Reading the tea leaves: Model uncertainty, robust forecasts, and the autocorrelation of analysts' forecast errors*

Juhani T. Linnainmaa Walter Torous James Yae

Abstract

We put forward a model in which analysts are uncertain about a firm's earnings process. Faced with the possibility of using a misspecified model, analysts issue forecasts that are robust to model misspecification. We estimate that this mechanism explains approximately 60% of the autocorrelation in analysts' forecast errors. The remainder stems from the cross-sectional variation in mean forecast errors and in analysts' estimation errors of the persistence of earnings growth shocks. Consistent with our model, we find that analysts learn about some features of the earnings process but not others, and this learning reduces, but does not eliminate, the autocorrelation of forecast errors as firms age. Other potential explanations for the autocorrelation of analyst's forecast errors are rejected. Our model of robust forecasting applies not only to analysts' forecasts but to all model-based forecasts.

Keywords: Model uncertainty; parameter uncertainty; forecasting; robustness; financial analysts
JEL classification: G14, G24

*Linnainmaa is at University of Chicago Booth School of Business and NBER, Torous is at MIT, and Yae is at University of Houston. We thank two anonymous referees, Alan Bester, Michael Brennan, Joseph Gerakos, Simon Gervais, Shingo Goto, Bing Han, David Hirshleifer, Matti Keloharju, Ralph Koijen, Christian Leuz, Ľuboš Pástor, Josh Rauh, Bill Schwert (the editor), Doug Skinner, Ross Valkanov, Pietro Veronesi, and the seminar and conference participants at the University of Chicago, University of Wisconsin-Madison, and the Western Finance Association meetings for helpful comments on earlier versions of this paper. We also thank Micah Allred and Brett Myers for their research assistance. Juhani Linnainmaa acknowledges financial support from the University of Chicago Booth School of Business.

1. Introduction

If analysts know the earnings process of a firm and seek to minimize the mean squared error of their forecasts, the forecast errors will be mean zero and serially uncorrelated. In the data they are neither. The average forecast error is positive and forecast errors are positively autocorrelated. This autocorrelation pattern is puzzling because it suggests that analysts do not learn from their past mistakes. While the literature agrees on these empirical facts, there has been a veritable explosion of research that has put forward a host of competing explanations for these seeming inefficiencies in analysts' forecasts.[1] The sheer magnitude of this research effort reflects the importance of security analysts in financial markets. This paper provides a framework that not only encompasses many of the leading explanations for the autocorrelation of analysts' forecast errors but can also be used to quantify their relative contributions. Relying on this framework, we find that the autocorrelations stem primarily from analysts' concern for model misspecification.

In our model, reported earnings equal the true earnings plus noise. This noise represents factors such as earnings smoothing, changes in the recognition of future cash flows, and one-off demand and price shocks. The analyst, however, faces model (or "Knightian") uncertainty about the noise in reported earnings and is concerned that his model is misspecified.[2] We show that the analyst's optimal strategy is to issue "robust" forecasts. The analyst achieves robustness by using a model which *ex post* likely overstates the amount of noise in reported earnings.

The intuition is simple. If the analyst assumes that there is little or no noise in reported earnings, his forecasts will be more *imprecise* than expected if the true model's reported earnings are very

[1] Abarbanell (1991), Mendenhall (1991), and Abarbanell and Bernard (1992), among many others, provide reliable evidence that analysts underreact to new information. To explain this finding, subsequent research explores the cross-sectional variation in analyst underreaction including, for example, the explanatory power of permanent versus transitory components of a firm's earnings Ali, Klein, and Rosenfeld (1992), and analyst specific characteristics such as analyst experience Mikhail, Walther, and Willis (2003). Easterwood and Nutt (1999) suggest that analysts underreact to negative information but overreact to positive information. However, Abarbanell and Lehavy (2003) show that this conclusion is sensitive to how outliers are treated. Shane and Brous (2001) document that analyst underreaction is corrected in subsequent forecasts. Markov and Tamayo (2006) argue that analysts are uncertain about the earnings process and underestimate its persistence. More recently, Williams (2013) investigates a behavioral explanation—"a false consensus effect"—of analyst underreaction.

[2] The distinction between Knightian and Bayesian uncertainty is important. If an agent knows the model but is uncertain about the parameters governing that model, Bayes rule describes how the agent updates his beliefs about the parameters. An agent who faces model uncertainty, by contrast, cannot assign a prior distribution to the universe of alternative models nor perform Bayesian updating to distinguish between models. We use the term "model uncertainty" throughout to refer to Knightian uncertainty about the model governing the earnings process. The term "parameter uncertainty" is reserved for Bayesian uncertainty about the parameters of a known model.

noisy. The analyst loses accuracy not only because there is now more noise in reported earnings and, therefore, the forecastable component is smaller but, importantly, because the analyst relies on a misspecified model to learn from historical data. If, on the other hand, the analyst assumes that the reported earnings are very noisy, then his forecasts will be more *precise* than expected if the true model has little or no noise in reported earnings. Therefore, a robust forecast in effect assumes that reported earnings are noisier than what the analyst's point estimate suggests. As a byproduct, this behavior leaves a trail of positively autocorrelated forecast errors in the data. However, this positive autocorrelation emerges not because the analyst uses information inefficiently but because the analyst guards against model misspecification. Using data on security analysts from Institutional Broker Estimate Systems (IBES), we estimate that model uncertainty accounts for 58% of the autocorrelation in analysts' forecast errors.

Having established the importance of model uncertainty, we proceed to show that two additional mechanisms also contribute to the autocorrelation in analysts' forecast errors. The first mechanism results from the variation in mean forecast errors. To illustrate how this variation generates positively autocorrelated forecast errors, consider, for simplicity, using data on two firms. Suppose that analysts issue too conservative forecasts about firm A and too optimistic forecasts about firm B. A pooled regression will return a positive estimate of the autocorrelation in forecast errors. The reason is that positive errors are likely followed by positive errors, and vice versa, but only because the positive errors are from firm A and the negative errors from firm B. We estimate that one-fifth of the autocorrelation in analysts' forecast errors is due to this variation.

Estimation errors about the persistence of earnings growth shocks also add to the autocorrelation of forecast errors. We show that if analysts receive signals about earnings growth shocks, then both positive and negative estimation errors about the persistence of these shocks will lead to positive autocorrelation in forecast errors. A stylized example illustrates the intuition. Suppose that a firm's earnings growth follows an AR(1) process, $y_t = \phi y_{t-1} + e_t$, with persistence $\phi > 0$ but that the analyst issues forecasts using $\hat{\phi} \neq \phi$. If we assume that the analyst learns the shock e_t before the firm announces earnings, his forecast equals $\hat{y}_t = \hat{\phi} y_{t-1} + e_t$.[3] The process of forecast *errors* is then but a scaled version of the y_t process itself, $\mathrm{FE}_t = y_t - \hat{y}_t = (\phi - \hat{\phi}) y_{t-1}$. The process

[3]We assume in this example that the analyst perfectly learns the shock e_t before the firm announces earnings. In Section 5.2's analysis of this mechanism, we dispose of this assumption.

of forecast errors is therefore positively autocorrelated regardless of whether the analyst over- or underestimates the persistence of earnings shocks:

$$\operatorname{cov}(\mathrm{FE}_t, \mathrm{FE}_{t-1}) = \operatorname{cov}\left[\left(\phi - \hat{\phi}\right) y_{t-1}, \left(\phi - \hat{\phi}\right) y_{t-2}\right] = \left(\phi - \hat{\phi}\right)^2 \phi \operatorname{var}(y_t) > 0. \quad (1)$$

The empirical contribution stemming from this channel is important. Although analysts, on average, use unbiased estimates of the persistence of earnings shocks when issuing forecasts, they have substantial estimation errors in $\hat{\phi}$ on a firm-by-firm basis. These estimation errors in turn generate one-fifth of the autocorrelation in their forecast errors.

Our decomposition results are easy to summarize. Analysts' uncertainty about the earnings process accounts for approximately 60% of the autocorrelation in analysts' forecast errors, with the remainder attributable, in roughly equal parts, to the variation in mean forecast errors and in analysts' estimation errors about the persistence of earnings growth shocks. These results therefore suggest that the uncertainty that analysts face about the firms they follow is of first-order importance in explaining the seeming inefficiencies of their forecasts.

Our results are important also because they reveal what mechanisms are absent from the data. One behavioral explanation for the autocorrelation in forecast errors is that analysts overestimate the precision of the non-earnings signals, and so they collectively do not update sufficiently given an earnings signal.[4] Put differently, their false trust in non-earnings signals crowds out earnings signals. Our model yields estimates of the degree to which analysts utilize both earnings and non-earnings signals and we find no evidence of analysts overestimating the precision of the non-earnings signals. Similarly, we show that learning itself—that is, *changes* in analysts' beliefs—does not generate autocorrelation in forecast errors. Rather, they are autocorrelated because of the uncertainty that analysts have about firms' earnings processes, and learning only reduces autocorrelations as these estimation errors dissipate.

This paper rationalizes the autocorrelations in analysts' forecast errors by arguing that analysts face significant uncertainty about firms' earnings processes. It is important to then ask whether the

[4]See, for example, Chen and Jiang (2006) and the references therein. Chen and Jiang (2006) note that analysts may deviate from efficient weighting not only because of behavioral biases but because they face incentives to do so. Bernhardt, Campello, and Kutsoati (2006) suggest that analysts "anti-herd:" they issue forecasts that overshoot the consensus forecast in the direction of their private signal.

levels of uncertainty implied by our estimates are reasonable. They are. We show that an analyst cannot rule out the possibility that there might be up to four times as much noise in reported earnings as what his point estimate suggests. The detection error probability—which measures the difficulty of distinguishing the "worst case model" from the analyst's best guess—is 17.2% for the model that matches the autocorrelation of forecast errors in the data. To put this result in perspective, the robust control literature argues that a decision maker should be concerned about alternative models whose detection error probabilities are at least 5%.

The assumption that analysts face model uncertainty about features of the earnings process is testable. The idea is simple. If analysts face only parameter uncertainty, then they learn over time and so the autocorrelations in forecast errors gradually dissipate. If analysts, by contrast, treat the uncertainty about the noise in reported earnings as a permanent state of affairs, the amount of autocorrelation emanating from this source will not diminish even after observing decades of earnings data. The time-series behavior of the autocorrelation patterns thus measures how much of the uncertainty is of the kind analysts can resolve through learning. We find that the positive autocorrelation of forecast errors in the data decreases as firms age. This rate of decay, however, also decreases with age, and so significant autocorrelations remain even among the oldest firms. We further show that the uncertainty that analysts resolve is about the persistence of the earnings growth shocks. Consistent with our assumptions, analysts appear to be unable to resolve their uncertainty about the noise in reported earnings.

Taken together, our results paint a more favorable view of security analysts' ability to forecast earnings than what is commonly reported in the literature. That analysts are quite sophisticated, however, is consistent with our empirical evidence of their remarkable forecast accuracy, notwithstanding the fact their forecast errors are predictable. For example, while an ARMA(1,1) model predicts earnings growth rates with an R^2 of 14%, analysts' forecasts explain 80% of this variation! If analysts are so remarkably accurate, it is reasonable to entertain the possibility that the positive autocorrelation of their forecast errors results from analysts' deliberate response to the substantial model uncertainty confronting them.

Our results contribute to three strands of the literature. Uppal and Wang (2003), Maenhout (2004), and Epstein and Schneider (2008) suggest that model uncertainty is of first-order importance for portfolio choice and asset pricing. We show that this conclusion extends to the behavior of security analysts. Our findings also relate to studies that examine the behavior of analysts for reasons other than investigating mechanisms that might drive the positive autocorrelation in forecast errors.[5] A complete model of analyst behavior should incorporate analysts' concerns for model misspecification. Finally, our results relate to studies that draw a connection between analysts and asset pricing anomalies. If analysts underreact to new information, perhaps the same can also be expected of investors, which could then help explain some of the anomalies in asset returns.[6] Our results suggest that the autocorrelation of forecast errors may be unrelated to those anomalies. Any connection between the two may simply then reflect the fact that investors themselves face similar misspecification concerns when pricing assets.

The rest of the paper is organized as follows. Section 2 puts forward our model of earnings and robust forecasting. Section 3 describes the data. Section 4 estimates the joint dynamics of firms' earnings processes and analysts' forecasts, and measures detection error probabilities. Section 5 decomposes the autocorrelation of forecast errors into components stemming from model uncertainty, variation in mean forecast errors, and variation in estimation errors about the persistence of earnings shocks. Section 6 measures the extent to which analysts resolve uncertainty over time. Section 7 concludes.

[5] See, for example, Hong and Kubik (2003) and Hilary and Hsu (2013). They note that understanding analyst behavior is of great interest because there exist considerable demand for analysts' earnings forecasts: investment bankers rely on analysts to procure investment banking deals while brokers rely on analysts to provide research services to attract order flow. The amounts by which security analysts are compensated support this view. The National Bureau of Labor Statistics reports that, in 2012, there were 253,000 financial analysts—buy and sell sides combined—with a median income of $76,950, indicating that the market annually spends approximately $19.5 billion dollars on security analysis.

[6] De Bondt and Thaler (1985), Jegadeesh and Titman (1993), Lakonishok, Shleifer, and Vishny (1994), and Daniel, Hirshleifer, and Subrahmanyam (1998), for example, are proponents of the view that long-term reversal, momentum, and value anomalies in asset prices stem from investors under- or overreacting to new information. There is an extensive literature in finance alone that study the connection between stock price behavior and analyst forecasts. See, for example, Abarbanell and Bernard (1992), La Porta (1996), Diether, Malloy, and Scherbina (2002), Jegadeesh, Kim, Krische, and Lee (2004), Loh and Stulz (2011), and So (2013).

2. Robust Forecasting

2.1. *The earnings process*

We assume that the earnings growth process has both persistent and temporary components:

$$\text{reported earnings growth:} \quad y_{t+1} = \mu + x_{t+1} + \alpha_{t+1}, \tag{2}$$

$$\text{persistent component of earnings growth:} \quad x_{t+1} = \phi x_t + e_{t+1}, \tag{3}$$

where y_{t+1} is the reported earnings growth, x_t is the persistent component of earnings growth and α_t is the noise in reported earnings. We assume that the analyst also receives additional information not contained in earnings numbers. This signal s_t informs the analyst about the shock to the persistent earnings growth component,

$$s_t = e_{t+1} + n_t, \tag{4}$$

where $n_t \sim N(0, \sigma_n^2)$. The shocks α_{t+1}, e_{t+1}, and n_t have zero cross-correlations, autocorrelations, and cross-autocorrelations. The analyst's objective at time t is to predict y_{t+1} using the full history of earnings $\{y_1, y_2, \ldots, y_t\}$ and signals $\{s_1, s_2, \ldots, s_t\}$.

This earnings process is an AR(1)-plus-noise process, which is observationally equivalent to an ARMA(1,1) process.[7] This modeling choice is motivated by the literature on firms' earnings processes. The work in this area—the most prominent papers are those by Griffin (1977), Foster (1977), Brown and Rozeff (1979), and Brown, Hagerman, Griffin, and Zmijewski (1987)—concludes, first, that there is a large seasonal component to earnings and, second, that an ARMA(1,1) model

[7] An ARMA(1,1) model with an autoregressive parameter ϕ, a moving-average parameter θ, and a variance of innovations σ_{e*}^2, can be expressed as an AR(1)-plus-noise model by setting $\sigma_\alpha^2 = -\frac{\theta}{\phi}\sigma_{e*}^2$ and $\sigma_e^2 = (1+\theta^2)\sigma_{e*}^2 - (1+\phi^2)\sigma_\alpha^2$.

provides a good description of the earnings process.[8] We discuss the empirical fit of the AR(1)-plus-noise model in Section 4.

2.2. The uncertainty environment

The reported earnings growth y_t in equation (2) is the sum of the true earnings growth process (denoted by x_t) and a "noise" component (denoted by α_t). We assume that analysts face model (or "Knightian") uncertainty about the noise term and that they know the parameters of the x_t process. We later relax the assumption about the parameters of the x_t process by letting analysts face Bayesian uncertainty about them.

This noise represents factors such as earnings smoothing, the treatment of accruals, changes in the recognition of future cash flows, and other managerial strategic considerations concerning a firm's reported earnings. The assumption that analysts have model uncertainty about α_t can be interpreted quite broadly. Simply put, analysts are unsure of the precise earnings process and realize that this AR(1)-plus-noise model is but an approximation. The α_t term then captures the sum of all the factors left outside the model.

Knightian uncertainty about the noise term has both a direct and indirect effect on the forecasting problem. First, uncertainty about α_t directly affects analysts' perceived accuracy because the reported earnings growth—the object which the analysts forecast—is the sum of x_t and α_t. The greater the analyst's uncertainty about α_t, the less accurate the analyst's forecast will be. Second, since analysts cannot separately identify x_t and α_t, uncertainty about the past values of α_t worsens the analyst's inferences about x_t. The analyst will be more uncertain about current true earnings growth and therefore his forecasts will, on average, be farther off the mark.

The noise term α_t is governed by some probability distribution but the analyst knows neither α_t nor its distribution. We assume that the analyst approximates the distribution of α_t by an IID

[8] The subsequent accounting literature uses either an ARMA(1,1) model or, for simplicity, an AR(1) model, and these choices are often accompanied by references to the studies referenced above. The post-earnings announcement drift study of Bernard and Thomas (1990, p. 313), for example, estimates an ARMA(1,1) model of seasonally adjusted quarterly earnings and states: "Based on prior research, we assume that the most accurate univariate description of the time-series process of earnings is provided by the Brown and Rozeff (1979) model, modified to include a trend term: $Q_t = \delta + Q_{t-4} + \phi(Q_{t-1} - Q_{t-5}) + \theta\varepsilon_{t-4} + \varepsilon_t, \ldots$" Our empirical analyses are also based on seasonally adjusted earnings growth—that is, we define $y_t = Q_t - Q_{t-4}$.

