TI 2015-032/III
Tinbergen Institute Discussion Paper

How Informative are the Unpredictable Components of Earnings Forecasts?

Bert de Bruijn
Philip Hans Franses

Erasmus School of Economics, Erasmus University Rotterdam, and Tinbergen Institute, the Netherlands.

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

More TI discussion papers can be downloaded at http://www.tinbergen.nl

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 1600

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Duisenberg school of finance is a collaboration of the Dutch financial sector and universities, with the ambition to support innovative research and offer top quality academic education in core areas of finance.

DSF research papers can be downloaded at: http://www.dsf.nl/

Duisenberg school of finance
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 8579

How informative are the unpredictable components of earnings forecasts?

Bert de Bruijn[a,b,*]
Philip Hans Franses[a,b,†]
[a] Econometric Institute, Erasmus School of Economics,
Erasmus University Rotterdam, the Netherlands
[b] Tinbergen Institute, the Netherlands

March 2015

An analysis of about 300000 earnings forecasts, created by 18000 individual forecasters for earnings of over 300 S&P listed firms, shows that these forecasts are predictable to a large extent using a statistical model that includes publicly available information. When we focus on the unpredictable components, which may be viewed as the personal expertise of the earnings forecasters, we see that small adjustments to the model forecasts lead to more forecast accuracy. Based on past track records, it is possible to predict the future track record of individual forecasters.

Keywords: Earnings Forecasts; Earnings Announcements; Financial Markets; Financial Analysts.
JEL classifications: G17, G24, M41.

1. Introduction

Earnings forecasts can provide useful information for investors. When investors rely on these forecasts, it is important to have insights into how earnings forecasters

*Econometric Institute, Erasmus School of Economics, PO Box 1738, NL-3000 DR Rotterdam, The Netherlands; phone: +3110 4088902; email: debruijn@ese.eur.nl
†Econometric Institute, Erasmus School of Economics, PO Box 1738, NL-3000 DR Rotterdam, The Netherlands; phone: +3110 4081273; fax: +3110 4089162; email: franses@ese.eur.nl

create their forecasts. Knowledge about the key drivers of the earnings forecasts is relevant as it allows for the analysis of the added value of earnings forecasters. More precise, one may want to disentangle the part of earnings forecasts that can be predicted using publicly available information from the part involving private knowledge that the earnings forecasters themselves have. In the present paper we will first estimate the first part and then focus on the usefulness of the second part. That is, to what extent does the contribution of an earnings forecaster lead to more forecast accuracy?

Our present study extends the important study of Stickel (1990) in various dimensions. He investigated whether the change in the forecast by an individual forecaster could be predicted by a change in the average forecast of other forecasters, the deviation of the forecasters previous forecast relative to that total average and the cumulative stock returns since a previous forecast. In Stickel (1990) it is concluded that virtually all explanatory power is associated with the first variable, that is, the change in the average forecast.

We update and extend in several ways. First, we use recent data for the period 1995 to 2011. Second, instead of looking at changes in forecasts we consider the actual levels, in particular as we want to zoom in on the unpredictable part of the forecasts. Third, we allow for the inclusion of more potential explanatory variables in our statistical model. And last, but most importantly, we focus on the added value of the earnings forecasts and seek to derive informative rules to discern the better forecasters from the lesser performing forecasters.

A summary of our findings is the following. A key predictor of the earnings forecasts appears to be the average of all available earnings forecasts concerning the same forecast event. A second predictor is the most recent difference between the individual forecaster's forecast and the average of the currently available forecasts. As the sign is positive, this means that a forecaster who previously was more optimistic about the earnings of a particular firm can be expected to persist in quoting above-average values. Other variables do have some predictive value in individual cases, but we do not find consistent effects. When we focus on the unpredictable components then one of our key findings is that a larger unpredictable component associates with less forecast accuracy. We also document that alternative weights to these unpredictable components can lead to more accuracy. Separating the data in an estimation sample and evaluation sample allows us to draw our final conclusion which is that past track records of forecasters have predictive value for future track records.

The outline of our paper is as follows. In Section 2 we provide a concise summary of the empirical evidence in the literature. In Section 3 we discuss the data and in Sections 4 and 5 we present our results. Section 6 summarizes our findings.

2. Literature review

Earnings forecasts have been the topic of interest for many academic studies. For an extensive discussion of research on earnings forecasts in the period 1992-2007, see Ramnath et al. (2008). For earlier overviews we refer to Schipper (1991) and Brown (1993).

One stream of earnings forecasts research has focused on relationships between forecast performance and forecaster characteristics. Performance can be measured by forecast accuracy and by forecast impact on stock market fluctuations. The characteristics of these performance measurements have been related to timeliness (Cooper et al., 2001; Kim et al., 2011), the number of firms that the forecaster follows (Kim et al., 2011; Bolliger, 2004), the firm-specific experience of the forecaster (Bolliger, 2004), age (Bolliger, 2004), the size of the firm for which the forecasts are created and the size of the company where the forecaster works (Kim et al., 2011; Bolliger, 2004), and whether the forecaster works individually or in a team (Brown and Hugon, 2009).

A second stream of research concerns the behaviour of an earnings forecaster and how it is related to what other forecasters do. In particular, herding behaviour is considered, which occurs when forecasters produce forecasts that converge towards the averages of those of the other forecasters. There have been efforts to categorize earnings forecasters into two groups, corresponding to leaders and followers or to innovators and herders (Jegadeesh and Kim, 2010; Clement and Tse, 2005). This is relevant for many reasons as such different forecasters may consult different sources of information, which in turn can be useful for investors to incorporate this information into their investment decisions. Indeed, a leading or innovative forecaster is perhaps more useful than a herding forecaster. This does not directly imply that leading forecasts are also more accurate, as it is documented that accuracy and the type of forecaster are not necessarily related. In fact, it has been documented that the aggregation of leading forecasts is a fruitful tactic to produce accurate forecasts (Kim et al., 2011).

Recently, Clement et al. (2011) have studied the effect of stock returns and other forecasters' forecasts on what forecasters do. In contrast to Jegadeesh and Kim (2010) and Clement and Tse (2005), Clement et al. (2011) do not consider categorizing the forecasters into different clusters. Instead, they consider how the first forecast revision after a forecast announcement is affected by how the stock market and other forecasters have reacted to that forecast announcement. Landsman et al. (2012) also look at how earnings announcements affect the stock market, where these authors focus on how mandatory IFRS adoption has moderated this effect. Sheng and Thevenot (2012) propose a new earnings forecast uncertainty measure, which they use to demonstrate that forecasters focus more on the information in the earnings announcement if there is more dispersion in the available earnings forecasts.

In sum, earnings forecasts have been studied concerning their performance and a few of their potential drivers. In this paper we extend the knowledge base by considering many more drivers of earnings forecasts, while we pay specific attention to the relevance of the unpredictable component of earnings forecasts.

For our study we go back to Stickel (1990) and seek to extend this important study in various dimensions. In that paper it is concluded that in a statistical model for predicting changes in earnings forecasts the key explanatory variable is the change in the average of all other forecasts. We extend this study by considering more and more recent data and also by including more variables in a model for the levels (and not the changes) of earnings forecasts. A key extension however is that we use the statistical model to disentangle the predictable component from the unpredictable component, and then we zoom in on the latter component. We do so to see to what extent the unobservable knowledge of forecasters contributes to the quality of the earnings forecasts. Also, we aim to examine if forecasters who successfully rely on their knowledge do so persistently. That is, is a past successful track record an indicator for future success?

