1-2018

Robust measures of earnings surprises

Jun TU
Singapore Management University, tujun@smu.edu.sg

Chin-Han CHIANG
Singapore Management University, chchiang@smu.edu.sg

Jianqing FAN

Harrison HONG

Jun TU

Citation

Jun TU; CHIANG, Chin-Han; FAN, Jianqing; HONG, Harrison; and TU, Jun. Robust measures of earnings surprises. (2018). *Journal of Finance*. 1-79. Research Collection Lee Kong Chian School Of Business.
Available at: http://ink.library.smu.edu.sg/lkcsb_research/5406

Robust Measures of Earnings Surprises*

Chin-Han Chiang[†] Wei Dai[‡] Jianqing Fan[§] Harrison Hong[¶]

Jun Tu [‖]

First Draft: June 20, 2014
This Draft: January 10, 2016

Abstract

Event studies of market efficiency measure an earnings surprise with the consensus error (CE), defined as earnings minus the average of professional forecasts. However, individual forecasts can be biased. CE and traditional robust statistics such as medians are not robust to such bias. We prove and show empirically that the fraction of forecasts that miss on the same side (FOM), by ignoring the size of the misses, is a more robust parameter-free estimate. It performs better than an approach that estimates and adjusts for individual forecast bias. We bound FOM's efficiency relative to a benchmark where bias parameters are known.

*We thank Bruce Grundy, Hongjun Yan and seminar participants at the European Finance Association Meetings 2015, Western Finance Association 2015 Meetings, China International Finance Conference 2015, Financial Management Association 2015 Meetings, Singapore Mangement University, Seoul National University and Emory University. This paper was previously titled "When Everyone Misses on the Same Side: Debiased Earnings Surprises and Stock Returns". Please address correspondences to Harrison Hong (hhong@princeton.edu) 26 Prospect Avenue, Princeton, New Jersey 08540.

[†]Singapore Management University

[‡]Princeton University

[§]Princeton University

[¶]Princeton University, NBER and CAFR

[‖]Singapore Management University

1 Introduction

Economists historically measure the degree to which the market is surprised by an earnings announcement with the consensus error. It is a simple parametric-free measure defined as the difference between the actual earnings and the consensus forecast, where the consensus is calculated using the mean of the available professional forecasts. The consensus error is widely used in financial markets. For instance, commentaries frequently allude to the extent to which earnings missed the consensus forecast as a key rationale for significant stock price movements. The consensus error is also a building block of event studies on how efficiently markets react to earnings news (see MacKinlay (1997), Lyon, Barber, and Tsai (1999), Kothari and Warner (2004)).

A canonical regression specification in such event studies is that of the cumulative abnormal return of a stock around the earnings announcement date (CAR) or subsequent to the announcement date ($POSTCAR$) on the consensus error (CE): the more positive the consensus error CE the higher is the CAR and also the higher is the $POSTCAR$ (see, e.g., Bernard and Thomas (1990)). These regressions indicate that markets react to earnings news gradually and have become a linchpin in the market efficiency debate (see Fama (1998)).

The ubiquitous use of this measure is premised on the consensus forecast being an unbiased measure of the market's expectation of earnings. But it is well known that a subset of professional earnings forecasts can sometimes be biased for a few reasons. One reason is conflicts of interest. For instance, the analysts of investment banks that have business with a company are likely to issue optimistically biased forecasts compared to peers working for banks without such a relationship (Michaely and Womack (1999), McNichols and O'Brien (1997), Lin and McNichols (1998), Lim (2001), Hong and Kubik (2003)). Another might be that it is optimal for analysts to strategically shade their forecasts, whether positively or negatively, away from their unbiased signal if the rewards to the forecasting tournament are sufficiently convex (see, e.g., Keane and Runkle (1998), Hong, Kubik, and Solomon (2000), DellaVigna and Gentzkow (2009)).

Since many investors, particularly institutions which comprise a significant fraction of the market, attempt to adjust for these strategic forecast biases in forming their earnings expectations (see, e.g., Iskoz (2003), Malmendier and Shanthikumar (2007), Mikhail, Walther, and Willis (2007)), the end result is that the consensus forecast is no longer an unbiased measure of the market's expectation of earnings. In other words, the consensus forecast by averaging in biased analyst forecasts systematically diverges from the true expectation of the market. In the context of the *CAR* and *POSTCAR* regressions, we ideally want an accurate and unbiased measure of the true market surprise on the right-hand side as the explanatory variable. If *CE* as a proxy for the true market surprise has substantial measurement error, this translates into poor explanatory power for *CAR* or *POSTCAR* in these canonical regressions, thereby leaving room for a better measure of the true market surprise.

The challenge from the point of view of the econometrician is how construct this better measure, which has the same advantage of being parametric-free as the consensus error but at the same time takes as given that the econometrician does not have the same information set as institutional investors. The usual robust statistics such as medians or winsorization cannot help much since these statistics are meant to deal with outliers and not systematic bias of forecasts. Importantly, it is difficult in practice to identify ex-ante which of the individual forecasts are compromised, i.e. to estimate the bias of different analysts, as we show below. Otherwise, one could pursue a parametric-dependent method, such as subtracting off the estimated bias parameters from the contaminated individual forecasts before calculating the consensus error.

To deal with this problem, we propose a new market surprise measure that is significantly more robust to such bias — the fraction of forecasts that miss on the same side or *FOM*, which shares the strength of *CE* in that it is parametric-free but is less sensitive to such biased forecasts and hence potentially superior to *CE*. Suppose that there are N forecasts and K is the number of forecasts less than the actual announced earnings A and M is the number of forecasts greater than A. Then the fraction of misses on the same side is given by

$$FOM = \frac{K}{N} - \frac{M}{N},$$

which takes on values between -1 and 1—the higher is FOM the more positive the earnings surprise. For instance, when $K = M$, then $FOM = 0$ and there are equal misses on both sides. When $K = N$ and $M = 0$, then $FOM = 1$ and the actual lies above the range of forecasts, which we will also denote by $I_{Actual>All} = 1$ (0 otherwise). In this case, the market is positively surprised and market returns positive around the announcement date. When $K = 0$ and $M = N$, $FOM = -1$ and everyone misses above the actual, which we also denote by $I_{Actual<All} = 1$ (0 otherwise) and the market is negatively surprised and market returns negative around the announcement date.

We show below by using a simple model that *FOM* better measures the true surprise than *CE* when the bias of some forecasts are potentially large and unobservable to the econometrician. When these biased errors are not a big concern, then *CE* is more accurate than *FOM*. In this model, we discuss why *FOM* is better than a number of alternatives such as using median instead of mean forecasts or winsorization in the presence of outliers. We use earnings forecasts to frame our model and motivate our empirical analysis but the methodology and ideas apply equally to other types of economic forecasts where bias is potentially important.

