} | E_{it}) + \frac{\sqrt{\alpha_1^2 + 4\alpha_2(r - c_j) - \alpha_1} \cdot 2\alpha_2}{2\alpha_2} \right)).
\]

Since \( \xi_j \) is standard normal, I can differentiate \( G(r) \) with respect to \( r \) (with the
appropriate application of the chain rule) to obtain the density function:

\[
g(r) = \phi \left[ (1 - W_j^P (\sigma_0, \bar{\tau}))^{-1} \tau_j^{-1} \left( \alpha_0 + c_j - \mathbb{E} (\mu_{it,j} | E_{it}) + \frac{\sqrt{\alpha_1^2 + 4\alpha_2(r - c_j)} - \alpha_1}{2\alpha_2} \right) \right] \\
\cdot \frac{(1 - W_j^P (\sigma_0, \bar{\tau}))^{-1} \tau_j^{-1}}{\sqrt{\alpha_1^2 + 4\alpha_2(r - c_j)}}.
\]

In deriving \( g(r) \) from \( G(r) \), the non-linearity of the transformation of \( \xi_j \) in \( f_j^* \) is reflected as a term representing the inverse of the Jacobian of this transformation. Using steps similar to those in the above analysis, I find that over the region where \( \mu_j < c_j - \alpha_0 \), the distribution function is:

\[
G(r) = \Pr \left[ \xi_j \leq (1 - W_j^P (\sigma_0, \bar{\tau}))^{-1} \tau_j^{-1} \left( c_j - \mathbb{E} (\mu_{it,j} | E_{it}) - \alpha_0 + \frac{\alpha_1 - \sqrt{\alpha_1^2 + 4\alpha_2(r - c_j)}}{2\alpha_2} \right) \right].
\]

After differentiating the above expression, with the appropriate application of the chain rule, I obtain the corresponding density function:

\[
g(r) = \phi \left[ (1 - W_j^P (\sigma_0, \bar{\tau}))^{-1} \tau_j^{-1} \left( c_j - \mathbb{E} (\mu_{it,j} | E_{it}) - \alpha_0 + \frac{\alpha_1 - \sqrt{\alpha_1^2 + 4\alpha_2(r - c_j)}}{2\alpha_2} \right) \right] \\
\cdot \frac{(1 - W_j^P (\sigma_0, \bar{\tau}))^{-1} \tau_j^{-1}}{\sqrt{\alpha_1^2 + 4\alpha_2(r - c_j)}}.
\]

When \( \mu_j \in [c_j - \alpha_0, c_j + \alpha_0] \), the likelihood function is equal to the difference between two probabilities:

\[
\Pr (\mu_j \leq c_j + \alpha_0) - \Pr (\mu_j < c_j - \alpha_0).
\]

Finally, the log-likelihood function can be obtained by combining the probability distribution functions over the regions where \( \mu_j < c_j - \alpha_0, \mu_j > c_j + \alpha_0 \) and \( \mu_j \in [c_j - \alpha_0, c_j + \alpha_0] \):
\ln \mathcal{L} (\alpha_0, \alpha_1, \alpha_1, \sigma_0, \tau) = \sum_{it} 1(\lfloor f_{it,j} - c_{it,j} \rceil \leq \alpha_0).

\begin{align*}
\ln \Phi &\left[ (1 - W_j^p (\sigma_0, \tau))^{-\frac{1}{2}} \tau_j^{-1} \left( \alpha_0 + c_{it,j} - \mathbb{E}(\mu_{it,j} | E_{it}) + \frac{\sqrt{\alpha_1^2 - 4\alpha_2 (f_{it,j} - c_{it,j}) - \alpha_1^2}}{2\alpha_2} \right) \right] \\
-\Phi &\left[ (1 - W_j^p (\sigma_0, \tau))^{-\frac{1}{2}} \tau_j^{-1} \left( c_{it,j} - \mathbb{E}(\mu_{it,j} | E_{it}) - \alpha_0 + \frac{\alpha_1 - \sqrt{\alpha_1^2 - 4\alpha_2 (f_{it,j} - c_{it,j})}}{2\alpha_2} \right) \right] \\
\sum_{it} 1(f_{it,j} - c_{it,j} > \alpha_0) \left\{ -\frac{1}{2} \ln \left[ (\alpha_1^2 + 4\alpha_2 (f_{it,j} - c_{it,j})) \tau_j^2 (1 - W_j^p (\sigma_0, \tau))^2 \pi \right] \\
- \left[ (1 - W_j^p (\sigma_0, \tau))^{-2} \tau_j^{-2} \left( \alpha_0 + c_{it,j} - \mathbb{E}(\mu_{it,j} | E_{it}) + \frac{\sqrt{\alpha_1^2 + 4\alpha_2 (f_{it,j} - c_{it,j}) - \alpha_1^2}}{2\alpha_2} \right) \right]^2 \right\} \\
\sum_{it} 1(f_{it,j} - c_{it,j} < -\alpha_0) \left\{ -\frac{1}{2} \ln \left[ (\alpha_1^2 - 4\alpha_2 (f_{it,j} - c_{it,j})) \tau_j^2 (1 - W_j^p (\sigma_0, \tau))^2 \pi \right] \\
- \left[ (1 - W_j^p (\sigma_0, \tau))^{-2} \tau_j^{-2} \left( c_{it,j} - \mathbb{E}(\mu_{it,j} | E_{it}) - \alpha_0 + \frac{\alpha_1 - \sqrt{\alpha_1^2 - 4\alpha_2 (f_{it,j} - c_{it,j})}}{2\alpha_2} \right) \right]^2 \right\}.
\end{align*}