$\mathcal{N}(0, \hat{\sigma}^2_{\alpha^*})$ distribution where we set $\hat{\sigma}^2_{\alpha^*}$ equal to the true variance of α_t, σ^2_α. This assumption ensures that the analyst's initial approximating model is very good.

The analyst accounts for possible model misspecification by always considering what would be the worst possible distribution of α_t. We define this worst case realization as a linear function of α_t^* in the analyst's initial approximating model without model uncertainty:

$$\alpha_t^{(w)} = \kappa_0 + \kappa_1 \alpha_t^*. \tag{5}$$

The realization $\alpha_t^{(w)}$ is therefore an IID random draw from $\mathcal{N}(\kappa_0, \kappa_1^2 \sigma^2_\alpha)$ where κ_0 is an arbitrary real number and $\kappa_1 \geq 0$ is an arbitrary nonnegative real number. The model uncertainty assumption can therefore be cast by saying that the analyst knows neither κ_0 nor κ_1, and, therefore, does not know the mean and variance of $\alpha_t^{(w)}$.

2.3. The robust forecasting problem

Our model is in the spirit of the robust control theory of Hansen and Sargent (2008) and we use their evil-agent device to characterize the analyst's problem. In this zero-sum game, the analyst chooses a forecast that minimizes mean squared error while the evil agent chooses parameters κ_0 and κ_1 to make the analyst look as bad as possible. This game is equivalent to a minimax problem in which the analyst chooses the forecast to minimize mean squared error while choosing κ_0 and κ_1 to maximize that error, anticipating that these would be the choices made by the evil agent. Although the evil agent could then choose different values of κ_0 and κ_1 to hurt the analyst, the unique solution is such that the analyst correctly anticipates the evil agent's choices of κ_0 and κ_1. The analyst's problem is then:

$$\min_{\hat{y}_{t|t-1}} \max_{(\kappa_0,\kappa_1)} \mathrm{E}\left[\left\{ y_t^{(w)} - \hat{y}_{t|t-1} \right\}^2 \,\middle|\, \mathcal{F}_{t-1} \right] \tag{6}$$
$$\text{subject to} \quad \mathrm{E}\left[\left\{ \underbrace{\left(\alpha_t^{(w)} - \alpha_t^*\right)}_{\substack{\text{Deviation of } \alpha_t^{(w)} \\ \text{from } \alpha_t^* \\ \text{(Direct effect)}}} + \underbrace{\left(\hat{x}^{(w)}_{t|t-1} - \hat{x}_{t|t-1}\right)}_{\substack{\text{Perceived bias in } x_t \\ \text{(Indirect effect)}}} \right\}^2 \,\middle|\, \mathcal{F}_{t-1} \right] \leq \eta^2 \sigma^2_\alpha,$$

where the constraint, as discussed below, measures the analyst's concerns for model misspecification. In our notation, $y_t^{(w)}$ is the earnings growth at time t with κ_0 and κ_1 chosen to result in the worst *ex ante* outcome; $\hat{y}_{t|t-1}$ is the analyst's optimal forecast; $\mathcal{F}_{t-1} = \{y_{1:t-1}, s_{1:t-1}, \mu, \phi, \sigma_e^2, \sigma_n^2\}$ represents the analyst's information set; $\hat{x}_{t|t-1}^{(w)}$ is the optimal forecast of x_t using a Kalman filter under the worst case choices of κ_0 and κ_1; and $\hat{x}_{t|t-1}$ is the optimal forecast of x_t under the analyst's expectation of the evil agent's choices of κ_0 and κ_1. As above, $\alpha_t^{(w)}$ is the worst case realization of α_t while α_t^* is the noise under the approximating model.

The left-hand side of the constraint decomposes the effects of the evil agent's distortion into two components. The evil agent tries to maximize these effects given the constraint. The first component, the expected squared deviation of $\alpha_t^{(w)}$ from α_t^*, measures the error induced by the evil agent's distortion of α_t. This is the direct effect of model uncertainty: an analyst's forecasts will be more imprecise than expected if the reported earnings are noisier. The second component captures the extra forecast error in x_t that is created by the evil agent when he distorts α_t. This is the indirect effect of model uncertainty: if the analyst learns from historical data using a misspecified model, the x_t he extracts from the data will deviate further from its true value, thereby increasing the forecast error.

The constraint in the minimax optimization problem penalizes the evil agent for distorting α_t.[9] Without this constraint, an equilibrium would not exist as the evil agent could take an arbitrarily extreme action Hansen and Sargent (2008). The right-hand side of the constraint, $\eta^2\sigma_\alpha^2$, determines the degree of robustness. The parameter η measures the agent's concern for model misspecification. If η is close to zero, the models that the analyst considers are nearly indistinguishable and so the analyst is as likely as not to detect the correct model resulting in a detection error probability of approximately 0.5. However, as η increases, we add models that are increasingly "different" from the approximating model and the detection error probability falls because these models are easier to tell apart in the data.

Two additional remarks are in order. First, the analyst's optimization problem is static; that

[9]Technically, the second component of the constraint in the analyst's optimization problem (equation 6) penalizes the evil agent for the indirect effect of the distortion and guarantees the existence of the equilibrium in the context of "robust filtering without commitment" Hansen and Sargent (2008). Without this second component, the equilibrium still exists in the context of "robust filtering with commitment" if (i) the analyst and the evil agent play the game only once and (ii) κ_0 is known at the beginning.

is, the analyst's forecast $y_{t|t-1}$ is independent of his past forecasts and the same solution applies at every date t. Second, because, by assumption, the analyst knows the parameters of the true earnings process x_t, his choices of $(\hat{\kappa}_0, \hat{\kappa}_1)$—which correspond to what he anticipates the evil agent will choose as (κ_0, κ_1)—completely determine $\hat{y}_{t|t-1}$. That is, after choosing $(\hat{\kappa}_0, \hat{\kappa}_1)$, the analyst obtains the optimal forecast using a Kalman filter. We can thus alternatively express the minimax problem in equation (6) using $(\hat{\kappa}_0, \hat{\kappa}_1)$ as the arguments of the problem's minimization part. Proposition 1 gives the solution to the analyst's forecasting problem:

Proposition 1. *The constrained minimax optimization problem* (6) *has a unique solution* $(\hat{\kappa}_0, \hat{\kappa}_1) = (\kappa_0, \kappa_1)$ *with* $\kappa_0 = 0$ *and* $\kappa_1 = \eta + 1$, *where* $\eta + 1$ *equals the maximum value of* κ_1 *that satisfies the constraint. The optimal forecast* $\hat{y}_{t|t-1}$ *is obtained via a Kalman filter with* $\hat{\kappa}_0 = 0$ *and* $\hat{\kappa}_1 = \eta + 1$. *The worst case outcome is therefore distributed as* $\alpha_t^{(w)} \sim \mathcal{N}\big(0, (\eta+1)^2\sigma_\alpha^2\big)$ *at the solution. This solution is also the pure-strategy equilibrium of the equivalent zero-sum game.*

Proof. See Appendix A. □

Proposition 1 states that the evil agent schemes against the analyst by inflating the noisiness of reported earnings from σ_α^2 to $(\eta+1)^2\sigma_\alpha^2$ but leaves the mean unchanged ($\kappa_0 = 0$). We detail the intuition for why this is the worst case model for the analyst in Section 2.5. In Section 2.6, we discuss how this model relates to asset pricing models in which investors face misspecification concerns.

It is worth noting that this result holds even if the analyst had an asymmetric objective function in which case the optimal forecast will deviate from the mean of the predictive distribution. This tilt would lead to forecasts that seem either pessimistic or optimistic relative to the forecasts of an analyst like ours who has a symmetric objective function Gu and Wu (2003). However, for an equilibrium to exist, it must be that the analyst would perfectly anticipate and offset the evil agent's attempt to shift the mean in light of this asymmetry, thereby eliminating the evil agent's incentive to do so. This argument holds irrespective of the shape of the analyst's objective function.

Of course, this shape clearly affects the location of the analyst's forecasts but not because the evil agent shifts the mean by choosing $\kappa_0 < 0$.[10]

Lemma 2.1 gives the optimal forecast as a function of parameters $\boldsymbol{\theta} = \{\phi, \sigma_\alpha^2, \sigma_e^2, \sigma_n^2\}$.

Lemma 2.1. *The predictive distribution of the earnings growth y_{t+1} given the information set $\mathcal{F}_t = \{y_{1:t}, s_{1:t}, \boldsymbol{\theta}\}$ is a normal distribution with mean and variance of*

$$\begin{aligned}
E[y_{t+1}|\mathcal{F}_t] &= (1-\phi)\mu + \phi\left\{E[y_t|\mathcal{F}_{t-1}] + K\left(y_t - E[y_t|\mathcal{F}_{t-1}]\right)\right\} + ws_t, && (7)\\
&= (1-\phi)\mu + \phi\left\{\hat{y}_t + K\left(y_t - \hat{y}_t\right)\right\} + ws_t,\\
var(y_{t+1}|\mathcal{F}_t) &= \pi_p^2 + \sigma_\alpha^2, && (8)\\
where\ \ w &= \left[1 + \frac{\sigma_n^2}{\sigma_e^2}\right]^{-1},\\
K &= \frac{cov(x_t, y_t|\mathcal{F}_{t-1})}{var(y_t|\mathcal{F}_{t-1})} = \frac{\pi_p^2}{\pi_p^2 + \sigma_\alpha^2},\\
\pi_p^2 &= var(x_{t+1}|\mathcal{F}_t) = \phi^2\pi_u^2 + \left[\sigma_e^{-2} + \sigma_n^{-2}\right]^{-1},\\
\pi_u^2 &= var(x_t|\mathcal{F}_t) = \left[\pi_p^{-2} + \sigma_\alpha^{-2}\right]^{-1}.
\end{aligned}$$

Proof. The results are obtained by applying Bayes theorem and a standard Kalman filter that has converged to the steady state. □

Lemma 2.1 shows that the analyst's forecast is a function of the previous forecast ($\hat{y}_t$), the previous forecast error ($y_t - \hat{y}_t$), and the additional signal s_t. Parameters K and w are important. The Kalman gain K measures how the analyst uses the previous forecast error to revise estimates of x_t. The intuition is that the analyst issued the forecast $\hat{y}_t$ with an estimate of x_t in mind. When

[10]The objective function's shape affects the location of the analyst's forecasts both directly and through its interaction with his misspecification concerns. To illustrate, suppose that, for example, because of an asymmetric objective function, the analyst's forecasting rule is of the form $\hat{y}_t = \mathrm{E}(y_t) - 0.1\mathrm{SD}(y_t)$, where the expectation and standard deviation are conditional on the analyst's information set. The analyst's forecast $\hat{y}_t$ is lower than the mean of predictive distribution even in the absence of misspecification concerns. However, if the analyst is concerned about model misspecification, this concern inflates $\mathrm{SD}(y_t)$ and so lowers the forecast further. Although asymmetry does not affect the mechanism we study, it is important in other models of uncertainty. Orlik and Veldkamp (2014), for example, examine the forecasting behavior of an agent who does not know the parameters governing the GDP process, but learns about these parameters as data accumulate. In this model, uncertainty coupled with the skewness of the GDP growth distribution generate large and counter-cyclical uncertainty fluctuations that match those observed in the data—there is no need to posit an exogenous disaster process to capture these features of the data. We revisit asymmetric objective functions in Section 5 where we examine the variation in mean forecast errors. In that analysis, we note that asymmetry in analysts' objective functions may contribute to variation in mean forecast errors.

the analyst now observes how much the realization deviates from his forecast, this deviation informs him about how far off the mark he was in terms of x_t. The weight w controls how much of the extra signal s_t is used to estimate e_{t+1}. Both K and w are functions of the parameter vector $\boldsymbol{\theta}$. If the parameters used by the analyst differ from the true parameter values, $\boldsymbol{\theta} \neq \hat{\boldsymbol{\theta}}$, then $\hat{K}$ and $\hat{w}$ are still given by Lemma 2.1 but with parameters $\hat{\boldsymbol{\theta}}$. For example, if the analyst sets the noise in reported earnings equal to $(\eta+1)^2\sigma_\alpha^2$ out of fear of model misspecification, this choice decreases the Kalman gain, $\hat{K} < K$.

2.4. Autocorrelation of forecast errors

The autocorrelation of forecast error depends on how the parameters used by the analyst ($\hat{\boldsymbol{\theta}}$) differ from the true parameters ($\boldsymbol{\theta}$). Proposition 2 characterizes the autocorrelation of forecast errors:

Proposition 2. *If the analyst uses parameter values $\hat{\boldsymbol{\theta}}$=($\hat{\phi}$, $\hat{\sigma}_\alpha^2$, $\hat{\sigma}_e^2$, $\hat{\sigma}_n^2$) in the forecasting problem when the true parameters are $\boldsymbol{\theta}$=(ϕ, σ_α^2, σ_e^2, σ_n^2), the analyst's forecast error $FE_t = (y_t - \hat{y}_{t|t-1})$ is a stationary time series and its first-order autocorrelation is given by*

$$corr(FE_{t+1}, FE_t) = \underbrace{\frac{A}{var(FE_t)}(\phi - \hat{\phi})}_{\substack{\text{Estimation} \\ \text{error in } \phi}} + \underbrace{\phi(K - \hat{K})}_{\substack{\text{Difference in} \\ \text{Kalman gain}}} + \underbrace{\phi(1-K)\left[1 - \frac{var(FE_t^*)}{var(FE_t)}\right]}_{\substack{\text{Any deviation from} \\ \text{MSE-optimal forecast}}}, \quad (9)$$

$$\begin{aligned} \textit{where } A &= cov\Big(\hat{K}y_t + \Big(1-\hat{K}\Big)\,\hat{y}_t, FE_t\Big), \\ FE_t^* &= \textit{forecast error when } \hat{\boldsymbol{\theta}} = \boldsymbol{\theta}. \end{aligned}$$

Proof. See Appendix B. □

If the analyst uses $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}$, the autocorrelation collapses to zero. In the robust forecasting model, the analyst knows all other parameters except for the distribution of α_t, and so the only difference between $\boldsymbol{\theta}$ and $\hat{\boldsymbol{\theta}}$ is that the analyst sets $\hat{\sigma}_\alpha^2 = (\eta+1)^2\sigma_\alpha^2$. In this case, the first term in equation (9) vanishes and both the second and third terms are strictly positive. The third term, which is positive as long as $\hat{\boldsymbol{\theta}} \neq \boldsymbol{\theta}$, is negligible relative to the second term.[11]

Figure 1 plots the autocorrelation of forecast errors when the analyst uses "wrong" parameters in the forecasting problem. The upper-right corner, in which $\hat{\sigma}_\alpha^2 \neq \sigma_\alpha^2$, applies to the robust forecasting model. We see that if analysts act as if the reported earnings are very noisy, forecast errors will be positively correlated. Other "mistakes" also affect the autocorrelation function. Forecast errors are positively autocorrelated if analysts either under- or overestimate the persistence of earnings growth shocks (ϕ), underestimate the variance of the permanent growth shocks (σ_e^2), or overestimate the precision of the non-earnings signal (σ_n^{-2}). In Section 4.2, we estimate the deviations between the true parameters and those used by the analysts, and discuss the interpretation of these other mistakes.

2.5. Robust forecasts: Intuition

An analyst concerned about model misspecification will issue forecasts that perform well even under the worst case model in which reported earnings have the highest possible variance. Equation (6) shows that noise in reported earnings hurts the analyst in two ways. First, the noisier reported earnings are, the less accurate the analyst's forecast will be. His task is to forecast *reported* earnings and the amount of variation that can be predicted decreases as the amount of noise increases. Second, the analyst's inferences about x_t depend on how noisy earnings are. The more noise in reported earnings, the farther away the analyst will, on average, be in his assessment of the current value of x_t.

The analyst will, therefore, want to overestimate the amount of noise in reported earnings. Doing so, his realized accuracy will be better than what he expected it to be. Whatever the analyst's forecast, it will be closer to reported earnings because there is less noise in these reports. Also, his assessment of x_t will be closer to the truth because, by definition, reported earnings (y_t) better track true earnings growth (x_t). If, on the other, the analyst has underestimated the amount of noise in reported earnings, both of these channels work against the analyst.

The analyst's desire for robustness comes at a cost, however. By assuming that reported earnings

[11] For example, if $\phi = 0.5$, $\sigma_\alpha = 0.1$, $\sigma_n = 0.5$, and $\sigma_e = 1.0$, then the third term is less than 1% of the second term even if the analyst "overestimates" the amount of noise in reported earnings by a factor of 100.

are noisier than what they probably are, he will, on average, pay less attention to historical earnings numbers. As a result, the analyst puts less weight on these numbers and, by extension, his previous forecast errors. This shift in his weighting scheme results in a positive autocorrelation in forecast errors, giving the appearance that the analyst underreacts to earnings numbers and forecast errors. The analyst's behavior is, nevertheless, intuitive and optimal. If he is unsure of the model generating reported earnings and wants to guard against the risk of making a large mistake, the analyst will rely on historical data only with caution. Doing so, he keeps in mind the possibility that some of the reported earnings numbers might have reflected large errors and, not wanting to be led astray, issues forecasts that are robust to such a possibility.

2.6. Discussion: Robustness in asset pricing versus forecasting

An analyst in our framework is concerned about the amount of noise in the reported earnings he is attempting to forecast. The analyst's objective is to minimize mean squared forecast error in the face of an evil agent when earnings are assumed to follow an AR(1)-plus-noise process. In this context, the worst case model is the maximally noisy model. This result stands in contrast to the asset pricing literature where investors want to make robust investment decisions. The worst case models there differ because investors have different objective functions than analysts and face uncertainty surrounding the consumption growth process.