First, to get an intuitive sense of why *FOM* is better than *CE*, consider the following example. Suppose the market expectations for stock A's and stock B's earnings are both 10 and that there are $N = 6$ analysts making forecasts for each stock. If a significant fraction of the forecasts are negatively biased, one might see forecasts like -11, -10, 9, 10, 11, and 12 for stock A and -10, -10, -10, 9, 10, and 11 for stock B. The large negative forecasts are the biased ones. The mean consensus is 3.5 for stock A and 0 for stock B. Suppose the actual earnings turns out to be 14 for both stock A and stock B. In other words, the true market surprise is 4 and the same for both stocks. But using the mean consensus, we get a *CE* of 10.5 for stock A and 14 for stock B. So using *CE* as a proxy, we would think there is more of a positive surprise in stock B's announcement than in stock A's announcement, which is an incorrect classification.

When we run the regression of *CAR* or *POSTCAR* on *CE*, in which *CE* is supposed to be a proxy for the true earnings surprise, we suffer from measurement error and hence the coefficient on *CE* will be downward biased. However, *FOM* is 1 for stock A and stock B, or $I_{Actual>All} = 1$, which is the correct classification in terms of both stocks A and B having the same true earnings surprise. Hence we expect that a regression of *CAR* or *POSTCAR* on *FOM* to have superior explanatory power relative to *CE*.

Essentially, when some of the forecasts are biased enough, it is better to discard the magnitudes of the misses and to simply count the fraction of misses on the same side. If everyone misses on the same side, we know that even unbiased forecasts missed on the same side as biased forecasts, which is enough to know that the market is truly surprised. Taking into account magnitudes, as the traditional consensus error measure does, when some forecasts are biased leads to sorting on bias as opposed to sorting on the true market surprise.

Second, observe that in the example above, using the median of the forecasts rather than the mean as the consensus does not help the *CE* measure. For stock A, the consensus error using the median is 4.5 for stock A and 14.5 for stock B which is a similarly bad classification as using the mean consensus. Medians deal with outliers but not when a significant fraction of the forecasts are biased. Third, in practice, event studies are implemented using a transformation of *CE* into a cross-sectional decile score from 1 to 10, which we call *Rank(CE)*. The *Rank(CE)* measure deals with outliers and offers a better fit for *CAR* and *POSTCAR* than *CE* (see Hirshleifer, Lim, and Teoh (2009)). But it is nonetheless weaker than our *FOM* measure as these rankings are a form of winsorization and deal with outliers but not biases which significantly affect the *CE* and the relative rankings of stocks that are considered positive or negative surprises.

Fourth, notice that the dispersion of forecasts in this example is also roughly equal for both stocks A and B. As a result, our findings are not driven by differences in the dispersion of forecasts across stocks and we show that this is indeed the case. As long as the fraction of biased forecasts stays constant with N, such biases will remain important regardless of N and we expect our FOM measure to be superior regardless of N.

Using annual forecasts of fiscal year-end earnings, the R^2 of a canonical regression of *CAR* (measured using the 3-day firm-size-adjusted return around the announcement date) on *CE* is 0.30% and on its decile rank score *Rank(CE)* is 2.8%. *CE* is constructed using the mean of the most recent forecasts for the annual year-end FY1 earnings. So every firm has one observation per year over the sample period from 1983 to 2011. A one standard deviation increase in *Rank(CE)* increases the *CAR* by 1.2%, a sizeable economic effect. For *POSTCAR*, the portfolio long positive earnings surprise (decile rank score 10) and short negative earnings surprise (decile rank score 1) yields a return of 1.7% over the subsequent six months (126 trading days to be exact) after the announcement date or 3.5% annualized.

Our *FOM* variable, however, performs better than *CE* or *Rank(CE)*. For instance, *FOM* variable gives an R^2 of 4.1%. A one standard deviation increase in *FOM* increases *CAR* by 1.5%. When we run a horse race of *FOM* and *Rank(CE)*, the coefficient in front of *FOM* is virtually unchanged whereas the one in front of *Rank(CE)* is no longer significant. For the *POSTCAR*, a portfolio long $FOM = 1$ stocks and short $FOM = -1$ stocks yields a six-month subsequent return of 3% or 6% annualized. Again, in a multiple regression to explain *POSTCAR*, our *FOM* measure remains significant, whereas *Rank(CE)* is insignificant.

We verify that our findings remain robust even when controlling for differences in the dispersion of forecasts across earnings events and that *FOM* works more consistently across different sub-samples of analyst coverage. In addition, we show that *FOM* also predicts revisions of the consensus forecast, although not as well as for stock returns, since the consensus forecast includes some biased forecasts which presumably need not adjust since they are driven by reasons other than accuracy. However, to the extent there are unbiased forecasters that adjust and learn from the announcement, we expect *FOM* to be informative about these revisions, which it is.

Finally, we conduct two additional sets of analyses which speak to the efficiency of *FOM* and the importance of having a parametric-free measure. First, we show that a parametric-dependent method involving trying to estimate individual analyst bias parameters based on their forecast histories and adjusting individual forecasts before calculating *CE* yields significantly worse results than even using the parametric-free *CE*, and needless to say much worst than *FOM*. Second, we provide a theoretical lower bound on the efficiency of *CE* and *FOM* relative to a full information benchmark. That is, we can construct an optimal measure of the consensus assuming we know individual bias parameters and precisions. Of course, the relative efficiency of *CE* to this full information benchmark goes to zero when the bias is large. But the lower bound on the relative efficiency *FOM* can be as high as 50% using plausible parameter values.

Our paper proceeds as follows. We present a simple model to contrast the accuracy of our *FOM* measure versus the traditional *CE* measure under various assumptions in Section 2. We describe our data and how we construct our key variables of interest in Section 3. We present our main empirical findings in Section 4. We conclude in Section 5. In the Appendix, we collect proofs. In the Supplementary Internet Appendix, we provide further discussions and extensions of our model to account for various aspects of the data.

2 Modeling the Performance of *CE* versus *FOM*

In this Section, we develop a stylized model to explain why *FOM* might be different from *CE* and *Rank(CE)* in terms of its effectiveness in capturing market surprises. Our argument relies on some fraction of analysts' forecasts being biased but the bias is not known to the econometrician. This is consistent with the empirical studies cited in the Introduction on the incentive reasons for why analyst forecasts might be biased. We are able to obtain analytic solutions and prove that *FOM* is better than *CE* when the bias is large enough, which we use to motivate our empirical work.

But for comparative statics, we need to use numerical calculations and will present these, after presenting the empirical findings, in the Supplementary Internet Appendix on Extensions and Simulations. Moreover, in this baseline model with bias, *CE* and *Rank(CE)* are essentially the same object and our arguments work for both. But in the data, the correlation of *CE* and *Rank(CE)* is very low due to outliers in *CE*. The *Rank(CE)* measure largely takes care of these outliers. We will add this element of outliers to our model in the Supplementary Internet Appendix so as to show that the effectiveness of *FOM* relative to *CE* and *Rank(CE)* extends to a more general setting with outliers.

