Opinion-producing agents: career concerns and exaggeration

Eric Zitzewitz[1]

Stanford Graduate School of Business

June 2001

[1]Address: Stanford Graduate School of Business, Stanford, CA 94305. Email: zitzewitz_eric@gsb.stanford.edu. The author would like to thank Glenn Ellison, Bengt Holmstrom, Dean Karlan, Leigh Linden, Sendhil Mullainathan, Nancy Rose, Jean Tirole, Ezra Zuckerman and seminar participants at Berkeley, Harvard, London Business School, Michigan, MIT, Northwestern, Stanford, Washington University-St. Louis, and Yale for helpful comments and suggestions. Particular thanks are due to my primary advisor, Robert Gibbons. All errors are my own. The financial support of an NSF Graduate Research Fellowship is gratefully acknowledged. I also thank I/B/E/S for providing me with access to their analyst data.

Abstract

This paper models the incentives created by career concerns for opinion-producing agents. We find that career concerns can create an incentive for exaggeration or anti-herding, since high-ability agents will have opinions that are more different from the consensus on average and potential clients will learn more quickly about how different an agent's opinions are from the consensus on average that about whether or not they are exaggerating. The model predicts that agents should exaggerate more when they are under-rated by their clients, when the realizations of the variables they are forecasting are expected to be especially noisy, and when they expect to make fewer future forecasts. We find that these predictions are consistent with the empirical data on equity analyst's earnings forecasts.

1 Introduction

Much of the information in the so-called information economy is not verifyable information as economists normally define it, but is rather opinion. Opinion goods such as forecasts, consulting advice, and product reviews are sold in markets, and the production of these and other types of opinions is also the primary job of many professionals within organizations. Unlike most traditional goods, the quality of information or opinion goods cannot be readily observed prior to purchase. In addition, opinions that are not verifiable can be manipulated by their producer. Reputational or career concerns are thus likely to be especially important to opinion producers.

This paper examines the relationship between the reputational or career concerns of opinion producers and their incentives to engage in a particular type of opinion manipulation, namely exaggerating their differences with the existing consensus. It essentially asks the question: do people exaggerate in order to appear smart? Most opinions can be thought of as forecasts of a random variable that will be realized in the future. We develop a model in which potential clients attempt to learn the ability of forecasters from their track record. In the model, an incentive to exaggerate arises because high-ability agents have access to more private information and thus have unbiased beliefs that are more different from the prior consensus on average. Clients learn more quickly about how different an agent's forecasts are from the consensus on average than about whether or not they are exaggerating, and thus agents can temporarily raise estimates of their ability by exaggerating.

The model also yields cross-sectional predictions about when we should expect agents to exaggerate more. This is potentially useful to consumers of opinions, since they will want to back out expected exaggeration to form unbiased beliefs. One source of variation arises from the fact that the difference in learning speed discussed above is more pronounced when forecast variable realizations are noisier, so agents should exaggerate more under these circumstances. Likewise, agents expecting a shorter future career length should exaggerate more. Agents should also exaggerate more when they have had bad luck in the past (and are consequently under-rated by the market) in order to increase the weighting on future observations of their ability. We find that both the general finding of exaggeration and these cross-sectional predictions are consistent with the empirical evidence when we examine the exaggeration of equity analysts forecasting earnings using the methodology developed in Zitzewitz (2001).

The model of career concerns and exaggeration in this paper draws on the literature on career concerns, reputational concerns for producers of goods of unobservable quality, and herding theory.[1] It is particularly related to three recent papers. Scharfstein and Stein (1990) provide a model of sequential investment decisions in which a herding equilibrium exists. A key assumption they make is that agents do not know their ability and thus low-ability agents (who receive noisier signals) do not know that their signals are noisier and do not know to reduce their reliance on their private information and therefore make reports that deviate more from the consensus. Repeating the consensus is therefore a signal of ability, and a herding equilibrium can result. Avery and Chevalier (1999) extend this equilibrium to allow agents to learn their own ability and find that anti-herding will be an equilibrium rather than herding once agents are expected to have a sufficiently precise knowledge of their own

[1]For general career concerns theory see Fama (1980), Holmstrom (1982), and Gibbons and Murphy (1992). The industrial organizations literature on reputational concerns includes Nelson (1974), Schmalensee (1978), Klein and Le- er (1981), and Shapiro (1983). Herding models include information cascade models (Banerjee, 1992; Bikhchandani, Hirshleifer, and Welch, 1992; Welch, 1999), incentive-concavity models (Holmstrom and Ricart i Costa, 1986; Zweibel, 1995; Chevalier and Ellison, 1997 and 1999; Laster, Bennett and Geoum, 1999), and career-concerns models (Scharfstein and Stein, 1990; Brandenburger and Polak, 1996; Trueman, 1994; Ehrbeck and Waldmann, 1996; Prendergast and Stole, 1996; Avery and Chevalier, 1999; Ottaviani and Sorensen, 2000; Effinger and Polborn, 2000).

ability. In Prendergast and Stole (1996), agents know their own ability perfectly and they anti-herd or exaggerate in order to signal ability. Prendergast and Stole (1996) also allow the set of possible investments to be continuous rather than restricting it to having only two possible values.

This paper is closest in setup to Prendergast and Stole (1996); its main departure is to examine an environment in which there is learning from investment outcomes (or, in the language of this paper, from the realizations of the variables being forecast). This seemingly trival extension yields the cross-sectional predictions about the degree of exaggeration mentioned above that we can test empirically. The extension comes at a cost, however, since modelling the learning from realizations is too complicated to handle in a fully Bayesian way. Rather than assuming that potential clients conduct a Bayesian estimate of an opinion seller's ability, we instead assume that they use an econometric methodology that approximates a maximum-likelihood estimate of ability. The results from this model can therefore be viewed either as an approximation of the results from a Bayesian model or as the results from a behavioral model in which we make the arguably realistic assumption that clients use econometric approximations when faced with a Bayesian problem that is too difficult to them to solve.