In the asset pricing literature, the nature of the investor's preferences dictate what constitutes the worst case model. For example, Hansen, Sargent, Turmuhambetova, and Williams (2006) and Anderson, Ghysels, and Juergens (2009) model investors as having time-separable preferences and show that the worst case model is the one in which the consumption growth rate's mean is shifted. Hansen and Sargent (2010) and Bidder and Dew-Becker (2014), on the other hand, model investors as having recursive utility with a preference for early resolution of uncertainty (Kreps and Porteus (1978), Epstein and Zin (1989), and Weil (1990)) and for these investors the worst case model is the one in which the persistence of the consumption growth process is increased.

In a forecasting problem, by contrast, a decision maker has a preference for accuracy. For example, Hansen and Sargent (2008, chapter 17) investigate a forecasting problem similar to ours in which a decision maker minimizes a mean squared error objective function. They conclude that the robust decision rule relies on the standard Kalman filter with "distorted beliefs" in which the decision maker acts as if the process is more volatile than what his approximating model suggests. The evil agent in a forecasting problem does not shift the mean of the process to be forecast. If an equilibrium exists, it must be that the decision maker is able to successfully offset the evil agent's mean shift.[12] Our model is similar to that of Hansen and Sargent (2008) except that, because we draw a distinction between earnings growth shocks and noise in reported earnings, the analyst in our model is concerned about the noise in reported earnings.

Of course, analysts in our model could, in principle, be concerned about other elements of their approximating model. Influenced by these asset pricing results, suppose that an analyst knows all model parameters except for the persistence of earnings growth shocks. If he is concerned about misspecification in ϕ, the analyst will issue forecasts using $\hat{\phi} > \phi$. The use of an inflated ϕ guards against the worst case scenario because the amount of uncertainty about the hidden state x_t—the true earnings growth—increases in ϕ. In Section 4.2, we estimate how the parameters that analysts use to issue their forecasts differ from the parameters of the true earnings process. We find that, on average, $\hat{\phi} \approx \phi$. That is, analysts do not behave in a way that would suggest that they are concerned about misspecification in ϕ.

The distinction between model and parameter uncertainty is important, and this distinction is also evident in the asset pricing literature. For example, Collin-Dufresne, Johannes, and Lochstoer (2013, 2015) show that if investors have recursive preferences, rational parameter learning generates *subjective* long-run risks. The shocks to rational beliefs are permanent and therefore impact consumption growth in all future periods, just as if we had shocks to long-run consumption growth in the standard model without learning. In deriving these results, investors are assumed to know the true model or can, at least, assign a prior on the universe of possible models. However, what they do not know are the values of the parameters governing this model. Anderson, Hansen, and Sargent (2000) delineate between these two approaches by noting that "the robust decision maker

[12]In an investment problem, an investor cannot "offset" the mean shift because future consumption growth is not a choice variable. That is, while the investor can account for the shifted mean in his optimal consumption and investment decisions, he cannot offset the utility loss that is due to the shifted mean. Although the evil agent could hurt the investor by increasing the variance of the consumption growth process, he would rather shift the mean or change the persistence of the process. To the evil agent, shifting the mean or modifying the persistence in these models is always cheaper than inflating the variance. Hansen and Sargent (2008, p. 42) note that, in a diffusion setting in continuous time, the evil agent agent chooses not to distort the volatility because doing so is infinitely costly in terms of relative entropy. See, also, Hansen et al. (2006) and Anderson et al. (2009).

accepts model misspecification as a permanent state of affairs, and devotes his thoughts to designing robust controls, rather than to using data to improve his model specification over time." However, for a decision maker to accept potential misspecification as a permanent state of affairs, he must find it difficult to tell his approximating model apart from the worst case model. We return to this question in Section 4.4 in which the detection error probabilities facing analysts are computed.

3. Data

3.1. Data sources and sample construction

We combine data from IBES, Compustat, and the Center for Research in Securities Prices (CRSP). We use analysts' earnings forecasts and actual earnings per share from January 1984 through December 2013 relying on the detailed history IBES files to avoid the rounding problems associated with the split-adjusted files.[13] To account for the seasonality in earnings, we define y_t as the year-to-year change in quarterly earnings per share, deflated by lagged stock price. A forecast stays in our sample only if it is issued or reconfirmed after the firm announced its previous quarter's earnings. This requirement excludes forecasts that have not been updated to reflect the latest available information. We denote the analysts' median forecast of y_t by $\hat{y}_t$.

We next merge in additional data from Compustat and CRSP. We require the firms in our sample to be matched against CRSP and Compustat and have CRSP share codes of 10 or 11 (common stock) and be listed on the NYSE, Nasdaq, or AMEX. We then impose standard sample selection rules[14] to alleviate the effects of outliers: (1) delete observations in which the beginning-of-quarter stock price is below \$5; (2) delete observations where the forecasted year-to-year change in the quarterly earnings per share is greater than \$10 in absolute value; (3) delete observations with extreme values of $(y_t, \hat{y}_t, y_t - \hat{y}_t)$ by trimming at their 1% and 99% levels; and (4) require a firm to have at least 20 quarterly observations of y_t and $\hat{y}_t$. The last restriction ensures that we have enough data to estimate firm-specific models of earnings and analyst forecasts.[15] The final sample with these restrictions has 185,420 firm-quarter observations on 3,804 firms.

[13] See, for example, Diether et al. (2002) for a description of this problem.

[14] See, for example, De Bondt and Thaler (1990), Lim (2001), Mikhail et al. (2003), and Raedy, Shane, and Yang (2006).

[15] In Table 1 below, we also provide some statistics for these deleted firms which we will call "short-lived firms."

3.2. Descriptive statistics

Table 1 provides descriptive statistics for the year-to-year quarterly earnings growth (y_t), forecasted earnings growth ($\hat{y}_t$), and resultant forecast errors. Analysts' forecasts are, on average, pessimistic. The fractions of negative, zero, and positive forecast errors are 32%, 10%, and 57%, respectively, and so positive errors outnumber negative errors by a two-to-one ratio. Although the average forecast error is close to zero—the heteroscedasticity and autocorrelation consistent t-value associated with the average is 2.07—the median forecast is significantly positive with a t-value of 13.37 based on a simple sign test. The difference between the mean and the median implies that large negative errors are more common than equivalently large positive errors. Lim (2001) and Hilary and Hsu (2013), among others, also find that average forecast errors are positive.

The last rows in Table 1 report autocorrelations in forecasts and forecast errors. We measure autocorrelations using AR(1) regressions:

$$\mathrm{FE}_{i,t+1} = a + \rho \mathrm{FE}_{i,t} + \varepsilon_{i,t+1}. \tag{10}$$

The *pooled* estimate of the autocorrelation in forecast errors, 0.216, is significant with a heteroscedasticity and autocorrelation consistent t-value of 28.87. This estimate is similar to other autocorrelation estimates reported in the literature, both qualitatively and quantitatively. Mendenhall (1991), for example, reports an estimate of 0.276 (t-value = 17.82) while the estimate in Abarbanell and Bernard (1992) is 0.20 (t-value = 15.77). These papers use different datasets and sample periods and so the similarity in these estimates speaks to the robustness of this result.

Table 1 also reports average firm-specific autocorrelations for all firms, short-lived firms, and long-lived firms.[16] The average firm-specific autocorrelation estimate across all 7,153 firms with sufficient data for estimating the autocorrelation is 0.102. These firm-specific estimates, however, are biased downwards Kendall (1954). The intuition for the bias is best illustrated by considering what the autocorrelation estimate would look like when estimated from a time series of two IID random numbers. Because we first estimate the sample average and then measure deviations from

[16] "Short-lived firms" are firms with fewer than 20 quarterly observations of earnings and forecasts. We drop these firms from the main sample which therefore consists of firms for which we have at least 20 quarterly observations ("long-lived firms"). The "all-firms" sample includes both short- and long-lived firms.

the average, a positive deviation must be followed by a negative deviation, or vice versa. That is, it must be that the autocorrelation estimate is negative even though the draws are IID.[17] This bias is the reason most papers in this literature, including those referenced above, rely on estimates from pooled regressions. We follow this tradition but later add fixed and random effects to soak up heterogeneity in the data.

The average bias-corrected autocorrelation estimate for all firms is 0.244. The remaining rows of the table show that, after correcting for the bias, autocorrelations are considerably smaller for long-lived firms than for short-lived firms, 0.196 versus 0.299. The 0.196 estimate for long-lived firms—which are the firms included in the main sample—is close to the pooled estimate of 0.216. In Section 6, we use the decay in autocorrelations, measured at the firm level, to estimate the rate at which analysts resolve uncertainty about the firms they follow.

4. Empirical Analysis of Analysts' Forecasts

4.1. The difficulty of model selection

In Section 2's model, earnings are assumed to follow an AR(1)-plus-noise model which is observationally equivalent to an ARMA(1,1) model. The earnings data strongly favor the ARMA(1,1) model over a MA(1) model. When these two models are taken to the data, the MA(1) model's adjusted R^2 is 11.5% while that of the ARMA(1,1) model is 14%. In a comparison of these two models, a likelihood-ratio test rejects the MA(1) model with a p-value below 0.001. We present the estimates of the AR(1)-plus-noise model for the earnings growth process in Section 4.2.

As discussed in Section 2, the earnings literature often defaults to an ARMA(1,1) specification because no model within this class of linear models matches the autocorrelation function of earnings growth much better than the ARMA(1,1) model.[18] In fact, this model provides an adequate description of the data for most firms in our sample. We can assess model fit by separately estimating this model for each firm using Generalized Method of Moments (GMM) and then testing the resultant over-identifying restrictions. Using the average earnings growth, the variance of

[17] Kendall (1954) shows that if the unadjusted autocorrelation estimated from a time series of T observations is $\hat{\rho}$, then the bias-adjusted estimate is $\hat{\rho}_{\text{bias-adjusted}} = \frac{\hat{\rho}(T-1)+1}{T-4}$. This bias is therefore larger for short-lived firms.

[18] For example, fitting an ARMA(5,1) model to our data gives an R^2 of only 22.6%.

earnings growth, and the autocovariances at the first four lags as the moment conditions, we reject the ARMA(1,1) model for 10.3% of the firms at the 5% level after applying a Bonferroni correction.

Furthermore, even if earnings growth follows a higher-order ARMA process, note that the analysts' information set in our forecasting model consists of an AR(1)-plus-noise model *and* an extra signal s_t. This extra signal can represent any useful information found in the higher-order lags. In fact, if we regress analysts' forecast errors on five lags of earnings growth, the resultant coefficients are all close to zero. This result implies that analysts' forecasts reflect information embedded in the higher-order lags or that those additional lags do not contain useful information.

Finally, it is important to emphasize that the problem confronting the analyst in Section 2's model is not choosing between, say, MA(1) and ARMA(1,1) models. Rather, in our formulation, the analyst assumes an AR(1)-plus-noise model but realizes that it is but an approximation stemming from his uncertainty surrounding the amount of noise in reported earnings. At one extreme, there may be little or no noise, in which case the earnings process is close to an AR(1) model. At the other extreme, many features of the reported earnings may not be captured by the analyst's approximating model, all of which is then subsumed by the noise term. An important consideration is the extent to which the analyst can use earnings data to distinguish this true model from among this spectrum of alternatives. We address this question in Section 4.4.

4.2. A joint model of earnings and analyst forecasts

We can express the dynamics of earnings together with their forecasts as a VARMA(1,1) system

$$Y_{t+1} = A + BY_t + C\varepsilon_{t+1} + D\varepsilon_t, \tag{11}$$

$$\text{where } Y_t = \begin{bmatrix} y_t \\ \hat{y}_t \end{bmatrix}, \quad \varepsilon_t = \begin{bmatrix} \alpha_t \\ e_t \\ n_{t-1} \end{bmatrix}, \quad \text{cov}(\varepsilon_t) = \begin{bmatrix} \sigma_\alpha^2 & 0 & 0 \\ 0 & \sigma_e^2 & 0 \\ 0 & 0 & \sigma_n^2 \end{bmatrix},$$

$$A = \begin{bmatrix} \mu(1-\phi) \\ \hat{\mu}(1-\hat{\phi}) \end{bmatrix}, \quad B = \begin{bmatrix} \phi & 0 \\ \hat{\phi}\hat{K} & \hat{\phi}\left(1-\hat{K}\right) \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 1 & 0 \\ 0 & \hat{w} & \hat{w} \end{bmatrix}, \quad D = \begin{bmatrix} -\phi & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

and μ and $\hat{\mu}$ are the long-term means of y_t and $\hat{y}_t$, respectively.[19]

We estimate this model in two steps. In the first step, we estimate the parameters of the earnings growth process—the AR(1)-plus-noise model—using maximum likelihood. In the second step, we hold these parameter values fixed and use conditional maximum likelihood to estimate the remaining parameters of the VARMA model. We use a block bootstrapping procedure, resampling *firms* to preserve the time-series properties of the data, to compute standard errors and to confirm that the parameter estimates obtained using the two-step procedure are unbiased.[20]

Table 2 reports the parameter estimates and the standard errors of the earnings and forecast processes. It also reports R^2s for the AR(1)-plus-noise model and analysts' forecasts to assess the accuracy of analyst forecasts. The latter compares the variance of forecast errors to the variance of earnings growth,

$$\text{Pseudo-}R^2 \text{ of analyst forecasts} = 1 - \frac{\text{var}(y_{t+1} - \hat{y}_{t+1})}{\text{var}(y_{t+1})}. \tag{12}$$

In our model, the analyst receives additional signals about earnings growth shocks. If analysts use only historical earnings to issue forecasts, the precision of their forecasts should be comparable to the R^2 of 14% of the ARMA(1,1) model. Analysts' forecasts, however, are considerably more accurate: the R^2 of their forecasts is 79.8%. Although the literature often portrays analysts as being poor forecasters[21], this characterization ignores the fact that analysts' forecasts are, on average, substantially more accurate than those obtained from time series models. The estimates of the VARMA(1,1) model capture the precision of these extra signals. This precision affects not only the estimate of σ_n but also other estimates, such as K and w, that depend on this precision.

[19] It is important to note that this VARMA(1,1) system is not a new specification of earnings and forecasts. It merely collects equations (2), (3), (4), and (A.1) (from the proof of Proposition (2)).

[20] Although we assume normality in estimating the VARMA(1,1) model, we verify that the departures from normality do not affect our inferences. We assess the robustness of our results as follows. We start from the parameter estimates of the VARMA(1,1) model reported in Table 2. We then generate a hypothetical data set with the same size as the real data set. However, instead of imputing normal shocks, we draw random shocks e_t, α_t, and n_t from the empirical distribution of the standardized earnings growth. This standardized distribution has mean zero and a standard deviation of 1 but, consistent with the estimated reported in Abarbanell and Lehavy (2003), it is negatively skewed (-0.866) and fat-tailed (kurtosis = 12.46). We then then re-estimate the VARMA(1,1) model using these simulated data, and compare the parameter estimates and standard errors to assess robustness. We find that the estimates are nearly unchanged. For example, consider the VARMA(1,1) estimates of $\sigma_\alpha/\sigma_e = 0.106$ and $\sigma_n/\sigma_e = 0.538$ reported in Table 2. If we simulate data by imputing shocks from the empirical earnings growth distribution, these parameters are nearly unchanged at 0.108 and 0.534, respectively.

[21] See Bradshaw (2011) for a discussion of the analysts-versus-time series models debate.

We also note an observational equivalence result that is useful in interpreting the estimates of the VARMA(1,1) model. Hansen and Sargent (2008, Section 17.8) show an observational equivalence between robust forecasts and ordinary forecasts made under "distorted beliefs." Importantly, this result implies that the estimates of the VARMA(1,1) model can be interpreted without specifying the motivations behind the analysts' forecasts. Put differently, the VARMA(1,1) model recovers from the data (i) the true parameters of the earnings process and (ii) the parameter values that represent distorted beliefs. For example, the result in Table 2 that $\phi \approx \hat{\phi}$, which we discuss below, holds regardless of whether or not the analyst fears model misspecification.

4.3. Which "mistakes" drive the autocorrelation in forecast errors?

The estimates in Table 2 assess how the true parameters of the earnings process differ from those relied upon by analysts and help us pinpoint the sources for the autocorrelation in forecast errors. Recall from Proposition 2 and Figure 1, forecast errors are positively autocorrelated if: (i) analysts either under- or overestimate the persistence of earnings growth shocks, ϕ; (ii) underestimate the variance of the earnings growth shocks, σ_e^2; (iii) overestimate the precision of the non-earnings signal, σ_n^{-2}; or (iv) overestimate the amount of noise in reported earnings, σ_α^2. Robust forecasting operates through the last of these channels with analysts optimally using "too high" a value of σ_α^2 to guard against the worst case scenario.

Table 2 recovers estimates of $\hat{\rho}$, $\hat{K}$, and $\hat{w}$ from the data. These parameters, in turn, depend on analysts' choices for the primitive parameters $(\hat{\phi}, \hat{\sigma}_\alpha, \hat{\sigma}_e, \hat{\sigma}_n)$. By equation (7), the analyst's optimal forecast is $\hat{y}_{t+1} = (1 - \hat{\phi})\hat{\mu} + \hat{\phi}\left\{\hat{y}_t + \hat{K}(\mathrm{FE}_t)\right\} + \hat{w}s_t$, where $\hat{\phi}$ represents the analyst's belief about the persistence of the earnings growth shocks, $\hat{K}$ summarizes his belief of the informativeness of the reported earnings growth, and $\hat{w}$ is the analyst's measure of the additional signal's informativeness. Because w is the weight the analyst places on the additional signal, the analyst would incorrectly estimate this parameter if he was overconfident in his private information ($\hat{w} \gg w$) or if he herded with other analysts and ignored the additional signal ($\hat{w} \ll w$). Also, the higher the value of σ_α the analyst uses, the lower the Kalman gain estimate $\hat{K}$.