2.1 Set-up

We start by assuming that actual (which we refer to as earnings through out but could as well be macro-variables like inflation or GDP) is given by

$$A = e + \epsilon_A, \tag{1}$$

where e is the *unobserved* market expectation and $\epsilon_A \sim \mathcal{N}(0, \sigma_A^2)$. The difference between the actual earnings and the market expectation is the market surprise, which is given by

$$S = A - e. \tag{2}$$

We then assume that an individual forecast F_i is the market expectation e plus some noise ϵ_i and a possible bias term b_i:

$$F_i = \begin{cases} e + \epsilon_i & \text{with prob. } \omega_0 \\ e + b_i + \epsilon_i & \text{with prob. } \omega_1 = 1 - \omega_0 \end{cases} \tag{3}$$

where $\epsilon_i \sim \mathcal{N}(0, \sigma_F^2)$ and is uncorrelated with ϵ_A. Each forecast is unbiased with probability ω_0, and is contaminated by an individual bias term b_i with probability $\omega_1 = 1 - \omega_0$.

We model the bias in the following manner. For each set of N forecasts an aggregated bias level $B \sim \mathcal{N}(0, \sigma_B^2)$ is drawn first, and conditional on this realized B individual bias b_i follows $\mathcal{N}(B, \sigma_b^2)$. Note that while ω_0 and ω_1 are fixed and do not change with N, the realized number of biased forecasts can be different from its expectation $\omega_1 N$. Therefore conditional on each set of N forecasts, on average a fraction of ω_1 of them are biased by a random magnitude. Note that we still have $\mathrm{E}[F_i] = e + \omega_1 \mathrm{E}[b_i] = e + \omega_1 \mathrm{E}[B] = e$ because B follows a symmetric distribution around zero.

We can motivate this set-up as the market is able to figure out which forecasts are biased and has access to information about the mean of earnings e beyond simply using analyst forecasts. ϵ_A is the unexpected shock to earnings which the market cannot know. The bias b_i can be derived in a number of ways. The simplest is as in Lim (2001). We show in the Supplementary Internet Appendix an extension where the market's expectation depends only on the analyst forecasts and we can derive similar results.

2.2 Definitions of CE and FOM

To construct our two proxies for market surprise, first note that the forecast error of the i^{th} forecaster is given by

$$U_i = A - F_i = S + Y_i \tag{4}$$

where

$$Y_i \sim \omega_0 \mathcal{N}(0, \sigma_F^2) + \omega_1 \mathcal{N}(b_i, \sigma_F^2) \tag{5}$$

is distributed as a mixture-normal of biased and unbiased forecasts. The consensus forecast error (CE) then is simply the average of the forecast errors:

$$CE = \frac{1}{N} \sum_{i=1}^{N} U_i. \tag{6}$$

In contrast, FOM is an average of the signs of the forecast errors:

$$FOM = \frac{1}{N}\sum_{i=1}^{N} \text{sgn}(U_i), \tag{7}$$

where $\text{sgn}(U_i)$ is 1, 0, or -1, depending whether $U_i > 0$, $U_i = 0$ or $U_i < 0$.

2.3 Unbiased Forecasts Benchmark: $\omega_1 = 0$

We begin with Proposition 1 which describes the unbiased forecasts benchmark. The comparison we want to make is between $\text{Cor}[CE, S]$ and $\text{Cor}[FOM, S]$. We can prove that in the case where N is big, $\text{Cor}[CE, S] > \text{Cor}[FOM, S]$, thereby making CE a better measure of the earnings surprise S than FOM.

Proposition 1: *Assume there is no bias ($\omega_1 = 0$). The correlation of the consensus error CE with the true market surprise S is given by:*

$$\text{Cor}[CE, S] = \frac{1}{\sqrt{1 + r_F^2/N}}, \tag{8}$$

where $r_F = \sigma_F/\sigma_A$ is the ratio between the standard deviation of forecasts and the actual. The correlation of FOM with S is

$$\text{Cor}[FOM, S] = \frac{r_F \text{E}[X \cdot \Phi(X)]}{\sqrt{\text{E}[\Phi(X)(1-\Phi(X))]/N + \text{Var}[\Phi(X)]}}, \tag{9}$$

where $X \sim \mathcal{N}(0, 1/r_F^2)$. CE is more correlated with S than is FOM when there is no bias and $N \to \infty$).

Proof: See Appendix.

It is worth noting that (8) and (9) only depend on N and r_F, namely the number of analysts and the ratio between the standard deviation of the forecasts and the actual (rather than their respective uncertainty levels). It is obvious that when we take $N \to \infty$ in (8) for any given r_F, the correlation between CE and S increases with N and decreases with r_F. Indeed, as N gets large, $\mathrm{Cor}[CE, S]$ goes to 1 as one would expect from the Law of Large Numbers. Also note that (9) goes to $l(r_F) = \frac{r_F \mathrm{E}[X \cdot \Phi(X)]}{\sqrt{\mathrm{Var}[\Phi(X)]}}$ as $N \to \infty$. The limit $l(r_F)$ can be shown to be strictly less than 1.

While it is analytically difficult to prove, we show using numerical calculations that $\mathrm{Cor}[CE, S] > \mathrm{Cor}[FOM, S]$ as one would expect when $\omega_1 = 0$ even for N set at realistic values from data, typically between 5 and 20 forecasts. When there is no bias, there is no reason to ignore the magnitude of misses.

2.4 Biased Forecasts: $\omega_1 \neq 0$

We now show that ignoring the magnitudes of misses is advantageous when forecasts can be biased and this bias potentially large in Proposition 2.

Proposition 2: *Assume there is bias ($0 < \omega_1 < 1$). The correlation of CE with S is given by:*

$$\mathrm{Cor}[CE, S] = \frac{1}{\sqrt{1 + (r_F^2 + \omega_0 \omega_1 r_B^2 + \omega_1 r_b^2)/N + \omega_1^2 r_B^2}}, \tag{10}$$

where $r_B = \sigma_B/\sigma_A$ and $r_b = \sigma_b/\sigma_A$ are the ratios between the standard deviation of the aggregate and individual bias to the actual, respectively. $\mathrm{Cor}[CE, S]$ *goes to* 0 *as r_B gets large, while* $\mathrm{Cor}[FOM, S]$ *is always bounded from below by*

$$2\omega_0 r_F \mathrm{E}[X \cdot \Phi(X)] > 0 \tag{11}$$

where $X \sim \mathcal{N}(0, 1/r_F^2)$. Hence, FOM is more correlated with S than CE when r_B the bias $\to \infty$.

Proof: See Appendix.

To get an intuition for this result, notice that the first claim follows from (10) by letting the bias distortion parameter $r_B \to \infty$. On the other hand, the lower bound for $\mathrm{Cor}[FOM, S]$ is given by $l(r_F, \omega_0) = 2\omega_0 r_F \mathrm{E}[X \cdot \Phi(X)] > 0$. Although very crude, it does not involve the bias components: no matter how bad the bias can be, at least a fraction of useful information is preserved. Since $\mathrm{Cor}[CE, S]$ goes to 0 with increasing r_B, for any value of $\omega_1 \in (0, 1)$ it will decrease to below $l(r_F, \omega_0)$ for a large enough r_B. That is, whatever value other parameters take, *FOM* will eventually outperform *CE* as the level of bias distortion increases.