The remainder of the paper is organized into two sections. The first develops the model and the cross-sectional predictions. The second section presents evidence that equity analyst's earnings forecasts are consistent with these predictions. A conclusion follows.

2 Career-concerns model

In this section, we model the incentives created by the career concerns of an agent who forecasts a series of random variables. We first present a two-period model in which agents issue forecasts in the first period and receive revenue in the second period proportional to clients' valuation of their forecasts based on their first period performance. We then extend the model to three periods in order to examine how past forecasting performance affects an agent's incentives to exaggerate. From the two-period model we conclude that agents have an incentive to exaggerate and this incentive to exaggerate will be greater when earnings realizations are expected to be noisy or when agents expect to make a limited number of forecasts in the future. From the three-period model, we conclude that agents will also exaggerate more when they have had bad luck in the past and thus are underrated by their clients.

2.1 Two-period model

The model has two periods. In the first period, the agent issues forecasts of J random variables after observing a private signal and a common prior. In the second period, the agent sells early access to their forecasts to clients, receiving revenue proportional to clients' valuation of their forecasts based on their first-period performance. Agents also face an exogenous incentive for forecast accuracy, and thus receive revenue proportional to $\widehat{v} - \lambda \cdot MSE$, where $\widehat{v}$ is the estimate of the value of new information in the agents forecasts and λ is the size of the incentive for accuracy. The role of the incentive for accuracy is to make the language analysts use relevant; we can think of this incentive as the result of clients' costs of translating exaggerated forecasts into unbiased expectations or of industry institutions that measure analysts based on mean squared error.

The timing of the model is as follows:

1. Nature chooses an ability, a, and an accuracy incentive, λ, for the agent from a prior distribution $g(a, \lambda)$. Both parameters are observed by the agent but are unknown to their potential customers.

2. The agent forecasts a series of J random variables A_j. For each A_j, the agent observes a public consensus prior that A_j is distributed $N(C_j, \Sigma^{-1})$, where Σ is the precision of the prior. Since C_j is public, we can think of the agent as forecasting $y_j = A_j - C_j$, i.e. the difference between the actual realization and its consensus prior expectation.

3. In addition, the agent observes an independent private signal $s_j \sim N(A_j, p^{-1})$, where $p = \frac{a\Sigma^2}{(1-a\Sigma)} > 0$ is the precision of the signal. p is an increasing function of a; higher ability agents receive higher precision private signals.

4. Based on the consensus priors and her private signals, the agent issues forecasts using a forecasting rule $\mathbf{x} = \mathbf{x}(\mathbf{s}, \mathbf{C}, p, \lambda)$, where $x_j = F_j - C_j$ is a forecast of y_j. All forecasts are made before any of the earnings variables are realized.

5. Clients with early access to forecasts can make investments in securities with returns that are proportional to y_j.

6. Current and potential clients observe the forecasts $\mathbf{x}$ and realizations $\mathbf{y}$ and estimate the value of the information content of the agent's forecasts using a valuation rule $\widehat{v} = v(\mathbf{x}, \mathbf{y})$.

7. Agents sell early access to their second-period forecasts and receive revenue proportional to $\widehat{v} - \lambda \cdot MSE$. Agents consume and experience linear utility.

2.1.1 Forecast information content and clients valuation of forecasts

After the agent observes the consensus prior and the private signal, her posterior expectation of y_j is

$$E(y_j|C_j, s_j) = (s_j - C_j)\frac{p}{p+\Sigma}, \tag{1}$$

where a higher-precision private signal receives a higher weight. Notice that the variance of the difference between an agent's unbiased beliefs and the prior consensus is increasing in p:

$$Var[E(y_j|s_j, C_j)] = \frac{p}{(p+\Sigma)\Sigma} = a. \tag{2}$$

That is, higher-ability agents have opinions that are more different from the consensus, on average. We will also refer to this variance as the information content of an agent's forecasts, since it is equal to the reduction in mean-squared error in the expection of y_j or A_j due to the agent's private information:[2]

$$Var[E(y_j|s_j, C_j)] = Var(y_j|C_j) - Var[y_j - E(y_j|s_j, C_j)]. \tag{3}$$

A mean-variance client with early access to a forecast investing in a security with returns that are linear in y_j will invest to maximize

$$\max_{I_j} I_j \cdot E(y_j|x_j) - I_j^2 \cdot \frac{r}{2} \cdot Var(y_j|x_j), \tag{4}$$

where I_j is the client's exposure to y_j and r is the coefficient of absolute risk aversion. The optimal investment is:

$$I_j^* = \frac{E(y_j|x_j)}{r \cdot Var(y_j|x_j)} \tag{5}$$

and such an investment has an ex-ante certainty-equivalent value of

$$CE = \frac{[E(y_j|x_j)]^2}{2r \cdot Var(y_j|x_j)}. \tag{6}$$

The value of early access to an agent's forecasts will be proportional to the variance of $E(y_j|x_j)$. If an agent's forecasts are fully revealing of her signal, this will be the same as $Var[E(y_j|s_j, C_j)] = a$, otherwise it will be a less a discount for uncertainty regarding an agent's forecasting strategy. For simplicity, we will assume that the

[2] This is true since by the law of iterated expectations, $y_j - E(y_j|s_j, C_j)$ must be uncorrelated with $E(y_j|s_j, C_j) - E(y_j|C_j)$.

information lost in communicating the signal is small in the long run and thus that clients are interested in estimating a as the long-run value of a client's forecasts, so $\widehat{v} = \widehat{a}$.[3]

2.1.2 Solution

We will look for a consistent-exaggeration forecasting equilibrium in which agents use the forecasting rule: $x_j = b(a, \lambda) \cdot E(y_j|s_j, C_j)$ with some constant exaggeration factor b that can depend on a or λ. This forecasting rule implies that agents exaggerate their differences with the consensus when $b > 1$, herd when $b < 1$, and report their expectation when $b = 1$.