The estimates in Table 2 are consistent with the autocorrelation in forecast errors stemming solely from the underestimated Kalman gain K. By contrast, the values of $\hat{\phi}$ and $\hat{w}$ that analyst rely on are remarkably close to their true values. This finding implies that analysts, on average, have correct beliefs about the precision of the additional signal, σ_n^2, and the variance of the shocks to the persistent component of earnings growth, σ_e^2.[22] These estimates suggest that the autocorrelation of forecast errors is therefore not due to some of the mechanisms that have been proposed in the literature. In Markov and Tamayo (2006), for example, the autocorrelation of forecast errors is positive because analysts underestimate the persistence of the earnings shocks, $\hat{\phi} \ll \phi$. Similarly, studies documenting that analysts herd or put excessive weight on their private signals[23] argue that the autocorrelation of forecast errors may stem from such mechanisms. The fact that $\hat{w} - w \approx 0$ in the data rules out these channels.

The underestimated Kalman gain K can stem from three different types of "mistakes." Analysts could overestimate the precision of the additional signal, $\hat{\sigma}_n^2 \ll \sigma_n^2$; they could underestimate the variance of the permanent growth shocks, $\hat{\sigma}_e^2 \ll \sigma_e^2$; or they could overestimate the noisiness of the reported earnings, $\hat{\sigma}_\alpha^2 \gg \sigma_\alpha^2$. However, the fact that $\hat{w} - w \approx 0$ implies that the underestimated Kalman gain can only result from analysts overestimating σ_α^2. This behavior is consistent with our model in which the analyst optimally makes this choice to ensure robustness against model misspecification. The Kalman gain used by analysts ($\hat{K} = 0.414$), however, is substantially lower than the actual Kalman gain ($K = 0.953$). Are analysts' concerns about model misspecification sufficiently large to lower the Kalman gain to such a degree? We turn to this question next.

4.4. Detection error probabilities

4.4.1. Definition

As discussed earlier, robust control theory assumes that the decision maker accepts model misspecification as a permanent state of affairs. The alternative models considered as plausible by the decision maker must therefore lie close to his approximating model. Otherwise, if the worst case model were too different from his approximating model, the decision maker could easily reject

[22]Because w is a function of the ratio σ_n^2/σ_e^2, the estimation results technically also permit for the possibility that analysts have wrong beliefs about both terms but that these mistakes are exactly of the same magnitude and thus offset each other in the ratio so that $w - \hat{w} \approx 0$. We assume in the discussion that follows that this knife's edge solution is an unlikely explanation for the autocorrelation of the forecast errors.

[23]See, for example, Trueman (1994), Welch (2000), Cooper, Day, and Lewis (2001), Chen and Jiang (2006), and Jegadeesh and Kim (2010).

it and thereby invalidate the assumption that he fears model misspecification. The detection error probability measures the likelihood that the decision maker errs in detecting the correct model and provides a measure of model uncertainty confronting him.

To calibrate the amount of model uncertainty confronting analysts, we follow Hansen, Sargent, and Wang's (2002) notation in defining these probabilities. Suppose that the analyst has two models in mind: the approximating model (A) and the worst case model (W). The analyst can use earnings data to formulate two likelihood ratio tests: (1) the probability of rejecting A when A is the correct model, and (2) the probability of rejecting W when W is the correct model. The two models are difficult to distinguish if the probability of mistakingly rejecting the correct model is high Maenhout (2004).

The detection error probability depends on the "distance" between the approximating model A and the worst case model W. In our formulation, this distance is controlled by the constraint parameter η (see equation 6). Attaching equal prior weights to A and W, the detection error probability $p(\eta)$ is the average probability of the two mistakes:

$$p(\eta) = \frac{1}{2} \left(\Pr(\text{mistake} \mid \text{A}) + \Pr(\text{mistake} \mid \text{W}) \right). \tag{13}$$

The higher $p(\eta)$, the more difficult it is for the analyst to distinguish between the models and so the more likely he errs in detecting the correct model. If η is close to zero, the worst case model is nearly indistinguishable from the approximating model and so the analyst is as likely as not to make an error in detecting the correct model, giving $p(\eta) \approx 0.5$. If, on the other hand, η is large, the worst case model will likely be easy to distinguish from the approximating model and so $p(\eta) \approx 0$. The robust control literature typically assumes that a decision maker wants to guard against alternative models whose detection probability is 5% or higher.

To determine the amount of uncertainty confronting the analyst, we must compute the detection error probabilities of alternative models and find the set of those that exceed this 5% threshold. However, detection error probabilities also depend on the amount of data relied upon by the analyst: the larger the sample, all else being equal, the more likely the analyst will detect the correct model and so the lower the detection error probability. Therefore, before turning our attention to computing detection error probabilities, we must measure the amount of data at analysts' disposal when forecasting earnings.

4.4.2. Measuring the effective sample size

Our data set has 185,420 firm-quarter observations with a median of 42 quarters of observations per firm. If analysts have access to similar data and use it to estimate firms' earnings processes, what is the amount of useful data per firm? The effective sample size lies between two extremes. At one extreme, if firms' earnings processes are completely heterogenous, then analysts can only use historical data specific to each firm. At the other extreme, if the parameters are identical across firms, then analysts can use the entire pooled data set.

We estimate the effective sample size as follows. We first assume that analysts can learn about a model parameter from other firms and possibly by using other additional information. We can then measure the uncertainty that analysts have about that parameter to back out the effective sample size. If, for example, analysts in the data issue forecasts using a very precise estimate of ϕ, that precision implies that the effective sample size is large. By contrast, if their estimate of ϕ is noisy, the effective sample size is small.

Panel A of Table 3 presents estimates of the estimation errors in long-term growth rates, measured by $\mathrm{SD}(\mu - \hat{\mu})$, and those in ϕ, as measured by $\mathrm{SD}(\phi - \hat{\phi})$.[24] We measure these standard deviations by estimating mixed-effects models of the earnings growth and forecast processes. In these models, both the intercepts (μ and $\hat{\mu}$) and autoregressive parameters (ϕ and $\hat{\phi}$) can vary across firms and as a function of calendar time and firm age. Appendix D details these estimation procedures.

To estimate the effective sample size, we begin with our estimate of $\mathrm{SD}(\phi - \hat{\phi}) = 0.0875$ in Panel A. Since the standard error of $\hat{\phi}$ when estimating the AR(1)-plus-noise model depends on the expected asymptotic Fisher information matrix and sample size, we can find that sample size N for which the standard error of $\hat{\phi}$ matches the 0.0875 estimate. In our data, the resultant effective sample size is $N = 465$ observations and measures the average amount of information that

[24] The parameter μ is the long-term growth rate of reported earnings in the AR(1)-plus-noise model (see equations 2 and 3). The standard deviation $\mathrm{SD}(\mu - \hat{\mu})$, however, also measures the variation in mean forecast errors, because $\mu - \hat{\mu} = \mathrm{E}(y_t) - \mathrm{E}(\hat{y}_{t \mid t-1}) = \mathrm{E}(y_t - \hat{y}_{t \mid t-1}) = \mathrm{E}(\mathrm{FE}_t)$, and therefore $\mathrm{SD}(\mu - \hat{\mu}) = \mathrm{SD}\left(\mathrm{E}(\mathrm{FE}_t)\right)$. We thus henceforth call $\mathrm{SD}(\mu - \hat{\mu})$ the "the variation in mean forecast errors."

an analyst relies upon when forecasting earnings. Since our estimate of the effective sample size exceeds the median number of observations available per firm, analysts appear to avail themselves to data on other firms when issuing their forecasts.[25]

4.4.3. Estimates of detection error probabilities

Panel B of Table 3 tabulates detection error probabilities.[26] We simulate data from an AR(1)-plus-noise model in which the parameters of the approximating model are those reported in Table 2, while the parameters of the corresponding worst case model are the same except that the amount of noise is inflated to the level shown in the column labeled σ_α/σ_e. The top part of Panel B computes detection error probabilities by simulating samples of $N = 465$ observations, the effective sample size previously estimated. Subsequently, we vary the effective sample size when computing the detection error probabilities.

The starting point in Panel B is the row $\sigma_\alpha/\sigma_e = 0.106$, which corresponds to the approximating model from Table 2. The worst case model here corresponds to the analyst's approximating model. In this case, the analyst faces no model uncertainty. The constraint parameter η is zero to indicate that there are no alternative models and the detection error probability is 50% because we are comparing two identical models. The autocorrelation of forecast errors must be zero because the parameters used by the analyst ($\hat{\boldsymbol{\theta}}$) coincide with the true parameter values ($\boldsymbol{\theta}$).

Notice that as we increase σ_α while holding the effective sample size fixed, detection error probabilities decrease because the analyst now entertains models that are increasingly different from his approximating model. For example, if we approximately double the noise σ_α so that $\sigma_\alpha/\sigma_e = 0.2$ in Panel B, the detection error probability is reduced to 38.3%. Correspondingly, η,

[25] We could also measure the effective sample size using the estimate of $\mathrm{SD}(\mu - \hat{\mu})$. Following the same procedure, the effective sample size becomes $N = 341$ observations. We use the higher estimate of $N = 465$ to be conservative; in Section 5.1 (see footnote 30), we note that variation in mean forecast errors may also reflect other mechanisms in addition to estimation error.

[26] Detection error probabilities, the constraint parameter η, the amount of noise in reported earnings under the worst case model ($\sigma_\alpha^{(w)}$), and the autocorrelations are all linked together. First, Proposition 1 shows that η relates σ_α to $\sigma_\alpha^{(w)}$ by $\sigma_\alpha^{(w)} = (\eta+1)\sigma_\alpha$. Second, Proposition 2 gives the autocorrelation of forecast errors as a function of the difference between the true parameters ($\boldsymbol{\theta}$) and those used by the analyst ($\hat{\boldsymbol{\theta}}$). If the analyst uses the true values for all parameters but σ_α—as when concerned with misspecification—this proposition relates $\sigma_\alpha^{(w)}$ to the autocorrelation of forecast errors. Third, detection error probabilities are determined as well because they are computed by comparing the data through the approximating model (using σ_α) and the worst case model (using $\sigma_\alpha^{(w)}$). These connections are important. If we fix any one of them—η, $\sigma_\alpha^{(w)}$, the autocorrelation of forecast errors, *or* the detection error probability—we fix the other parameters as well.

which controls the distance between the approximating and the worst case models has increased to 0.89. The autocorrelation in forecast errors is now positive because the analyst is concerned that there might be more noise in reported earnings than suggested by his approximating model. By Proposition 2, this positive serial correlation in forecast errors reflects the increase in σ_α that lowers Kalman gain K.[27]

The detection error probability remains at least 5%, the threshold for which the analyst should guard against possible misspecification, up until $\sigma_\alpha/\sigma_e = 0.438$ for $N = 465$ observations. At this threshold, the analyst faces so much model uncertainty that he is plausibly concerned that the amount of noise in reported earnings could be as much as four times as high as in his approximating model. If we set this model as the worst case model, the analyst's desire for robustness leads him to issue forecasts that have an autocorrelation of 0.184.

Recall that the pooled estimate of the autocorrelation in analysts' forecast errors is 0.216. From Panel B, this amount of autocorrelation is consistent with a detection error probability of only 1.7%, smaller than the 5% threshold, indicating that model uncertainty would not be an issue as the analyst should feel comfortable in ruling out this alternative model. This computation, however, is too conservative as it assumes all of the estimated autocorrelation is due to model uncertainty. In Section 5, we show that just over half of the 0.216 estimate—0.125, to be precise—is due to analysts' concerns for model misspecification. Panel B shows the worst case model that generates an autocorrelation of 0.125 (on row $\sigma_\alpha/\sigma_e = 0.330$) has a detection error probability of 17.2%, well above the 5% threshold.

The last two rows of Panel B fix the autocorrelation in forecast errors at 0.125 but change the amount of data available to an analyst. These computations are important because the effective sample size estimate of 465 observations applies to the average forecast about the average firm. If the analyst is only able to make use of a smaller sample because of a firm's uniqueness, he would be less likely to to distinguish between the approximating and worst case models. For example, in Panel B, if the effective sample size is only 200 observations, then the detection error probability of the $\sigma_\alpha/\sigma_e = 0.330$ model increases to 26.7%. Conversely, if the analyst can avail himself to a

[27] We can also use detection error probabilities to investigate how likely an analyst would be able to distinguish an AR(1) model ($\sigma_\alpha = 0$) as his approximating model from an ARMA(1,1) model with parameters as those reported in Table 2. The corresponding detection error probability is 45.4% indicating that the two models are nearly indistinguishable in our data and highlighting just how much uncertainty an analyst has about the earnings process.

larger sample, his ability to distinguish between the approximating and worst case models increases. However, from Panel B note that the effective sample size would have to increase by a factor of three, to 1,417 observations, before the detection error probability drops to the 5% threshold.[28]

5. Decomposing the autocorrelation in forecast errors

To clearly discern the role of model uncertainty in explaining the observed autocorrelation of forecast errors requires that we purge the effects of all other contributing mechanisms. In addition to analysts' misspecification concerns, two mechanisms contribute to the autocorrelation of forecast errors. Section 5.1 shows that variation in mean forecast errors generate positive autocorrelation of forecast errors. Section 5.2 shows that both positive and negative estimation errors about the persistence of earnings growth shocks can also add to the autocorrelation of forecast errors.

After detailing these mechanisms, we decompose the total autocorrelation of forecast errors into three parts: analysts' concerns for model misspecification and the contributions from these two mechanisms. The autocorrelation estimate of 0.216 in Table 1 is the sum of these three parts. By assuming parameter homogeneity, however, the VARMA(1,1) model in Table 2 attributes all of this autocorrelation to analysts' concerns for model misspecification by lowering the Kalman gain $\hat{K}$.[29]

[28] Although the estimates in Table 2 suggest that analysts are not concerned about misspecification in ϕ, we could use the same methodology as in Panel B of Table 3 to measure detection error probabilities for models indexed by ϕ. For example, the detection error probability comparing the approximating model with $\phi = 0.472$ (from Table 2) to a model with $\phi = 0.6$ is 1.6%.

[29] See Appendix D.

5.1. *Autocorrelation from the variation in the mean forecast error*

Table 1 shows that analysts are, on average, too pessimistic in their forecasts and there is substantial variation in their forecast errors.[30] Before providing a formal proof, Figure 2 illustrates how this variation induces positive autocorrelation in forecast errors. In this figure, we assume that forecast errors are IID draws from normal distributions with different means for two groups: +0.1 for group 1 and −0.1 for group 2. Their standard deviations are both 0.1. The within-group autocorrelations are zero because of the IID assumption. In a pooled regression, however, the autocorrelation of forecast errors is 0.5. The intuition is simple: a positive error is likely followed by another positive error, and vice versa, but only because observations are drawn from groups with different means. Proposition 3 formalizes this result:

Proposition 3. *Suppose that the observations of forecast errors $FE_{i,t}$ in panel data divide into $m = 1, 2, \ldots, M$ groups. The fraction of observations from group m is w_m so that $\sum_{m=1}^{M} w_m = 1$. Each group includes firms whose forecast errors are drawn from a covariance stationary process with group-specific mean forecast error $E[FE_{m,t}] = b_m$. Let $\rho^*_{pooled} = corr(FE_{t+1}, FE_t)$ denote the pooled first-order autocorrelation of forecast errors when there is no variation in mean forecast errors, $var(b_m) = 0$. The pooled first-order autocorrelation of forecast errors ρ_{pooled} when $var(b_m) \geq 0$ is*

$$\begin{aligned}
\rho_{pooled} &= (1-\lambda)\cdot\rho^*_{pooled} + \lambda\cdot 1 \geq \rho^*_{pooled}, \ \ \textit{where} \qquad (14)\\
\lambda &= \frac{var(b_m)}{var(FE_t)}, \ \ 0 \leq \lambda \leq 1, \ \ \textit{and} \qquad (15)\\
var(b_m) &= \sum_{m=1}^{M} w_m b_m^2 - \left(\sum_{m=1}^{M} w_m b_m\right)^2.
\end{aligned}$$

[30] The literature has put forth several mechanisms to explain the variation in mean forecast errors. Analysts may, for example, systematically underestimate the growth rate of earnings, that is, $\hat{\mu} < \mu$. Alternatively, analysts may strategically issue low forecasts to gain better access to management and to thus increase the accuracy of their forecasts Lim (2001). Hong and Kubik (2003) and Jackson (2005) show that analysts who are more optimistic relative to the consensus are more likely to experience favorable job separations and to generate more trade for their brokerage firms. Hilary and Hsu (2013) find that analysts who are more consistent in their forecast errors tend to give more pessimistic forecasts, and that the market rewards this behavior by paying more attention to such analysts. Gu and Wu (2003) note that if analysts' loss function is not quadratic, then analysts' forecasts may be biased even though they are efficient. Moreover, as discussed in Section 2.3, analysts' objective functions could be asymmetric. Analysts, for example, may prefer to understate earnings to generate investment banking revenue, and this preference for positive forecast errors leads to forecasts that an outsider might view as being unjustifiably pessimistic. We do not attempt to distinguish between these alternatives but instead measure to what extent variation in mean forecast errors induces autocorrelation in forecast errors.

Proof. See Appendix C. □

A "group" in Proposition 3 can be firms, time periods, or a combination of these. If mean forecast errors vary in any of these dimensions, the estimated autocorrelation is biased upwards. For example, if the autocorrelation of forecast errors in the absence of this variation is $\rho^*_{\text{pooled}} = 0$, then the bias equals

$$\text{corr}(\text{FE}_{t+1}, \text{FE}_t) = \frac{\text{var}(b_m)}{\text{var}(\text{FE}_t)} = \frac{\text{var}(\mu - \hat{\mu})}{\text{var}(\text{FE}_t)}.$$

That is, the autocorrelation becomes a measure of how much of the total variation in forecast errors emanates from the variation in mean forecast errors. This is important. If analysts are accurate but issue systematically too low or high forecasts for some firms, estimated autocorrelations can be significantly upward biased.