3. Data and sample selection

Our data have been collected from WRDS[1], using the I/B/E/S database for the analyst forecasts and the CRSP data for the stock prices and stock returns.

Concerning the earnings forecasts, we have collected data for all firms which have been part of the S&P500 during the period 1995 to 2011. Sometimes the sample size was too small, and in other cases we could not properly link the forecasts with the firms, so in the end we have useful information concerning 316 firms with some 270000 earnings forecasts. We focus on the within-year annual earnings forecasts, that is, the forecasts that are produced to forecast the earnings of the current year.

The structure of the data is illustrated in Figure 1. This figure shows an x for the moment when a forecaster makes a forecast available, which is not necessarily the same moment that other forecasters give their quotes as not all forecasters have the same frequency of quotes. This figure also shows the variables which we measure at the highest frequency and these are the daily observed stock returns. Finally, this figure shows vertical lines depicting the moments of the earnings announcements, at which point the realization occurs of the variable that is to be forecasted. We only use the within-year earnings forecasts, which means that we only include forecasts for the next upcoming annual earnings announcement, and hence we abstain from forecasts for year T made in year T-1.

Some descriptives of the data are shown in Table 1. The data until and including 2005 cover the estimation sample, and the data from 2006 onwards constitute the evaluation sample. We make this distinction in order to examine if past track records have predictive value for future track records. And, we also want to see if estimated

[1]Wharton Research Data Services (WRDS) was used in preparing this paper. This service and the data available thereon constitute valuable intellectual property and trade secrets of WRDS and/or its third-party suppliers. `http://wrds-web.wharton.upenn.edu/wrds/`

parameters in the estimation sample provide reasonably constant inference in a post-estimation period.

All data are used to create and evaluate the statistical models for the earnings forecasts. For that purpose we have data on 18338 forecasters and more than 270000 forecasts, with the latter about equally spread over the estimation and evaluation samples. When it comes to forecaster-specific regressions and correlations, we need enough data points to run these computations, and then our sample size drops to about one-third of the forecasts. Still, this is a large database and therefore we are confident that our results below are informative.

4. Predicting earnings forecasts

To create the unpredictable components of earnings forecasts, we first have to create the predictable components. For this we put forward a statistical model to predict earnings forecasts using information publicly available up until the day before the publication of the earnings forecast. In this section we first introduce the statistical model that we use to make predictions of the earnings forecasts. We present the explanatory variables and the relevant estimation results. Next, we apply a correction method to account for the firms for which we have a small number of forecasts.

4.1. The statistical model

For predicting the earnings forecasts we use a linear regression model. The list of explanatory variables is presented in Table 2.

Following Stickel (1990), we expect earnings forecasters to look at the recent forecasts of competing forecasters. We thus include in our model the average of all most recent forecasts across individual forecasters. Note that we only include forecasts that have been made within the same year for the same forecast event.

We also include in our model several variables that are related to this average forecast. First, the average forecast may contain more useful information when it is based on a larger number of forecasters. To see whether this holds true, we include an interaction term of the average forecast with an indicator function that is 1 if the number of forecasters is below 10 and that is 0 otherwise. The contribution to the fit of average forecast may also increase when the moment of the actual announcement of the true value of the earnings comes closer. We therefore include an interaction term with an indicator function which is 1 when the data concern the last two weeks before the announcement, and 0 otherwise. The final explanatory variable related to the average forecast is the day-to-day change in this average forecast. Indeed, when the average forecast has increased on one day, then individual forecasters could be tempted to extrapolate this growth to the next day.

The second set of explanatory variables concerns the own previous forecasts of a forecaster. We include the most recent forecast and the difference between this previous forecast and the average forecast at that particular moment in time. These two variables can allow for persistence in the opinions of a forecaster, implying that forecasters can be more optimistic or pessimistic for some period of time.

Finally, the third set of explanatory variables concerns the stock market. We include the most recent stock price of the firm for which its earnings are predicted. Also, recent changes in the stock price can be relevant, and for that purpose we include the daily returns and the returns relative to the most recent moment when an individual forecaster produced a forecast. Next to these three firm-related stock prices, we include similar variables for the entire S&P500 stock exchange.

We estimate the parameters in the regression models using Ordinary Least Squares (OLS) for data for each of the 316 firms, and summaries of the estimation results across these 316 firms are presented in Table 3. The first five columns show results on the OLS based estimates, including the mean of the estimated parameters, the median and their standard deviation and also the 5% and a 95% percentiles of these estimates. The next two columns concern a summary of the standardized estimates, which are the estimates that are found if the variables are first all standardized by subtracting the mean and scaling the variance to 1. Such standardized estimates can be helpful when comparing the contribution of each of the variables to the overall fit. In the last column of Table 3 we present these contributions as percentages.

The results in Table 3 show that, on average, the coefficient of the recent average forecast is about 1. The distribution of this effect across firms, indicated by standard deviation and percentiles) indicates that the sign of this effect is consistently positive. When we scroll down the table, we see that none of the other variables have this property. Also, looking at the contribution to the fit, it is clear that the average forecast is most important, and that the previous forecast and its difference to the average forecast are a distant second and third useful explanatory variable.

To continue with these regression models, Table 4 shows summary statistics on the t-statistic values for each of the explanatory variables. The last column of this table shows that all variables are significant for at least 20% of the firms, but it also repeats the finding that most of the variables are not consistent in the sign of their effect (and thus, the sign of their t-statistic). Again, the average forecast is seen to be most relevant as it is associated with the largest percentage of significant cases. Additionally, the difference of the previous forecast to the average stands out with a higher percentage significant and a high median value of the t-statistic (78% of the cases).

4.2. Correcting for small sample sizes

The results in Tables 3 and 4 show that various explanatory variables do have a statistically significant effect, but at the same time this effect does not have a consistent sign. The latter causes the finding in Table 3 that on average these effects are equal to 0. Now it could be that this finding is a small-sample effect, as for some firms we only have a small number of earnings forecasts.

To correct for these small sample sizes, we employ the following method that is detailed in Appendix A. This method amounts to an assumption that the collection of firm-specific (population) parameters for one of the variables corresponds to a normal distribution. Suppose that the parameters of this distribution are known. As a consequence, there are two sources of information for the value of each individual estimated parameter, and these are the estimated OLS coefficient and the parameters of this common distribution. The optimal choice is a weighted average of these two values, with weights determined by the standard error of the estimated coefficient and the standard deviation of the underlying distribution. For firms with only a few observations, the weight for the estimated coefficient most likely will be low, and the best estimate will thus be relatively close to the mean of the common distribution. On the other hand, for firms with many observations the weight of the estimated coefficient will be high and the best estimate will not deviate much from the OLS estimated parameters.

In our application, we of course do not know the values of the common distribution in advance. We therefore apply an iterative process. First, the two parameter values are initialized on the sample mean and standard deviation of all OLS estimates. Then, we adjust the estimates using the weights. After adjustment, we use the weighted mean and weighted standard deviation to construct a new value of the two parameters, with weights equal to the reciprocal of the estimated standard error. This is again followed by a new adjustment of the estimated parameters, and then again the calculation of a new set of parameters. We do this until convergence.