The standard solution approach would be to solve for a Bayesian equilibrium in which the analyst chooses her forecasts to maximize expected utility, given the clients' valuation rule, and the valuation rule produces an unbiased and efficient estimate, given the analyst's forecasting rule. Unfortunately, the standard Bayesian estimate of $v(\mathbf{x}, \mathbf{y}) = E[a|\mathbf{x}, \mathbf{y}, g(), b()]$ is very intractable and conjugate prior distribution families that improve tractability do not exist. Instead of assuming that clients make a such a difficult calculation, we will instead assume that they use an econometric estimation approach that is consistent but not necessarily efficient, with the inefficiency coming from the incorporation of the information from the prior distribution in an approximate rather than a fully Bayesian way. We will describe an equlibrium in which clients do econometric estimation that anticipates consistent exaggeration and then show that consistent exaggeration is in fact optimal for the analysts.[4] This ex-

[3] If we relaxed this assumption, agents would face an additional incentive for limiting the uncertainty regarding their exaggeration strategy, since uncertainty about exaggeration creates a gap between $a = Var[E(y_j|s_j, C_j)]$ and $v = Var[E(y_j|x_j)]$. Relative to the solution described below, the agent would choose an exaggeration factor slightly closer to their clients' prior expectation of exaggeration.

[4] Actually, a consistent-exaggeration equilibrium will be the only equilibrium whenever $\lambda > 0$. To

ercise can be viewed as the derivation of an Bayesian equilibrium in which clients are constrained by bounded rationality and thus use an unbiased and tractable but less efficient estimation approach. Alternatively, it can be viewed as merely an analysis of the incentives created for analysts if clients estimate their ability using a particular econometric procedure.

We proceed by specifying the clients' estimator of analyst ability, solving the analyst's problem given the clients' valuation rule, and verifying that consistent-exaggeration forecasting is an equilibrium.

Clients' problem If analysts follow the consistent-exaggeration forecasting rule described above, they will issue forecasts such that:

$$
\begin{aligned}
E(y_j|s_j, C_j) &= \frac{p}{p+\Sigma}(s_j - C_j) \sim N(0, a) \\
x_j &= b \cdot \frac{p}{p+\Sigma}(s_j - C_j) \sim N(0, b^2 a) \\
y_j &= b^{-1} x_j + \varepsilon_j \\
\varepsilon_j &= y_j - E(y_j|s_j, C_j).
\end{aligned}
$$

As a result, x_j and y_j are distributed joint normally:

$$
\begin{bmatrix} y_j \\ x_j \end{bmatrix} \sim N(0, \begin{bmatrix} a + V + (\Sigma + p)^{-1} & ba \\ ba & b^2 a \end{bmatrix}).
$$

Note that a is the information content of the analyst's forecasts $Var[E(y_j|x_j)]$; clients are therefore trying to estimate a.

see this consider an alternative equlibrium where clients expect an analyst to exaggerate by $b \cdot f(j)$, where f varies predictably from observation to observation. In this case, the first-order condition in (12) below will be $b = [E(\widehat{\beta}_{CL}) + f^{-1}\lambda][E(\widehat{\beta}_{CL})^2 - \gamma\beta_0^2 + \lambda]^{-1}$, so the analyst will choose a lower b when f is high and vice versa, i.e. will choose a strategy closer to consistent exaggeration than they are expected to. It follows that a consistent exaggeration is the only equilibrium (when clients deviate from Bayesian estimation as described above).

Given the regression-like setup, it will be convenient to discuss $\beta = b^{-1}$ as the change in the expectation of y for a given change in x. A natural set of estimators for $\beta = b^{-1}$ and a are the classical statistical estimators that are used to estimate exaggeration and forecast information content in Zitzewitz (2001):

$$\widehat{\beta}_{CL} = \frac{\sum_{j=1}^{J} x_j y_j}{\sum_{j=1}^{J} x_j^2} \tag{7}$$

$$\widehat{a}_{CL} = \widehat{\beta}_{CL}^2 \cdot \widehat{Var}(x_j) = \frac{(\sum_{j=1}^{J} x_j y_j)^2}{(J-1)\sum_{j=1}^{J} x_j^2}. \tag{8}$$

These estimators are consistent and unbiased, but they are inefficient if clients have prior information about a or $b(a,\lambda)$. A client can improve the efficiency of an estimate of a by averaging the observed $\widehat{a}_{CL}$ with the mean of her prior distribution. In addition, since the estimate $\widehat{\beta}_{CL}$ is likely to be noisier than the estimate $\widehat{Var}(x_j)$, especially when V is high, clients can improve on the efficiency of their estimate by constructing their $\widehat{a}$ with an average of the observed $\widehat{\beta}_{CL}$ and β_0, client's prior expectation of b^{-1} given the distribution $g(a,\lambda)$ and the function $b(a,\lambda)$. We therefore assume that clients estimate analysts' ability as:

$$\widehat{a}_P = \frac{(\widehat{\beta}_{CL}^2 + \gamma \cdot \beta_0^2) \cdot \widehat{Var}(x_j) + \delta \cdot a_0}{1+\gamma+\delta}, \tag{9}$$

where δ is the weight given the prior expectation of ability and $\gamma > 0$ is the weight placed on β_0.

The γ term in (9) captures an important feature of any optimal estimator, namely that prior information about exaggeration should be relied on when estimates of exaggeration are noisy. Equation (9) is similar in structure to a maximum-likelihood estimator of ability; an analysis of the maximum likelihood estimator in Appendix A yields some intuitive predictions about the determinants of γ. The weight γ placed on β_0 should be higher when the estimate $\widehat{\beta}$ is noisier i.e. when earnings realizations are noisier (and V is higher) or when the number of observations J is low.[5]

[5] We can also think of γ as capturing the potential for exaggeration to signal high or low ability.

