The University of Chicago Booth School of Business

Working Paper No. 10-04
CRSP Working Paper

Reading the Tea Leaves: Why Serial Correlation Patterns in Analysts' Forecast Errors Are Not Evidence of Inefficient Information Processing

Juhani Linnainmaa
University of Chicago Booth School of Business

Walter N. Torous
University of California, Los Angeles

Reading the Tea Leaves: Why Serial Correlation Patterns in Analysts' Forecast Errors Are Not Evidence of Inefficient Information Processing

JUHANI T. LINNAINMAA and WALTER N. TOROUS*

January 20, 2010

ABSTRACT

Forecast errors are serially uncorrelated when agents possess full information about a firm. However, if there are multiple firm types and types are unobservable, agents update as if dealing with the *average* firm. Consequently, agents underreact to signals from some firms and overreact to signals from others. Applying this explanation to analysts' forecasts, our model predicts that, first, analysts must overreact as well as underreact to earnings announcements; second, serial correlation patterns in analysts' forecast errors diminish over time; third, the largest reduction in serial correlation occurs with the receipt of initial signals; fourth, analysts' forecasts are initially biased but the bias also diminishes over time; and fifth, because signal size is informative about firm type, analysts update differentially in response to small versus large signals. We confirm these predictions using analyst forecast data from I/B/E/S. Our findings suggest that an absence of serial correlation in forecast errors is not the appropriate benchmark for rational analyst behavior.

Keywords: Parameter uncertainty, learning, financial analysts

JEL classification: G14, G24

*Linnainmaa is at University of Chicago Booth School of Business, 5807 S Woodlawn Ave Suite 450, Chicago, IL 60637, juhani.linnainmaa@chicagobooth.edu, (773) 834-3176. Torous is at UCLA Anderson School of Management, University of California, Los Angeles, 110 Westwood Plaza, Los Angeles, CA 90995, walter.n.torous@anderson.ucla.edu, (310) 825-4059. We thank Alan Bester, Michael Brennan, Simon Gervais, Shingo Goto, Bing Han, David Hirshleifer, Matti Keloharju, Ralph Koijen, Christian Leuz, Ľuboš Pástor, Joshua Rauh, Doug Skinner, Ross Valkanov, Pietro Veronesi, and the seminar participants at the University of Chicago and University of Wisconsin-Madison for helpful comments. An earlier version of this paper was presented at the 2005 WFA meetings. We also thank Micah Allred and Brett Myers for their research assistance. Juhani Linnainmaa gratefully acknowledges financial support from the University of Chicago Booth School of Business. All errors are our own.

How analysts forecast is at the heart of the debate on over- and underreaction in financial markets. For example, while a long-term reversal pattern in stock prices is consistent with overreaction (De Bondt and Thaler 1985), a rational mechanism can also generate this pattern (Fama and French 1996). Because stock prices fluctuate in response to discount factor news as well as cash flow news, changes in expectations about future earnings cannot be cleanly isolated from returns data. Numerous studies therefore investigate the efficiency with which market participants process new information by focusing on security analysts' forecasts (De Bondt and Thaler 1990). Rationality in these studies stipulates that analysts' forecast errors and forecast revisions be serially uncorrelated (Easterwood and Nutt 1999). Any evidence of serial correlation or biases in analysts' forecast errors is consequently interpreted as evidence of inefficient information processing.

This paper, however, demonstrates that forecast errors will be serially correlated whenever a rational analyst faces uncertainty about a firm's earnings generating process. Consequently, an absence of serial correlation in forecast errors is not an appropriate benchmark for rational analyst behavior.

To illustrate this idea, suppose there are two types of firms with the shocks to the earnings of one firm being persistent while the shocks to the other firm's earnings are temporary. If these firm types are *a priori* indistinguishable, an analyst rationally behaves as if analyzing the average firm. If, *ex post*, the firm's earnings shocks are actually persistent, then the analyst's forecast errors will be positively serially correlated. In this case, the analyst will be seen as not making full use of past earnings information and so underreacting to the firm's earnings announcement. Alternatively, if, *ex post*, the firm's earning shocks are actually temporary, the analyst would appear to be overreacting as the optimal forecast now relies too heavily on the earnings announcement itself.

While this example is intended to be simplistic, this paper formulates a model in which rational analysts optimally learn about a firm's earnings process in the face of uncertainty about firm type.

We then provide empirical evidence that the serial correlation patterns and biases in forecast errors documented by others arise as a consequence of analysts optimally learning and resolving this uncertainty confronting them.

The forecast errors of a rational analyst will be serially uncorrelated if the analyst has full information about a firm's earnings generating process. However, if analysts do not have full information then forecast errors typically are serially correlated. For example, in Markov and Tamayo (2006), analysts learn over time but unknowingly believe that a firm's earnings are less persistent than they truly are. By assuming that analysts' beliefs are systematically biased, Markov and Tamayo are able to generate positively serially correlated forecast errors. Our explanation of underreaction, by contrast, does not require that analysts systematically make mistakes. Rather it hinges on the sensible requirement that analysts *a priori* do not know with certainty the type of firm being followed. This is consistent with, for example, Sloan (1996), who argues that the fundamental problem confronting analysts is distinguishing between the persistent versus temporary components of earnings signals.

We assume that firm type is differentiated by a parameter characterizing a firm's profitability. As before, suppose analysts function in a setting where there are only two types of firms now distinguished by this profitability parameter. Conditional on the firm type, the analyst has a prior on the parameter and with the receipt of signals is able to learn the true parameter value by applying Bayes' rule.[1] However, if firm types are unobservable then the analyst faces a more difficult updating problem. Not knowing which of the two types of firms is being analyzed, the analyst's prior will now

[1]Bayes' rule describes how rational agents update their beliefs after receiving new information. A prior distribution summarizes our knowledge about an unknown parameter before any data have been observed. An agent's prior beliefs about some parameter θ could, for example, be normally distributed with a mean of zero and a variance of one. Upon receiving new information, the agent relies on the likelihood function. The likelihood function gives the probability of observing the new signal conditional on θ being equal to some particular value θ'. Finally, Bayes' rule prescribes exactly how an agent should combine a priori beliefs, summarized by the prior distribution, with the information contained in the new observation, given by the likelihood function, to update beliefs. We view security analysts' behavior as being governed by optimal statistical inference in which analysts correctly combine new data with an accurate probabilistic model of the firm. This is consistent with recent evidence (e.g., Griffiths and Tenenbaum (2007)) suggesting that human judgment in everyday tasks corresponds closely to the optimal predictions produced by Bayesian models. By contrast, studies such as those by Kahneman and Tversky (1973) characterize cognitive judgments as the result of heuristics that are insensitive to prior beliefs.

be a mixture of the priors characterizing each of the firm types. This prior must be too narrow for one type of firm and too wide for the other type of firm. Conditional on the signals actually coming from a firm for which the analyst's prior is too narrow, the seeming precision of this prior means that the analyst's rational updating will underweight the data and lead to positively correlated forecast errors. Alternatively, if the signals come from the other type of firm for which the analyst's prior is too wide, the seeming imprecision of this prior will now lead to negatively correlated forecast errors as the analyst's rational updating overweights the data. In either case, however, the *expected* serial correlation in forecast errors from the analyst's perspective is zero. Even though firm types are unobservable, there is no *a priori* requirement that the proportions of each firm type are equal. This implies that if, for example, an analyst has greater uncertainty about the profitability parameter of, say, three-quarters of all new firms but less about the remaining firms, the analyst's prior will be too narrow for the majority of these firms. As a result, the analyst will be seen to underreact to the earnings announcements of most firms.

This explanation is reminiscent of the arguments of Timmermann (1993, 1996) and Lewellen and Shanken (2002) which rely on parameter uncertainty to rationalize asset pricing anomalies and apparent deviations from market efficiency. While returns might appear predictable to an econometrician, because of parameter uncertainty, investors can neither perceive nor exploit this predictability.

Any theory of analysts' forecasting behavior must be consistent with the extensive empirical evidence of systematic serial correlation patterns in their forecast errors. The most prominent of this evidence is that analysts' forecast errors are positively serially correlated, or equivalently, future earnings changes are positively correlated with past forecast errors. This finding is interpreted as evidence of underreaction on the part of analysts. Analysts do not appear to fully incorporate all available information in forming their forecasts if future earnings can be predicted by their past forecast errors. In addition to this serial correlation pattern, other regularities in analysts' behavior cast doubt on the optimality of their forecasts. For example, analysts' earnings forecasts appear to be systematically over-optimistic. Also, analysts appear to revise their earnings forecasts less in the face of large earnings surprises than small earnings surprises.

Motivated by this and other empirical evidence, we use forecast panel data from I/B/E/S (Institutional Broker Estimate Systems) to test a number of implications stemming from our model uncertainty explanation of analysts' forecasting behavior. First, we document that analysts underreact to the earnings announcements of approximately three out of every five firms and overreact in the remaining two out of every five firms. Analysts in our model overreact to the announcements of some firms while underreacting to the signals of other firms. Second, the serial correlation in analysts' forecast errors diminishes over time. The serial correlation is also weaker for firms which were older when analysts began following them. Third, the reduction in the serial correlation of analysts' forecast errors is largest with the receipt of the initial signals. These results are all in agreement with a learning argument. For example, learning implies that signals are the most informative when analysts' priors are the most dispersed. Fourth, our model also predicts that while analysts' forecasts are on average unbiased, they are biased in the case of an individual firm itself. We document that, as predicted by the model, these optimism/pessimism patterns in the data decay over time. Finally, we find that analyst forecast revisions are nonlinear functions of new signals. This result is also consistent with our model. Because the analyst's prior is a mixture of distributions, the size of a signal is informative about firm type. Thus, an analyst's optimal updating rule is always nonlinear.

Another novel contribution of this paper is to document the extent to which a time-series investigation of analysts' forecasts is sensitive to survivorship bias. We show that if firm profitability is correlated with firm survival and analysts learn about firm profitability, an econometrician is likely to conclude that underreaction pattern is stronger than it actually is. This bias arises because a surviving firm must have been better than the average firm. We confirm the direction of this bias in the I/B/E/S data. When we exclude firms for which there are only a few observations—for example, to obtain sharper parameter estimates or just to be able to estimate the time-series model at all—the underreaction pattern appears stronger than what it is in the full sample.

Our paper is organized as follows. Section I provides the background for the study. Section II presents a learning model in which agents face uncertainty, provides testable implications, and illustrates these implications using numerical examples. Section III describes the data. Section IV shows that analysts' forecast revisions and the serial correlation patterns in analyst forecast errors are consistent with the model. Section V concludes.

I. Background

Security prices exhibit well-documented overreaction and underreaction patterns. For example, when portfolios are formed on the basis of three- to five-year past returns, stock prices display evidence of overreaction (De Bondt and Thaler 1985). However, when portfolios are based on either short-term past returns or on significant corporate events, such as earnings announcements, stock prices appear to underreact.[2]

Unfortunately, consensus has not yet been reached as to whether these patterns reflect investor irrationality or a rational alternative, such as learning in the presence of structural uncertainty. Because many competing theories yield similar implications for the behavior of security prices, it is difficult to distinguish between these theories on the basis of price data (Brav and Heaton 2002). To circumvent this difficulty, numerous studies test explanations of under- and overreaction by focusing on financial analysts' forecasts (De Bondt and Thaler 1990; Zhang 2006a). Investigating how efficiently market participants process new information avoids reliance on security price data and so provides a cleaner setting to investigate under- and overreaction in securities markets.

Studies investigating how efficiently financial analysts process new information provide mixed

[2]See, for example, Ball and Brown (1968), Abarbanell and Bernard (1992), and Jegadeesh and Titman (1993). Daniel, Hirshleifer, and Subrahmanyam (1998) survey the literature.

evidence. For example, De Bondt and Thaler (1990) conclude that analysts overreact to new information. By contrast, Abarbanell and Bernard (1992), Ali, Klein, and Rosenfeld (1992), Elliott, Philbrick, and Wiedman (1995), Shane and Brous (2001), Raedy, Shane, and Yang (2006), and others find evidence consistent with analyst underreaction. Easterwood and Nutt (1999), on the other hand, suggest that analysts underreact to negative information and overreact to positive information.[3] Other studies have examined analysts' over- and underreaction patterns by documenting analysts' differential responses to small and large signals as well as to positive and negative signals.[4] In addition to these under- and overreaction patterns, analysts' forecasts also appear to be biased. Stickel (1990), Abarbanell (1991), Brown (1997), and Das, Levine, and Sivaramakrishnan (1998) find that analysts' earnings forecasts are systematically over-optimistic.

To concentrate solely on the role of learning, we shut down two channels that appear to influence analysts' behavior. First, analysts' incentives may affect their forecasting behavior. For example, sell-side analysts who provide more favorable forecasts generate more trades and business for the brokerage firms and investment banks which employ them.[5] Similarly, an analyst may rationally underweight his private signals in the presence of unemployment risk because the analyst may get fired if a bold forecast misses the mark.[6]. Second, we ignore the feedback loop between analysts and the firms they follow. Firms may manage their earnings to meet or beat market expectations[7] or manage market expectations to meet forecasts.[8] Although both of these channels undoubtedly influence analysts' forecasting behavior, we show that even in their absence, analysts' forecast errors can still be serially correlated and their forecasts biased.

[3]However, Abarbanell and Lehavy (2003), Ahmed, Lobo, and Zhang (2000), and Mikhail, Walther, and Willis (2003), among others, note that the Easterwood and Nutt (1999) results are sensitive to how outliers are classified in the data.

[4]See, for example, Easterwood and Nutt (1999) and Skinner and Sloan (2002).

[5]See, for example, Jackson (2005), Ljungqvist, Marston, and William J. Wilhelm (2006), Ljungqvist, Marston, Starks, Wei, and Yan (2007), and Fang and Yasuda (2009). Mehran and Stulz (2007) survey the literature.

[6]See, for example, Hong and Kubik (2003) and Clarke and Subramanian (2006)

[7]See, for example, Burgstahler and Dichev (1997), Kasznik (1999), Abarbanell and Lehavy (2003), Burgstahler and Eames (2003).

[8]See, for example, Matsumoto (2002), Richardson, Teoh, and Wysocki (2004), and Cotter, Tuna, and Wysocki (2006).