3 Data

The data on analysts' earnings forecasts are taken from the Institutional Brokers Estimate System (I/B/E/S). We conduct our analysis using the Unadjusted Detailed files. We focus on forecasts of the fiscal year-end earnings (FY1) from 1983 to 2011. Stock returns, prices, and number of outstanding shares are drawn from the Center for Research in Securities Prices (CRSP) Daily Stocks file. The forecast data are merged with actual earnings obtained from I/B/E/S and the daily stock price data from CRSP. Observations are dropped if forecast data, earnings data, or stock data are missing.

For each analyst in a given forecast period, we restrict every forecast to be made within 90 days to the annual earnings announcements. If an analyst makes more than one forecast within 90 days to the earnings announcement, we keep the latest forecast before the earnings announcements. In some records, the revision date precedes the original forecast date, which is considered an error on the part of I/B/E/S. In this case, we use the original forecast date. We then calculate the mean, standard deviation, median, minimum and maximum value of these individual forecasts for each stock in a given fiscal period. In addition, the FY1 earnings announcements need to fall between 15 to 90 calendar days following the fiscal period end date. Otherwise, we drop the observations.[1] We remove penny stocks with a price of less

[1]We also consider forecasts of quarterly earnings of the same sample period as a robustness exercise. For each analyst in a given forecast period, we keep the latest forecast before the quarterly earnings announce-

than \$5. To control for stock splits, we delete observations where the number of shares outstanding at date t when the variables are calculated is larger than the number of shares 20 days prior to the earnings announcement. We also require an earnings announcement to have at least 5 analyst forecasts.

Following the literature, we define consensus error (*CE*) as the difference between the actual FY1 earnings and the consensus forecast scaled by the stock price 20 days prior to the earnings announcement (*price(-20)*). We consider both mean consensus (arithmetic mean across individual forecasts) and median consensus (50th percentile of individual forecasts) in formulating *CE*. We sort *CE* into deciles and assign a rank score from 1 to 10 to *CE* based on mean consensus. As for *CE* based on median consensus, which has a value of 0 for over 20% of the data, we apply a more coarse sort by ranking *CE* into only 6 groups. Analyst forecast dispersion (*DISP*) is defined as the standard deviation of analyst forecasts scaled by *price(-20)*. We further sort *DISP* into deciles and assign a rank score from 1 to 10 to each batch (*Rank(DISP)*).

We use two indicator functions, $I_{Actual<All}$ and $I_{Actual>All}$, to denote when all analysts completely miss on the same side. $I_{Actual<All}$ equals 1 if the minimum forecast is higher than the actual earnings. In this case, all analysts are being too positive and make forecasts higher than the actual earnings. In contrast, $I_{Actual>All}$ equals 1 if all analysts are too pessimistic and the maximum forecast is lower than the actual earnings.

The fraction of misses (*FOM*) is defined as follows:

$$FOM = \frac{K}{N} - \frac{M}{N}, \tag{12}$$

where K is the number of forecasts strictly smaller than the actual earnings, and M is the number of forecasts strictly greater than the actual earnings. N is the total number of analyst forecasts for stock i. Notice that $K + M$ does not necessarily equal N. By construction, *FOM* equals 1 if $I_{Actual>All}$ is 1 and -1 if $I_{Actual<All}$ is 1.

ments. Relevant summary statistics based on qualified quarterly forecasts are then calculated.

Using CRSP, we calculate cumulative abnormal returns (CAR) as follows:

$$CAR(i,y) = \prod_{t=t_0}^{t_1} (1 + (R(i,t) - R(p,t))) - 1, \tag{13}$$

where $R(i,t)$ is the daily returns of stock i on date t around an earnings announcement in year y. The window to calculate the cumulative abnormal return begins at date t_0 and ends at date t_1. $R(p,t)$ is the daily return of the size portfolio to which stock i belongs. The size deciles are based on CRSP Portfolio Statistics Capitalization Deciles file.

We concentrate on two time windows relative to earnings announcements. The first are returns cumulated over the three-day window from one trading day before until one day after the earnings release date (CAR). The second is the cumulative post-announcement returns ($POSTCAR$) using trading days +2 to +126 after the earnings announcement.

Table 1 provides the summary statistics of the variables. In Panel A, notice that the *CE* (using the mean consensus) has a mean of -0.0031, consistent with the positive bias in the consensus forecast found in the literature, and a standard deviation of 0.043. The *CE* using the median consensus has similar magnitudes. *Rank(CE)* using mean consensus has a mean of 5.49 and a standard deviation of 2.87. *Rank(CE)* using median consensus has a mean of 3.46 and a standard deviation of 1.7.[2] Our *FOM* has a mean of .1454 and a standard deviation of .7. $I_{Actual<All}$ has a mean of 0.125 and $I_{Actual>All}$ has a mean of 0.2. In other words, everyone misses on the same side for 32% of the earnings announcement observations in our sample. Moreover, notice that *CE* based on either median or mean consensus have a correlation of 0.8447 with each other.

In Figure 1, we plot the distribution of *FOM* across the entire sample. On the x-axis are the bins for various values of *FOM*. Notice that 12% of our sample is in the -1 bin (which denotes $I_{Actual<All}$) and 20% in the 1 bin (which denotes $I_{Actual>All}$). For the bins in the

[2]The standard deviation of the *Rank(CE)* using the median consensus is smaller because as we noted above we only use 1-6 groups as opposed to 1-10 deciles. The reason is that the median consensus has 20% of the observations concentrated at 0.

middle, we have a bin width of 0.25. The bins with positive *FOM*'s have 10% each of the observations. The bins with negative *FOM*'s have a somewhat smaller representation at 5% each.

In Figure 2, we show the *FOM* distribution conditional on the number of analysts N. When $N = 5$ to $N = 9$, which represents 53% of the sample observations, the fraction of observations where everyone misses on the same side is 39%. The analogous number for $N = 10$ to $N = 19$, which is 35% of the sample, is 27%. The number for $N \geq 20$, which is 12% of the sample, is 19%. In all these situations, the fraction of everyone missing on the same side is a non-trivial fraction of the observations.

In Figure 3, it is also interesting to see that the time series of misses on the same side varies over our period of study from 1983 to 2011. While the total misses on the same side is consistently high at 30%, the misses all above the actual have been declining, while the misses all below the actual have been increasing.