(B.2.1)

Using the log-likelihood functions in Equation (B.2.1), I search over the parameters \(\lambda, \sigma_0\) and \(\tau_j\) which maximizes the expression subject to the restrictions that all three parameters have to be strictly positive.


Figure I: Forecast Strategy from the Absolute Value Objective Function

(i) Forward-looking Measure, $j < J + 1$  
(ii) Static measure or $j = J + 1$

This figure illustrates the equilibrium arising the absolute value objective function discussed in Section 4.2. The horizontal axis is analyst $j$’s earnings prediction $\mu_j$ and the vertical axis $f_j^*$ is the model-implied forecast. Both plots are centered on consensus forecast $c_j$. Plot (i) assumes that peer forecast is a forward-looking measure and that the analyst is not the last to forecast. In Plot (ii), I assume that either peer forecast is a static measure or the analyst is forecasting last out of a total of $J + 1$ analysts.
The plots in this figure illustrate the approximation of forecast strategies resulting from a model with an absolute value objective function, discussed in detail in Section 5.2. The horizontal axis is the analyst’s prediction about earnings while the vertical axis is the model-predicted forecast. Each plot is centered on consensus forecast $c_j$. The region over which $\mu_j < c_j$ is not shown but the strategy is symmetric to strategy over the $\mu_j > c_j$ region. The dotted curve in each plot represents the exact forecast strategies computed using a numerical solver over the analyst’s objective function and the solid curve represents the approximation function. The parameters $\alpha_0, \alpha_1$ and $\alpha_2$ are approximation parameters related to $\lambda, \sigma_0$ and $\tau_j$ in the model.
Table 1: Summary Statistics

This table reports summary statistics about forecast error and the deviation from consensus for a sample constructed from two-year ahead earnings per share forecasts announced within 30 days of annual earnings announcements between 1990 and 2012. Observations with forecast error of greater than five dollars are removed. Forecast error is computed as the difference between the observed forecast and actual earnings. Consensus is computed as the average over all preceding analysts’ forecasts starting from the earnings announcement. In Panel A, observations are pooled based on their order, where each row corresponds to the summary statistics for all forecasts announcing in the j-th order. In Panel B, observations are pooled based on the size of the analyst coverage, where each row corresponds to the summary statistics for all firms with analyst coverage of J.

**Panel A: Summary Statistics by Order**

<table>
<thead>
<tr>
<th>j=</th>
<th>Forecast Error</th>
<th>Deviation from Consensus</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. Dev.</td>
</tr>
<tr>
<td>2</td>
<td>0.198</td>
<td>0.606</td>
</tr>
<tr>
<td>3</td>
<td>0.188</td>
<td>0.617</td>
</tr>
<tr>
<td>4</td>
<td>0.186</td>
<td>0.616</td>
</tr>
<tr>
<td>5</td>
<td>0.172</td>
<td>0.622</td>
</tr>
<tr>
<td>6</td>
<td>0.169</td>
<td>0.626</td>
</tr>
<tr>
<td>7</td>
<td>0.173</td>
<td>0.628</td>
</tr>
<tr>
<td>8</td>
<td>0.171</td>
<td>0.642</td>
</tr>
<tr>
<td>9</td>
<td>0.159</td>
<td>0.644</td>
</tr>
<tr>
<td>10</td>
<td>0.149</td>
<td>0.652</td>
</tr>
</tbody>
</table>

