Journal of Empirical Finance 68 (2022) 133–159

Contents lists available at ScienceDirect

Journal of Empirical Finance

journal homepage: www.elsevier.com/locate/jempfin

Forecasting earnings with combination of analyst forecasts

Hai Lin [a], Xinyuan Tao [b,*], Chunchi Wu [c]

[a] School of Economics and Finance, Victoria University of Wellington, RH320, Rutherford House, 23 Lambton Quay, New Zealand
[b] Martin Tuchman School of Management, New Jersey Institute of Technology, University Heights CAB 4028, Newark, NJ 70102, United States of America
[c] School of Management, State University of New York at Buffalo, 335A Jacobs Management Center, Buffalo, NY 14260-4000, United States of America

ARTICLE INFO

JEL classification:
G12
G13

Keywords:
Forecast combination
Consensus forecast
Forecast bias and dispersion
Earnings response coefficients
Post-earnings-announcement drift
Profitability factor

ABSTRACT

We propose a regression-based method for combining analyst forecasts to improve forecasting efficiency. This method significantly reduces the bias in earnings forecasts, and generates forecasts that consistently outperform consensus forecasts over time and across firms of different characteristics. Incorporating firm-level and macroeconomic information in the model further improves earnings forecasting performance. Forecasting gains increase with the dispersion and bias of analyst forecasts, and the degree of under/overreactions to earnings news. Moreover, the combination forecast produces larger earnings response coefficients, weakens the anomaly of post-earnings-announcement drift, and provides a better expected profitability measure that has higher power to predict stock returns.

1. Introduction

Financial analysts play an important role in producing and disseminating information to the investment community. An important output from analysts' information production is earnings forecasts. Earnings forecasts convey the perspective of a firm's future cash flows, which are essential for security valuation and are the fundamental basis for analysts to make investment recommendations. Given the special role of analyst forecasts, understanding the properties of these forecasts, and how to best use this set of information is important for academics and practitioners.

Firms covered by financial information intermediaries typically have multiple analysts and so a natural question is how to select or combine individual forecasts in such a way that can provide the optimal forecast for future earnings. In practice, the consensus forecast (averaged across analysts) has been used as a proxy of the market's expectations for future earnings. However, the literature has shown that the consensus forecast inefficiently aggregates information by assigning too much weight to analysts' common information (Kim et al., 2001; Kirk et al., 2014). Also, analysts' forecasts are biased, inefficient, and on average too optimistic (Brown, 1993; Kothari et al., 2016). Therefore, it is important to cope with these problems when combining forecasts to provide an optimal earnings forecast.

It is well known that the time-series process of earnings is complex (see Fama and French, 2000; Bradshaw et al., 2012; Gerakos and Gramacy, 2013). Changes in technology, corporate policy, consumer demand and production costs, and other economic uncertainty can significantly affect firms' profitability. Structural instability resulting from fundamental shocks gives rise to a constantly evolving earnings-generating process that is too complicated for any single forecaster to approximate. Analyst forecasts also exhibit significant cross-sectional variations due to analyst characteristics (O'Brien, 1990; Stickel, 1990; Sinha et al., 1997; Mikhail et al., 1997; Clement, 1999; Gerakos and Gramacy, 2013). Under these conditions, combining individual analyst forecasts

* Corresponding author.
E-mail addresses: hai.lin@vuw.ac.nz (H. Lin), xinyuan.tao@njit.edu (X. Tao), chunchiw@buffalo.edu (C. Wu).

https://doi.org/10.1016/j.jempfin.2022.07.003
Received 26 July 2020; Received in revised form 18 April 2022; Accepted 3 July 2022
Available online 9 July 2022
0927-5398/

offers a solution for reducing the instability risk associated with reliance on a subset of forecasters and enhancing the information content of earnings forecasts (see Timmermann, 2006; Rapach et al., 2010; Lin et al., 2018).

This paper proposes to improve earnings forecasts using advanced methods that build on a sound statistical foundation and have proven success in economic and financial forecasts. Capistrán and Timmermann (2009) show that individual forecasts are inherently biased, and averages of biased forecasts, such as the conventional consensus forecast used by the finance profession, will remain biased. To resolve this problem, they propose a regression-based method to adjust the bias in the mean forecast. They demonstrate this method is very powerful for correcting the bias in survey forecasts. Built on this finding, Lin et al. (2018) propose an iterated regression method to further improve the performance of the single-step mean-adjusted regression method. This method features a two-step procedure that first corrects the bias and inefficiency of an individual forecaster, and then runs a second-step regression similar to Capistrán and Timmermann (2009) to obtain an optimal forecast. These studies show that the regression-based combination forecast methods can substantially improve out-of-sample forecasts of security returns and macroeconomic performance. While these methods show considerable promise, it remains unclear whether they can significantly improve the forecast for future earnings, which exhibit a distinct data-generating process. This paper evaluates the efficacy of these methods in enhancing earnings forecasts and mitigating the analyst forecast aggregation problem.

Using a comprehensive data sample, we find that the regression-based forecast combination consistently outperforms the consensus forecast and other methods in forecasting firms' future earnings. The bias-adjusted mean (BAM) method of Capistrán and Timmermann (2009) corrects a substantial amount of bias in consensus analyst forecasts, and the iterated regressed mean (IMC) and weighted-average (IWC) combination methods (Lin et al., 2018) further improve it whenever sufficient historical data are available to refine earnings forecasting. The regression-based combination methods outperform other earnings forecast methods by substantial margins over time and across firms of different characteristics. The benefit of regressed combination is greater when analyst forecasts are disperse, overly optimistic, or severely over or under reacting to earnings-related information.

Additional tests show the robustness of the regressed combination approach to different forecast specifications. Jame et al. (2016) and Ball and Ghysels (2018) find that including other publicly available earnings-relevant information at the firm level and market/macroeconomic information can improve earnings forecasts. In light of the literature, we incorporate these variables into our forecast model and find that combining these variables with analyst forecasts further improves out-of-sample forecasts. Using characteristic weights via Bradley et al. (2017) to account for heterogeneity in analyst experience,[1] we find that adding this weighting scheme in the model helps improve the performance of combination forecasts. These findings suggest that the regression-based combination forecast approach is quite versatile and can be easily adapted to different forecasting information and decision rules to refine the earnings forecast. Furthermore, we compare the performance of the regression-based forecast with other novel forecasting methods such as principal component analysis (PCA) and machine learning methods, Lasso and Elastic-Net. We find that the regression-based forecast performs favorably relative to these alternative methods. The results suggest that in the presence of the unbalanced panel data in analyst surveys and low-frequency earnings forecasts, the regression-based methods have an edge over machine learnings methods and PCA.

Our finding that the regressed combination method produces earnings forecasts superior to consensus analyst forecasts has important capital market implications. We use our earnings forecast as a proxy for the market's expectations of earnings to estimate earnings response coefficients (ERCs) and to re-examine the issue of post-earnings-announcement drift (PEAD). The earnings response coefficient is a more direct way of evaluating how closely the model-based and consensus analyst forecasts line up with market expectations (see Hou et al., 2012). We find that using the regressed combination earnings forecast generates much larger and more significant ERCs than using the consensus forecast.[2] This finding strongly suggests that the regressed combination earnings forecast is a better proxy for expected earnings. Using our earnings forecasts as a measure for the latent earnings expectations, we find that the anomaly of post-earnings-announcement drift is substantially weakened. More importantly, the expected ROE generated from the regressed combination forecasts significantly improves the predictive power of the profitability factor in the cross-section for future stock returns. The literature suggests that profitability is an important pricing factor for stocks, and the realized ROE is commonly used as a proxy for a firm's future profitability (see Hou et al., 2015). An issue in asset pricing tests with a profitability factor is that the accounting-based ROE is released at a low frequency and contains stale information between announcement dates. Using the ROE forecast by our method generates a much higher stock return predictability than the low-frequency realized accounting ROE. Overall, our finding suggests that the choice of an earnings expectation measure has a significant impact on inferences about the post-earnings-announcement drifts and the cross-section of expected stock returns. Our combination forecast method extracts better expected earnings information from analyst forecasts and using the ROE forecasts generated from our model substantially improves the predictive power of the factor model in the cross-section for future stock returns.

The remainder of this paper is organized as follows. Section 2 discusses different approaches for combining or selecting analyst forecasts. Section 3 presents the results of Monte Carlo simulations for evaluating the performance of selected combination methods. Section 4 discusses the data and Section 5 reports empirical results. Section 6 examines the implications of efficient earnings forecasts for asset pricing. Finally, Section 7 summarizes our main results and concludes the paper.

[1] This method gives more (less) weight to the analysts whose forecasts have higher (lower) accuracy.

[2] Consistent with the error-in-variable (EIV) theory, more precise measures of expected and unexpected earnings lead to significantly larger earnings response coefficients.

2. Forecast combination methods

Numerous factors affect corporate profitability, and changes in these factors can result in a complex earnings generating process. In such case, while reliance on a "superior" analyst may sometimes yield reasonable earnings forecasts, it is unlikely to generate reliable forecasts consistently over time. Combining individual analyst forecasts can thus reduce instability risk of earnings forecasts. In this section, we discuss prototype forecast combination methods and their usage.

Let $\hat{Y}^i_{j,t+1}$ be the forecast conditional on the information at time t by analyst i for the earnings of firm j at $t+1$, $\overline{Y}_{j,t+1}$ be the combined analyst forecast from individual forecasts, and $Y_{j,t+1}$ be the realized/actual earnings. Earnings are measured as earnings per share scaled by book equity per share (ROE).[3] While we focus on the one-quarter-ahead earnings forecast, the forecast horizon can be longer or shorter. Forecast combination aims to find the optimal weights among analysts to generate an efficiently combined forecast for future earnings $\overline{Y}_{j,t+1} = \sum_{i=1}^{N_{j,t}} \omega^i_{j,t} \times \hat{Y}^i_{j,t+1}$ such that the mean square errors of forecasts are minimized.[4] $\omega^i_{j,t}$ is the weight assigned to analyst i who follows firm j at time t, the sum of $\omega^i_{j,t}$ equals one, and $N_{j,t}$ is the number of analysts following firm j at time t.

Equal-Weighted Average (EW)

The most common way to combine analyst forecasts is to take the simple average forecast where each individual forecast is assigned an equal weight, $\overline{Y}_{j,t+1} = \frac{1}{N_{j,t}} \sum_{i=1}^{N_{j,t}} \hat{Y}^i_{j,t+1}$. Equal-weighted mean earnings forecast is dubbed the consensus forecast by professionals. The main advantage of average forecast lies in its ability to reduce noise and increase stability of forecasts. Ashton and Ashton (1985) find that simple averages of forecasts improve accuracy and reduce variability of accuracy. Conroy and Harris (1987) argue that averaging forecasts has a "portfolio effect" when errors are less than perfectly correlated, thereby producing smaller forecast errors than any of the constituent forecasts. Statistically, when there is a common factor in the forecasts, the first principal component can be approximated by mean combination ($\omega^i_{j,t} = 1/N_{j,t}$). This explains why the simple average method works quite well and is difficult to beat (Timmermann, 2006).

The equal-weighted average is optimal in terms of mean-square-error minimization when individual forecasts are unbiased and their forecast error variances are the same and have identical pairwise correlations (or homogeneous). Nevertheless, it will generally be suboptimal when these assumptions are violated. The literature has suggested that analyst forecasts are heterogeneous and biased (see Kirk et al., 2014, and the references therein). In this situation, the mean combination produces biased forecasts.[5] Also, averaging forecasts may wash out valuable private information in individual analyst forecasts.

Previous Best Forecast (PBest)

Rather than assigning a weight to each analyst, the previous best forecast (PBest) method simply selects the forecast by the analyst with the best track record to predict future earnings. This implies that the weight for the "star" analyst is one, and zero for other analysts. Analytically, the past forecast performance is measured by an analyst's historical mean square error (MSE), $\frac{1}{t-1} \sum_{\tau=1}^{t-1} \left(e^i_{j,\tau} \right)^2$, where $e^i_{j,\tau} = Y_{j,\tau+1} - \hat{Y}^i_{j,\tau+1}$. The PBest chooses the forecast by the analyst with the smallest MSE, $\hat{Y}^{PBest}_{j,t+1}$, as the optimal forecast $\overline{Y}_{j,t+1}$ for future earnings.

A drawback of the previous best forecast is that it relies too heavily on a single analyst. Empirical evidence shows that no analysts can be consistently better than others (O'Brien, 1990) and it is highly unlikely that an analyst will always have superior information. To the extent that individual forecasts are biased, unstable and incomplete, this method generates inconsistent and risky/unstable forecasts over time.

Inverse MSE (IMSE)

Unlike the equal-weight method, the inverse MSE method, IMSE, uses the weighted average across analysts as the optimal forecast where the weight is inversely proportional to each analyst's historical MSE, i.e., $\omega^i_{j,t} = 1/[\frac{1}{t-1} \sum_{\tau=1}^{t-1} \left(e^i_{j,\tau} \right)^2]$ and normalized so that the sum of weights is equal to one. The intuition is that by giving a low weight to the analysts with poor past performance, one can improve the efficiency of forecast combination. The IMSE method yields a combined forecast $\overline{Y}_{j,t+1} = \sum_{i=1}^{N_{j,t}} (\omega^i_{j,t} \hat{Y}^i_{j,t+1})$ where the weight depends on past forecast errors.

Odds Matrix Approach (Odds)

Similar to IMSE, the Odds matrix method uses a weighted average as the combined forecast. Unlike IMSE, the weight of the Odds method is based on the odd ratio. More specifically, weights are derived from a matrix of pairwise odds ratios, $\boldsymbol{O}$, which contains pairwise probabilities $o_{m,n} = P_{m,n}/P_{n,m}$ where $P_{m,n} = \frac{K_{m,n}}{(K_{m,n}+K_{n,m})}$, $K_{m,n}$ is the number of times analyst m has a smaller absolute forecast error than analyst n in the past, and $K_{n,m}$ is the opposite. The weight vector, $\boldsymbol{\omega}$, is computed from the solution $\left(\boldsymbol{O} - N_{j,t} \boldsymbol{I} \right) \boldsymbol{\omega} = \boldsymbol{0}$ where $\boldsymbol{I}$ is the identical matrix, and the normalized eigenvector associated with the largest eigenvalue of $\boldsymbol{O}$ is used. In implementation, one first computes the weight vector, $\boldsymbol{\omega}$, for each firm j at time t and then assigns the weights to individual analyst forecasts to generate a combined forecast of $\overline{Y}_{j,t+1} = \sum_{i=1}^{N_{j,t}} (\omega^i_{j,t} \hat{Y}^i_{j,t+1})$.

[3] The methods discussed in this section can be applied to any other scaled or unscaled earnings figures.

[4] Mean squared errors are measured as $\frac{1}{T} \sum_{t=1}^{T} (Y_{j,t+1} - \overline{Y}_{j,t+1})^2$ where $(Y_{j,t+1} - \overline{Y}_{j,t+1})^2$ stands for one-quarter ahead squared errors of out-of-sample forecast for firm j at t, and T is the total number of firm-quarter out-of-sample forecasts.

[5] Kim et al. (2001) show that in the aggregation process, the mean analyst estimate overweights analysts' public information and underweights their private information.

None of the above methods uses a regression technique to obtain an optimal forecast. A common problem with these methods is that if analyst forecasts are biased, combining individual forecasts using these methods will continue to be biased. The regression-based combination is designed to tackle this problem. In the following, we present three regression-based combination methods, each including at least one round of regression adjustment to correct the bias in analyst forecasts.

Bias-Adjusted Mean (BAM)

Capistrán and Timmermann (2009) suggest an intuitive way to correct the bias in the equal-weighted forecast by running a simple regression to adjust the mean of individual forecasts:

$$Y_{j,t} = \alpha + \beta \overline{Y}_{j,t} + \varepsilon_{j,t}. \tag{1}$$

Capistrán and Timmermann (2009) indicate that at any given point in time, there always exists a scaling factor such that the product of β and $\overline{Y}_{j,t}$ is unbiased, and to further ensure that the combined forecast is on average unbiased, an intercept α is included.

To implement the bias-adjusted mean method, BAM, we first obtain an equal-weighted average forecast $\overline{Y}_{j,\tau} = \frac{1}{N_{j,\tau-1}}\left(\hat{Y}^1_{j,\tau} + \hat{Y}^2_{j,\tau} + \cdots + \hat{Y}^{N_{j,\tau-1}}_{j,\tau}\right)$ in each period (e.g., quarter) τ from 1 to t and run the regression of actual earnings $Y_{j,\tau}$ on the equal-weighted average forecast $\overline{Y}_{j,\tau}$ to estimate $\hat{\alpha}$ and $\hat{\beta}$. The estimated parameters at t are then used to generate an adjusted forecast $\hat{\alpha} + \hat{\beta}\overline{Y}_{j,t+1}$ to predict the firm's earnings at $t+1$ where the mean forecast $\overline{Y}_{j,t+1}$ is obtained before the firm's earnings announcement. This procedure is recursive, that is, in the next period, we use all historical data from $\tau = 1$ to $t+1$ to obtain the equal-weighted average forecast for each quarter and parameter estimates $\hat{\alpha}$ and $\hat{\beta}$ to predict earnings at $t+2$. This procedure continues until we reach the end of the forecast period $T-1$ where T is the sample size.

Like the consensus forecast, the BAM exploits the cross-sectional information provided by individual analysts by averaging individual forecasts. However, the mean combination (EW) alone cannot correct the bias in individual analyst forecasts. To correct this problem, the BAM method runs a regression to adjust the bias in the mean forecast. As the regression approach uses the full set of individual forecasts, both in the cross section and time series, it does not entail a loss of information relative to the consensus forecast. In the special case that $\alpha = 0$ and $\beta = 1$, the BAM degenerates to the consensus forecast. Thus, the BAM method is more general and theoretically will encompass the consensus forecast (EW) method in out-of-sample forecasts.[6]

An advantage of the BAM method is that it involves only two parameters α and β and therefore, is highly parsimonious and less affected by estimation errors. The BAM is also robust to the unbalanced panel data problem associated with the uneven records of forecasts due to entry and exits of analysts. Moreover, the BAM does not rely on the covariance–variance matrix estimated from historical data to perform forecast combination.

Iterated Mean Combination (IMC)

The iterated mean combination (IMC) performs two steps of regressions (Lin et al., 2018). The second-step regression is the same as that used in the BAM method but in the first step, there is an additional regression adjustment on each individual forecast before constructing the equal-weighted average $\overline{Y}_{j,t}$. The purpose of the first-step regression is to correct the problems in an individual forecast, such as bias, inefficiency and uncertainty of forecasts associated with analyst characteristics and incentives, before aggregating individual forecasts. In principle, the first-step adjustment accounts for the effects of analysts' characteristics and learning experience from the past.