Several studies have examined the implications of learning on financial analysts' behavior. Mikhail, Walther, and Willis (2003) find that the serial correlation in analyst forecast errors decreases in an analyst's experience. Bolliger (2004) finds similar results for forecast accuracy. He also concludes that it is firm-specific experience which matters and that general analyst experience appears to be unrelated to forecast accuracy. In a related study, Markov and Tamayo (2006) show that analysts' uncertainty about a firm's underlying earnings generating process can induce serial correlation into analysts' forecast errors. However, in contrast to our framework, Markov and Tamayo assume analysts have systematically mistaken prior beliefs about a firm's earnings generating process. If this bias were eliminated from their prior beliefs, analysts' forecast errors in the Markov and Tamayo model would once again become serially uncorrelated.

II. Theory

A. Serial Correlation in Forecast Errors

Forecast errors must be serially uncorrelated if the parameters of the underlying data generating process are known (Harvey 1989). This result holds regardless of the distributional assumptions placed on the error terms. The intuition is that if forecast errors were serially correlated, a rational agent's updating rule can be improved by taking this serial correlation into account. For example, suppose that an agent uses an updating rule which results in positively serially correlated forecast errors. Because the agent knows the underlying data generating process, this expected serial correlation can be computed. Noticing that the forecast errors are positively correlated, the agent recognizes that the current updating rule's response to each signal is insufficient. As a result, the updating rule is altered to give more weight to each new signal. The agent then recomputes the expected serial correlation in the forecast errors and makes further adjustments until the resultant forecast errors become serially uncorrelated.

However, if the agent does not know parameters of the underlying data generating process and, as a result, uses wrong values in an updating rule, the resultant forecast errors become serially correlated. To demonstrate this result, consider the simplest possible setup: an agent is uncertain about a fixed location parameter μ and both the agent's prior and likelihood are normally distributed.[9] Formally, the agent's prior on μ is normally distributed with mean m and variance σ_B^2. Suppose that μ is actually drawn from a normal distribution with mean m but variance σ_A^2, $\tilde{\mu} \sim N(m, \sigma_A^2)$. Notice that the agent's prior is unbiased because the prior's mean coincides with the mean of $\tilde{\mu}$. For convenience, we can write $\tilde{\mu}$ as $m + \tilde{e}_A$, where $\tilde{e}_A$ is a zero mean random variable with variance σ_A^2.

Now suppose that the agent receives two successive signals about μ:

$$\begin{aligned} \tilde{s}_1 &= m + \tilde{e}_A + \tilde{e}_1, \\ \tilde{s}_2 &= m + \tilde{e}_A + \tilde{e}_2, \end{aligned} \tag{1}$$

where $\tilde{e}_1$ and $\tilde{e}_2$ are normally distributed *i.i.d.* zero mean noise terms with variance σ^2. These errors are also uncorrelated with $\tilde{e}_A$. With this notation, the forecast errors, $\tilde{\xi}_1$ and $\tilde{\xi}_2$, can be written as

$$\begin{aligned} \tilde{\xi}_1 &= \tilde{s}_1 - m = \tilde{e}_A + \tilde{e}_1, \\ \tilde{\xi}_2 &= \tilde{s}_2 - m' = m + \tilde{e}_A + \tilde{e}_2 - m', \end{aligned} \tag{2}$$

where m' is the agent's posterior mean after observing the first signal $\tilde{s}_1$. Given the normality assumption, the posterior mean m' is

$$m' = \frac{\frac{m}{\sigma_B^2} + \frac{m + \tilde{e}_A + \tilde{e}_1}{\sigma^2}}{\frac{1}{\sigma_B^2} + \frac{1}{\sigma^2}} = (1-w)m + w(m + \tilde{e}_A + \tilde{e}_1), \text{ where } w \equiv \frac{\frac{1}{\sigma^2}}{\frac{1}{\sigma_B^2} + \frac{1}{\sigma^2}} = \frac{\sigma_B^2}{\sigma_B^2 + \sigma^2}. \tag{3}$$

[9] Our results apply as well to more general prediction problems. For example, suppose that a firm's earnings-per-share (EPS) are generated by an $AR(1)$ process with autoregressive parameter b. If an analyst is uncertain about b, the analyst's beliefs about b will be updated with the receipt of each new earnings report. Our analysis suggests that if the analyst's prior on b coincides with b's population distribution, the analyst's EPS forecast errors, will be serially uncorrelated. If, however, the analyst faces model uncertainty and does not know, a priori, which distribution b is drawn from, then forecast errors will be serially correlated.

Because the prior is unbiased, we have that $E(\tilde{\xi}_1) = E(\tilde{\xi}_2) = 0$ in expression (2). As a result, the serial covariance in forecast errors is

$$\begin{aligned} \text{cov}(\tilde{\xi}_1, \tilde{\xi}_2) &= \sigma_A^2 - E[(\tilde{e}_A + \tilde{e}_1)m'] && (4) \\ &= \sigma_A^2 - w(\sigma_A^2 + \sigma^2) = \frac{\sigma^2}{\sigma^2 + \sigma_B^2}(\sigma_A^2 - \sigma_B^2), && (5) \end{aligned}$$

where the second line uses the assumption of normality. Expression (5) demonstrates that if the prior distribution is too narrow relative to the distribution of μ ($\sigma_B^2 < \sigma_A^2$), forecast errors are positively serially correlated. Intuitively, the agent is placing too much weight on the prior and too little weight on the data. Conversely, if the prior distribution is too wide ($\sigma_B^2 > \sigma_A^2$), forecast errors are negatively serially correlated because too much weight is being placed on the data. Forecast errors are serially uncorrelated only if the variance of the prior distribution coincides with μ's true variance.

In this example, forecast errors are serially correlated when the agent's prior does not coincide with the unknown parameter's true population distribution. This is likely to be the case when an agent faces model uncertainty. For example, suppose that the agent is uncertain as to which one of two possible models generate the observed data. In the first model, μ is drawn from a distribution with variance σ_A^2 while in the second model the variance is σ_B^2 where $\sigma_A^2 > \sigma_B^2$. If the agent cannot distinguish between these models, the resultant prior never coincides with the truth: the agent updates by relying on a weighted average of the two models. This implies that the agent's forecast errors will exhibit positive serial correlation if the first model is true and negative serial correlation if the second model is true. We now formalize this intuitive idea by studying the problem of an analyst who cannot distinguish between firms.

B. Updating with Two Types of Firms

We assume that firms in the economy differ according to their underlying profitability as measured by an unobserved location parameter $\tilde{\mu}$. A fraction ω of the firms in the economy are type A and the remainder are type B. The profitability of a type A firm is drawn from a normal distribution $N(m_A, \sigma_A^2)$ while the profitability of a type B firm is drawn from a normal distribution $N(m_B, \sigma_B^2)$. Suppose that $\sigma_A^2 > \sigma_B^2$. Firms are initially indistinguishable from one another so that an analyst only learns over time a particular firm's type by observing new signals.

The analyst who begins following a new firm has a prior distribution on its profitability $p(\mu) = \omega_t \times N(m_A, \sigma_A^2) + (1 - \omega_t) \times N(m_B, \sigma_B^2)$ where ω_t is the prior mixture probability.[10] Before receiving the first signal, the mixture probability coincides with ω. With the receipt of each new signal, the analyst's beliefs about the following quantities are updated:

1. The probability that the firm is of type A. This is the mixture probability ω_t.

2. The distribution of μ conditional on the firm actually being type A. This is the distribution $N(m_{A,t}, \sigma_{A,t}^2)$.

3. The distribution of μ conditional on the firm actually being type B. This is the distribution $N(m_{B,t}, \sigma_{B,t}^2)$.

Any updates to the analyst's beliefs about these quantities will impact the serial correlation pattern in the analyst's forecast errors.

[10]Let us use a coin-flip example to illustrate why the prior distribution has this particular form. Suppose that coin A is biased towards heads, Pr(Heads|A) = 0.6, and that coin B is biased towards tails, Pr(Heads|B) = 0.4. If we randomly pick a coin from a pile that has seven 'A' coins and three 'B' coins, our prior assessment of the probability of getting heads on the first throw is $\frac{7}{10}$ Pr(Heads|A) + $\frac{3}{10}$ Pr(Heads|B) = 0.54. If the first outcome is heads, the prior mixture probability increases from $\omega = \frac{7}{10}$ because coin A is more likely to produce such an outcome. See, for example, Rossi, Allenby, and McCulloch (2006) and Robert (2007) for a discussion on problems where the prior distribution is a mixture of distributions. A number of papers in the asset pricing literature have relied on an investor's prior being a mixture of distributions. For example, the investor in Brennan and Xia (2001) who is not sure which asset pricing model holds has a prior which is a mixture of normals. Detemple (1991) also examines an asset pricing model where an investor's posterior belief distribution is a mixture of two normals.

Forecast errors will be serially correlated in this setup because the analyst's prior is never exactly right for any particular firm. If a firm is of type A, the firm's true profitability is drawn from a distribution $N(m_A, \sigma_A^2)$, but the analyst's prior is narrower because it mixes in density $N(m_B, \sigma_B^2)$ in proportion $1 - \omega_t$. Similarly, if the firm is of type B, the analyst's prior is too wide.

Suppose that the analyst receives a signal $y_t = \mu + \epsilon_t$, where ϵ_t is a normally distributed *i.i.d.* zero mean noise term, $\epsilon_t \sim N(0, \sigma^2)$ for all t. The posterior distribution for μ after observing y_t is

$$\begin{aligned} p(\mu|y_t) &\propto p(y_t|\mu)p(\mu) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(y_t-\mu)^2}{\sigma^2}} \left(\frac{\omega_t}{\sqrt{2\pi}\sigma_{A,t}} e^{-\frac{1}{2}\frac{(\mu-m_{A,t})^2}{\sigma_{A,t}^2}} + \frac{1-\omega_t}{\sqrt{2\pi}\sigma_{B,t}} e^{-\frac{1}{2}\frac{(\mu-m_{B,t})^2}{\sigma_{B,t}^2}} \right) \\ &= \frac{\omega_t}{\sqrt{2\pi(\sigma^2+\sigma_{A,t}^2)}} e^{-\frac{1}{2}\frac{(y_t-m_{A,t})^2}{\sigma^2+\sigma_{A,t}^2}} \frac{1}{\sqrt{2\pi}\sigma_{A,t+1}} e^{-\frac{1}{2}\frac{(\mu-m_{A,t+1})^2}{\sigma_{A,t+1}^2}} \\ &\quad + \frac{1-\omega_t}{\sqrt{2\pi(\sigma^2+\sigma_{B,t}^2)}} e^{-\frac{1}{2}\frac{(y_t-m_{B,t})^2}{\sigma^2+\sigma_{B,t}^2}} \frac{1}{\sqrt{2\pi}\sigma_{B,t+1}} e^{-\frac{1}{2}\frac{(\mu-m_{B,t+1})^2}{\sigma_{B,t+1}^2}}, \end{aligned} \tag{6}$$

where $m_{A,t+1} \equiv \frac{\frac{m_{A,t}}{\sigma_{A,t}^2}+\frac{y_t}{\sigma^2}}{\frac{1}{\sigma_{A,t}^2}+\frac{1}{\sigma^2}}$, $\sigma_{A,t+1}^2 \equiv \frac{1}{\frac{1}{\sigma^2}+\frac{1}{\sigma_{A,t}^2}}$, and similarly for $m_{B,t+1}$ and $\sigma_{B,t+1}^2$. Therefore, the analyst's posterior distribution is a weighted average of two normal distributions:

$$p(\mu|y_t) = \omega_{t+1} N(m_{A,t+1}, \sigma_{A,t+1}^2) + (1-\omega_{t+1}) N(m_{B,t+1}, \sigma_{B,t+1}^2), \tag{7}$$

where the date $t+1$ mixture probability ω_{t+1} is now

$$\omega_{t+1} = \frac{\frac{\omega_t}{\sqrt{\sigma^2+\sigma_{A,t}^2}} e^{-\frac{1}{2}\frac{(y_t-m_{A,t})^2}{\sigma^2+\sigma_{A,t}^2}}}{\frac{\omega_t}{\sqrt{\sigma^2+\sigma_{A,t}^2}} e^{-\frac{1}{2}\frac{(y_t-m_{A,t})^2}{\sigma^2+\sigma_{A,t}^2}} + \frac{1-\omega_t}{\sqrt{\sigma^2+\sigma_{B,t}^2}} e^{-\frac{1}{2}\frac{(y_t-m_{B,t})^2}{\sigma^2+\sigma_{B,t}^2}}}. \tag{8}$$

Table I investigates the properties of the serial correlation in analysts' forecast errors arising from model uncertainty. We conduct simulations assuming the variance of the likelihood function

σ^2 equals one as does σ_B^2. However, σ_A^2 is varied along the vertical axis between one and ten while the proportion of type A firms in the economy ω is varied along the horizontal axis between zero and one. We report the results of numerically computing the correlation between the resultant forecast errors $\tilde{\xi}_1$ and $\tilde{\xi}_2$.

[Table I here]

In Panel A, the firm is actually type A. Since $\sigma_A^2 \geq \sigma_B^2$, the serial correlation in forecast errors is non-negative throughout. In particular, positively serially correlated forecast errors obtain for $\sigma_A^2 \neq 1$ and $\omega \neq 1$. By contrast, the firm is type B in Panel B. Negatively serially correlated forecast errors now obtain for $\sigma_A^2 \neq 1$ and $\omega \neq 0$.

The mixture probability determines how often analysts under- and overreact to new signals. For example, if an analyst has greater uncertainty about the profitability parameter of, say, three-quarters of all new firms but less about the remaining firms, the analyst's prior will be too narrow for the majority of these firms. As a result, the analyst will be seen to underreact to the earnings announcements of three-quarters of firms and overreact to the announcements of the remaining firms. An econometrician who analyzes data originating from such a world would detect positively serially correlated forecast errors three-quarters of the time. The estimated serial correlation in forecast errors would be zero only in a pooled cross section-time series regression that uses all observations across all firms and the composition of the empirical data matches the analyst's information set. For example, if the mixture of firms in the empirical data differs from the true mixture of firms due to survivorship problems, even a pooled regression would return a non-zero estimate for the serial correlation in forecast errors.