When we apply this method we obtain the summarized results in Table 5. Comparing the numbers in this table with those in Table 3, we can see that the average and median values have not changed much. In contrast, and as expected, the standard deviation and the width of the 90% interval have clearly decreased. There are now more variables that are (almost) consistent in their estimated sign, and among them are the parameters for firm-specific stock price and the S&P500 stock market index. At the same time, however, the contribution to the fit as reported in the last column has stayed about the same.

To conclude, whether we employ a small-sample correction or not, the key result is that the recent past average forecast is the main explanatory variable for current earnings forecasts. At the same time, for many individual cases (out of the 316 cases) we find various other variables to be relevant, and we will use these variables in our analyses below. Like Stickel (1990) we find that earnings forecasts can be predicted, and as such we substantiate these earlier findings.

The estimation results for the statistical models so far are informative in their own right, but for our present study they mainly serve to each time disentangle a predictable component from an unpredictable component. This last component will become the focus of our interest in the rest of this paper.

5. How useful are the predictable and unpredictable components?

Now we have seen that earnings forecasts can be predicted to quite some extent, we will now analyse to what extent earnings forecasters add some value to the statistical model that can be constructed using publicly available data. We adopt three focus points. The first concerns all forecasts, then we see if we can evaluate individual forecasters against each other, and finally we consider the forecasts from the same forecaster and compare these with his or her own other forecasts.

5.1. All forecasts

We start with an examination of the predictive accuracy and compare the performance of the forecasts of the earnings forecasters (which are of course equal to the sum of the predictable and unpredictable components) with the statistical model forecasts (which are just the predictable components). Next, we zoom in on the size of the unpredictable components and examine if larger deviations from the statistical model forecasts are better or not. Finally, we look at whether we can use the unpredictable component in an alternative and perhaps better way by using different weights.

5.1.1. Do earnings forecasts improve on statistical model forecasts?

Table 6 shows some statistics on a newly created variable that seeks to highlight the differences across the two sets of forecasts. This variable is the median value (across 316 firms) of the ratio of squared earnings forecast errors over squared model forecast errors. The difference between these two sets of forecasts is the unpredictable component, so if this performance ratio is different from 1 in either direction then that must be due to this unpredictable component. The table presents this median ratio for both the estimation sample and the evaluation sample. In the evaluation sample we use the model parameters as they have been estimated using the estimation sample.

The bottom panel of Table 6 shows that the performance ratio is below 1 in about

65% of the cases. Hence, for 65% of the 316 firms, the earnings forecasts created by the earnings forecasters provide more accuracy than the predictable component from the statistical model. The results across the estimation sample and the evaluation sample are similar. Note that this thus means that for 35% of the firms one could easily rely on the statistical model forecasts.

Table 7 concerns the outcomes for the same ratio, but now for different parts of the year. These parts correspond with the four periods between the quarterly announcements and the three periods around the quarterly announcements (except for the quarterly announcement that coincides with the yearly announcement). The results in this table show that the performance ratio increases throughout the year, meaning that the unpredictable component leads to more accuracy in the beginning than towards the end. This might be due to an increase in the accuracy of the statistical model simply because the predictable components are then based on more observations and thus the sample period which may require added expertise from the earnings forecaster becomes smaller. All in all, we can conclude that the earnings forecasters can substantially contribute to the quality of the final forecasts, which is most obvious from the two bottom rows in Table 7.

5.1.2. Is there an optimal size of the unpredictable component?

As the unpredictable component does seem to help for improved forecast accuracy we now want to know what kind of added value of an earnings forecaster makes the difference. For this, we regress the squared earnings forecast error on a constant, on the unpredictable component and on the squared unpredictable component. We do this three times, once using the raw data and twice using two standardization approaches. Standardization might be relevant as it may occur that earnings are difficult to forecast as the data may be unstable, and this could then have an effect on both the squared earnings forecast error and the unpredictable components. The first standardization employs the variance of the predictable component for a firm, whereas the second considers the variance of the unpredictable component. The results are presented in Table 8.

The left-hand columns of Table 8 show the unstandardized results, while the other columns concern the standardized results. In all cases, even across the estimation and evaluation samples, the result for the squared unpredictable component is the same. That is, the larger is the squared unpredictable component, the larger is the squared forecast error of the earnings forecast. So, in general, earnings forecasts that are close to what a statistical model would predict are most accurate.

The story for whether an earnings forecast is better off by being higher or lower than the predictable component is not so clear, as can be seen from the right-hand side columns of Table 8. Using no standardization or the first standardization suggests that negative unpredictable components perform better (see the positive parameters on UC). However, the second type of standardization gives no relationship (in the estimation sample) or the opposite relationship (in the evaluation sample). In all cases, however, the parameter for UC^2 stays close to 1. So, we do not find evidence that systematically adjusting upwards or downwards leads to more accuracy.

5.1.3. How useful is the unpredictable component?

Table 9 presents our OLS-based estimation results of the regression of the actual (the true earnings observations) on various functions of the predictable and unpredictable components. We include interaction terms with the number of forecasts, as the predictable component might be more accurate when it is based on more forecasts. We also include interaction terms with the time until the announcement, as forecasts just before the announcement might have already incorporated all information into the predictable component, and as such leaving not much room for extra expertise of the earnings forecaster.

Several results in this table are interesting. The estimated parameters for the predictable and unpredictable components in the estimation sample seem to suggest that they need to be made more important than what they are in the actual forecast. In the latter their weights are equal to 1 by construction, but the table suggests that alternative weights could be beneficial. Note that these larger weights are downplayed by the interaction terms with the number of forecasts and the time until announcement, which are two variables that are both strictly positive and have an associated negative parameter estimate. To visualize this findings, consider Figure 2 which shows the effective parameters for both components throughout the year, both in the estimation and in the evaluation sample. This figure demonstrates that the optimal weight for each of the components is always below 1. For the predictable component, the contribution is relatively stable throughout the year, whereas for the unpredictable component the contribution is highest at the beginning of the year.

Another result from Table 9 is that the predictable component parameters are all estimated more accurately than their unpredictable component counterparts. Also, and not unexpectedly, the estimation results are more reliable in the estimation sample than in the evaluation sample.

These results altogether suggest that the optimal contribution of the unpredictable component can be less than 1. Hence, in other words, perhaps the earnings forecasters are adding too much of their unobservable expertise on top of what a statistical model already could achieve. This is not to say that the contribution of this expertise should be set at 0, as the results in Table 10 clearly indicate that this unpredictable component matters. This table presents the results on the F-test for the joint statistical relevance of the four variables that are associated with the unpredictable component. In both samples, the median F-statistic is larger than 20, and the 5% based F-test rejects no significant effect in more than 90% of the cases. Hence, there are clear signs that the unpredictable component does add useful information.

The next step is to examine how much the unpredictable component actually contributes to forecast accuracy. Table 10 also shows the R^2 values when using the predictable component variables and also when additionally including the unpredictable component variables. The increase in the median R^2 is about 2 to 3 percent, which is not that much. On the other hand, the median R^2 using only the predictable variables is already around 90% so there is not much left to be explained.

