5-2013

An Examination of the Statistical Significance and Economic Implications of Model-Based and Analyst Earnings Forecasts

Kevin Ow Yong
Singapore Management University, kevinowyong@smu.edu.sg

M. Evans

K Njoroge

An Examination of the Statistical Significance and Economic Implications of Model-Based and Analyst Earnings Forecasts

Mark E. Evans*
(Indiana University)

Kenneth Njoroge
(University of Oregon)

Kevin Ow Yong
(Singapore Management University)

Draft: December 2012

* Corresponding author. email: evansme@indiana.edu

An earlier draft of this paper was entitled "Bias and Accuracy in Long-Horizon Earnings Forecasts: Does a Cross-Sectional Model Improve Analysts' Forecasts?". We have received many helpful comments from the following sources: Christine Botosan, Asher Curtis, Ro Gutierrez, Linda Krull, Steve Matsunaga, Sarah McVay, Dale Morse, Stephen Penman, Marlene Plumlee, Katherine Schipper, Cathy Schrand, Teri Lombardi Yohn, and workshop participants at the College of William and Mary, University of Miami, University of Oregon, University of Utah, the 2010 Midwest Accounting Research Conference, and the 2010 Northwest Accounting Research Conference. We also gratefully acknowledge financial support from Kelley School of Business at Indiana University, Lundquist College of Business at the University of Oregon and the School of Accountancy at Singapore Management University.

An Examination of the Statistical Significance and Economic Implications of Model-Based and Analyst Earnings Forecasts

ABSTRACT

We address the demand for model-based earnings forecasts by proposing a cross-sectional model which incorporates three salient ideas. First, firm performance converges to expected levels over time; second, amounts from current financial statements are robust predictors of future performance; and third, ordinary least squares (OLS) estimation is unreliable in samples including extreme values. Accordingly, we estimate a cross-sectional earnings forecasting model based on least absolute deviations analysis (LAD), and include profitability drivers derived from financial statements as predictors. In terms of statistical significance, we find that these forecasts are more accurate than forecasts from three extant prediction models and consensus analysts' forecasts. In terms of economic implications, we find that forecasts from our model have greater predictive ability for future abnormal returns than consensus analysts' forecasts. Overall, our results are important because they document the usefulness of a cross-sectional earnings forecasting model for a broad range of diverse firms, including those with little or no analyst coverage.

Keywords: Earnings Forecasts; Financial Statement Analysis; Security Analysts

JEL Codes: M40; G11; G17

Data Availability: All data are publicly available from the sources identified in the paper.

1. Introduction

In this paper, we address the demand for model-based earnings forecasts by developing a cross-sectional earnings forecasting model that is suitable for a broad set of diverse firms. Specifically, we propose and test a model which is useful when reliance on consensus analysts' forecasts is difficult (because of concerns about optimism or incentives) or impossible (because of lack of coverage). We first evaluate our model in terms of statistical significance, and find that forecasts from our proposed model are more accurate than three extant models' forecasts at every forecast horizon that we considered in our tests (one to five years). We also find that our model's forecasts are more accurate than consensus analysts' forecasts for forecast horizons ranging from two to five years. We next evaluate our model in terms of economic implications, and show that our model's forecasts have predictive ability with regard to future returns, incremental to signals provided by analysts, the market, and fundamental analysis. Taken together, our results highlight the usefulness of model-based earnings forecasting models when reliance on analysts is problematic.

Forecasted earnings is important because of its use as a key input in valuation models (Richardson, Tuna, and Wysocki 2010; Chee, Sloan, and Uysal 2012). In recent years, important contributions have been made in the use of statistical forecasting models instead of—or in addition to—analyst forecasts (e.g., Bradshaw et al. 2012; Hou, van Dijk, and Zhang 2012). Demand for statistical forecasts exists because research has shown that analysts' forecasts are biased—especially for forecast horizons greater than one year (e.g., Harris 1999; Chan et al. 2003)—and less accurate than forecasts implied by stock returns (Hughes, Liu, and Su 2008). In addition, many firms are not followed by analysts even though these firms constitute an important part of the economy. For example, in our sample for the year 2009, aggregate market value and sales for uncovered firms is approximately $3 trillion. Finally, machine readable databases typically contain only one year ahead (or shorter) analyst forecasts.

Despite these concerns, researchers frequently use consensus analysts' forecasts rather than statistical forecasts to capture expectations about future earnings (e.g., Kothari 2001; Bradshaw 2011). Prior research suggests that consensus analysts' forecasts should outperform statistical models in terms of forecast accuracy for several reasons. First, analysts are able to influence the choice of firm coverage and forecast horizon (e.g., Francis, Chen, Philbrick, and Willis 2004). For example, analysts may prefer to follow firms with more predictable earnings streams. Second, analysts have a timing advantage that allows them to update their forecasts between earnings report dates (O'Brien 1988). Thus, they are able to quickly revise forecasts in response to unexpected earnings shocks (e.g., strikes, lawsuits, mergers, management turnover). In contrast, statistical models restrict coefficient estimates to be either constant over time, constant across firms, or both. Third, analysts have an information advantage over statistical models because inputs to statistical models are likely to utilize a smaller subset of available information relative to the analysts' overall information set. In particular, a thriving analyst profession provides evidence that analysts add value to statistical models. Thus, it would be surprising if forecasting models are able to outperform analysts' forecasts.

Notwithstanding these advantages, we find that out-of-sample forecasts from our proposed model are less biased and more accurate, on average, than analysts' consensus forecasts for two, three, four, and five year forecast horizons. Moreover, we show that our forecasts are more accurate than forecasts based on a random walk model (e.g., Bradshaw et al. 2012), autoregressive forecasting model (i.e., AR1 forecasts) and an extant cross-sectional forecasting model (e.g., Hou et al. 2012) in out-of-sample tests at all horizons considered.

Specifically, we document forecast accuracy improvements based on our model's forecasts ranging from 0.07% of stock price at the one-year horizon to 1.03% of stock price at the five-year horizon over these competing models. If we consider, for example, a hypothetical firm that is reporting $1 earnings per share and trading at $15 (i.e., P/E ratio = 15), the forecast improvements would range from approximately $0.01 per share to $0.15 per share, or 3% to 38% of the median firm's EPS.

In market tests, we find that forming hedge portfolios using our model's earnings forecasts generates positive abnormal returns up to four quarters ahead in the range of 3-4% abnormal returns, compared to analysts' forecasts, which generate insignificant abnormal returns in most instances. In addition, we show that forecasts from our model predict future abnormal returns up to four quarters ahead in the range of 1-2% for a subsample of firms not covered by analysts. In subsequent tests, we also show that the signals produced by our model's forecasts are positively associated with future returns, incremental to (i.e., not subsumed by) twelve other investment signals established in the literature as factors predicting future returns. In addition, signals produced by analysts' forecasts are *negatively* associated with future returns when included in the same regression as signals produced by our forecasting model. Taken together, our results suggest that researchers can potentially incorporate a significantly larger sample (upwards of 30-40% larger) with nontrivial economic significance by using our proposed model for firms that are not covered by analysts, or for subsamples of analyst-covered firms with significant concerns about selection bias, optimism, and forecast inaccuracy.

Consistent with other cross-sectional prediction models, our model exploits the fact that earnings revert to expected levels over time (Fama and French 2000). However, our forecasting model differs from extant cross-sectional models on two dimensions. First, we include a broader set of predictors. For example, in addition to including prior realizations of earnings and firm size, we include measures related to firms' debt and equity financing decisions and stock splits, as well as levels of special items. Second, we refine the estimation methodology of cross-sectional models by using least absolute deviations (LAD) analysis instead of ordinary least squares (OLS) analysis. Consistent with previous research, we find that the cross-sectional distribution of pooled earnings per share (EPS) is negatively skewed in our sample, and has a higher peak and fatter tails than the normal distribution (see Figure 1). Accordingly, the influence of outliers on the median is smaller than that on the mean and, therefore, we utilize LAD estimation in order to minimize the influence of outliers in the distribution.[1] Importantly, our design choice is supported, *ex post*, by narrower confidence intervals surrounding our out-of-sample LAD forecasts relative to OLS forecasts (see Figure 2).

We make three contributions to the extant literature. First, we address concerns raised by Richardson, Tuna, and Wysocki (2010) by applying more structure to the earnings forecasting framework. Specifically, we propose and test a model that incorporates the economic notion of reversion to expected levels using known profitability drivers derived from financial statements; in addition, we incorporate an estimation methodology that is appropriate for broad and diverse samples of firms. Second, we provide evidence on both the statistical significance of our forecasts relative to models and analysts (forecast accuracy, bias), and the economic implications of these forecasts to investors. In particular, we show that forecasts from our model have predictive ability for future returns incremental to analyst signals, market signals, and fundamental signals. Third, we provide evidence documenting the characteristics of analysts and

[1] OLS minimizes the squared forecast error, thus minimizing deviation from the mean. In contrast, LAD minimizes the absolute forecast error, thus minimizing deviation from the median. See Gu and Wu (2003) and Basu and Markov (2004) for further discussion regarding both estimation methodologies. Also, Blouin, Core, and Guay (2010) implement a non-parametric approach for estimating expected earnings used in marginal tax rate calculations.

firms for which our model's forecasts outperform consensus analysts' forecasts. These results have the potential to inform future researchers in determining the necessity of model-based forecasts for particular samples.

The rest of the paper is organized as follows. In section 2, we describe our forecasting model as well as competing forecasting models. In section 3, we discuss our sample. We present the results of our statistical tests in section 4 and the results of our market tests in section 5. We conclude in section 6.

2. Model development and relation to competing forecasts

In this section, we describe and explain our proposed earnings prediction model and other extant earnings forecasting models that are commonly used in recent accounting research.[2] Specifically, we compare and explain common similarities and differences among these models.

2.1 Proposed cross-sectional earnings forecasting model

Our proposed earnings prediction model originates from an extension of early cross-sectional profitability models (e.g., Fama and French 2000; Hou and Robinson 2006). Fama and French (2000) use a two-stage approach to estimate a partial adjustment model of profitability in the cross-section. They find that their model is able to significantly explain variation in the expected profitability across firms. Specifically, they use OLS in the first stage of the two-stage approach to estimate a cross-sectional regression of return on assets (ROA) on the ratio of dividends to book equity, a dummy for dividend payers and the market-value-to-assets ratio. The fitted value of ROA from this regression is then used as a proxy for the expected level of

[2] We do not include time-series forecasting models fitted to individual firms for the following reasons. First, time-series models induce survivor bias that may be more problematic than the lack of analyst coverage. Moreover, the ability to predict earnings for long-term survivors may not be representative to a broad sample of firms. Additionally, the precise estimation of these models requires a long time-series, such that even twenty observations of annual earnings may produce imprecise forecasts, especially for long horizon forecasts (see Fama and French 2000). Finally, the out-of-sample performance when these models are fit using quarterly data is typically outperformed by analyst forecasts (see Brown 1993).

profitability in the second stage. The second stage estimates the rate of reversion to the expected level of profitability. That is, the second stage estimates an OLS regression of next-year change in ROA on ROA, the proxy for expected profitability from the first stage and change in ROA.

If there is earnings reversion to expected levels, the coefficient on ROA should be positive while the coefficient on the fitted ROA should be negative. Additionally, the two coefficients should be of equal absolute magnitude if the model is properly specified. Hence, the magnitude of both coefficients provides an estimate of the average rate of profitability reversion to expected levels. More importantly, Fama and French note that there is further potential to capture predictable variation in earnings beyond that which is captured by their earnings profitability model. Specifically, they observe that negative changes in earnings and extreme changes reverse faster than what is predicted in their cross-sectional model (p.163).

We extend their base model by including additional variables that are not previously considered in the literature in Stage 1 of the forecasting model to estimate the expected earnings per share as follows:

<u>Stage 1: Estimating the expected earnings per share</u>

$$EPS_t = \alpha_0 + \alpha_1 EPS_{t-1} + \alpha_2 LOSS_{t-1} * EPS_{t-1} + \alpha_3 \Delta EPS_{t-1} + \alpha_4 DEBT\_DIST_{t-1} + \alpha_5 EQUITY\_DIST_{t-1} + \alpha_6 SPLIT\_DUM_{t-1} + \alpha_7 DIV\_DUM_{t-1} + \alpha_8 SPEC\_ITEMS_{t-1} + \alpha_9 lnSIZE_{t-1} + \varepsilon_t \qquad (1)$$

where *EPS* is earning per share, before extraordinary items, *LOSS* is an indicator equal to 1 if *EPS* is negative and zero otherwise, *ΔEPS* is change in *EPS*, *DEBT_DIST* is net distributions to debt holders, *EQUITY_DIST* is net distributions to equity holders, *SPLIT_DUM* is a dummy for stock splits, *DIV_DUM* is a dummy for dividend payers, *SPEC_ITEMS* is special and extraordinary items, and *lnSIZE* is the natural log of total assets. Calculation details, including Compstat variable names, are included in the Appendix.

Equation (1) models expected EPS as a function of past performance and signals about future performance. Accordingly, our first group of predictors includes *EPS* and *ΔEPS*, while allowing for differential persistence of earnings conditional on losses (*LOSS*EPS*). We expect a positive coefficient on *EPS* and a negative coefficient on *LOSS*EPS* and *ΔEPS*.

With regard to signals about future performance, we follow Nissim and Penman (2001) and Dechow, Richardson, and Sloan (2008) and include lagged equity distributions (*EQUITY_DIST*) and lagged debt distributions (*DEBT_DIST)* as predictors of future earnings. This research contends that distributions to equity and debt holders are more likely to come from core (persistent) earnings rather than non-core (transitory) earnings. Thus, we expect positive coefficients for *EQUITY_DIST* and *DEBT_DIST*. Also, we include indicator variables for stock splits (*SPLIT_DUM*) and dividend payers (*DIV_DUM*), as credible management signals of future profitability. We expect positive coefficients for both variables (Fama and French 2000). We also include the lag of special items (*SPEC_ITEMS*). We expect a negative coefficient for this variable, consistent with the persistence of core profit margins and the transitory nature of special items (Fairfield, Kitching, and Tang 2009; Jones and Smith 2011). That is, holding bottom-line earnings constant, firms with more negative special items will experience higher future net income, relative to firms with less negative special items. Finally, we control for size using total assets (*lnSIZE*), which is expected to be positively associated with future earnings (Fama and French 2006).

We validate our Stage 1 regression's ability to explain cross-sectional variation in expected earnings using the Stage 2 partial adjustment model presented below:

Stage 2: Partial Adjustment Model

$$EPS_{t+1} - EPS_t = \beta_0 + \beta_1 E[EPS_t] + \beta_2 EPS_t + \beta_3(EPS_t - EPS_{t-1}) + \epsilon_{t+1} \qquad (2)$$

where $E[EPS_t]$ is the fitted value from the Stage 1 regression, and it proxies for expected earnings. If the first stage is well specified— in the sense that it reasonably captures the cross-sectional variation in expected EPS, reversion of EPS to the expected level implies that $\beta_1 = -\beta_2$. Effectively, the stage 2 model validates the extent to which our forecasting model captures cross-sectional variation in expected EPS. Consistent with Fama and French (2000), we include $(EPS_t - EPS_{t-1})$ as an explanatory variable to capture additional variation in earnings not captured by the partial adjustment term.