To implement this method, for each individual analyst *i*, we first run a time-series regression of actual earnings on the forecast of analyst i for firm j, $Y_{j,\tau} = a_i + b_i\hat{Y}^i_{j,\tau} + \varepsilon^i_{j,\tau}$, in the estimation period from $\tau = 1$ to t to obtain the parameters $\hat{a}_i$ and $\hat{b}_i$, which are used to generate an adjusted forecast $\tilde{Y}^i_{j,\tau} = \hat{a}_i + \hat{b}_i\hat{Y}^i_{j,\tau}$ at the individual analyst level. The equal-weighted average $\overline{Y}_{j,\tau}$ is then obtained using these adjusted individual forecasts $\tilde{Y}^i_{j,\tau}$ instead of the observed analyst forecast $\hat{Y}^i_{j,\tau}$ as in BAM; that is, $\overline{Y}_{j,\tau} = \frac{1}{N_{j,\tau-1}}\left(\tilde{Y}^1_{j,\tau} + \tilde{Y}^2_{j,\tau} + \cdots + \tilde{Y}^{N_{j,\tau-1}}_{j,\tau}\right)$. The next step is to run a time-series regression of actual earnings against the mean of adjusted-earnings forecasts, $Y_{j,\tau} = \alpha + \beta\overline{Y}_{j,\tau} + \varepsilon_{j,\tau}$ to perform another round of adjustment to ensure that unconditionally the combined forecast is unbiased at any given point in time. Similar to the BAM, the IMC forecast procedure is recursive.

In the special case that $a_i = 0$ and $b_i = 1$ the IMC degenerates to the BAM. More generally, when $a_i \neq 0$ and/or $b_i \neq 1$, the IMC tends to be more efficient as it considers historical information that influences an analyst's forecast. To see this merit, we can rewrite the second-step regression of IMC as

$$Y_{j,t+1} = \hat{\alpha} + \hat{\beta}\sum_{i=1}^{N_{j,t}}\left(\hat{a}_i + \hat{b}_i\hat{Y}^i_{j,t+1}\right)/N_{j,t} = \hat{\alpha} + \hat{\beta}\overline{a} + \frac{\hat{\beta}}{N_{j,t}}\sum_{i=1}^{N_{j,t}}\hat{b}_i\hat{Y}^i_{j,t+1}. \tag{2}$$

This expression reveals that the IMC takes into account the characteristics and properties of analyst forecasts in aggregating individual analyst forecast for each firm. As indicated, $\hat{b}_i\hat{Y}^i_{j,t+1}$ adjusts for the potential bias arising from the differences in individual analysts' experience, style, characteristics and incentives. The IMC tends to perform better when analysts are more heterogeneous.

Iterated Weighted-Mean Combination (IWC)

The iterated weighted-average combination method, IWC, is similar to the IMC except that it assigns a weight to each forecast when combining the adjusted analyst forecasts. Rather than taking the simple (equal-weighted) average of adjusted analyst forecasts $\tilde{Y}^i_{j,\tau}$ from the first step regression, the IWC calculates the weighted average using the method suggested by Bates and Granger (1969).

[6] There could be sample-dependent estimation errors and the BAM may not always outperform the EW empirically.

The weight depends on the inverse of residual variances ($\hat{\sigma}_i^2$) from the first-step regression for each analyst *i*. More specifically, $\overline{Y}_{j,\tau} = \frac{1/\widehat{\sigma_1}^2}{\sum_{i=1}^{N_{j,\tau-1}} 1/\hat{\sigma}_i^2}\tilde{Y}_{j,\tau}^1 + \frac{1/\widehat{\sigma_2}^2}{\sum_{i=1}^{N_{j,\tau-1}} 1/\hat{\sigma}_i^2}\tilde{Y}_{j,\tau}^2 + \cdots + \frac{1/\widehat{\sigma_{N_{j,\tau-1}}}^2}{\sum_{i=1}^{N_{j,\tau-1}} 1/\hat{\sigma}_i^2}\tilde{Y}_{j,\tau}^{N_{j,\tau-1}}$. If an analyst has less accurate forecasts in the past, her current forecast will be down weighted to account for the past poor performance. After obtaining the weighted average $\overline{Y}_{j,\tau}$, we run the regression $Y_{j,\tau} = \alpha + \beta\overline{Y}_{j,\tau} + \varepsilon_{j,\tau}$ as is in the BAM and IMC methods. This procedure can be refined using more a sophisticated Bayesian weighting scheme as suggested by Pettenuzzo et al. (2014).

In summary, a variety of methods can be used for combining individual forecasts. These methods have different properties and their performance depends on the attributes of data and forecasters. To assess the performance of these methods under various conditions, we employ Monte Carlo simulations in a control experiment setting.

3. Monte Carlo simulations

We conduct Monte Carlo simulations using a dynamic factor model that permits biases in individual forecasts and heterogeneity in individual forecasters' characteristics and skills (see the Online Appendix for detail). In simulations, we allow for cross-sectional dispersion and time variations in analyst performance to account for the stylized fact of analyst forecasts.

There are several interesting results from Monte Carlo simulations. First, all regression-based methods generate better earnings forecasts as they are effective in correcting the bias in individual forecasts. Second, the performance of single- and two-step regression methods is similar when analysts are relatively homogeneous and variations (both temporal and cross-sectional) in individual forecasts are low. The two-step regression methods (IMC and IWC) perform better than the BAM in the presence of high forecast dispersion and severe under or over reactions to earnings news by analysts. Third, given the number of analysts (N), a larger sample size (T) generates more reliable forecasts. The regression-based methods benefit more from a larger sample size (T) that improves the precision of regression estimates.

In Appendix A, we show that theoretically the two-step regression method will never perform worse than the single-step regression method, given an equal number of analysts and the same distribution of analyst forecasts, and the benefit of using these methods increases when earnings forecasts are more disperse and analysts underreact more to fundamental (earnings) news. Nevertheless, in applications, there is no guarantee that the IMC (IWC) will always outperform the BAM as there are estimation errors, which are sample dependent.

Collectively, Monte Carlo simulations suggest that the regression-based forecast combination outperforms the consensus forecast (EW) and other methods in the presence of bias and heterogeneity in analyst forecasts. The two-step regression methods (IMC and IWC) have an edge over the single-step regression adjustment method (BAM) when time variations and cross-sectional dispersion in forecasts are high. Under the normal condition that analyst forecasts are temporally stable and no excessive dispersion, the performance BAM is close to the iterated regression method.

A caveat in empirical analysis is that iterated regression combination requires a sufficient length of forecast records for each analyst in the first-step regression to achieve efficiency. This data requirement may limit the sample to the well-established firms with more consistent analyst following. By contrast, the single-step regression method does not have such limitation, which allows researchers to work on a larger sample with higher statistical power. In empirical investigation, we perform tests for both single- and two-step regression methods but focus on the former to achieve a larger sample size of firms with more power.

4. Data

Our data come from three major sources: Thomson Reuters' Institutional Brokers' Estimate System (I/B/E/S), Compustat and CRSP (Center for Research in Security Prices). Quarterly earnings per share (in dollars) and analyst forecasts are collected from I/B/E/S, annual firms' accounting data are from Compustat, and monthly stock prices and number of outstanding shares are from CRSP. The sample covers the period from September 1982 to June 2016.

We collect quarterly one-period-ahead analyst forecasts of EPS from I/B/E/S. We first drop the records with missing actual EPS and missing announcement date of quarterly earnings. For individual analyst forecasts, we only keep the most recent one for each firm at a given quarter. To ensure reliability in empirical estimation, we require a minimum of ten quarterly earnings observations for each firm. A firm is included in the sample as long as it has at least one analyst following. As raw earnings data are affected by the size of firms, we scale EPS by the book value of equity per share at the previous quarter (ROE), stock price and EPS volatility (standard deviation).

Our base sample includes 237,837 out-of-sample forecasts of quarterly ROEs by 8185 firms from June 1985 to June 2016. Our baseline analysis focuses on the performance of the bias adjustment method (BAM) relative to other forecast combination methods. We also report results for two-step regression adjustment methods, which have more stringent data requirement.[7] In extended analysis, we compare the performance of the BAM forecast with the principal component analysis (PCA) and other novel forecast methods such as Lasso and Elastic-Net.

Predictability of earnings and precision of analyst forecasts vary by characteristic (see, for example, Fama and French, 2006; Hou et al., 2012; So, 2013; Li and Mohanram, 2014; Call et al., 2016). We examine the effects of characteristics on the efficiency of analyst forecast combination. We consider four major categories of characteristics as control variables. The first set of variables includes characteristics of analyst forecasts and earnings. Past studies have shown that characteristics of analysts and firms' earnings affect

[7] See Section 5.6 for more detailed explanation.

forecast accuracy (e.g., Clement, 1999; Kim et al., 2011). We compare the performance of forecast combination methods across characteristics such as the number of analysts, dispersion of forecasts, the number of forecast revisions, the magnitude of most recent forecast revision, volatility of EPS, variations in forecast accuracy, and surprise in announced earnings.

The second set of control variables includes firm characteristics. Bradshaw et al. (2012) re-examine the issue of whether analysts provide more accurate forecasts than naïve time-series models, given that the former seems to have advantages in information and timing. They find that analysts' advantage diminishes for firms with small size and a low number of analysts. These firms tend to have high information asymmetry and are more opaque. In addition, when information is more complex and more difficult to interpret, errors of analyst forecasts increase. We examine the performance of forecast combination methods for firms with different capitalization and information complexity. Following the literature, we consider firm characteristics such as size, book-to-market ratio, age, and accruals quality as control variables.

The third set of control variables includes sector-related characteristics. Firms in the same industry tend to have similar operating and financial characteristics. To capture the sector-related characteristics, we divide firms into ten groups based on their first 2-digit standard industry classification (SIC) code. These include: (1) Agriculture, forestry and fishing, (2) Mining, (3) Construction, (4) Manufacturing, (5) Transportation and public utilities, (6) Wholesale trade, (7) Retail trade, (8) Finance, insurance and real estate, (9) Service, and (10) Non-classifiable.

Lastly, we control for the effects of market environment, economic regime and sentiment. Earnings forecasts are forward evaluation of a firm's future performance. When the market is more uncertain, the subsequent cash flow of a firm will be more difficult to forecast. Forecast errors tend to increase during the recession when there is high uncertainty about future economic activities and recovery. On the other hand, analysts tend to be more cautious during the bad time, which could reduce earnings overestimation. Therefore, the forecast errors can go either way when market uncertainty is high.

Appendix B provides a description for the variables used in empirical tests and their data sources. In empirical investigation, we perform subsample analysis to examine the performance of different forecast methods for firms with different characteristics and across industries and economic regimes.

5. Empirical results

In this section, we first provide a diagnosis for the problem in forecasting unscaled earnings and show that it is important to apply the forecast models to scaled earnings in order to have a consistent comparison for the performance of these models. We report the diagnostic results of forecasts based on three scaled earnings measures: EPS divided by EPS-volatility, price per share (EP) and book equity (ROE). Based on the performance of these scaled earnings forecasts, we choose ROE as our primary scaled earnings measure to compare the performance of different forecast combination methods. We find that the single-step regression adjustment method (BAM) outperforms the consensus forecast (EW) and other conventional methods by substantial margins. Following this, we show iterated regression methods (IMC and IWC) can further improve the earnings forecast when there are sufficient historical data to implement these methods efficiently.

5.1. Out-of-sample combination forecasts

5.1.1. Diagnosis of the scale effect in forecasting earnings

Prior research on analyst earnings forecasts typically uses unscaled earnings measures to evaluate the performance of earnings forecast models (see, for example, Ball and Ghysels, 2018). However, a potential problem of using firm-level quarterly accounting earnings or earnings per share (EPS) is that these data are subject to the scale effect, which may result in inconsistent statistical inference. For example, firms with stock price of $100 per share are likely to have higher EPS and earnings volatility than firms with stock price of $10 per share. Even for the same firm, if it splits one share into two, the EPS will mechanically falls by half. The cross-sectional and time-series non-stationarity in EPS can distort empirical tests. As an example, researchers typically use mean-square-errors to evaluate the performance of an earnings forecast model. Mean-square-errors are likely larger for firms with larger share price or capitalization. If we do not adjust for the scale effect of earnings, it may lead to a biased inference that a forecast model performs poorly for a firm with high stock price or for an industry dominated by firms with high share prices.

To illustrate the problem, we first use our models to forecast unscaled earnings per share (see the Online Appendix for the details of these results). We then sort firm-quarter observations evenly into 20 groups by the magnitude of consensus analyst forecasts (EW) and report the mean-square-errors (MSEs) of EW forecasts for each group.[8] Fig. 1 shows that the distribution of MSEs by EW forecasts is highly right-skewed. The result informs a mechanically generated cross-sectional variation in MSE due to the unscaled EPS effect and suggests a need to adjust this effect in empirical tests.

We next sort all firms into 10 groups on the MSE of EW forecasts where group 1 includes firms with the lowest MSE of EW and group 10 with the highest MSE. We then compare the performance of regression-based combination (BAM) with EW by computing the mean MSE values of EW and BAM forecasts for each decile group and report the differences and the associated t values. Table C.1 in Appendix C reports these results. As shown, MSEs of both EW and BAM exhibit a monotonic increase over the deciles and the differences between them exhibit an uptrend. The results continue to show a mechanical cross-sectional variation in earnings forecast errors which can distort the inference for firms with different characteristics (e.g., earnings volatility). Again, this finding suggests a need to scale EPS in order to provide a consistent comparison for the performance of forecast models.

[8] We are very grateful to an anonymous referee for this insightful suggestion.

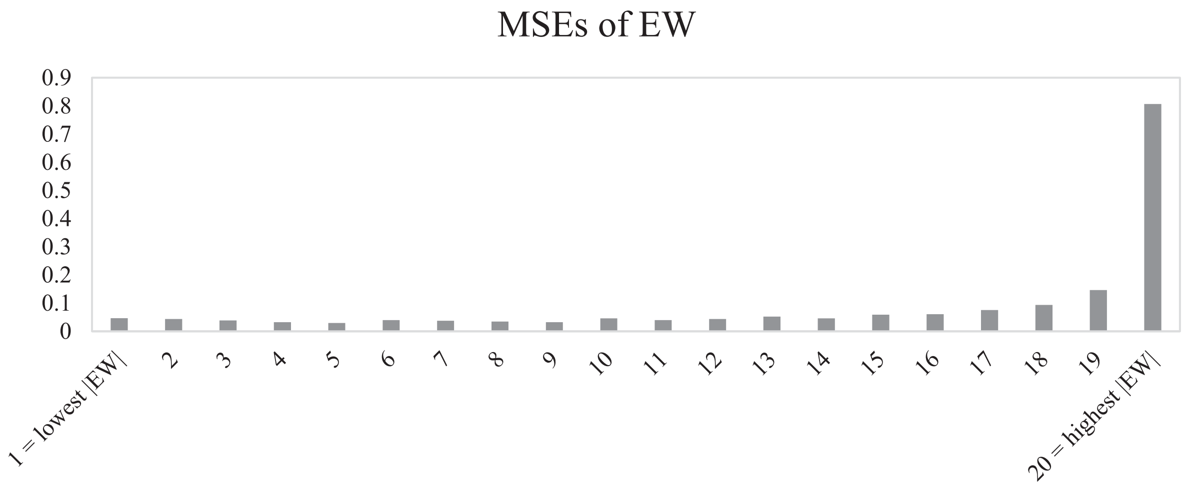

Fig. 1. Distribution of MSEs against EW. We sort firm-quarter observations evenly into 20 groups by the magnitude of consensus analyst forecasts (EW) and report the mean-square-errors (MSEs) of EW forecasts for each group.

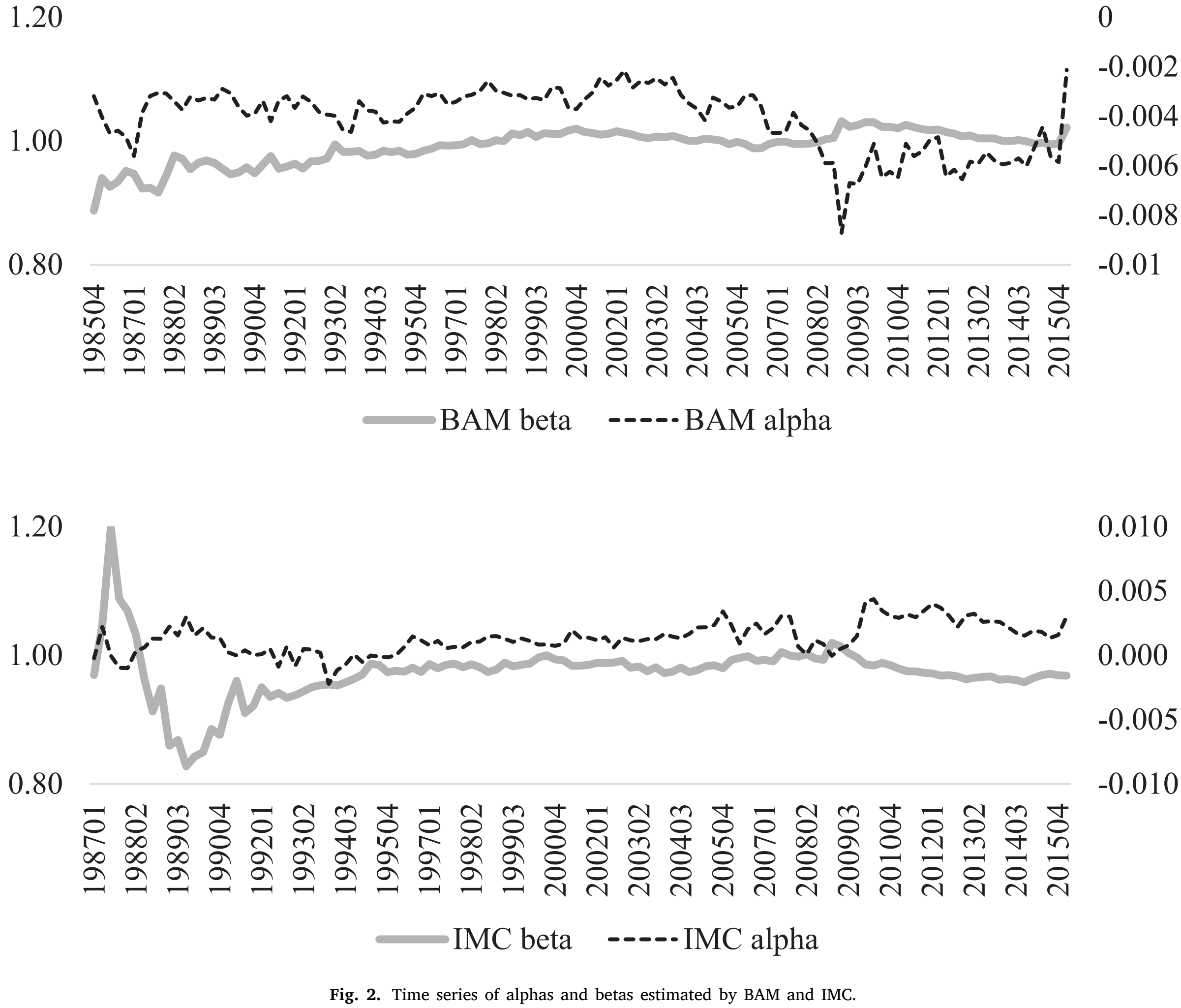

Fig. 2. Time series of alphas and betas estimated by BAM and IMC.