In Panel B of Table 1, we report the correlation matrix for our variables of interest in which *CE* is based on the mean consensus. Notice that the correlation of *CE* with *Rank(CE)* is 0.28 and the correlation of *FOM* and *CE* is .24. As we show in Section A.5 of the Supplementary Internet Appendix, *CE* has fat-tails which drive down these correlations. The correlation of *FOM* with *Rank(CE)* is higher at 0.81 but it is far from perfectly correlated. As a result, it will be interesting to see which of these two measures is more informative for stock returns. We will in our extended model below try to capture this difference in correlations which is absent from our baseline model above. This distinction is not relevant for our main empirical results but we show that our model can be extended to capture this additional feature of the data. In any event, *FOM* will have different information about market surprises than *CE* and *Rank(CE)*. Results in Panel C using the median as the consensus forecast are similar. One thing to note is that the outliers make *CE* not as effective a measure of market surprises as *Rank(CE)*. But *FOM* does better than both as we show below.

It is also useful to do a decomposition of *FOM* on firm and time characteristics. In a table which we omit for brevity, we report the R^2 of *FOM* regressions on firm and year dummies. We consider three models: (1) *FOM* on firm dummies only, (2) *FOM* on year dummies only, and (3) *FOM* on firm and year dummies. The R^2 for specification (1) is 0.12, for specification (2) is 0.045, and for specification (3) is 0.14. In other words, firm fixed effects or year dummies explain little of the variation in *FOM*. *FOM* is mostly driven by idiosyncratic events, consistent with the premise of our model.

4 Empirical Findings

4.1 *FOM* and *CAR*

In Table 2, we run the canonical earnings announcement event study regression with *CAR* as the dependent variable and various permutations of *CE*, *FOM*, $I_{Actual<All}$, and $I_{Actual>All}$ as independent variables. All regressions include Year Dummies. In column (1), we see that the coefficient on *CE* is positive as expected but the R^2 is low at 0.3%. In column (2) *FOM* attracts a coefficient of 0.0210 with a t-statistic of 33.86 and an R^2 of 4.1%. A one standard deviation increase in *FOM* increases *CAR* by 1.5%, which is a sizeable 3-day move in stock returns. Using the everyone-misses-on-the-same-side measures, we find that the coefficient in front of $I_{Actual<All}$ is as expected negative with a coefficient of -0.021 and a t-statistic of -15.21. For $I_{Actual>All}$, it is positive at 0.0265 with a t-statistic of 24.37. The market's reaction is fairly symmetric when everyone misses on the same side, whether it is too high or too low. Again, the market reactions are sizeable — roughly a 2.1% decrease in stock prices over 3 days when all analysts miss too high and a 2.6% 3-day increase when all analysts miss too low. In column (4), we find that *FOM* is far more informative for *CAR* than *CE* when we put both variables together in a multiple regression. The coefficient of *CE* goes to zero while the coefficient in front of *FOM* is unchanged. It is in this sense that *FOM* dominates *CE*. The same holds true in column (5) when we compare *CE* to the everyone-misses-on-the-same-side indicators.

In columns (6)-(10), we repeat the same specifications using the median of the forecasts as the proxy for the consensus. In every case, the results are virtually unchanged. Using the median consensus does not help for the reasons we gave above in that the issue is not so much outliers but systemic bias which the median or winsorization more generally cannot solve.

In Panel B, we compare the relative power of *Rank(CE)* and *FOM* for explaining *CAR*. In column (1), we find that *Rank(CE)* attracts a coefficient of 0.004 with a t-statistic of 27.15. The R^2 is 0.028. As expected, it performs much better than *CE* because *CE* has outliers. The coefficient in column (2) for *FOM* is the same as that of Panel A with an R^2 of 0.041, which is higher than that of *Rank(CE)*. The coefficients in front of the everyone-misses-on-the-same-side indicators in column (3) are the same as in Panel A. In column (4), when we combine both of these explanatory variables, we see that the coefficient in front *FOM* is largely unchanged, increasing from 0.021 to 0.0213 with a t-statistic of 20.45. The coefficient for *Rank(CE)* is close to zero and is no longer significant. Moreover, the R^2 remains the same as when *FOM* is by itself in the regression. In column (5), we find that adding in a horserace of *Rank(CE)* with the everyone-misses-on-the-same-side indicators, *Rank(CE)* retains more explanatory power. This indicates that *FOM* does better than *Rank(CE)* not only when everyone misses on the same side. There is also information not captured in *Rank(CE)* in intermediate values of *FOM*.

In columns (6)-(10), we consider *Rank(CE)* but using the median forecast as the consensus forecast. The coefficient in front of *Rank(CE)* is 0.008 with a t-statistic of 30.84 and an R^2 of 0.035, which is better than *Rank(CE)* using mean forecasts. But when we combine *Rank(CE)* with *FOM*, we see again that the coefficient in front of *Rank(CE)* falls to 0.00218 with a t-statistic of 0.0028, while the coefficient in front of *FOM* is 0.0164 with a t-statistic of 13.89.

One way to compare the economic magnitudes is to ask how a one standard deviation increase in *Rank(CE)* or *FOM* increases the *CAR*. For *Rank(CE)*, its standard deviation is 1.7, while for FOM, it is .72. The implied *CAR* effect of *Rank(CE)* is just 0.0038 compared to the implied *CAR* effect for *FOM*, which is 0.014. The *FOM* effect is 3 to 4 times as large as the *Rank(CE)* effect using the median forecast for the consensus. It is not surprising that the R^2 does not change much from the univariate *FOM* specification when we add *Rank(CE)*. In column (10), we combine the *Rank(CE)* and the everyone-misses-on-the-same-side measures. The results are similar to the case of the mean consensus. So overall, while using a median consensus in conjunction with taking a rank of *CE* improves performance, *FOM* is still the best univariate measure by a substantial margin. This will become even more apparent when we consider *POSTCAR* next.

But before then, it is helpful to visualize these regressions in Figure 4, where we plot the average *CAR* for different values of *Rank(CE)* and in Figure 5, where we plot the average *CAR* for the different bins of *FOM*. Notice that an effective earnings surprise measure should generate a strong positive monotonic relationship between the measure on the x-axis and *CAR* on the y-axis. In both cases, we see an upward sloping sloping curve. But *FOM* actually generates a much bigger spread in *CAR* than *Rank(CE)*—from bin -1 to bin 1, we see a movement in the *CAR* of -0.021 to 0.0265, consistent with our estimates for the everyone-misses-on-the-same-side indicators from Table 2. In contrast, *Rank(CE)* only generates an analogous movement from decile 1 to decile 10 of -0.015 to 0.02 in *CAR*. Also, *Rank(CE)* generates a much more muted increase in *CAR* for deciles scores 1 to 3.