While our model is motivated by Fama and French (2000), we note that ours is distinct from their model in four ways. First, we use additional profitability drivers identified in previous research. This improves our model's ability to capture cross-sectional variation in expected EPS. Second, we use least absolute deviation (LAD) instead of OLS to estimate regressions (1) and (2) to alleviate the influence of outliers. As a result, our model is able to accommodate numerous small firms and firms with frequent or large losses.[3] Third, our model excludes the market value of equity or stock returns as predictors. One reason for this exclusion concerns the primary objective our study, which is to compare our model forecasts to analysts' forecasts. To the extent that market values and stock returns are related to analyst forecasts, including market values in the model will bias the model toward outperforming analysts—this is similar to using analyst forecasts as an input to our model. Additionally, an equally important objective in this paper is to

[3] We justify this approach based on the sample distribution of pooled EPS in our sample firms. Specifically, Figure 1 shows that the distribution has a higher peak and fatter tails (leptokurtic distribution) and is negatively skewed, relative to the normal distribution. Thus, extreme negative observations will occur more frequently for this distribution than under the normal distribution.

compare the future abnormal returns predicted by our model's EPS forecast to that implied by analyst forecasts. Using market values or returns in our model would complicate any inference we can draw from such tests.[4] Finally, we scale earnings by weighted shares outstanding instead of assets. This yields earnings-per-share forecasts which we can directly compare with analysts' forecasts. After validating our Stage 1 model, we use it to forecast earnings-per-share for *h* periods ahead using the following model, which re-states equation 1 in terms of multi-period forecasts:

$$EPS_{t+h} = \alpha_0 + \alpha_1 EPS_t + \alpha_2 LOSS_t * EPS_t + \alpha_3 \Delta EPS_t + \alpha_4 DEBT\_DIST_t + \alpha_5 EQUITY\_DIST_t + \alpha_6 SPLIT\_DUM_t + \alpha_7 DIV\_DUM_t + \alpha_8 SPEC\_ITEMS_t + \alpha_9 lnSIZE_t + \varepsilon_{t+h} \quad (3)$$

2.2 Competing Forecast Models

Random Walk Model

If the earnings time series follows random walk, then current EPS is the best predictor of its future value. Specifically:

$$EPS_{i,t+h} = EPS_{i,t} + \varepsilon_{t+h},$$

where $h \in \{1,2,3,4,5\}$. The conditioning information set for the random walk model is limited to current EPS and the model implies that future change in EPS is unpredictable. Random walk forecasts are attractive since selection bias is minimal and the estimation is simple. One disadvantage of the random walk model is that it ignores growth, resulting in pessimistic forecasts for growth firms. To the extent that growth is difficult to predict, developing models that outperform random walk forecasts out-of-sample is nontrivial, especially over long forecast horizons. In standard applications, economics research frequently uses the random walk model as

[4] Notwithstanding the rationale for excluding market values, in unreported tests, we find that including market value as a predictor does not significantly change the model's out-of-sample performance but it limits the model to publicly traded firms.

a benchmark in assessing the performance of proposed forecast models. In addition, early accounting research that uses sophisticated time-series models fit separately to individual firms are frequently unable to reject the hypothesis that earnings follow random walk, out-of-sample. Finally, Bradshaw et al. (2012) show that the random walk model outperforms analysts over long horizons for small firms, and for firms with greater potential for growth. For these reasons, we use the random walk model as a benchmark in comparing forecasting models.

Autoregressive model

The autoregressive model (hereafter AR) estimates a slope coefficient based on a variable's current value to forecast its future values. Formally:

$$EPS_{i,t+h} = \alpha_0 + \alpha_1 EPS_{i,t} + \varepsilon_{t+h} \tag{4}$$

Like the random walk model, the conditioning information set for the AR model is limited to current EPS. Unlike the random walk model, the AR model implies that current EPS can predict future change in EPS. In particular, when the absolute value of the slope coefficient is between zero and one (i.e.,$0 < |\alpha_1| < 1$), EPS is mean reverting. Hence, EPS accommodates earnings reversion to expected levels, but assumes that current EPS subsumes all information that is useful in predicting future performance. In the pooled cross-section, the AR model is attractive for its minimal selection bias and simple estimation. To the extent that EPS subsumes other predictors, the AR model can be suitable. The AR model is also frequently used as a benchmark model in economics and accounting research. Additionally, it can accommodate additional lags of earnings depending on the dynamics of the earnings process. However, we find that including additional lags of EPS slightly increases the model's in-sample R^2, but does not significantly improve its out-of-sample accuracy. For this reason, we exclude additional lags of EPS from the model.

Cross-sectional earnings forecasting model (Hou, van Dijk, and Zhang 2012)

Hou et al. (2012) contribute to the earnings forecasting literature by using the comparative advantages of statistical power and minimal survivorship requirements for cross-sectional models, relative to time-series models. They propose a cross-sectional earnings model that captures significant variation in earnings performance across firms. The main focus of their study is to compare the implied cost of capital estimates derived from their model's forecasts to those derived from analyst earnings forecasts—rather than to make statistical comparisons of out-of-sample forecast accuracy. However, they also compare their model's forecasts to analyst forecasts and document that: (1) consensus analyst forecasts are, on average, significantly more accurate than their model's forecasts for forecast horizons of one to three years ahead; (2) the forecasts from their cross-sectional model, on average, have lower forecast bias and higher earnings response coefficients, relative to analysts' forecasts; (3) forecasts from their model produce more reliable implied cost of capital estimates, relative to forecasts from analysts.

The model proposed by Hou et al., (2012) is:

$$E_{i,t+h} = \alpha_0 + \alpha_1 E_{i,t} + \alpha_2 NegE_{i,t} + \alpha_3 AC_{i,t} + \alpha_4 A_{i,t} + \alpha_5 D_{i,t} + \alpha_6 DD_{i,t} + \varepsilon_{t+h}, \tag{5}$$

where $E_{i,t+h}$ is unscaled dollar earnings and the forecast horizon is $h \in \{1,2,3,4,5\}$. $NegE_{i,t}$ is a dummy for negative earnings, $AC_{i,t}$ is operating accruals, $A_{i,t}$ is total assets, $D_{i,t}$ is the dividend payment and $DD_{i,t}$ is a dummy for dividend payers. The model is estimated using a 10-year rolling window to predict dollar earnings, out-of-sample.[5] The conditioning information set for Hou et al. (2012) model includes non-earnings information—dividends and total assets. In addition, the model allows the persistence of negative earnings to be different from that of positive earnings and allows persistence of accruals to be different from that of cash flows. To the extent that these variables have incremental predictive ability over earnings then this model should, in expectation, outperform both random walk model and the AR model.

Our study is distinct from Hou et al. along three dimensions. First, the model we propose is different from Hou et al.'s model: (1) our model accommodates different predictors (e.g., distributions to debt holders, special items, lagged change in EPS, and an indicator for stock splits); (2) we scale our regression variables to directly predict EPS, whereas Hou et al. use unscaled variables to predict unscaled earnings;[6] and (3) we estimate the model using least absolute deviations, while Hou et al. use OLS. As mentioned earlier, Hou et al. find that consensus analyst forecasts outperform their model's out-of-sample forecasts on accuracy; however, whether consensus analyst forecasts can also outperform our model's forecasts is an empirical question. Second, Hou et al. compare their model's forecasts with analyst forecasts with respect to association with contemporaneous returns (i.e., ERC comparisons). In contrast, we compare our model's forecasts to analyst forecasts with respect to association with future abnormal returns, after controlling for other determinants of future returns. Third, one focus of our study is to compare the out-of sample model bias and accuracy to benchmark models over

[5] Because we compare EPS forecasts rather than unscaled dollar earnings forecasts, we deflate the predicted value of earnings via equation (5) by common shares used to calculate basic EPS (Compustat Item CSHPRI). We then compare this value to actual EPS and predicted EPS from our model to compare forecast accuracy.

[6] Hou et al (2012) assert that analysts EPS forecasts are comparable with dollar earnings. The authors scale the model's dollar earnings forecasts ex post by each firm's end-of-June market equity and divide analysts' EPS forecasts by end-of-June stock price and compare the two variables. The in-sample R^2 of Hou et al.'s model is higher than ours because their variables are unscaled. However, whether this high R^2—combined with ex post scaling of the predicted dollar earnings—translates to superior out-of-sample performance is an empirical question.

various forecast horizons. In contrast, the focus in Hou et al. is to compare the reliability of the implied cost of capital estimates derived from their model relative to those derived from analyst forecasts.

In summary, relative to the three aforementioned extant model-based forecasts (random walk, AR, Hou et al.), we expect superior forecasts from our proposed model for the following reasons. First, random walk and autoregressive forecasts are typically used in the interest of parsimony (e.g., Bradshaw et al. 2012). However, both the random walk and autoregressive models limit the conditioning information set to current-period earnings. In contrast, our model is based on a conditioning information set that is incremental to current period earnings. Second, scaling regression variables presents outlier problems, which, when combined with OLS estimation, requires researchers to use sample selection criteria that may exacerbate selection bias. For example, Fama and French (2000) exclude firms with less than \$10 million in assets or \$5 million in book equity to mitigate the effect of outliers and Hou et al. (2012) rely on unscaled variables to mitigate the impact of outlier observations on scaled earnings (p. 507).[7] The use of OLS estimation thus limits the generalizability of the sample firms by excluding small firms or firms with losses from the estimation sample. To address this issue, we use least absolute deviation (LAD) estimation and find that our estimation procedure has lower measurement error (Figure 2) and is able to generate more accurate earnings forecasts in our out-of-sample forecasting tests.

2.3 Analysts' Forecasts

---

[7] In robustness tests, Hou et al. scale their variables and include additional variables, noting no differences in their results. However, the primary focus of their study is precision in implied costs of capital and, while they do scale, they eliminate firms with less than \$5 million in assets. Our methodology accommodates small firms and we show that this accommodation yields significant improvement in forecast accuracy and bias.

We also compare forecasts from our model to forecasts from security analysts. While the aforementioned models are virtually unaffected by concerns about sampling bias, analysts choose their coverage and, therefore, these samples are subject to selection bias. In order to document these potential concerns regarding analyst coverage, we present coverage and descriptive statistics over four consecutive five-year intervals in Table 1. We emphasize several insights from this analysis.[8] First, a significant portion of firms with available CRSP and Compustat data are not covered by analysts. Specifically, from 1990 to 2004, about one-third of potential observations are excluded when requiring analyst coverage. This percentage decreases in the 2005 to 2009 period, but still remains large (20%).[9] Second, a disproportionate number of firms without analyst coverage are NASDAQ and AMEX firms; specifically, roughly 80% (55%) of the sample not covered by analysts (covered by analysts) are either NASDAQ or AMEX. In addition, untabulated results show that these differences are not concentrated within a particular industry; rather, they come from multiple industries. In addition, firms not covered by analysts are smaller, growth firms with relatively poor performance; however, they are economically important—for example, in 2009, these uncovered firms are aggregately worth $2.86 trillion in market value and they report $3.25 trillion in sales. Taken together, these results

[8] We use the IBES database to capture analyst coverage, which has the widest coverage available over our sample period. Alternative databases (e.g., Zack's or Value Line) likely suffer from similar, and perhaps more severe, selection bias concerns. See Francis, Chen, Philbrick, and Willis (2004) for a more complete discussion of analyst forecast databases.

[9] Due to these coverage concerns—which are exacerbated for long-horizon forecasts—previous accounting research typically enlarges the sample by extracting long-run earnings forecasts from analysts' two-year-ahead earnings forecasts and IBES analysts' forecasts of earnings long-term growth rates (e.g., Frankel and Lee 1998; Bradshaw et al. 2012; Hou et al. 2012). However, analysts' forecasts of long-term earnings growth rates (LTG) are problematic because their forecast horizons, and the corresponding growth forecast errors, are difficult to specify. In addition, it is problematic to interpret LTG when the base fiscal years are loss years. This problem forces researchers to drop loss years from the sample, leading to additional selection bias. Also, Harris (1999) and Chan et al. (2003) show that analyst forecasts of long-term growth lack predictive ability and are overly optimistic. Accordingly, studies exclusively using analysts' forecasts face a severe selection concern, especially for forecasts with greater than two year horizons.

highlight the significance of firms not covered by analysts—in terms of sample size as well as economic magnitude.

3. Sample

Our sample covers firms with required data from 1966 to 2010. Descriptive statistics for the explanatory variables used in our model are presented in Table 2, Panel A. We discuss several aspects of this description. First, there is considerable variability around the mean for each of these variables, as indicated by the large standard deviations. This lends credence to our use of LAD instead of OLS. In particular, Figure 1 compares the actual distribution of EPS in our sample to the normal distribution. This pictorial evidence clearly suggests that the distribution of EPS in our sample is not normal and methodologies assuming the normal distribution should be modified. Second, close to one quarter of the firm-year observations have EPS less than zero (24.0%). This sizable percentage of loss-years is likely due to the lack of requirement for analyst coverage; i.e., analysts tend to follow more successful firms. Third, mean values for debt and equity distributions are negative (-0.024 and -0.058, respectively), indicating that firms, on average, are net issuers of capital. These values are consistent with those reported by Dechow, Richardson, and Sloan (2008). Fourth, about one-half of the firm-years in our sample issue dividends (52.1%) and more than one-third of the firm-years in our sample split stock (37.5%).

4. Results on Statistical Significance of Model-Based Forecasts

4.1 In-Sample Estimation of Cross-Sectional Earnings Forecasting Model

The first column in Panel B of Table 2 shows results for the stage 1 regression, which estimates expected EPS at the one-year horizon (estimates for longer horizons are presented in Panel C). The model has significant explanatory power (R-squared = 39%) and with the exception of the dividend payer dummy, the coefficient estimates are consistent with our predictions and significant in the expected direction, demonstrating the incremental contribution of each predictor variable. The results for the Stage 2 model are presented in the second column. Note that (1) the coefficient estimates on $E[EPS_t]$ and EPS_t are equal in magnitude and opposite in sign, at 0.660 and -0.671, respectively, each significant at the 1% level; and (2) the coefficient on change in EPS is 0.314 and significant at the 1% level. We interpret this result as follows. First, the result supports the hypothesis that EPS reverts to its expected level. In particular, this result suggests that on average EPS reverts to an economy wide expected level at a rate of about 66-67% per year—that is, on average, it takes less than 2 years for EPS to revert to expected levels. Fama and French (2000) report that profitability reverts to an economy-wide mean at a rate of roughly 38%. These rates are substantially different and could be due to any or all of the following: different firms (Fama and French select a sample that excludes small firms and many loss firms), different sample periods—the incidence of losses have increased over time, different stage 1 models, different deflators, or use of LAD instead of OLS. Second, the result that $\beta_2 = -\beta_2$ suggests that our partial adjustment model is reasonably specified, providing empirical support for the model's ability to capture cross-sectional variation in expected EPS. Third, the coefficient estimate for the lagged change in EPS in the stage 2 model is 0.314 (significant at the 1% level), suggesting that the partial adjustment model does not fully capture all variation in earnings, consistent with Fama and French (2000).