Table 1
Out-of-sample performance of different combination forecast models.

Panel A: The full sample

	Forecast models					Difference between other models and EW			
	PBest	EW	IMSE	Odds	BAM	PBest	IMSE	Odds	BAM
MSE	0.054	0.030	0.040	0.040	0.026	0.024	0.010	0.010	−0.004
t-value						(3.34)	(1.72)	(1.79)	(−3.76)
R^2_{OS} (%)						−78.42%	−32.68%	−32.54%	14.48%
t-value						(3.03)	(3.27)	(3.28)	(4.00)
DM test						(6.31)	(3.27)	(3.26)	(−2.50)
Bootstrap						1%	1%	1%	10%

Panel B: The subsample with iterated regression combination

	Forecast models							Difference between other models and EW					
	PBest	EW	IMSE	Odds	BAM	IMC	IWC	PBest	IMSE	Odds	BAM	IMC	IWC
MSE	0.0069	0.0034	0.0058	0.0062	0.0032	0.0030	0.0029	0.0035	0.0024	0.0028	−0.0002	−0.0004	−0.0005
t-value								(2.19)	(1.81)	(1.14)	(−1.70)	(−2.04)	(−1.93)
R^2_{OS} (%)								−102.18%	−71.30%	−82.38%	6.87%	11.03%	13.49%
t-value								(1.97)	(0.80)	(0.75)	(1.77)	(1.83)	(1.67)
DM test								(2.44)	(1.34)	(1.53)	(−1.12)	(−2.68)	(−1.59)
Bootstrap								10%	–	–	10%	1%	1%

This table reports the average value of mean-square-errors (MSEs) of ROE forecasts for each forecast combination model and their differences. For each forecast, we require at least eight historical quarterly ROE records and one analyst following the firm where ROE is computed as EPS scaled by book value of equity per share. Difference between other forecasting models and EW is computed as other models minus EW and a more negative difference means a greater decrease in forecast errors. The out-of-sample R squares statistics (R^2_{OS}) are computed to evaluate the out-of-sample performance of four models relative to equally weighted model (EW). The higher and positive R^2_{OS}, the better out-of-sample forecast performance than the benchmark EW. The t-values that test the significance of mean MSE difference and R^2_{OS} are reported in parentheses and adjusted for firm-clustered standard errors. We also report the DM test. The bootstrap value indicates the significance level of DM tests compared to bootstrap cutoffs where 1%, 5% and 10% denote the significance level and "−" indicates insignificance. Panel A reports the results of the full sample which includes 237,837 out-of-sample forecasts of quarterly ROE by 8185 firms from June 1985 to June 2016. In Panel B, for each forecast, we require at least 20 historical quarterly ROE records and six analysts following the firm to generate IMC and IWC forecasts. As a result, the sample size decreases to 21,696 out-of-sample forecasts of quarterly EPS by 1106 firms.

To adjust the scale effect of earnings per share, we use three scaled EPS measures: earnings per share divided by EPS volatility (standardized EPS), price per share (EP), and book equity (ROE). We then compare the forecast performance of the EW and BAM models using these scaled earnings measures. Table C.2 in Appendix C reports these results. The results show that BAM consistently outperforms EW in earnings forecasts across all scaled earnings measures. Among the three scaled EPS measures, ROE allows us to better differentiate the performance between the two forecast methods in terms of out-of-sample R-squared (R^2_{os}). We hence choose the ROE measure as the primary measure of scaled earnings in the remainder of this paper to compare the predictive performance of different forecasting models. Using the ROE measure has an additional advantage of enabling us to explore the role of the expected ROE in the cross-section of expected stock returns in the spirit of Hou et al. (2015).

5.1.2. The ROE forecasts

We begin our analysis by comparing the ability of different models to forecast the firm-level ROE. Panel A of Table 1 reports average MSEs of forecasts by each combination method and their magnitude relative to the equal-weighted average (EW) or consensus forecast which is used as the benchmark for performance evaluation. MSEs are the mean squared differences between the actual and predicted ROEs. Using the consensus forecast as a benchmark for forecasting performance evaluation is rightful as it is commonly used to summarize analyst earnings forecasts in practice.

The top of Panel A reports average MSEs of out-of-sample forecasts for all methods. For ease of comparison, we calculate the differences in MSEs between the consensus and other forecasts. The results show that the BAM outperforms other methods in out-of-sample earnings forecasts. The MSE of BAM is lower than that of consensus forecast (EW) by 0.004 or about a 14% reduction (see row 1 on the right side). Other forecast methods do not perform better than consensus forecast. The MSEs of odds ratio (Odds) and inverse MSE (IMSE) methods are somewhat larger than that of the consensus forecasts (EW) while the previous best forecast (PBest) has the worst MSE, which is 0.024 larger than that of consensus forecast.

To give a formal evaluation on the forecast performance, we calculate the following out-of-sample R^2 statistic:

$$R^2_{OS} = 1 - \frac{\sum_{t=t_m}^{T-k}\left(Y_{t+k} - \hat{Y}_{t+k}\right)^2}{\sum_{t=t_m}^{T-k}\left(Y_{t+k} - \overline{Y}_{t+k}\right)^2} \tag{3}$$

where Y_{t+k} is actual earnings at $t+k$, $\overline{Y}_{t+k}$ is the consensus forecast (EW), $\hat{Y}_{t+k}$ is the earnings forecast by any other methods, T is the sample size, t_m indicates the time that the forecast is made and k is the number of periods ahead in the forecast (for simplicity, we drop the subscript for firms). As we focus on the one-period-ahead forecast, we set k equal to one in the quarterly forecast. R^2_{OS} measures the improvement in mean squared prediction errors (MSPE) of a forecast combination method over the consensus forecast (benchmark). When $R^2_{OS} > 0$, the former outperforms the latter, and the larger R^2_{OS}, the greater is the improvement over the consensus forecast. As shown in past studies, it is difficult to outperform the consensus forecast (Timmermann, 2006).

We test the significance of R^2_{OS} using the MSPE-adjusted statistic of Clark and West (2007). This is a one-sided test of the null hypothesis that expected square prediction errors of the EW method and the other method are equal, against the alternative that the latter has lower squared prediction errors than the former. To obtain the MSPE-adjusted statistic, we first compute the following squared error difference:

$$g_{t+k} = \left(Y_{t+k} - \overline{Y}_{t+k}\right)^2 - \left[\left(Y_{t+k} - \hat{Y}_{t+k}\right)^2 - \left(\overline{Y}_{t+k} - \hat{Y}_{t+k}\right)^2\right] \tag{4}$$

Regressing g_{t+k} on a constant, we have the t-statistic that gives a p-value for the one-sided (upper tail) test under the standard normal distribution.

The third row on the right side of Panel A reports the average R^2_{OS} of all firm-quarter out-of-sample forecasts. As shown, only the BAM has a positive R^2_{OS}. The MSPE t-statistics (in parentheses) are adjusted for firm-clustered standard errors. The result shows that the improvement in earnings forecasts by the BAM over the consensus forecast (EW) is significant at the 1% level. The R^2_{OS} values of other three methods are negative. The results show that the BAM method is superior to all other methods.

We also check the robustness of our results using Diebold–Mariano (DM, 1995) test. The DM test is an asymptotic z-test with the null hypothesis that there is equal forecast accuracy between two groups. Specifically, $DM_{AB} = \frac{\overline{d}_{AB}}{\hat{\sigma}_{\overline{d}_{AB}}} \xrightarrow{d} N(0,1)$ where $\overline{d}_{AB}$ is the sample mean loss differential between groups A and B, and $\hat{\sigma}_{\overline{d}_{AB}}$ is a consistent estimate of the standard deviation of loss differential d_{AB}. In the present case, $\overline{d}_{AB}$ is the mean difference in MSEs between two forecast models, and $\hat{\sigma}_{\overline{d}_{AB}}$ is a consistent estimate of the standard deviation of loss differential ($d_t = \left(e_{At}\right)^2 - \left(e_{Bt}\right)^2$), which accounts for serial correlation and heteroscedasticity of d_t. To ensure the difference in the forecast accuracy between the two is significant in population, not by chance, we compare DM test with the bootstrap estimation inference based on Huang et al. (2015). Assume that we have N observations of actual earnings (Y_t) and their forecasts by model 1 ($\hat{Y}_{1t}$) and model 2 ($\hat{Y}_{2t}$). The forecast errors of two models are $e_{1t} = Y_t - \hat{Y}_{1t}$ and $e_{2t} = Y_t - \hat{Y}_{2t}$, respectively. Denote their mean squared errors by μ_1 and μ_2, respectively; that is, $\mu_1 = \frac{1}{N}\sum_{t=1}^{N} e_{1t}^2$ and $\mu_2 = \frac{1}{N}\sum_{t=1}^{N} e_{2t}^2$.

To conduct the bootstrap test, we follow several steps:

Step 1. Pair these residuals from $t = 1$ to N: $\left(e_{11}, e_{21}\right), \ldots\ldots, \left(e_{1N}, e_{2N}\right)$.

Step 2. Generate a random sample of N paired observations (with replacement) using the N pairs of residuals in step 1, $\left(z_{11}, z_{21}\right), \ldots\ldots, \left(z_{1N}, z_{2N}\right)$.

Step 3. Generate the normalized $z_1(z_1^S)$ by $z_{1t}^S = z_{1t} \times \frac{\mu_1+\mu_2}{2\mu_1}$ and the normalized $z_2\left(z_2^S\right)$ by $z_{2t}^S = z_{2t} \times \frac{\mu_1+\mu_2}{2\mu_2}$. This is consistent with the null hypothesis that $E(e_{1t}^2) = E(e_{2t}^2)$.

Step 4. Calculate the DM statistics of the pseudo sample z_1^S and z_2^S.

We run 10,000 times of steps 2 to 4, generate the 10,000 pseudo samples and calculate their corresponding DM statistics. We use the generated 10,000 DM statistics to obtain the 1%, 5%, and 10% critical values. They are then used to test the statistical significance of the sample DM statistics.

The results of DM tests and bootstrap significance levels are reported in the last two rows in the right side of Panel A in Table 1. The robust DM test confirms that the BAM forecast is better than the EW forecast at the 10 percent significance level.

5.2. Decomposition of forecast errors

An interesting question is what determine forecast errors. Analyzing the composition of forecast errors sheds light on the strength of different forecast combination methods and why some methods perform better than others. To ascertain the determinants of forecast errors, we perform a decomposition of mean square errors for adjusted earnings (ROE) forecasts.

Mean square errors can be decomposed into three components (see Mincer and Zarnowitz, 1969):

$$MSE = \mathrm{E}\left(Y - \hat{Y}\right)^2 = \left(E(Y) - E\left(\hat{Y}\right)\right)^2 + (1-\hat{\beta})^2\sigma^2_{\hat{Y}} + \left(1 - \rho^2_{Y,\hat{Y}}\right)\sigma^2_Y \tag{5}$$

where Y is the actual adjusted earnings, $\hat{Y}$ is the forecasted value, σ^2_Y and $\sigma^2_{\hat{Y}}$ are the variance of actual and forecasted adjusted earnings, $\rho_{Y,\hat{Y}}$ is the correlation between Y and $\hat{Y}$, $\hat{\beta}$ is the estimated slope coefficient from the regression $Y = \alpha + \beta\hat{Y} + \varepsilon$ (for simplicity, we drop the subscripts for firms). The first component is the squared difference between the mean realized and forecast ROE values, which captures the bias in the earnings forecast. The second component is the measure of inefficiency of the forecast. When $\beta = 1$, this component reduces to zero, indicating perfect efficiency in forecast. The third component is associated with random errors. As the strength of a forecast combination method mainly lies in its ability to reduce bias and inefficiency, the first two error components are our focus. To performing decomposition of forecast errors, we run the regression of actual earnings against the forecasted earnings for each firm by imposing a restriction that a firm must have at least five out-of-sample forecasts.[9]

Panel A of Table 2 reports the mean-square-error components and the forecast regression parameters averaged across firms using the full sample. The regression-based combination method substantially reduces the bias in the consensus earnings forecast. For example, the magnitude of the bias components is 0.009 for the consensus forecasts (EW) whereas it is only 0.006 for the BAM, about one third lower. The bias reduction is significant at five percent with a t-value of −2.03. The results show effectiveness of the BAM in reducing biasedness in analyst earnings forecasts. The BAM method also reduces the forecast inefficiency of the EW method by 14.78% (0.001). The t value of this improvement is −1.56, which falls short of significance at the five percent level.

Overall, the results of forecast error decomposition show that the regression-based combination method reduces bias and inefficiency of out-of-sample forecasts. This explains why the BAM provides more accurate prediction of earnings than other forecasts.

[9] This imposition reduces the sample size somewhat and affects mean MSE for each forecast method.

Table 2
Decomposition of forecast errors.

Panel A: The full sample

		PBest	EW	IMSE	Odds	BAM
Mean MSE		0.116	0.064	0.076	0.076	0.046
Mean relative MSE		1.806	1.000	1.182	1.185	0.712
Bias	Mean value	0.015	0.009	0.011	0.010	0.006
	Diff. from EW	0.006[b]		0.002[b]	0.001[b]	−0.003[b]
	Diff. %	68.62%		22.51%	12.79%	−37.09%
	t-test	(2.26)		(2.06)	(2.02)	(−2.03)
	Mean %	12.87%	13.79%	14.29%	13.12%	12.19%
Inefficiency	Mean value	0.014	0.008	0.009	0.009	0.007
	Diff. from EW	0.006[b]		0.002	0.001	−0.001
	Diff. %	83.94%		20.59%	18.28%	−14.78%
	t-test	(2.027)		(0.483)	(0.456)	(−1.56)
	Mean %	12.18%	11.96%	12.19%	11.93%	14.31%
Random errors	Mean value	0.087	0.048	0.056	0.057	0.033
	Mean %	74.95%	74.25%	73.52%	74.95%	73.50%

Panel B: The subsample results with iterated regression combination

		PBest	EW	IMSE	Odds	BAM	IMC	IWC
Mean MSE		0.0186	0.0162	0.0181	0.0186	0.0144	0.0124	0.0127
Mean relative MSE		1.1439	1.0000	1.1150	1.1453	0.8896	0.7621	0.7815
Bias	Mean value	0.0038	0.0032	0.0035	0.0037	0.0028	0.0025	0.0027
	Diff. from EW	0.0006[c]		0.0003	0.0005	−0.0004	−0.0007[c]	−0.0005[b]
	Diff. %	19.07%		10.35%	15.21%	−11.19%	−22.02%	−15.11%
	t-test	(1.70)		(1.27)	(1.31)	(−1.60)	(−1.81)	(−2.19)
	Mean %	20.43%	19.63%	19.43%	19.75%	19.60%	20.09%	21.32%
Inefficiency	Mean value	0.0009	0.0008	0.0009	0.0009	0.0008	0.0005	0.0008
	Diff. from EW	0.0001		0.0001	0.0001	0.0000	−0.0003	0.0000
	Diff. %	12.29%		18.75%	19.08%	3.14%	−36.20%	5.34%
	t-test	(1.180)		(0.986)	(1.346)	(0.00)	(−1.40)	(0.06)
	Mean %	4.69%	4.77%	5.09%	4.96%	5.54%	4.00%	6.44%
Random errors	Mean value	0.014	0.012	0.014	0.014	0.011	0.009	0.009
	Mean %	74.88%	75.59%	75.49%	75.29%	74.87%	75.91%	72.24%

This table summarizes the results of decomposing the mean square error (MSE). MSE can be decomposed into three components: $\mathrm{MSE} = \mathrm{E}\left(Y-\hat{Y}\right)^2 = \left(E(Y)-E\left(\hat{Y}\right)\right)^2 + (1-\hat{\beta})^2\sigma^2_{\hat{Y}} + \left(1-\rho^2_{Y,\hat{Y}}\right)\sigma^2_Y$ where Y is the actual adjusted earnings (ROE), $\hat{Y}$ is the forecasted value, σ^2_Y is the variance of actual adjusted earnings, $\sigma^2_{\hat{Y}}$ is the variance of forecasted value, $\rho_{Y,\hat{Y}}$ is the correlation between Y and $\hat{Y}$, and β is from the regression $\mathrm{Y} = \alpha + \beta\hat{Y} + \varepsilon$. The first component represents the bias of forecast, the second component measures inefficiency, and the third one captures random errors. Diff. from EW is the difference between EW and other models, and difference in percentage to EW is denoted as Diff. %. Mean % is the proportion of each error component to the total MSE. Panel A reports the results based on the full sample and Panel B is based the subsample that meets the data requirement for iterated regression combination. The signs [a], [b] and [c] denote the significance at 1%, 5% and 10%, respectively. In this decomposition analysis, we require at least 5 out-of-sample forecast observations for each firm and therefore, sample size is smaller than that in Table 1 and results can be different.

5.3. Analyst/firm characteristics and forecast accuracy

The literature has suggested that analysts' forecast accuracy depends on characteristics of firms and properties of analysts' forecasts (Brown et al., 2015; Kothari et al., 2016). Factors such as analysts' ability, available resources, and portfolio complexity influence forecast accuracy. Forecast accuracy tends to increase with analysts' experience (a proxy for ability) and employer size (a proxy for available resources), and decrease with the number of firms and industries followed by an analyst. Forecast accuracy also decreases with uncertainty as measured by volatility of earnings and stock prices, and complexity of firms. In addition, individual analyst characteristics and incentives affect the bias, accuracy, and timeliness of analysts' forecasts.