**Panel B: Summary Statistics by Coverage**

<table>
<thead>
<tr>
<th>J=</th>
<th>Forecast Error</th>
<th>Deviation from Consensus</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. Dev.</td>
</tr>
<tr>
<td>2</td>
<td>0.224</td>
<td>0.583</td>
</tr>
<tr>
<td>3</td>
<td>0.207</td>
<td>0.609</td>
</tr>
<tr>
<td>4</td>
<td>0.212</td>
<td>0.607</td>
</tr>
<tr>
<td>5</td>
<td>0.190</td>
<td>0.612</td>
</tr>
<tr>
<td>6</td>
<td>0.162</td>
<td>0.616</td>
</tr>
<tr>
<td>7</td>
<td>0.176</td>
<td>0.613</td>
</tr>
<tr>
<td>8</td>
<td>0.171</td>
<td>0.636</td>
</tr>
<tr>
<td>9</td>
<td>0.168</td>
<td>0.637</td>
</tr>
<tr>
<td>10</td>
<td>0.158</td>
<td>0.637</td>
</tr>
</tbody>
</table>
Table 2: Estimation Based on Strategy from Quadratic Objective Function

This table reports results from estimating a linear model of forecasts on consensus, actual earnings and the previous analyst’s forecast. Panel A (Panel B) reports the estimated weight on actual earnings (consensus forecast), where the weight on the previous analyst’s forecast has been restricted such that all three weights sum to 1. The number reported in the \( j \)-th column and the \( J \)-th row corresponds to the weight specific to forecasts announced \( j \)-th out of \( J \) total analysts. Standard errors are reported in parentheses below the point estimate. Panel C reports the log-likelihood from estimation with no restrictions as well as the inequality and equality restrictions predicted by the model. The sample is comprised of two year-ahead earnings per share forecasts issued within 30 days after earnings announcements between 1990 and 2012 with coverage between 2 and 10 analysts.

| Panel A: Weight on Earnings \( \omega_E \) |
|-----------------|---|---|---|---|---|---|---|---|---|---|
|      | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 |
| \( J \) | \( \omega_E \) | \( \omega_E \) | \( \omega_E \) | \( \omega_E \) | \( \omega_E \) | \( \omega_E \) | \( \omega_E \) | \( \omega_E \) | \( \omega_E \) | \( \omega_E \) |
| 2    | 0.26 |   |   |   |   |   |   |   |   |   |
|      | (0.01) | | | | | | | | | |
| 3    | 0.27 | 0.20 |   |   |   |   |   |   |   |   |
|      | (0.01) | (0.01) | | | | | | | | |
| 4    | 0.23 | 0.14 | 0.11 |   |   |   |   |   |   |   |
|      | (0.01) | (0.01) | (0.01) | | | | | | | |
| 5    | 0.29 | 0.14 | 0.13 | 0.16 |   |   |   |   |   |   |
|      | (0.01) | (0.01) | (0.01) | (0.01) | | | | | | |
| 6    | 0.26 | 0.19 | 0.12 | 0.16 | 0.11 |   |   |   |   |   |
|      | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | | | | | |
| 7    | 0.20 | 0.19 | 0.14 | 0.12 | 0.13 | 0.13 |   |   |   |   |
|      | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | | | | |
| 8    | 0.24 | 0.14 | 0.16 | 0.12 | 0.09 | 0.09 | 0.11 |   |   |   |
|      | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | | | |
| 9    | 0.21 | 0.10 | 0.17 | 0.10 | 0.09 | 0.13 | 0.15 | 0.07 |   |   |
|      | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | | |
| 10   | 0.20 | 0.13 | 0.13 | 0.23 | 0.18 | 0.11 | 0.11 | 0.08 | 0.11 |   |
|      | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | |
Table 2 (Continued): Estimation Based on Strategy from Quadratic Objective Function

Panel B: Weight on Consensus $\omega_c$

<table>
<thead>
<tr>
<th></th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.64</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.91</td>
<td>0.93</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.02)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.84</td>
<td>0.88</td>
<td>0.74</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.84</td>
<td>0.88</td>
<td>0.90</td>
<td>0.85</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.53</td>
<td>0.56</td>
<td>0.76</td>
<td>0.99</td>
<td>0.82</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.02)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1.01</td>
<td>0.89</td>
<td>0.98</td>
<td>0.94</td>
<td>0.81</td>
<td>0.75</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.03)</td>
<td>(0.02)</td>
<td>(0.03)</td>
<td>(0.02)</td>
<td>(0.03)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.57</td>
<td>0.79</td>
<td>0.85</td>
<td>0.99</td>
<td>1.11</td>
<td>0.59</td>
<td>0.69</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.03)</td>
<td>(0.02)</td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.03)</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.80</td>
<td>0.92</td>
<td>1.05</td>
<td>0.68</td>
<td>0.85</td>
<td>1.08</td>
<td>0.86</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.04)</td>
<td>(0.04)</td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.03)</td>
<td>(0.03)</td>
</tr>
</tbody>
</table>