4.2 *FOM* and *POSTCAR*

In Table 3, we have as the dependent variable *POSTCAR*. In Panel A, we compare *FOM* to the unranked *CE*. In column (1), we see that *CE* again attracts a positive coefficient of .606 but is not statistically significant. In column (2), the coefficient in front of *FOM* is 0.0135 with a t-statistic of 6.27. In column (3), we see that the coefficients in front of the indicators when everyone misses on the same side are -0.0137 with a t-statistic of 2.87 and 0.0151 with a t-statistic of 3.83. These two coefficients are economically interesting since we can interpret these as the returns of shorting a portfolio where everyone misses too high (negative surprise) and longing a portfolio where everyone misses too low (positive surprise). The six-month return is 3%, which translates to an annualized return of 6%, quite an economically interesting magnitude. When we run the multiple regression, we see that *FOM* is more informative about *POSTCAR* than *CE*. The coefficient in front of *CE* gets cut dramatically, while the coefficient in front of *FOM* is virtually unchanged. In column (4), we find that *FOM* best explains *POSTCAR*. A similar conclusion holds in column (5) with the everyone-misses-on-the-same-side indicators. In column (6)-(10), we use the median forecast to create *CE* and find virtually identical results.

In Panel B, we compare *FOM* to the *Rank(CE)* using means and medians for explaining the *POSTCAR*. In column (1), we see that *Rank(CE)* comes in significantly with a coefficient of 0.00188 and a t-statistic of 3.45. Columns (2) and (3) are the same as to those in Panel A. In column (4), where we combine *Rank(CE)* and *FOM*, *Rank(CE)* has the wrong sign, while the *FOM* is even more significant and in the right direction. The coefficient is 0.0229 with a t-statistic of 5.98. So here moving from an *FOM* of -1 to 1 would lead to an increase in the *POSTCAR* of nearly 5% per six months or nearly 10% annualized. In column (5) where we examine how the indicators of everyone-missing-on-the-same-side do compared to *Rank(CE)*, we see that *Rank(CE)* is no longer significant and the coefficient in front of the indicators are virtually unchanged. In columns (6)-(10), we use the median forecast to construct *Rank(CE)* instead of the mean forecast and find very similar results.

To visualize these *POSTCAR* regressions, we show in Figure 6 the average *POSTCAR* for different values of *Rank(CE)* and in Figure 7 the average *POSTCAR* for the different bins of *FOM*. Again, we want our earnings surprise measure to generate a monotonic or upward sloping *POSTCAR*. Notice that *FOM* generates a much more upward-sloping and monotonic *POSTCAR* than *Rank(CE)* and also generates a much more sizeable spread in *POSTCAR*, consistent with Table 3.

4.3 Controlling for Dispersion of Forecasts

In Table 4, we add into our baseline regression specifications the dispersion of analysts' forecasts (*DISP*) and *CE* interacted with *DISP* to see if more complicated models of *CE* might take away the explanatory power of *FOM*. The idea is that the effect of *CE* on returns is lower when there is more uncertainty or disagreement in the forecasts.

Note that we have already established the power of *FOM* over *CE* and *Rank(CE)* in all cases. It is interesting to nonetheless consider whether more complicated *Rank(CE)* models might attenuate the univariate power of *FOM*. More precisely, we implement our regression using *Rank(DISP)* and *Rank(CE)*. This is indeed what we find since the coefficient in front of the interaction term with $DISP$ is negative. However, the coefficients on *FOM* are little changed from before. The coefficients in front of $I_{Actual<All}$ and $I_{Actual>All}$ are also significant. This is true for mean and median consensus forecasts. The overall picture is that *FOM* remains significant throughout.

4.4 Cuts by Analyst Coverage

In Table 5, we run our baseline specifications for sub-groups of stocks with different levels of analyst coverage. Recall that we require a minimum of 5 analysts to begin with. We divide our sample into 4 groups: 5 to 9 analysts, 10 to 14 analysts, 15 to 19 analysts and greater than or equal to 20 analysts. In Panel A, we consider the mean consensus. In the first row, we see that the effect of *Rank(CE)* is fairly similar across all the sub-groups. *FOM* also has fairly similar effects for all the sub-groups in the second row. But notice that in each case, the R^2 of FOM is higher than that of *Rank(CE)*. So the baseline effects that we established are not concentrated in a particular sub-group of stocks. The same applies for the everyone-misses-on-the-same-side indicators in the third row. In fourth row, we run a horse race between *Rank(CE)* and *FOM* and find again that *Rank(CE)* is not significant in any of the sub-groups once we have *FOM* in the regression. The coefficients on *FOM* are in contrast unchanged. In the fifth row, we run a horse race of *Rank(CE)* and the everyone-misses-on-the-same-side indicators. We obtain similar effects to the baseline; *Rank(CE)* is weakened but not as much as if we had *FOM* in the regression.

In Panel B, we conduct the same analysis but now using the median consensus to calculate *Rank(CE)*. Our conclusions are largely similar. Interestingly, focusing on row (4) where we run a horserace between *Rank(CE)* and *FOM*, we see that *FOM* does much better and *Rank(CE)* is insignificant except for the N equals 5 to 9 case. In other words, recall from our Table 2 Panel B that the *Rank(CE)* measure constructed by taking the median forecast as the consensus did slightly better compared to the *Rank(CE)* measure using the mean forecast as the consensus when compared to *FOM*. Whereas *Rank(CE)* using the mean forecast as the consensus was entirely wiped out in the horserace, *Rank(CE)* using the median forecast as the consensus survived a bit though the *FOM* effect was three times as big. We see here that this differential was coming only from the group with the fewest analysts. For N big, *FOM* is much better, which fits with the intuition we developed in the model. When there is a big N, if everyone misses on the same side, it is very indicative that there was a big surprise since even the unbiased forecasts are also missing on the same side. Recall that the fraction of biased forecasts stays constant with N in our model and hence bias remains just as important for N big as N small.

In Table 6, we consider the same exercise but using *POSTCAR*. Here the results are noisier but we can still discern that *FOM* is much more robust than *Rank(CE)* in explaining *POSTCAR*. In Panel A, we again use the mean forecast to calculate *Rank(CE)*. Notice in the first row, *Rank(CE)* is only sporadically significant across the four sub-groups. *FOM* in the second row is much more consistent in its performance. In the third row, the indicators for everyone-missing-on-the-same-side are also less consistent compared to *FOM*. In the fourth row, we see that *FOM* takes out the significance of *Rank(CE)* in explaining *POSTCAR*. The only significant coefficient for *Rank(CE)* goes the wrong way in the first sub-group.

However, for N greater than or equal to 20, even *FOM* has limited explanatory power. So most of the power of *FOM* is coming from stocks with fewer analysts. This is not surprising since there is far less drift in stock prices for stocks with more analyst coverage to begin with as these stocks are likely to be better arbitraged (see DellaVigna and Pollet (2009)) Of course, also note that *FOM* does work very well for the sub-group with lots of analysts using *CAR* since this captures the reaction of the market to the surprise. The *POSTCAR* is the delayed reaction or inefficiency in the market. In Panel B, we reach very similar conclusions for the *POSTCAR*.