5.2. *Autocorrelation from estimation errors in the persistence of the earnings growth shocks*

Analysts' estimation errors about the persistence of earnings growth shocks ϕ also contribute to the observed autocorrelation of forecast errors. This effect is due to the first term of Proposition 2 which can be expressed as

$$\frac{\text{cov}(\hat{K}y_t + (1-\hat{K})\hat{y}_t, \text{FE}_t)}{\text{var}(\text{FE}_t)}(\phi - \hat{\phi}). \tag{16}$$

If $\phi > \hat{\phi}$, then the covariance term is always positive, and so the autocorrelation of forecast errors is increasing in $\phi - \hat{\phi}$. If, alternatively, $\phi < \hat{\phi}$, this covariance becomes *negative* if the analyst's additional signal is sufficiently precise.[31] Therefore, if the additional signal is sufficiently precise, both positive and negative estimation errors in $\hat{\phi}$ add to the autocorrelation of forecast errors. The stylized example in the introduction, summarized in equation (1), illustrates this result

[31] The covariance term can be written as

$$\begin{aligned}
\text{cov}\big(\hat{x}_{t|t}, y_t - \hat{y}_t\big) &= \text{cov}\Big(\hat{x}_{t|t}, \phi y_{t-1} + e_t + \alpha_t - \phi\alpha_{t-1} - \hat{\phi}\hat{x}_{t-1|t-1} - \hat{w}(e_t + n_{t-1})\Big) \\
&= \text{cov}\Big(\hat{x}_{t|t}, (\phi - \hat{\phi})\hat{x}_{t-1|t-1} + \phi(x_{t-1} - \hat{x}_{t-1|t-1}) + \alpha_t + (1-\hat{w})e_t - \hat{w}n_{t-1}\Big) \\
&= \Big(\phi - \hat{\phi}\Big)\,\text{cov}\big(\hat{x}_{t|t}, \hat{x}_{t-1|t-1}\big) + \phi\text{cov}\big(\hat{x}_{t|t}, x_{t-1} - \hat{x}_{t-1|t-1}\big) + \text{cov}\big(\hat{x}_{t|t}, (1-\hat{w})e_t - \hat{w}n_{t-1}\big)\,.
\end{aligned}$$

As the additional signal on e_t becomes increasingly precise, $\sigma_n \to 0$, the first term converges to $(\phi - \hat{\phi})\text{cov}(x_t, x_{t-1}) = \phi(\phi - \hat{\phi})\text{var}(x_t)$, and the second and third terms converge to zero because $\hat{x}_{t|t} \to x_t$, $\hat{w} \to 1$, and $n_t \to 0$.

under the assumption that the additional signal is perfect. In that example, the autocorrelation of forecast errors increases in $(\phi - \hat{\phi})^2$. Panel A of Figure 1 clearly illustrates how this result depends on the precision of the additional signal. When the additional signal is sufficiently precise, the autocorrelation of forecast errors is a U-shaped function of $\phi - \hat{\phi}$. Therefore, even if $\mathrm{E}(\phi - \hat{\phi}) \approx 0$—that is, analysts *on average* correctly estimate the persistence of earnings growth shocks—variation in $\phi - \hat{\phi}$ will contribute to the autocorrelation in their forecast errors.

5.3. Decomposition estimates

Sections 5.1 and 5.2 show that variation in mean forecast errors and in estimation errors in $\hat{\phi}$ both contribute to our pooled estimate of the autocorrelation of forecast errors (0.216). Table 4 decomposes this pooled estimate into three components: (1) the autocorrelation due to the variation in mean forecast errors, (2) the autocorrelation due to estimation errors in $\hat{\phi}$, and, by subtracting these two additional components from the overall pooled estimate, (3) the autocorrelation due to analysts' concerns for model misspecification.

This decomposition is based on the estimates given in Panel A of Table 3. There we report the amount of variation in mean forecast errors, $\mathrm{SD}(\mu - \hat{\mu})$, and in estimation errors, $\mathrm{SD}(\phi - \hat{\phi})$. Estimating how much of the pooled autocorrelation estimate of 0.216 is due to the variation in mean forecast errors is straightforward. We estimate the pooled regression given by equation (10) while letting the intercept vary across firms and as a function of calendar time and firm age. The change in the slope coefficient measures the contribution of the mean-forecast error channel. Similarly, we measure the contribution of the estimation errors in $\hat{\phi}$ but this computation is slightly more involved because we need to iterate for a fixed point. We provide the computational details in Appendix D.

The estimates in Table 4 show that the variation in mean forecast errors accounts for approximately one-fifth (19.7%) of the autocorrelation in forecast errors. That is, while the original pooled autocorrelation estimate is 0.216, it decreases to 0.173 once we remove the effects of calendar time, age, and cross-sectional variation across firms in mean forecast errors. Most (71%) of this additional autocorrelation emanates from the cross-sectional variation across firms. That is, analysts are, on average, pessimistic or optimistic about some firms, and these differences contribute to the autocorrelation in forecast errors. It should be noted that this "firm" effect captures both analysts' uncertainty about the long-term growth rate μ and cross-sectional variation to the extent to which analysts purposefully issue too low or high forecasts. The calendar-time effect—aggregate mean forecast errors trend upwards during the sample period—contributes one-quarter (28%) of the additional autocorrelation. This effect captures, for example, any persistent market-wide factors in earnings surprises; for example, macroeconomic shocks that surprise all analysts. The effect of firms aging is negligible.

The estimation errors-in-ϕ channel is equally important. Table 4 shows that total variation in $\phi - \hat{\phi}$ also accounts for approximately one-fifth (21.9%) of the pooled autocorrelation estimate. That is, the pooled autocorrelation estimate decreases by 0.047 once these effects are removed from the data. This decrease is almost entirely due to the cross-sectional variation in $\phi - \hat{\phi}$. The calendar-time and age effects in $\phi - \hat{\phi}$ account for just 2% of the total reduction in autocorrelation.

The last row of Table 4 shows that after removing the effects of mean forecast errors and estimation errors in $\hat{\phi}$, the autocorrelation of 0.125 that remains stems from analysts' concerns for model misspecification. These concerns are warranted. Recall from Panel B of Table 3 that an autocorrelation of 0.125 in analysts' forecast errors is consistent with a detection error probability of 17.2%, well above the 5% threshold. Therefore, the analyst will find it difficult to distinguish his approximating model from this worst case model in which earnings are three times as noisy. The autocorrelation of 0.125 that we observe reflects analysts guarding against this very real possibility of model misspecification.

6. Learning and its Limits

Underlying our analysis to this point is the assumption that analysts cannot learn about the noisiness of a firm's reported earnings. We now turn our attention to empirically investigating this claim. As a byproduct, we also shed light on which features of the earnings process analysts appear to learn about.

6.1. *Changes in autocorrelations as firms age*

Table 5 measures the change in autocorrelations as firms age. We use a two-stage procedure. We first estimate for each firm-quarter the autocorrelation of forecast errors and then run a linear regression to explain variation in these estimates

$$\widehat{\text{corr}}(\text{FE}_{i,t}, \text{FE}_{i,t-1}) = \beta_0 + \beta_t + \gamma \text{Age}_{i,t-1} + \mathbf{X}_{i,t-1}\boldsymbol{\beta} + e_{i,t}, \tag{17}$$

where the dependent variable is estimated using a five-year window centered around quarter t and adjusted for the Kendall (1954) bias; β_t are calendar-time fixed effects; and firm age is measured as the number of years since a firm's first appearance on CRSP, entered linearly or in logs. Because firm age is potentially confounded with other variables that can influence the autocorrelation of forecast errors, we include as controls in $\mathbf{X}_{i,t-1}$ the standard deviation of analyst forecasts scaled by the prior quarter's stock price, book-to-market, firm size, analyst coverage, and the standard deviation of daily stock returns. We compute heteroscedasticity and autocorrelation consistent standard errors because the dependent variable is constructed using overlapping data. Because we estimate the dependent variable (autocorrelations) in equation (17) on a firm-by-firm basis, the estimates are free of the issues discussed in Section 5.1.

Table 5 shows that the autocorrelation of forecast errors decreases significantly as firms age. The log-specification of age fits the data better than the linear version. These results suggests that autocorrelations decrease as firms age but at a decreasing rate and are thus consistent with analysts learning about *something*: analysts learn more from the first observation than from the second, more from the second than the third, and so forth. The estimates in the second column are useful for quantifying the magnitude of this aging effect. The age variables are scaled so that their slopes represent the reduction in autocorrelations over the first 50 years of a firm's life. The point estimate of β_0 indicates that the average autocorrelation at the time of birth is 0.21. The log-age coefficient of -0.054 then implies that, over the first 50 years of a firm's life, analysts would learn enough about its earnings process to decrease the autocorrelation of forecast by 26%. This learning pattern is also consistent with the simple descriptive statistics reported in Table 1: autocorrelations are considerably higher for short-lived firms than what they are for long-lived firms.

These results suggest that there are limits to what analysts can learn. Although analysts learn quickly at the outset, a firm would need to exist for hundreds of years for the autocorrelations in analysts' forecast errors to vanish. This limits-to-learning result is consistent with our assumption that analysts can learn about certain features of the earnings process but not about others.[32]

6.2. What do analysts learn about?

Our assumption that analysts can learn about the persistence of earnings growth shocks but not about the noisiness of reported earnings is empirically testable.[33] We begin by measuring the amount of uncertainty analysts have about the persistence of the earnings growth shocks, ϕ, as a function of firm age. We follow the same procedure as that used in Panel A of Table 3. The only difference is that, instead of estimating $\mathrm{SD}(\phi - \hat{\phi})$ using the full data set, we estimate this standard deviation for firms of different age using five-year windows around every quarter t.

Panel A of Figure 3 shows that analysts' precision about ϕ improves significantly as firms age. The estimates show that with 15 years of quarterly data, analysts' uncertainty about ϕ is approximately one-half of what it was at the time of the first observation, and the estimation errors continue dissipating afterwards.

Panel B of Figure 3 confirms that the autocorrelation of forecast errors decreases because estimation errors about ϕ attenuate—and not because analysts learn about the noisiness of reported earnings and $\hat{K}$ changes. This panel plots how much of the autocorrelation is attributable to analysts' uncertainty about ϕ versus the extent to which it results from model uncertainty ($\hat{K} < K$). These estimates are scaled by their age-zero values. The y-axis therefore measures the amount of autocorrelation that stems from each channel relative to that channel's importance at birth. We

[32] In unreported results, we find that autocorrelations attenuate faster among firms that go public at a young age. We excluded from the sample firms that satisfy the following two requirements: (1) they have their IPO during the sample period and (2) they are below the median in age at the time of their IPO relative to the founding dates provided by Jay Ritter at http://bear.warrington.ufl.edu/ritter/FoundingDates.htm. Both the intercepts and the slopes on $\log(1 + \mathrm{Age})$ are closer to zero in this restricted sample than they are in Table 5's full sample. These results suggest that analysts are more uncertain about the earnings processes of firms that go public at a young age.

[33] It is important to note that analysts could also learn about μ in which case mean forecast errors might dissipate over time and thus autocorrelations would diminish. Table 4, however, shows that mean forecast errors in the pooled estimation do not vary predictably as a function of firm age. Moreover, as described above, the regressions in Table 5 shut down this channel by estimating firm-by-firm regressions in the first stage. The remaining age-autocorrelation patterns must therefore depend on K or ϕ.

use the same procedure as earlier (see Appendix D) to estimate the amount of autocorrelation that is due to each channel.

The estimates in Panel B show that analysts are unable to learn about the noisiness of reported earnings. The amount of autocorrelation due to this component is very similar across the very youngest and oldest firms. Although there is a dip in the middle, such a temporary dip—even if not a chance finding—is not consistent with analysts resolving uncertainty about the noisiness of reported earnings. The dashed line, by contrast, shows that the autocorrelation of forecast errors falls as analysts learn about the persistence of earnings growth shocks.

7. Conclusion

Analysts guard against model misspecification by relying on a model which *ex post* likely overstates the noisiness of reported earnings. Intuitively, if the true earnings process is noisy but the analyst erroneously assumes that there will be little or no noise in reported earnings, his resultant forecasts will be grossly inaccurate. By contrast, if the analyst assumes that reported earnings are noisy when the true process actually has no noise, his resultant forecasts will be remarkably accurate. An analyst therefore achieves robustness by exaggerating the noisiness of reported earnings. This behavior leaves a trail of positively correlated forecast errors. This autocorrelation pattern, however, does not imply that analysts use information inefficiently.

Our empirical tests are consistent with this mechanism explaining a large part of the autocorrelation of analysts' forecast errors. Two additional mechanisms, stemming from variation in mean forecast errors and estimation errors in the persistence of earnings growth shocks, also contribute to the autocorrelation of analysts' forecast errors.

The assumptions underlying our model are supported by the data. First, the data confirm that an analyst would struggle to distinguish his approximating model even from a model that has four times as much noise in reported earnings. Analysts thus have reason to be concerned about model misspecification. Second, the accuracy of analysts' forecasts greatly exceeds those obtained from time-series models, suggesting that they are privy to additional non-earnings signals. Third, we show that analysts learn about the persistence of earnings growth shocks and that this learning reduces the autocorrelation of forecast errors. By contrast, the amount of autocorrelation that emanates from the robustness channel remains nearly unchanged as firms age.

Our results suggest a number of directions for future research. First, investors face the same model uncertainty that confronts analysts. Their predictions of future cash flows depend on how they interpret new earnings numbers and so, like analysts, investors also need to take into account model misspecification. An interesting exercise would be to use analyst forecasts to measure model uncertainty and then embed this measure into asset pricing tests.

Second, future research should further explore the role of model uncertainty versus parameter uncertainty in explaining the behavior of decision makers. In the case of analysts' forecasting behavior, a model based on parameter uncertainty, similar to that in Orlik and Veldkamp (2014), can also generate "underreaction". In particular, consider an analyst facing uncertainty about the noise in reported earnings α_t such that his beliefs about α_t are unbiased. If we choose a prior such that the Kalman gain K is a concave function of the unknown parameters then, because of Jensen's inequality, $\hat{K} < K$. The analyst now "underreacts" to new information. This effect can be further amplified by making the distribution of α_t skewed and by giving the analyst an asymmetric objective function. Skewness would then play the same role as analysts' misspecification concerns in our model in altering the analysts forecasting behavior.[34] The difference between the model uncertainty and parameter uncertainty approaches is that the former provides a clear rationale for why $\hat{K} < K$—analysts do so to ensure that their forecasts are reasonably accurate even if their model is misspecified.

Third, our theoretical results apply to all model-based forecasting tasks. The same underreaction pattern is found in non-security analyst settings as well. Deschamps and Ioannidis (2013), for example, survey the evidence on professional forecasters' GDP forecasts and note that "forecasters on average underreact to new information" (p. 146). Some of the explanations for this pattern in the macroeconomics literature evoke arguments about incentives and the fact that forecasters may not be compensated solely on the basis of their accuracy. Our model posits an alternative. If forecasters are uncertain about the model generating GDP numbers and want to guard against using a possibly misspecified model, they will issue forecasts that "underreact" to new information.

[34] We thank the referee for pointing out this alternative mechanism.

Appendix

Appendix A. Proof of Proposition 1

We use the following Lemma to prove Proposition 1:

Lemma A.1. *Given the evil agent's choice* (κ_0, κ_1)*, the analyst obtains the optimal forecast* $\hat{y}_{t|t-1}$ *using the Kalman filter with* $(\hat{\kappa}_0, \hat{\kappa}_1) = (\kappa_0, \kappa_1)$*. Any solution to the constrained minimax optimization* (6) *will therefore satisfy* $(\hat{\kappa}_0, \hat{\kappa}_1) = (\kappa_0, \kappa_1)$*.*

We now prove Proposition 1.

Proof. The objective function is expressed as follows:

$$\begin{aligned}
\mathrm{E}\Big[\{y_t^{(w)} - \hat{y}_{t|t-1}\}^2 \mid \mathcal{F}_{t-1}\Big] &= \mathrm{E}\left[\left\{(y_t^{(w)} - \hat{y}_{t|t-1}^{(w)}) + (\hat{y}_{t|t-1}^{(w)} - \hat{y}_{t|t-1})\right\}^2 \,\middle|\, \mathcal{F}_{t-1}\right] \\
&= \mathrm{E}\left[\left\{y_t^{(w)} - \hat{y}_{t|t-1}^{(w)}\right\}^2 \,\middle|\, \mathcal{F}_{t-1}\right] + \left(\hat{y}_{t|t-1}^{(w)} - \hat{y}_{t|t-1}\right)^2 \\
&= \left(\hat{\pi}_p^{(w)}\right)^2 + \kappa_1^2\sigma_\alpha^2 + \left(\hat{x}_{t|t-1}^{(w)} - \hat{x}_{t|t-1} + \kappa_0 - \hat{\kappa}_0\right)^2
\end{aligned}$$

where $\mathcal{F}_{t-1} = \{y_{1:t-1}, s_{1:t-1}, \boldsymbol{\theta}\}$, $(\hat{\pi}_p^{(w)})^2$ is $\mathrm{var}(x_t^{(w)}|\mathcal{F}_{t-1})$ after the Kalman filter has converged to the steady state, and $\hat{y}_{t|t-1}^{(w)}$ is the optimal forecast about y_t using the Kalman filter under the evil agent's worst-case choices of κ_0 and κ_1. The constraint is expressed as follows:

$$\begin{aligned}
\mathrm{E}\left[\left\{(\hat{x}_{t|t-1}^{(w)} - \hat{x}_{t|t-1}) + (\alpha_t^{(w)} - \alpha_t^*)\right\}^2 \,\middle|\, \mathcal{F}_{t-1}\right] &= \mathrm{E}\left[\left\{(\hat{x}_{t|t-1}^{(w)} - \hat{x}_{t|t-1} + \kappa_0) + (\kappa_1 - 1)\alpha_t^*\right\}^2 \,\middle|\, \mathcal{F}_{t-1}\right] \\
&= \left(\hat{x}_{t|t-1}^{(w)} - \hat{x}_{t|t-1} + \kappa_0\right)^2 + (\kappa_1 - 1)^2\sigma_\alpha^2 = \eta^2\sigma_\alpha^2 - \xi,
\end{aligned}$$

where $\xi \geq 0$ is the slack in the constraint.