Panel C: Model Fit

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Log-likelihood at unrestricted estimate</td>
<td>-25,025</td>
</tr>
<tr>
<td>Log-likelihood with inequality restrictions</td>
<td>-25,883</td>
</tr>
<tr>
<td>Log-likelihood with equality restrictions</td>
<td>-26,010</td>
</tr>
</tbody>
</table>
Table 3: Estimation Based on Strategy from Absolute Value Objective Function

This table reports results from estimating a non-linear approximation of the strategy resulting from a model based on the Absolute Value Objective Function which is discussed in detail in Section 5.2. In the approximation, the forecast strategy is assumed to follow an error function over the range where the analyst’s earnings prediction and consensus forecast is between \(-\alpha_0\) and \(\alpha_0\) and linear beyond this interval. The parameter \(\alpha_1\) determines the slope of the error function. In Panel A (Panel B), I report the estimated value of \(\alpha_0\) (\(\alpha_1\)). The number reported in the \(j\)-th column and the \(J\)-th row corresponds to the weight specific to forecasts announced \(j\)-th out of \(J\) total analysts. Panel C reports the log-likelihood from estimation with no restrictions as well as the inequality and equality restrictions predicted by the model. The sample is comprised of two year-ahead earnings per share forecasts issued within 30 days after earnings announcements between 1990 and 2012 with coverage between 2 and 10 analysts. Standard errors are reported in parentheses below the point estimate.

<p>| Panel A: Approximation Parameter (\alpha_0) |
|---------------------------------|------|------|------|------|------|------|------|------|------|------|</p>
<table>
<thead>
<tr>
<th>(J)</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.001</td>
<td>0.043</td>
<td>0.001</td>
<td>0.012</td>
<td>0.007</td>
<td>0.009</td>
<td>0.184</td>
<td>0.240</td>
<td>0.341</td>
</tr>
<tr>
<td></td>
<td>(1.61)</td>
<td>(0.69)</td>
<td>(1.45)</td>
<td>(4.32)</td>
<td>(2.70)</td>
<td>(1.52)</td>
<td>(0.18)</td>
<td>(0.12)</td>
<td>(0.09)</td>
</tr>
<tr>
<td></td>
<td>0.004</td>
<td>0.006</td>
<td>0.245</td>
<td>0.001</td>
<td>0.025</td>
<td>0.002</td>
<td>0.002</td>
<td>0.295</td>
<td>0.377</td>
</tr>
<tr>
<td></td>
<td>(1.22)</td>
<td>(0.29)</td>
<td>(7.07)</td>
<td>(1.42)</td>
<td>(1.89)</td>
<td>(1.74)</td>
<td>(2.05)</td>
<td>(0.62)</td>
<td>(0.15)</td>
</tr>
<tr>
<td></td>
<td>0.245</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
<td>0.002</td>
<td>0.000</td>
<td>0.004</td>
<td>0.163</td>
<td>0.213</td>
</tr>
<tr>
<td></td>
<td>(7.07)</td>
<td>(1.22)</td>
<td>(1.42)</td>
<td>(1.34)</td>
<td>(1.89)</td>
<td>(1.74)</td>
<td>(2.05)</td>
<td>(1.66)</td>
<td>(1.22)</td>
</tr>
<tr>
<td></td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.004</td>
<td>0.018</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>(1.45)</td>
<td>(1.22)</td>
<td>(1.34)</td>
<td>(1.33)</td>
<td>(1.66)</td>
<td>(1.33)</td>
<td>(2.00)</td>
<td>(2.05)</td>
<td>(2.22)</td>
</tr>
<tr>
<td></td>
<td>0.002</td>
<td>0.018</td>
<td>0.007</td>
<td>0.015</td>
<td>0.015</td>
<td>0.011</td>
<td>0.163</td>
<td>0.280</td>
<td>0.203</td>
</tr>
<tr>
<td></td>
<td>(1.22)</td>
<td>(2.05)</td>
<td>(2.00)</td>
<td>(2.05)</td>
<td>(2.00)</td>
<td>(2.05)</td>
<td>(1.66)</td>
<td>(2.00)</td>
<td>(1.22)</td>
</tr>
<tr>
<td></td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.185</td>
<td>0.280</td>
<td>0.203</td>
</tr>
<tr>
<td></td>
<td>(1.33)</td>
<td>(2.05)</td>
<td>(2.00)</td>
<td>(2.05)</td>
<td>(2.00)</td>
<td>(2.05)</td>
<td>(1.66)</td>
<td>(2.00)</td>
<td>(1.22)</td>
</tr>
<tr>
<td></td>
<td>0.002</td>
<td>0.002</td>
<td>0.002</td>
<td>0.002</td>
<td>0.002</td>
<td>0.002</td>
<td>0.061</td>
<td>0.041</td>
<td>0.034</td>
</tr>
<tr>
<td></td>
<td>(2.22)</td>
<td>(2.22)</td>
<td>(2.22)</td>
<td>(2.22)</td>
<td>(2.22)</td>
<td>(2.22)</td>
<td>(2.22)</td>
<td>(2.22)</td>
<td>(2.22)</td>
</tr>
<tr>
<td></td>
<td>0.002</td>
<td>0.002</td>
<td>0.002</td>
<td>0.002</td>
<td>0.002</td>
<td>0.002</td>
<td>0.041</td>
<td>0.041</td>
<td>0.041</td>
</tr>
<tr>
<td></td>
<td>(2.22)</td>
<td>(2.22)</td>
<td>(2.22)</td>
<td>(2.22)</td>
<td>(2.22)</td>
<td>(2.22)</td>
<td>(2.22)</td>
<td>(2.22)</td>
<td>(2.22)</td>
</tr>
<tr>
<td></td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.256</td>
<td>0.256</td>
<td>0.256</td>
</tr>
<tr>
<td></td>
<td>(2.22)</td>
<td>(2.22)</td>
<td>(2.22)</td>
<td>(2.22)</td>
<td>(2.22)</td>
<td>(2.22)</td>
<td>(2.22)</td>
<td>(2.22)</td>
<td>(2.22)</td>
</tr>
</tbody>
</table>
Table [Continued]: Estimation Based on Strategy from Absolute Value
Objective Function