C. Approximate Serial Covariances with Two Types of Firms

There is no analytical expression for the serial covariance in forecast errors because the posterior mixture probability ω_{t+1} is a nonlinear function of the signal y_t. We can, however, approximate the serial covariance by expanding ω_{t+1} in expression (8) around the prior mean, $y_t = m_t$. For simplicity, both normally distributed prior components are assumed to have the same mean, $m_A = m_B = m_t$. Because $\frac{\partial \omega_{t+1}}{\partial y_t}\Big|_{y_t=m_t} = 0$, the unexpected component of the signal, $y_t - m_t$, has only a second-order effect on the posterior mixture probability. By ignoring higher-order terms, the approximate posterior mixture probability can be expressed as

$$\omega_{t+1} \approx \hat{\omega}_{t+1} = \frac{\frac{\omega_t}{\sqrt{\sigma^2+\sigma^2_{A,t}}}}{\frac{\omega_t}{\sqrt{\sigma^2+\sigma^2_{A,t}}} + \frac{1-\omega_t}{\sqrt{\sigma^2+\sigma^2_{B,t}}}} = \frac{1}{1 + \frac{1-\omega_t}{\omega_t}\sqrt{\frac{\sigma^2+\sigma^2_{A,t}}{\sigma^2+\sigma^2_{B,t}}}}. \tag{9}$$

If the analyst receives a signal $y_t = m_t$, the probability of this signal is evaluated under each of the two components of the prior distribution. If one of the components has a higher variance, the probability $\hat{\omega}_{t+1}$ shifts towards the lower variance distribution because this distribution has more density at point m_t. For example, if $\sigma_{A,t} > \sigma_{B,t}$, the probability of a type A firm decreases, $\hat{\omega}_{t+1} < \omega_t$.

The exact posterior mean is $m_{t+1} = \omega_{t+1} m_{A,t+1} + (1-\omega_{t+1}) m_{B,t+1}$, where $m_{j,t+1} = (1-w_j)m_t + w_j(m_t + \tilde{e}_j + \tilde{e}_1)$ and $w_j \equiv \frac{\sigma_j^2}{\sigma_j^2+\sigma^2}$ for both firms j. By comparison, the approximate posterior mean is

$$\begin{aligned} \hat{m}_{t+1} &\approx \hat{\omega}_{t+1} m'_A + (1-\hat{\omega}_{t+1}) m'_B \\ &= m_t + \big(\hat{\omega}_{t+1} w_A + (1-\hat{\omega}_{t+1}) w_B\big)(\tilde{e}_j + \tilde{e}_1) \\ &= m_t + \frac{1}{\sigma^2+\sigma^2_{B,t}} \left(\sigma^2_{B,t} + \hat{\omega}_{t+1}\sigma^2 \frac{\sigma^2_{A,t} - \sigma^2_{B,t}}{\sigma^2+\sigma^2_{A,t}} \right)(\tilde{e}_j + \tilde{e}_1). \end{aligned} \tag{10}$$

The approximate covariance of the analyst's forecast errors if the firm is actually type A is

$$\begin{aligned}
\operatorname{cov}(\tilde{\xi}_1, \tilde{\xi}_2 \mid \text{Type A Firm}) &\approx \sigma^2_{A,t} - E[(\tilde{e}_A + \tilde{e}_1)\hat{m}_{t+1}] \\
&= (1 - \hat{\omega}_{t+1}) \frac{\sigma^2_{A,t} - \sigma^2_{B,t}}{\sigma^2_{B,t} + \sigma^2} \sigma^2 \\
&= \frac{1 - \omega_t}{(1-\omega_t) + \omega_t \sqrt{\frac{\sigma^2 + \sigma^2_{B,t}}{\sigma^2 + \sigma^2_{A,t}}}} \frac{\sigma^2_{A,t} - \sigma^2_{B,t}}{\sigma^2_{B,t} + \sigma^2} \sigma^2,
\end{aligned} \tag{11}$$

and, by symmetry, the approximate covariance of analyst's forecast errors if the firm is actually type B is

$$\operatorname{cov}(\tilde{\xi}_1, \tilde{\xi}_2 \mid \text{Type B Firm}) \approx \frac{1 - \omega_t}{(1 - \omega_t) + \omega_t \sqrt{\frac{\sigma^2 + \sigma^2_{A,t}}{\sigma^2 + \sigma^2_{B,t}}}} \frac{\sigma^2_{B,t} - \sigma^2_{A,t}}{\sigma^2_{A,t} + \sigma^2} \sigma^2. \tag{12}$$

Several intuitive properties of the serial covariances of forecast errors can be derived from expression (11). First, the absolute magnitude of the serial covariance of forecast errors is decreasing in ω. This decrease reflects the fact that as $\omega \to 1$, the analyst's prior converges towards the correct distribution. Conversely, as $\omega \to 0$ then almost all firms are of type B, and so the aggregate prior is a very poor fit for a type A firm. Second, expression (11) implies that the absolute magnitude of the serial covariance of forecast errors increases in the signal variance σ^2. In particular, as $\sigma^2 \to 0$, serial covariance disappears. The reason is that if signals are almost noise-free, the analyst learns μ almost immediately. Finally, we see that the serial covariance in forecast errors increases in the spread between the two prior variances, $|\sigma^2_{A,t} - \sigma^2_{B,t}|$. If the two types of firms are almost identical, $\sigma^2_{A,t} \approx \sigma^2_{B,t}$, and the aggregate prior is almost correct for each firm. If, on the other hand, the spread in variances is large, the aggregate prior fits neither firm very well.

Table I confirms that the serial correlation in forecast errors varies significantly in the mixture probability ω. The fewer type A firms in the economy, the larger is the serial correlation in forecast errors for these firms. Serial correlation can also be seen to increase in the spread $\sigma^2_{A,t} - \sigma^2_{B,t}$. The effect is somewhat non-monotonic, however, as an increase in this spread beyond some point decreases the serial correlation.

D. Optimism and Pessimism in Analysts' Forecasts

To this point we have assumed that the distributions associated with type A and type B firms have the same mean. Any difference in means is irrelevant to the serial covariance properties of forecast errors which only depends on a difference in variances. However, if the means of the two distributions differ, the model also generates seemingly predictable biases in analysts' forecasts. This is an important additional implication because a number of studies have documented what appears to be persistent optimism (or pessimism) on the part of analysts.

The model generates biased forecasts through a mechanism similar to that responsible for the serial correlation patterns. For example, suppose that the profitability of a type A firm is drawn from a distribution with a higher mean than the distribution associated with a type B firm. If firm types are unobservable, the mean of the aggregate prior will be too low for type A firms. This implies that the analyst will make a too pessimistic forecast about a type A firm. This pessimism persists until the analyst resolves enough uncertainty. The opposite holds for type B firms for which the analyst's forecasts will initially be too optimistic.

This result follows from the properties of the posterior mean in the mixture prior setup. If we take expression (7) and assume that (i) $\sigma^2_{A,t} = \sigma^2_{A,t} \equiv \sigma^2_p$ and (ii) $m_{A,t} \neq m_{B,t}$, the posterior mean becomes

$$E(\mu|y_t) = \frac{\frac{1}{\sigma^2}}{\frac{1}{\sigma^2_p} + \frac{1}{\sigma^2}} y_t + \frac{\frac{1}{\sigma^2_p}}{\frac{1}{\sigma^2_p} + \frac{1}{\sigma^2}} (\omega_{t+1} m_{A,t} + (1 - \omega_{t+1}) m_{B,t}). \tag{13}$$

If, unbeknownst to the analyst, the firm is actually type A, the expected signal is $E(y_t|\text{Type A}) = m_{A,t}$ by the unbiased prior assumption. The expected posterior mean conditioned on information about firm type, which is not available to the analyst, can be written as

$$E(E(\mu|y_t)\,|\,\text{Type A}) = m_{A,t} + (1-\omega_{t+1})\frac{\frac{1}{\sigma_p^2}}{\frac{1}{\sigma_p^2}+\frac{1}{\sigma^2}}(m_{B,t} - m_{A,t}). \tag{14}$$

Expression (14) demonstrates that because of model uncertainty the posterior mean no longer equals the prior mean. In this case, the posterior mean is biased towards a type B firm until the analyst either learns the firm type or the means of the two prior distribution components converge.

[Figure 3 here]

Figure 3 summarizes the results of simulations illustrating this persistence in the optimism/pessimism patterns of analysts' forecasts. In these simulations, the mean of type A firms is set equal to one while the mean of type B firms is zero. In the case of type A firms, until analysts resolve type uncertainty, their forecasts will be systematically too low. Although the convergence towards zero takes place very quickly at the outset, its speed tapers off. For type B firms, analysts' forecasts are systematically too high. This pessimism gives rise to persistent negative forecast errors.

E. Learning and Survivorship Bias

If analysts learn about a firm's earnings generating process then an empirical investigation of their forecast errors is subject to survivorship bias. This conclusion holds even in the absence of model uncertainty when all firms are *a priori* identical. For example, suppose that an analyst's prior on a firm's profitability parameter is a standard normal distribution as is the likelihood function. In this simple setup, if all firms survive then, according to expression (5), an analyst's forecast errors are serially uncorrelated. If, however, firms which experience sufficiently poor profitability are liquidated then these firms will no longer be followed by analysts and will drop out of the econometrician's sample.

We simplify our argument by assuming a firm's profitability parameter summarizes all available information about the firm's survivorship. To fix matters, suppose that an analyst's posterior mean of the profitability parameter falling below zero coincides with a firm's liquidation. In this case, the analyst's updating will remain unchanged because knowing whether a firm fails or not is uninformative in light of the firm's earnings announcement. Surviving firms will now be seen to differ from firms that do not survive. A surviving firm will consistently announce earnings which exceed analysts' expectations while a firm whose earnings repeatedly fall short of analysts' expectations is more likely not to survive. This pattern means that an econometrician will document positively serially correlated forecast errors amongst the surviving firms. Any selection criteria which skew the sample towards surviving firms is more likely to result in the econometrician documenting positively correlated analysts' forecast errors.

To verify this intuition, we simulate a maximum of five years of quarterly data for each of 100,000 firms assuming a standard normal prior distribution and likelihood function. Initially, a firm is liquidated in these simulations when the analyst's posterior mean falls below zero. As a result, firms in the simulation have a minimum of two quarterly observations. If we include all available data on each of the simulated firms, a pooled regression of date $t+1$ forecast error on date t forecast errors results in a slope coefficient that is statistically indistinguishable from zero. However, if we restrict attention to firms which have at least four quarters of data, then the resultant sample is skewed towards surviving firms and the pooled regression gives an estimated slope coefficient of 0.005 with a t-statistic of 13.2. This bias increases as the survival rate decreases and the sample selection criteria results in more data being thrown away. For example, if a firm survives as long as the analyst's posterior mean remains above one, the same pooled regression with a minimum of four quarterly observations now gives an estimated slope coefficient of 0.053 with a t-statistic of 54.1.

Survivorship bias is a concern in empirical analyses of analyst behavior in which a large number of observations are needed to obtain reliable firm-specific estimates. Markov and Tamayo (2006), for example, restrict their sample to firms with at least sixteen years of data, acknowledging that by doing so, they cannot extrapolate from their sample to the population as a whole.

While sample selection criteria used by researchers will tend to result in positively correlated analysts' forecast errors being estimated, analysts' coverage of U.S. firms is also limited to begin with. For example, while CRSP has return data for 6,987 firms in 2006, analysts in I/B/E/S issued forecasts for only 5,211 of these firms in that year. Therefore, analysts do not cover approximately one-quarter of all publicly-traded U.S. firms. Analysts also do not randomly choose which firms to cover. Studies such as those by O'Brien and Bhushan (1990), Pearson (1992), and Rajan and Servaes (1997) find that analyst coverage increases in firm size, the size of the firm's industry, market beta, idiosyncratic risk from the market model, institutional ownership, industry growth, and the underpricing in the initial public offering.

Although we cannot entirely circumvent the issue of survivorship bias due to the nature of the analyst data, we take a number of steps in our empirical work to minimize the problem. First, we broaden our sample by including even those firms for which we have just enough data to estimate the parameters of interest, regardless of the resultant estimates' noisiness. Second, we examine survivorship bias explicitly by comparing standard firm-specific estimates with estimates obtained from a structural model. The structural model is estimated using data on all firms for which there are any observations available. Finally, throughout this paper, we investigate the validity of our learning model by testing its firm-specific predictions. We focus on predictions about how forecast errors behave over time because survivorship bias is largely concerned with the composition of the cross-section of firms and not about what happens to a particular firm.

F. Empirical Predictions

When a rational analyst faces model uncertainty, the resultant forecast errors and, consequently, forecast revisions can be positively or negatively serially correlated. This model uncertainty explanation for analyst behavior provides a number of testable predictions.

Prediction 1. *If there are N firms in the economy and a fraction ω of the firms are of the high variance type, analysts' forecast errors are positively correlated for ωN firms (underreaction pattern) and negatively correlated for the remaining $(1-\omega)N$ firms (overreaction pattern).*

An analyst's prior is correct for the weighted average of all firms in the economy. If analysts underreact to signals from one type of firm, they must overreact to signals from the other type of firm.

The results tabulated in Table I illustrate prediction 1. For example, when $\sigma_A^2 = 1.9$ and $\omega = 0.1$, forecast errors exhibit positive correlation for 10% of the firms and negative serial correlation for the remaining 90% of the firms. At the same time, the magnitude of the serial correlation in the case of the 10% of firms is larger, +0.17, than for the 90% of firms, −0.02.

Prediction 2. *The serial correlation in forecast errors diminishes over time.*

Analysts' forecast errors exhibit serial correlation because they face model uncertainty. As analysts observe more signals, they resolve this uncertainty and, consequently, the serial correlation in forecast errors diminishes.