4.5 *FOM* and Revision of Consensus

In Table 7, we compare the relative performance of *Rank(CE)* and *FOM* in explaining revisions of analysts expectations (between two adjacent fiscal years) in the same direction as market returns. In other words, if both *Rank(CE)* and *FOM* are picking up surprises, we should see that positive surprises are followed by positive revisions of the consensus forecast. But when it comes to comparing which is more powerful, any conclusion becomes more involved since we know from our analysis that a subset of analyst forecasts are biased and that these biased forecasts influence the consensus. So it really also depends on how the biased analysts revise their expectations, which is difficult to say. In any event, since part of the consensus is unbiased and similar to the market, we expect *FOM* to still have power to predict the revision of the consensus. This is indeed what we find. If we look at the economic significance of the coefficients in front of *Rank(CE)* and *FOM* and perform our comparative statics of a one standard deviation shock to these two variables and see what it implies for the consensus revision, we still find that *FOM* is stronger than *Rank(CE)* in both Panels A and B. But the difference is far smaller than when it comes to predicting stock returns. In sum, it is comforting that *FOM* and the everyone-misses-on-the-same-side indicators are picking up revisions of the consensus.

4.6 Additional Robustness Exercises

In our baseline results, we focus on forecasts for year-end earnings that has to be within 90 days before the announcement date. We obtain similar results when there are no such screens, using quarterly instead of annual year-end earnings forecasts, and where the benchmark excess return is the Daniel, Grinblatt, Titman, and Wermers (1997) returns accounting for size, book-to-market and momentum. These tables are omitted for brevity and can be obtained from the authors.

4.7 Parametric-Dependent Alternative

Up to this point, we have focused on parametric-free measures, but one can consider a parametric-dependent alternative. A natural approach is to find the optimal weights, which minimize the square consensus error, for each individual forecast. To do so, we start by breaking down square consensus error into bias and variance:

$$SCE = \left(\sum_{i=1}^{N} \omega_i b_i\right)^2 + \sum_{i=1}^{N} \omega_i^2 \sigma_i^2 = \mathbf{w}\mathbf{b}\mathbf{b}'\mathbf{w}' + \mathbf{w}'\mathbf{D}\mathbf{w} \tag{14}$$

where SCE is the square consensus error, ω_i is the optimal weight of individual forecast i, b_i is the bias of individual analyst i, and σ_i^2 is the square debiased forecast error of analyst i, and $D = diag\{\sigma_1^2, \sigma_2^2, ..., \sigma_n^2\}$. In this setting, we allow for heterogeneous individual forecast precision as opposed to the above baseline model which assumed these were all equal.

To find w_i, consider the following Lagrange function:

$$L(\lambda) = SCE - \lambda \mathbf{w}'\mathbf{1}. \tag{15}$$

By taking the first order derivative with respect to $\mathbf{w}$, we get

$$2(\mathbf{b}\mathbf{b}' + \mathbf{D})\mathbf{w} - \lambda \mathbf{1} = 0, \tag{16}$$

which we can then use to solve for $\mathbf{w}$ as

$$\mathbf{w} = \frac{\lambda}{2}(\mathbf{bb}' + \mathbf{D})^{-1}\mathbf{1}. \tag{17}$$

By applying the Sherman-Morrison formula, we can replace $(\mathbf{bb}' + \mathbf{D})^{-1}$ with

$$(\mathbf{bb}' + \mathbf{D})^{-1} = \mathbf{D^{-1}} - \frac{\mathbf{D^{-1}bb'D^{-1}}}{1 + \mathbf{b'D^{-1}b}}, \tag{18}$$

and $\mathbf{w}$ then becomes

$$\mathbf{w} = \frac{\lambda}{2}\mathbf{D^{-1}1} - \frac{\lambda}{2 + 2\mathbf{b'D^{-1}b}}\mathbf{D^{-1}bb'D^{-1}1}. \tag{19}$$

Therefore, for each ω_i we have

$$\omega_i = \frac{\lambda}{2}\left(1 - \frac{b_i \sum_{k=1}^{N} b_k \sigma_k^{-2}}{1 + \sum_{k-1}^{N} b_k^2 \sigma_k^{-2}}\right)\sigma_i^{-2}, i = 1, 2, ..., N. \tag{20}$$

We can then calculate the optimal-weighted consensus as $\sum_{i=1}^{N} \omega_i F_i$ and optimal-weighted consensus error (optimal-weighted CE).

Empirically, we estimate individual forecast bias (b_i) by averaging their forecast errors over the past T forecasting periods. That is,

$$b_i = \frac{1}{T}\sum_{t=-T}^{-1}(F_{i,t} - Actual_{i,t}). \tag{21}$$

We require each analyst to have at least 2 forecasts in the past 5 years. If an analyst does not have a forecast history long enough (less than 2 observations over the past three years) to estimate b_i, we replace it with the mean bias, which is the sample average taken across all available b_i in fiscal year t. The variance of analyst i (σ_i^2) is the individual square forecast error:

$$\sigma_i^2 = (F_i - Actual_i - b_i)^2. \tag{22}$$

We repeat the same regressions as in Subsection 4.1 and 4.2 by replacing *CE* and *Rank(CE)* with optimal-weighted *CE* and *Rank(CE)*. The results are reported in Table 8 with regard to earnings announcement returns. First, notice that in Panel A, the parametric-dependent *CE* is not statistically significant and performs actually worst than simply using *CE*. The reason of course is that the bias and precision of the individual forecasts are estimated with error and this can negate the advantage of trying to optimally adjust for these parameters in calculating the consensus forecast. The point estimate on *FOM* in column (2) is nearly identical to that of Table 2 but with a smaller t-statistic because we have fewer observations as we require 5 years of initial data to calculate the bias and precision parameters. In Panel B, we use the parametric-dependent *Rank(CE)* instead of *Rank(CE)* and we obtain similar conclusions as before. In sum, the difficulty of estimating bias and precision points to the value of a parametric-free method.

4.8 Relative Efficiency

In this final section, we provide a lower bound on the efficiency of *FOM* relative to a benchmark where we assume that the bias and precision parameters governing individual forecasts are known. Recall that the earning surprise $S \sim \mathcal{N}(0, \sigma_A^2)$ and the error of the i^{th} forecast is

$$U_i = A - F_i = S + Y_i, \qquad Y_i \sim \omega_0 \mathcal{N}(0, \sigma_F^2) + \omega_1 \mathcal{N}(b_i, \sigma_F^2),$$

where $\omega_1 = 1 - \omega_0$, and b_i's are drawn from $\mathcal{N}(B, \sigma_b^2)$ conditional on the realization of $B \sim \mathcal{N}(0, \sigma_B^2)$.