To complete our story on weights of the two components that could constitute an accurate forecast, we look at the comparison of the accuracy of optimally weighted forecasts to its constituent earnings forecasts and statistical model forecasts, and we report the results in Table 11.

From this table we see that the ratios that include the errors of the optimal forecast (OFE) are smaller than 1 for the samples for which the optimality is based on the parameters in the estimation sample for the estimation sample data or on the parameters found in the evaluation sample for the evaluation sample data. In contrast, when we use the estimation sample parameters for the evaluation sample data, the relevant ratio is larger than 1 compared to both the model forecast and the earnings forecast. This suggests that the optimal weights do not yield a stable performance over time and that they apparently need to be re-estimated on a regular basis.

5.2. Comparison across forecasters

The final part of our empirical analysis concerns an examination into which properties of individual earnings forecasters make them to display superior forecast performance. We first look at the key aspects that make earnings forecasters outperform a statistical model, and next we zoom in on optimal properties of the added expertise of the forecaster.

5.2.1. How to find outperforming earnings forecasters?

To investigate which earnings forecasters do best, we introduce a new measure, which is the balanced relative difference defined by $BRD(E,P) = \frac{EFE^2 - PCE^2}{EFE^2 + PCE^2}$, where EFE refers to the forecast error of the earnings forecaster and PCE refers to the forecast error of the predictable component. The variable to be explained concerns the data in the evaluation sample. Table 12 presents the results of a regression of the balanced relative difference between the earnings forecasts and the predictable component thereof (in the evaluation sample) on an intercept, the ratio of the squared unpredictable component to the squared predictable component and three balanced relative differences. These latter three variables are the BRD(E,P) itself and the $BRD(U,P) = \frac{UC^2 - PCE^2}{UC^2 + PCE^2}$ and the $BRD(O,P) = \frac{OFE^2 - PCE^2}{OFE^2 + PCE^2}$, where UC denotes the unpredictable component and OFE refers to the optimal forecast. Three variables show significant results. First, the BRD(E,P) in the evaluation sample is significantly related to its previous value in the estimation sample. So, the past track record seems to have predictive value for the future track record. Next to this, the previous value of the relative size of the unpredictable component to the predictable component and to the previous value of BRD(E,P) also contain predictive information.

These results suggest that the forecasters who will predict best in the evaluation sample are those that have predicted best in the estimation sample (autoregressive), who have a small unpredictable component relative to the predictable component and who have a small unpredictable component relative to the error of the predictable component. Of these three, the autoregressive type variable has the highest statistical significance.

We can now use the above regressions to produce forecasts for the median balanced relative difference of each forecaster. Next, we can then compare the actual errors of the 50% forecasters that who we predict to have the best performance to the 50% forecasters who we predict to perform the worst. The ratio of the median squared error of the best 50% to the median squared error of the worst 50% turns out to be 0.600. Also, the predicted probabilities of having a negative balanced relative difference, which are the probabilities of outperforming the statistical model, are on average 80.8% and 61.9% for the best and worst half, respectively. This indicates that it is indeed possible to select a subset of all forecasters who will perform better in future.

5.2.2. Which forecasters have most expertise?

We use a similar approach to investigate whether it is possible to select forecasters who have more useful information in their unpredictable component, where we define this situation as where the optimal forecast performs best. We again use balanced relative differences. The variable to be explained now is BRD(O,P) in the evaluation sample. The results are presented in Table 12 in the right-hand side panel.

Again, the autoregressive type variable is statistically most significant, whereas the other two significant regressors are the other two balanced relative differences, that is, BRD(U,P) and BRD(E,P). Hence, the forecasters with the most useful information (meaning low values of BRD(O,P)) in the evaluation sample are those with the most useful information in the estimation sample, who are most accurate in the estimation sample and, surprisingly, who have a large unpredictable component.

When we compare the actual optimal forecast errors of two groups that are predicted to have the most and the least information, we get that the relative median squared optimal forecast error is 0.566. Hence, it is indeed possible to select a subset of the forecasters that contains those who have more informative unpredictable components.

5.2.3. Do the best performing forecasters have most expertise?

One may now wonder if there is an overlap between the best-performing forecasters and those who have most (unobservable) expertise. To investigate this, we calculate the hit rate, which is the percentage of cases in which a forecaster is categorized in the same cluster for both measures. It so turns out that this hit rate is 85.4%, which to us indicates that the question in the title can be answered affirmatively.

5.3. Comparison within forecasters

In this last subsection, we take a look at individual forecasts and compare their properties to other forecasts by the same earnings forecaster. Indeed, a large unpredictable component might be much more surprising if produced by a forecaster who usually has small unpredictable components than if produced by someone else who usually has large unpredictable components. In the first case, this single forecast may be based on unique and important information, but it might also mean that the forecaster quoted at random.

We compute for each forecaster the correlation between the size of the unpredictable component and the three balanced relative difference variables, which are BRD(E,P), BRD(O,P) and BRD(O,E), of which the latter is defined as $BRD(O,E) = \frac{OFE^2 - EFE^2}{OFE^2 + EFE^2}$. This last measure can be interpreted as how much the earnings forecaster could improve his forecast if he would optimally use his available information. As measures for the size of the unpredictable component we use both $|UC|$ and UC^2. A summary of the results across all forecasters is presented in Table 13.

Table 13 shows that only negative correlations are found. The negative correlations between the size variables of UC and BRD(E,P) indicate that large unpredictable components for a particular forecaster are associated with a better performance compared to the statistical model. Similarly, the negative correlations with BRD(O,M) show that large unpredictable components are associated with more information in that unpredictable component. Finally, the negative correlations with BRD(O,E) show that large unpredictable components are associated with a better optimal forecast than the actual earnings forecasts, and thus with less optimal use of the unpredictable component by the earnings forecaster.

Table 13 also covers the evaluation sample. In this case, not all correlations are negative. The correlations with BRD(E,P) and BRD(O,P) result in the same qualitative conclusion as before, that is, large unpredictable components are associated with a better performance and more information than smaller unpredictable components produced by the same forecaster. The positive correlation of BRD(O,E) with the size of the unpredictable component indicates that in this case, on average, large unpredictable components coincide with less opportunity to set optimal weights in the combination of the unpredictable component with the model forecast. This finding may be due to unstable weights over time.

Overall, we find that in general small-sized added expertise of an earnings forecaster to a statistical model forecast is beneficial. At the same time, when an individual forecaster with a track record of small-sized added expertise suddenly makes large adjustments, then this usually leads to an increased accuracy of the earnings forecasts.

6. Conclusion

Earnings forecasts are an important factor in the decision making process of investors. In this paper we have shown that earnings forecasts can be predicted, which allows investors to already incorporate the predictable part in their investment decision. Furthermore, we also show that the unpredictable part of an earnings forecast can be used. One way to use it, is to improve the forecast based on just the predictable part. This is especially beneficial in the beginning of the year. Another use of the predictable and unpredictable components concerns the selection of earnings forecasters, which can be relevant if an investor wants to ignore the forecasters with a poor track record. We have shown that there is persistence in the performance of forecasters compared to the predictable component, that is, earnings forecasters who perform better in our estimation sample, also perform better, on average, in the evaluation sample. Similarly, the information in the unpredictable component, that can be used to improve the optimal forecast, is also persistent, that is, earnings forecasters whose unpredictable components are more useful in the estimation sample also have this property in the evaluation sample.