An issue of interest is whether forecast combination produces a greater benefit when analyst forecasts are more biased and unstable. To address this issue, we examine the performance of forecast combination methods for the groups with different analyst and firm characteristics, industries and market conditions.

We first perform subsample analysis based on analyst and firm characteristics. In each quarter, firms are sorted into deciles from low to high by each characteristic, where $D = 1$ and 10 refer to the lowest and highest deciles, respectively. We use two dummy variables, DIV^d and MGT^d, to capture the effects of dividend policy and management forecasts of earnings. $DIV^d = 1$ if a firm distributes dividends in a given quarter and 0, otherwise. Similarly, we set $MGT^d = 1$ when a firm's management provides an earnings forecast and 0, otherwise.

Table 3 shows that when sorted by characteristic of analyst forecast and firm earnings, earnings forecast errors are lower for firms with a high number of analyst forecasts (*N-analyst*), high frequency (*NREV*) of forecast revisions, low analyst forecast dispersion (*DISP*), small earnings forecast revision prior to the announcement of actual earnings (*REV*), low volatility of earnings (*VOLEPS*) and earnings forecast errors (*VOLERR*), and less earnings surprise (*SUR*). These results are consistent with the view that analyst forecasts are more accurate when there is less information uncertainty, the information environment is more favorable, and more analysts provide forecasts. Combination forecast performance reflects these properties. As shown, the difference between lowest and highest deciles is larger for portfolios sorted by analysts' dispersion (*DISP*), volatility of earnings (*VOLEPS*), earnings forecast errors (*VOLERR*) and earnings surprise (*SUR*).

To ensure that the difference in forecast accuracy between high and low deciles is not due to randomness, we employ the bootstrap method to examine its significance. For each forecast model, we select a random sample for each group, $D = 1$ and D = 10, and compute mean-difference t statistics between the two groups. We repeat this procedure for 10,000 times and generate critical values of t statistics at the 1%, 5%, and 10% significance levels, which are then used to determine the significance level for the group mean differences.

Table 3 (the third to fifth rows for each subsample test) shows that the differences in the MSEs between the highest and lowest deciles (10 – 1) are overwhelmingly significant. The results strongly suggest that out-of-sample forecast accuracy depends on the characteristics of analyst forecasts and the earnings generating process. The BAM delivers much better forecast performance than the consensus forecast when heterogeneity of analysts is high, or the information and earnings are more uncertain.

The right side of Table 3 shows the improvement over the EW method for high and low decile groups sorted by characteristic. The largest decreases in MSE of the BAM relative to that of EW occur for firms with fewer analysts (N-*analyst*), high dispersion in analyst forecast (*DISP*), fewer revision (*NREV*), large forecast revisions (*REV*), high volatility of earnings (*VOLEPS*) and earnings forecast errors (*VOLERR*) and larger earnings surprise (*SUR*). The BAM delivers the largest gain in out-of-sample forecasts when information and earnings are more uncertain and analyst forecasts are more diverse, consistent with our findings from Monte Carlo simulations (see the Online Appendix). For example, for the firms with low analyst following (N-*analyst*, $D = 1$) and high analyst forecast dispersion (*DISP*, $D = 10$), the BAM is able to reduce the MSEs of consensus forecasts by 28% and 27%, respectively. A similar pattern is found for firms with larger earnings revision, high volatility of earnings and forecast errors and large earnings surprise. The results confirm that the benefit of using the BAM method is larger when analysts are more heterogeneous or the information is more uncertain and harder to interpret.

For the subsample analysis based on firm characteristics, we also find that the forecast performance depends on these characteristics. Table 4 shows that forecast errors are generally smaller for firms with large size (*SIZE*), high book-to-market ratio (*BM*), older age (*AGE*), having dividend payout (DIV^d) and management earnings forecast announcement (MGT^d), and higher institutional ownership (*IO*). Also, firms with high quality of earnings (low *AQ2*) have lower MSEs. The BAM delivers the best out-of-sample forecasts in most cases. The results suggest that regression-based adjustment works better for firms that provide less information (or signals) to the public or when they are more opaque.

5.4. Industry analysis

In this section, we examine forecast performance for firms in different sectors. Firms in different industries often have distinct business and operation characteristics and different choices of accounting procedure, which may contribute to variations in accuracy of earnings forecasts. To investigate this possibility, we group firms into ten industry groups by the first two-digit standard industry classification (SIC) code.

Table 5 reports the results of out-of-sample adjusted earnings (ROE) forecasts by industry (sector). The results show variations in forecasting accuracy across sectors. Firms in Construction, Manufacturing and retail trade industries have low forecasting accuracy and those in Agriculture, Forestry & Fishing, and Finance, Insurance & Real Estate have high forecasting accuracy. To see if there are significant differences in forecasting accuracy across industries, we employ the ANOVA F test. The null hypothesis is that mean-square-errors are the same across industries for each combination method. Test results are reported at the bottom of the table. The null hypothesis is overwhelmingly rejected at the 1% significance level for all combination methods, suggesting that there are significant differences in forecasting accuracy across industries. Importantly, the BAM performs better or at least as well as other methods across all industries. The greatest improvement is for firms in Construction group, where the BAM reduces the average MSE of EW by about 50%, suggesting that the regression-based method is more advantageous when firm type is vague.

5.5. Forecasting performance under different regimes

Firms' profitability and earnings stability depend on business conditions. Firms typically face higher uncertainty in earnings outcomes when the market is uncertain or the economy is in recession. The uncertain environment makes it more difficult for analysts to deliver accurate earnings forecasts. Combining the noisier forecasts would likely result in lower forecast accuracy in general but an important question is which combination method is more robust to noisy analyst forecasts.

To see the effect of market conditions on forecasting performance, we compare out-of-sample forecasts for subperiods with different market uncertainty and economic conditions. We use CBOE *VIX* as a measure for market uncertainty. This option-based index measures volatility in the stock market. In addition, we use the cyclically adjusted aggregate P/E ratio (*CAPE*) to capture the performance in the stock market. High P/E usually is associated with a bull market and low P/E with a bear market. The market tends to be more uncertain in a bear market.[10] We sort the time-series observations based on these market uncertainty measures into deciles from low ($D = 1$) to high ($D = 10$). For macroeconomic conditions, we use the seasonal and inflation adjusted growth

[10] Extremely high CAPE suggests that securities are more likely overvalued, often associated with financial market booming and subsequent market crash. For example, since 1881, there are only three periods with CAPE over 25: 1929, 1999, and 2007.

Table 3
Out-of-sample performance of forecast combination models by analyst and earnings characteristics.

		Models					Difference between models and EW			
		PBest	EW	IMSE	Odds	BAM	PBest	IMSE	Odds	BAM
N-analyst	D = 1	0.103	0.074	0.078	0.078	0.053	0.030	0.005	0.005	−0.021
	D = 10	0.012	0.005	0.012	0.012	0.005	0.007	0.007	0.007	0.000
	10 − 1	−0.091	−0.068	−0.066	−0.066	−0.048				
	t test	(−5.13)	(−3.44)	(−5.93)	(−5.85)	(−2.71)				
	Bootstrap	1%	1%	1%	1%	1%				
DISP	D = 1	0.011	0.003	0.006	0.006	0.002	0.008	0.003	0.003	0.000
	D = 10	0.391	0.248	0.227	0.230	0.180	0.143	−0.021	−0.017	−0.068
	10 − 1	0.380	0.245	0.221	0.224	0.177				
	t test	(7.29)	(5.97)	(10.43)	(10.30)	(5.09)				
	Bootstrap	1%	1%	1%	1%	1%				
NREV	D = 1	0.095	0.054	0.062	0.058	0.041	0.041	0.008	0.004	−0.013
	D = 10	0.034	0.013	0.030	0.031	0.011	0.021	0.017	0.019	−0.002
	10 − 1	−0.061	−0.041	−0.032	−0.027	−0.030				
	t test	(−3.08)	(−2.17)	(−2.80)	(−2.58)	(−1.74)				
	Bootstrap	1%	1%	1%	1%	1%				
REV	D = 1	0.008	0.002	0.008	0.008	0.002	0.005	0.005	0.005	−0.001
	D = 10	0.335	0.202	0.189	0.193	0.141	0.133	−0.013	−0.009	−0.061
	10 − 1	0.327	0.199	0.181	0.185	0.139				
	t test	(5.71)	(5.49)	(8.74)	(8.58)	(4.65)				
	Bootstrap	1%	1%	1%	1%	1%				
VOLEPS	D = 1	0.021	0.008	0.017	0.017	0.006	0.014	0.009	0.009	−0.002
	D = 10	0.293	0.208	0.163	0.168	0.142	0.085	−0.045	−0.041	−0.066
	10 − 1	0.271	0.200	0.146	0.151	0.137				
	t test	(6.13)	(5.70)	(8.86)	(8.76)	(4.86)				
	Bootstrap	1%	1%	1%	1%	1%				
VOLERR	D = 1	0.004	0.001	0.004	0.004	0.001	0.003	0.003	0.003	0.000
	D = 10	0.378	0.237	0.223	0.227	0.173	0.141	−0.014	−0.010	−0.064
	10 − 1	0.373	0.236	0.219	0.223	0.172				
	t test	(7.44)	(6.04)	(10.60)	(10.48)	(5.18)				
	Bootstrap	1%	1%	1%	1%	1%				
SUR	D = 1	0.014	0.000	0.014	0.014	0.001	0.014	0.014	0.014	0.001
	D = 10	0.361	0.280	0.211	0.215	0.212	0.081	−0.069	−0.066	−0.068
	10 − 1	0.347	0.280	0.198	0.201	0.211				
	t test	(7.10)	(6.87)	(10.46)	(10.32)	(6.00)				
	Bootstrap	1%	1%	1%	1%	1%				

This table reports the mean MSE and its difference by characteristics of analyst forecasts and earnings. For each characteristic, the difference between deciles is noted as 10 − 1, D = 10 (high) minus D = 1 (low). We apply the mean difference t-test and bootstrap to examine the significance of the difference of forecast accuracy between two groups. The bootstrap value indicates the significance level of t tests relative to bootstrap cutoffs where 1%, 5% and 10% denote the significance level. The right columns report the differences in MSE between EW and other forecasts.

rates of gross national products, measured by both annual growth rate (*GDP1*) and mean quarterly growth rate (annualized, *GDP2*). We sort these two GDP series into quartiles and define the bottom 25% as the bad economy and the top 25% as the good economy. In addition, we consider two sentiment indicators, one is investor sentiment (INV) from Baker and Wurgler (2006) and another is manager sentiment (MGT) from Jiang et al. (2019). We sort the time-series observations based on these two sentiment indexes into deciles from low ($D = 1$) to high ($D = 10$).

Table 6 reports the results of forecasts under different economic regimes. Across all forecast methods, earnings forecasting performance tends to worsen during the periods of high market volatility, low P/E, and economic recession. The tests show that the difference between two time periods is more significant using VIX and two GDP measures but the results are mixed using CAPE, INV and MGT.[11] This finding suggests that it is generally more difficult to forecast future earnings when market is volatile and economy is bad. In contrast, investor and manager sentiments have a limited impact on forecasts accuracy. More importantly, the results continue to show that the BAM consistently delivers superior earnings forecasting performance relative to the consensus forecast. Mean square errors are always lower for the BAM, regardless of the market and macroeconomic conditions.

[11] The results show that investor sentiment plays a role for PBest and value-weighted forecasts (IMSE and Odds) but not for EW and BAM. When investor sentiment is high, forecasts errors using PBest, IMSE and Odds are smaller. A possible reason is that when sentiment is high, investors are more active and analysts are more willing to spend time to provide forecasts for firms that receive less attention in normal times. This could improve the performance of PBest, IMSE and Odds forecast methods, which put more weights on the forecasts of better analysts. On the other hand, EW and BAM forecast accuracy is less dependent on investor sentiment. The EW method aggregates the forecasts over a larger group of analysts, instead of a few star analysts. Given that analysts are generally more rational than individual investors, EW forecasts are more robust to investor sentiment. Since BAM forecasts are refinement of EW forecasts, they further improve the forecast efficiency and therefore are less sensitive to investor sentiment. Also, manager sentiment seems to have a limited impact on forecasts accuracy. It could be due to the data availability. As manager sentiment is only available from 2003, covering only about 42% of the period in our study.

Table 4
Out-of-sample performance of forecast combination models by firm characteristic.

		Models					Difference between models and EW			
		PBest	EW	IMSE	Odds	BAM	PBest	IMSE	Odds	BAM
SIZE	D = 1	0.065	0.062	0.057	0.055	0.042	0.003	−0.005	−0.007	−0.020
	D = 10	0.004	0.004	0.002	0.002	0.003	0.000	−0.002	−0.002	−0.001
	10 − 1	−0.062	−0.058	−0.055	−0.053	−0.039				
	t test	(−6.00)	(−5.33)	(−6.80)	(−6.60)	(−4.71)				
	Bootstrap	1%	1%	1%	1%	1%				
BM	D = 1	0.110	0.066	0.086	0.089	0.050	0.044	0.020	0.023	−0.016
	D = 10	0.055	0.015	0.023	0.023	0.010	0.041	0.008	0.008	−0.005
	10 − 1	−0.055	−0.051	−0.063	−0.066	−0.040				
	t test	(−1.82)	(−3.84)	(−5.34)	(−5.19)	(−4.11)				
	Bootstrap	10%	1%	1%	1%	1%				
NREALIZ	D = 1	0.085	0.020	0.052	0.052	0.014	0.065	0.032	0.032	−0.007
	D = 10	0.078	0.021	0.031	0.032	0.012	0.057	0.011	0.011	−0.009
	10 − 1	−0.007	0.001	−0.021	−0.020	−0.001				
	t test	(−0.20)	(0.09)	(−1.67)	(−1.57)	(−0.26)				
	Bootstrap	–	–	5%	5%	–				
AGE	D = 1	0.025	0.025	0.022	0.022	0.016	0.000	−0.002	−0.002	−0.008
	D = 10	0.004	0.002	0.004	0.004	0.003	0.002	0.002	0.002	0.000
	10 − 1	−0.021	−0.022	−0.018	−0.018	−0.014				
	t test	(−6.62)	(−4.17)	(−6.48)	(−5.88)	(−3.90)				
	Bootstrap	1%	1%	1%	1%	1%				
DIV^d	D = 1	0.028	0.015	0.014	0.014	0.012	0.013	−0.001	−0.001	−0.003
	D = 0	0.119	0.058	0.083	0.084	0.043	0.062	0.025	0.026	−0.014
	10 − 1	0.091	0.043	0.069	0.070	0.032				
	t test	(7.16)	(7.32)	(11.99)	(11.99)	(4.12)				
	Bootstrap	1%	1%	1%	1%	1%				
DIV	D = 1	0.029	0.016	0.029	0.029	0.014	0.012	0.013	0.013	−0.002
	D = 10	0.025	0.012	0.020	0.019	0.009	0.013	0.008	0.007	−0.003
	10 − 1	−0.004	−0.004	−0.009	−0.010	−0.005				
	t test	(−0.44)	(−0.70)	(−0.95)	(−1.05)	(−0.92)				
	Bootstrap	–	–	–	–	–				
MGT^d	D = 1	0.007	0.003	0.006	0.007	0.003	0.004	0.003	0.004	0.000
	D = 0	0.082	0.040	0.053	0.053	0.030	0.042	0.012	0.013	−0.010
	10 − 1	0.075	0.038	0.046	0.047	0.027				
	t test	(10.18)	(7.32)	(13.90)	(13.60)	(6.25)				
	Bootstrap	1%	1%	1%	1%	1%				
IO	D = 1	0.024	0.022	0.016	0.015	0.018	0.002	−0.006	−0.007	−0.004
	D = 10	0.010	0.005	0.010	0.010	0.004	0.005	0.005	0.006	−0.001
	10 − 1	−0.013	−0.017	−0.006	−0.005	−0.015				
	t test	(−1.32)	(−1.99)	(−1.11)	(−0.90)	(−1.96)				
	Bootstrap	–	5%	–	–	5%				
AQ1	D = 1	0.079	0.032	0.064	0.063	0.024	0.047	0.032	0.031	−0.008
	D = 10	0.107	0.044	0.070	0.069	0.037	0.062	0.026	0.025	−0.008
	10 − 1	0.028	0.013	0.006	0.007	0.013				
	t test	(0.87)	(0.85)	(0.39)	(0.41)	(1.07)				
	Bootstrap	–	–	–	–	–				
AQ2	D = 1	0.074	0.017	0.060	0.059	0.013	0.056	0.042	0.042	−0.004
	D = 10	0.107	0.047	0.068	0.068	0.039	0.060	0.022	0.021	−0.007
	10 − 1	0.033	0.029	0.009	0.009	0.026				
	t test	(1.03)	(2.09)	(0.51)	(0.49)	(2.13)				
	Bootstrap	–	1%	–	–	1%				

This table reports the mean MSE and the differences in mean MSEs by variables of firm-specific characteristics. At each quarter, firms are sorted into decile groups based on one characteristics where D = 1 stands for lowest value and D = 10 stands for highest value. DIV^d and MGT^d are two dummy variables, where value = 1 if a firm at quarter t has dividend payment and management forecasts, respectively. For each characteristic, the difference between the top and bottom group is noted as 10 − 1. We apply the mean difference t-test and bootstrap to examine the significance of the difference of forecast accuracy between two groups. The bootstrap value indicates the significance level of t tests relative to bootstrap cutoffs where 1%, 5% and 10% denote the significance level and "–" indicates insignificance. The right columns report the differences in MSE between EW and other forecasts.

5.6. Forecasts using iterated regression methods

In the preceding analysis, we choose the BAM method to represent the regression adjustment approach. This method is easy to implement and less data demanding, which enables us to draw a large sample with more statistical power. Recent research has suggested an iterated (or two-step) regression approach to improve forecast combination (Lin et al., 2018). This approach makes use of past information for each forecaster (analyst) to make a finer adjustment. As more information is used than in the single-step

Table 5
Out-of-sample performance of forecast combinations by industry.