Panel C of Table 2 reports results of the proposed model's EPS prediction, in-sample. All the coefficient estimates remain significant in the expected direction, while the R-squared decreases monotonically from 39% (one year ahead) to 16% (five years ahead). In addition, while the coefficient for DIV_DUM is insignificant for one year ahead forecasts, the coefficient is positive and significant, as expected, for longer horizons. These results provide further support for the stability of the proposed model and its ability to explain future earnings, up to five years ahead.

4.2 Incremental Contribution of LAD Estimation Relative to OLS Estimation

In order to support our use of LAD rather than OLS, we compare the 95% confidence interval of LAD estimation to that of OLS estimation, out-of-sample, in Figure 2. Results are presented for one, two, and three years ahead. We highlight two aspects of this analysis. First, the confidence intervals are non-linear, suggesting asymmetry in model performance between profit and loss firms. Specifically, the confidence interval is narrower when current-period EPS is positive, suggesting that forecasting future EPS for firms that report a loss is more difficult than for firms that report a profit. More importantly, the LAD estimation captures this nonlinearity much better than OLS does. Second, the LAD confidence interval is narrower than the OLS confidence interval across all forecast horizons. This evidence suggests that estimation by LAD is more precise than OLS. Third, the confidence interval becomes narrower as the forecast horizons becomes shorter, highlighting that model's out-of-sample accuracy decreases with the length of forecast horizon. Taken together, these results provide further support for the utilization of LAD in our research setting.

4.3 Out-of-Sample Comparison of Cross-Sectional Earnings Forecasting Model vs. Other Models

We first report results comparing the out-of-sample accuracy of the cross-sectional model that we propose with the other models described in section 2—random walk, AR, and Hou et al.'s cross-sectional model. We estimate all the models using a 10-year rolling window with in-sample data beginning in 1966, and out-of-sample forecasts beginning in 1981.[10] We calculate absolute forecast error as: $AFE_i^k = \left(\left|100 * (eps_{i,t+h} - forecast_{i,t+h}^k)/price_{i,t}\right|\right)$. Following previous research, and for purposes of the comparability of our results, we report the Wilcoxon signed-rank test to make accuracy comparisons.

Table 3, Panel A, reports results of accuracy comparisons based on the Wilcoxon signed-rank test. We subtract the AFE of the model indicated in the column from the AFE of the model indicated in the row and report the median of this difference. Under the null that two models are equally accurate, the median difference should not be significantly different from zero and the proportion of negative (positive) differences should not be significantly different from half. Accordingly, a negative median difference indicates the model indicated in the column is more accurate than the model indicated in the row. Statistical significance is based on the Wilcoxon signed-rank test. At one (five) year ahead horizon the median difference in absolute forecast error between the proposed model and random walk is -0.066% (-0.340%) of the stock price in favor of the proposed model, significant at the 1% level—moreover, 76,454 (51,905) of the differences are negative, compared to 64,721 (38,556) positive differences. The median difference between our model and the AR model is -0.365% (-0.483%) of the stock price at one (five) year horizon in favor of the proposed model, significant at the 1% level. Finally, the median difference in absolute forecast error between our model and the Hou et al. model is -

[10] For example, in predicting EPS for five years ahead for 1986, we first estimate coefficients for the 1966 – 1975 time period on EPS five years ahead ending in 1971 – 1980. We then apply these coefficients to 1981 variables to predict 1986 EPS, out-of-sample:

$$\begin{bmatrix} Y_{1971} \\ \vdots \\ Y_{1980} \end{bmatrix} = \boldsymbol{\beta} \begin{bmatrix} \boldsymbol{X_{1966}} \\ \vdots \\ \boldsymbol{X_{1975}} \end{bmatrix} + \begin{bmatrix} e_{1966} \\ \vdots \\ e_{1975} \end{bmatrix}$$

$$\hat{Y}_{1986} = \boldsymbol{\beta}[\boldsymbol{X_{1981}}]$$

0.185% (-1.032%) of the stock price at the one (five) year horizon in favor of proposed the model, significant at the 1% level. Our results suggest that the AR model is more accurate than the random walk model and both models are more accurate than the Hou et al model, out-of-sample.[11] For robustness and completeness, we also compare accuracy using the root mean squared error (Stock and Watson 2002; Gerakos and Gramacy 2012). The results of this test are presented in Panel B and are consistent with the Wilcoxon test results reported in Panel A. Overall, these results provide evidence that our model is significantly more accurate than each of the other three models over every forecast horizon. In addition, the accuracy improvement from our model increases with the length of the forecast horizon.

4.4 Cross-Sectional Model vs. Analysts

In this section, we compare the model's out-of-sample forecasts to consensus analysts' forecasts. While analysts possess both timing and information advantages over statistical approaches, analysts' forecasts are subject to incentive and behavioral biases. Accordingly, whether analysts outperform the proposed model is an empirical question. In all analyses, we use consensus analysts' forecasts rather than individual analyst forecasts because cost of capital and valuation studies commonly use consensus forecasts in their tests.

4.4.1 Out-of-Sample Comparison of Proposed Model to Analysts

Panel A of Table 4 describes the distributions of signed forecast errors (i.e., bias) by the proposed model and analysts, for one to five year forecast horizons.[12] (Figure 3 presents horizontal histograms that further illustrate these distributions.) Consistent with prior research,

---

[11] However, it is important to note that the focus of Hou et al.'s model was not to improve statistical accuracy, but rather to more appropriately capture market expectations and improve implied cost of capital estimates.

[12] In reported results, we use GAAP EPS (per Compustat) to capture actual realizations of earnings for in-sample estimation and out-of-sample comparison. In unreported robustness tests, we use IBES EPS instead of Compustat (GAAP) EPS for in-sample estimation and out-of-sample comparisons with analyst forecasts and obtain qualitatively similar results.

we find that analysts are optimistic—in contrast, the model is slightly pessimistic. For the median firm, at the one (five) year horizon, the bias in consensus analyst forecasts is -0.51% (-2.62%) of the stock price, compared to 0.04% (0.34%) for model forecasts. Furthermore, analyst optimism is much bigger than model pessimism—based on the medians, at one (five) years ahead, analyst optimism is 12.4 (7.6) times bigger than model pessimism. This result is not restricted to the median. For example, at one (five) years ahead, analyst optimism is -13.70% (-20.30%) of the stock price at the 10^{th} percentile, compared to 4.81% (6.93%) for model pessimism at the 90^{th} percentile. Accordingly, as Figure 3 shows, for the mean firm, analyst optimism is also much bigger than the model's pessimism.

Panel B and Figure 4 compares the absolute forecast errors (i.e., accuracy) of consensus forecasts to that of the proposed model, out-of-sample. We find that the model forecasts are more accurate than consensus analyst forecasts (significant at the 1% level), for two to five year forecast horizons. Overall, analysts are significantly more accurate at one year ahead—the median difference in absolute forecast error ahead is 0.048% of the stock price, in favor of the consensus forecast. This accuracy advantage is, however, considerably smaller than the model's accuracy advantage in horizon years two through five. Specifically, the median accuracy advantage of model forecasts over the consensus forecast increases monotonically from -0.092% to -0.841% of the stock price, from two to five years ahead. If we consider, for example, a firm with a price-to-earnings ratio of 15 (say, 15-to-1), this represents an EPS accuracy increase ranging from \$0.01 per share to \$0.13 per share. Figure 4 presents horizontal bar charts that compare the distribution of absolute forecast errors of the model to that of the consensus analyst forecast, at different percentiles. The charts show that below the 25^{th} percentile, differences in accuracy between model forecasts and consensus analyst forecasts are minimal, across all forecast horizons. However, above the 75th percentile, the model is much more accurate, except at the one year horizon. Accordingly, for the mean firm, the model is also more accurate than consensus analyst forecasts. These accuracy comparison results are particularly useful given the use of longer term analysts' forecasts in cost of capital and valuation studies. Taken together, our results suggest that the model we propose has the potential to alleviate sample selection bias in research settings that require analyst forecasts.

4.4.2 Cross-Sectional Determinants of Model Superiority

Researchers' sample selection criteria are context specific and sample selection bias is often the byproduct of unavoidable tradeoffs. In addition, while bias and accuracy are important factors that influence the choice of whether to use statistical or analyst forecasts, researchers consider numerous other factors—such as the timeliness of forecasts and the association of forecasts with stock returns. Accordingly, in this section, we examine the conditions under which our model is likely to outperform analysts. We use a logistic regression to estimate the average marginal effects of firm characteristics on the probability that the proposed model will be more accurate than analysts. The model's dependent variable is a dummy variable equal to one if the absolute forecast error of the model is smaller than that of the consensus median analyst forecast, and zero otherwise. Based on prior literature's findings that analysts fail to fully incorporate contrarian investment signals (e.g., Jegadeesh, Kim, Krische, and Lee 2004), we consider the following firm-specific factors. First, if analysts are unable to fully anticipate the transitory nature of extreme accruals or special items, then the model is more likely to outperform analysts when the magnitude of current-period absolute total accruals (defined as change in net operating assets) and special items is high. Second, if analysts are overoptimistic about: (1) the sustainability of performance for dividend paying firms; and (2) the implications of past growth for future growth, the model is more likely to outperform analysts for firms that pay dividends and for firms with high sales growth and high growth opportunities. Third, if analysts use a broader information set to uncover the underlying causes for a loss or a stock split, then we expect that analysts will on average anticipate the implications of such events for future performance more accurately than our model does. Finally, we control for size (a proxy for the firm's information environment) using the log of total assets. Consistent with these predictions, the results, presented in Table 5, reveal that our model is more likely to outperform analysts for small firms, growth firms, firms with large magnitudes of accruals and special items, firms with fewer losses, firms with fewer stock splits and firms that pay dividends. These results are stable across every horizon except for Tobin's Q, which is negatively associated with the model's success at the one-year horizon, uncorrelated with the model's success at the two-year horizon, and positively associated with the model's success at longer horizons.

Next, we design tests to assess whether analyst characteristics are correlated with our model's success. Specifically, we expect that the following analyst characteristics are associated with forecast accuracy (e.g., Bradshaw 2004). First, we expect that analysts affiliated with a brokerage that has vast experience in following a firm are more likely to outperform the model. Second, we expect the model to outperform analysts when consensus among analysts in low. Third, we expect analysts to outperform the model for firms that receive frequent forecast issuance. Fourth, we expect analysts to outperform the model when forecast revision is large—to the extent that analyst forecast revision reflects new information between earnings report dates, it captures the timing advantage analysts have over the model. Fifth, we control for the number of firms covered by the analyst. Consistent with these predictions, the results presented in Table 6, reveal that our model is more likely to outperform analysts for firms where consensus among analysts is low, forecast revision is small, brokerage experience is low and the frequency of forecast issuance is low. Overall, cross-sectional differences in model superiority have the potential to inform researchers about sample selection choices which are likely context-specific. In particular, our model performs especially well in samples comprised of small, high-growth firms with large magnitudes of accruals, and, when analyst ability is limited, or analyst agreement is low.

5. Results on the Economic Implications of Model-Based Forecasts

In addition to evaluating the statistical performance (i.e., accuracy and bias) of the proposed model relative to three extant models and analysts, we evaluate the economic implications of the model by comparing selected forecasts to future returns. We view the predictive ability of earnings forecasts for future returns as evidence consistent with their economic implications, and their potential to play a price discovery role in the capital markets (Poon and Granger 2003; Lev, Li, and Sougiannis 2010). Accordingly, this section examines the extent to which forecasts from the proposed model can be used to predict the future performance of stocks in capital markets, relative to consensus analyst forecasts, while controlling for other determinants of stock return performance.

Our first set of tests follows Daniel, Grinblatt, Titman and Wermers (1997), hereafter DGTW. Specifically, we rely on a measure of abnormal returns that controls for size, book-to-market equity and momentum and by calculating a benchmark portfolio for each stock each quarter. We form the benchmark portfolios using the following three-way sorting method. First, we sort all CRSP stocks into quintiles based on their market value of equity at the end of June, using breakpoints from NYSE stocks. Second, within each size quintile, we sort stocks into quintiles based on their book-to-market ratios—the ratios are updated each June, with fiscal year end book equity from the preceding calendar year and market value from the prior December. Third, each calendar quarter-end, we sort stocks within each of the 25 size and book-to-market groups into quintiles based on the preceding 12-month return window ending a month prior to the calendar quarter-end, to form 125 (5 x 5 x 5) benchmark portfolios. We then calculate the equally weighted returns for each of the 125 benchmark portfolios over the subsequent quarter. The DGTW characteristics-adjusted abnormal return is obtained by subtracting each specific stock's corresponding benchmark return from the stock's quarterly return.

To compare our model's forecasts to analysts' forecasts with respect to ability to predict future returns, we first note that the implied EPS "growth forecast" is a more appropriate investment signal than the raw EPS level forecast per se, because growth comparisons naturally incorporate current period EPS as a benchmark. Accordingly, we calculate the implied EPS growth forecast as EPS forecast minus current EPS and deflated by the absolute value of current EPS. In unreported robustness tests, we confirm that deflating by book value of equity per share, instead of absolute EPS, gives qualitatively similar results in all our market return tests. Specifically, for each calendar quarter-end from 1981 to 2010, we rank stocks into quintiles based on EPS growth forecasts; we calculate growth as: $(EPS\ forecast_{t+h} - EPS_t)/abs(EPS_t)$, where $EPS\ forecast_{t+h}$ is the h-years ahead consensus EPS forecast (or out-of-sample model forecast) most recently available as of the calendar quarter-end. We denote the event-quarter during which we do this ranking as Quarter 0. Stocks in each growth quintile form an equally weighted portfolio. For each portfolio, we calculate the equally weighted buy and hold DGTW abnormal quarterly return for Quarter 0 as well as for each of the two preceding (Quarters -2, and -1) and four subsequent (Quarters 1, 2, 3 and 4) event quarters. Finally, we calculate the time-series mean and standard errors of the abnormal stock return based on calendar-time portfolios.

We begin by restricting the sample to firms that have both model and analyst forecasts. Table 7 presents the mean DGTW abnormal quarterly return, reported in percent. Panel A reports the results for the 1-year-ahead EPS growth forecasts. First, the results show that two (one) quarters prior to portfolio formation, the difference in abnormal returns between the highest and lowest growth quintiles is 6.30% (7.33%) based on analysts compared to -0.67% (1.60%) based on the model. This result is also true in the quarter of portfolio formation. This result is consistent with analysts' forecasts being driven by momentum; that is, analysts expect past winners to continue to outperform past losers. Our model's forecasts do not appear to be driven by momentum. Second, we find that the ability of analyst based EPS growth to predict future abnormal returns is limited to below six months—one quarter after portfolio formation, the difference in abnormal returns between the highest and lowest quintiles is 0.67%, and it becomes insignificant in subsequent quarters. In comparison, the ability of model based EPS growth to predict future abnormal returns is strongly evident even after 4 quarters ahead—the difference in abnormal returns between the highest and the lowest growth quintiles is 3.16% after one quarter, and 2.59% (significant at the 1% level) after 4 quarters. As panels B and C report, this result remains robust, and becomes even stronger, when we apply EPS forecast horizons of two or three years ahead to calculate and rank growth. Effectively, after controlling for size, book-to-market and momentum, we find that the EPS growth implied by the model's out-of-sample forecasts is positively related to future abnormal returns up to 4 quarters ahead—significant at the 1% level.