In general, large unpredictable components seem to be a bad sign, as they are associated with large relative forecast errors. This is not the case if the earnings forecaster normally produces small unpredictable components. In that case, a large unpredictable component is a sign of both good performance and more useful information in this unpredictable component.

A. Small-sample error correction method

In Section 4 we use a model to predict earnings forecasts, including a correction to account for small-sample error. In this appendix, we present the mathematical definition of the model.

We will describe the regression by using the familiar notation

$$y_{i,j,t} = X_{i,j,t}\beta_j + \varepsilon_{i,j,t}, \tag{1}$$

with subscript i denoting the individual forecaster, j the firm for which the earnings are forecasted and t the day on which the forecast is produced. The parameter coefficients are denoted by β_j, which is a vector consisting of $\beta_{j,k}$ for $k = 1, .., K$, one parameter for each variable in $X_{i,j,t}$. We will let the vector of coefficients differ per firm, but not per individual nor for different time periods. Also, the error variance $\sigma^2_{\varepsilon,j}$ differs per firm. This is the model without the small-sample error correction.

Now we introduce the small-sample error correction, for which we use a latent variable model for β_j. We can use this latent variable model to correct estimates that have been estimated with a small number of data points and which are thus less accurate and more prone to outliers. These estimates can be adjusted towards the overall mean of that respective parameter, and we do that in such a way that estimates based on more than thousand observations are hardly affected. As a necessary assumption for this model we use

$$\beta_j \sim N(\beta^*, \Sigma_\beta) \tag{2}$$

which means that the latent parameter vector β_j (the estimated parameters for firm j) is related to the overall mean parameter vector β^*. For simplicity, we will assume the covariance matrix Σ_β to be diagonal. Then we employ the following steps:

1. The elements of β^* and Σ_β are estimated by taking the weighted average and weighted variance of all individual estimates.

2. We update each individual estimate by taking a weighted average:

$$\beta_{j,k}^{(u)} = w_{j,k}\beta_k^* + (1 - w_{j,k})\beta_{j,k} \tag{3}$$

$$w_{j,k} = \frac{\frac{1}{\sigma_{\beta,k}}}{\frac{1}{\sigma_{\beta,k}} + \frac{n_k}{\sigma_{\varepsilon,j}}} \tag{4}$$

   The weights are calculated using the inverses of the latent variable standard deviation and the standard error of the regression, as these determine how accurate both sources of information on the $\beta_{j,k}$ estimate are.

We will repeat (3) and (4) until convergence.

Tables and figures

Table 1: The number of firms, forecasters and forecasts for each upcoming section and subsection. The number of forecasts is shown separately for the estimation sample, which is up until 2005, and the evaluation sample, which is from 2006 onwards.

	Number of firms	Number of forecasters	Number of forecasts	
			Estimation sample	Evaluation sample
Section 4 and 5.1	316	18338	146319	126651
Section 5.2	316	1835	52236	36403
Section 5.3	316	4541	90190	28000

Table 2: The variables that are used to forecast the earnings forecasts. These variables enter a linear model. They all use one-day lagged information. Several variables are based on historic forecasters' behaviour, others are based on stock market data.

Variable	Description
Intercept	
Average Forecast	The average of all most recent forecasts of every forecaster, until the previous day
Average Forecast is also included in an interaction term with two indicator variables:	
	1. Whether the number of forecasters is lower than 10 or not, $I[nrF < 10]$
	2. Whether the time until the announcement of the earnings is more than two weeks or not, $I[TUA > 10]$
Δ Average Forecast	First difference in Average Forecast
Δ Previous Forecast	The difference between the previous forecast of the forecaster, and the average forecast at that time
Previous Forecast	The previous forecast of the individual forecaster
Stock Price Firm	The daily stock price of the firm for which the earnings are forecasted
Stock Returns Firm	The daily returns of the firm for which the earnings are forecasted
Cumulative Stock Returns Firm	Stock returns of the firm since the day of the previous forecast by this forecaster
Stock Index S&P500	The daily stock market index of the S&P500
Stock Returns S&P500	The daily stock market returns of the S&P500
Cumulative Stock Returns S&P500	Stock market returns of the S&P500 since the day of the previous forecast by this forecaster

Table 3: A summary of estimation results of forecasting earnings forecasts. Results are for the estimation sample, which amounts to 316 firms, 18338 forecasters and 146319 forecasts. The variable to be explained concerns the earnings forecasts. As explanatory variables we include the variables mentioned in Table 2. The regression is run individually for each firm, and the table shows statistics which summarize these results. The first five columns contain summary results on the regular parameter estimates (average, median, standard deviation and bounds of a 90% interval). The last three columns show results for the standardized estimate, which is included to compare contributions to the fit. The standardized estimate is defined as the estimate that would have been obtained had the regressor been standardized beforehand (which is a transformation to having an average of 0 and a standard deviation of 1).

	Estimate					Standardized estimate		
	Average	Median	Standard Deviation	Bounds of 90% interval		Median	Median of absolute	Contribution to total fit
Intercept	-0,028	-0,026	0,278	-0,395	0,229			
Average Forecast	1,050	1,077	0,384	0,557	1,483	0,490	0,491	96,9%
Average Forecast x $I[nrF < 10]$	0,018	0,005	0,172	-0,053	0,078	0,001	0,004	0,0%
Average Forecast x $I[TUA > 14]$	-0,038	-0,034	0,292	-0,212	0,058	-0,017	0,025	0,3%
Δ Average Forecast	0,765	0,607	1,053	-0,526	2,438	0,005	0,006	0,0%
Δ Previous Forecast	0,498	0,548	0,359	-0,151	1,051	0,029	0,030	0,4%
Previous Forecast	-0,046	-0,078	0,273	-0,388	0,375	-0,028	0,076	2,3%
Stock Price Firm	0,002	0,001	0,004	-0,001	0,007	0,009	0,013	0,1%
Stock Returns Firm	0,302	0,150	1,039	-0,264	1,291	0,005	0,007	0,0%
Cumulative Stock Returns Firm	0,054	0,018	0,206	-0,092	0,285	0,003	0,005	0,0%
Stock Index S&P500	0,000	0,000	0,002	-0,001	0,002	0,001	0,009	0,0%
Stock Returns S&P500	-0,214	-0,096	1,226	-2,155	1,421	-0,001	0,004	0,0%
Cumulative Stock Returns S&P500	-0,032	-0,006	0,238	-0,408	0,256	0,000	0,005	0,0%

Table 4: A summary of t-statistics when forecasting earnings forecasts. Results are for the estimation sample, which amounts to 316 firms, 18338 forecasters and 146319 forecasts. The variable to be explained concerns the earnings forecasts. As explanatory variables we include the variables mentioned in Table 2. The regression is run individually for each firm, and the table shows statistics which summarize these results.