Candidate Solution. Suppose there is a solution for the constrained minimax optimization problem (6). Lemma (A.1) states that the solution satisfies $(\hat{\kappa}_0, \hat{\kappa}_1) = (\kappa_0, \kappa_1)$. The objective function and the constraint in the optimization problem (6) therefore reduce at the solution to

$$\left(\hat{\pi}_p^{(w)}\right)^2 + \kappa_1^2\sigma_\alpha^2 \quad \textit{subject to} \quad \kappa_0^2 + (\kappa_1 - 1)^2\sigma_\alpha^2 = \eta^2\sigma_\alpha^2 - \xi.$$

Note that the terms $(\hat{\pi}_p^{(w)})^2$ and $\kappa_1^2\sigma_\alpha^2$ are both increasing functions of κ_1 and independent of κ_0.

An evil agent who maximizes this objective function therefore chooses $\kappa_1 = \bar{\kappa}_1$ where $\bar{\kappa}_1$ is the maximum of κ_1 satisfying the constraint. This maximum $\kappa_1 = \bar{\kappa}_1$ is attained when $\kappa_0 = 0$ and $\xi = 0$ since η and σ_α^2 are constant.[35] The only candidate solution is therefore $(\hat{\kappa}_0, \hat{\kappa}_1) = (\kappa_0, \kappa_1) = (0, \bar{\kappa}_1)$, and we have $\bar{\kappa}_1 = \eta + 1$ from the constraint with the candidate solution.

Verification. We now verify that the candidate solution $(\hat{\kappa}_0, \hat{\kappa}_1) = (\kappa_0, \kappa_1) = (0, \bar{\kappa}_1)$ is optimal for the analyst and the evil agent and that neither player has an incentive to deviate. Suppose that the analyst chooses $(\hat{\kappa}_0, \hat{\kappa}_1) = (0, \bar{\kappa}_1)$. The evil agent's objective function becomes

$$\left(\hat{\pi}_p^{(w)}\right)^2 + \kappa_1^2\sigma_\alpha^2 + \left(\hat{x}_{t|t-1}^{(w)} - \hat{x}_{t|t-1} + \kappa_0\right)^2.$$

The expression for the constraint remains the same although the value of $\hat{x}_{t|t-1}$ is chosen by the analyst. Replacing the last term in the evil agent's objective function by the constraint, we rewrite the objective function as follows:

$$(2\kappa_1 - 1)\sigma_\alpha^2 + \left(\hat{\pi}_p^{(w)}\right)^2 + \eta^2\sigma_\alpha^2 - \xi.$$

The first and second terms are increasing functions of κ_1 and independent of κ_0. The third term is constant and the last term is maximized at $\kappa = \bar{\kappa}_1$. The objective function is thus maximized at $\kappa_1 = \bar{\kappa}_1$ and $\xi = 0$, which implies that $\kappa_0 = 0$. The evil agent therefore does not deviate at the candidate solution. The analyst will choose $(\hat{\kappa}_0, \hat{\kappa}_1) = (0, \bar{\kappa}_1)$ given the evil agent's choice $(\kappa_0, \kappa_1) = (0, \bar{\kappa}_1)$ as shown in Lemma (A.1). The candidate solution is therefore the unique solution for the constrained minimax optimization (6). □

[35] Note that even if κ_1 is allowed to be negative, the minimum κ_1 satisfying the constraint is $\underline{\kappa}_1 = 1 - \bar{\kappa}_1$, and the value of the objective function at $\underline{\kappa}_1$ is smaller than the maximum when $\underline{\kappa}_1 < 0$. We assume that $\kappa_1 \geq 0$ because the analyst views $\hat{\kappa}_1 = \kappa_1$ and $\hat{\kappa}_1 = -\kappa_1$ the same.

Appendix B. Proof of Proposition 2

Lemma B.1. *The analyst's forecast for y_{t+1} at time t is its predictive mean given the information set $\hat{\mathcal{F}}_t = \{y^t, s^t, \hat{\theta}\}$:*

$$
\begin{aligned}
\hat{y}_{t+1|t} = E[y_{t+1}|\mathcal{F}_t] &= (1-\hat{\phi})\hat{\mu} + \hat{\phi}\left\{E[y_t|\mathcal{F}_{t-1}] + \hat{K}\cdot(y_t - E[y_t|\mathcal{F}_{t-1}])\right\} + \hat{w}\cdot s_t, \qquad \text{(A.1)}\\
\textit{where } \hat{w} &= \frac{\hat{\sigma}_e^2}{\hat{\sigma}_e^2 + \hat{\sigma}_n^2}\\
\hat{K} &= \frac{cov(x_t, y_t|\hat{\mathcal{F}}_{t-1})}{var(y_t|\hat{\mathcal{F}}_{t-1})} = \frac{\hat{\pi}_p^2}{\hat{\pi}_p^2 + \hat{\sigma}_\alpha^2}\\
\hat{\pi}_p^2 &= var(x_{t+1}|\hat{\mathcal{F}}_t) = \hat{\phi}^2\hat{\pi}_u^2 + \left[\hat{\sigma}_e^{-2} + \hat{\sigma}_n^{-2}\right]^{-1}\\
\hat{\pi}_u^2 &= var(x_t|\hat{\mathcal{F}}_t) = \left[\hat{\pi}_p^{-2} + \hat{\sigma}_\alpha^{-2}\right]^{-1}.
\end{aligned}
$$

Proof. Replacing θ by $\hat{\theta}$ in Lemma 2.1 proves the lemma. The predictive mean of y_{t+1} is the point forecast because the analyst's objective is to minimize the mean squared error. □

Proof of Proposition 2

Proof. First, let FE_t denote the forecast error when the analyst uses the Kalman filter with the analysts' estimates $\hat{\theta}$ of the parameter θ. Define Γ_0 as the variance of forecast errors, $\mathrm{var}(\mathrm{FE}_t)$. Similarly, let FE_t^* denote the forecast error when the analyst uses the Kalman filter with the correct parameter values, that is, $\hat{\theta} = \theta$. The variance of this forecast error, $\Gamma_0^* = \mathrm{var}(\mathrm{FE}_t^*)$, is therefore the minimum value that $\Gamma_0 = \mathrm{var}(\mathrm{FE}_t)$ can take since the forecast based on $\hat{\theta} = \theta$ is unbiased and minimizes the mean squared errors. That is,

$$\Gamma_0 = \mathrm{var}(\mathrm{FE}_t) \geq \mathrm{var}(\mathrm{FE}_t^*) = \Gamma_0^*.$$

Second, the equation for K in Lemma 2.1 implies $(1-K)\Gamma_0^* = \sigma_\alpha^2$ when $\hat{\theta} = \theta$ because

$$1 - K = 1 - \frac{\pi_p^2}{\pi_p^2 + \sigma_\alpha^2} = \frac{\sigma_\alpha^2}{\pi_p^2 + \sigma_\alpha^2} = \frac{\sigma_\alpha^2}{\mathrm{var}(\mathrm{FE}_t^*)} = \frac{\sigma_\alpha^2}{\Gamma_0^*}.$$

The dynamics of earnings growth y_t and analysts' forecasts $\hat{y}_t$ can be expressed as

$$
\begin{aligned}
y_{t+1} &= (1-\phi)\mu + \phi y_t + e_{t+1} + \alpha_{t+1} - \phi\alpha_t, \\
\hat{y}_{t+1} &= (1-\hat{\phi})\hat{\mu} + \hat{\phi}\left(1-\hat{K}\right)\hat{y}_t + \hat{\phi}\hat{K}y_t + \hat{w}(e_{t+1}+n_t).
\end{aligned}
$$

We then calculate the first-order autocovariance function as follows:

$$
\begin{aligned}
\Gamma_1 = \mathrm{cov}(\mathrm{FE}_{t+1}, \mathrm{FE}_t) &= \mathrm{cov}(y_{t+1} - \hat{y}_{t+1}, y_t - \hat{y}_t) \\
&= \mathrm{cov}\Big(\phi y_t + e_{t+1} + \alpha_{t+1} - \phi\alpha_t - \hat{\phi}\Big(1-\hat{K}\Big)\hat{y}_t - \hat{\phi}\hat{K}y_t - \hat{w}(e_{t+1}+n_t), \mathrm{FE}_t\Big) \\
&= \mathrm{cov}\Big(\Big(\phi - \hat{\phi}\hat{K}\Big)y_t - \hat{\phi}\Big(1-\hat{K}\Big)\hat{y}_t, y_t - \hat{y}_t\Big) - \phi\sigma_\alpha^2 \\
&= \mathrm{cov}\Big(\phi\Big(1-\hat{K}\Big)\mathrm{FE}_t + (\phi-\hat{\phi})\hat{K}y_t + (\phi-\hat{\phi})\Big(1-\hat{K}\Big)\hat{y}_t, \mathrm{FE}_t\Big) - \phi\sigma_\alpha^2 \\
&= \phi\Big(1-\hat{K}\Big)\mathrm{var}(\mathrm{FE}_t) - \phi\sigma_\alpha^2 + (\phi-\hat{\phi})\mathrm{cov}\Big(\hat{K}y_t + \Big(1-\hat{K}\Big)\hat{y}_t, \mathrm{FE}_t\Big) \\
&= \phi\Big(1-\hat{K}\Big)\Gamma_0 - \phi\sigma_\alpha^2 + (\phi-\hat{\phi})A \\
&= \phi\Big(1-\hat{K}\Big)\Gamma_0 - \phi(1-K)\Gamma_0^* + (\phi-\hat{\phi})A \\
&= \phi(1-K)(\Gamma_0 - \Gamma_0^*) + \phi(K-\hat{K})\Gamma_0 + (\phi-\hat{\phi})A,
\end{aligned}
$$

where $A \triangleq \mathrm{cov}\Big(\hat{K}y_t + \Big(1-\hat{K}\Big)\hat{y}_t, \mathrm{FE}_t\Big) = \mathrm{cov}\big(\hat{x}_{t|t}, \mathrm{FE}_t\big)$. Finally, we compute the first-order autocorrelation as follows:

$$
\mathrm{corr}(\mathrm{FE}_{t+1}, \mathrm{FE}_t) = \frac{\Gamma_1}{\Gamma_0} = \frac{A}{\Gamma_0}(\phi-\hat{\phi}) + \phi\Big(K-\hat{K}\Big) + \phi(1-K)\left[1 - \frac{\Gamma_0^*}{\Gamma_0}\right],
$$

which is equation (9) in Proposition 2. The variance of forecast errors and the term A can be computed as:

$$
\begin{aligned}
\Gamma_0 = \mathrm{var}(\mathrm{FE}_t) &= \mathrm{var}(y_t - \hat{y}_t) = \mathrm{var}(y_t) + \mathrm{var}(\hat{y}_t) - 2\mathrm{cov}(y_t, \hat{y}_t), \\
A &= \hat{K}\mathrm{var}(\mathrm{FE}_t) + \mathrm{cov}(y_t, \hat{y}_t) - \mathrm{var}(\hat{y}_t), \text{ where}
\end{aligned}
$$

$$\begin{aligned}
\mathrm{var}(y_t) &= \mathrm{var}(x_t) + \sigma_\alpha^2 = \frac{\sigma_e^2}{1-\phi^2} + \sigma_\alpha^2, \\
\mathrm{cov}(y_t, \hat{y}_t) &= \phi\hat{\phi}\left(1-\hat{K}\right)\mathrm{cov}(y_t, \hat{y}_t) + \phi\hat{\phi}\hat{K}\mathrm{var}(y_t) - \phi\hat{\phi}\hat{K}\sigma_\alpha^2 + \hat{w}\sigma_e^2 \\
&= \frac{1}{1-\phi\hat{\phi}\left(1-\hat{K}\right)}\left\{\phi\hat{\phi}\hat{K}\left(\mathrm{var}(y_t) - \sigma_\alpha^2\right) + \hat{w}\sigma_e^2\right\}, \\
\mathrm{var}(\hat{y}_t) &= \hat{\phi}^2\left(1-\hat{K}\right)^2\mathrm{var}(\hat{y}_t) + \hat{\phi}^2\hat{K}^2\mathrm{var}(y_t) + \hat{w}^2(\sigma_e^2+\sigma_n^2) + 2\hat{\phi}^2\hat{K}\left(1-\hat{K}\right)\mathrm{cov}(y_t, \hat{y}_t) \\
&= \frac{1}{1-\hat{\phi}^2\left(1-\hat{K}\right)^2}\left\{\hat{\phi}^2\hat{K}\left(\hat{K}\mathrm{var}(y_t) + 2\left(1-\hat{K}\right)\mathrm{cov}(y_t, \hat{y}_t)\right) + \hat{w}^2(\sigma_e^2+\sigma_n^2)\right\}.
\end{aligned}$$

□

Appendix C. Proof of Proposition 3

Proof.

$$\begin{aligned}
\mathrm{E}\left[\mathrm{FE}_{m,t+1}\mathrm{FE}_{m,t} \mid m^{\mathrm{th}}\text{ group}\right] &= \mathrm{E}\left[\mathrm{FE}_{m,t+1} \mid m^{\mathrm{th}}\text{ group}\right]\mathrm{E}\left[\mathrm{FE}_{m,t} \mid m^{\mathrm{th}}\text{ group}\right] \\
&\quad + \mathrm{cov}\left(\mathrm{FE}_{m,t+1}, \mathrm{FE}_{m,t} \mid m^{\mathrm{th}}\text{ group}\right) = b_m^2 + \Gamma_{m,1}.
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}[\mathrm{FE}_t] &= \mathrm{E}\left\{\mathrm{E}\left[\mathrm{FE}_{m,t} \mid m^{\mathrm{th}}\text{ group}\right]\right\} = \mathrm{E}[b_m]. \\
\mathrm{cov}(\mathrm{FE}_{t+1}, \mathrm{FE}_t) &= \mathrm{E}[\mathrm{FE}_{t+1}\mathrm{FE}_t] - \mathrm{E}[\mathrm{FE}_{t+1}]\,\mathrm{E}[\mathrm{FE}_t] \\
&= \mathrm{E}\left\{\mathrm{E}\left[\mathrm{FE}_{m,t+1}\mathrm{FE}_{m,t} \mid m^{\mathrm{th}}\text{ group}\right]\right\} - \{\mathrm{E}[b_m]\}^2 \\
&= \mathrm{E}\left[b_m^2 + \Gamma_{m,1}\right] - \{\mathrm{E}[b_m]\}^2 \\
&= \mathrm{var}(b_m) + \mathrm{E}[\Gamma_{m,1}] = \mathrm{var}(b_m) + \mathrm{E}[\rho_m\Gamma_{m,0}]. \\
\mathrm{var}[\mathrm{FE}_t] &= \mathrm{E}\left[(\mathrm{FE}_t)^2\right] - \{\mathrm{E}[\mathrm{FE}_t]\}^2 = \mathrm{E}\left\{\mathrm{E}\left[(\mathrm{FE}_{m,t})^2 \mid m^{\mathrm{th}}\text{ group}\right]\right\} - \{\mathrm{E}[b_m]\}^2 \\
&= \mathrm{E}\left[b_m^2 + \Gamma_{m,0}\right] - \{\mathrm{E}[b_m]\}^2 = \mathrm{E}\left[b_m^2\right] - \{\mathrm{E}[b_m]\}^2 + \mathrm{E}[\Gamma_{m,0}] \\
&= \mathrm{var}(b_m) + \mathrm{E}[\Gamma_{m,0}].
\end{aligned}$$

$$
\begin{aligned}
\operatorname{corr}(\mathrm{FE}_{t+1}, \mathrm{FE}_t) &= \frac{\operatorname{cov}(\mathrm{FE}_{t+1}, \mathrm{FE}_t)}{\operatorname{var}(\mathrm{FE}_t)} = \frac{\operatorname{var}(b_m) + \mathrm{E}[\rho_m \Gamma_{m,0}]}{\operatorname{var}(\mathrm{FE}_t)} \\
&= \frac{\operatorname{var}(b_m)}{\operatorname{var}(\mathrm{FE}_t)} + \frac{\mathrm{E}[\rho_m \Gamma_{m,0}]}{\mathrm{E}[\Gamma_{m,0}]} \frac{\mathrm{E}[\Gamma_{m,0}]}{\operatorname{var}(\mathrm{FE}_t)} \\
&= \frac{\operatorname{var}(b_m)}{\operatorname{var}(\mathrm{FE}_t)} \cdot 1 + \mathrm{E}\left[\rho_m \cdot \frac{\Gamma_{m,0}}{\mathrm{E}[\Gamma_{m,0}]}\right] \left(1 - \frac{\operatorname{var}(b_m)}{\operatorname{var}(\mathrm{FE}_t)}\right) \\
&= (1-\lambda) \cdot \rho^*_{\text{pooled}} + \lambda \cdot 1. \\
\text{where } \rho^*_{\text{pooled}} &= \mathrm{E}\left[\rho_m \cdot \frac{\Gamma_{m,0}}{\mathrm{E}[\Gamma_{m,0}]}\right] \equiv \sum_{m=1}^{M} \rho_m \cdot \frac{w_m \Gamma_{m,0}}{\mathrm{E}[\Gamma_{m,0}]}, \\
\mathrm{E}\left[\Gamma_{m,0}\right] &\equiv \sum_{m=1}^{M} w_m \Gamma_{m,0}, \quad \rho_m \equiv \operatorname{corr}(\mathrm{FE}_{m,t+1}, \ \mathrm{FE}_{m,t} \mid m^{\text{th}} \text{ group}) = \frac{\Gamma_{m,1}}{\Gamma_{m,0}}, \\
\Gamma_{m,0} &\equiv \operatorname{var}(\mathrm{FE}_{m,t} \mid m^{\text{th}} \text{ group}), \quad \Gamma_{m,1} \equiv \operatorname{cov}(\mathrm{FE}_{m,t+1}, \ \mathrm{FE}_{m,t} \mid m^{\text{th}} \text{ group}), \\
\lambda &= \frac{\operatorname{var}(b_m)}{\operatorname{var}[\mathrm{FE}_t]} = \frac{\operatorname{var}(b_m)}{\operatorname{var}(b_m) + \mathrm{E}[\Gamma_{m,0}]}, \text{ and} \\
\operatorname{var}(b_m) &\equiv \sum_{m=1}^{M} w_m b_m^2 - \left(\sum_{m=1}^{M} w_m b_m\right)^2.
\end{aligned}
$$

Note that ρ^*_{pooled} is the weighted average of group-specific first-order autocorrelation of forecast errors ρ_m. The weights $\frac{w_m \Gamma_{m,0}}{\sum_{m=1}^{M} w_m \Gamma_{m,0}}$ are proportional to each group's size (w_m) and unconditional variance ($\Gamma_{m,0}$). □

Appendix D. Decomposing the autocorrelation of analysts' forecast errors and estimating variation in mean forecast errors and estimation errors in $\hat{\phi}$

The *observed* autocorrelation of analysts' forecast errors, ρ_{pooled}, arises not only because of analysts' concerns for model misspecification but also because of the variation in mean forecast errors and variation in errors estimating ϕ:

$$
\rho_{\text{pooled}} = \rho_\mu + \rho_\phi + \rho_r, \tag{A.2}
$$

where ρ_μ, ρ_ϕ, and ρ_r are the contributions to the observed autocorrelation due to variation in mean forecast errors, variation in firm-level estimation errors in $\hat{\phi}$, and analysts' concern for model misspecification, respectively.