We also examine whether the model's ability to predict future abnormal returns is restricted to the sample of firms with both analyst and model forecasts. Accordingly, we repeat the above analysis for the subsample of firms covered by the model, but not by analysts. Table 8 reports that two (one) quarters prior to portfolio formation, the difference in abnormal returns between the highest and lowest growth quintiles is -5.50% (-2.59%), significant at the 1% level. Interestingly, however, this negative abnormal return reverses around Quarter 0 and becomes significantly positive in subsequent quarters. Specifically, the difference in abnormal returns between the highest and the lowest growth quintiles is 1.34% after one quarter, which increases to 1.73% (significant at the 1% level) after 4 quarters. This result also remains robust, and becomes even stronger, when we use EPS forecast horizons of two or three years ahead. We interpret this reversal as suggesting that the model can potentially play a useful price discovery role for stocks that are not covered by analysts. As a whole, these market-based tests provide evidence that the growth forecasts from the model we propose are economically meaningful.

While the DGTW characteristics adjust returns for size, book-to-market and momentum, Jegadeesh, Kim, Krische, and Lee (2004) argue for the inclusion of nine additional investment signals that are known to predict future returns. In our final set of tests, we incorporate these additional investment signals. The signals include earnings momentum (in addition to returns momentum), as well as growth indicators and fundamental signals, among others. These analyses also incorporate characteristic-based controls rather than factor sensitivity-based controls. Accordingly, these tests are designed to test the robustness of our result that forecasts based on our proposed model are economically significant, as measured by future abnormal stock returns. Consistent with Jegadeesh et al., we include the following investment signals as controls in our analysis. (Variable definitions are included in the Appendix.) First, we control for price momentum using RETP, the cumulative market-adjusted return for 6 months before the relevant forecast and RETP2, the cumulative market-adjusted returns for 12 months before the relevant forecast. Second, we include earnings momentum using FREV, analyst forecast revisions for six months prior to the forecast and SUE, which is the unexpected earnings, scaled by standard deviation of earnings over the previous eight quarters. Third, we control for trading volume using TURN, average daily volume turnover for the six months prior to the forecast. Fourth, we include valuation multiples: BP (book-to-price ratio) and EP (earnings-to-price ratio). Fifth, we include SG (sales growth) as a growth indicator. [13] Sixth, we include SIZE, measured as the natural log of market capitalization. Finally, we include fundamental indicators: TA (total accruals divided by total assets), CAPEX (capital expenditures divided by total assets), and LOSS (indicator equal to 1 if EPS is less zero, and 0 otherwise).

Results are presented in Table 9. Column 1 presents results for a baseline model, before including signals provided by either our model or analysts. Results are generally consistent with arguments and results from prior research (e.g., Jegadeesh and Titman 1993; Lakonishok, Shelifer, and Vishny 1994; Lee and Swaminathan 2000). Specifically, returns momentum and earnings momentum are positively associated with future abnormal returns, while share turnover, total accruals, and sales growth are negative associated with future abnormal returns. Coefficients for capital expenditures, size, and valuation multiples are insignificant. Columns 2 and 3 include only the analyst forecasting signal, with or without FREV as an independent variable. The analyst forecasting signal is negative and insignificant in both specifications. In contrast, Column 4 includes only our model's signal, and reveals a positive and significant

[13] We do not include analysts' long-term growth forecasts (LTG) as a growth indicator because it substantially reduces sample size. However, in unreported robustness tests, we include LTG and it does not load significantly. Also, including LTG as an additional investment signal does not alter our conclusions about our model's economic implications.

coefficient (coefficient = 0.022, t = 12.51), indicating that firms in the highest forecasted growth quintile earn a return of 8.8% (4 times 2.2) higher than firms in the lowest forecasted growth quintile. Column 5 includes both the model's and analysts' signal; the coefficient for our model's signal is positive and significant (coefficient = 0.024, t = 13.07), while the coefficient for the analyst signal is negative and marginally significant (coefficient = -0.003, t = -1.73). In summary, these results provide robust evidence that signals about future growth using our cross-sectional earnings forecasting model has predictive ability for future returns, incremental to analyst signals, market signals, and fundamental signals.

6. Conclusion

In this paper, we propose a cross-sectional earnings forecasting model which incorporates the reversion of earnings per share to expected levels using methods that alleviate the influence of outliers. We compare the model's out-of-sample accuracy compared to that of three extant models and consensus analysts' forecasts, while also examining the predictive ability of our model's forecasts for future abnormal returns. We show that our model is significantly more accurate than a random walk model, an AR(1) model, and a cross-sectional model proposed by Hou et al. (2012) at forecast horizons of one to five years ahead. In addition, the improvements increase with the forecast horizon. We find that the proposed model's forecasts outperform consensus analyst forecasts with respect to bias at one to five years ahead. Furthermore, while consensus forecasts have an accuracy advantage over our proposed model's forecasts at the one-year-ahead horizon, we show that our model has a much larger accuracy advantage over consensus forecasts from two to five years ahead. We also document that the forecasts generated from our model have predictive ability for future abnormal returns, relative to analysts' forecasts and other investment signals, which suggests our forecasts have economic implications to investors.

Our results suggest that the proposed model has the potential to alleviate concerns about selection bias. In particular, researchers can use our model to make out-of-sample forecasts for the sample of firms not covered by IBES analysts. The model can also be particularly helpful in settings where there is reason for concern regarding the quality of analyst forecasts. Specifically, we document that our proposed model's forecasts are more likely to outperform analysts for small firms, growth firms, firms with large absolute accruals and special items, firms with few losses, dividend paying firms, and firms with fewer stock splits. Our model is also more likely to outperform analysts when consensus among analysts is low, analyst forecast revision is small, brokerage experience is low and the frequency of analyst forecast issuance is low.

It is important to emphasize that our results do not suggest that our model's forecasts should replace analyst forecasts. Such a suggestion is unwarranted for at least three reasons. First, we compare the model to the consensus median analyst and not to individual analysts; there are individual, or star, analysts whose average forecast accuracy is higher than our model's forecasts. Second, our accuracy comparisons restrict analysts from having a timing advantage over the model. In many applications, this restriction is likely unwarranted. Third, numerous research questions are best answered using timely analyst forecasts —for example, event studies that examine short window market surprises.

Overall, our research helps answer the call from Richardson et al. (2010) for more structure in researchers' forecasting frameworks. Specifically, the use of alternative archival research techniques provides a forecasting model that is relatively simple, in the interest of parsimony, and assumes that earnings revert to expected economy-wide levels. Recent research suggests roles for industry information (Fairfield, Ramnath, and Yohn 2009), macroeconomic factors (Richardson et al. 2010), and a combination of signals from analysts, the market, and financial statements (Gao and Wu 2012) in predicting performance. In addition, Fama and French (2000) show that the reversion of profitability to expected levels is nonlinear and occurs at different speeds depending on the distance from realized to expected earnings. Finally, previous research suggests that quarterly data can be used to improve the timeliness of models that rely on annual data which could alleviate the timing advantage of analysts over models. We do not address these refinements to our model but acknowledge these are fruitful avenues for future research.

REFERENCES

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

Blouin, J., J. Core, and W. Guay. 2010. Have the tax benefits of debt been overestimated? *Journal of Financial Economics* 98: 195-213.

Bradshaw, M. 2004. How do analysts use their earnings forecasts in generating stock recommendations? *The Accounting Review* 79, 1: 25-50.

Bradshaw, M. 2011. Analysts' forecasts: what do we know after decades of work? Working paper, Boston College.

Bradshaw, M., M. Drake, J. Myers, and L. Myers. 2012. A re-examination of analysts' superiority over time-series forecasts of annual earnings. Forthcoming, *Review of Accounting Studies*.

Brooks, L. and D. Buckmaster. 1976. Further evidence on the time series properties of accounting income. *Journal of Finance* 31, 5: 1359-1373.

Brown, L. 1993. Earnings forecasting research: its implications for capital markets research. *International Journal of Forecasting,* 9, 295-320.

Chan, L., Karceski, J., & Lakonishok, J. 2003. The level and persistence of growth rates. *Journal of Finance,* 58, 643-684.

Chee, S., R. Sloan, and A. Uysal. 2012. A framework for value investing. Working paper, University of Oregon and University of California (Berkeley).

Daniel, K., M. Grinblatt, S. Titman, and R. Wermers. 1997. Measuring mutual fund performance with characteristic-based benchmarks. The Journal of Finance LII, 3: 1035-1058.

Dechow, P., S. Richardson, and R. Sloan. 2008. The persistence and pricing of the cash component of earnings. *Journal of Accounting Research* 46, 3: 537-566.

Easton, P. and G. Sommers. 2007. Effect of analysts' optimism on estimates of the expected rate of return implied by earnings forecasts. *Journal of Accounting Research* 45, 5: 983-1015.

Fairfield, P., K. Kitching, and V. Tang. 2009. Are special items informative about future profit margins? *Review of Accounting Studies* 14: 204-236.

Fairfield, P., Ramnath, S., & Yohn, T. 2009. Do industry-level forecasts improve forecasts of financial performance? *Journal of Accounting Research,* 47, 147-178.

Fama, E. and K. French. 1997. Industry costs of equity. *Journal of Financial Economics* 43: 153-193.

Fama, E. and K. French. 2000. Forecasting profitability and earnings. *Journal of Business* 73, 2: 161-175.

Fama, E. and K. French. 2006. Profitability, investment and average returns. *Journal of Financial Economics* 82: 491-518.

Francis, J., Chen, Q., Philbrick, D., & Willis, R. 2004. *Security Analyst Independence*. CFA Institute.

Frankel, R. and C. Lee. 1998. Accounting valuation, market expectation, and cross-sectional stock returns. *Journal of Accounting and Economics* 25, 3: 283-319.

Freeman, R., J. Ohlson, and S. Penman. 1982. Book rate-of-return and prediction of earnings changes: an empirical investigation. *Journal of Accounting Research* 20, 2:639-653.

Gao, Z. and W. Wu. 2012. Predicting long-term earnings growth from multiple information sources. Working paper, Lancaster University and University of Massachusetts (Boston).

Gerakos, J. and R. Gramacy. 2012. Regression-based earnings forecasts. Working paper, University of Chicago.

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

Harris, R.D.F. 1999. The accuracy, bias and efficiency of analysts‟ long run earnings growth forecasts. *Journal of Business Finance & Accounting*, 26, 725-755.

Hou, K., M. van Dijk, and Y. Zhang. 2012. The implied cost of capital: a new approach. *Journal of Accounting and Economics*, forthcoming, doi:10.1016/j.jacceco.2011.12.001

Hughes, J., J. Liu, and W. Su. 2008. On the relation between predictable market returns and predictable analyst forecast errors. *Review of Accounting Studies* 13: 266-291.

Jegadeesh, N., J. Kim, S. Krische, and C.M.C. Lee. 2004. Analyzing the analysts: when do recommendations add value? *Journal of Finance* LIX, 3: 1083-1124.

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

Jones, D. and K. Smith. 2011. Comparing the value relevance, predictive value, and persistence of other comprehensive income and special items. *The Accounting Review*, forthcoming.

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

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

Lee, C. and B. Swaminathan. 2000. Price momentum and trading volume. *Journal of Finance* 55: 2017-2070.

Lev, B. 1983. Some economic determinants of time-series properties of earnings. *Journal of Accounting and Economics* 5: 31-48.

Lev, B., S. Li, and T. Sougiannis. 2010. The usefulness of accounting estimates for predicting cash flows and earnings. *Review of Accounting Studies* 15: 779-807.

Nissim, D. and S. Penman. 2001. Ratio analysis and equity valuation: from research to practice. *Review of Accounting Studies* 6: 109-154.

Poon, S. and C. Granger. 2003. Forecasting volatility in financial markets. *Journal of Economic Literature* XLI: 478-539.

Ramakrishnan, R. and J. Thomas. 1992. Valuation of permanent, transitory, and price-irrelevant components of reported earnings. *Journal of Accounting, Auditing, and Finance* 301-336.

Richardson, S., I. Tuna, and P. Wysocki. 2010. Accounting anomalies and fundamental analysis: a review of recent research advances. *Journal of Accounting and Economics* 50: 410-454.

Stock, J. and M. Watson. 2002. Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics* 20, 2: 147-162.