	Median t-statistic	Median absolute of t-statistic	Percentage significant at 5% level
Intercept	-0,986	1,920	48,4%
Average Forecast	9,865	9,865	96,4%
Average Forecast x $I[nrF < 10]$	0,530	1,118	27,9%
Average Forecast x $I[TUA > 14]$	-1,662	2,212	51,6%
Δ Average Forecast	1,680	1,766	47,2%
Δ Previous Forecast	4,794	4,794	78,0%
Previous Forecast	-0,804	1,599	38,6%
Stock Price Firm	1,928	2,402	55,8%
Stock Returns Firm	1,378	1,653	40,1%
Cumulative Stock Returns Firm	0,730	1,329	32,9%
Stock Index S&P500	0,151	1,928	49,3%
Stock Returns S&P500	-0,395	1,192	24,9%
Cumulative Stock Returns S&P500	-0,110	1,196	26,7%

Table 5: A summary of estimation results of forecasting earnings forecasts, after using the correction method to account for small-sample error. Results are for the estimation sample, which amounts to 316 firms, 18338 forecasters and 146319 forecasts. The variable to be explained concerns the earnings forecasts. As explanatory variables we include the variables mentioned in Table 2. The regression is run individually for each firm, and the table shows statistics which summarize these results. The first five columns contain summary results on the regular parameter estimates (average, median, standard deviation and bounds of a 90% interval). The last three columns show results for the standardized estimate, which is included to compare contributions to the fit. The standardized estimate is defined as the estimate that would have been obtained had the regressor been standardized beforehand (which is a transformation to having an average of 0 and a standard deviation of 1). The correction method is based on the assumption of an underlying normal distribution out of which each parameter (for the different firms) is drawn. This provides additional information on the firm-specific estimate especially in the case when the firm has only a few observations.

		Estimate					Standardized estimate		
		Average	Median	Standard Deviation	Bounds of 90% interval		Median	Median of absolute	Contribution to total fit
Analyst-based	Intercept	-0,019	-0,021	0,051	-0,113	0,068			
	Average Forecast	1,097	1,094	0,134	0,880	1,287	0,479	0,479	98,5%
	Average Forecast x $I[nrF < 10]$	0,003	0,004	0,011	-0,012	0,019	0,001	0,002	0,0%
	Average Forecast x $I[TUA > 14]$	-0,048	-0,040	0,063	-0,161	0,018	-0,019	0,022	0,2%
	Δ Average Forecast	0,741	0,687	0,409	0,148	1,444	0,005	0,005	0,0%
	Δ Previous Forecast	0,553	0,564	0,192	0,206	0,865	0,029	0,029	0,4%
	Previous Forecast	-0,081	-0,087	0,120	-0,253	0,117	-0,031	0,044	0,8%
Stock market	Stock Index Firm	0,001	0,001	0,001	0,000	0,002	0,009	0,009	0,0%
	Stock Returns Firm	0,160	0,120	0,175	-0,043	0,508	0,005	0,005	0,0%
	Cumulative Stock Returns Firm	0,016	0,013	0,022	-0,013	0,055	0,002	0,003	0,0%
	Stock Index S&P500	0,000	0,000	0,000	0,000	0,000	0,000	0,005	0,0%
	Stock Returns S&P500	-0,087	-0,081	0,214	-0,410	0,251	-0,001	0,001	0,0%
	Cumulative Stock Returns S&P500	-0,002	-0,002	0,036	-0,065	0,060	0,000	0,002	0,0%

Table 6: A summary of results on median $\frac{EFE^2}{PCE^2}$, the median ratio of squared earnings forecast error over squared predictable component error. The predictable component error is the error made if we use the part of the earnings forecast that we can predict using a statistical model. This ratio shows whether the inclusion of the unpredictable component results in an improvement or not. The last two rows show the average and median for the percentage of the forecasts for which the ratio is smaller than 1. We show results for 18338 forecasters across 316 firms, separated for the estimation (146319 forecasts) and evaluation (126651 forecasts) samples.

Period	Estimation sample	Evaluation sample
Average	0,609	0,638
Median	0,655	0,631
Standard Deviation	0,304	0,571
5% percentile	0,071	0,061
95% percentile	1,031	1,207
Percentage $<$ 1, average	65,1%	66,4%
Percentage $<$ 1, median	63,7%	64,4%

Table 7: A summary of results on median $\frac{EFE^2}{PCE^2}$, the median ratio of squared earnings forecast error over squared predictable component error. The predictable component error is the error made if we use the part of the earnings forecast that we can predict using a statistical model. This ratio shows whether the inclusion of the unpredictable component results in an improvement or not. The last two rows show the average and median for the percentage of the forecasts for which the ratio is smaller than 1. We show results for 18338 forecasters across 316 firms, for a total number of 272970 observations spread over seven periods in the year leading up to the earnings announcement. The seven periods roughly correspond to the periods around the quarterly earnings announcement (excluding the fourth quarter, which coincides with the announcement of the earnings of interest) and the four periods in-between.

	During Q1	Announcement Q1	During Q2	Announcement Q2	During Q3	Announcement Q3	During Q4
Average	0,455	0,414	0,632	0,613	0,797	0,789	0,921
Median	0,335	0,350	0,640	0,625	0,808	0,750	0,887
Standard Deviation	0,464	0,329	0,351	0,379	0,411	0,683	0,552
5% percentile	0,020	0,025	0,077	0,088	0,166	0,180	0,266
95% percentile	1,170	1,032	1,136	1,209	1,463	1,378	1,581
Percentage < 1, average	69,7%	72,3%	64,4%	66,6%	60,2%	62,1%	58,1%
Percentage < 1, median	69,0%	71,8%	62,8%	66,2%	60,0%	62,1%	57,1%

Table 8: Regression of squared forecast error of the earnings forecasts on the unpredictable component and its square: $FE^2 = \beta_0 + \beta_1 UC + \beta_2 UC^2$. We do this for all 316 firms and 18338 forecasters simultaneously in two regressions, one for the estimation sample (n=146319) and one for the evaluation sample (n=126651). Next to OLS estimation of the above linear model, we also use two standardization methods to account for firm differences in the size of earnings and the uncertainty of earnings. Standardization 1 uses the variance of the predictable component per firm. Standardization 2 uses the variance of the unpredictable component per firm. Standard errors in parentheses.

		No standardization		Standardization 1		Standardization 2	
Estimation sample	intercept	0,116	(0,003)	0,079	(0,002)	2,427	(0,034)
	UC	0,539	(0,020)	0,257	(0,012)	0,151	(0,044)
	UC^2	0,891	(0,011)	0,940	(0,008)	0,934	(0,012)
	R^2	0,044		0,085		0,040	
Evaluation sample	intercept	0,777	(0,024)	0,163	(0,004)	5,955	(0,061)
	UC	0,292	(0,054)	0,250	(0,015)	-1,182	(0,042)
	UC^2	0,980	(0,002)	1,006	(0,001)	0,879	(0,007)
	R^2	0,625		0,810		0,107	

Table 9: A summary of results of the regression of the actual earnings on predictable and unpredictable component variables: $Actual = \alpha + \beta PCV + \gamma UCV$. PCV does not include only the predictable component itself, but also interaction terms of the predictable component with logNF, the logarithm of the number of forecasts on which Average Forecast is based at that moment, and with logTUA, the logarithm of the number of days until the announcement. In a similar way UCV is based on the unpredictable component and interaction terms of the unpredictable component with logNF and logTUA. We run these regressions for each firm separately (of the 316 firms) but pool the results of all 18338 forecasters. The total number of observations in the regressions across all firms is 146319 in the estimation sample and 126651 in the evaluation sample. We present several statistics (average, median, standard deviation, 90% interval) on the estimated parameters and also the average and median of the standard error of the parameters.