The contributions of the ρ_μ and ρ_ϕ components reflect parameter heterogeneity. This decomposition, therefore, cannot be addressed in the context of our pooled VARMA(1,1) estimation which assumes parameter homogeneity. Rather, it would require a VARMA(1,1) model with parameter heterogeneity. However, the estimation and statistical properties of such a model, unfortunately, have not heretofore been investigated. Our empirical strategy then is to obtain separately unbiased estimates of each component's contribution to the pooled autocorrelation in equation (A.2). We only need to obtain estimates of ρ_μ and ρ_ϕ. The autocorrelation that is due to analysts' misspecification concerns, ρ_r, is what remains after removing ρ_μ and ρ_ϕ from ρ_{pooled}.

We use a recursive procedure to decompose the pooled first-order autocorrelation of forecast errors ρ_{pooled}. We fix the estimates of VARMA(1,1) model from Table 2 except for $\hat{K}$, and then iterate for a fixed point. These iterations are required because $\hat{K}$ in the VARMA(1,1) model is biased downwards if $\text{SD}(\mu - \hat{\mu}) > 0$ or $\text{SD}(\phi - \hat{\phi}) > 0$.[36] Our recursive method proceeds as follows:

(a) Pick an initial value of ρ_r and compute the implied value of $\hat{K}$ using Proposition 2;

(b) Estimate $\text{SD}(\mu - \hat{\mu})$ and $\text{SD}(\phi - \hat{\phi})$ with mixed models (detailed below), and then compute the values of ρ_μ and ρ_ϕ. These values are computed by simulating data from the VARMA(1,1) model together with the value of $\hat{K}$ from step (a);

(c) Update $\rho_r = \rho_{\text{pooled}} - \rho_\mu - \rho_\phi$ and recompute the implied value of $\hat{K}$;

(d) Repeat steps (b) and (c) until ρ_r and $\hat{K}$ converge.

Below, we describe how we estimate $\text{SD}(\phi - \hat{\phi})$, $\text{SD}(\mu - \hat{\mu})$, ρ_ϕ, and ρ_μ in this procedure.

Estimation of $\text{SD}(\phi - \hat{\phi})$. To estimate $\text{SD}(\phi - \hat{\phi})$ in step (b), we first simplify the VARMA(1,1) model by setting $\sigma_\alpha^2 = 0$ so that the demeaned earnings growth y_t and analyst forecast $\hat{y}_{t\,|\,t-1}$ have

[36] We confirm this property of the pooled estimator of the VARMA(1,1) model via the simulations described below.

the following dynamics:[37]

$$y_t = \phi y_{t-1} + \epsilon_t, \quad (A.3)$$

$$\hat{y}_{t\,|\,t-1} = \hat{\phi} y_{t-1} + \hat{w} s_{t-1}. \quad (A.4)$$

The difference between equations (A.4) and (A.3) is

$$\text{FE}_t = (\phi - \hat{\phi}) y_{t-1} + \varepsilon_t, \quad (A.5)$$

where $\varepsilon_t = \epsilon_t - \hat{w} s_{t-1}$. We can therefore estimate $\text{SD}(\phi - \hat{\phi})$ by running the regression in equation (A.5) by letting the coefficient on y_{t-1} to vary across firms (a random effect) and as a function of calendar time and firm age (fixed effects). The iterations of steps (b) and (c) adjust for any bias in $\text{SD}(\phi - \hat{\phi})$ that emerges from approximating the AR(1)-plus-noise model by an AR(1) model when $\rho_r \neq 0$.[38]

Estimation of $\boldsymbol{\rho_\phi}$. The ρ_ϕ component of the pooled autocorrelation estimate depends not only on $\text{SD}(\phi - \hat{\phi})$ but also $\text{SD}(\phi)$ and $\text{SD}(\hat{\phi})$. Following the same mixed-effects procedure used to estimate $\text{SD}(\phi - \hat{\phi})$ from equation (A.5), we estimate $\text{SD}(\phi)$ from equation (A.3) and $\text{SD}(\hat{\phi})$ from equation (A.4). That is, we again let the coefficients on y_{t-1} to vary across firms (a random effect) and as a function of calendar time and firm age (fixed effects).

We now repeatedly simulate a panel data set with the same dimensions as the actual data set and compute the pooled autocorrelation of forecast errors. In these simulations, we use the estimates of $\text{SD}(\phi)$, $\text{SD}(\hat{\phi})$, and $\text{SD}(\phi - \hat{\phi})$ that we just computed; we use the current value of $\hat{K}$ in the recursive procedure; and we set the remaining parameters to the values given in Table 2. We continue these simulations until we obtain a sufficiently precise autocorrelation estimate. In this procedure, the resulting autocorrelation estimate, by construction, equals $\rho_\phi + \rho_r$. Given the current estimate of ρ_r from step (a), we can therefore back out the ρ_ϕ component of the pooled autocorrelation estimate. We can further decompose ρ_ϕ into three different sources of variation

[37] We demean the time-series by estimating equation (A.5) with calendar-time and firm-age fixed effects and firm random effects. We use *random* rather than fixed firm effects to avoid the Kendall (1954) bias.

[38] We re-estimate $\text{SD}(\phi - \hat{\phi})$ within each loop of the recursive procedure because the size of the bias in $\text{SD}(\phi - \hat{\phi})$ depends on the ρ_r component of the pooled autocorrelation estimate.

(across firms, calendar time, and firm age) by setting $\mathrm{SD}(\phi) = \mathrm{SD}(\hat{\phi}) = \mathrm{SD}(\phi - \hat{\phi}) = 0$ selectively for each source and by measuring the resulting changes in the autocorrelation estimate.

Estimation of $\mathrm{SD}(\mu - \hat{\mu})$ and ρ_μ. We can estimate $\mathrm{SD}(\mu - \hat{\mu})$ and its contribution to the autocorrelation of forecast errors, ρ_μ, by using a model-free procedure. This procedure does not require any simulations or recursions. To illustrate this alternative approach, we first rewrite the regression (see equation 10) that is used to estimate ρ_{pooled}:

$$\begin{aligned}
\mathrm{FE}_{i,t+1} &= a + \rho \mathrm{FE}_{i,t} + e_{i,t+1} \\
&= \mathrm{E}[\mathrm{FE}_{i,t}](1-\rho) + \rho \mathrm{FE}_{i,t} + e_{i,t+1} \\
&= (\mu - \hat{\mu})(1-\rho) + \rho \mathrm{FE}_{i,t} + e_{i,t+1}, \qquad \text{(A.6)}
\end{aligned}$$

because $a = \mathrm{E}[\mathrm{FE}_{i,t}](1-\rho)$ for a stationary AR(1) process and $\mu - \hat{\mu} = \mathrm{E}[\mathrm{FE}_{i,t}]$ (see footnote 24). We obtain an estimate of ρ_{pooled} by estimating this regression since the OLS estimator is a consistent estimator of $\rho_{\text{pooled}} = \mathrm{ACF}(1)/\mathrm{ACF}(0)$ when $\mathrm{SD}(\mu - \hat{\mu}) = 0$ irrespective of whether the higher-order ACFs are zero or not. Put differently, we examine this regression *only* because of this property of its OLS estimator. This approach is model free in that this regression is not a time-series model implied by any particular earnings and forecast processes. Now, Proposition 3 shows that the estimate ρ_{pooled} is higher when $\mathrm{SD}(\mu - \hat{\mu}) \neq 0$ than when $\mathrm{SD}(\mu - \hat{\mu}) = 0$. We can jointly estimate $\rho^*_{\text{pooled}} = \rho_{\text{pooled}} - \rho_\mu$ and $\mathrm{SD}(\mu - \hat{\mu})$ by estimating a mixed-effects version of the regression in equation (A.6), in which the intercept varies across firms (a random effect) and as a function of calendar time and firm age (fixed effects).[39] We can now recover $\rho_\mu = \rho_{\text{pooled}} - \rho^*_{\text{pooled}}$.

[39]Please note that, consistent with equation (A.6), the random and fixed effects measure $\mathrm{SD}((\mu - \hat{\mu})(1-\rho))$ instead of $\mathrm{SD}(\mu - \hat{\mu})$. This distinction is potentially important because variation in $(1-\rho)$ could then affect the estimates of $\mathrm{SD}(\mu - \hat{\mu})$ and, by extension, that of ρ_μ. However, we confirm using simulations that the variation in $(1-\rho)$ has but a negligible effect on the $\mathrm{SD}(\mu - \hat{\mu})$ estimates.

References

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

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

Abarbanell, J. S., Lehavy, R., 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, 105–146.

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

Anderson, E. W., Ghysels, E., Juergens, J. L., 2009. The impact of risk and uncertainty on expected returns. Journal of Financial Economics 94, 233–263.

Anderson, E. W., Hansen, L. P., Sargent, T. J., 2000. Robustness, detection and the price of risk, University of Chicago working paper.

Bernard, V. L., Thomas, J. K., 1990. Evidence that stock prices do not fully reflect the implications of current earnings for future earnings. Journal of Accounting and Economics 13, 305–340.

Bernhardt, D., Campello, M., Kutsoati, E., 2006. Who herds? Journal of Financial Economics 80, 657–675.

Bidder, R., Dew-Becker, I., 2014. Long-run risk is the worst-case scenario, Northwestern University working paper.

Bradshaw, M. T., 2011. Analysts' forecasts: what do we know after decades of work? Boston College working paper.

Brown, L. D., Hagerman, R. L., Griffin, P. A., Zmijewski, M. E., 1987. Security analyst superiority relative to univariate time-series models in forecasting quarterly earnings. Journal of Accounting and Economics 9, 61–87.

Brown, L. D., Rozeff, M. S., 1979. Univariate time-series models of quarterly accounting earnings per share: a proposed model. Journal of Accounting Research 17, 179–189.

Chen, Q., Jiang, W., 2006. Analysts' weighting of private and public information. Review of Financial Studies 19, 319–355.

Collin-Dufresne, P., Johannes, M., Lochstoer, L. A., 2013. Parameter learning in general equilibrium: the asset pricing implications, Columbia Business School working paper.

Collin-Dufresne, P., Johannes, M., Lochstoer, L. A., 2015. Asset pricing when 'This Time is Different', Columbia Business School working paper.

Cooper, R. A., Day, T. E., Lewis, C. M., 2001. Following the leader: a study of individual analysts' earnings forecasts. Journal of Financial Economics 61, 383–416.

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

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

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

Deschamps, B., Ioannidis, C., 2013. Can rational stubborness explain forecast biases? Journal of Economic Behavior & Organization 92, 141–151.

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

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

Epstein, L. G., Schneider, M., 2008. Ambiguity, information quality, and asset pricing. Journal of Finance 63, 197–228.

Epstein, L. G., Zin, S. E., 1989. Subtitution, risk aversion, and the temporal behavior of consumption and asset returns: a theoretical framework. Econometrica 57, 937–968.

Foster, G., 1977. Quarterly accounting data: time-series properties and predictive-ability results. Accounting Review 52, 1–21.

Griffin, P. A., 1977. The time-series behavior of quarterly earnings: preliminary evidence. Journal of Accounting Research 15, 71–83.

Gu, Z., Wu, J. S., 2003. Earnings skewness and analyst forecast bias. Journal of Accounting and Economics 35, 5–29.

Hansen, L. P., Sargent, T. J., 2008. Robustness. Princeton University Press, Princeton, New Jersey.

Hansen, L. P., Sargent, T. J., 2010. Fragile beliefs and the price of uncertainty. Quantitative Economics 1, 129–162.

Hansen, L. P., Sargent, T. J., Turmuhambetova, G., Williams, N., 2006. Robust control and model misspecification. Journal of Economic Theory 128, 45–90.

Hansen, L. P., Sargent, T. J., Wang, N. E., 2002. Robust permanent income and pricing with filtering. Macroeconomic Dynamics 6, 40–84.

Hilary, G., Hsu, C., 2013. Analyst forecast consistency. Journal of Finance 68, 271–297.

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

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

Jegadeesh, N., Kim, J., Krische, S. D., Lee, C. M. C., 2004. Analyzing the analysts: when do recommendations add value? Journal of Finance 59, 1083–1124.

Jegadeesh, N., Kim, W., 2010. Do analysts herd? an analysis of recommendations and market reactions. Review of Financial Studies 23, 901–937.

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

Kendall, M. G., 1954. Note on bias in the estimation of autocorrelation. Biometrika 41, 403–404.

Kreps, D. M., Porteus, E. L., 1978. Temporal resolution of uncertainty and dynamic choice theory. Econometrica 46, 185–200.

La Porta, R., 1996. Expectations and the cross-section of stock returns. Journal of Finance 51, 1715–1742.

Lakonishok, J., Shleifer, A., Vishny, R., 1994. Contrarian investment, extrapolation and investment risk. Journal of Finance 49, 1541–1578.

Lim, T., 2001. Rationality and analysts' forecast bias. Journal of Finance 56, 369–385.

Loh, R. K., Stulz, R. M., 2011. When are analyst recommendation changes influential? Review of Financial Studies 24, 593–627.

Maenhout, P. J., 2004. Robust portfolio rules and asset pricing. Review of Financial Studies 17, 951–983.

Markov, S., Tamayo, A., 2006. Predictability in financial analyst forecast errors: learning or irrationality? Journal of Accounting Research 44, 725–761.

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

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

Newey, W. K., West, K. D., 1987. A simple, positive semi-definite, heteroscedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.

Newey, W. K., West, K. D., 1994. Automatic lag selection in covariance matrix estimation. Review of Economic Studies 61, 631–653.

Orlik, A., Veldkamp, L., 2014. Understanding uncertainty shocks and the role of black swans, NBER Working Paper No. 20445.

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

Shane, P., Brous, P., 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, 387–404.

So, E. C., 2013. A new approach to predicting analyst forecast errors: do investors overweight analyst forecasts? Journal of Financial Economics 108, 615–640.

Trueman, B., 1994. Analyst forecasts and herding behavior. Review of Financial Studies 7, 97–124.

Uppal, R., Wang, T., 2003. Model misspecification and underdiversification. Journal of Finance 58, 2465–2486.

Weil, P., 1990. Nonexpected utility in macroeconomics. Quarterly Journal of Economics 105, 29–42.

Welch, I., 2000. Herding among security analysts. Journal of Financial Economics 58, 369–396.

Williams, J., 2013. Financial analysts and the false consensus effect. Journal of Accounting Research 51, 855–907.