	Models					Difference between models and EW			
	PBest	EW	IMSE	Odds	BAM	PBest	IMSE	Odds	BAM
Agriculture, Forestry & Fishing	0.0006	0.0005	0.0006	0.0005	0.0006	0.0001	0.0001	0.0000	0.0001
Mining	0.0734	0.0206	0.0514	0.0531	0.0183	0.0528	0.0308	0.0325	−0.0024
Construction	0.3195	0.0762	0.0774	0.0766	0.0378	0.2432	0.0012	0.0004	−0.0384
Manufacturing	0.1312	0.0624	0.0830	0.0833	0.0474	0.0688	0.0206	0.0209	−0.0150
Transportation & Public Utilities	0.0290	0.0242	0.0265	0.0265	0.0149	0.0048	0.0022	0.0023	−0.0093
Wholesale Trade	0.0062	0.0057	0.0035	0.0041	0.0050	0.0005	−0.0022	−0.0016	−0.0007
Retail Trade	0.0217	0.0092	0.0221	0.0227	0.0080	0.0125	0.0129	0.0135	−0.0012
Finance, Insurance & Real Estate	0.0035	0.0037	0.0034	0.0036	0.0033	−0.0003	−0.0004	−0.0001	−0.0004
Service	0.0181	0.0170	0.0168	0.0179	0.0142	0.0012	−0.0001	0.0009	−0.0028
Non-classifiable	0.0589	0.0418	0.0529	0.0537	0.0419	0.0171	0.0111	0.0119	0.0001
F value	11.8622	4.0114	18.2745	17.4953	3.1182				
p-value	0.0000	0.0000	0.0000	0.0000	0.0009				

This table reports the mean MSE (left) and the difference of mean MSEs between other forecast models and EW (right) by industries. All firms are grouped into ten industries by the first 2-digit SIC code. We apply one-way ANOVA F test to examine the null hypothesis that the mean of squared errors of each industry is same. The last row reports the p-values of the F tests.

Table 6
Out-of-sample forecasts under different regimes.

		Models					Difference between models and EW			
		PBest	EW	IMSE	Odds	BAM	PBest	IMSE	Odds	BAM
VIX	D = 1	0.043	0.020	0.033	0.033	0.015	0.023	0.013	0.013	−0.005
	D = 10	0.118	0.041	0.059	0.059	0.027	0.077	0.019	0.018	−0.013
	10 − 1	0.074	0.021	0.026	0.026	0.012				
	t test	(2.23)	(2.01)	(2.27)	(2.23)	(1.77)				
	Bootstrap	5%	10%	5%	5%	10%				
CAPE	D = 1	0.099	0.028	0.053	0.052	0.020	0.072	0.026	0.024	−0.008
	D = 10	0.018	0.016	0.016	0.020	0.014	0.001	0.000	0.004	−0.003
	10 − 1	−0.081	−0.011	−0.037	−0.031	−0.006				
	t test	(−2.24)	(−0.61)	(−3.36)	(−2.61)	(−0.52)				
	Bootstrap	1%	–	1%	1%	–				
GDP1	Bad	0.103	0.054	0.059	0.060	0.038	0.050	0.005	0.007	−0.016
	Good	0.036	0.016	0.031	0.032	0.012	0.020	0.015	0.016	−0.004
	Good - Bad	−0.067	−0.037	−0.028	−0.028	−0.026				
	t test	(−3.40)	(−3.80)	(−3.58)	(−3.38)	(−3.48)				
	Bootstrap	1%	1%	1%	1%	1%				
GDP2	Bad	0.113	0.053	0.056	0.056	0.037	0.060	0.002	0.003	−0.016
	Good	0.019	0.014	0.018	0.019	0.012	0.005	0.003	0.005	−0.003
	Good - Bad	−0.094	−0.039	−0.038	−0.037	−0.026				
	t test	(−5.02)	(−3.59)	(−5.86)	(−5.34)	(−2.92)				
	Bootstrap	1%	1%	1%	1%	1%				
INV	D = 1	0.104	0.036	0.065	0.065	0.024	0.068	0.029	0.029	−0.012
	D = 10	0.017	0.036	0.015	0.015	0.029	−0.018	−0.020	−0.020	−0.007
	10 − 1	−0.087	−0.001	−0.050	−0.050	0.005				
	t test	(−3.19)	(−0.03)	(−4.91)	(−4.95)	(0.26)				
	Bootstrap	1%	–	1%	1%	–				
MGT	D = 1	0.096	0.018	0.081	0.073	0.013	0.077	0.063	0.054	−0.005
	D = 10	0.084	0.034	0.058	0.058	0.025	0.050	0.025	0.024	−0.009
	10 − 1	−0.012	0.016	−0.022	−0.015	0.011				
	t test	(−0.39)	(1.36)	(−1.01)	(−0.75)	(1.31)				
	Bootstrap	–	–	–	–	–				

This table reports the mean MSE (left) and the MSE differences between other forecast models and EW (right) under different market and macroeconomic conditions. For *VIX*, the cyclically adjusted P/E ratio (*CAPE*), investor sentiment (INV) and manager sentiment (MGT), we sort all time-series observations by each of these four variables into deciles and calculate the difference in forecast errors between high ($D = 10$) and low ($D = 1$) deciles (10 − 1). For the GDP growth annual (*GDP1*) and quarterly (*GDP2*), we sort these two time series into quartiles and denote the bottom 25% as the bad economy and the top 25% as the good economy to calculate the difference in forecasting errors between good and bad economies (Good–Bad). We apply the mean difference t-test, and bootstrap to examine the significance of the difference of forecast accuracy between two groups. The bootstrap value indicates the significance level of t tests relative to bootstrap cutoffs where 1%, 5% and 10% denote the significance level and "–" indicates insignificance.

regression, it has a potential to further improve the forecasting performance (see Appendix A). In this section, we further investigate more sophisticated regression-adjustment methods (IMC and IWC).

To ensure more reliable parameter estimation in the first-step time-series regression, we impose a minimum of 20 quarterly historical and forecast earnings records. In addition, we require at least six analysts following each firm in a given forecast period to allow cross-sectional variations. This data requirement results in a sample with 1106 firms covering 21,696 firm-quarter out-of-sample forecasts.[12] Later, we examine the sensitivity of our results to the sample size due to this data requirement (see the discussion and results in the Online Appendix). Using this refined sample, we re-estimate the results for all forecast methods.

Panel B of Table 1 reports the results for all combination methods based on the data sample that suits IMC and IWC methods. The biased-adjusted mean (BAM) and iterated regression (IMC and IWC) methods outperform the consensus forecast (EW) method. The MSEs of BAM, IMC and IWC methods are all significantly lower than that of EW. The three regression-based combination forecast methods have R^2_{OS} values of 6.87%, 11.03% and 13.49%, respectively (see row 3 on the right side). Among the three regression-based methods, IMC and IWC perform better than BAM in percentage terms by reducing MSEs by 6.25% and 9.38%, respectively.[13] The MSPE t-statistics (in parentheses) show that the improvements in out-of-sample earnings forecasts by the regressed combination methods over the consensus forecast (EW) are significant at least at the 10% level. The DM tests in row 5 confirm the superiority of the regressed combination forecasts. The bootstrap DM tests indicate that the improvement by IMC and IWC are significant at the 1% level. The results show that the iterated regression methods further improve the single-step regression adjustment method.

We next examine the composition of forecast errors for iterated regression combination relative to other methods. Panel B of Table 2 reports the MSE components for all methods including IMC and IWC based on the refined sample. The results again show that the regression-based methods substantially reduce the bias in consensus earnings forecast. For example, the magnitude of the bias component is 0.0032 for the consensus forecast (EW) whereas it is only 0.0028, 0.0025 and 0.0027 for BAM, IMC and IWC, respectively. On average, the three regression-based methods reduce the bias in consensus forecasts by 11%, 22%, and 15%, respectively. On the other hand, the methods of previous best (PBest), odd ratios (Odds) and inverse MSE (IMSE) fail to improve the consensus forecast.

Collectively, the results suggest that the regression-based combination methods deliver better forecasting performance and they are more effective in reducing bias than other forecasts. The two-step regression methods further improve the single-step regression method in reducing the bias of earnings forecasts. However, the incremental improvement of the IMC and IWC over the BAM is often moderate and comes with a price of a smaller sample size.

Fig. 2 plots the time series of alphas and betas estimated by the BAM and IMC over the sample period. The vertical axis on the left side indicates the value for betas and on the right side for alphas. There is evidence that alpha values drop in the 1987 market crash and 1992–93 recession. A more pronounced drop occurs during the subprime crisis. The results suggest that the analyst earnings forecasts were too optimistic (upward biased) in these periods of downturns, and the decrease in alpha corrects this analyst-specific upward bias. The IMC alpha shows a similar pattern but less drastic drops during the market downturn and subprime crisis, reflecting the strength of this forecast method to produce more stable forecasts. On the other hand, betas are relatively smooth over the sample period,[14] except for a visible upward adjustment during the 1987 stock market crash and the subprime crisis. These increases in beta are attributable to higher earnings uncertainty at the aggregate level. Beta increases when the market is more uncertain and analyst forecasts are noisier to perform an error correction.[15]

5.7. The role of analyst heterogeneity and other information

The analysis above suggests that the performance of a forecasting method depends on heterogeneity in analyst forecast accuracy, firm characteristics and information, and economic conditions. In this section, we consider these factors in forecast combination.

The literature has suggested that earnings forecasts can be improved by combining analysts with other forecasting information (see, for example, Jame et al., 2016; Ball and Ghysels, 2018). In light of the literature, we can modify the forecasting model as follows:

$$Y_{j,t} = a_i + b_i \cdot \hat{Y}^i_{j,t} + \sum_{k=1}^{K} \beta_{k,i} \cdot FV_{k,t-1} + \varepsilon_{it} \tag{6}$$

where $Y_{j,t}$ is the realized earnings of firm j, $\hat{Y}^i_{j,t}$ is the analyst forecast at t-1 for earnings at t, and $FV_{k,t-1}$ consists of other forecasting information, such as firm-level and macroeconomic data, available before firm's earnings announcement at time t, and $\beta_{k,i}$ reflects the value of FV for refining the forecast of analyst i. If FV indeed contains useful information for future earnings above and beyond the analyst's forecast, incorporating this information set will improve earnings forecasts.

Following the literature, we include accounting and financial variables related to firm profitability and macroeconomic variables (see Appendix B) in the predictive regression. The parameter estimates are then used to adjust the analyst's forecast, that is,

[12] The forecast period is from September 1991 to June 2016.

[13] These figures are based on the results on the left side of Panel B where 6.25% reduction is from (0.0030−0.0032)/0.0032 and 9.38% reduction is from (0.0029−0.0032)/0.0032.

[14] The IMC beta estimates are volatile before 1990 because this two-step regression method needs more data to produce stable parameter estimates. The results confirm that to achieve satisfactory forecasts, the two-step regression requires more time-series forecast data.

[15] Capistrán and Timmermann (2009, p. 431) derive the beta formula = $N\sigma^2_F/(\sigma^2_\varepsilon + N\sigma^2_F)$ where N is the number of analyst forecasts, σ^2_F and σ^2_ε are the variances of the market factor and analyst forecast errors, respectively, under the assumption that analysts fully respond to market news F. Beta increases with market uncertainty σ^2_F. Under the more realistic assumption when analysts may underact to market news or respond slowly to the market factor change, it can be shown that beta = $N\sigma^2_F/(\sigma^2_\varepsilon + Nb^2\sigma^2_F)$ where b is the response coefficient of analysts to F. When the market becomes more uncertain, analysts may become more conservative (underreact) and b becomes smaller. In such case, the beta value will be even higher when the market uncertainty heightens as in the financial crisis or market crash.

Table 7
Out-of-sample performance of forecasts accounting for heterogeneity and other forecasting information.

	PBest	EW	IMSE	Odds	BAM	BAM*	PBest - EW	IMSE - EW	Odds - EW	BAM - EW	BAM* −EW	BAM* −BAM
MSE	0.0074	0.0055	0.0065	0.0063	0.0048	0.0036	0.0020	0.0010	0.0008	−0.0019	−0.0045	−0.0012
t-value							(2.33)	(1.70)	(1.54)	(−1.84)	(−2.71)	(−2.04)
R^2_{OS} (%)							−35.85%	−17.93%	−14.16%	12.32%	34.60%	25.53%
t-value							(2.20)	(1.74)	(2.01)	(2.31)	(6.21)	(6.42)

This table compares the mean MSEs and $R^2_{OS}s$ of other model forecasts for ROEs with those of BAM* forecasts which are generated from combining analyst forecasts and firm-level and macroeconomic forecast variables and adjusted for heterogeneity in analyst forecast accuracy. A more negative difference means a greater decrease in forecast errors and a higher positive R^2_{OS} indicates a better out-of-sample forecast performance. The t-values are in parentheses.
Results are different from the one in Table 1. Because some firms included in Table 1 do not have sufficient accounting information and observations to conduct BAM forecasts.

$\tilde{Y}^i_{j,t+1} = \hat{a}_i + \hat{b}_i \cdot \hat{Y}^i_{j,t+1} + \sum_{k=1}^{K} \hat{\beta}_{k,i} \cdot FV_{k,t}$. We estimate (6) by the stepwise regression which select a subset of predictors that maximize the explanatory power.

Next, to account for the effect of heterogeneity in forecast accuracy, we use a characteristic-weighted average method to adjust for the mean forecast. We first run the following regression of individual analyst forecast errors against characteristic variables:

$$PAFE_{i,t} = c_{0,i} + \sum_{q=1}^{Q} c_{q,i} Z_{q,i,t} + \upsilon_{i,t} \tag{7}$$

where $PAFE_{i,t}$ is the percentage of analyst forecast error at time t and $Z_{q,i,t}$ contains characteristic variables that are relevant to analysts' forecast accuracy. $PAFE_{i,t}$ is an analyst i's absolute forecast error scaled by the mean of absolute forecast errors across all analysts following a firm. Following Jame et al. (2016), we consider the analyst's past forecast precision, tenure, age, and the numbers of firms and SICs covered by the analyst (see Appendix B). We estimate the forecast error that is related to analyst characteristics $\widehat{PAFE}_{i,t} = \hat{c}_{0,i} + \sum_{q=1}^{Q} \hat{c}_{q,i} \cdot Z_{q,i,t}$. We then calculate the weighted average forecast $\overline{Y}_{j,t+1} = \sum_i w_i \tilde{Y}^i_{j,t+1}$ where the weight is the inverse of $\widehat{PAFE}_{i,t}$ and use this adjusted mean forecast to run the BAM regression.

Table 7 compares the BAM* forecasts with others. The results show that accounting for the differences in analyst forecast accuracy and incorporating the information of firm-level financial and market variables further improves the performance of the BAM combination forecasts. This finding implies that the usefulness of the regressed combination method proposed in this paper goes beyond just combining analyst forecasts. In particular, our evidence reveals that the BAM method present an enormous amount of potential to improve earnings forecasts by further combining analyst forecasts with other valuables such as accounting, financial market, macroeconomic, media and internet data, whatever information is relevant to firm's future profitability.

5.8. Comparison with other novel forecast methods

Principal component analysis (PCA) and more recently machine learning methods have also been used extensively in the literature to forecast returns and other financial variables. In this section, we compare the performance of BAM model with these methods. To forecast using the principal components, we use all available analyst forecast for firm j up to time t as the predictors and construct their first principal component. Since this setting requires predictors to have a balanced panel, we replace the missing values with their cross-sectional average analyst forecast of the same period if the information of one analyst forecast is not available. We construct the first principal component using the raw forecast data ($\widehat{PCA}$) or the standardized forecast data ($\widehat{SPCA}$). We then forecast the future earnings from the following predictive regression:

$$Y_{jt} = \alpha_j + \beta_j \widehat{X}_{jt} + \varepsilon_{jt}, \tag{8}$$

where $\widehat{X}_{jt} = \widehat{PCA}_{jt}$ or $\widehat{X}_t = \widehat{SPCA}_{jt}$.

To forecast using the machine learning method, we run the following multiple predictive regressions:

$$Y_{jt} = \hat{\mathbf{Y}}_{\mathbf{j,t}}{}' \boldsymbol{\gamma} + \varepsilon_{jt}, \tag{9}$$

where $\hat{\mathbf{Y}}_{\mathbf{j}}, \mathbf{t} = \left[1 Y^1_{j,t} \dots Y^{N_{j,t-1}}_{j,t} \right]'$ is a $(N_{j,t-1} + 1) \times 1$ vector of predictors and $\boldsymbol{\gamma}$ is the parameter vector. Similar to the principal component, we replace the missing values of $\hat{\mathbf{Y}}_{\mathbf{j}}, \mathbf{t}$ with their cross-sectional average analyst forecast of the same period.

Due to the large number of correlated regressors used in the conventional predictive regressions, the conventional ordinary least squares (OLS) approach may fail to generate stable parameter estimation and result in poor out-of-sample prediction. The recent advance in the machine learning literature has suggested a way to circumvent this estimation problem. We follow the literature and estimate the coefficients by minimizing the following objective function:

$$\left\| \boldsymbol{Y}_j - \hat{\mathbf{Y}}_{\mathbf{j}} \right\|^2 + \lambda \left\| \boldsymbol{\gamma} \right\| + (1 - \lambda) \left\| \boldsymbol{\gamma} \right\|^2, \tag{10}$$

where λ is the regularization parameter corresponding to the Lasso norm (ι_1 penalty term) and $1 - \lambda$ is the weight placed on the ridge norm (ι_2 penalty term). We estimate the coefficients using two different settings. First, we use the Lasso method and set $\lambda = 1$. Second, we use the elastic-net approach proposed by Zou and Hastie (2005). We let λ be data-driven and set the number of folds used in cross-validating λ to 5.

Table 8
Out-of-sample performance of forecasts using machine learning techniques.

	Models					Difference between novel models and BAM			
	PCA	SPCA	E-Net	Lasso	BAM	PCA	SPCA	E-Net	Lasso
MSE	0.0060	0.0060	0.0072	0.0092	0.0052	0.0008	0.0008	0.0020	0.0040
t-value						(0.82)	(0.74)	(1.72)	(1.73)
R^2_{OS} (%)						−16.08%	−16.35%	−39.35%	−78.18%
t-value						(−1.00)	(−0.99)	(−1.73)	(−1.88)

This table compares the MSEs (mean square errors) of BAM forecasts for ROEs with those of other model forecasts including PCA, SPCA, E-Net and Lasso. The difference between other forecasting models and BAM is computed as the average MSEs of other models minus that of BAM and a more negative difference means a greater decrease in forecast errors. The out-of-sample R-squared (R^2_{OS}) statistics are computed to evaluate the out-of-sample performance of the four models relative to BAM. The higher positive R^2_{OS}, the better out-of-sample forecast performance. The t-values that test the significance of mean MSE and R^2_{OS} differences are reported in parentheses and adjusted for firm-clustered standard errors.