Analyst's problem Analysts choose their x_j to maximize their expectation of $\widehat{a}_P - \lambda \cdot MSE$, their clients' estimation of their forecast value plus their incentive for absolute accuracy. If clients use the estimation approach outlined above, the analyst's problem is:

$$\max_{\mathbf{x}} E[\frac{(\sum_{j=1}^{J} x_j y_j)^2}{(J-1)\sum_{j=1}^{J} x_j^2}] + \gamma\beta_0^2 \frac{\sum_{j=1}^{J} x_j^2}{(J-1)} - \lambda E[\frac{\sum_{j=1}^{J}(y_j - x_j)^2}{(J-1)}], \tag{10}$$

where the expectations are the analyst's before $\mathbf{y}$ is realized. The first-order condition for each x_j is:[6]

$$x_j = \frac{E(y_j \widehat{\beta}_{CL}) + E(y_j)\cdot\lambda}{E(\widehat{\beta}_{CL}^2) - \gamma\beta_0^2 + \lambda} = E(y_j)\cdot\frac{E(\widehat{\beta}_{CL}) + \lambda}{E(\widehat{\beta}_{CL})^2 - \gamma\beta_0^2 + \lambda}. \tag{11}$$

So a consistent-exaggeration strategy of the type assumed above is in fact optimal. The analyst chooses b such that:

$$\beta = b^{-1} = \frac{E(\widehat{\beta}_{CL})^2 - \gamma\beta_0^2 + \lambda}{\widehat{\beta}_{CL} + \lambda}. \tag{12}$$

From (7) above we can see that rational expectations on the part of the analyst imply that $E(\widehat{\beta}_{CL}) = \beta$. This condition together with (8) implies the following relationships between accuracy incentives (λ), the weight placed on β_0 (γ), and exaggeration (β):

- When $\gamma = 0$ (no prior information about β) and $\lambda = 0$ (no incentive for accuracy), any value of β is possible. This is a cheap talk result: if there is no incentive for absolute accuracy, any language is as good as the next so long as the clients do not have prior beliefs about β.

When the distribution $g(a, \lambda)$ and the function $b(a, \lambda)$ are such that the priors on a and $b(a, \lambda)$ are positively (negatively) correlated, then exaggeration signals low (high) ability. Clients can account for this in their estimation by lowering (raising) γ relative to its value when the priors on a and $b(a, \lambda)$ are uncorrelated.

[6] Although it might appear that this simplification was made using the incorrect assumption that $E(y_j\widehat{\beta}_{CL}) = E(y_j)\cdot\widehat{\beta}_{CL}$ and $E(\widehat{\beta}_{CL}^2) = E(\widehat{\beta}_{CL})^2$, actually, the two cross terms exactly cancel.

- When $\gamma = 0$ and $\lambda > 0$, $\beta = 1$. Adding even a small incentive for absolute accuracy to the cheap talk situation makes unbiased forecasting optimal.

- When $\gamma > 0$ and $\lambda \leq \gamma\beta_0^2$, $\beta = 0$ and agents exaggerate by an infinite factor. With no or a limited incentive for absolute accuracy and with clients placing some weight on their prior belief, analysts can always increase estimates of their ability by exaggerating more. Analysts essentially report a binary forecast (i.e., above or below the consensus) and cannot credibly communicate the strength of their beliefs.[7]

- When $\gamma > 0$ and $\lambda > \gamma\beta_0^2$, $\beta = \lambda^{-1}(\lambda - \gamma\beta_0^2) < 1$. Some exaggeration occurs, but it is limited to a finite amount by the incentive for absolute accuracy. As this incentive increases, the amount of exaggeration decreases. Likewise, as J decreases or random variable realizations become noisier and thus clients rely more on their prior beliefs about β_0, exaggeration increases.[8]

2.2 Three-period model

In the two-period model above, the analyst makes all J forecasts before seeing any of the realizations. In this section, we analyze how an analyst's past forecasting performance affects her future exaggeration. We extend the model by assuming that agents observe the realizations of the first J variables and then forecast a second group

[7] This result is similar to the infinite exaggeration result in Ottaviani and Sorensen (2000).

[8] Notice that, given the estimation approach assumed above, $b(a, \lambda)$ is a function of only λ. This implies that when clients' prior beliefs $g(a, \lambda)$ imply that a and λ are uncorrelated, then exaggeration will signal neither high or low ability. If instead clients' believe that high-ability agents face more (less) exogenous incentives to exaggerate, then exaggeration will signal high (low) ability. When ability (a) and exogenous incentives to exaggerate (λ) are more positively correlated, exaggeration signals high ability and γ and the equilibrium amount of exaggeration increase.

of K variables. Clients expect consistent exaggeration within a group of random variables, but not necessarily across groups.

Clients estimate ability as before, except that they allow for different exaggeration in the two sets of observations:

$$\begin{aligned}\widehat{a}_P &= (1+\gamma+\delta)^{-1}\frac{(\sum_{j=1}^{J} wx_j y_j + \sum_{j=J+1}^{J+K} x_j y_j)^2}{\sum_{j=1}^{J} w^2 x_j^2 + \sum_{j=J+1}^{J+K} x_j^2} \\ &\quad + \frac{\gamma\beta_0^2}{1+\gamma+\delta}\frac{\sum_{j=1}^{J} x_j^2 + \sum_{j=J+1}^{J+K} x_j^2}{(J+K-1)} + \frac{\delta}{1+\gamma+\delta}\cdot a_0 \\ w &= \frac{\widehat{\beta}_J}{E(\beta_K|\widehat{\beta}_J)}\end{aligned}$$

where β_J and β_K refer to the β for the first J and the second K random variables, respectively. The weight w is the ratio between the observed exaggeration in the first set and the clients' expectation of exaggeration in the second set of random variables, conditional on the observed $\widehat{\beta}_J$. The first order condition for the agent when forecasting the second set of variables reduces to:

$$\beta_K = \frac{wE(\widehat{\beta}_{JK}|\widehat{\beta}_J)^2 - \gamma\beta_0^2 + \lambda}{wE(\widehat{\beta}_{JK}|\widehat{\beta}_J) + \lambda}, \tag{13}$$

where $\widehat{\beta}_{JK}$ is the exaggeration factor estimated across both sets of observations, which will be an average of $w^{-1}\widehat{\beta}_J$ and β_K.