An additional prediction is that the serial correlation in forecast errors should decrease with firm age. For example, if a firm has existed for, say, fifty years prior to an analyst beginning to observe signals, the analyst can resolve type uncertainty by studying the firm's history.[11]

Prediction 3. *The reduction in serial correlation of analysts' forecast errors is largest with the receipt of the initial signals.*

Serial correlation patterns change in response to learning. If analysts have observed many signals,

[11] This prediction assumes that there are no significant structural breaks in the firm's data generating process. If such breaks are present then the analyst may not be able to rely on past data to resolve type uncertainty.

their prior distributions narrow and they pay less attention to new data. This deceleration-of-learning is a common feature of learning models.[12]

[Figure 1 here]

Figure 1 illustrates the behavior over time of the serial correlation in analysts' forecast errors. We simulate data for both type A and B firms and for each realization compute the correlation coefficient between forecast errors $\tilde{\xi}_t$ and $\tilde{\xi}_{t+1}$. In panel A the signal variance is assumed to be high, $\sigma^2 = 10$, leading to relatively persistent serial correlation in forecast errors. In this case, we see that it requires the arrival of approximately ten signals before the serial correlation for type B firms approaches zero (below 0.01) and approximately twenty signals in the case of type A firms. By contrast, Panel B lowers the signal variance to $\sigma^2 = 1$, making the signals more informative. The serial correlation for type B firms now approaches zero with the receipt of approximately five signals and approximately ten signals are needed in the case of type A firms.

The patterns evident in Figure 1 illustrate predictions 2 and 3. First, serial correlation disappears asymptotically as the number of signals increases. Secondly, because signals are relatively more informative at the outset when the analyst's prior is more dispersed, the convergence towards zero is faster with the receipt of the earlier as opposed to later signals.

Prediction 4. *Biases in analysts' forecasts diminish over time.*

The model also predicts that analysts appear to be "optimistic" about the prospects of some firms and "pessimistic" about the prospects of other firms. These patterns arise because analysts do not know from which distribution each firm's profitability parameter is drawn from. As analysts observe more signals and learn, these optimism and pessimism patterns diminish.

Prediction 5. *An analyst's updating rule is nonlinear as forecast revisions are no longer a constant multiple of forecast errors regardless of the size of the error.*

[12] See, for example, Harvey (1989) and Pratt, Raiffa, and Schlaifer (1995).

The posterior mean is linear in the signal when the prior is conjugate to the likelihood function.[13] Many common models fall into this category, the most popular being models in which prior beliefs and signal errors are each normally distributed. The tractability of these models follows from the fact that, by definition, the resultant posterior belongs to the same family of distributions as the prior.

However, if the prior is not conjugate to the likelihood, the updating rule is no longer linear in the signal. In our model uncertainty framework, we have assumed that an analyst's prior is a *mixture* of normals which is no longer conjugate to a normal likelihood. The nonlinearity of the analyst's resultant updating rule can be seen in expressions (7) and (8).

To illustrate prediction 5, Figure 2 displays the behavior of the posterior mean as a function of the signal received by the analyst. The solid line corresponds to the case of a conjugate prior because both components of the prior distribution have the same variance ($\sigma_A^2 = \sigma_B^2 = 1$). In this case, the posterior mean is a linear function of the signal. The other two lines correspond to cases where we increase firm type A variance to $\sigma_A^2 = 5$ and then to $\sigma_A^2 = 20$. As a consequence, the analyst's prior is no longer conjugate to the normal likelihood. The posterior mean can now be seen to be a nonlinear function of the signal. The relation is almost linear for the very smallest signals. But for more extreme signal values, the posterior mean places more weight on the signal.

[Figure 2 here]

Finally, it is important to note that these predictions are not specific to the assumption that the prior distribution is a mixture of normals. To see this, Table II computes correlation coefficients for the case where the components of the prior distribution are t-distributed with different degrees of freedom. The degrees of freedom for type B firms is set to $\nu_B = 1,000$ in these simulations while the degrees of freedom for type A firms ranges from $\nu_A = 3$ to $\nu_A = 33$. When $\omega < 1$, the aggregate

[13] See, for example, DeGroot (1970) for a discussion on conjugate priors.

prior is too narrow for type A firms, leading to positive serial correlation in forecasts errors.[14] When $\omega > 0$, the aggregate prior is too wide for type B firms and forecast errors exhibit negative serial correlation.

[Table II here]

III. Data

Our primary data requirement is analysts' earnings forecasts. To better understand the determinants of analysts' under- and overreaction patterns, forecast data will be supplemented by additional data on firm characteristics and stock prices.

I/B/E/S Analyst Data. We rely on analyst consensus earnings forecasts for all U.S. firms. These data are from the quarterly Institutional Broker Estimate Systems (I/B/E/S) files beginning in January 1984 through February 2008. We use the unadjusted I/B/E/S files to avoid the rounding error problems associated with the split-adjusted files (Diether, Malloy, and Scherbina 2002).

The raw data contain 334,314 firm-quarter observations. Our tests require consecutive firm-quarter observations because we investigate the time-series properties of analysts' forecast errors. We also remove potentially erroneous observations by requiring that analysts' earnings per share (EPS) forecasts cannot increase or decrease by more than \$10 from quarter t to $t+1$ (De Bondt and Thaler 1990). We also require that each firm be matched against CRSP stock price data and that stock price data be available for one year preceding a forecast. These requirements leave us with a sample of 291,270 firm-quarter observations for 11,410 firms. It should be noted that the time-series is very short for many of the firms in our sample. For example, the median number of sequential quarterly observations is just eighteen or, equivalently, four and a half years of quarterly data.

Compustat Annual Files. We compute the following measures from the merged annual

[14] The variance of a variable t-distributed with ν degrees of freedom is $\frac{\nu}{\nu-2}$.

CRSP/Compustat data set: (i) market value defined as the number of shares outstanding times the annual closing stock price, (ii) book-to-market ratio calculated as the book-to-market per share reported in Compustat divided by the annual closing stock price, and (iii) cash-flow volatility computed as in Zhang (2006b).[15]

CRSP Monthly Data. In addition to providing stock price data, we obtain firm age, stock volatility, and major SIC industry membership from the monthly files of the Center for Research in Security Prices (CRSP). We compute firm age as the difference between each I/B/E/S report date and the firm's first appearance in the CRSP database. We compute stock volatility as the sample standard deviation of returns using the preceding five years of monthly returns terminating one month before each earnings announcement.

IV. Empirical Evidence

A. Under- and Overreaction Patterns in the Cross-Section of Firms

A common approach to test for analysts' under- and overreaction is to run a regression of quarter $t+1$'s forecast error on a constant and quarter t's forecast error:

$$\mathrm{FE}_{i,t+1} = a + b\,\mathrm{FE}_{i,t} + e_{i,t}. \tag{15}$$

The forecast error $\mathrm{FE}_{i,t}$ is defined as firm i's actual EPS in quarter t minus the analysts' corresponding consensus forecast scaled by the average stock price:

$$\mathrm{FE}_{i,t+1} \equiv \frac{\mathrm{EPS}_{i,t} - E^{*}_{t^{-}}[\mathrm{EPS}_{i,t}]}{\bar{P}_{i,t-1}}, \tag{16}$$

[15]Zhang (2006b) measures the standard deviation of cash flow from operations relying on the past five years of data but requiring a minimum of three years. Cash flow from operations is defined as earnings before extraordinary items minus total accruals, scaled by average total assets, where total accruals are equal to changes in current assets minus changes in cash, changes in current liabilities, and depreciation expense plus changes in short-term debt.

where $E^*_{t-}[\text{EPS}_{i,t}]$ is the latest consensus estimate from I/B/E/S prior to quarter t's earnings announcement and $\bar{P}_{i,t-1}$ is the average stock price calculated over the preceding one year period concluding one month before quarter t ends. We scale forecast errors by average stock price to make them comparable across firms. Also, scaling by average stock price rather than the current stock price $P_{i,t}$ mitigates any concern that reversal or momentum patterns in stock prices could induce serial correlation in $\text{FE}_{i,t+1}$.

We standardize forecast errors before estimating expression (15). To do so, we first rank all negative forecast errors into one set of quintiles and all positive forecast errors into another set of quintiles. We repeat this ranking separately for each quarter and then form deciles by merging the two quintiles. As a result, the categories are symmetric with deciles one through five containing negative forecast errors and deciles six through ten containing positive forecast errors. We place zero forecast errors in a separate middle category.

A pooled regression of expression (15) using all 249,979 consecutive firm-quarter observations gives an estimated slope coefficient of $\hat{b} = 0.287$ with a standard error of 0.003 and has an adjusted R^2 of 0.08. This slope coefficient estimate implies underreaction on the part of analysts: positive (negative) forecast errors on average are followed by positive (negative) forecast errors. These estimation results are consistent with the results of, among others, Mendenhall (1991), Abarbanell and Bernard (1992), and Shane and Brous (2001).

The highly significant slope coefficient estimate in expression (15), however, masks the fact that analysts often exhibit the opposite behavior by seemingly overreacting to earnings announcements. That is, positive (negative) forecast errors tend often to be followed by negative (positive) forecast errors. To see this, instead of estimating expression (15) as a pooled regression using data on all firms, we can estimate this expression separately for each individual firm. There are enough observations to identify parameters a and b in expression (15) using OLS for 10,128 firms. For this full sample, the estimated slope coefficient $\hat{b}$ is positive for 67.3% of the firms and negative in the remaining 32.7% of the firms.

It is important to investigate whether this distribution of estimated slope coefficients is consistent with the null hypothesis of serially uncorrelated forecast errors or, alternatively, with Markov and Tamayo (2006)'s model in which analysts possess biased priors, as opposed to our explanation in which analysts resolve firm type uncertainty. To begin with, it should be noted that the resultant estimated slope coefficients are not t-distributed as they would be under the null hypothesis of no serial correlation. To verify this, we use the method of maximum likelihood to fit the slope coefficient estimates to a t distribution and obtain a degrees-of-freedom parameter estimate of 2.09 (SE $= 0.04$). The Kolmogorov-Smirnov test, however, rejects this benchmark t distribution with a p-value of $< .01$. This rejection suggests that the empirical deviations that we obtain from the null hypothesis that $b = 0$ represent more than just sampling errors. Furthermore, the distribution of estimated slope coefficients is not consistent with a model in which analysts possess biased priors. If this were the case, the estimated slope coefficients would be largely positive and the resultant distribution would be unimodal. By contrast, our model in which analysts resolve uncertainty between two firm types would generate a bimodal distribution consistent with the data: significantly positive slope coefficients for some firms and significantly negative slope coefficients for other firms.

These firm-specific estimates indicate that although analysts underreact to new information from a majority of firms, analysts overreact to new information from approximately one-third of firms. The presence of *both* under- and overreaction is consistent with prediction 1. The learning explanation to analyst behavior suggests that, to the extent that analysts underreact to earnings announcements from certain firm, they must overreact to the announcements from others.

Estimating expression (15) on a firm-by-firm basis also allows us to explore the issue of survivorship bias. By restricting attention to firms having fewer observations, we see less evidence of underreaction in the resultant sample. For example, for the 2,832 firms for which we have no more than three years of quarterly data[16], the estimated slope coefficient $\hat{b}$ is positive for only 45.2% of the firms. Overreaction is now the more prevalent feature of this sample.

In general, time-series analyses of the analyst data are very sensitive to firm survival. By discarding firms which have only a few observations, the econometrician would conclude that the underreaction pattern is stronger than it actually is. This follows because "overreaction" firms have empirically a lower survival rate than "underreaction" firms.

B. Time-Series Models of Analysts' Forecast Errors

We now turn our attention to estimating time-series models of forecast errors to better understand the temporal behavior of analysts' under- and overreaction patterns. In particular, we test predictions 2 and 3 which characterize the convergence properties of the serial correlation in analyst's forecast errors. Prediction 2 states that as analysts learn and resolve model uncertainty, serial correlation in their forecast errors will diminish. The reduction in serial correlation is largest at the outset of analysts' information gathering efforts according to prediction 3.

The first time-series model that we estimate is a simple regime switching model. We posit an AR(1) model with drift in which the autoregressive parameter assumes a value of ρ_1 up to some switching time τ after which it changes to a value of $\rho_1 + \rho_2$:

$$\mathrm{FE}_{i,t+1} = a + [\rho_1 + \mathbf{1}(t \geq \tau)\ \rho_2]\ \mathrm{FE}_{i,t} + e_{i,t}, \tag{17}$$

where $\mathbf{1}(t \geq \tau)$ is an indicator variable which equals one if $t \geq \tau$ and equals zero otherwise. The forecast error $\mathrm{FE}_{i,t}$ is defined as in expression (16) except that it is not scaled by the average stock price because we now carry out estimation on a firm by firm basis. This specification allows us to

[16]But for which we can still estimate the parameters of expression (15).

ask whether the subsequent serial dependence in analysts' forecast errors, $|\rho_1 + \rho_2|$, differs from the earlier serial dependence, $|\rho_1|$.

We estimate expression (17) for each firm in I/B/E/S with at least eight consecutive quarterly observations. We use nonlinear least squares and impose the restrictions that $|\rho_1| \leq 1$ and $|\rho_1+\rho_2| \leq 1$. Without imposing stationarity, the parameters a, ρ_1 and ρ_2 can still be estimated consistently. Because τ must be an integer value, we fit the model $T_i - 1$ times for each firm i with sample size T_i, letting τ range from $t = 2$ to $t = T_i$. We choose that τ and corresponding parameter estimates which minimize the sum of squared residuals over the $T_i - 1$ models.

Having estimated expression (17) separately for each firm, we then study the cross-sectional distributions of the $\hat{\rho}_1$ and $\hat{\rho}_2$ estimates. This approach is similar to Fama and MacBeth (1973) because it allows us to avoid the problem of determining the distributions of firm-specific parameter estimates.