We perform adjusted earnings (ROE) forecasts using PCA, SPCA (scaled PCA), Lasso and Elastic-Net (E-Net) and compare their performance with BAM.[16] Table 8 reports these results.[17] The results show that BAM continues to outperform these methods in terms of both MSE and out-of-sample R-squared values. The outperformance over machine learnings methods is significant at the 10 percent level. This finding is surprising as machine learning methods usually work quite well in the stock return forecast literature. However, earnings (ROE) data are different from stock return data in two major aspects, which may explain the results in Table 8. First, unlike high-frequency return data, accounting earnings are low-frequency data (quarterly or annually). Machine learning methods tend to work better for high-frequency data with a large number of data points. Second, earnings forecasts are unbalanced panel data. An analyst typically does not provide update forecasts regularly (uneven intervals). In addition, there are frequent entry and exits of analysts in this profession, which result in unbalanced panel surveys. In the presence of unbalanced panel data, the sample must be trimmed to get a balanced subset of forecasts or the missing observations have to be filled by some ways (in our case, we replace the missing observation by the cross-sectional mean) from which Lasso and E-Net algorithms can operate.[18] This step entails a loss of information relative to the BAM method which is based on the complete set of real analyst forecasts. Under this condition, Capistrán and Timmermann (2009) theoretically demonstrate that the BAM is a more robust forecast method for correcting the bias in survey forecasts. Our results are consistent with Capistrán and Timmermann's theoretical prediction.

6. Additional tests

In this section, we conduct additional tests on the implications of better earnings forecasts for asset pricing. We first examine how the use of a better measure of market expectations for future earnings affects the estimation of earnings response coefficient and the inference on the earnings momentum. Following this, we explore the role of the expected ROEs based on the EW and BAM forecasts in asset pricing of stocks.

6.1. Earnings response coefficient (ERC)

Asset pricing theory suggests that the stock price is the present value of the firm's expected cash flow. Stock prices react to news about the firm's cash flows, or the unexpected component of the earnings release. Quantifying the effects of earnings news requires a reliable measure for earnings expectations. The preceding analysis shows that the regression-based combination model generates more accurate forecasts for future earnings than the consensus forecast. This implies that the regression-based combination earnings forecast provides a better measure for the market's expectation for future earnings. To see whether this is the case, we examine how well different earnings forecasts approximate market expectations.

In the literature, the earnings response coefficient (*ERC*) is commonly used to evaluate how closely a forecast of earnings lines up with the market's expectations. The *ERC* captures the reactions of stock prices to unexpected earnings, which should be greater and more significant for a better proxy for the market's earnings expectations (see Brown et al., 1987; Hou et al., 2012). Thus, a direct way to evaluate forecast models is to compare the *ERCs* based on their earnings forecasts. A larger *ERC* implies that an earnings forecast is a better measure for market expectations of earnings.

To estimate *ERCs*, for each fiscal quarter, we run the cross-sectional regressions of three-day earnings announcement risk-adjusted returns (from day −1 to day +1 where 0 is the day of the earnings announcement) against the quarterly standardized unexpected earnings (SUE) as follows.

$$R_{j,\tau} = a + ERC_k \times SUE^k_{j,\tau} + \varepsilon_{j,\tau}, \tag{11}$$

[16] Note we did not provide PLS forecasts in this analysis because Lin et al. (2018) have shown that PLS is the same as the IMC, which has been examined in this paper. We use standardized variables to process PCA and SPCA is the scaled PCA in Huang et al. (2022).

[17] Similar to the sample for iterated regressions (IMC and IWC), we require each firm has at least 20-quarter ROE records and at least six analysts' following. However, the final sample of out-of-sample observations is different. Because IMC and IWC requires two time-series regressions to compute the adjusted alpha and beta, one at individual analyst level and another at the firm level, many observations used for computing are not included in the out-of-sample results. By contrast, there is only one time-series estimation in Table 8 and its sample size is larger than Table 1 Panel B with different BAM results.

[18] There has been new development in data science on how to run machine learning regressions of unbalanced data. It will be interesting to investigate how these new methods perform in future research.

where k stands for different models (EW, BAM, BAM*), $SUE^{k}_{j,\tau}$ is actual EPS minus model-k forecasted EPS for firm j at τ scaled by standard deviations of actual EPS in the last eight quarters of firm j. $R_{j,\tau}$ is the risk-adjusted return, which is the stock return adjusted for risk factors.[19] Following Fama and French (2008), we sort firms independently into 5 × 5 portfolios by market capitalization (size) and the book-to-market ratio at the end of June in each year n. Value-weighted monthly portfolio returns are calculated for each of the 25 portfolios from July of year n to June of year $n + 1$. Adjusted returns are stock returns from CRSP in excess of the matched portfolio returns.

We choose the BAM earnings forecasts to represent the regression-based forecast as it permits the largest number of firms among all regression combination models. This choice minimizes the sample selection bias[20] and generates more representative results with statistical power. Panel A in Table 9 reports the results for stock price reactions to unexpected earnings. On average, the *ERC* associated with the BAM earnings forecast exceeds that with the consensus forecast (EW) by 0.467 (38%), which is significant at the one percent level ($t = 4.643$). The median difference between the *ERCs* of BAM and EW is larger, which is 0.628 (58%). The Mood's median test shows that the difference in the median *ERCs* is significantly different at the five percent level (p value in parentheses). The results suggest that the BAM forecast is a better measure for the market's earnings expectations than the consensus forecast.

The results are even stronger when we use the BAM* earnings forecast, which consider firm-level and aggregate forecasting information to improve analyst forecasts. The last three columns show that the *ERC* associated with the BAM* forecast surpasses that of BAM, which combine only analyst forecasts. The mean (median) difference in the *ERC* between BAM* and BAM is 0.298 (0.243) or 18% (14%) which is significant at the five percent level or above. The results again show the benefits of using additional information to improve the BAM forecast. Taken together, there is evidence that the performance of the regression-based combination forecast can be enhanced by incorporating other available information than analyst forecasts.

6.2. Return predictability of unexpected earnings

One of the most pronounced anomalies uncovered in asset pricing tests is that unexpected earnings can predict future returns (Bernard and Thomas, 1989, 1990), which is dubbed earnings momentum in the literature. If an earnings forecast is an unbiased measure for the market's expectations of earnings, the difference between actual and forecasted earnings reflects true new information, and the stock price will fully respond to the new information immediately, provided that the market is efficient. Under this condition, unexpected earnings should have no predictive power for future stock returns. However, if a biased earnings forecast is used as a measure for the market's expectation for earnings, the difference between the actual earnings and biased forecasted earnings would contain fundamental information (bias) which can predict future stock returns. For example, if analysts underestimate earnings, the difference between actual and analyst forecast earnings (or earnings surprise) would be an inaccurate measure of the true earnings shock and may therefore induce a spurious correlation with future stock returns as the surprise component contains information for true earnings.

To elaborate this point, we use a simple model to provide an intuitive explanation. The literature has shown that analysts often underreact to earnings news. Suppose that econometricians update their earnings forecast Y_t^a in a way similar to analysts. Then, we can express this earnings forecast process as

$$Y_t^a = Y_{t-1}^a + \lambda(Y_{t-1} - Y_{t-1}^a) + e_t, \tag{12}$$

where Y_{t-1} is the actually earnings reported at t-1, and $0 < \lambda < 1$ captures the underreaction to earnings news. For convenience, we drop the subscript for individual forecaster i. The above specification resembles the familiar partial adjustment process, which is used to capture the phenomenon that analysts underreact to the earnings news. When there is an earnings surprise, the projected earnings are revised only partially to the difference between the actual and forecasted earnings by analysts due to underreaction. The value of λ reflects the extent of inertia, a higher value of which indicates a more speedy adjustment or less underreaction to the earnings news.

Assume that the stock price responds to the unexpected earnings based on the true earnings process Y_t. Given an earnings response coefficient (*ERC*) π, the resulting abnormal returns (*AR*) can be expressed as

$$AR_t = \pi(Y_t - E(Y_t)) \tag{13}$$

where $E(Y_t)$ is the true expected earnings. Using (12), we can rewrite (10) as

$$\begin{aligned} AR_t &= \pi[(Y_t - E(Y_t^a)) + (E(Y_t^a) - E(Y_t))] \\ &= \pi(Y_t - E(Y_t^a)) + \pi(Y_{t-1}^a - E(Y_t)) + \lambda\pi(Y_{t-1} - Y_{t-1}^a) \end{aligned} \tag{14}$$

where $E(Y_t^a)$ is the consensus forecast (EW) for earnings in period t and $Y_t - E(Y_t^a)$ is the earnings surprise based on the consensus forecast. The above equation shows that when there is an earnings surprise at t-1 against the analyst (consensus) forecast Y_{t-1}^a, it will affect the abnormal stock return of earnings announcement at time t due to the analyst underreaction. This effect depends on the coefficient π or the response to earnings news, and the earnings forecast error.

The specification in (12) posits that the econometrician considers only the earnings surprise in the last period to adjust the earnings forecast. More generally, a forecaster may consider the earnings forecast errors in multiple lags to adjust the earnings

[19] We report the results of *ERC* and *SUE* using value-weighted adjusted returns of portfolios. Results are robust to using equal-weighted portfolio returns.

[20] Because the two-step regression methods require sufficient analyst forecasts in the first step time-series regression, the sample tends to cover the well-established firms. Small firms and financially distressed firms are underrepresented.

Table 9
Earnings response coefficients (*ERCs*) and earnings momentum.

Panel A: Earnings response coefficients

	ERC_{EW}	ERC_{BAM}	ERC_{BAM*}	BAM - EW			BAM* −BAM		
				Diff	Diff (%)	Test stat	Diff	Diff (%)	Test stat
Mean	1.220	1.687	1.985	0.467	38.265%	(4.643)	0.298	17.645%	(2.631)
Median	1.069	1.697	1.940	0.628	58.723%	(0.028)	0.243	14.348%	(0.032)

Pane B: Return predictability of standardized unexpected earnings

	Future return								
	3 months			6 months			12 months		
βSUE_{EW}	0.486			0.617			0.634		
	(4.835)			(3.632)			(2.995)		
βSUE_{BAM}		0.162			0.187			0.188	
		(1.524)			(1.034)			(0.638)	
βSUE_{BAM*}			0.090			0.063			0.168
			(0.503)			(0.130)			(0.300)

Panel C: Earnings response coefficients using ROE as a scaled earnings measure

	ERC_{EW}	ERC_{BAM}	ERC_{BAM*}	BAM - EW			BAM* −BAM		
				Diff	Diff (%)	Test stat	Diff	Diff (%)	Test stat
Mean	1.463	1.639	1.815	0.176	12.061%	(2.151)	0.175	10.704%	(2.046)
Median	1.346	1.735	1.837	0.390	28.969%	(0.011)	0.101	5.842%	(0.032)

Pane D: Return predictability of the unexpected ROE

	Future return								
	3 months			6 months			12 months		
βSUE_{EW}	0.421			0.513			0.768		
	(3.975)			(3.634)			(3.241)		
βSUE_{BAM}		0.258			0.282			0.286	
		(1.644)			(1.034)			(0.477)	
βSUE_{BAM*}			0.136			0.109			0.005
			(1.002)			(0.130)			(0.300)

Panel A summarizes the *ERCs* for the full sample. *ERCs* are the coefficient of the cross-sectional regression of the three-day earnings announcement (risk-adjusted) returns against the standardized unexpected earnings (SUE). *SUE* is measured as the difference between actual and forecasted earnings scaled by the standard deviation of actual earnings over last eight quarters. The first two rows report the mean and median *ERCs* over fiscal quarters. Mean difference test t-statistics and median difference test p-values are in parentheses. For the median difference, Mood's median test is applied. Panel B reports the average coefficients of SUE over fiscal quarters. In a given fiscal quarter, we run the cross-sectional regression of holding period stock returns over next 3, 6, and 12 months (skip the month of earnings announcement) on *SUE*. SUE_{EW}, SUE_{BAM} and SUE_{BAM*} denote the *SUE* measures using EW, BAM and BAM* forecasts, respectively. Besides analyst forecasts, BAM* considers firm-level accounting/financial and aggregate information to generate the optimal earnings forecast. Fama–MacBeth t-statistics are in parentheses. The sample period is from 1985 Q3 to 2016 Q1 for both panels. Panels C and D repeat the analysis using ROE as an alternative measure of scaled earnings and show similar results.

forecast. In such case, (14) will include more lagged earnings surprises, i.e., t-2, t-3, …, t-k. On the flip side, this suggests that the past earnings surprise can predict future stock returns over various horizons.

Next, suppose that the BAM forecast, Y_t^{BAM} is an unbiased estimator for the true earnings or $E(Y_t) = Y_t^{BAM}$. Then, Eq. (13) can be rewritten as

$$AR_t = \pi(Y_t - Y_t^{BAM}) = \pi\varepsilon_t \tag{15}$$

where ε_t is the unexpected earnings based on the BAM forecast. Since the BAM forecast is an unbiased and efficient forecast, the lagged forecast error term will not show up in (14). In an efficient market, stock prices will respond fully to the unbiased earnings surprise and thus the unexpected earnings ε_t will have no relation with future stock returns.

The above analysis assumes that the regressed combination forecast is an unbiased measure for the true earnings and the market is perfectly efficient. Any violation of these two assumptions can lead to a predictive relation between unexpected earnings and future stock returns. For example, market participants may underreact to the true earnings surprise (or the market is inefficient). As the stock price gradually adjusts to the new information, this underreaction problem will be corrected. This adjustment process will induce a predictive relation between current earnings surprise and future stock returns, or there is earnings momentum. Alternatively, if the regressed combination forecast is not an unbiased measure of the true earnings or is subject to a similar lagged adjustment problem of earnings forecast revision as portrayed in (12), the difference between the actual earnings and the BAM forecast will no longer represent the true earnings surprise. In such case, the past errors of the regressed combination forecast or its unexpected earnings component would show up in (14) and empirically we may still find a positive relation between the current earnings surprise (measured against the BAM forecast) and future stock returns.

In general, if the regressed combination forecast is a better measure for the market's expectations of earnings than the consensus forecast, the former's unexpected earnings component will have less predictive power for future returns because its earnings proxy error is smaller. To test this hypothesis, we can examine the relation between the standardized unexpected earnings (*SUE*) and future stock returns. We run the cross-sectional regression of holding period stock returns over the next 3-, 6-, and 12-month horizons (skip the month of earnings announcement) on the firm's current *SUE* as follows.

$$HPR_{j,t+1,t+1+n} = \alpha + \beta SUE_{k,t} \times SUE^{k}_{j,t} + \epsilon_{j,t}, \tag{16}$$

where $HPR_{j,t+1,t+1+n}$ is the holding period return from month $t+1$ to $t+1+$ n ($n = 3, 6, 12$) for firm *j*, adjusted for risk factors as the return measure in (11), *k* stands for different models (EW, BAM and BAM*), and $SUE^{k}_{j,t}$ is actual EPS minus model-*k* forecasted EPS for firm *j* at *t* scaled by standard deviations of actual EPS in the last eight quarters of firm *j*.

If the BAM method produces a better measure for the expected earnings than the consensus (EW) forecast, we should observe a less significant coefficient of *SUE* measured by the former forecast in the cross-sectional predictive regression. For comparative purposes, we consider both BAM and BAM* forecasts, where the latter considers other forecasting information in addition to analyst forecasts to improve earnings forecasts. Likewise, if the BAM* forecast is a more efficient measure of market's earnings expectations, we expect that the coefficient of *SUE* based on the BAM* earnings forecast will be even weaker.

Panel B of Table 9 reports the average cross-sectional estimates of βSUE_{EW}, βSUE_{BAM} and βSUE_{BAM*} over time, respectively. Regardless of the length of future holding periods (3 to 12 months), βSUE_{EW} is significantly positive in all cases, suggesting that SUE_{EW} has predictive power for future stock returns. This result is consistent with the findings in the literature, which are typically based on the consensus forecast (EW) or its variants to calculate the earnings surprise. In contrast, the coefficients of SUE_{BAM} fail to be significant. This finding is consistent with the contention that the BAM forecast provides a more accurate measure of earnings expectations. The results are even stronger when we use the BAM* forecast as the measure for the expected earnings. The coefficients of SUE_{BAM*} are all statistically and economically insignificant. The results show that earnings momentum weakens significantly when using a better measure for the expected earnings. This finding suggests that the choice of an appropriate measure for the market's expectations of future earnings has a significant impact on inferences about the cross-section of expected stock returns and the post-earnings-announcement drifts.

For robustness, we also perform the ERC analysis and return predictability using the ROE as an alternative measure for scaled earnings. The results are reported in Panels C and D, which show a similar pattern as those based on standardized earnings. Thus, our findings are robust to different unexpected earnings measures.

6.3. Portfolio sorts on the realized and expected ROEs

Recent research (Fama and French, 2015; Hou et al., 2015) suggests that profitability is an important pricing factor for stocks, and the realized ROE is commonly used as a proxy for a firm's future profitability. An issue in the asset pricing test with a profitability factor is that the accounting-based ROE is released at a low frequency, i.e., quarterly or even annually. This implies that researchers may have to use the information with a three-month lag or longer to predict future returns in the cross-section. Given our finding that analyst forecasts convey future earnings information, a natural question is whether the expected ROE based on updated analyst earnings forecasts contains additional important information for asset pricing before the next earnings announcement and therefore, has higher predictive power for future stock returns than the profitability factor constructed from the realized ROE.

To investigate this possibility, we employ an approach similar to Hou et al. (2015) to evaluate the predict power of the expected ROE. Following Hou et al. (2015), we construct 2-by-3-by-3 portfolios based on size, investments-to-assets (I/A) and profitability measures. At the end of June of year *n*, we use the median NYSE firm size (stock price per share times by the number of outstanding shares from CRSP) to split stocks into two size groups for the period from July of year *n* to June of year $n + 1$. Similarly and independently, at the end of June of year *n*, we use the NYSE cutoffs for the low 30%, middle 40%, and top 30% of I/A (annual changes in total assets divided by previous total assets) to split stocks into three groups for the period from July of year *n* to June of year $n + 1$. Then, independently, at the beginning of each month, we divide stocks into three groups based on NYSE cutoffs for the low 30%, middle 40%, and top 30% of the profitability measures. Finally, we compute the value-weighted returns for $2 \times 3 \times 3 = 18$ portfolios each month.[21]

We use multiple profitability measures to evaluate their predictive power for future stock returns. The first measure is the most recent reported accounting ROE, and the second and third measures are the expected ROEs constructed from the monthly updated EW and BAM forecasts, respectively.[22] We form test portfolios based on these ROE measures. As an example, assume that a firm's quarterly earnings in year *n* Q2 is announced in August of year *n*. We skip one month to form portfolios based on the three profitability measures starting from October.[23] At the beginning of October, we sort portfolios on the latest actual ROE announced in August, and the expected ROEs based on EW and BAM forecasts in September for future earnings. At the beginning of November, we form portfolios still using the same actual ROE, but use the updated EW and BAM forecasts in October to form portfolios based on the expected (or predicted) ROE. Similarly, in December, we form portfolios using the same actual ROE and the updated EW and BAM forecasts for ROEs in November if the actual earnings of year *n* Q3 are not yet announced. We repeat this procedure throughout the sample period.