**Panel B: Approximation Parameter $\alpha_1$**

<table>
<thead>
<tr>
<th>J=</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.29</td>
<td>(0.10)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.49</td>
<td>0.40</td>
<td>(0.14)</td>
<td>(0.19)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.63</td>
<td>0.49</td>
<td>0.47</td>
<td>(0.10)</td>
<td>(0.13)</td>
<td>(0.13)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.60</td>
<td>0.40</td>
<td>0.45</td>
<td>0.35</td>
<td>(0.14)</td>
<td>(0.13)</td>
<td>(0.13)</td>
<td>(0.12)</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.49</td>
<td>0.46</td>
<td>0.51</td>
<td>0.65</td>
<td>1.24</td>
<td>(0.15)</td>
<td>(0.13)</td>
<td>(0.12)</td>
<td>(0.09)</td>
</tr>
<tr>
<td>7</td>
<td>0.68</td>
<td>0.44</td>
<td>0.49</td>
<td>0.57</td>
<td>0.50</td>
<td>0.60</td>
<td>(0.11)</td>
<td>(0.21)</td>
<td>(0.14)</td>
</tr>
<tr>
<td>8</td>
<td>0.51</td>
<td>0.60</td>
<td>0.46</td>
<td>0.45</td>
<td>0.53</td>
<td>0.42</td>
<td>0.57</td>
<td>(0.14)</td>
<td>(0.09)</td>
</tr>
<tr>
<td>9</td>
<td>0.52</td>
<td>0.69</td>
<td>0.41</td>
<td>0.51</td>
<td>0.64</td>
<td>0.44</td>
<td>0.38</td>
<td>0.41</td>
<td>(0.17)</td>
</tr>
<tr>
<td>10</td>
<td>0.48</td>
<td>0.29</td>
<td>0.48</td>
<td>0.45</td>
<td>0.31</td>
<td>0.52</td>
<td>0.64</td>
<td>0.56</td>
<td>0.45</td>
</tr>
</tbody>
</table>

**Panel C: Model Fit**

<table>
<thead>
<tr>
<th></th>
<th>Log-likelihood at unrestricted estimate</th>
<th>-14,955</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log-likelihood with inequality restrictions</td>
<td>-16,877</td>
<td></td>
</tr>
<tr>
<td>Log-likelihood with equality restrictions</td>
<td>-22,940</td>
<td></td>
</tr>
</tbody>
</table>