The resultant 7,954 firm-specific estimates of expression (17) indicate that the serial correlation in analysts' forecast errors does indeed diminish with time. Prediction 2 requires that $|\hat{\rho}_1+\hat{\rho}_2|-|\hat{\rho}_1| < 0$ and, consistent with this hypothesis, this difference is negative in 56% of the regressions. We can reject the null hypothesis of no change in serial correlation with a z-value of -10.3. The median change in the absolute value of the autoregressive parameter is $|\hat{\rho}_1 + \hat{\rho}_2| - |\hat{\rho}_1| = -0.20$ with a bootstrapped 99% confidence interval of $[-0.23, -0.16]$.[17]

We next investigate whether this decrease in the serial correlation of analysts' forecast errors is more pronounced with earlier data arrivals. To do so, we test whether the speed of convergence in analysts' forecast errors attenuates exponentially versus linearly with time. In the first case, we allow the autoregressive parameter to increase or decrease exponentially in t:

$$\mathrm{FE}_{i,t+1} = a + \rho_0 \, e^{-\kappa \, (t-1)} \, \mathrm{FE}_{i,t} + e_{i,t}, \tag{18}$$

[17] We resample the coefficient estimates 100,000 times with replacement in computing this confidence interval.

where κ is the convergence parameter. The functional form of expression (18) is consistent with the serial correlation pattern of analysts facing model uncertainty (see Figure 1). Alternatively, the autoregressive parameter increases or decreases linearly:

$$\mathrm{FE}_{i,t+1} = a + \rho_0 \max[(1 - \kappa\ (t-1)), 0]\ \mathrm{FE}_{i,t} + e_{i,t}. \tag{19}$$

In this model, ρ increases in t for $\kappa < 0$ and decreases in t for $\kappa > 0$.[18]

Because both models have the same number of parameters, we base our inference on a comparison of the sum of squared residuals of the competing models. Based on this criteria, the exponential decay model fits the data better than the linear decay model for 64% of the firms. This leads us to reliably reject the null hypothesis of no difference between the models with a z-value of 27.5. In other words, not only does the serial correlation in analysts' forecast errors diminish with time, the reduction in serial correlation is largest with the initial data arrivals.

C. Serial Correlation and Optimism/Pessimism Patterns in the Cross-Section of Firms

The final specification we consider is an exponential model of analysts' forecast errors similar to expression (18) except that to test prediction 4 we also allow the intercept to decay exponentially over time:

$$\mathrm{FE}_{i,t+1} = a\ e^{-\kappa_1\ (t-1)} + \rho_0\ e^{-\kappa_2\ (t-1)}\ \mathrm{FE}_{i,t} + e_{i,t}. \tag{20}$$

We estimate this model on a firm-by-firm basis. To better understand the observed patterns in analysts' over- and underreaction, Table III reports summary statistics on the resultant firm-specific estimates across a number of different cross-sections. These include sorting by a firm's book-to-market ratio, size as measured by market capitalization, age, stock volatility, cash flow volatility, and membership in a major SIC industry group. A time series of at least twelve quarterly observations

[18]Until ρ becomes zero for $t \geq \frac{1}{\kappa}$.

is required for a firm to be included. Firms are sorted into bins based on the corresponding values prevailing at the time of their first I/B/E/S observation. For example, the age-based groups sort firms by how old each firm was at the time it first appeared in the analyst data.

For comparison purposes, we also estimate a standard AR(1) model of forecast errors

$$\mathrm{FE}_{i,t+1} = a + \rho\, \mathrm{FE}_{i,t} + e_{i,t}. \tag{21}$$

Firm-specific estimates of ρ are also provided in Table III.

[Table III here]

Panel A tabulates summary statistics on the date 0 autoregressive parameter estimates, $\hat{\rho}_0$, of expression (18). We see that forecast errors are initially negative serially correlated for approximately one out of every three firms (36%). This result holds for each of the sampled industry groups as well. The mean estimate of ρ_0 for the full sample of firms is $\hat{\rho}_0 = 0.15$, which is slightly lower than the autoregressive parameter estimate from the standard AR(1) model of $\hat{\rho} = 0.17$ in Panel C. However, in light of the evidence of serial dependence changing with time, the standard AR(1) model is misspecified. From Panel C we see that by attempting to match the data with a constant ρ, the misspecified model understates the proportion of firms with negatively serially correlated forecast errors.

A number of interesting cross-sectional patterns in the distribution of $\hat{\rho}_0$ also emerge in Panel A. For example, these estimates can be seen to decrease almost monotonically in firm age: the average coefficient is $\hat{\rho}_0 = 0.19$ for the quintile of youngest firms but is less than half this value, $\hat{\rho}_0 = 0.07$, for the quintile of oldest firms, the difference being statistically significant with a t-value of -5.8. This result supports prediction 2 and complements our previous evidence. Even if analysts have not previously followed a particular firm, the serial correlation in forecast errors should be weaker for firms that have been in existence for a longer period of time. By contrast, from Panel C we see that firm age has no discernable effect on the estimated autoregressive parameter $\hat{\rho}$ of the standard AR(1) model. This model is misspecified and simply cannot capture the statistically significant role of firm age on the serial correlation properties of analysts' forecast errors. We also have reliable evidence in Panel A that firms with larger market capitalizations have weaker initial serial correlation in analysts' forecast errors as do firms with lower stock or cash flow volatility.

Panel B reports summary statistics on the firm-specific estimates of the speed of adjustment κ_1 of expression (18)'s intercept. This parameter measures the speed with which any optimism or pessimism on the part of analysts dissipates. The average estimate for the full sample of firms is $\hat{\kappa}_1 = 0.12$ and is statistically significant with a t-value of 5.4. Equivalently, this estimate implies a half-life for the adjustment in the average firm's intercept of 6.0 years.[19] The average estimates of κ_1 are also positive for all but a few of the subsamples. When negative, however, the average estimates are statistically insignificant. All of these results are consistent with prediction 4 which argues that optimism and pessimism patterns in analysts' forecasts should dissipate with time. The speed with which optimism and pessimism dissipate is quicker for the oldest as opposed to the youngest firms and value firms (highest book-to-market quintile) as opposed to growth firms (lowest book-to-market quintile).

Summary statistics on the firm-specific estimates of the speed of adjustment in the serial correlation of analysts' forecast errors, measured by κ_2 in expression (18), are provided in Panel C of Table III. The average estimate of this parameter for the full sample of firms is $\hat{\kappa}_2 = 0.26$ and is statistically significant with a t-value of 10.9. This estimate implies a half-life of 2.7 years in the adjustment of the serial correlation for the average firm. Taken together with the results from Panel A, this half-life estimate suggests that, on average, the autoregressive parameter of analysts' forecast errors decreases from an initial value of $\hat{\rho}_0 = 0.15$ to $\hat{\rho}_1 = 0.12$ after one year and to $\hat{\rho}_2 = 0.09$ after

[19] Determined by solving $1 \times e^{-\kappa_1(t-1)} = \frac{1}{2}$.

two years. It would, however, take more than ten years of data arrivals before this autoregressive parameter would fall below 0.01. Therefore, although the serial correlation pattern in analyst forecast errors dissipates quickly at the outset, a long period of time is required until this pattern is eliminated.

From Panel C we also see that the average κ_2 estimates are positive not only in the full sample but in all except one of the subsamples where it is statistically insignificant. The fact that these estimates are positive across subsamples is important because it suggests that the adjustment in the serial correlation patterns of analysts' forecast errors is not specific to any particular group of firms.

To supplement these estimation results, we also conducted a number of robustness tests and additional analyses not reported here. First, we used annual I/B/E/S data and replaced mean consensus forecasts with medians. Second, we formed additional subsamples based on other firm characteristics such as dividend yield, intangibles, and historical analyst coverage. Third, we defined forecasts errors by scaling with then current stock prices. However, all of these additional analyses yielded results similar to those reported in Table III.

D. Nonlinearities in Forecast Revisions

If model uncertainty generates the observed serial correlation patterns in analysts' forecast errors then according to prediction 5 the resultant forecast revisions should be a nonlinear function of the signals observed by analysts. Intuitively, an analyst's beliefs about which model generates an observed signal will change as the size of the signal changes.

We investigate how analysts revise their forecasts of quarter $t+1$'s EPS in light of quarter t's earnings signal measured by expression (16). To test for the presence of nonlinearities, we regress forecast revisions against the signal itself and a signed second-order term:

$$E^*_{t^+}[\text{EPS}_{i,t+1}] - E^*_{t^-}[\text{EPS}_{i,t+1}] = a + b_1\ s_{i,t} + b_2\ \text{sign}(s_{i,t})\ s^2_{i,t} + e_{i,t}, \tag{22}$$

where $s_{i,t} \equiv \text{EPS}_t - E^*_{t^-}[\text{EPS}_t]$.[20]

If forecast revisions are a nonlinear function of the signal then $b_2 \neq 0$. In the case where the analyst's prior is a mixture of normal distributions and the likelihood is normally distributed, as hypothesized in our earlier discussion, then it can be shown that $b_2 > 0$ (see Figure 2). However, this conclusion is not robust to distributional assumptions. For example, $b_2 < 0$ for a normally distributed prior but a t-distributed likelihood. Here analysts update less in the direction of a large signal, either positive or negative, because such signals are more likely due to a large error drawn from a fatter-tailed likelihood. Alternatively, $b_2 > 0$ for a t-distributed prior and a normally distributed likelihood. Now analysts update more in the direction of the large signal because the prior is fatter-tailed and large signals are more likely to be an informative signal of the firm's type.[21]

Initially, we estimate expression (22) separately for each firm in the sample having at least three years of data (twelve quarterly observations). The average b_2 estimate is negative $\hat{b}_2 = -0.49$ but is not significantly different from zero with a t-value of -0.99. Using a 5% significance level, however, we find that the resultant b_2 estimates are significantly negative in 8.7% of the regressions and significantly positive in 20.5% of the regressions (N = 4,891). Although finding 24.2% ($\equiv$ 8.7% + 20.5% − 5%) "too many" estimates which are statistically significant at the 5% level is consistent with the hypothesis that $b_2 \neq 0$, it raises concerns about the size of this test. The true size of the test may be higher than 5% because, first, forecast error distributions are markedly non-normal (Abarbanell and Lehavy 2003) and, second, because many firms in our sample have a very short time-series of data. At the other extreme, had we insisted that the sampled firms have a long time-series of data then we would potentially worsen the selection bias.

To address these issues, we use a hierarchical model to obtain the distribution of corresponding b_2 estimates for all firms in our sample. This requires that we assume each firm's b_2 parameter

[20] As before, the subscript t^- denotes the time of the latest consensus analyst forecast which predates quarter t's EPS announcement while t^+ is the time of the first consensus analyst forecast available after quarter t's EPS announcement.

[21] See O'Hagan (1979) for more details.

is drawn from a hyperdistribution and then estimate the parameters of this distribution through expression (22). In doing so, we depart from a standard random coefficients model (Hildreth and Houck 1968) by letting the hyperdistribution be a mixture of ten normal distributions. With this specification, we can separately test whether the distribution of b_2 estimates is centered at zero and whether it exhibits fat tails (excess kurtosis).

A hierarchical model is well-suited for this application because we have limited information about a large number of firms ($N = 11,410$). Even firms with few observations contain useful information about the hyperdistribution (Chamberlain 1980). A hierarchical model is better suited than firm-specific regressions for testing the prediction that $b_2 \neq 0$ because we only estimate the parameters of the hyperdistribution and do not attempt to separately estimate the parameters of expression (22) for each firm.

Formally, the model we estimate is

$$\begin{aligned} E^*_{t+}[\mathrm{EPS}_{i,t+1}] - E^*_{t-}[\mathrm{EPS}_{i,t+1}] &\;=\; \beta_{i,0} + \beta_{i,1}\, s_{i,t} + \beta_{i,2}\, \mathrm{sign}(s_{i,t})\, s^2_{i,t} + \varepsilon_{i,t}, \\ \beta_j &\;\sim\; N(\mu_{\mathrm{ind}_i}, \Sigma_{\mathrm{ind}_i}), \\ \mathrm{ind}_i &\;\sim\; \mathrm{Multinomial}_{10}(\mathrm{pvec}), \end{aligned} \tag{23}$$

where $\mathrm{ind}_i = 1, \ldots, 10$ is a latent indicator variable which indicates from which component observation i is drawn from and pvec is a 10×1 vector of mixture probabilities. Thus, for each firm i, parameters $\beta_{i,0}$, $\beta_{i,1}$, and $\beta_{i,2}$ are drawn from distinct ten-component hyperdistributions:

$$\begin{aligned} \mathrm{pvec} &\;\sim\; \mathrm{Dirichlet}(\alpha), \\ \mu_k &\;\sim\; N(\bar{\mu}, \Sigma_k \otimes a_\mu^{-1}), \\ \Sigma_k &\;\sim\; \text{Inverse-Wishart}(v, V). \end{aligned} \tag{24}$$

We use Markov chain Monte Carlo (MCMC) methods to estimate this model, specifying a prior that all components are centered at zero with equal variances.[22]

Figure 4 displays the distributions of firm-specific estimates $\hat{\beta}_{i,1}$ and $\hat{\beta}_{i,2}$ based on twenty thousand runs of a Gibbs sampler in which every tenth draw of the parameters is retained. Notice that the distribution of the estimated nonlinearity parameter $\hat{\beta}_{i,2}$, has the bulk of its mass around the point $\hat{\beta}_2 = -0.1$. However, this distribution is markedly non-normal with a very fat right tail. As a result, for firms having a negative $\beta_{i,2}$ estimate, the average value is $\hat{\beta}_{i,2} = -0.10$, but for firms having a positive estimate, the average value is $\hat{\beta}_{i,2} = 0.42$. Due to this asymmetry, the mean of the distribution is, in fact, slightly positive (0.00063) although it is not significantly different from zero with a t-value of 0.69. Both the location and the shape of this distribution support prediction 5. If analysts responded in the same fashion to small and large signals, the estimated $\beta_{i,2}$ coefficients would be distributed normally around zero because they would only pick up noise in forecast revisions. Figure 4 suggests that the estimated $\beta_{i,2}$ coefficients are close to zero only for a very limited number of firms.

The distributions in Figure 4 lead to very different conclusions about analyst behavior than those drawn from corresponding firm-specific regressions. For example, the distribution of the firm specific $\beta_{i,2}$ estimates is consistent with $\beta_{i,2}$ being positive, implying that analysts tend to update more in response to small signals than to large signals. By contrast, the hierarchical model indicates that the most common nonlinearity pattern, evident for 82% of the firms, is consistent with a modestly negative value of $\beta_{i,2}$. This implies that analysts commonly update less in response to large signals than they do to small signals. Figure 4 also reveals significant heterogeneity across firms in how analysts respond to signals. This heterogeneity implies that a pooled model with fixed parameters would be unsuitable for analyzing analyst behavior.

[22]We follow Rossi, Allenby, and McCulloch's (2006) implementation of this model and use their default parameter values for α, $\bar{\mu}$, Σ_k, a_μ, v, and V to specify the prior distributions.