Proposition 1 *Equation (13) implies that agents will choose a β_K that is between the one-period optimal $\beta_1 = \lambda^{-1}(\lambda - \gamma\beta_0^2)$ and their client's expectation $E(\beta_K|\widehat{\beta}_J)$.*

Proof. In Appendix B ■

Agents who have had bad luck in the past and realized a lower $\widehat{\beta}_J$ than the β_J they intended will choose a lower β_K. This can be interpreted as the agents who have had bad luck in the past and are thus under-rated will exaggerate more in order to increase the relative weight of the later observations.[9]

[9] This prediction that under-rated agents should exaggerate more is also made by a different model in Graham (1999).

3 Testing predictions of the model

The model in section 2 has three cross-sectional predictions about when we should expect more exaggeration. Agents should exaggerate more when they are underrated by their clients, when earnings realizations are expected to be noisy, and when they expect to make a limited number of future forecasts. In this section of the paper we test these predictions using the I/B/E/S analyst earnings forecast dataset and the methodology for measuring exaggeration outlined in Zitzewitz (2001).[10] Specifically, we estimate the average exaggeration coefficient for a specific group of forecasts using the regression

$$A - C = \alpha + \beta(F - C) + \varepsilon \tag{14}$$

as in the classical estimator described in section 2.1.2, where A is the I/B/E/S actual earnings for a given firm-quarter combination, F is a forecast of earnings, and C is an econometric expectation of earnings based on prior forecasts for that firm-quarter.[11] We test the predictions for how exaggeration should vary with a specific variable by interacting the right-hand side of (14) with the variable of interest.

As Zitzewitz (2001) shows, the regression in (14) produces an unbiased estimate of the inverse of the exaggeration factor, $\beta = b^{-1}$, because the error term is the analyst's

[10] As Zitzewitz (2001) argues, inferring exaggeration or herding from forecast dispersion, a popular methodology (e.g., Hong, Kubik, and Solomon, 2000), is problematic in that it does not control for forecast information content. In other words, a forecaster can forecast close to the consensus for two reasons: 1) if she is herding, but also 2) if she does not have very much independent private information. Since the amount of independent private information many vary with our cross-sectional variables, we use the methodology in Zitzewitz (2001) which controls for it.

[11] The exaggeration measurement methodology, including the methodology for measuring C, is described in more detail in Zitzewitz (2001). A non-econometric methodology for estimating C (such as using the mean of all outstanding or the most recent forecasts) can be substituted without materially affecting the results. All earnings variables are normalized by the share price.

expectational error at time of forecasting, $\varepsilon = A - E(A|s, C)$, and expectational errors must be mean zero with respect to all variables known at time of forecasting, including $F - C$. In order for interaction versions of this regression to be valid, the analyst's expectational error must be mean zero with respect to the interaction variable as well. This must be true for all variables that are known at time of forecasting, but for variables that incorporate the econometricians knowledge of the future, we will need to verify that the orthogonality condition still holds.

3.1 Under-rated analysts

Since it is impossible to directly observe which analysts have true ability that is higher than their measured ability, we are forced to use our knowledge of the future to help identify under-rated analysts. In particular, we divide analysts with a given past forecast information content into those whose performance eventually rises and those whose performance falls and assume that, on average, the analysts whose performance rises were under-rated in the past.

Specifically, we rank analysts with at least 50 past forecasts based on two variables: their past forecast information content from observation 1 to $j-1$ and the difference between their past information content and their information content from observation $j+1$ to $j+50$ (or fewer if the analyst leaves the sample). We then interact these rankings with the right-hand side of equation (14).

Table 1 presents the results of such an interaction regression. The results suggest that under-rated analysts exaggerate more, whether or not past performance is controlled for. The results also do not change if we control for the analyst's career length or the size of their brokerage, both of which have significant positive effects on β.

In constructing this test, we took two steps to avoid violating the orthogonality condition discussed above. First, we used different observations to measure performance improvement (observations $j+1$ to $j+50$) and exaggeration (observation j). This is important since if an analyst "gets lucky" and gets surprised in the direction of their deviation from the consensus (i.e., $A-E$ and $F-C$ positively correlated), β will be overestimated, exaggeration will be underestimated, and analyst performance will be overestimated. Using different observations avoids this potential problem. Second, we used $\widehat{a}_{CL} = \beta^2 Var(x)$ as our performance measure, a measure that is robust to exaggeration, so even if exaggeration in observation j were correlated with exaggeration in observations $j+1$ to $j+50$ or 1 to $j-1$, this would not create a correlation with our measure of the change in performance.

3.2 Expected earnings uncertainty

We test for whether analysts exaggerate more when earnings are uncertain using a two-step process (Table 2). In the first step, we predict the average absolute earnings surprise (actual less consensus) for a particular firm-quarter based on market cap, the standard deviation of prior outstanding forecasts, and the prior average absolute earnings surprise for the firm in question. We hypothesize and find that average absolute earnings surprise is higher for small-cap firms, when past forecasts are dispersed, and for firms for which average earnings surprise has been large in the past. In the second step, we test the effect of expected earnings surprise on β using an interaction regression, finding that there is significantly more exaggeration when predicted absolute earnings surprise is higher. Notice that in this analysis all of the interaction variables are known at time of forecasting; thus the orthogonality condition should be satisfied.

3.3 Expected career length

The prediction that analysts should exaggerate less when they expect to make more future forecasts is more difficult to test. We can use the actual number of future forecasts as our interaction variable, and when we do this, we find less exaggeration by analysts who make more future forecasts (Table 3, Panel A). A problem with this analysis, however, is that analysts who have good luck should both survive longer and have measured exaggeration that is less than what they intended.