APPENDIX:
Variable Definitions

Panel A: Variables Used in Model-Based Forecasts[a]

Proposed Model, Random Walk, AR:

EPS	=	Annual Earnings per Share (EPSPX / AJEX)
LOSS	=	Dummy equal to 1 if EPS is negative and zero otherwise
ΔEPS	=	Annual change in *EPS*
AVG_CSE	=	Average book value of common equity [0.5 * (CEQL + lagCEQL)]
DEBT_DIST	=	Net distributions to Debt holders [- (ΔDLC + ΔDLTT)/AVG_CSE]
EQUITY_DIST	=	Net distributions to Equity holders [- (ΔAT – ΔLT – IB)/AVG_CSE]
SPLIT_DUM	=	Dummy for Stock splits (Dummy equals 1 if AJEX > 1 and zero otherwise)
DIV_DUM	=	Dummy for positive dividends (dummy=1 if DV is positive and zero o/w)
SPEC_ITEMS	=	Special and extraordinary items [(SPI + XIDO)/AVG_CSE)]
lnSIZE	=	Natural logarithm of lagged assets [log (lag AT)]

Hou et al. Model:

E	=	Unscaled earnings before extraordinary items (IB)
NegE	=	Dummy for negative earnings (dummy = 1 if IB is negative, and 0 otherwise)
AC	=	Working capital accruals: If OANCF missing, Δ((ACT – CHE) – (LCT – DLC)) – DP, otherwise IB - OANCF
V	=	Market value of equity (PRCC_F*CSHO)
D	=	Unscaled Cash dividends (DV or DVC if DV is missing)
DD	=	Dummy for positive dividends (dummy =1 if D is positive, and zero otherwise)

Panel B: Variables Representing and Firm (Table 5) and Analyst (Table 6) Characteristics[a]

Firm Characteristics

Absolute acccruals	=	Absolute change in NOA, deflated by total assets: [(AT –CHE) – (LT-DLC-DLTT-PSTK)] / AT
Tobin's Q	=	Market-to-assets: (PRCC_F*CSHO+AT-SEQ-TXDB) / AT
Sales growth	=	One-year sales growth: ln(SALE) – ln (lagSALE)
Special items	=	Special items, deflated by total assets: (SPI + XIDO) / AT
Loss dummy	=	Dummy equal to 1 if EPS is negative and zero otherwise
Dividend dummy	=	Dummy equal to 1 if dividends are positive and zero otherwise
Stock split dummy	=	Dummy equal to 1 for a stock split and zero otherwise

Analyst Characteristics

Analyst forecast dispersion	=	Annual mean of standard deviation of one-year ahead EPS forecasts, deflated by absolute value of consensus estimate
Analyst forecast revision	=	Newest median EPS forecasts less oldest median EPS forecast
No. of firm-EPS-forecasts	=	No of forecasts issued by median analyst following firm
No of firms covered	=	No of firms covered by analyst (one-year horizon)
Brokerage-firm experience	=	No of years median brokerage has covered firm

APPENDIX (continued)

Panel C: Variables Representing Investment Signals in Tests of Future Abnormal Returns[b]

Variable	Description	Calculation
RETP	Cumulative market-adjusted return for the preceding 6 months	$\{[\prod_{i=m-6}^{m=1}(1 + \text{monthly ret})] - 1\} - \{[\prod_{i=m-6}^{m=1}(1 + \text{value} - \text{weighted mkt monthly ret})] - 1\}$, where $\text{m} = \text{month} - \text{end of quarter t}$
RET2P	Cumulative market-adjusted returns for the second preceding 6 months	$\{[\prod_{i=m-12}^{m=7}(1 + \text{monthly ret})] - 1\} - \{[\prod_{i=m-12}^{m=7}(1 + \text{value} - \text{weighted mkt monthly ret})] - 1\}$, where $\text{m} = \text{month} - \text{end of quarter t}$
TURN	Average daily volume turnover	Percentile rank of the average of (daily volume / shares outstanding), by exchange, over the previous six months. Daily volume and shares outstanding obtained from CRSP.
SIZE	Natural log of market cap	ln (PRCCQ * CSHOQ)
FREV	Analyst earnings forecast revisions to price	Rolling sum of preceding six months revision-to-price ratios, where revision is measured as the mean consensus IBES forecast for this month minus last month.
LOSS	Loss dummy	Dummy equal to 1 if EPS is negative and zero otherwise
SUE	Standardized unexpected earnings	Change in (EPSFXQ / AJEXQ), deflated by standard deviation of change in earnings over previous eight quarters
SG	Sales growth	Sum of SALEQ for preceding four quarters, deflated by sum of SALEQ for the second preceding four quarters
TA	Total accruals to total assets	ΔNOA, deflated by average total assets, where NOA = (ATQ – CHEQ) – (LTG – DLCQ – DLTTQ – PSDQ)
CAPEX	Capital expenditures to total assets	(CAPEXQ + lag1CAPEXQ + lag2CAPEXQ + lag3CAPEXQ) / Average ATQ where CAPEXQ=CAPXY if first fiscal quarter, otherwise CAPEXQ = Δ CAPEXY
BP	Book to price	CEQQ / (PRCCQ * CSHOQ)
EP	Earnings to price	EPSPXQ / PRCCQ

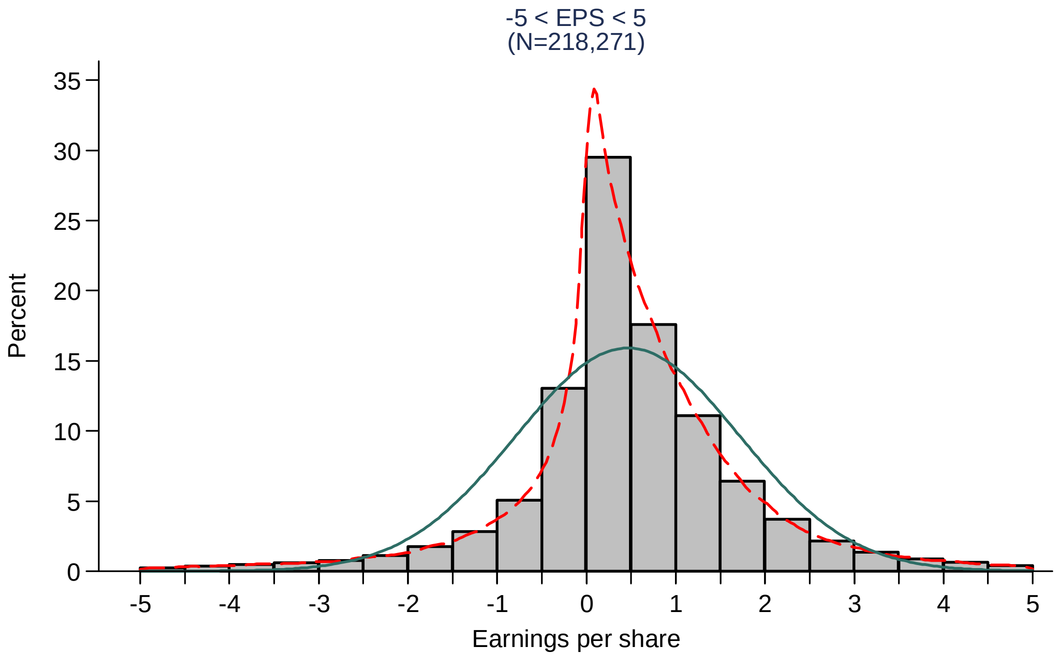
APPENDIX (continued)

Panel D: Compustat Variable Definitions[c]

EPSPX	=	Earnings Per Share (Basic) - Excluding Extraordinary Items
AJEX	=	Adjustment Factor (Company) - Cumulative by Ex-Date
AT	=	Assets – Total
CHE	=	Cash and Short-Term Investments
LT	=	Liabilities – Total
DLC	=	Debt in Current Liabilities – Total
DLTT	=	Long-Term Debt – Total
OANCF	=	Net Cash Flow from Operating Activities
DP	=	Depreciation and Amortization
PSTK	=	Preferred/Preference Stock (Capital) – Total
CEQL	=	Common Equity - Liquidation Value
IVAO	=	Investment and Advances – Other
TSTKP	=	Treasury Stock – Preferrred
DVPA	=	Preferred Dividends in Arrears
IB	=	Income Before Extraordinary Items
SALE	=	Sales/Turnover (Net)
SPI	=	Special Items
XIDO	=	Extraordinary Items and Discontinued Operations
ACT	=	Current Assets – Total
LCT	=	Current Liabilities – Total
PRCC_F	=	Price Close - Annual – Fiscal
PRCCQ	=	Price Close – Quarter
CSHO	=	Common Shares Outstanding
DV	=	Cash Dividends (Cash Flow)
DVC	=	Dividends Common/Ordinary
CAPXY	=	Capital Expenditures
TXDI	=	Income Taxes – Deferred

a – Acronyms represent COMPUSTAT variable names.
b – Investment signals and variable definitions consistent with Jegadeesh et al. 2004.
c – Quarterly versions of variables are amended with "Q" in Panel C.

Figure 1: Comparison of EPS distribution to the normal distribution. This figure presents the distribution of EPS, pooled for the years 1966 to 2010, requiring COMPUSTAT and CRSP data. The continuous solid line overlays the normal distribution while the dotted line overlays the actual distribution.



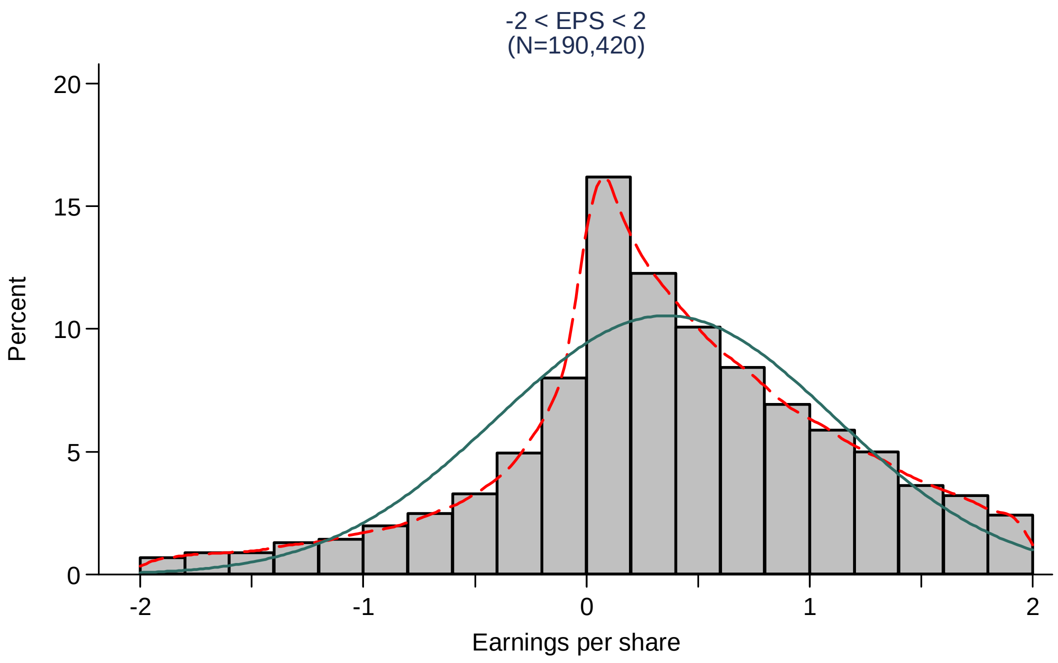


Figure 2: 95% Confidence Intervals for Out-of-Sample Forecasts. This figure plots actual EPS vs. forecasted EPS from our proposed model, separately estimated via LAD and OLS.

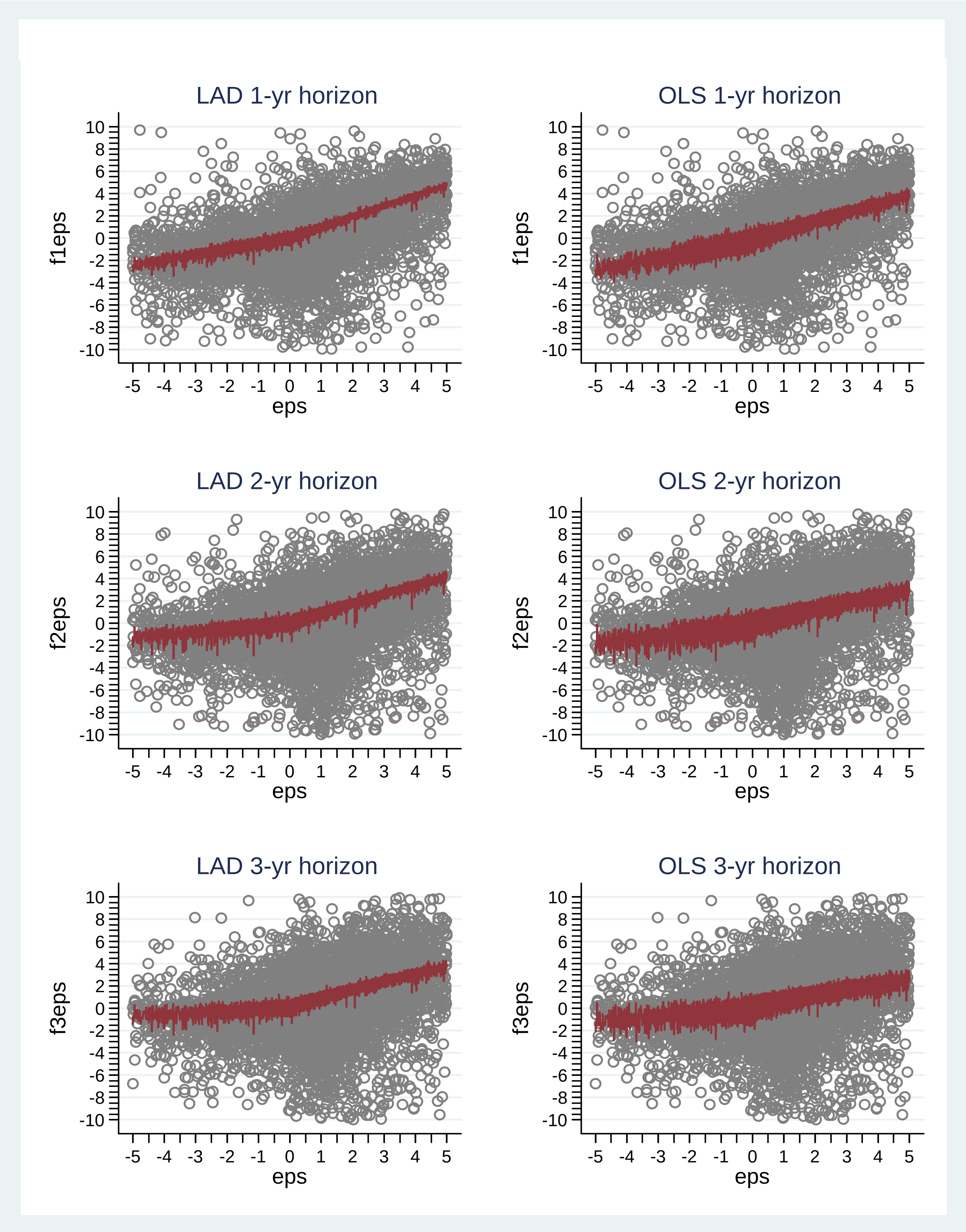


Figure 3. Distribution of Signed Forecast Errors. This table presents frequency histograms of signed forecast errors from our model vs. consensus analyst forecasts. Numerical results and further description of the test is presented in Table 4, Panel A.