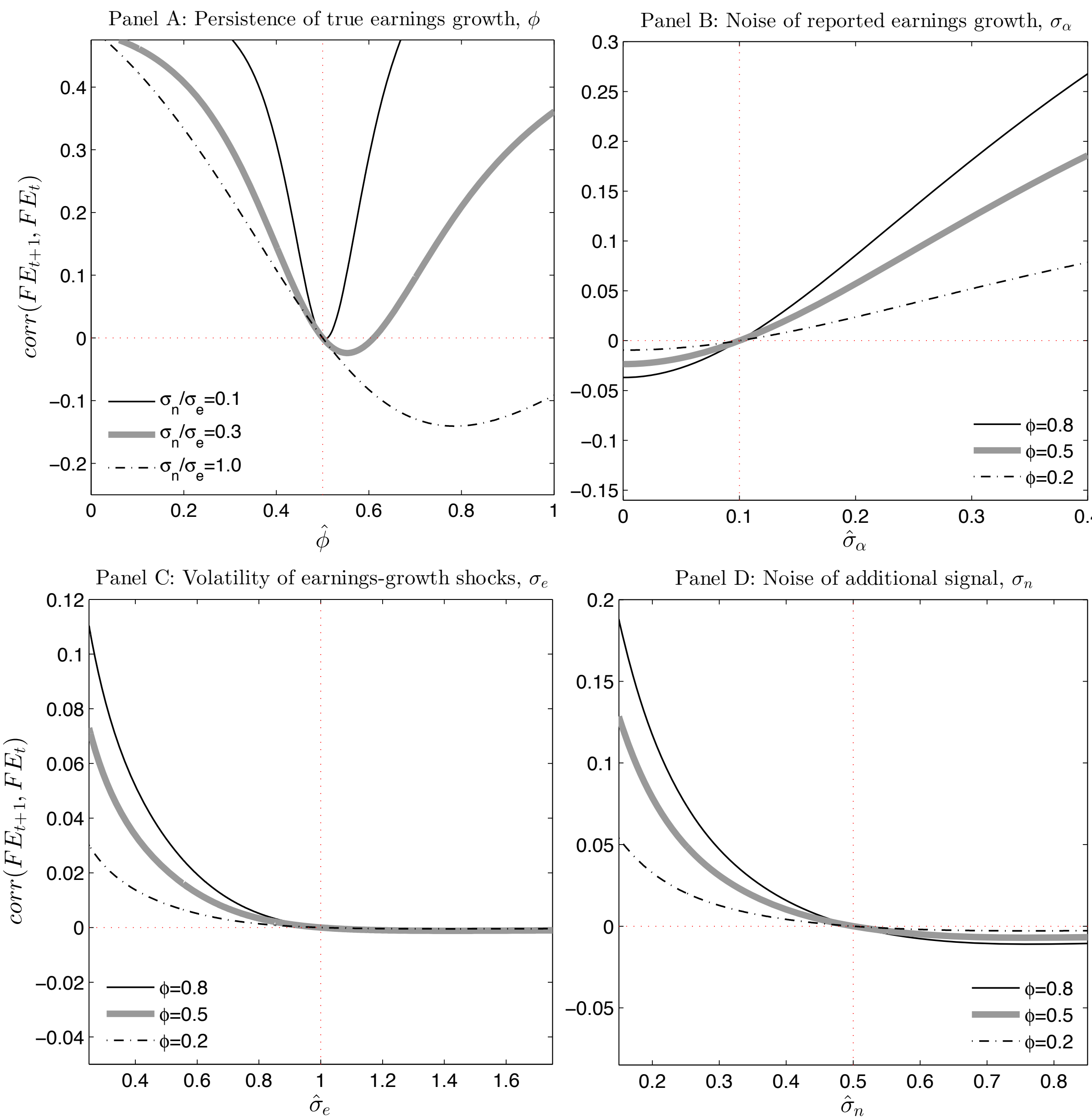

Fig. 1. Serial correlation in forecast errors as a function of model parameters. Earnings growth y_{t+1} follows an AR(1)-plus-noise process and the analyst observes an additional signal that informs him about the shock to the persistent earnings-growth component. The true parameters are $\phi = 0.5$, $\sigma_\alpha^2 = (0.1)^2$, $\sigma_e^2 = (1)^2$, and $\sigma_n^2 = (0.5)^2$. This figure plots the autocorrelation of forecast errors as a function of the parameters used by the analyst: $\hat{\phi}$, $\hat{\sigma}_\alpha^2$, $\hat{\sigma}_e^2$, and $\hat{\sigma}_n^2$. The vertical line in each panel denotes the true parameter value.

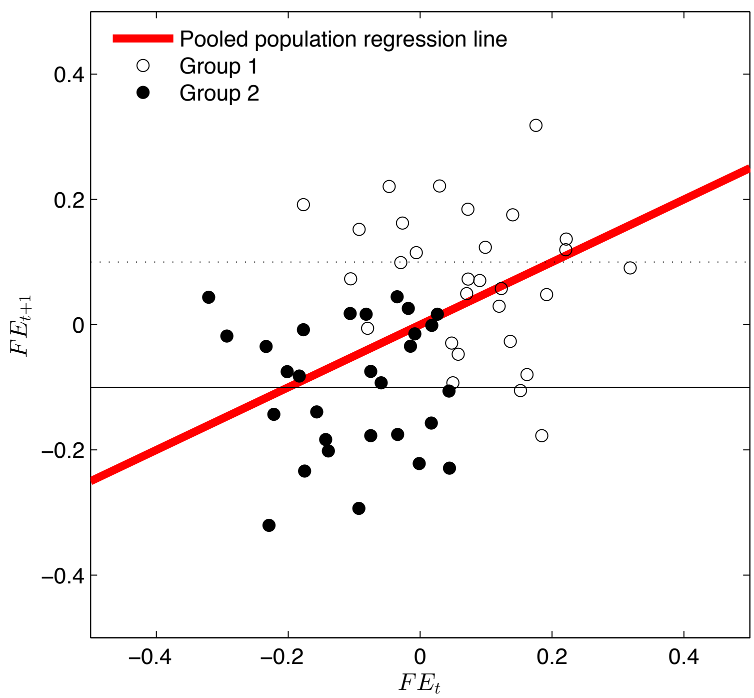

Fig. 2. Variation in mean forecast errors and its effect on the autocorrelations of forecast errors. This figure illustrates Proposition 3 by assuming that forecast errors are IID draws from two normal distributions with means of +0.1 (group 1) and −0.1 (group 2) and standard deviations of 0.1. Therefore, $\text{var}(b_m) = 0.01$ and $\text{var}(\text{FE}_t) = \text{var}(b_m) + \text{var}(\text{FE}_t \mid \text{group}) = 0.02$. The solid thin line and the dashed line are the population regression lines for the two groups. The thick solid line is the regression line for a sample that pools observations from groups 1 and 2.

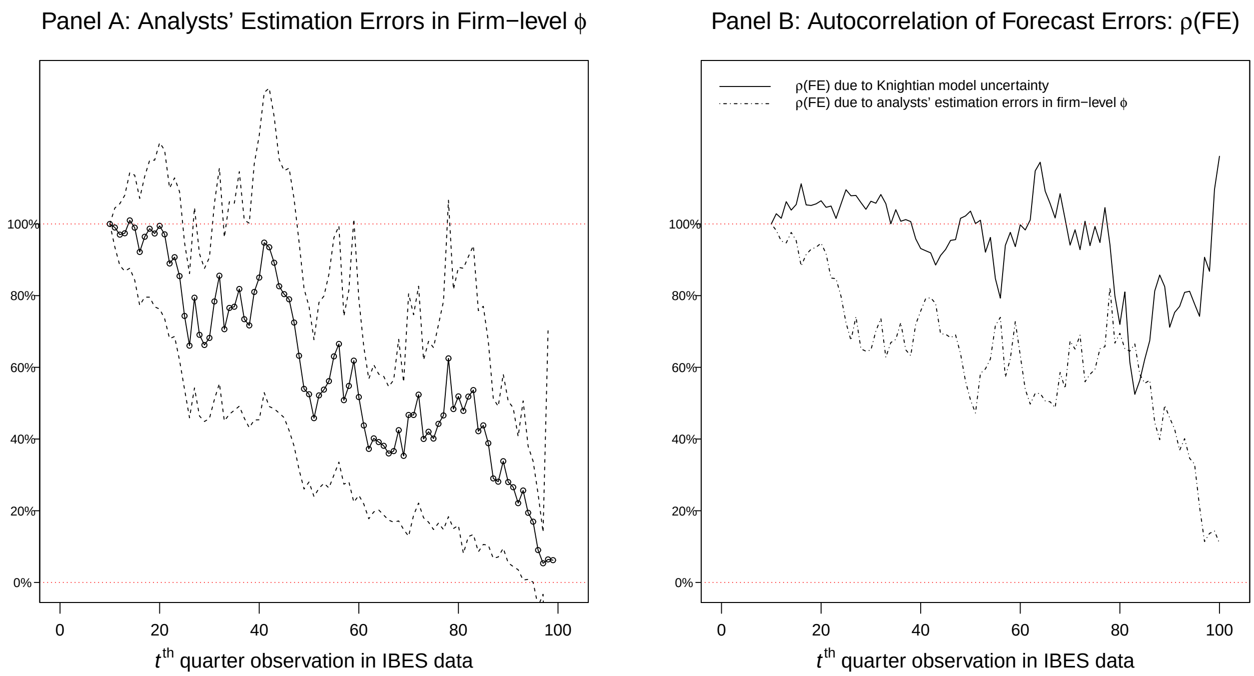

Fig. 3. Learning about the persistence of earnings-growth shocks and time-variation in model uncertainty. Panel A estimates how analysts' uncertainty about the persistence of the earnings-growth shocks changes as firms age. The solid line in Panel A represents analysts' uncertainty about ϕ_i after observing t quarters of IBES data. The estimates are scaled so that the uncertainty is relative to the uncertainty at the time of the first observation. The 95% confidence interval are computed by block bootstrapping the data by firms. Panel B measures how much of the autocorrelation of forecast errors is due to (1) uncertainty about the noisiness of reported earnings (solid line) and (2) uncertainty about ϕ (dashed line). The two autocorrelation components are scaled so that the estimates are relative to their sizes at the time of the first observation.

Table 1
Descriptive statistics, 1984–2013

This table reports the distributions of year-to-year quarterly earnings growth (y_t), forecasted earnings growth ($\hat{y}_t$), and forecast errors. The data combine IBES, Compustat, and CRSP data from 1984 through 2013. See the text for details on sample construction. Autocorrelations are estimated from AR(1) regressions. The pooled estimate uses data on all firms and firm-specific estimates are the average estimates from firm-specific regressions. Short-lived firms (N = 3,349) are firms with fewer than 20 quarterly observations. Long-lived firms (N = 3,804) are firms with at least 20 quarterly observations. Except for these firm-specific autocorrelation estimates, the short-lived firms are not part of the main sample. The main sample therefore has 185,420 firm-quarter observations on 3,804 firms that survive for at least five years. The bias-adjusted autocorrelation estimates $\hat{\rho}_{\text{bias-adjusted}}$ correct raw estimates $\hat{\rho}$ for Kendall's (1954) small-sample bias, $\hat{\rho}_{\text{bias-adjusted}} = \frac{\hat{\rho}(T-1)+1}{T-4}$, where T is the number of observations.

Statistic	Year-to-year earnings growth, y_t	Forecasted growth, $\hat{y}_t$	Forecast error, $y_t - \hat{y}_t$: Estimate	t-value
Mean	0.004	0.000	0.004	2.07
SD	1.113	0.992	0.501	
Percentiles				
5%	−1.615	−1.398	−0.607	
25%	−0.250	−0.246	−0.037	
50%	0.030	0.015	0.020	13.37
75%	0.340	0.289	0.119	
95%	1.442	1.294	0.563	
Negative	46.0%	47.7%	32.4%	
Zero	0.2%	0.1%	10.3%	
Positive	53.8%	52.2%	57.3%	
Pooled autocorrelation	0.429	0.434	0.216	28.87
Firm-specific autocorrelations				
All firms			0.102	20.12
All firms (bias-adjusted)			0.244	23.36
Short-lived firms			0.042	4.49
Short-lived firms (bias-adjusted)			0.299	13.79
Long-lived firms			0.153	31.80
Long-lived firms (bias-adjusted)			0.196	36.88

Table 2

Estimates of a VARMA(1,1) model of earnings and analyst forecasts

This table presents parameter estimates from a VARMA(1,1) model that describes the evolution of firms' earnings and analyst forecasts. The data are analysts' earnings forecasts and actual earnings per share from 1984 through 2013 from IBES. We estimate the VARMA model in two steps. First, we fit the earnings dynamics with an AR(1)-plus-noise model using maximum likelihood estimation and a Kalman filter. Second, we use these estimates of the earnings process within the VARMA model to estimate the remaining parameters using maximum likelihood estimation and a Kalman filter. We report bootstrapped standard errors that draw firms as blocks with replacement. Rows labeled True report the estimated parameters of the earnings process; Implied are the parameters used by the analysts, as implied by their forecasts. The bottom part reports R^2s for the AR(1)-plus-noise model and for analysts' forecasts. The latter compares the variance of forecast errors to the variance of earnings growth, $R^2 = 1 - \text{var}(y_{t+1} - \hat{y}_{t+1})/\text{var}(y_{t+1})$.

Parameter	Estimate	SE
Persistence of the earnings-growth shocks		
True, ρ	0.472	0.006
Implied, $\hat{\rho}$	0.470	0.005
SD(noise term) / SD(earnings growth shock), σ_α/σ_e	0.106	0.073
SD(additional signal) / SD(earnings growth shock), σ_n/σ_e	0.538	0.022
Kalman gain		
True, K	0.953	0.059
Implied, $\hat{K}$	0.414	0.049
Weight placed on the additional signal		
True, w	0.775	0.014
Implied, $\hat{w}$	0.783	0.014
R^2s for predicting earnings growth		
AR(1)-plus-noise model	14.0%	
Analysts' median forecast (IBES)	79.8%	

Table 3
Variation in mean forecast errors, estimation errors in $\hat{\phi}$, and detection error probabilities

Panel A reports estimates of variation in mean forecast errors, $\mathrm{SD}(\mu - \hat{\mu})$, and estimation errors in the persistence of the earnings-growth shocks, $\mathrm{SD}(\phi - \hat{\phi})$. We measure variation from three sources: calendar-time variation; variation as a function of firm age; and variation across firms. Appendix D describes the computational details. Effective sample size is estimated by equating $\mathrm{SD}(\mu-\hat{\mu})$ or $\mathrm{SD}(\phi-\hat{\phi})$ to the standard errors of $\hat{\mu}$ and $\hat{\phi}$ in the AR(1)-plus-noise model of earnings growth. Panel B reports detection error probabilities as a function of σ_α (the noisiness of reported earnings) and sample size. Detection error probability measures the difficulty of distinguishing the approximating model from the worst-case model. These probabilities range from 50% (when the models are observationally equivalent) to 0% (when one model is wholly inconsistent with the data). In Section 2's model, detection error probabilities, the noisiness of reported earnings in the worst-case model ($\sigma_\alpha^{(w)}$), the constraint-parameter η, and the autocorrelation of forecast errors are related so that fixing one of them uniquely determines the others.

Panel A: Variation in mean forecast errors and estimation errors in $\hat{\phi}$

Parameter	Estimate	SE	Effective sample size
$100 \times \mathrm{SD}(\mu - \hat{\mu})$	**0.1026**	**0.0034**	341
Calendar-time variation	0.0451	0.0028	
Age variation	0.0112	0.0026	
Variation across firms	0.0915	0.0036	
$\mathrm{SD}(\phi - \hat{\phi})$	**0.0875**	**0.0021**	465
Calendar-time variation	0.0185	0.0012	
Age variation	0.0088	0.0010	
Variation across firms	0.0851	0.0026	

Panel B: Detection error probability as a function of σ_α and sample size

σ_α/σ_e	Detection error probability	Auto-correlation of FEs	η	Sample size
0.106 ← true model	50.0%	0.000	0.00	465
0.200	38.3%	0.047	0.89	465
0.330	17.2%	0.125	2.11	465
0.438	5.0%	0.184	3.13	465
0.508	1.7%	0.216	3.76	465
0.330	26.7%	0.125	2.11	200
0.330	5.0%	0.125	2.11	1,417

Table 4
Decomposing the autocorrelation of forecast errors

This table decomposes the autocorrelation of forecast errors into three main components: (1) autocorrelation due to variation in mean forecast errors, (2) autocorrelation due to estimation errors in $\hat{\phi}$, and (3) autocorrelation due to analysts' concerns for model misspecification. These three components add up to the total autocorrelation of forecast errors estimated from a pooled regression. The first two components are further decomposed by the source of heterogeneity. Mean forecast errors, for example, vary as a function of calendar time (year), firm age, and firm, and this table reports how much the variation in each dimension contributes to the autocorrelation of forecast errors. Standard errors associated with the variation-in-mean forecast errors channel are are heteroskedasticity and autocorrelation consistent Newey and West (1987) with the number of lags selected using Newey and West (1994). Standard errors associated with the estimation errors-in-$\hat{\phi}$ channel are computed using a parametric bootstrap.

Autocorrelation estimate	Estimate	SE
Total autocorrelation of forecast errors, (1) + (2) + (3)	**0.216**	**0.008**
(1) Autocorrelation due to variation in mean forecast errors	**0.043**	**0.008**
Calendar-time variation	0.012	0.009
Age variation	0.000	0.008
Variation across firms	0.030	0.009
(2) Autocorrelation due to estimation errors in $\hat{\phi}$	**0.047**	**0.006**
Calendar-time variation	0.001	0.002
Age variation	0.000	0.003
Variation across firms	0.046	0.010
(3) Autocorrelation due to analysts' concerns for model misspecification	**0.125**	

Table 5

Explaining cross-sectional variation in the autocorrelation of forecast errors

This table reports estimates from pooled regressions $\widehat{\text{corr}}(\text{FE}_{i,t}, \text{FE}_{i,t-1}) = \beta_0 + \beta_t + \gamma \text{Age}_{i,t-1} + \mathbf{X}_{i,t-1}\boldsymbol{\beta} + e_{i,t}$, in which the dependent variable is estimated using a five-year window around quarter t, β_t represent year fixed effects, and firm age is measured from the firm's first appearance in CRSP. The characteristics $\mathbf{X}_{i,t}$ are defined as: (1) Return volatility is computed from daily returns over the prior quarter; (2) Forecast dispersion is the standard deviation of analyst forecasts scaled by the stock price in the prior quarter; (3) BE/ME is the firm's book-to-market ratio in the prior quarter; (4) Firm size is the market value of equity in the prior quarter; and (5) Coverage is the number of analysts covering the firm. The standard errors are heteroscedasticity and autocorrelation consistent. All characteristics except age are demeaned by subtracting their median and scaled by their standard deviation. The age variables are scaled so that their coefficients represent the decrease in autocorrelation for 50 years since a firm's birth.

Independent	Regression				
variable	(1)	(2)	(3)	(4)	(5)
β_0	0.24	0.21	0.19	0.18	0.22
	(10.37)	(8.65)	(10.71)	(9.75)	(6.75)
Age			−0.04	−0.02	0.01
			(−3.92)	(−2.08)	(0.35)
$\log(1 + \text{Age})$	−0.09	−0.05			−0.07
	(−4.53)	(−2.55)			(−1.41)
Return volatility		0.02		0.02	0.02
		(3.80)		(4.16)	(3.77)
Forecast dispersion		0.00		0.00	0.00
		(0.16)		(0.14)	(0.14)
BE/ME		−0.01		−0.01	−0.01
		(−1.39)		(−1.45)	(−1.40)
Firm size		0.00		0.00	0.00
		(−0.50)		(0.00)	(−0.12)
Coverage		0.00		0.00	0.00
		(0.73)		(0.67)	(0.74)
Year FEs	Yes	Yes	Yes	Yes	Yes
R^2	0.5%	0.6%	0.5%	0.6%	0.7%