An alternative approach is to use variables that are known at time of forecasting that predict an analyst's longevity. Probit regressions that predict an analyst's leaving the sample and not returning for at least 2 years after a given forecast find longer survival is expected for analysts who have made a large number of past forecasts, analysts who work at larger (usually the more prestigous) brokerages, and analysts who have had better forecast accuracy in the past (Table 3, Panel B).[12]

In Table 1, we found that analysts with more forecasting experience and analysts at larger brokerage firms exaggerated less. The probit regressions suggest that one potential explanation for this result is that these analysts have longer expected careers, and the optimal exaggeration rate for these analysts is lower. Alternative explanations exist, however. Analysts may become less overconfident in their own information or better calibrated with experience. In the model we assumed that analysts' utility is linear in the market valuation of their forecasts; if the concavity of analyst's incentives varies with brokerage size or career length, this may also explain the results. Inexperienced analysts may face greater outside options and thus more convex incentives (i.e., they can gamble and then leave if it does work out), and this may explain their greater exaggeration. Analysts at larger firms may be given more concave incentives by their firm to reduce exaggeration, such as a risk of getting fired

[12] Forecast information content, in turn, does not appear to play a role in predicting exits from the I/B/E/S sample. One potential explanation for this result is that analysts can leave the I/B/E/S sample for either good reasons (moving to lucrative proprietary researcg positions) or bad reasons (getting fired). We do find that for analysts who stay in the profession, forecast information content helps explain which analysts are ranked highly by Institutional Investor (Zitzewitz, 2001, Table 7).

for deviating from the consensus and being wrong that is not fully compensated by the reward for deviating from the consensus and being right. A model in which firms have a collective reputation for exaggeration might predict this, since larger firms will make more forecasts in the future than smaller firms and thus would prefer that their analysts exaggerate less.

In summary, the empirical evidence that is available is consistent with the prediction that analysts who expect to make more future forecasts should exaggerate less, but alternative explanations for the results exist.

4 Conclusion

The evidence presented in Zitzewitz (2001) suggests that there are persistent differences in analyst's exaggeration factors and forecast information content and that the best predictor of the future value of an analyst's forecasts is the value of her past forecasts. This suggests that potential clients should use an analyst's track record to determine how much to pay her. This paper investigates the incentives for exaggeration created by clients attempting to learn ability from forecasting record in a financial market environment where forecasts are valuable for their new information content.

We find that career concerns can create an incentive for agents to exaggerate, or overweight their private information. This incentive exists because high-ability analysts have viewpoints that are more different from the consensus on average and since potential clients learn more quickly about an analyst's average difference with the consensus than about whether she is exaggerating. The equilibrium exaggeration rate is finite so long as there is a sufficiently large external incentive for absolute forecast accuracy. The model also predicts that agents should exaggerate more when earnings are expected to be noisy, when they expect to make a limited number of future forecasts, or when they are under-rated by the market, and we find that these predictions are consistent with the equity analyst forecast data.

Although the evidence in the paper is for equity analysts, the issues examined in this paper potentially apply to other opinion-producing agents. A large number of agents produce opinions that can be thought of as forecasts of random variables. Especially when the actions taken based on the opinions are strategic substitutes, the value of privileged access to an opinion depends on its information content relative to the consensus. Whenever the realized values of random variables are noisy, agents will learn more quickly about the average difference between an agent's opinion and the consensus that they will about whether the agent is exaggerating, an the agent will be able to raise estimates of her ability by exaggerating. This incentive to exaggerate will be greater when the realization of the random variable is expected to be more noisy, which makes exaggeration harder to detect, when the agent expects to leave the profession soon, or when the agent perceives that she is under-rated by the market. These predictions, together with the empirical support for them in the analyst data, are potentially useful for consumers attempting to account for exaggerate in their interpretation of opinions or for firms attempting to reduce exaggeration in the incentives they design for opinion-producers.

A Maximum likelihood estimation

In this appendix we assume that rather than calculating expectations for a and b, clients calculate maximum likelihood estimates. We can think about combining the prior distribution $g(a, \lambda)$ and the function $b(a, \lambda)$ into a prior on a and $\beta = b^{-1}$, which we will call $f(a, \beta)$. For tractability, we will also assume that the prior distribution $g(a, \lambda)$ is such that the distribution $f(a, \beta)$ is concave in logs for both variables, i.e. $\frac{d^2 \ln f(a,\beta)}{da^2} \leq 0$ and $\frac{d^2 \ln f(a,\beta)}{d\beta^2} \leq 0 \ \forall x$, a condition that is satisfied by the normal and chi-squared distributions, for example. We also assume that $V = Var(y_j|x_j)$ is known.

Given these assumptions, the log likelihood function is:

$$\ln L(x, y, a, \beta) = \sum_{j=1}^{J} \ln \phi(\frac{\beta x_j}{a^{1/2}}) + \sum_{j=1}^{J} \ln \phi(\frac{y_j - x_j\beta}{V^{1/2}}) + \ln[f(a, \beta)],$$

where $\phi(\cdot)$ is the standard normal p.d.f. The maximum likelihood estimators of a and b satisfy the conditions:

$$\begin{aligned} \widehat{a}_{MLE} &= \widehat{\beta}^2_{MLE} \frac{\sum_{j=1}^{J} x_j^2}{J} + \frac{f_a(\widehat{a}_{MLE}, \widehat{\beta}_{MLE})}{f(\widehat{a}_{MLE}, \widehat{\beta}_{MLE})} \cdot \frac{\widehat{a}^{3/2}_{MLE}}{J} \\ \widehat{\beta}_{MLE} &= (1 + \frac{V}{\widehat{a}_{MLE}})^{-1}[\frac{\sum_{j=1}^{J} x_j y_j}{\sum_{j=1}^{J} x_j^2} + \frac{f_\beta(\widehat{a}_{MLE}, \widehat{\beta}_{MLE})}{f(\widehat{a}_{MLE}, \widehat{\beta}_{MLE})} \cdot \frac{V}{\sum_{j=1}^{J} x_j^2}]. \end{aligned}$$

We have assumed that $\frac{f_a}{f}$ is monotonically decreasing in a, so the second term in the expression for a will be positive (negative) when a is less (greater) than the mode of the prior distribution. The maximum likelihood estimate of a will thus be between $\widehat{\beta}^2_{MLE}\widehat{Var}(x_j)$ and the mode of the prior. The first factor in the maximum likelihood estimate of β induces a bias toward zero that is a artifact of our taking a maximum likelihood approach to estimation. Ignoring this factor, the maximum likelihood estimate of β will be between $\widehat{\beta}_{OLS}$ and the mode of the prior, with more weight being placed on the prior when V is large.