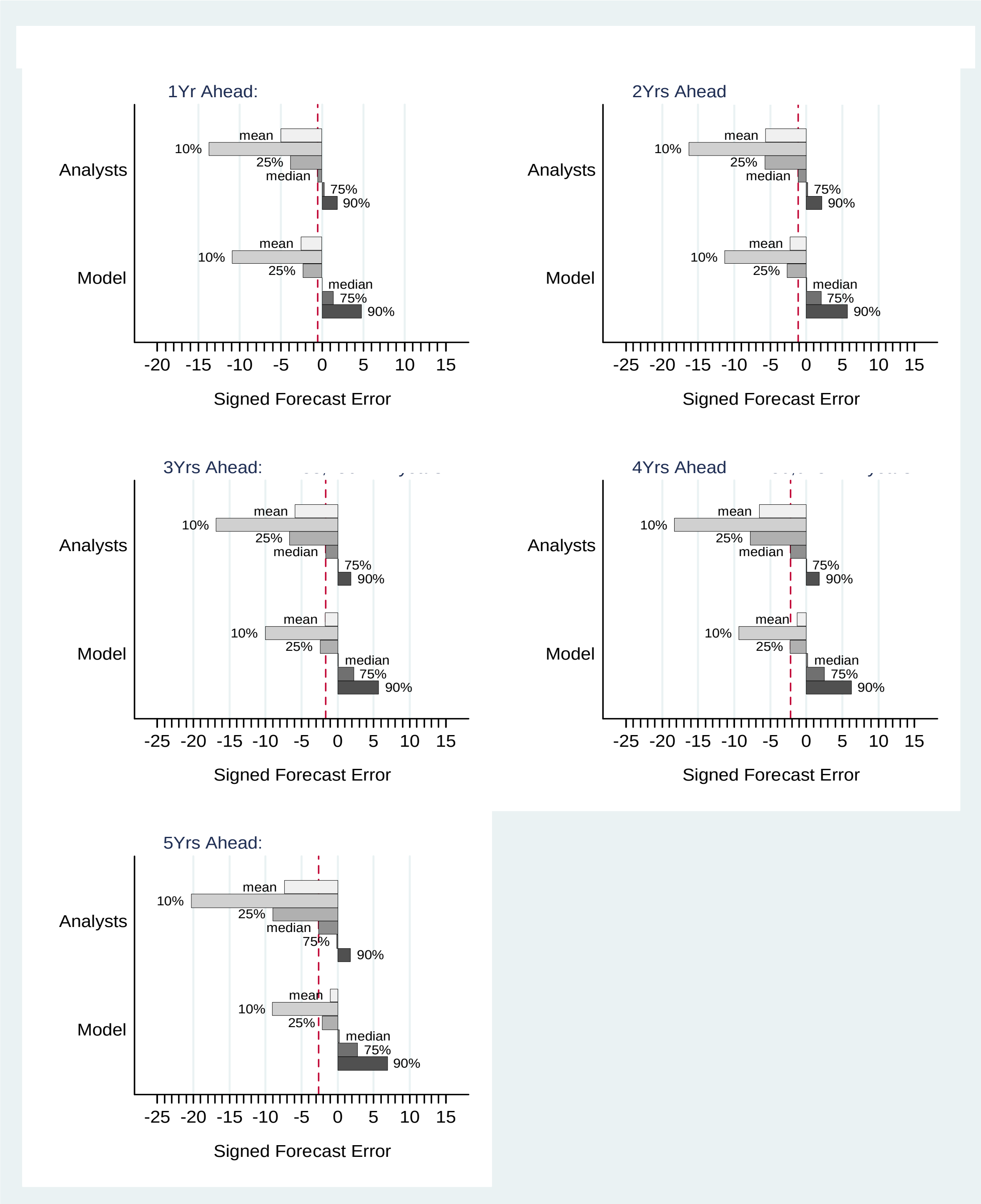


Figure 4. Distribution of Absolute Forecast Errors. This table presents frequency histograms of absolute forecast errors from our model vs. consensus analyst forecasts. Numerical results and further description of the test is presented in Table 4, Panel B.

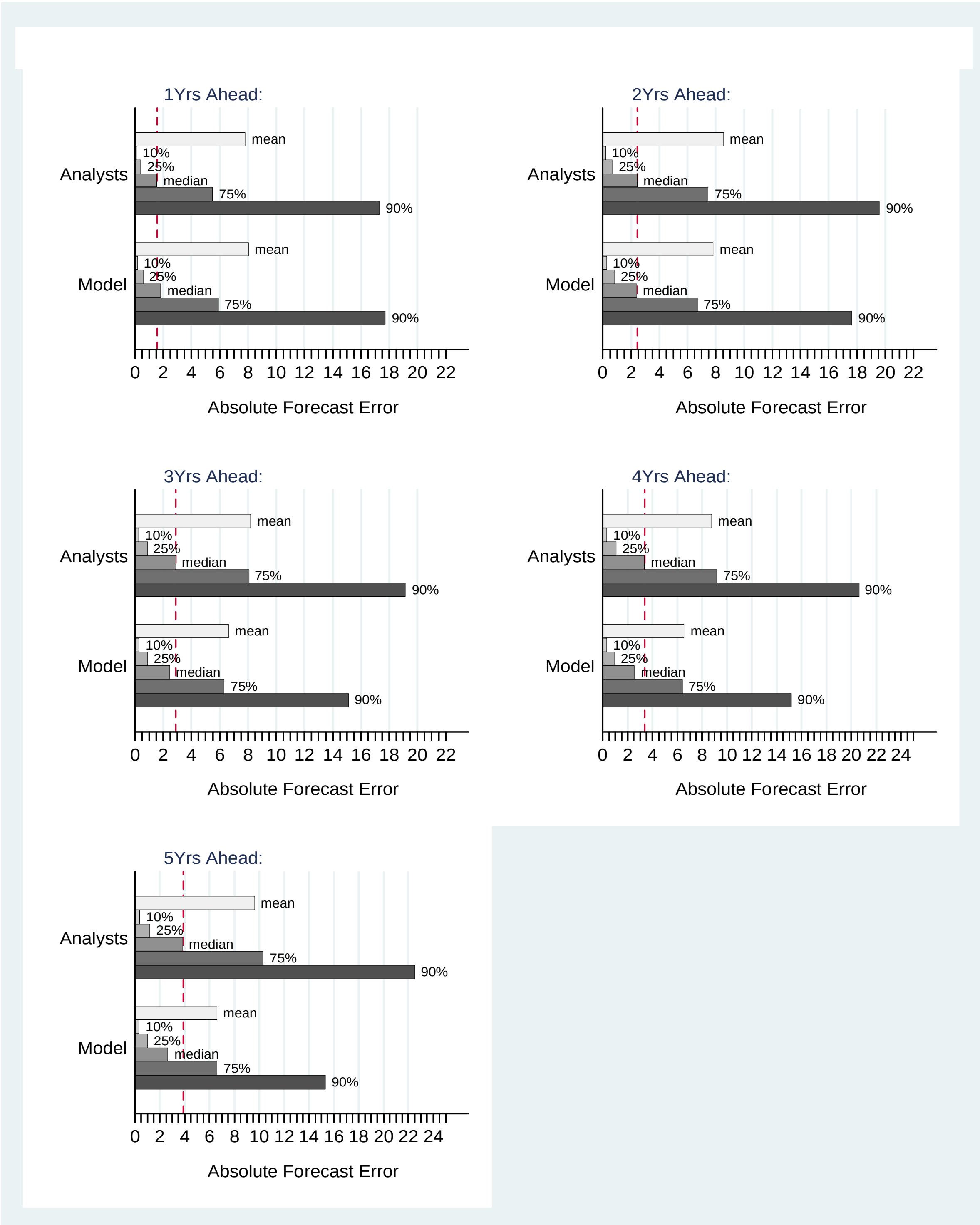


Table 1: Firm characteristics by model and analyst coverage. This table compares characteristics of firms covered by analysts to characteristics of firms covered (out-of-sample) by the proposed model. Coverage is based on 1-year-ahead earnings per share forecasts and the comparison uses firm-years with both COMPUSTAT and CRSP data. The proposed model covers all firms covered by analysts. In contrast, analysts cover a subset of the firms covered by the model. MVE is market value of equity (PRCC_F*CSHO), Earnings is before extraordinary items (IB) and BTM is ratio of book value of equity to market value of equity. $|Accruals|$ is absolute accruals, calculated as absolute change in net operating assets. Net operating assets is non-cash assets (AT - CHE) less non-debt liabilities (LT – DLC – DLTT – PSTK).

PANEL A: 1990 to 1994 (N = 25,159)

	Firms with analyst coverage (N = 16,440)				Firms without analyst coverage (N = 8,719)			
	Assets	MVE	EPS	BTM	Assets	MVE	EPS	BTM
mean	2,584.6	1,176.2	0.416	0.696	333.3	145.9	-0.368	0.987
25%	64.6	63.6	0.099	0.336	8.1	6.8	-0.440	0.313
median	239.6	188.7	0.480	0.582	24.6	16.2	0.027	0.748
75%	1,218.5	751.4	1.040	0.905	86.1	43.6	0.370	1.328
	NASDAQ	NYSE	AMEX	ARCA	NASDAQ	NYSE	AMEX	ARCA
	7,566	8,046	824	4	5,542	1,362	1,810	5

PANEL B: 1995 to 1999 (N = 32,907)

	Firms with analyst coverage (N =23,828)				Firms without analyst coverage (N = 9,079)			
	Assets	MVE	EPS	BTM	Assets	MVE	EPS	BTM
mean	2,795.7	2,099.7	0.425	0.578	576.5	341.6	-0.129	0.790
25%	81.8	83.8	0.030	0.264	14.5	13.2	-0.380	0.325
median	309.2	271.3	0.550	0.478	46.2	31.4	0.120	0.664
75%	1,280.7	1,029.4	1.180	0.772	211.8	103.0	0.655	1.080
	NASDAQ	NYSE	AMEX	ARCA	NASDAQ	NYSE	AMEX	ARCA
	13,337	9,501	981	9	5,941	1,629	1,506	3

PANEL C: 2000 to 2004 (N = 27,904)

	Firms with analyst coverage (N = 19,723)				Firms without analyst coverage (N = 8,181)			
	Assets	MVE	EPS	BTM	Assets	MVE	EPS	BTM
mean	5,771.2	3,545.6	0.128	0.664	1,374.6	654.8	-0.337	0.980
25%	169.8	153.4	-0.180	0.307	24.4	16.3	-0.470	0.407
median	648.6	513.4	0.560	0.524	88.8	41.1	0.130	0.766
75%	2,494.0	1,834.7	1.340	0.822	352.2	118.5	0.780	1.279
	NASDAQ	NYSE	AMEX	ARCA	NASDAQ	NYSE	AMEX	ARCA
	10,792	8,443	477	11	5,285	1,410	1,483	3

PANEL D: 2005 to 2009 (N = 24,262)

	Firms with analyst coverage (N = 19,563)				Firms without analyst coverage (N = 4,699)			
	Assets	MVE	EPS	BTM	Assets	MVE	EPS	BTM
mean	8,262.4	4,706.7	0.661	0.700	2,315.1	886.1	0.077	0.936
25%	223.8	196.1	-0.130	0.313	27.8	22.7	-0.370	0.378
median	879.1	648.3	0.700	0.541	95.7	54.4	0.130	0.694
75%	3,354.2	2,443.5	1.710	0.853	481.2	157.6	0.830	1.130
	NASDAQ	NYSE	AMEX	ARCA	NASDAQ	NYSE	AMEX	ARCA
	10,312	8,510	728	13	3,006	665	1,014	14

Table 2: Cross-Sectional Earnings Forecasting Model. Panel A of this table reports descriptive statistics for variables used for in-sample estimation of the cross-sectional earnings forecasting model. Data coverage is 1966 to 2010. Panel B reports a test that uses two stages to estimate the reversion of earnings per share (EPS) to expected levels. The first stage estimates a fitted EPS conditional on proxies intended to capture cross-sectional differences in expected profitability. The second stage estimates the rate of EPS reversion to its expected value by regressing next-period change in EPS on current-period EPS, the fitted EPS from the first-stage regression and current-period change in EPS. The second stage regression reports the average rate EPS reversion using the coefficients on EPS and E[EPS] – which should be of equal magnitude but of opposite sign, if the first stage regression is reasonably specified. Regression variables beyond the 1st and 99th percentiles are treated as missing. Estimation in both the first and second stage uses least absolute deviations (LAD) as opposed to OLS. Bootstrap standard errors are reported in parenthesis. ***, ** or * indicate significance at the 1, 5 or 10% levels, respectively.

PANEL A: Descriptive Statistics for Variables Used in EPS Reversion Model

	N	Mean	Std. Dev.	Q1	Median	Q3
EPS	188,696	0.455	1.667	0.010	0.400	1.070
Δ EPS	188,696	0.040	1.440	-0.160	0.050	0.280
SPEC_ITEMS	188,696	-0.008	0.045	-0.002	0.000	0.000
EQUITY_DIST	188,696	-0.058	0.232	-0.030	0.000	0.021
DEBT_DIST	188,696	-0.024	0.117	-0.054	0.000	0.019
SIZE	188,696	5.120	2.213	3.466	4.948	6.672
		LOSS	SPLIT_D	DIV_D		
	%	24.0%	37.5%	52.1%		

PANEL B: EPS Reversion Model

Stage 1 Regression		Stage 2 Regression	
Dependent variable is current period EPS		Dep. Variable is next period change in EPS	
lag EPS	0.991***	***EPS***	**-0.671*****
	(0.001)		**(0.001)**
*lag LOSS*EPS*	-0.512***	***E[EPS] fitted value from stage1***	**0.660*****
	(0.001)		(0.002)
lag ΔEPS	-0.042***	*Δ EPS*	0.314***
	(0.001)		(0.001)
lag DEBT_DIST	0.178***		--
	(0.006)		--
lag EQUITY_DIST	0.142***		--
	(0.003)		--
lag SPLIT_DUM	0.065***		--
	(0.002)		--
lag DIV_DUM	-0.002		--
	(0.002)		--
lag SPEC_ITEMS	-1.459***		--
	(0.017)		--
ln(lag SIZE)	0.011***		--
	(0.000)		--
Intercept	-0.056***		0.034***
	(0.002)		(0.001)
No. Observations	194,279		177,654
Pseudo R-Squared	0.39		0.04

Table 2 (continued): Cross-Sectional Earnings Forecasting Model. Panel B of this table reports in-sample forecasting regressions of the proposed model. The dependent variable is earnings per share, one to five years ahead. In each regression, continuous variables beyond the 1^{st} and 99^{th} percentiles are treated as missing. Each regression is estimated using least absolute deviations (LAD) as opposed to OLS. Bootstrap standard errors are reported in parenthesis. ***, ** and * indicate significance at the 1%, 5% or 10% levels, respectively.

PANEL C: EPS Reversion Model Across Five Forecast Horizons

Explanatory variables	Forecast Horizon				
(all in the current period)	1yr	2yrs	3yrs	4yrs	5yrs
EPS	0.991***	0.891***	0.811***	0.737***	0.689***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
*LOSS*EPS*	-0.511***	-0.641***	-0.661***	-0.629***	-0.601***
	(0.001)	(0.002)	(0.002)	(0.003)	(0.003)
ΔEPS	-0.043***	-0.073***	-0.075***	-0.065***	-0.070***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
DEBT_DIST	0.184***	0.291***	0.281***	0.287***	0.272***
	(0.006)	(0.010)	(0.013)	(0.015)	(0.016)
EQUITY_DIST	0.137***	0.223***	0.227***	0.221***	0.217***
	(0.003)	(0.005)	(0.007)	(0.008)	(0.009)
SPLIT_DUM	0.066***	0.113***	0.146***	0.171***	0.182***
	(0.002)	(0.002)	(0.003)	(0.004)	(0.004)
DIV_DUM	0.003	0.019***	0.036***	0.048***	0.058***
	(0.002)	(0.003)	(0.004)	(0.004)	(0.004)
SPEC_ITEMS	-1.451***	-1.089***	-0.791***	-0.715***	-0.556***
	(0.017)	(0.027)	(0.036)	(0.044)	(0.049)
ln (SIZE)	0.012***	0.029***	0.042***	0.055***	0.066***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
Intercept	-0.061***	-0.114***	-0.155***	-0.188***	-0.208***
	(0.002)	(0.003)	(0.004)	(0.005)	(0.005)
No. Observations	188,696	172,100	156,910	142,980	130,275
Pseudo R-Squared	0.39	0.27	0.21	0.18	0.16

Table 3: Out-of–sample comparison to extant models. Panel A of this table uses pair wise comparisons of differences in absolute forecast error to estimate the out-of-sample accuracy of random walk (RWK), autoregressive (AR1), Hou et al. and the proposed model. The out-of-sample absolute forecast error by model k is calculated as: $AFE_i^k = |100 * (eps_{i,t+h} - forecast_{i,t+h})/price_{i,t}|$, where $eps_{i,t+h}$ is firm i's actual GAAP earnings per share for fiscal-year $t + h$; $forecast_{i,t}^k$ is the out-of-sample forecast of $eps_{i,t+h}$ by model *k*; and $price_{i,t}$ is firm i's stock price at the end of fiscal-year *t*. We subtract the AFE of the model in the first column from the AFE of the model in the top row and report the cross-sectional median of this difference (med-dif) as a percentage of the stock price. Statistical significance is based on the Wilcoxon signed-rank test. The table also reports the number of negative differences (neg-dif) and the number of positive differences (pos-dif). Under the null that the matched model pairs are equally accurate, the median difference should not be significantly different from zero and the proportion of positive (negative) differences should be one-half. If the median of the differences is significantly negative (positive) then the model indicated in the top row is more (less) accurate than the one in the first column. ***, ** or * indicate significance at the 1%, 5% or 10% levels. For example, when the forecast horizon is two years ahead, the $median(AFE_i^{proposed\ model} - AFE_i^{RWK})$ is -0.149% of the stock price, and 70,326 (55,025) of the AFE differences are negative (positive), which rejects the null that the proposed model and random walk are equally accurate, at the 1% level.