21 We verify our procedure using the data of Hou et al. (2015) and find that our procedure generates average return for the ROE factor very close to theirs, suggesting that our procedure is accurate.

22 To compute monthly BAM from monthly EW, we use the most recent alpha and beta from quarterly regression.

23 This is because right after the earnings announcement, the actual/realized ROE from the newly announced earnings will surely contain the best information and it is not meaningful to compare it with the ROEs based on analyst forecasts.

Table 10
Portfolio sorts using the realized and expected ROEs.

	25P	50P	Mean	75P
Actual ROE				
Low	−2.362	1.812	1.817	5.829
High	−1.269	1.874	2.205	5.131
Diff			0.388	
t-stat			(2.39)	
Monthly updated EW				
Low	−2.469	1.745	1.876	6.037
High	−1.407	1.842	2.352	5.236
Diff			0.475	
t-stat			(3.20)	
Monthly updated BAM				
Low	−2.698	1.716	1.812	5.994
High	−1.323	1.909	2.368	5.276
Diff			0.557	
t-stat			(3.57)	
Difference-to-difference test				
$\text{Diff}_{\text{EW}} - \text{Diff}_{\text{actual ROE}}$	−1.945	−0.012	0.087	2.026
t-stat			(2.05)	
$\text{Diff}_{\text{BAM}} - \text{Diff}_{\text{EW}}$	−0.704	0.061	0.081	0.880
t-stat			(1.77)	

This table reports the distribution of value-weighted portfolio returns in percentage for groups with high and low profitability and its difference. Following Hou et al. (2015), we construct 2-by-3-by-3 portfolios based on size, investments-to-assets (I/A) and profitability measures. At the end of June of year n, we use the median NYSE size (stock price per share times by the number of outstanding share from CRSP) to split stocks into two groups for the period from July of year n to June of year $n+1$. Similarly and independently, at the end of June of year n, we use the NYSE cutoffs for the low 30%, middle 40%, and top 30% of I/A (annual changes in total assets divided by previous total assets) to split stocks into three groups for the period from July of year n to June of year $n+1$. Then, independently, at the beginning of each month, we divide stocks into three groups based on NYSE cutoffs for the low 30%, middle 40%, and top 30% of profitability measures. Following this, we compute the value weighted returns for 18 portfolios at each month and report the distribution of portfolios with low and high profitability. Profitability measures include the most recently available actual ROE, and the monthly updated EW and BAM ROE forecasts up to three months following the last earnings announcement date. The larger and more positive portfolio return spreads (Diff = High − Low) indicate better predictability. Mean difference t-statistics are in parentheses.

We then compare the return prediction using the actual (realized) ROE with those using the expected ROEs based on EW and BAM forecasts over the three-month future horizon. The realized ROE remains the same for three months once earnings are announced but the predicted ROEs change when analysts update their forecasts. Therefore, we expect the predicted ROEs based on EW and BAM updated forecasts will on average have higher predictive power in the cross-section for future stock returns than the actual ROE which may become stale as time moves on. Moreover, if the BAM earnings forecast is a better proxy for expected earnings, the BAM ROE forecast should have higher predictive power in the cross-section for future returns than the EW ROE forecast.

Our results in Table 10 are consistent with this expectation. As shown in the bottom panel, the average return spread of portfolios formed by the EW expected ROE over the three-month horizon is 0.087% ($t = 2.05$) larger than those formed by the realized ROE. The results strongly suggest that analyst earnings forecasts contain valuable information for expected future stock returns. Furthermore, the average return spread of portfolios formed by the BAM expected ROE is 0.081% ($t = 1.77$) larger than the return spread of portfolios formed by the EW expected ROE. Consistent with the finding of the superiority of BAM earnings forecasts, the BAM expected ROE has significantly higher predictive power in the cross-section for future stock returns than the EW expected ROE. The results show that the BAM delivers the highest predictive power for future stock returns as it improves the predictive information content in the EW forecast and historical ROE for future earnings.

Overall, there is clear evidence that the bias-adjusted mean method provides a much better measure of the market's expectations for future earnings. Using the ROE forecast by the BAM method generates a higher stock return predictability than the forecast by the EW method and the realized ROE. The results strongly suggest that analyst earnings forecasts contain important information for future stock returns, and the BAM combination forecast method extracts better expected earnings information from analyst forecasts, thereby further improving the predictive power of the profitability factor in the cross-section for future stock returns.

7. Conclusion

Forecasting future earnings is an issue of fundamental importance to both academics and practitioners in accounting and finance. Earnings forecasts are useful for asset pricing as they provide information for the expectations of future earnings and security valuation. A vast literature has been devoted to studying the methods and factors that may affect forecast accuracy and efficiency. The enormous amount of research efforts in the past several decades have expanded our understanding for the roles of models and analyst forecasts in improving earnings forecast performance and the strength and weakness of these forecasts.

This paper contributes to the literature by employing a combination approach to forecast earnings and evaluating the efficacy of different methods to obtain the optimal earnings forecast. Analyst forecasts are biased and noisy, the earnings generating process is constantly evolving, and the information environment is often uncertain and unstable. We argue that in these circumstances forecast combination methods can avoid excessively noisy forecasts and reduce the uncertainty and instability risk associated with an individual analyst.

Empirical evidence shows that forecast combination of analyst forecasts provides substantially better earnings forecasts than other methods. In particular, we find the regression-based combination methods produce better forecasts of future earnings than the consensus forecast commonly used in the accounting and finance literature. The single-step regression bias-adjusted mean method does a remarkable job to correct the bias and instability of forecasts by individual analysts and the two-step regression methods further improve the out-of-sample forecast performance. While the two-step regression methods generally produce better forecast results than the single-step regression adjustment method, the latter has an advantage of lowering the time-series data requirement and therefore can be applied to a large sample of firms, including young and small firms with shorter earnings records and fewer analysts.

We examine the stock prices response to unexpected earnings estimated by the regressed combination forecast and consensus methods. We find that the earnings response coefficient is much larger for the unexpected earnings measure based on the regressed combination forecast. This finding provides direct evidence that the regressed combination forecast is a better measure for the latent expectations of future earnings. Further, we find that using this better proxy for expected earnings significantly weakens the anomaly of post-earnings-announcement drift. More importantly, we find that the expected ROE based on analyst forecasts improves asset pricing performance of the multifactor model that includes a profitability factor, and this improvement is much stronger when using the BAM method to forecast ROE.

Overall, we find that the regression-based forecast combination methods show promise of reducing the bias and improving the accuracy of earnings forecasts. Regression-based combination forecasts outperform other forecasts by significant margins on a reasonably consistent basis over time and across firms of different characteristics. The regression-based bias-adjusted mean method has distinctive advantages of providing efficient forecasts in the presence of unbalanced panels characterized by irregularly spaced surveys or forecasts with entry and exits of analysts. Moreover, combining analyst forecasts with firm-level and macroeconomic information further improves earnings forecasts. These findings suggest that the regressed combination approach will be a useful strategy for earnings forecasts in future research.

CRediT authorship contribution statement

Hai Lin: Methodology, Software, Validation. **Xinyuan Tao:** Conceptualization, Methodology, Formal analysis, Writing – original draft. **Chunchi Wu:** Conceptualization, Methodology, Writing – review & editing, Supervision.

Appendix A

In this appendix, we demonstrate that theoretically the IMC will perform better than the BAM in that the MSE of the iterated mean regression combination will never be larger than that of the bias-adjustment method, given an equal number of analysts and the same distribution of analyst forecasts. We demonstrate this property under Experiment 2 with the assumption that analyst forecasts are heterogeneous and time-varying. Our results also hold under Experiment 1. The main difference between Experiments 1 and 2 is that in the latter, the weights are not sum to one, and under this condition, the equal weighted average combination is no longer optimal (see the Online Appendix for detail).

For simplicity and without loss of generality, we assume that there are only two analysts ($N = 2$) and the random effect (dispersion and time variations of forecasts) only occurs to the factor loadings of the second analyst ($\gamma_2 \sim N\left(\mu_2, \sigma_2^2\right)$) but not to the first analyst ($\gamma_1 \sim N(1,0)$). In this case, conditions (a)–(g) of Experiment 2 are satisfied and Eqs. (1)–(3) in the Online Appendix become

$$Y_t = \boldsymbol{\iota}' F_t + \epsilon_{y,t}, \epsilon_{y,t} \sim N(0,1), \tag{A.1}$$

$$\begin{cases} \hat{Y}_t^1 = \frac{1}{2}\boldsymbol{\iota}' F_t + \epsilon_{1,t}, \epsilon_{1,t} \sim N\left(0, \frac{1}{2}\right) \\ \hat{Y}_t^2 = \gamma_2 \frac{1}{2}\boldsymbol{\iota}' F_t + \epsilon_{2,t}, \epsilon_{2,t} \sim N\left(0, \frac{1}{2}\right) \end{cases}, \tag{A.2}$$

$$F_t = \epsilon_{F_t}, \epsilon_{F_t} \sim N\left(0, D_{\epsilon_F}\right). \tag{A.3}$$

The variance and covariance are

$$var\left(\hat{Y}_t^1\right) = var\left(\frac{1}{2}\boldsymbol{\iota}' F_t + \epsilon_{1,t}\right) = \frac{1}{4} var\left(\boldsymbol{\iota}' F_t\right) + var\left(\epsilon_{1,t}\right) = \frac{1}{4}(1+1) + \frac{1}{2} = 1, \tag{A.4}$$

$$var\left(\hat{Y}_t^2\right) = var\left(\gamma_2 \frac{1}{2}\boldsymbol{\iota}' F_t + \epsilon_{2,t}\right) = \gamma_2{}^2 \frac{1}{4} var\left(\boldsymbol{\iota}' F_t\right) + var\left(\epsilon_{2,t}\right) = \gamma_2{}^2 \frac{1}{4}(1+1) + \frac{1}{2} = \frac{1}{2}(\gamma_2{}^2 + 1), \tag{A.5}$$

$$cov\left(Y_t, \hat{Y}_t^1\right) = cov\left(\boldsymbol{\iota}' F_t + \epsilon_{y,t}, \frac{1}{2}\boldsymbol{\iota}' F_t + \epsilon_{1,t}\right) = \frac{1}{2} var\left(\boldsymbol{\iota}' F_t\right) = 1, \tag{A.6}$$

$$cov\left(Y_t, \hat{Y}_t^2\right) = cov\left(\boldsymbol{\iota}' F_t + \epsilon_{y,t}, \gamma_2 \frac{1}{2}\boldsymbol{\iota}' F_t + \epsilon_{2,t}\right) = \gamma_2 \frac{1}{2} var\left(\boldsymbol{\iota}' F_t\right) = \gamma_2, \tag{A.7}$$

$$cov\left(\hat{Y}_t^1, \hat{Y}_t^2\right) = cov\left(\frac{1}{2}\boldsymbol{\iota}' F_t + \epsilon_{1,t}, \gamma_2 \frac{1}{2}\boldsymbol{\iota}' F_t + \epsilon_{2,t}\right) = \gamma_2 \frac{1}{4} 2 = \frac{1}{2}\gamma_2 \tag{A.8}$$

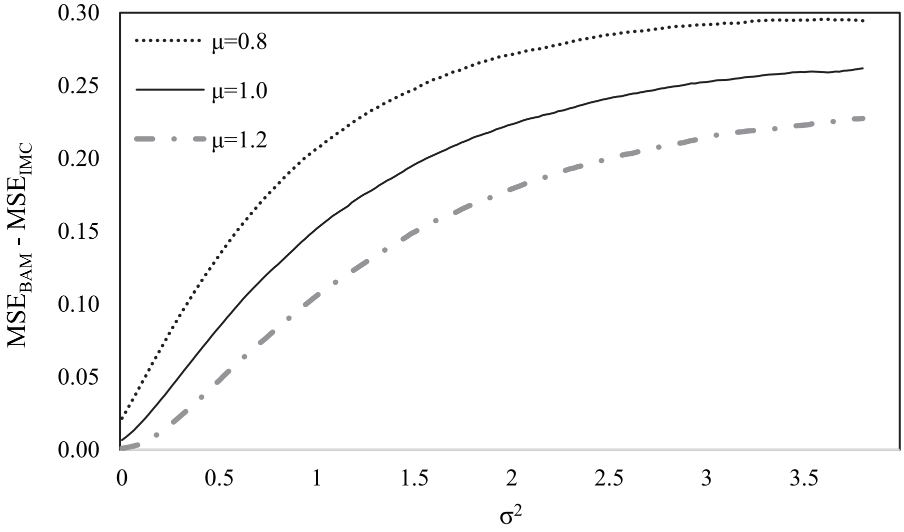

Fig. A.1. The performance of IMC versus BAM.

For the BAM, we regress the true earnings on the average forecast value,

$$Y_t = \alpha + \beta \overline{Y}_t + \varepsilon_t = \alpha + \beta \frac{1}{2}\left(\hat{Y}_t^1 + \hat{Y}_t^2\right) + \varepsilon_t^{BAM}$$

and the MSE is

$$\begin{aligned} MSE_{BAM} &= E\left[\left(Y_t - \alpha - \beta\frac{1}{2}\left(\hat{Y}_t^1 + \hat{Y}_t^2\right)\right)^2\right] = \left[E\left(\varepsilon_t^{BAM}\right)\right]^2 + var\left(\varepsilon_t^{BAM}\right) = var\left(\varepsilon_t^{BAM}\right) \\ &= var\left(Y_t - \alpha - \beta\frac{1}{2}\left(\hat{Y}_t^1 + \hat{Y}_t^2\right)\right) where\ \beta = \frac{cov\left(Y_t, \frac{1}{2}\left(\hat{Y}_t^1 + \hat{Y}_t^2\right)\right)}{var\left(\frac{1}{2}\left(\hat{Y}_t^1 + \hat{Y}_t^2\right)\right)} \\ &= var\left(Y_t\right) - \frac{\left(1+\gamma_2\right)^2}{1 + \frac{1}{2}\left(1+\gamma_2\right)^2}. \end{aligned} \tag{A.9}$$

For the IMC, we first run a time-series regression for the first analyst,[24]

$$Y_t = a_1 + b_1\hat{Y}_t^1 + \varepsilon_t^1 \text{ where } b_1 = \frac{cov\left(Y_t, \hat{Y}_t^1\right)}{var\left(\hat{Y}_t^1\right)} = 1 \text{ and } a_1 = 0$$

to get the first adjusted individual forecast $\tilde{Y}_t^1 = \hat{Y}_t^1$, and for the second analyst,

$$Y_t = a_2 + b_2\hat{Y}_t^2 + \varepsilon_t^2 \text{ where } b_2 = \frac{cov\left(Y_t, \hat{Y}_t^2\right)}{var\left(\hat{Y}_t^2\right)} = \frac{2\gamma_2}{\gamma_2^2+1}, \text{ and } a_2 = 0$$

to get the second adjusted individual forecast $\tilde{Y}_t^2 = \frac{2\gamma_2}{\gamma_2^2+1}\hat{Y}_t^2$. After the adjustment, the mean forecast $\overline{Y}_t$ becomes $\frac{1}{2}\left(\tilde{Y}_t^1 + \tilde{Y}_t^2\right) = \frac{1}{2}\left(\hat{Y}_t^1 + \frac{2\gamma_2}{\gamma_2^2+1}\hat{Y}_t^2\right)$, which is different from that for the BAM.

We then regress the true earnings on the mean forecast and calculate MSE

$$\begin{aligned} MSE_{IMC} &= E\left[\left(Y_t - \alpha - \beta\frac{1}{2}\left(\hat{Y}_t^1 + \frac{2\gamma_2}{\gamma_2^2+1}\hat{Y}_t^2\right)\right)^2\right] = var\left(\varepsilon_t^{IMC}\right) \\ &= var\left(Y_t - \alpha - \beta\frac{1}{2}\left(\hat{Y}_t^1 + \frac{2\gamma_2}{\gamma_2^2+1}\hat{Y}_t^2\right)\right) where\ \beta = \frac{cov\left(Y_t, \frac{1}{2}\left(\hat{Y}_t^1 + \frac{2\gamma_2}{\gamma_2^2+1}\hat{Y}_t^2\right)\right)}{var\left(\frac{1}{2}\left(\hat{Y}_t^1 + \frac{2\gamma_2}{\gamma_2^2+1}\hat{Y}_t^2\right)\right)} \\ &= var\left(Y_t\right) - \frac{\frac{1}{4}\left(1 + \frac{2\gamma_2^2}{\gamma_2^2+1}\right)^2}{\frac{1}{4} + \frac{\gamma_2^2}{\gamma_2^2+1}}. \end{aligned} \tag{A.10}$$

[24] From (A.1) and (A.3), we have $\mathrm{E}\left[Y_t\right] = 0$ and from (A.2) and (A.3), we have $\mathrm{E}\left[\hat{Y}_t^1\right] = 0$ and $\mathrm{E}\left[\hat{Y}_t^2\right] = 0$. Therefore, $\mathrm{E}\left[Y_t\right] = E\left[a_1 + b_1\hat{Y}_t^1 + \varepsilon_t^1\right] = a_1 + b_1\mathrm{E}\left[\hat{Y}_t^1\right] = 0 \Rightarrow a_1 = 0$. Similarly, we can show that $a_2 = 0$.

Table B.1
Description of variables in subsample analysis..