The distribution of the estimated coefficients of the linear term, $\hat{\beta}_1$, in Figure 4 confirms that analysts almost always revise their estimates of quarter $t+1$'s EPS in the same direction as quarter t's EPS surprise. The average estimated coefficient for the linear term in expression (23) is $\frac{1}{N}\sum_{i=1}^{N}\hat{\beta}_{i,1} = 0.20$.

Taken together, the effects of nonlinearity on an analyst's updating rule are economically significant. By way of example, consider the average firm for which $\beta_{i,2}$ is positive. If the EPS surprise is small, say, $s = 10$ cents, analysts would revise their $q+1$ consensus forecast upwards by 2.4 cents. This is very close to the revision of 2.0 cents that we would observe if $\beta_{i,2}$ were zero. However, if the EPS surprise is large, say, $s = 50$ cents, the actual revision of 20.2 cents is twice as large as the revision of 9.8 cents we would observe if $\beta_{i,2}$ were zero. These pronounced nonlinearities are consistent with prediction 5 and suggest that analysts do rely on signal size to draw inferences about the underlying firm type.

V. Conclusion

Serial correlation in analysts' forecast errors and optimism/pessimism patterns in analysts' forecasts are often interpreted as evidence that analysts are inefficient in the processing of new information. This evidence has been relied upon to cast doubt on the efficiency of financial markets. If knowledgeable market participants like financial analysts make systematic mistakes, it would not be surprising that security prices do not always reflect fundamentals but rather under- and overshoot from time to time. We demonstrate that such a link from financial analysts' behavior to market efficiency is tenuous at best. If analysts are uncertain about the types of firms they follow, serial correlation patterns and biases arise even if analysts update their beliefs optimally. We use data on security analysts' forecasts to show that their seemingly irrational behavior is consistent with a number of predictions stemming from our firm type uncertainty explanation.

Our results are central to the discussion of how efficiently markets process new information. Because their forecasts are directly observable, analysts would appear to be ideal candidates to investigate how expectations change in response to the arrival of new information. However, this task is more complicated if analysts do not know the population distributions from which the earnings process parameters are drawn from. Because of this complication, relying on security analysts' forecasts to distinguish between rational and behavioral alternatives is subject to problems reminiscent to when security prices are used to investigate market efficiency.

Pástor and Veronesi (2009) recently suggested that learning is of first-order importance in a variety of finance applications. The anomalous behavior of various financial quantities may not necessarily reflect irrationality or inefficiency but, rather, may be outcomes of rational learning. As a case in point, many have concluded from the serial correlation patterns detected in the analyst data that analysts' forecasts are inefficient. By contrast, echoing Pástor and Veronesi's (2009) message, we argue that this apparent predictability is the outcome of rational learning by analysts.

REFERENCES

Abarbanell, J. S. (1991). Do analysts' earnings forecasts incorporate information in prior stock price changes? *Journal of Accounting and Economics 14*(2), 147–165.

Abarbanell, J. S. and V. L. Bernard (1992). Tests of analysts' overreaction / underreaction to earnings information as an explanation for anomalous stock price behavior. *Journal of Finance 47*(3), 1181–1207.

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

Ahmed, A., G. Lobo, and F. X. Zhang (2000). Do analysts under-react to bad news and over-react to good news? Syracuse University and University of Chicago working paper.

Ali, A., A. Klein, and J. Rosenfeld (1992). Analysts' use of information about permanent and transitory earnings components in forecasting annual EPS. *The Accounting Review 67*(1), 183–198.

Ball, R. and P. Brown (1968). An empirical evaluation of accounting income numbers. *Journal of Accounting Research 6*(2), 159–178.

Bolliger, G. (2004). The characteristics of individual analysts' forecasts in Europe. *Journal of Banking and Finance 28*(9), 2283–2309.

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

Brennan, M. J. and Y. Xia (2001). Assessing asset pricing anomalies. *Review of Financial Studies 14*(4), 905–942.

Brown, L. D. (1997). Analyst forecasting errors: Additional evidence. *Financial Analysts Journal 53*(6), 81–88.

Burgstahler, D. C. and I. Dichev (1997). Earnings management to avoid earnings decreases and losses. *Journal of Accounting and Economics 24*(1), 99–126.

Burgstahler, D. C. and M. J. Eames (2003). Earnings management to avoid losses and earnings decreases: Are analysts fooled? *Contemporary Accounting Research 20*(2), 253–294.

Chamberlain, G. (1980). Analysis of covariance with qualitative data. *Review of Economics and Statistics 47*(1), 225–238.

Clarke, J. and A. Subramanian (2006). Dynamic forecasting behavior by analysts: Theory and evidence. *Journal of Financial Economics 80*(1), 81–113.

Cotter, J., İ. Tuna, and P. D. Wysocki (2006). Expectations management and beatable targets: How do analysts react to explicit earnings guidance? *Contemporary Accounting Research 23*(3), 593–624.

Daniel, K., D. A. Hirshleifer, and A. Subrahmanyam (1998). Investor psychology and security market under- and overreaction. *Journal of Finance 53*(6), 1839–1885.

Das, S., C. B. Levine, and S. Sivaramakrishnan (1998). Earnings predictability and bias in analysts' earnings forecasts. *The Accounting Review 73*(2), 277–294.

De Bondt, W. F. M. and R. H. Thaler (1985). Does the stock market overreact? *Journal of Finance 40*(3), 379–395.

De Bondt, W. F. M. and R. H. Thaler (1990). Do security analysts overreact? *American Economic Review 80*(2), 52–57.

DeGroot, M. H. (1970). *Optimal Statistical Decisions*. New York: McGraw-Hill College.

Detemple, J. B. (1991). Further results on asset pricing with incomplete information. *Journal of Economic Dynamics and Control 15*(3), 425–453.

Diether, K. B., C. Malloy, and A. Scherbina (2002). Differences of opinion and the cross section of stock returns. *Journal of Finance 57*(5), 2113–2141.

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

Elliott, J. A., D. R. Philbrick, and C. I. Wiedman (1995). Evidence from archival data on the relation between security analysts' forecast errors and prior forecast revisions. *Contemporary Accounting Research 11*(2), 919–938.

Fama, E. J. and K. French (1996). Multifactor explanations of asset pricing anomalies. *Journal of Finance 51*(1), 55–84.

Fama, E. J. and J. D. MacBeth (1973). Risk, return and equilibrium: Empirical tests. *Journal of Political Economy 71*(3), 607–636.

Fang, L. and A. Yasuda (2009). The effectiveness of reputation as a disciplinary mechanism in sell-side research. *Review of Financial Studies 22*(9), 3735–3777.

Griffiths, T. L. and J. B. Tenenbaum (2007). Optimal predictions in everyday cognition. *Psychological Science 17*(9), 767–773.

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and The Kalman Filter*. Cambridge: Cambridge University Press.

Hildreth, C. and J. P. Houck (1968). Some estimators for a linear model with random coefficients. *Journal of the American Statistical Association 63*(322), 584–595.

Hong, H. and J. D. Kubik (2003). Analyzing the analysts: Career concerns and biased earnings forecasts. *Journal of Finance 58*(1), 313–351.

Jackson, A. R. (2005). Trade generation, reputation, and sell-side analysts. *Journal of Finance 60*(2), 673–717.

Jegadeesh, N. and S. Titman (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance 48*(1), 65–91.

Kahneman, D. and A. Tversky (1973). On the psychology of prediction. *Psychological Review 80*(4), 237–251.

Kasznik, R. (1999). On the association between voluntary disclosure and earnings management. *Journal of Accounting Research 37*(1), 57–81.

Lewellen, J. and J. Shanken (2002). Learning, asset-pricing tests, and market efficiency. *Journal of Finance 57*(3), 1113–1145.

Ljungqvist, A., F. Marston, L. T. Starks, K. D. Wei, and H. Yan (2007). Conflicts of interest in sell-side research and the moderating role of institutional investors. *Journal of Financial Economics 85*(2), 420–456.

Ljungqvist, A., F. Marston, and J. William J. Wilhelm (2006). Competing for securities underwriting mandates: Banking relationships and analyst recommendations. *Journal of Finance 61*(1), 301–340.

Markov, S. and A. Tamayo (2006). Predictability in financial analyst forecast errors: Learning or irrationality? *Journal of Accounting Research 44*(4), 725–761.

Matsumoto, D. A. (2002). Management's incentives to avoid negative eatrnings surprises. *The Accounting Review 77*(3), 483–514.

Mehran, H. and R. M. Stulz (2007). The economics of conflicts of interest in financial institutions. *Journal of Financial Economics 85*(2), 267–296.

Mendenhall, R. R. (1991). Evidence on the possible underweighting of earnings-related information. *Journal of Accounting Research 29*(1), 170–179.

Mikhail, M. B., B. R. Walther, and R. H. Willis (2003). The effect of experience on security analyst underreaction. *Journal of Accounting and Economics 35*(1), 101–116.

O'Brien, P. C. and R. Bhushan (1990). Analyst following and institutional ownership. *Journal of Accounting Research 28*(Supplement), 259–286.

O'Hagan, A. (1979). On outlier rejection phenomena in Bayesian inference. *Journal of the Royal Statistical Society, Series B (Methodological), 41*(3), 358–367.

Pástor, Ľ. and P. Veronesi (2009). Learning in financial markets. *Annual Review of Financial Economics 1*, 361–381.

Pearson, N. D. (1992). Determinants of the production of information. University of Rochester working paper.

Pratt, J. W., H. Raiffa, and R. Schlaifer (1995). *Introduction to Statistical Decision Theory*. Cambridge, Massachusetts: The MIT Press.

Raedy, J. S., P. Shane, and Y. Yang (2006). Horizon-dependent underreaction in financial analysts' earnings forecasts. *Contemporary Accounting Research 23*(1), 291–322.

Rajan, R. and H. Servaes (1997). Analyst following of initial public offerings. *Journal of Finance 52*(2), 507–529.

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

Robert, C. P. (2007). *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*. New York, NY: Springer.

Rossi, P. E., G. M. Allenby, and R. McCulloch (2006). *Bayesian Statistics and Marketing*. West Sussex, England: John Wiley and Sons.

Shane, P. and P. Brous (2001). Investor and (value line) analyst underreaction to information about future earnings: The corrective role of non-earnings surprise information. *Journal of Accounting Research 39*(2), 387–404.

Skinner, D. J. and R. G. Sloan (2002). Earnings surprises, growth expectations, and stock returns, or, don't let an earnings torpedo sink your portfolio. *Review of Accounting Studies 7*(2-3), 289–312.

Sloan, R. G. (1996). Do stock prices fully reflect information in accruals and cash flows about future earnings? *The Accounting Review 71*(3), 289–315.

Stickel, S. (1990). Predicting individual analyst earnings forecasts. *Journal of Accounting Research 28*(2), 409–417.

Timmermann, A. G. (1993). How learning in financial markets generates excess volatility and predictability in stock prices. *Quarterly Journal of Economics 108*(4), 1135–1145.

Timmermann, A. G. (1996). Excess volatility and predictability of stock prices in autoregressive dividend models with learning. *Review of Economics and Statistics 63*(4), 523–557.

Zhang, X. F. (2006a). Information uncertainty and analyst forecast behavior. *Contemporary Accounting Research 23*(2), 565–590.

Zhang, X. F. (2006b). Information uncertainty and stock returns. *Journal of Finance 61*(1), 105–137.

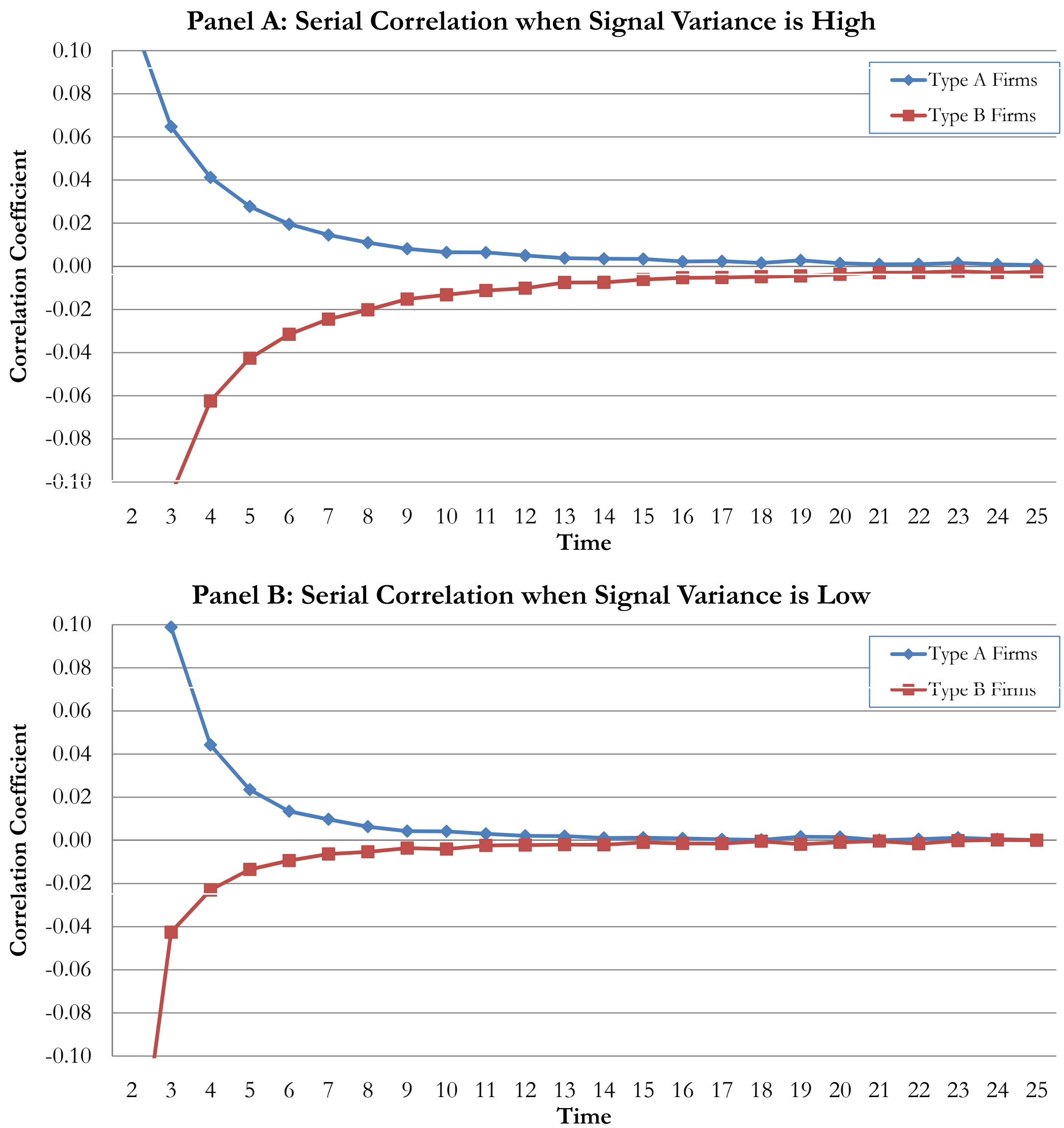

Figure 1. Serial correlation in forecast errors conditional on firm type and time. This Figure plots the serial correlation coefficients between successive forecast errors for type A and type B firms. We separately compute the correlation coefficient cross-sectionally for each new signal. For example, in Panel A for time $t = 2$, the correlation coefficient between the first and second forecast errors is 0.13 for a type A firm and -0.24 for a type B firm. The following parameters are held fixed: the prior mixture probability $\omega = 0.5$, the variance of a type A firm $\sigma_A^2 = 10$, the variance of a type B firm $\sigma_B^2 = 1$, and both types of firms have a prior mean of 0. The signal variance is $\sigma^2 = 10$ in the first panel and $\sigma^2 = 1$ in the second panel. Each correlation coefficient is based on 2.5 million simulations.