PANEL A: Difference in absolute forecast errors

	proposed model		
RWK	med-dif	neg-dif	pos-dif
1yr	-0.066***	76,454	64,721
2yrs	-0.149***	70,326	55,025
3yrs	-0.230***	64,403	48,003
4yrs	-0.297***	58,010	42,854
5yrs	-0.340***	51,905	38,556

	proposed model			RWK		
AR1	med-dif	neg-dif	pos-dif	med-dif	neg-dif	pos-dif
1yr	-0.365***	86,370	54,805	-0.265***	81,191	59,984
2yrs	-0.440***	75,492	49,859	-0.208***	68,033	57,318
3yrs	-0.474***	67,617	44,789	-0.152***	59,428	52,978
4yrs	-0.486***	60,848	40,016	-0.118***	52,547	48,317
5yrs	-0.483***	54,712	35,749	-0.087***	46,532	43,929

	proposed model			RWK			AR1		
Hou	med-dif	neg-dif	pos-dif	med-dif	neg-dif	pos-dif	med-dif	neg-dif	pos-dif
1yr	-0.185***	82,957	58,218	-0.112***	77,692	63,483	0.088***	67,028	74,147
2yrs	-0.308***	75,268	50,083	-0.124***	67,937	57,414	-0.021***	63,169	62,182
3yrs	-0.511***	69,699	42,707	-0.164***	61,297	51,109	-0.165***	59,272	53,134
4yrs	-0.769***	64,515	36,349	-0.259***	56,121	44,743	-0.379***	55,828	45,036
5yrs	-1.032***	59,117	31,344	-0.433***	51,085	39,376	-0.589***	51,929	38,532

Table 3 (continued): Out-of-sample comparison to extant models. Panel B of this table uses comparisons of relative squared errors (RSFE) to estimate the out-of-sample accuracy of random walk, autoregressive, Hou et al. (2012) and the proposed model. We calculate the squared forecast error by model k as: $SFE^{k}_{i,t+h} = \left(eps_{i,t+h} - forecast_{i,t+h}\right)^{2}$, where $eps_{i,t+h}$ is firm *i*'s actual GAAP earnings per share, h-years-ahead and $forecast^{k}_{i,t}$ is the out-of-sample by model *k*. In this table we subtract, the SFE of the model in the first columns from the SFE of the model in the top row and report the cross-sectional median of this difference. We also report the number of observations in which this difference is smaller (bigger) than zero. Under the null that two models are equally accurate, the median SFE should not be significantly different from one and the number of observations in which the difference is smaller than zero and for 50% of the observations, the difference should be smaller (bigger) than zero. However, if the reported median SFE is smaller (bigger) than zero, then the model in the row is more (less) accurate than model in the first. For example, when the forecast horizon is two years ahead, the median SFE of the proposed model to the random walk model is -0.069, and 76,480 (64,779) of the SFE differences are negative (positive). ***, ** or * indicate significance at the 1%, 5% or 10% levels.

PANEL B: Difference in squared forecast errors

	proposed model		
RWK	med-dif	neg-dif	pos-dif
1yr	-0.040***	76,480	64,779
2yrs	-0.069***	70,284	54,997
3yrs	-0.074***	64,281	47,970
4yrs	-0.073***	57,867	42,807
5yrs	-0.075***	51,773	38,488

	proposed model			RWK		
AR1	med-dif	neg-dif	pos-dif	med-dif	neg-dif	pos-dif
1yr	-0.004***	86,621	54,638	0.036***	81,348	59,911
2yrs	-0.008***	75,696	49,585	0.060***	68,129	57,152
3yrs	-0.012***	67,709	44,542	0.061***	59,474	52,777
4yrs	-0.016***	60,911	39,763	0.058***	52,548	48,126
5yrs	-0.020***	54,733	35,528	0.054***	46,541	43,720

	proposed model			RWK			AR1		
Hou	med-dif	neg-dif	pos-dif	med-dif	neg-dif	pos-dif	med-dif	neg-dif	pos-dif
1yr	-0.060***	83,006	58,253	-0.019***	77,690	63,569	-0.056***	66,914	74,345
2yrs	-0.114***	75,198	50,083	-0.045***	67,837	57,444	-0.105***	62,933	62,348
3yrs	-0.164***	69,543	42,708	-0.090***	61,155	51,096	-0.151***	59,014	53,237
4yrs	-0.220***	64,323	36,351	-0.147***	55,957	44,717	-0.205***	55,582	45,092
5yrs	-0.277***	58,925	31,336	-0.203***	50,914	39,347	-0.257***	51,678	38,583

Table 4: Out-of-sample comparison to analysts. Panel A reports the distribution of unsigned forecast error: $FE_i = 100 * (eps_{i,t+h} - forecast_{i,t+h})/price_{i,t}$, where $eps_{i,t+h}$ is firm *i*'s actual GAAP earnings per share for fiscal-year $t + h$; $forecast_{i,t+h|t}$ is either the median analyst forecast or the out-of-sample forecast by the proposed model and $price_{i,t}$ is firm *i*'s stock price at the end of fiscal-year *t*. The unsigned forecast error captures the forecast bias as a percentage of the stock price—negative (positive) forecast error indicates optimism (pessimism). Model and analyst absolute forecast errors are separately trimmed at 1% and 99%. To be included in this test, each observation is required to have the model forecast and the analyst forecast. To mitigate the timing advantage of analysts over the model, the earliest IBES consensus median forecast is used. In addition, for a h-years ahead forecast, we require that this consensus median forecast be within $5 + h$ months following the preceding fiscal year. For example, for 3-years-ahead forecasts, the consensus median forecast must be within the 8 months following the preceding fiscal year end. Panel B reports the tests of equal accuracy between the proposed model's out-of-sample forecasts and consensus median analyst forecasts, using matched pairs of forecasts. Each matched pair consists of the model's and the median analyst's absolute forecast error: $AFE_i = |FE_i|$. Column (3) reports the cross-sectional medians of absolute forecast error by model and by analyst. Column (4) reports $median\left(AFE_i^{proposed\ model} - AFE_i^{analyst}\right)$, while column(5) reports the Z-statistic based on the Wilcoxon signed-rank test. Column (6) reports the number of negative differences while column (7) reports the number of positive difference. Finally, column (8) reports the sum of negative differences while column (9) reports the sum of positive differences. Under the null that the model and consensus median analyst are equally accurate, the median of paired differences should be insignificantly different from zero and the proportion of negative (positive) differences should be insignificantly different from one-half. ***, ** or * indicate significance at the 1%, 5% or 10% levels.

PANEL A: Bias

		10%		25%		median		75%		90%	
year	No. Obs	model	analyst	model	analyst	model	analyst	model	analyst	model	analyst
1yr	90,907	-10.92	-13.70	-2.33	-3.84	0.04	-0.51	1.42	0.30	4.81	1.82
2yrs	74,526	-11.33	-16.30	-2.67	-5.76	0.11	-1.15	2.05	0.25	5.78	2.16
3yrs	58,130	-10.04	-16.85	-2.44	-6.66	0.17	-1.67	2.25	0.12	5.78	1.92
4yrs	50,078	-9.37	-18.27	-2.25	-7.79	0.27	-2.15	2.58	0.02	6.33	1.90
5yrs	43,765	-9.02	-20.30	-2.12	-9.01	0.34	-2.62	2.83	-0.08	6.93	1.86

PANEL B: Accuracy

(1)	(2)	(3)		(4)	(5)	(6)	(7)	(8)	(9)
		median AFE						sum-neg	sum-pos
horizon	No. Obs	model	analyst	med-dif	z-stat	neg-dif	pos-dif	(in mill)	(in mill)
1yr	90,373	1.856	1.575	0.048***	8.848	42,433	47,940	1,970	2,110
2yrs	74,066	2.405	2.446	-0.092***	-29.218	39,038	35,028	1,540	1,200
3yrs	57,773	2.454	2.892	-0.347***	-50.906	32,932	24,841	1,040	630
4yrs	49,779	2.585	3.381	-0.594***	-60.082	29,474	20,305	812	427
5yrs	43,488	2.673	3.868	-0.841***	-64.906	26,577	16,911	643	303

Table 5: The effect of firm characteristics on model's ability to outperform analysts, out of sample
This table uses a logistic regression to estimate the average marginal effects of firm characteristics on the probability that the proposed model outperforms analysts. The dependent variable is a dummy variable equal to one if the absolute forecast error of the model is smaller than that of the consensus median analyst forecast, and zero otherwise. To mitigate the timing advantage of analysts over the model, the earliest IBES consensus median forecast is used. In addition, for a *h*-years ahead forecast, we require that this consensus median forecast be within 5 + *h* months following the preceding fiscal year. For example, for 3-years-ahead forecasts, the consensus median forecast must be within the 8 months following the preceding fiscal year end. Accruals are calculated as change in net operating assets ((AT – CHE) – (LT – DLC – DLTT – PSTK)); special items is SPI + XIDO; Loss equals one if net income is negative, and zero otherwise. Sales growth is calculated as ln(SALE) – ln(lagSALE). Tobin's Q is calculated as (PRCC_F*CSHO + AT - SEQ - TXDB)/AT. The dividend payer dummy equals one if the firm is a dividend payer and zero otherwise; the stock split dummy equals one if AJEX > 1 and zero otherwise. Size is total assets (AT). Absolute accruals and special items are deflated by average total assets. To mitigate the effect of outliers, the continuous explanatory variables are trimmed at 1% and 99% levels. The table reports t-statistics in parenthesis. The standard errors are clustered at the firm level. The notation *, ** and *** indicates statistical significance at the 1%, 5% and 5% level, respectively.

	Average Marginal Effect on Prob(model is more accurate than analysts)				
	1yr-ahead	2yrs-ahead	3yrs-ahead	4yrs-ahead	5yrs-ahead
Absolute Accruals	0.064***	0.091***	0.176***	0.180***	0.219***
	(3.907)	(4.694)	(7.461)	(6.774)	(7.164)
Special Items	0.255***	0.318***	0.360***	0.434***	0.402***
	(5.019)	(5.311)	(4.554)	(4.736)	(4.008)
Loss dummy	-0.008	-0.032***	-0.033***	-0.028***	-0.019*
	(-1.503)	(-4.787)	(-4.049)	(-3.015)	(-1.801)
Sales growth	0.028***	0.069***	0.090***	0.105***	0.137***
	(3.697)	(7.543)	(7.773)	(7.663)	(8.837)
Tobin's Q	-0.018***	-0.002	0.005**	0.010***	0.012***
	(-11.205)	(-1.035)	(2.352)	(3.718)	(3.795)
Dividend payer dummy	0.026***	0.027***	0.025***	0.029***	0.029***
	(5.559)	(4.735)	(3.681)	(3.743)	(3.290)
Stock split dummy	-0.069***	-0.136***	-0.157***	-0.158***	-0.148***
	(-15.954)	(-26.864)	(-26.291)	(-23.785)	(-20.382)
ln(Size)	-0.027***	-0.026***	-0.026***	-0.032***	-0.031***
	(-22.660)	(-18.483)	(-15.066)	(-16.046)	(-13.977)
No. Observations	82,382	67,500	52,748	45,491	39,761
Pseudo R-Squared	0.010	0.020	0.028	0.034	0.035

Table 6: The effect of analyst attributes on model's ability to outperform analysts, out of sample
This table uses a logistic regression to estimate the average marginal effects of analyst attributes on the probability that the proposed model outperforms analysts. The dependent variable is a dummy variable equal to one if the absolute forecast error of the model is smaller than that of the consensus median analyst forecast, and zero otherwise. To mitigate the timing advantage of analysts over the model, the earliest IBES consensus median forecast is used. In addition, for a *h*-years ahead forecast, we require that this consensus median forecast be within 5 + *h* months following the preceding fiscal year. For example, for 3-years-ahead forecasts, the consensus median forecast must be within the 8 months following the preceding fiscal year end. Analyst forecast dispersion is calculated as the annual mean of the standard deviation divided by absolute mean analyst forecast (STDEV/abs(MEANEST)). Analyst forecast revision is the newest median EPS forecast less the earliest median EPS forecast, divided by stock price at the beginning of the current fiscal year. Number of firm-EPS-forecasts is the number forecasts issued by the median analyst following a firm. Number of firms covered by analyst is calculated at the 1-year-ahead horizon. Brokerage-firm experience is the number of years in which the median brokerage has been following a firm. Forecast dispersion, revision, number of firm-EPS forecasts, number of firms covered by analyst and brokerage-firm experience are calculated at the 1-year-ahed forecast horizon. To mitigate the effect of outliers, the explanatory variables are trimmed at 1% and 99% levels. T-statistics are in parenthesis. The standard errors are clustered at the firm level. The notation *, ** and *** indicates statistical significance at the 1%, 5% and 5% level, respectively.