Thus the MLE is similar in structure to the estimating approach assumed in (9):

$$\widehat{a}_P = \frac{(\widehat{\beta}^2_{CL} + \gamma \cdot \beta_0^2) \cdot \widehat{Var}(x_j) + \delta \cdot a_0}{1 + \gamma + \delta}.$$

The estimate of β is some weighted average of the observed β_{CL} and the mean of the prior distribution β_0. Ability is estimated in turn as some weighted average of $\widehat{\beta}^2_{MLE}\widehat{Var}(x_j)$ and the mean of the prior distribution. The weight placed on the prior belief about exaggeration, γ, is increasing in V and decreasing in J, i.e. it is higher when realizations are noisy or when the number of observations is small.

B Proof of Proposition 1

We know that $E(\widehat{\beta}_{JK}|\widehat{\beta}_J)$ is between $w^{-1}\widehat{\beta}_J$ and β_K. Since $w^{-1}\widehat{\beta}_J = E(\beta_K|\widehat{\beta}_J)$, this implies that β_K will be above $E(\widehat{\beta}_{JK}|\widehat{\beta}_J)$ when it is above client's expectation based on the observed $\widehat{\beta}_J$. If we define $\triangle\beta = \beta_K - E(\widehat{\beta}_{JK}|\widehat{\beta}_J)$, we can rewrite the first order condition as

$$
\begin{aligned}
E(\widehat{\beta}_{JK}|\widehat{\beta}_J) + \triangle\beta &= \frac{wE(\widehat{\beta}_{JK}|\widehat{\beta}_J)^2 - \gamma\beta_0^2 + \lambda}{wE(\widehat{\beta}_{JK}|\widehat{\beta}_J) + \lambda} \\
E(\widehat{\beta}_{JK}|\widehat{\beta}_J) &= \frac{\lambda(1 - \triangle\beta) - \gamma\beta_0^2}{\lambda + w \cdot \triangle\beta} \\
\beta_K &= E(\widehat{\beta}_{JK}|\widehat{\beta}_J) + \triangle\beta = \frac{\lambda - \gamma\beta_0^2 + w(\triangle\beta)^2}{\lambda + w \cdot \triangle\beta} \\
\beta_K &= \beta_1 - \frac{wE(\widehat{\beta}_{JK}|\widehat{\beta}_J) \cdot \triangle\beta}{\lambda}.
\end{aligned}
$$

If we define $\beta_1 = \lambda^{-1}(\lambda - \gamma\beta_0^2)$ to be the optimal one-period beta, we can multiply both sides of the above expression by $\lambda^{-1}(\lambda + w \cdot \triangle\beta)$to get:

$$
\beta_K = \beta_1 - \frac{wE(\widehat{\beta}_{JK}|\widehat{\beta}_J) \cdot \triangle\beta}{\lambda}
$$

This implies that β_K is less than β_1 if and only if it is greater than $E(\widehat{\beta}_{JK}|\widehat{\beta}_J)$ and $E(\beta_K|\widehat{\beta}_J)$.

References

Avery, Christopher N. and Judith A. Chevalier, "Herding Over the Career," *Economics Letters*, 1999, *63*, 327–333.

Banerjee, Abhijit, "A Simple Model of Herd Behavior," *Quarterly Journal of Economics*, 1992, *107*, 797–817.

Bikchandani, Sushil, David Hirshleifer, and Ivo Welch, "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades," *Journal of Political Economy*, 1992, *100*, 992–1026.

Brandenburger, Adam and Ben Polak, "When Managers Cover Their Posteriors: Making Decisions the Market Wants to See," *RAND Journal of Economics*, 1996, *27*, 523–541.

Chevalier, Judith A. and Glenn D. Ellison, "Risk Taking by Mutual Funds as a Response to Incentives," *Journal of Political Economy*, 1997, *105*, 1167–1200.

___ and ___ , "Career Concerns of Mutual Fund Managers," *Quarterly Journal of Economics*, 1999, *114*, 389–432.

Cooper, Rick A., Theodore E. Day, and Craig M. Lewis, "Following the Leader: A Study of Individual Analysts Earnings Forecasts," 1999. Vanderbilt University Mimeo.

Effinger, Matthias R. and Mattias K. Polborn, "Herding and Anti-herding: a model of reputational differentiation," 2000. University of Munich Mimeo.

Ehrbeck, T. and R. Waldmann, "Why Are Professional Forecasts Biased? Agency Versus Behavioral Explanations," *Quarterly Journal of Economics*, 1996, *111*, 21–40.

Fama, Eugene, "Agency Problems and the Theory of the Firm," *Journal of Political Economy*, 1980, *88*, 288–307.

Gibbons, Robert and Kevin J. Murphy, "Optimal Incentive Contracts in the Presence of Career Concerns: Theory and Evidence," *Journal of Political Economy*, 1992, *100*, 468–505.

Graham, J. R., "Herding among Investment Newsletters," *Journal of Finance*, 1999, *54*, 237–268.

Holmstrom, Bengt, "Managerial Incentive Problems: A Dynamic Perspective," *Review of Economic Studies*, 1999, *66*, 169–182.

——— **and Joan Ricart i Costa**, "Managerial Incentives and Capital Management," *Quarterly Journal of Economics*, 1986, *101*, 835–860.