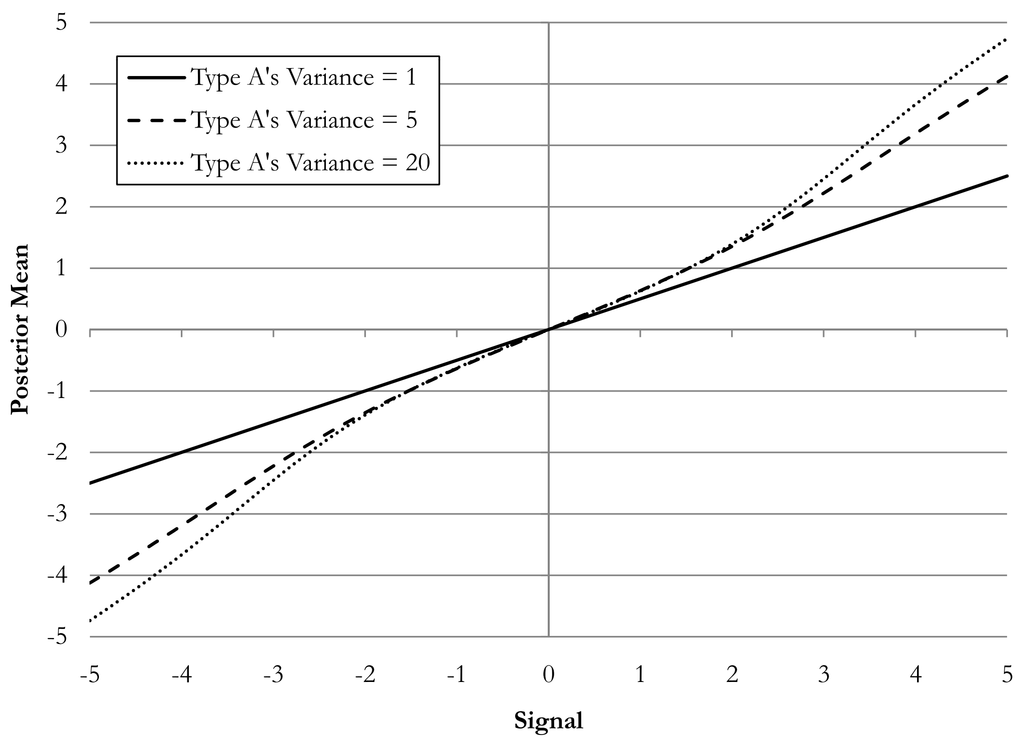

Figure 2. Posterior mean as a function of the signal when the prior is a mixture of two normal distributions. This Figure plots posterior means as a function of the signal. The prior distribution is a mixture of two zero mean normal distributions, with both components having equal weight. The variance of the second distribution (firm type B) is fixed to 1. The variance of the first distribution (firm type A) is either one (solid line), five (dashed line), or twenty (dotted line).

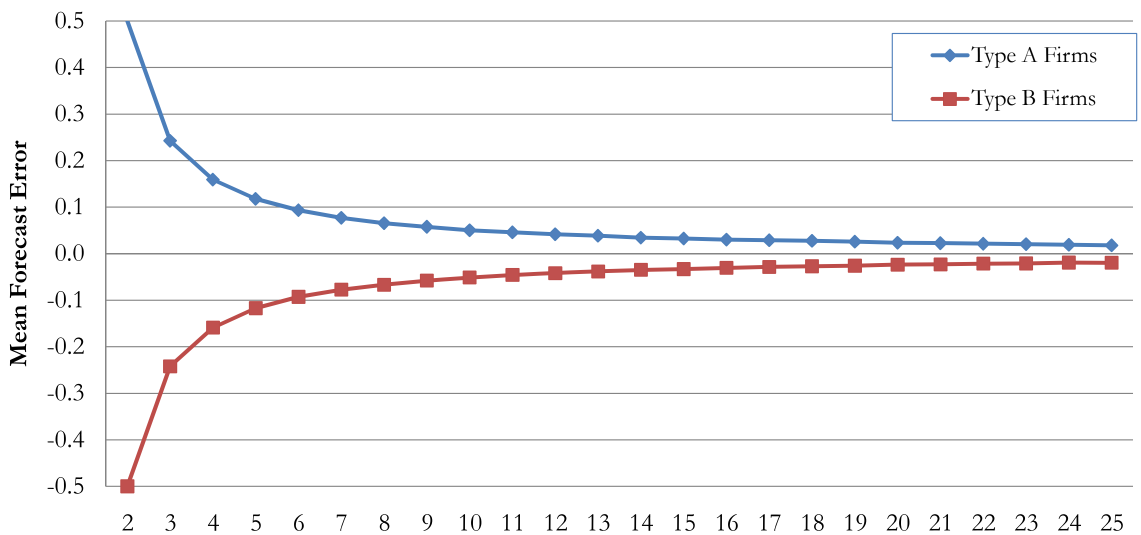

Figure 3. Optimism and pessimism in analyst forecasts conditional on firm type and time. This Figure plots mean forecast errors for type A and type B firms conditional on the date of the signal. The prior mixture probability is $\omega = 0.5$, the variances of the population distributions for the two firm types are $\sigma_p^2 = 1$, and the signal variance is $\sigma^2 = 1$. The mean of the population distribution for type A firms is $m_{A,0} = 1$ while the mean of the population distribution for type B firms is $m_{B,0} = 0$. Each mean forecast error is computed separately for each firm type and is based on 2.5 million simulations.

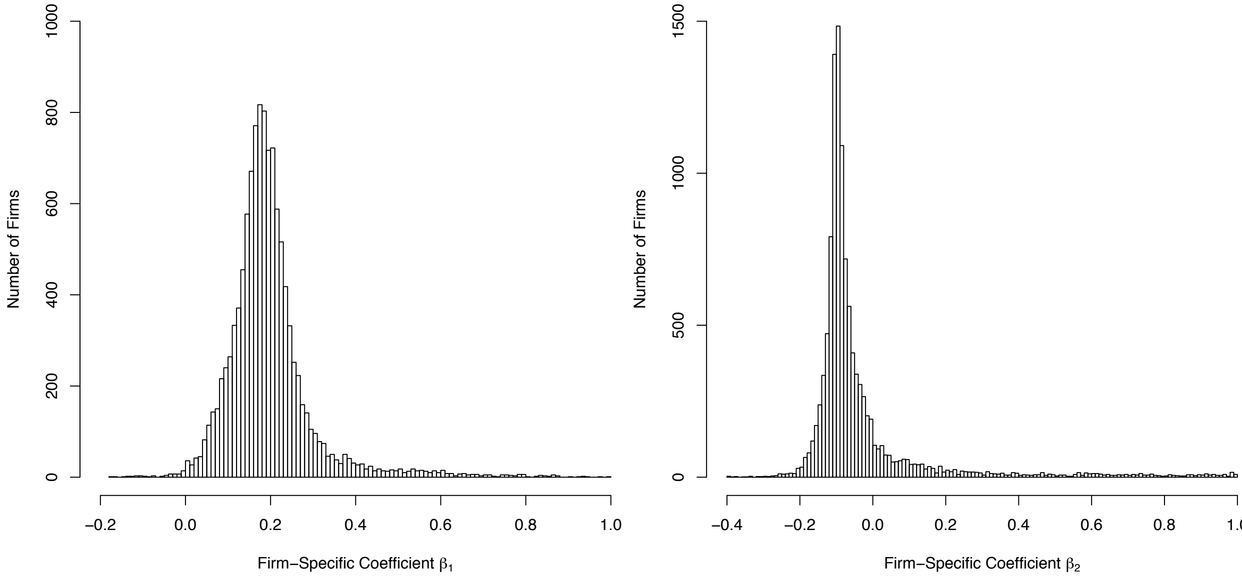

Figure 4. Distribution of firm-specific coefficients in a regression of analyst forecast revision on $s_{i,t}$ and sign$(s_{i,t}) \times s_{i,t}^2$. This Figure plots distributions for the coefficients $\hat{\beta}_{i,1}$ and $\hat{\beta}_{i,2}$ from a hierarchical model. For each firm i, the second stage model's dependent variable is the revision in the consensus forecast about quarter $t + 1$'s earnings around quarter t's earnings announcement. This revision is explained by quarter t's earnings surprise, defined as the actual EPS minus the expected EPS, as well as the signed second-order term of the earnings surprise. In the first stage, each firm i's coefficients $\beta_{i,0}$, $\beta_{i,1}$, and $\beta_{i,2}$ are drawn from hyperdistributions. Each of these hyperdistributions is a mixture of ten normal distributions. The parameters of these hyperdistributions are estimated using I/B/E/S data from January 1984 through February 2008. We estimate the parameters using MCMC, specifying prior distributions for all parameters that are centered around zero.

Table I
Serial Correlation in Forecast Errors as a Function of Firm Type, ω, and σ_A^2

This Table documents serial correlations in analyst forecast errors as a function of firm type, ω, and σ_A^2. Analysts make predictions about a fixed profitability parameter. A fraction ω of the firms in the economy are type A and their profitability μ is drawn from a normal distribution $N(m, \sigma_A^2)$. A fraction $1-\omega$ of firms are type B and their profitability is drawn from a normal distribution $N(m, \sigma_B^2)$. In this Table we vary σ_A^2 and the mixture probability ω and compute the serial correlation between the first two forecast errors for type A (Panel A) and type B (Panel B) firms. Parameters $\sigma_B^2 = 1$ and $\sigma^2 = 1$ are fixed, where σ^2 is the variance of the likelihood function. We calculate serial correlations numerically by simulating data for 2.5 million firms of both types.

Panel A: Serial Correlation in Forecast Errors for Type A Firms

	Prior Mixture Probability, ω										
σ_A^2	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
1.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1.9	0.20	0.17	0.14	0.12	0.10	0.08	0.06	0.05	0.03	0.02	0.00
2.8	0.33	0.25	0.20	0.17	0.14	0.11	0.08	0.06	0.04	0.02	0.00
3.7	0.42	0.28	0.22	0.18	0.15	0.12	0.09	0.07	0.04	0.02	0.00
4.6	0.49	0.29	0.23	0.18	0.15	0.12	0.09	0.07	0.04	0.02	0.00
5.5	0.55	0.29	0.23	0.18	0.15	0.12	0.09	0.07	0.04	0.02	0.00
6.4	0.59	0.29	0.22	0.18	0.14	0.11	0.09	0.06	0.04	0.02	0.00
7.3	0.62	0.28	0.22	0.17	0.14	0.11	0.09	0.06	0.04	0.02	0.00
8.2	0.65	0.27	0.21	0.17	0.14	0.11	0.08	0.06	0.04	0.02	0.00
9.1	0.68	0.26	0.20	0.16	0.13	0.10	0.08	0.06	0.04	0.02	0.00
10.0	0.70	0.26	0.20	0.16	0.13	0.10	0.08	0.06	0.04	0.02	0.00

Panel B: Serial Correlation in Forecast Errors for Type B Firms

	Prior Mixture Probability, ω										
σ_A^2	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
1.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1.9	0.00	−0.02	−0.05	−0.07	−0.08	−0.10	−0.12	−0.13	−0.15	−0.16	−0.18
2.8	0.00	−0.04	−0.08	−0.11	−0.13	−0.16	−0.18	−0.20	−0.22	−0.24	−0.26
3.7	0.00	−0.05	−0.10	−0.13	−0.16	−0.19	−0.22	−0.24	−0.27	−0.29	−0.31
4.6	0.00	−0.06	−0.11	−0.15	−0.18	−0.21	−0.24	−0.27	−0.30	−0.32	−0.35
5.5	0.00	−0.07	−0.12	−0.16	−0.20	−0.23	−0.26	−0.29	−0.32	−0.34	−0.37
6.4	0.00	−0.07	−0.12	−0.17	−0.20	−0.24	−0.27	−0.30	−0.33	−0.36	−0.39
7.3	0.00	−0.08	−0.13	−0.17	−0.21	−0.25	−0.28	−0.31	−0.34	−0.37	−0.40
8.2	0.00	−0.08	−0.13	−0.18	−0.21	−0.25	−0.28	−0.32	−0.35	−0.38	−0.41
9.1	0.00	−0.08	−0.13	−0.18	−0.22	−0.25	−0.29	−0.32	−0.36	−0.39	−0.42
10.0	0.00	−0.08	−0.13	−0.18	−0.22	−0.26	−0.29	−0.33	−0.36	−0.39	−0.43

Table II
Serial Correlation in Forecast Errors when the Prior is a Mixture of Two t Distributions

This Table documents serial correlations in analyst forecast errors as a function of firm type, ω, and σ_A^2. Analysts make predictions about a fixed profitability parameter. A fraction ω of the firms in the economy are type A and their profitability μ is drawn from a t distribution with ν_A degrees of freedom. A fraction $1-\omega$ of firms are type B and their profitability is drawn from a t distribution with ν_B degrees of freedom. In this Table the parameter ν_B is fixed to 1,000 and we vary ν_A and the mixture probability ω and computes the serial correlation between the first two forecast errors for type A (Panel A) and type B (Panel B) firms. We calculate serial correlations numerically by simulating data for 2.5 million firms of both types. The mean of the posterior distribution is computed by evaluating the integrals in the Bayes rule using the composite version of the Simpson's rule with 200 nodes. The error term in the signal is normally distributed with a variance $\sigma^2 = 1$.