	Average Marginal Effect on Prob (model is more accurate than analyst)				
	1yr-ahead	2yrs-ahead	3yrs-ahead	4yrs-ahead	5yrs-ahead
Analyst forecast dispersion	0.1087***	0.0663***	0.0481***	0.0336***	0.0254**
	(14.440)	(8.535)	(5.311)	(3.167)	(2.112)
Analyst forecast revision	-0.1263***	-0.1042***	-0.0874***	-0.0774***	-0.0544***
	(-27.574)	(-23.946)	(-19.662)	(-16.399)	(-11.804)
Number of Firm-EPS-forecasts	-0.0006***	-0.0003***	-0.0003***	-0.0006***	-0.0007***
	(-6.458)	(-3.046)	(-2.856)	(-4.512)	(-4.867)
Number of firms covered by analyst	-0.0002	-0.0011***	-0.0013***	-0.0010***	-0.0010***
	(-0.940)	(-5.065)	(-4.841)	(-3.403)	(-3.382)
Brokerage firm-experience	-0.0025***	-0.0063***	-0.0069***	-0.0103***	-0.0113***
	(-3.451)	(-7.794)	(-7.563)	(-9.693)	(-9.674)
No. Observations	75491	64739	55565	47900	41812
Pseudo R-Squared	0.031	0.020	0.014	0.013	0.011

Table 7: Stock performance grouped by growth for firms covered by both analysts and model. We require each observation to have an IBES median analyst earnings per share (EPS) forecast and an out-of-sample EPS forecast from the proposed model. Each calendar quarter from 1981 to 2010, we independently sort stocks into quintiles based on earnings per share (EPS) growth forecasts calculated using either the proposed model's or median analyst EPS forecasts. We calculate EPS growth forecast as: $(EPS\ forecast_{t+h} - EPS_t)/abs(EPS_t)$, where $EPS\ forecast_{t+h}$ is either the analysts' or the model's out-of-sample EPS forecast, h-years ahead. We denote the event-quarter during which the sorting is done as Qtr 0. We sort the EPS growth forecast based on the most recent median analyst forecast issued during the event-quarter and the most recent out-of-sample model forecast available as of the event-quarter. Stocks in each quintile form an equally weighted portfolio. For each stock in this portfolio, we calculate the characteristics adjusted buy-and-hold quarterly abnormal stock return (adjusted for size, book-to-market equity and prior one-year-return using 5 x 5 x 5 portfolios) for each of the preceding two event-quarters and for each of the subsequent four event-quarters. The characteristics adjusted abnormal stock return is reported in percent. The notation *, **, and *** indicates statistical significance at the 10%, 5% and 1% levels, respectively. Time-series t-statistics are in parentheses. Panels A, B and C report the abnormal stock returns for the 1-year-ahead, 2-years-ahead and 3-years-ahead EPS growth forecasts, respectively.

PANEL A (N =344,524 firm-qtrs): Stock performance grouped by 1-yr-ahead forecasts of earnings growth

Growth forecast				**Event-Quarter**			
Model	Qtr -2	Qtr -1	Qtr 0	Qtr 1	Qtr 2	Qtr 3	Qtr 4
Quintile1 (Low)	0.326	-1.078***	-1.851***	-2.113***	-1.840***	-1.543***	-1.342***
Quintile2	-0.173	-0.597***	-1.035***	-1.188***	-1.132***	-1.011***	-0.979***
Quntile3	1.221***	0.906***	0.492**	0.369*	0.318*	0.27	0.096
Quintile4	1.801***	1.792***	1.559***	1.295***	1.331***	1.194***	1.137***
Quintile5 (High)	-0.342	0.521	0.960***	1.049***	1.097***	1.242***	1.246***
High - Low	**-0.667***	**1.599*****	**2.811*****	**3.161*****	**2.937*****	**2.785*****	**2.589*****
	(-1.736)	**(4.297)**	**(7.826)**	**(8.893)**	**(8.894)**	**(7.918)**	**(7.533)**
Analysts							
Quintile1 (Low)	-3.884***	-4.614***	-3.423***	-0.734***	-0.129	-0.22	-0.316*
Quintile2	-0.564**	-1.012***	-0.738***	-0.195	-0.152	0.006	0.054
Quntile3	2.009***	1.677***	0.829***	-0.064	-0.113	0.202	0.335
Quintile4	2.865***	2.787***	1.734***	0.470**	0.447**	0.419**	0.155
Quintile5 (High)	2.415***	2.711***	1.724***	-0.067	-0.281	-0.258	-0.071
High - Low	**6.299*****	**7.325*****	**5.146*****	**0.667*****	**-0.152**	**-0.038**	**0.245**
	(15.218)	**(20.026)**	**(16.156)**	**(2.726)**	**(-0.624)**	**(-0.136)**	**(0.889)**

Table 7 (continued): The performance of stocks with different earnings growth forecasts.

PANEL B (N =273,703 firm-qtrs): Stock performance grouped by 2-yrs-ahead forecasts of earnings growth

Growth forecast				**Event-Quarter**			
Model	Qtr -2	Qtr -1	Qtr 0	Qtr 1	Qtr 2	Qtr 3	Qtr 4
Quintile1 (Low)	0.352*	-1.055***	-1.990***	-2.399***	-2.069***	-1.961***	-1.904***
Quintile2	0.016	-0.443**	-0.891***	-1.131***	-1.207***	-1.083***	-1.135***
Quintile3	1.367***	1.168***	0.715***	0.409**	0.414**	0.218	0.038
Quintile4	2.041***	2.178***	1.997***	1.535***	1.494***	1.245***	1.144***
Quintile5 (High)	0.78**	1.552***	1.711***	1.554***	1.555***	1.631***	1.452***
High - Low	**0.428**	**2.606*** **	**3.701*** **	**3.953*** **	**3.624*** **	**3.592*** **	**3.356*** **
	(1.153)	**(6.448)**	**(9.537)**	**(9.874)**	**(9.645)**	**(9.731)**	**(9.546)**
Analysts							
Quintile1 (Low)	-3.131***	-3.743***	-2.545***	-0.398**	-0.068	-0.199	-0.26
Quintile2	-0.437**	-0.682***	-0.633***	-0.359	-0.162	-0.072	-0.043
Quintile3	2.151***	1.651***	0.841***	-0.011	0.024	0.098	-0.075
Quintile4	2.656***	2.843***	1.920***	0.627**	0.483**	0.345	0.155
Quintile5 (High)	3.322***	3.336***	1.963***	0.105	-0.094	-0.126	-0.187
High - Low	**6.453*** **	**7.079*** **	**4.508*** **	**0.503**	**-0.026**	**0.073**	**0.074**
	(14.135)	**(16.616)**	**(11.506)**	**(1.483)**	**(-0.079)**	**(0.215)**	**(0.237)**

PANEL C (N =223,361 firm-qtrs): Stock performance grouped by 3-yrs-ahead forecasts of earnings growth

Growth forecast				**Event-Quarter**			
Model	Qtr -2	Qtr -1	Qtr 0	Qtr 1	Qtr 2	Qtr 3	Qtr 4
Quintile1 (Low)	0.102	-1.056***	-1.853***	-2.248***	-1.975***	-1.964***	-1.941***
Quintile2	0.158	-0.292***	-0.615***	-0.762***	-0.916***	-0.818***	-0.891***
Quintile3	1.294***	1.122***	0.819***	0.570***	0.532***	0.416**	0.204
Quintile4	2.282***	2.283***	2.03***	1.565***	1.513***	1.454***	1.288***
Quintile5 (High)	1.800***	2.323***	2.405***	1.925***	1.754***	1.859***	1.864***
High - Low	**1.697*** **	**3.308*** **	**4.258*** **	**4.172*** **	**3.729*** **	**3.823*** **	**3.805*** **
	(4.661)	**(8.676)**	**(10.614)**	**(10.535)**	**(9.812)**	**(10.747)**	**(11.175)**
Analysts							
Quintile1 (Low)	-2.652***	-3.058***	-1.829***	-0.156	-0.035	-0.04	-0.083
Quintile2	-0.346*	-0.639***	-0.400*	0.039	0.091	0.08	0.136
Quintile3	1.712***	1.423***	0.717***	0.15	0.258	0.278	0.098
Quintile4	3.032***	3.039***	1.992***	0.645***	0.397*	0.434	0.202
Quintile5 (High)	3.894***	3.620***	2.307***	0.366	0.192	0.192	0.168
High - Low	**6.546*** **	**6.678*** **	**4.136*** **	**0.522**	**0.227**	**0.232**	**0.251**
	(12.456)	**(13.421)**	**(8.740)**	**(1.264)**	**(0.591)**	**(0.565)**	**(0.663)**

Table 8: Stock performance grouped by growth for firms covered by model but not by analysts. We require each observation to be without an earnings per share (EPS) forecast issued by an IBES analyst, but to have an out-of-sample forecast from the proposed model. Each calendar quarter from 1981 to 2010, we sort stocks into quintiles based on earnings growth forecasts. We calculate earnings growth forecast as: $\left(Model\ EPS\ forecast_{t+h} - EPS_t\right)/abs(EPS_t)$, where $Model\ EPS\ forecast_{t+h}$ is the model's out-of-sample EPS forecast, h-years ahead. We denote the event-quarter during which the sorting is done as Qtr 0. We sort the EPS growth forecast based on the most recent median analyst forecast issued during the event-quarter and the most recent out-of-sample model forecast available as of the event-quarter. Stocks in each quintile form an equally weighted portfolio. For each stock in this portfolio, we calculate the characteristics adjusted buy-and-hold quarterly abnormal stock return (adjusted for size, book-to-market equity and prior one-year-return using 5 x 5 x 5 portfolios) for each of the preceding two event-quarters and for each of the subsequent four event-quarters. The characteristics adjusted abnormal stock return is reported in percent. The notation *, **, and *** indicates significance at the 10%, 5% and 1% levels, respectively. Time-series t-statistics are in parentheses. Panels A, B and C report the abnormal returns grouped by 1-year-ahead, 2-years-ahead and 3-years-ahead EPS growth forecasts, respectively.

Growth forecast	**Event-Quarter**						
1yr-ahead (N = 201,996)	Qtr -2	Qtr -1	Qtr 0	Qtr 1	Qtr 2	Qtr 3	Qtr 4
Quintile1 (Low)	1.561***	0.643**	-0.16	-0.528	-0.692	-0.485	-0.428
Quintile2	0.684**	0.531*	0.161	-0.491	-0.64	-0.984	-1.125
Quntile3	0.279	0.382	0.394	0.401	0.43	0.344	0.281
Quintile4	-3.519***	-2.536***	-1.473***	-0.779*	-0.341	0.278	0.443
Quintile5 (High)	-3.938***	-1.946***	-0.049	0.820**	1.178***	1.298***	1.303***
High - Low	**-5.499*** **	**-2.589*** **	**0.111**	**1.348*** **	**1.870*** **	**1.783*** **	**1.730*** **
	(-13.658)	**(-6.769)**	**(0.272)**	**(3.051)**	**(3.910)**	**(3.949)**	**(4.459)**
2yrs-ahead (N = 237,351)							
Quintile1 (Low)	1.468***	0.337	-0.749***	- 1.111***	- 1.081***	-0.711**	-0.652**
Quintile2	0.411	0.18	-0.008	-0.266	-0.295	-0.258	-0.227
Quntile3	0.168	0.096	0.177	0.174	0.271	0.299	0.337
Quintile4	-3.798***	-2.785***	-1.790***	-1.040**	-0.518	0.157	0.515
Quintile5 (High)	-3.843***	-2.242***	-0.756***	0.143	0.710**	0.949***	1.079***
High - Low	**-5.312*** **	**-2.579*** **	**-0.007**	**1.254*** **	**1.791*** **	**1.659*** **	**1.731*** **
	(-15.347)	**(-7.801)**	**(-0.022)**	**(3.325)**	**(4.398)**	**(4.050)**	**(4.857)**
3yrs-ahead (N = 277,384)							
Quintile1 (Low)	1.490***	0.252	-0.961***	- 1.357***	- 1.298***	- 1.113***	- 1.027***
Quintile2	0.226	-0.112	-0.278	-0.525	-0.57	-0.537*	-0.657**
Quntile3	0.338*	0.227	0.201	0.239	0.246	0.202	0.007
Quintile4	-3.801***	-2.643***	-1.632***	-1.070**	-0.495	-0.038	0.255
Quintile5 (High)	-3.347***	-1.874***	-0.684***	0.026	0.491*	0.818***	1.102***
High - Low	**-4.837*** **	**-2.126*** **	**0.276**	**1.382*** **	**1.789*** **	**1.930*** **	**2.128*** **
	(-16.333)	**(-7.376)**	**(1.019)**	**(4.088)**	**(5.238)**	**(5.023)**	**(5.853)**

Table 9: Mutlivariate regressions of future returns on investment signals. We require each observation to have an IBES median analyst earnings per share (EPS) forecast and an out-of-sample EPS forecast from the proposed model. We calculate EPS growth forecast as: $(EPS\ forecast_{t+h} - EPS_t)/abs(EPS_t)$, where $EPS\ forecast_{t+h}$ is either the analysts' or the model's out-of-sample EPS forecast, h-years ahead, quintile-ranked. Investment signals are consistent with those used by Jegadeesh, Kim, Krische, and Lee (2004) are defined in the Appendix. The dependent variable is future returns: market-adjusted return in the six months after the forecast date. t-statistics are reported in parentheses. . ***, ** or * indicate significance at the 1, 5 or 10% levels, respectively.

	Baseline (1)	Analysts (excl. FREV) (2)	Analysts (incl. FREV) (3)	Model (4)	Model & Analysts (5)
MODEL	--	--	--	0.022***	0.023***
	--	--	--	(12.513)	(12.436)
ANALYST	--	0.001	-0.000	--	-0.004**
	--	(0.567)	(-0.017)	--	(-2.222)
SIZE	-0.001	-0.001	-0.001	-0.003	-0.003
	(-0.261)	(-0.300)	(-0.348)	(-1.330)	(-1.421)
RETP	0.037**	0.047***	0.037**	0.027*	0.029*
	(2.278)	(2.793)	(2.238)	(1.694)	(1.798)
RET2P	0.001	0.004	0.000	-0.005	-0.004
	(0.050)	(0.363)	(0.007)	(-0.476)	(-0.364)
FREV	0.012***	--	0.012***	0.011***	0.012***
	(4.342)	--	(4.374)	(4.066)	(4.605)
SUE	0.002**	0.004***	0.002**	0.001	0.002*
	(2.077)	(3.396)	(2.257)	(1.108)	(1.955)
EP	0.024	0.024	0.024	0.054*	0.051*
	(0.822)	(0.828)	(0.846)	(1.908)	(1.851)
BP	0.016	0.015	0.015	0.017*	0.017*
	(1.581)	(1.544)	(1.538)	(1.725)	(1.714)
TURN	-0.002**	-0.003**	-0.002**	-0.002**	-0.002*
	(-2.366)	(-2.461)	(-2.371)	(-1.993)	(-1.963)
SG	-0.017***	-0.016**	-0.016***	-0.010*	-0.008
	(-2.833)	(-2.565)	(-2.802)	(-1.851)	(-1.558)
TA	-0.114***	-0.116***	-0.114***	-0.089***	-0.089***
	(-8.035)	(-8.039)	(-7.961)	(-6.549)	(-6.632)
CAPEX	0.015	0.013	0.019	0.013	0.017
	(0.423)	(0.347)	(0.528)	(0.360)	(0.477)
LOSS	-0.022***	-0.023***	-0.022***	-0.064***	-0.061***
	(-2.977)	(-2.927)	(-2.717)	(-8.263)	(-7.613)
Constant	0.032	0.028	0.034	-0.025	-0.017
	(1.376)	(1.151)	(1.416)	(-1.073)	(-0.742)
N	178618	178618	178618	178618	178618
R-squared	0.10	0.10	0.10	0.11	0.11