Hong, Harrison, Jeffrey D. Kubik, and Amit Solomon, "Security Analysts' Career Concerns and Herding of Earnings Forecasts," *RAND Journal of Economics*, 2000, *31*, 121–144.

Klein, Benjamin and Keith B. Le- er, "The Role of Market Forces in Assuring Contractual Performance," *Journal of Political Economy*, 1981, *89*, 615–641.

Kutsoati, E. and D. Bernhardt, "Can Relative Performance Evaluation Explain Analysts' Forecasts of Earnings," 1999. Tufts University Mimeo.

Laster, David, Paul Bennett, and In Sun Geoum, "Rational Bias in Macroeconomic Forecasts," *Quarterly Journal of Economics*, 1999, *114*, 293–318.

Nelson, Philip, "Advertising as Information," *Journal of Political Economy*, 1974, *82*, 729–754.

Ottaviani, Marco and Peter Sorensen, "Professional Advice," 2000. University College London Mimeo.

Prendergast, Canice and Lars Stole, "Impetuous Youngsters and Jaded Old-Timers: Acquiring a Reputation for Learning," *Journal of Political Economy*, 1996, *104*, 1105–1134.

Scharfstein, David and Jeremy Stein, "Herd Behavior and Investment," *American Economic Review*, 1990, *80*, 465–479.

Schmalensee, Richard, "A Model of Advertising and Product Quality," *Journal of Political Economy*, 1978, *86*, 485–503.

Shapiro, Carl, "Optimal Pricing of Experience Goods," *Bell Journal of Economics*, 1983, *14*, 497–507.

Trueman, Brett, "Analyst Forecasts and Herding Behavior," *Review of Financial Studies*, 1994, *7*, 97–124.

Welch, Ivo, "Herding Among Security Analysts," 1999. UCLA Mimeo.

Zitzewitz, Eric, "Measuring herding and exaggeration by equity analysts," 2001. MIT Mimeo.

Zweibel, J., "Corporate Conservativism and Relative Performance Compensation," *Journal of Political Economy*, 1992, *103*, 1–25.

Table 1. Exaggeration by under or over-rated analysts
Dependent variable: (ACT - CONS)

This table reports interaction coefficients from a version of equation (14) with the variables listed interacted with the right-hand side. Regressions also include (FOR - CONS) and the interaction variables. Regressions include forecasts made by analysts with at least 50 past forecasts in the 1993-99. The UNDER and INFO variables are rankings, scaled to 0 to 1, of the analyst based on past performance and performance change over the next 50 forecasts, respectively.

		Interactions with (FOR - CONS)							
		UNDER		INFO		FNUM/100		LN(BROKSIZE)	
Spec.	Obs.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
1	329,401	-0.458	0.141						
2	329,401	-0.503	0.133	-0.156	0.120				
3	329,401	-0.503	0.130	-0.140	0.117	0.048	0.016		
4	329,401	-0.519	0.133	-0.198	0.112	0.045	0.016	0.092	0.035
5	329,401			-0.029	0.136	0.046	0.018	0.084	0.038
6	329,401					0.046	0.018	0.082	0.041
7	329,401			-0.046	0.141			0.087	0.037
8	329,401							0.084	0.040

Variable defintions

ACT	Actual earnings per share
FOR	Analyst's forecast of earnings per share
CONS	Prior expectation of earnings per share, as estimated in Table 3
UNDER	Ranking based on change in analyst forecast value over the next 50 forecasts, scaled 0 to 1.
INFO	Ranking of analyst's historical forecast value, scaled 0 to 1.
FNUM	Forecast number in analyst's career
BROKSIZE	Number of analysts at analyst's brokerage

Notes:
1. Standard errors are heteroskedasticity robust and adjusted for clustering within firm-quarter combinations.

Table 2. Exaggeration and uncertainty

The effect of expected earnings uncertainty is examined by predicting the absolute earnings surprise for a firm-quarter combination based on its market cap, the standard deviation of price-normalized forecasts, and past earnings surprise for the firm, and average earnings surprise in the prior 90 days. We then interact predicted earnings surprise with the right-hand side of (2) to measure the effect on exaggeration. The negative interaction coefficient reported implies more exaggeration when expected earnings surprise is high.

	Coeff.	S.E.
First stage regression		
Dependent variable: Abs(ACT - CONS)		
Indepdent variables:		
Ln(Market Cap)	-0.0184	0.0009
Ln(SD-Price ratio)	0.0092	0.0004
Avg past abs(ACT - CONS) for firm	0.710	0.015
Avg abs(ACT - CONS) in quarter	0.136	0.007
Second stage		
Interaction coefficient from exaggeration regression		
Predicted abs. forecast error	-0.059	0.022

Notes:

1. Standard errors are heteroskedasticity robust and adjusted for clustering within firm-quarter combinations. Standard error in second stage is adjusted for the use of a predicted value on the right-hand side.

Table 3. Exaggeration and expected future forecasts

In panel A, the right hand side of (14) is interacted with the actual number of future forecasts an analyst makes between the current forecast and the end of 1999. Forecasts for the years 1993-97 and for analysts who have already made 50 forecasts are included in the sample. In Panel B, exit from the I/B/E/S sample (defined as making a final forecast and not reappearing in the sample for 2 years) is predicted for each forecast in the 1993-97 period. Past average relative forecast ranking is the average of a 0-1 ranking of analysts' relative forecast accuracy for each firm-quarter in which they forecast.

Panel A. Interaction regression with actual future forecasts
Dependent variable: actual earnings less consensus

	Coeff.	S.E.
Forecast less consensus	0.251	0.067
(FOR - CONS)*(Actual future forecasts/100)	0.117	0.037
Constant (in basis points)	-0.159	0.202
Actual future forecasts/100 (in basis points)	-0.134	0.140
Observations	198,909	

Panel B. Probit regression predicting exit from sample

	Coeff.	S.E.
Forecasts in career/100	-0.251	0.014
Ln(Analysts at brokerage)	-0.090	0.010
Ln(Forecast information content)	0.001	0.003
Observations	334,388	