Panel A: Serial Correlation in Forecast Errors for Type A Firms

	Prior Mixture Probability, ω										
ν_A	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
3.0	0.33	0.13	0.10	0.08	0.07	0.05	0.04	0.03	0.02	0.01	0.01
6.0	0.12	0.09	0.07	0.06	0.05	0.04	0.03	0.02	0.02	0.01	0.00
9.0	0.08	0.06	0.05	0.05	0.04	0.03	0.02	0.02	0.01	0.00	0.00
12.0	0.05	0.05	0.04	0.03	0.03	0.02	0.02	0.01	0.01	0.00	0.00
15.0	0.04	0.03	0.03	0.03	0.02	0.02	0.02	0.01	0.01	0.00	0.00
18.0	0.04	0.03	0.03	0.02	0.02	0.02	0.02	0.01	0.01	0.00	0.00
21.0	0.03	0.03	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.00	0.00
24.0	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.00	0.00
27.0	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.00	0.00	0.00
30.0	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.00	0.01	0.00	0.00
33.0	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.00

Panel B: Serial Correlation in Forecast Errors for Type B Firms

	Prior Mixture Probability, ω										
ν_A	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
3.0	0.00	−0.02	−0.04	−0.05	−0.06	−0.08	−0.09	−0.10	−0.11	−0.12	−0.13
6.0	0.00	−0.01	−0.02	−0.03	−0.04	−0.05	−0.05	−0.06	−0.07	−0.07	−0.08
9.0	0.00	−0.01	−0.02	−0.02	−0.02	−0.03	−0.04	−0.04	−0.05	−0.05	−0.06
12.0	0.00	−0.01	−0.01	−0.01	−0.02	−0.03	−0.03	−0.03	−0.04	−0.04	−0.05
15.0	0.00	0.00	−0.01	−0.01	−0.02	−0.02	−0.03	−0.03	−0.03	−0.03	−0.04
18.0	0.00	0.00	0.00	−0.01	−0.01	−0.02	−0.02	−0.02	−0.02	−0.03	−0.03
21.0	0.00	0.00	−0.01	−0.01	−0.01	−0.01	−0.02	−0.02	−0.02	−0.02	−0.03
24.0	0.00	0.00	0.00	−0.01	−0.01	−0.01	−0.01	−0.01	−0.02	−0.02	−0.02
27.0	0.00	0.00	0.00	−0.01	−0.01	−0.01	−0.01	−0.01	−0.02	−0.02	−0.02
30.0	0.00	0.00	0.00	−0.01	−0.01	−0.01	−0.01	−0.01	−0.02	−0.02	−0.02
33.0	0.00	0.00	0.00	0.00	−0.01	−0.01	−0.01	−0.01	−0.01	−0.01	−0.02

Table III
Firm-Specific Estimates of the Autoregressive Parameters in Analyst Forecast Errors

This Table provides firm-specific estimates of two time-series models of analysts' forecast errors. The first model regresses quarter $t+1$'s forecast error on an intercept and quarter t's forecast error, but both the intercept and the autoregressive parameter change exponentially in t at rates κ_1 and κ_2, respectively. Panels A, B, and C report on initial autoregressive parameter (ρ_0) and the two exponential change rates (κ_1 and κ_2), respectively. The second model is a standard AR(1) model of the forecast errors with fixed parameters. Panel D reports the autoregressive parameter estimates from this model. Each model is estimated separately for each firm with at least 12 quarterly observations. The data are the unadjusted quarterly I/B/E/S data from January 1984 through February 2008. This Table reports average coefficients and the associated t-values for various subsamples. These subsamples are based on each firm's book-to-market ratio, size (market capitalization), age, stock volatility, cash flow volatility, and major SIC industry group membership. Firms are sorted into bins based on the corresponding values prevailing at the time of their first I/B/E/S observation. The Column labeled % Positive reports the fraction of positive parameter estimates across all firms in each subsample. The last column in Panels B and C convert the κ_1 and κ_2 estimates into half-life estimates. Negative half-life estimates are replaced with missing values. These estimates measure, in years, how long it takes for the intercept and the autoregressive parameter, respectively, to halve in value.

Panel A: Firm-Specific Estimates of ρ_0

Sample	N	Mean	t-value	% Positive
All Firms	6,526	0.153	23.6	63.6
Book-To-Market				
Q1 (Low)	1,234	0.180	11.7	64.7
Q2	1,235	0.168	11.5	65.6
Q3	1,235	0.163	11.1	63.7
Q4	1,235	0.167	11.5	64.0
Q5 (High)	1,237	0.101	6.6	61.3
Q5 Minus Q1	2,471	−0.079	−3.7	
Firm Size				
Q1 (Small)	1,236	0.178	11.5	67.2
Q2	1,237	0.151	10.2	63.5
Q3	1,238	0.155	10.7	63.8
Q4	1,237	0.191	13.3	65.9
Q5 (Large)	1,239	0.105	6.9	58.8
Q5 Minus Q1	2,475	−0.073	−3.3	
Firm Age				
Q1 (Young)	1,277	0.187	12.7	61.1
Q2	1,309	0.189	13.5	64.1
Q3	1,299	0.177	12.5	63.6
Q4	1,296	0.151	10.2	65.4
Q5 (Old)	1,297	0.065	4.4	63.9
Q5 Minus Q1	2,574	−0.122	−5.8	
Stock Volatility				
Q1 (Low)	1,289	0.089	6.2	59.5
Q2	1,291	0.113	7.7	64.0
Q3	1,291	0.189	13.0	67.7
Q4	1,291	0.197	14.2	64.8
Q5 (High)	1,292	0.184	12.0	62.7
Q5 Minus Q1	2,581	0.095	4.5	
Cash Flow Volatility				
Q1 (Low)	1,197	0.114	7.8	63.0
Q2	1,199	0.139	9.4	63.1
Q3	1,199	0.157	10.4	65.3
Q4	1,199	0.189	12.4	66.6
Q5 (High)	1,200	0.198	12.8	64.1
Q5 Minus Q1	2,397	0.083	3.9	
Agriculture (SIC 01-09)	14	0.070	0.5	57.1
Minerals (10-14)	187	0.110	2.7	63.6
Construction (15-17)	65	0.155	2.6	63.1
Manufacturing (20-39)	2,529	0.169	16.5	65.4
Transportation (41-49)	537	0.090	3.9	60.7
Wholesale Trade (50-51)	189	0.176	4.8	67.2
Retail Trade (52-59)	399	0.173	6.9	66.7
Finance (60-67)	953	0.121	7.0	60.3
Service Industries (70-89)	1,123	0.187	11.8	63.9

Panel B: Firm-Specific Estimates of κ_1

Sample	Mean	t-value	% Positive	Half-Life
All Firms	0.116	5.4	50.4	6.0
Book-To-Market				
Q1 (Low)	0.075	1.7	48.9	9.3
Q2	0.009	0.2	47.4	79.1
Q3	0.103	2.1	49.1	6.7
Q4	0.126	2.5	51.5	5.5
Q5 (High)	0.258	5.0	55.3	2.7
Q5 Minus Q1	0.183	2.7		
Firm Size				
Q1 (Small)	0.140	2.9	51.3	4.9
Q2	0.136	2.9	52.9	5.1
Q3	0.155	3.0	50.2	4.5
Q4	0.003	0.1	49.9	212.7
Q5 (Large)	0.146	2.9	48.2	4.7
Q5 Minus Q1	0.006	0.1		
Firm Age				
Q1 (Young)	−0.068	−1.4	45.3	·
Q2	−0.054	−1.2	44.3	·
Q3	0.003	0.1	45.9	248.4
Q4	0.265	5.7	56.7	2.6
Q5 (Old)	0.430	8.4	59.7	1.6
Q5 Minus Q1	0.499	7.0		
Stock Volatility				
Q1 (Low)	0.102	2.3	53.0	6.8
Q2	0.288	5.3	52.3	2.4
Q3	0.073	1.6	47.8	9.5
Q4	0.040	0.9	49.8	17.4
Q5 (High)	0.101	2.1	49.5	6.9
Q5 Minus Q1	−0.002	0.0		
Cash Flow Volatility				
Q1 (Low)	0.134	2.7	49.7	5.2
Q2	0.124	2.7	51.2	5.6
Q3	0.101	2.0	50.0	6.9
Q4	0.090	2.0	48.6	7.7
Q5 (High)	0.152	3.0	52.7	4.5
Q5 Minus Q1	0.018	0.3		
Agriculture (SIC 01-09)	0.920	0.9	42.9	0.8
Minerals (10-14)	0.131	1.1	54.0	5.3
Construction (15-17)	−0.057	−0.3	40.0	·
Manufacturing (20-39)	0.157	4.7	50.7	4.4
Transportation (41-49)	0.321	4.0	54.7	2.2
Wholesale Trade (50-51)	0.236	1.8	56.1	2.9
Retail Trade (52-59)	0.082	1.0	46.1	8.5
Finance (60-67)	0.050	0.8	48.5	13.9
Service Industries (70-89)	−0.042	−1.0	48.5	·

Panel C: Firm-Specific Estimates of κ_2

Sample	Mean	t-value	% Positive	Half-Life
All Firms	0.255	10.9	55.0	2.7
Book-To-Market				
Q1 (Low)	0.275	5.4	56.4	2.5
Q2	0.199	3.7	54.4	3.5
Q3	0.296	5.2	55.8	2.3
Q4	0.229	4.6	54.6	3.0
Q5 (High)	0.265	5.0	54.2	2.6
Q5 Minus Q1	−0.010	−0.1		
Firm Size				
Q1 (Small)	0.312	5.8	55.2	2.2
Q2	0.186	3.8	54.7	3.7
Q3	0.231	4.5	53.9	3.0
Q4	0.166	3.3	54.8	4.2
Q5 (Large)	0.376	6.2	56.8	1.8
Q5 Minus Q1	0.064	0.8		
Firm Age				
Q1 (Young)	0.165	2.8	56.1	4.2
Q2	0.088	1.9	54.1	7.9
Q3	0.226	4.4	52.4	3.1
Q4	0.317	6.4	56.2	2.2
Q5 (Old)	0.477	8.7	56.1	1.5
Q5 Minus Q1	0.311	3.9		
Stock Volatility				
Q1 (Low)	0.238	4.8	53.2	2.9
Q2	0.351	6.5	55.7	2.0
Q3	0.211	4.3	56.1	3.3
Q4	0.145	2.8	52.5	4.8
Q5 (High)	0.343	6.0	57.4	2.0
Q5 Minus Q1	0.105	1.4		
Cash Flow Volatility				
Q1 (Low)	0.205	4.1	54.2	3.4
Q2	0.162	3.5	53.8	4.3
Q3	0.344	5.8	55.5	2.0
Q4	0.270	5.4	56.5	2.6
Q5 (High)	0.250	4.5	55.4	2.8
Q5 Minus Q1	0.046	0.6		
Agriculture (SIC 01-09)	0.484	1.0	50.0	1.4
Minerals (10-14)	0.260	2.1	61.0	2.7
Construction (15-17)	−0.192	−1.2	43.1	.
Manufacturing (20-39)	0.266	7.3	56.6	2.6
Transportation (41-49)	0.386	4.4	54.9	1.8
Wholesale Trade (50-51)	0.065	0.5	48.7	10.7
Retail Trade (52-59)	0.200	2.5	50.4	3.5
Finance (60-67)	0.191	3.2	52.7	3.6
Service Industries (70-89)	0.283	4.6	55.5	2.4

Panel D: Firm-Specific Estimates of ρ in AR(1) Model

Sample	Mean	t-value	% Positive
All Firms	0.172	28.8	71.8
Book-To-Market			
Q1 (Low)	0.195	13.5	72.9
Q2	0.176	11.5	73.2
Q3	0.153	12.6	71.1
Q4	0.181	11.5	72.7
Q5 (High)	0.153	14.1	69.7
Q5 Minus Q1	−0.042	−2.3	
Firm Size			
Q1 (Small)	0.202	15.4	74.4
Q2	0.171	13.3	71.4
Q3	0.155	14.3	71.6
Q4	0.188	11.0	73.2
Q5 (Large)	0.142	9.8	69.0
Q5 Minus Q1	−0.061	−3.1	
Firm Age			
Q1 (Young)	0.165	13.0	72.2
Q2	0.178	12.5	72.2
Q3	0.191	12.8	71.1
Q4	0.174	17.7	72.5
Q5 (Old)	0.144	11.4	71.0
Q5 Minus Q1	−0.021	−1.2	
Stock Volatility			
Q1 (Low)	0.124	12.3	67.0
Q2	0.159	13.0	72.7
Q3	0.183	14.2	75.5
Q4	0.208	12.9	71.7
Q5 (High)	0.188	12.3	72.4
Q5 Minus Q1	0.063	3.5	
Cash Flow Volatility			
Q1 (Low)	0.136	11.6	70.8
Q2	0.162	17.5	70.7
Q3	0.175	10.1	73.5
Q4	0.202	16.8	75.1
Q5 (High)	0.192	10.8	71.6
Q5 Minus Q1	0.056	2.6	
Agriculture (SIC 01-09)	0.111	1.3	71.4
Minerals (10-14)	0.141	4.6	72.7
Construction (15-17)	0.292	4.8	78.5
Manufacturing (20-39)	0.187	17.7	73.1
Transportation (41-49)	0.130	6.2	68.9
Wholesale Trade (50-51)	0.200	8.4	73.0
Retail Trade (52-59)	0.185	10.4	75.4
Finance (60-67)	0.163	12.3	69.0
Service Industries (70-89)	0.162	12.0	71.8