Category	Symbol	Name of variable	Variable description
Properties of analyst forecasts and earnings	*N*	Number of analysts	Number of analysts who post EPS estimates for a firm during a given quarter
	DISP	Forecast dispersion	Cross-sectional standard deviation of the most recently revised EPS estimates among analysts for a firm during a given quarter, scaled by the firm's stock price measured at the last month preceding the month of actual EPS announcement
	NREV	Number of forecast revisions	Average number of the revision by a given analyst for a firm during a given quarter
	REV	Magnitude of revision	Absolute change in the mean value of analyst forecasts during the last two months preceding the month of actual EPS announcement
	VOLEPS	Volatility of EPS	Time-series of standard deviation of actual EPS over last 8 quarters, scaled by stock price. Stock price is measured at the last month preceding the month of actual EPS announcement
	VOLERR	Volatility of forecast errors	Time-series of standard deviation of forecast errors over the last 8 quarters. Forecast errors are defined as the absolute differences between average analyst forecasts and actual EPS
	SUR	Surprise in EPS	Absolute differences between average analyst forecasts and actual EPS of the last forecast period scaled by stock price
Firm characteristics	*SIZE*	Firm size	Natural logarithm of stock price times the total number of outstanding shares divided by 1000. Price and shares are values in the preceding month of actual EPS announcement
	BM	Book-to-market ratio	Book value of equity divided by market value of equity
	NREALIZ	Number of EPS realization	Number of years since the first actual quarterly EPS is announced in I/B/E/S
	AGE	Firm age	Number of years since the first appearing on Compustat
	DIV^d	Dividend dummy	$\mathrm{DIV^d} = 1$ if the firm in a given quarter has dividend information and dividend is greater than zero, otherwise $\mathrm{DIV^d} = 0$
	DIV2	Dividend	Total dividend scaled by firm size
	MGT^d	Management forecasts dummy	$\mathrm{MGT^d} = 1$ if the firm at a given quarter has management forecasts, otherwise $\mathrm{MGT^d} = 0$
	IO	Institutional ownership	Institutional share holdings
	AQ1	Discretionary accruals	Estimated from the Jones (1991) model. High discretionary accruals are associated with low quality of earnings
	AQ2		Estimated from the Kothari et al. (2005) model
Industry	*INDS*	Firm industry	Industry category is based on the first 2-digit SIC code 1 = Agriculture, Forestry & Fishing 2 = Mining 3 = Construction 4 = Manufacturing 5 = Transportation & Public Utilities 6 = Wholesale Trade 7 = Retail Trade 8 = Finance, Insurance & Real Estate 9 = Service 10 = Others (Nonclassifiable)
Market-wide and economic regimes	*VIX*	CBOE Volatility Index	The implied volatility of S&P 500 index options, from Yahoo Finance
	CAPE	Cyclically adjusted price-to-earnings ratio	The price divided by the average of ten years of earnings (moving average), adjusted for inflation, from Robert J. Shiller's online data
	GDP1	Real GDP growth rate	Annual GDP growth rates from Bureau of Economic Analysis (BEA)
	GDP2	Real GDP growth rate	Mean of quarterly annualized GDP growth rates from Bureau of Economic Analysis (BEA)
	INV	Investor sentiment	From Baker and Wurgler (2006)
	MGT	Manager sentiment	From Jiang et al. (2019)

It follows that the difference in MSEs between BAM and IMC is

$$MSE_{BAM} - MSE_{IMC} = -\frac{\left(1+\gamma_2\right)^2}{1+\frac{1}{2}\left(1+\gamma_2\right)^2} + \frac{\frac{1}{4}\left(1+\frac{2\gamma_2{}^2}{\gamma_2{}^2+1}\right)^2}{\frac{1}{4}+\frac{\gamma_2{}^2}{\gamma_2{}^2+1}}. \tag{A.11}$$

Table B.2
Description of variables used in the adjusted earnings forecasts..

Index, k	Description	References
Analyst characteristics		
PAFE	Difference between the absolute forecast error for analyst i for firm j at time t and the mean absolute forecast error for firm j at time t scaled by the mean absolute forecast error for firm j at time t	Clement (1999) Bradley et al. (2017)
Z_1	The age of analyst i's forecast minus the age of the average analyst's forecast following firm j at time t, where age is the age of the forecast in days at the minimum forecast horizon date	
Z_2	The number of years (including time t) that analyst i appeared in the data set minus the average number of years analysts following firm j at time t appeared in the data set	
Z_3	The number of years (including time t) that analyst i supplied a forecast for firm j minus the average number of years analysts following firm j had supplied forecasts	
Z_4	The number of companies followed by analyst i at time t minus the average number of companies followed by an analyst following firm j at time t	
Z_5	The number of 2 digit SICs followed by analyst i at time t minus the average number of two-digit SICs followed by an analyst following firm j at time t	
Quarterly firm-level accounting variables		
FV_1	Inventory, Δ Inventory (INVTQ) - Δ Revenue (REVTQ)	Ball and Ghysels (2018)
FV_2	Account receivable, Δ Account receivable (RECTQ) - Δ Revenue	
FV_3	Capital expenditure, Δ Industry CAPX - Δ Firm CAPX (CAPXQ)	
FV_4	Gross margin, Δ Revenue - Δ Gross margin (REVTQ – COGSQ)	
FV_5	SG&A expenses, ΔSG&A (XSGAQ) - Δ Revenue	
FV_6	Total assets (ATQ)	Fama and French (2006) Hou et al. (2012)
FV_7	Dividend payment (DVY)	
FV_8	Divd = 1 if the firm at a given quarter has dividend information and dividend payment is greater than zero, otherwise 0	
FV_9	Net income (IBQ)	
FV_{10}	negNI = 1 if the firm at a given quarter has negative earnings, otherwise 0	
FV_{11}	Accruals, Δ ACTQ - Δ CHEQ - Δ LCTQ + Δ DLCQ – DPQ (before 1988); IBQ - OANCFQ (since 1988)	
FV_{12}	Investment, ΔATQ	
FV_{13}	BM, log (CEQ/(PRCCQ × CSHOQ))	
FV_{14}	ME, log (PRCCQ × CSHOQ)	
Firm-level stock market variables		
FV_{15}	Firm-specific stock return from CRSP during month m less the same-industry portfolio return in month m before actual earnings announcement t	Ball and Ghysels (2018)
FV_{16}	Average of squared daily firm-level stock returns from CRSP during month m before actual earnings announcement t	
FV_{17}	Most recent daily close price before the date of earnings announcement t from CRSP	
Monthly macroeconomic variables		
FV_{18}	Year-over-year growth rate of seasonally adjusted monthly industrial production index that is observed at the end of month t, which is equal to $\mathrm{IND}_{t-1}/\ \mathrm{IND}_{t-13} - 1$	Ball and Ghysels (2018)
FV_{19}	Year-over-year growth rate of seasonally adjusted monthly consumer price index that is observed at the end of month t, which is equal to $\mathrm{CPI}_{t-1}/\ \mathrm{CPI}_{t-13} - 1$	
FV_{20}	Year-over-year growth rate of monthly crude oil prices that is observed at the end of month t, which is equal to $\mathrm{OIL}_{t-1}/\ \mathrm{OIL}_{t-13} - 1$	
FV_{21}	Monthly change in yields on 3-month treasury bills that is observed at the end of month t, which is equal to $\mathrm{Y3M}_{t-1}$ minus $\mathrm{Y3M}_{t-2}$	
FV_{22}	Monthly change in the yield spread between 10-year treasury bonds and 3-month treasury bills that is observed at the end of month t, which is equal to TERM_{t-1} minus TERM_{t-2} where TERM = Y10Y – Y3M	
FV_{23}	Monthly change in the yield spread between BAA corporate bonds and AAA corporate bonds that is observed at the end of month t, which is equal to SPRD_{t-1} minus SPRD_{t-2} where SRPD = BAA - AAA	

which is a function of random variable $\gamma_2 \sim N\left(\mu_2, \sigma_2^2\right)$, The function in (A.11) is plotted in Fig. A.1 below where the x-axis denotes the variance of γ_2 (a measure of analyst forecasts cross-sectional and time-series variations) and y-axis denotes the difference in

Table C.1
MSEs for EW and BAM for firms sorted by EW forecast errors of unscaled EPS.

Group	EW	BAM	BAM - EW	t-test
1	0.0275	0.0289	0.0014	(1.98)
2	0.0823	0.0760	−0.0063	(−0.04)
3	0.1686	0.1670	−0.0016	(−0.47)
4	0.3088	0.3026	−0.0061	(−1.05)
5	0.5485	0.5481	−0.0004	(−0.03)
6	0.9873	0.9840	−0.0033	(−0.15)
7	1.9286	1.8392	−0.0894	(−1.06)
8	4.1501	3.9876	−0.1625	(−1.49)
9	12.2621	11.3067	−0.9554	(−2.44)
10	159.9582	84.7559	−75.2024	(−2.50)

This table reports the average values of mean-square-errors (MSEs) of unscaled EPS for different groups. We sort all firms into deciles according to the MSE of EW forecasts where group 1 includes firms with the lowest MSE and group 10 with the highest MSE. We then compare the performance of the regression-based combination (BAM) with EW and report their MSE differences and associated t values.
*MSEs are multiplied by 100.

Table C.2
MSEs of EW and BAM forecasts using scaled EPS.

	Scaled by EPS volatility (Standardized EPS)	Scaled by stock price (EP ratio)	Scaled by book value of equity per share (ROE)
EW	3.231	0.081	0.030
BAM	3.048	0.072	0.026
BAM - EW	−0.183	−0.009	−0.004
t-value	(−4.50)	(−5.38)	(−3.76)
R^2_{os}	5.677%	10.211%	14.475%
t-value	(8.95)	(6.07)	(4.00)

This table reports the average value of mean-square-errors (MSEs) and the out-of-sample R-squared (R^2_{OS}) of scaled EPS. Standardized EPS is computed as the current EPS scaled by the standard deviation of actual EPS in the previous eight quarter; EP ratio is the EPS scaled by stock price on the day before earnings announcement; and ROE is the EPS scaled by book value of equity per share. Negative BAM – EW indicates the reduction of MSE using BAM relative to EW. We compute R^2_{OS} to evaluate the out-of-sample performance of a forecast model relative to the consensus forecast (EW). The higher positive R^2_{OS}, the better out-of-sample forecast performance than EW. The t-values that test the significance of mean MSE difference and R^2_{OS} are reported in parentheses and adjusted for firm-clustered standard errors.

MSEs between the two models. We plot MSE differences for three scenarios: under-reaction of analysts to news ($\mu = 0.8 < 1$), over-reaction ($\mu = 1.2 > 1$) and no biased reaction ($\mu = 1$). For simplicity, we drop the subscript for the mean (μ_2) and variance (σ_2^2) of γ_2.

As shown above, the MSE is always lower for the IMC. When the variance of factor loading increases, which reflect the time-series and cross-sectional variations in analysts' forecasts, the difference in MSEs between the BAM and IMC increases. The results suggest that when heterogeneity among analysts and forecast dispersion are high, the iterated regression method have more advantages. This finding is consistent with the simulation results in Panels B – F of TABLE A2 under Experiment 2 (see the Online Appendix). In addition, the results show that when analysts underact to earnings news (e.g., $\mu = 0.8$), the iterated regression method (IMC) can correct this bias more effectively and produce lower MSEs.

Appendix B

See Tables B.1 and B.2.

Appendix C. Diagnosis of unscaled and scaled earnings per share

See Tables C.1 and C.2.

Appendix D. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jempfin.2022.07.003.

References

Ashton, Alison H., Ashton, Robert H., 1985. Aggregating subjective forecasts: Some empirical results. Manage. Sci. 31, 1499–1508.
Baker, Malcolm, Wurgler, Jeffrey, 2006. Investor sentiment and the cross-section of stock returns. J. Finance 61, 1645–1680.
Ball, Ryan T., Ghysels, Eric, 2018. Automated earnings forecasts: beat analysts or combine and conquer? Manage. Sci. 64, 4936–4952.
Bates, J.M., Granger, Clive W.J., 1969. The combination of forecasts. Oper. Res. Soc. 20, 451–468.

Bernard, Victor, Thomas, Jacob K., 1989. Post-earnings-announcement drift: Delayed price response or risk premium? J. Account. Res. Suppl. 27, l–36.
Bernard, Victor, Thomas, Jacob K., 1990. Evidence that stock prices do not fully reflect the implications of current earnings for future earnings. J. Account. Econ. 13, 305–340.
Bradley, Daniel, Gokkaya, Sinan, Liu, Xi, 2017. Before an analyst becomes an analyst: Does industry experience matter? J. Finance 72, 751–792.
Bradshaw, Mark T., Drake, Michael S., Myers, James N., Myers, Linda A., 2012. A re-examination of analysts' superiority over time-series forecasts of annual earnings. Rev. Account. Stud. 17, 944–968.
Brown, Lawrence D., 1993. Earnings forecasting research: Its implications for capital markets research. Int. J. Forecast. 9, 295–320.
Brown, Lawrence D., Call, Andrew C., Clement, Michael B., Sharp, Nathan Y., 2015. Inside the black box of sell-side financial analysts. J. Account. Res. 53, 1–47.
Brown, Lawrence D., Griffin, Paul, Hagerman, Robert L., Zmijewski, Mark E., 1987. Security analyst superiority relative to univariate time-series models in forecasting quarterly earnings. J. Account. Econ. 9, 61–87.
Call, Andrew C., Hewitt, Max, Shevlin, Terry, Yohn, Teri Lombardi, 2016. Firm-specific estimates of differential persistence and their incremental usefulness for forecasting and valuation. Account. Rev. 91, 811–833.
Capistrán, Carlos, Timmermann, Allan, 2009. Forecast combination with entry and exit of experts. J. Bus. Econom. Statist. 27, 428–440.
Clark, Todd E., West, Kenneth D., 2007. Approximately normal tests for equal predictive accuracy in nested models. J. Econometrics 138, 291–311.
Clement, Michael B., 1999. Analyst forecast accuracy: Do ability, resources, and portfolio complexity matter? J. Account. Econ. 27, 285–303.
Conroy, Robert, Harris, Robert, 1987. Consensus forecasts of corporate earnings: Analysts' forecasts and time series methods. Manage. Sci. 33, 725–738.
Diebold, Francis X., Mariano, Roberto S., 1995. Comparing predictive accuracy. J. Bus. Econom. Statist. 13, 253–263.
Fama, Eugene F., French, Kenneth R., 2000. Forecasting profitability and earnings. J. Bus. 73, 161–175.
Fama, Eugene F., French, Kenneth R., 2006. Profitability, investment and average returns. J. Financ. Econ. 82, 491–518.
Fama, Eugene F., French, Kenneth R., 2008. Dissecting anomalies. J. Finance 63, 1653–1678.
Fama, Eugene F., French, Kenneth R., 2015. A five-factor asset pricing model. J. Financ. Econ. 116, 1–22.
Gerakos, Joseph, Gramacy, Robert B., 2013. Regression-Based Earnings Forecasts. Working Paper, University of Chicago.
Hou, Kewei, van Dijk, Mathijs A., Zhang, Yinglei, 2012. The implied cost of capital: A new approach. J. Account. Econ. 53, 504–526.
Hou, Kewei, Xue, Chen, Zhang, Lu, 2015. Digesting anomalies: An investment approach. Rev. Financ. Stud. 28, 650–705.
Huang, Dashan, Jiang, Fuwei, Li, Kunpeng, Tong, Guoshi, Zhou, Guofu, 2022. Scaled PCA: A new approach to dimension reduction. Manage. Sci. 68, 1678–1695.
Huang, Dashan, Jiang, Fuwei, Tu, Jun, Zhou, Goufu, 2015. Investor sentiment aligned: A powerful predictor of stock returns. Rev. Financ. Stud. 28, 791–837.
Jame, Russell, Johnson, Rick, Markov, Stanimir, Wolfe, Michael C., 2016. The value of crowdsourced earnings forecasts. J. Account. Res. 54, 1077–1110.
Jiang, Fuwei, Lee, Joshua, Martin, Xiumin, Zhou, Guofu, 2019. Manager sentiment and stock returns. J. Financ. Econ. 132, 126–149.
Jones, Jennifer J., 1991. Earnings management during import relief investigations. J. Account. Res. 29, 193–228.
Kim, Oliver, Lim, Steve C., Shaw, Kenneth W., 2001. The inefficiency of the mean analyst forecast as a summary forecast of earnings. J. Account. Res. 39, 329–335.
Kim, Yongtae, Lobo, Gerald J., Song, Minsup, 2011. Analyst characteristics, timing of forecast revisions, and analyst forecasting ability. J. Bank. Financ. 35, 2158–2168.
Kirk, Marcus P., Reppenhagen, David A., Tucker, Jennifer W., 2014. Meeting individual analyst expectations. Account. Rev. 89, 2203–2231.
Kothari, S.P., Leone, Andrew J., Wasley, Charles E., 2005. Performance matched discretionary accrual measures. J. Account. Econ. 39, 163–197.
Kothari, S.P., So, Eric, Verdi, Rodrigo, 2016. Analysts' forecast and asset pricing: a survey. Annu. Rev. Financ. Econ. 8, 197–219.
Li, Kevin K., Mohanram, Partha, 2014. Evaluating cross-sectional forecasting models for implied cost of capital. Rev. Account. Stud. 19, 1152–1185.
Lin, Hai, Wu, Chunchi, Zhou, Guofu, 2018. Forecasting corporate bond returns with a large set of predictors: an iterated combination approach. Manage. Sci. 64, 4218–4238.
Mikhail, Michael B., Walther, Beverly R., Willis, Richard H., 1997. Do security analysts improve their performance with experience? J. Account. Res. 35, 131–157.
Mincer, Jacob A., Zarnowitz, Victor, 1969. The evaluation of economic forecasts (NBER).
O'Brien, Patricia C., 1990. Forecast accuracy of individual analysts in nine industries. J. Account. Res. 28, 286–304.
Pettenuzzo, Davide, Timmermann, Allan, Valkanov, Rossen, 2014. Forecasting stock returns under economic constraints. J. Financ. Econ. 114, 517–553.
Rapach, David, Strauss, Jack, Zhou, Guofu, 2010. Out-of-sample equity premium prediction: Combination forecast and links to the real economy. Rev. Financ. Stud. 23, 821–862.
Sinha, Praveen, Brown, Lawrence D., Das, Somnath, 1997. A re-examination of financial analysts' differential earnings forecast accuracy. Contem. Account. Res. 14, 1–42.
So, Eric C., 2013. A new approach to predicting analyst forecast errors: Do investors overweight analyst forecasts? J. Financ. Econ. 108, 615–640.
Stickel, Scott E., 1990. Predicting individual analyst earnings forecasts. J. Account. Res. 28, 409–417.
Timmermann, Allan, 2006. Forecast combinations. In: Elliott, G., Granger, C.W.J., Timmermann, A. (Eds.), HandBook of Economic Forecasting. Elsevier, Amsterdam, The Netherlands.
Zou, Hui, Hastie, Trevor, 2005. Regularization and variable selection via the elastic net. J. R. Stat. Soc. Ser. B Stat. Methodol. 67, 301–320.