		Estimated coefficient					Standard error	
		Average	Median	Standard deviation	Bounds of 90% interval		Average	Median
Estimation sample	intercept	0,067	0,021	0,348	-0,243	0,509	0,034	0,020
	PC	1,060	1,099	3,815	-3,274	4,489	0,663	0,444
	PC x logNF	-0,018	-0,019	1,176	-0,980	1,240	0,233	0,152
	PC x logTUA	-0,020	-0,016	0,704	-0,582	0,834	0,122	0,081
	PC x logNF x logTUA	0,004	0,004	0,215	-0,249	0,182	0,043	0,027
	UC	2,400	1,506	24,167	-34,384	40,469	11,212	10,087
	UC x logNF	-0,663	-0,456	8,279	-13,757	12,491	4,005	3,460
	UC x logTUA	-0,198	0,028	4,458	-7,161	6,155	2,068	1,836
	UC x logNF x logTUA	0,082	0,029	1,543	-2,411	2,530	0,744	0,646
Evaluation sample	intercept	0,394	0,278	0,972	-0,498	1,704	0,067	0,046
	PC	0,054	0,711	5,900	-8,086	5,740	0,950	0,506
	PC x logNF	0,242	0,068	1,888	-1,590	2,451	0,326	0,174
	PC x logTUA	0,084	0,027	0,988	-0,959	1,272	0,172	0,095
	PC x logNF x logTUA	-0,024	-0,006	0,321	-0,367	0,343	0,059	0,033
	UC	-2,370	-2,213	43,991	-52,879	50,791	13,371	9,822
	UC x logNF	0,712	0,793	14,559	-18,921	16,258	4,661	3,402
	UC x logTUA	0,811	0,655	7,789	-9,058	9,518	2,435	1,832
	UC x logNF x logTUA	-0,224	-0,229	2,584	-3,003	3,002	0,853	0,631

Table 10: A summary of results on the comparison between the regressions of (1) $Actual = \alpha + \beta PC$, the actual earnings on only predictable component variables and (2) $Actual = \alpha + \beta PCV + \gamma UCV$, the actual earnings on both the predictable and unpredictable component variables. We run these regressions for each firm separately (of the 316 firms) but pool the results of all 18338 forecasters. The total number of observations in the regressions across all firms is 146319 in the estimation sample and 126651 in the evaluation sample. The F-statistic is based on the test for the joint significance of γ, the parameters of the unpredictable component variables, and the results for the associated p-value are shown in the column labelled p-value. Also shown are results for the R^2 values for both the restricted and the unrestricted model.

	Estimation sample				Evaluation sample			
	F-statistic	p-value	R^2 without UC	R^2 with UC	F-statistic	p-value	R^2 without UC	R^2 with UC
Average	33,776	0,015	0,868	0,893	29,292	0,039	0,817	0,850
Median	21,554	0,000	0,918	0,935	20,794	0,000	0,878	0,901
Standard Deviation	44,568	0,103	0,155	0,129	30,484	0,164	0,180	0,155
5% percentile	3,066	0,000	0,554	0,610	1,442	0,000	0,422	0,508
95% percentile	101,095	0,032	0,997	0,997	85,891	0,233	0,988	0,993
Significant at 5% level	96,2%				91,8%			

Table 11: A summary of results on the median ratios between two squared errors. We use combinations of the following: FE^2, the squared earnings forecast error, PCE^2, the squared error of using the predictable component as forecast, and OE^2, the squared error of the optimal combination of the predictable component and unpredictable component variables. We calculate these median ratios for each firm separately (of the 316 firms) but pool the ratios of all 18338 forecasters. The total number of observations across all firms is 146319 in the estimation sample and 126651 in the evaluation sample. We calculate some ratios in the evaluation sample twice: once with the weights (used in the construction of the optimal forecast) as estimated in the estimation sample, and once using weights based on the evaluation sample itself.

	Estimation sample with estimation sample weights			Evaluation sample with estimation sample weights			Evaluation sample with evaluation sample weights	
	$\frac{EFE^2}{PCE^2}$	$\frac{OFE^2}{PCE^2}$	$\frac{OFE^2}{EFE^2}$	$\frac{EFE^2}{PCE^2}$	$\frac{OFE^2}{PCE^2}$	$\frac{OFE^2}{EFE^2}$	$\frac{OFE^2}{PCE^2}$	$\frac{OFE^2}{EFE^2}$
Average	0,609	0,669	1,499	0,638	2,878	7,463	0,570	1,123
Median	0,655	0,532	0,877	0,631	1,107	2,165	0,311	0,591
Standard Deviation	0,304	1,206	4,425	0,571	5,703	18,399	3,097	5,330
5% percentile	0,071	0,059	0,279	0,061	0,142	0,642	0,031	0,164
95% percentile	1,031	1,371	3,437	1,207	10,102	33,621	0,943	1,757

Table 12: The results for the regressions to predict better analysts in the evaluation sample using variables in the estimation sample. This is based on 1835 forecasters (since we only include forecasters with a minimum of 10 observations in both sample periods) with a total of 52236 forecasts in the estimation sample and 36403 forecasts in the evaluation sample. We put the data across all firms in one regression. We use two interpretations for what a better earnings forecaster is: a forecaster who has a smaller forecast error compared to the predicted component ("better performing") and a forecaster whose associated optimally constructed forecasts have smaller forecast errors compared to the predicted component error ("having more information"). These might overlap if the forecasters with more information also use them well (so if the optimal forecast is similar to the earnings forecast), but there could also be forecasters that do not use their information well, which is why we separate these measures. In these regressions we use the balanced relative difference: $BRD(x, y) = \frac{x-y}{x+y}$ with x and y being combinations of E (for the earnings forecast error, EFE^2), P (for the predictable component error, PCE^2), O (for the optimal forecast error, OFE^2) and U (for the squared unpredictable component, UC^2). As performance variable, we use $BRD(E, P)$, while we use $BRD(O, P)$ as information variable. The variables to be explained are measured in the evaluation sample, while the regressors are measured in the estimation sample. Standard errors are shown in parentheses.

	Variable to explain			
	$BRD(E, P)$		$BRD(O, P)$	
intercept	-0,174	(0,020)	-0,003	(0,022)
$\frac{UC^2}{PC^2}$	1,152	(0,398)	0,244	(0,440)
BRD(U,P)	-0,098	(0,029)	-0,148	(0,032)
BRD(E,P)	0,407	(0,035)	0,302	(0,038)
BRD(O,P)	0,042	(0,026)	0,266	(0,029)

Table 13: A summary of results on the correlation between three balanced relative difference variables and two unpredictable component variables, calculated per individual forecaster. This is based on 4541 forecasters, with 90190 forecasts in the estimation sample and 28000 in the evaluation sample. We calculate the correlation of the UC variables with three balanced relative difference variables, with the definition $BRD(x, y) = \frac{x-y}{x+y}$ with x and y being combinations of E (for the earnings forecast error, EFE^2), P (for the predictable component error, PCE^2) and O (for the optimal forecast error, OFE^2).

		Correlation with $\|UC\|$			Correlation with UC^2		
		BRD(E,P)	BRD(O,P)	BRD(O,E)	BRD(E,P)	BRD(O,P)	BRD(O,E)
Estimation sample	Average	-0,096	-0,185	-0,116	-0,069	-0,166	-0,121
	Median	-0,125	-0,214	-0,133	-0,124	-0,210	-0,146
	Standard Deviation	0,322	0,279	0,273	0,335	0,278	0,272
	5% percentile	-0,582	-0,598	-0,535	-0,559	-0,556	-0,526
	95% percentile	0,454	0,324	0,359	0,506	0,354	0,358
Evaluation sample	Average	-0,146	-0,122	0,033	-0,125	-0,105	0,030
	Median	-0,173	-0,129	0,049	-0,177	-0,116	0,046
	Standard Deviation	0,477	0,490	0,472	0,481	0,487	0,468
	5% percentile	-0,953	-0,967	-0,877	-0,954	-0,963	-0,863
	95% percentile	0,790	0,852	0,909	0,816	0,835	0,892

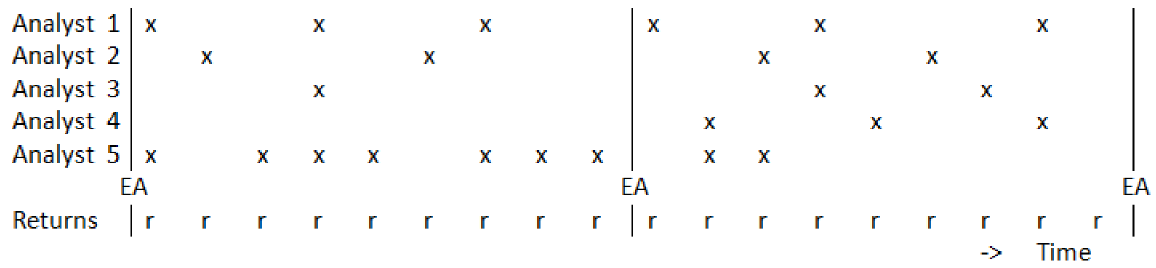

Figure 1: An example of the data format, with an x indicating an earnings forecast and EA indicating when a new yearly earnings announcement takes place. This figure shows for five forecasters for two years a variety of hypothetical patterns of forecasts, including analysts that follow a very regular forecasting pattern, or the opposite, and including forecasters that quit producing forecasts or that join during a later year.

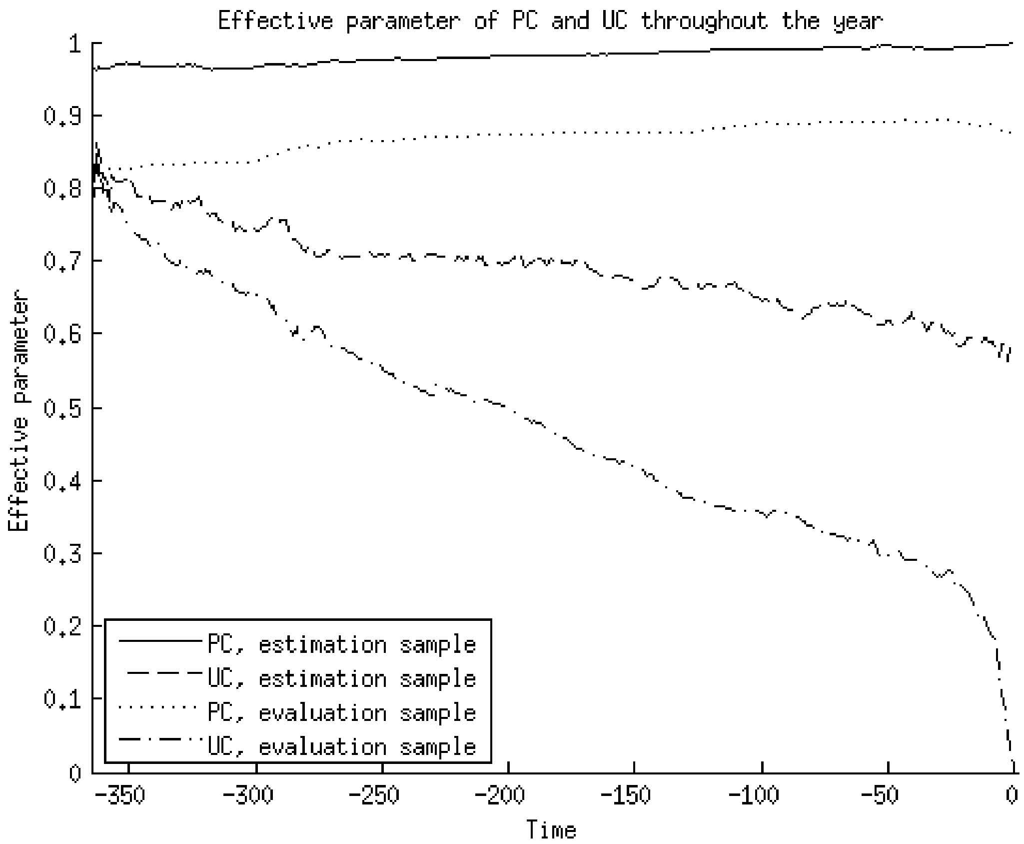

Figure 2: The effective parameter of the predictable and unpredictable component in forecasting the actual earnings throughout the year (with earnings announcement at t=0), after filling in average actual values for the number of forecasts and time until announcement across all firms and all years in the estimation or evaluation sample.

References

Bolliger, G., 2004. The characteristics of individual analysts forecasts in Europe. Journal of Banking & Finance 28, 2283–2309. doi:10.1016/j.jbankfin.2003.09.011.

Brown, L.D., 1993. Earnings forecasting research: its implications for capital markets research. International Journal of Forecasting 9, 295–320. doi:10.1016/0169-2070(93)90023-G.

Brown, L.D., Hugon, A., 2009. Team earnings forecasting. Review of Accounting Studies 14, 587–607.

Clement, M.B., Hales, J., Xue, Y., 2011. Understanding analysts' use of stock returns and other analysts' revisions when forecasting earnings. Journal of Accounting and Economics 51, 279–299.

Clement, M.B., Tse, S.Y., 2005. Financial analyst characteristics and herding behavior in forecasting. The Journal of Finance 60, 307–341.

Cooper, R.A., Day, T.E., Lewis, C.M., 2001. Following the leader: a study of individual analysts earnings forecasts. Journal of Financial Economics 61, 383–416.

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

Kim, Y., Lobo, G.J., Song, M., 2011. Analyst characteristics, timing of forecast revisions, and analyst forecasting ability. Journal of Banking & Finance 35, 2158–2168.

Landsman, W.R., Maydew, E.L., Thornock, J.R., 2012. The information content of annual earnings announcements and mandatory adoption of IFRS. Journal of Accounting and Economics 53, 34–54.

Ramnath, S., Rock, S., Shane, P., 2008. The financial analyst forecasting literature: A taxonomy with suggestions for further research. International Journal of Forecasting 24, 34–75.

Schipper, K., 1991. Analysts forecasts. Accounting Horizons 5, 105–121.

Sheng, X., Thevenot, M., 2012. A new measure of earnings forecast uncertainty. Journal of Accounting and Economics 53, 21–33.

Stickel, S.E., 1990. Predicting individual analyst earnings forecasts. Journal of Accounting Research 28, 409–417.