To probe the unobservable surprise S, we look for a measure that is highly correlated with it. The ideal measure should be a function of U_i's that satisfies the following,

$$f^* = \arg\max_f |\mathrm{Cor}(f(U_1, U_2, \cdots, U_N), S)|. \tag{23}$$

It is well-known that f^* is the nonlinear projection of S onto the space of U_i's:

$$f^*(U_1, U_2, \cdots, U_N) = \mathrm{E}[S|\mathbf{U}] \qquad \text{with } \mathbf{U} = (U_1, U_2, \cdots, U_N). \tag{24}$$

This conditional expectation can be written as

$$f^*(\mathbf{U}) = \mathrm{E}[S|\mathbf{U}] = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} Sg(S, B, \mathbf{U})dSdB}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(S, B, \mathbf{U})dSdB}, \tag{25}$$

where $g(S, B, \mathbf{U})$ is the joint density of S, B and $\mathbf{U}$. It can be derived by integrating out $\{b_i\}$ such that

$$g(S, B, \mathbf{U}) = C\phi(S; \sigma_A^2)\phi(B; \sigma_B^2)\prod_{i=1}^{N}\left[\omega_0\phi(U_i - S, \sigma_F^2) + \omega_1\phi(U_i - S - B, \sigma_F^2 + \sigma_b^2)\right], \tag{26}$$

where C is a normalization constant, and $\phi(x; \sigma^2)$ denotes the density function of $\mathcal{N}(0, \sigma^2)$.

Note that in the previous section, we derived an optimal choice in the linear class of functions and demonstrated that it did worse than using the parametric-free *CE*. The optimal choice $f^*(\mathbf{U})$ in this more general derivation is highly nonlinear and cannot be analytically derived. It depends on many unknown parameters (ω_0, σ_B^2, σ_b^2, σ_F^2 and σ_A^2).

But we can nonetheless provide a lower bound of the efficiency of our simple parameter-free measure *FOM* relative to this optimal measure $f^*(\mathbf{U})$. Our derivation assumes that, for a given set of N forecasts, we can condition on both $\mathbf{U}$ and additional information on the subset of forecasts that are unbiased. We call this subset of unbiased forecasts $\mathcal{A}$ and denote the size of this set by $|\mathcal{A}|$, which is an integer between 0 (no forecasts are unbiased) and N (all forecasts are unbiased).

That is, we are now considering any functions of $h(\mathbf{U}, \mathcal{A})$ (which also includes $f^*(\mathbf{U})$ as a special case) to find a function that correlates better with S. If we add more information (the unbiased subset A), the maximal achievable correlation with S using a function of $\mathbf{U}$ and $\mathcal{A}$ should be larger:

$$\mathrm{Cor}[f^*(\mathbf{U}), S] \leq \max_h \mathrm{Cor}[S, h(\mathbf{U}, \mathcal{A})]. \tag{27}$$

We can calculate analytically the maximal correlation of the optimal h with S assuming we can condition also on $\mathcal{A}$. We will compare the relative efficiency of *FOM* to this maximal correlation and it is in this sense that this is a lower bound since $f^*(\mathbf{U})$ need not generate this maximal correlation.

Proposition 3. *Let $f^*(\mathbf{U})$ be the optimal measure of the forecast error assuming all parameters are known. The relative efficiency of FOM to the optimal but infeasible measure $f^*(\mathbf{U})$ is given by*

$$\frac{\mathrm{Cor}[FOM, S]}{\mathrm{Cor}[f^*(\mathbf{U}), S]}. \tag{28}$$

In the limit when bias approaches ∞,

$$\begin{aligned} \lim_{\sigma_B^2 \to \infty} \frac{\mathrm{Cor}[FOM, S]}{\mathrm{Cor}[f^*(\mathbf{U}), S]} &\geq \lim_{\sigma_B^2 \to \infty} \frac{\mathrm{Cor}[FOM, S]}{\max_h \mathrm{Cor}[S, h(\mathbf{U}, \mathcal{A})]} \\ &= \frac{2\omega_0 \mathrm{E}[Z\Phi(Z/r_F)]}{\sqrt{\mathrm{E}\left[\frac{1}{1+r_F^2/|\mathcal{A}|}\right]}}, \end{aligned} \tag{29}$$

where $\mathcal{A}$ is the subset of unbiased forecasts in a given draw of N forecasts and $|\mathcal{A}|$ is the size of the $\mathcal{A}$. In contrast, the relative efficiency of CE can be seriously compromised in the presence of bias,

$$\lim_{\sigma_B^2 \to \infty} \frac{\mathrm{Cor}[CE, S]}{\mathrm{Cor}[f^*(\mathbf{U}), S]} = 0.$$

The numerator of the third term in (29) can be evaluated analytically.[3] Using this and the fact that $|\mathcal{A}| \sim \mathrm{Binomial}\,(N, \omega_0)$ we can evaluate (29), which only depends on N, r_F and ω_0.

[3]By taking derivative w.r.t. r_F, we have $\frac{d}{dr_F} E[Z\Phi(Z/r_F)] = -\frac{1}{r_F^2} E[Z^2\phi(Z/r_F)] = -\frac{1}{\sqrt{2\pi}} \cdot \frac{r_F}{(1+r_F^2)^{3/2}}$. Integrating the derivative back, we have $E[Z\Phi(Z/r_F)] = \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{1+r_F^2}}$.

Table 9 summarizes the numerical values of the lower bound for $N = 10$, which is a typical number of forecasts, and various values of w_0 (the fraction of unbiased forecasts) and r_F (the ratio of the standard deviation of analysts' forecast errors to the standard deviation of the earnings shock). The relative efficiency increases with w_0, holding fixed r_F. The relative efficiency increases with r_F holding fixed w_0. Recall again that this lower bound assumes we know exactly which subset of forecasts are unbiased. But even against this unrealistic benchmark, *FOM* can be relatively efficient when w_0 is not too small, that is when not too many forecasts are biased, and r_F low, that is when analyst forecasts errors are not too variable relative to the actual earnings uncertainty. Reasonable values for w_0 ought to be between 0.7 to 0.9. That is, 10% to 30% of the analysts forecasting are biased from incentives or conflicts of interest. For r_F, parameters values around .5 to 2 seem plausible as the standard deviation of analyst forecasts are on par with how much actual uncertainty there is about the stock. At these parameter values, the lower bound is .5. That is, *FOM*, the parametric-free measure achieves roughly 50% of the efficiency as a measure conditioning on knowing all the parameter values and also the set of unbiased forecasts for any given draw.

5 Conclusion

An important part of event studies of earnings announcements is capturing whether or not the market is surprised. The traditional measure is the difference between realized earnings and the consensus forecast, defined as the average or median of individual forecasts. We argue, however, that the fraction of forecasts that miss on the same side does a superior job of explaining stock returns than the consensus error. We develop a model to show that the reason is that when analysts forecasts are biased the consensus forecast is more sensitive to this bias than the fraction of same-sided misses. While our paper has focused on earnings forecasts, the methodology we have laid out can be applied equally well to any type of forecasts such as on macro-variables. We believe that our new methodology can be used to improve the precision of event studies of capital market efficiency which are a most basic tool for economists.

References

Bernard, V. L., and J. K. Thomas, 1990, "Evidence that stock prices do not fully reflect the implications of current earnings for future earnings," *Journal of Accounting and Economics*, 13(4), 305–340.

Daniel, K., M. Grinblatt, S. Titman, and R. Wermers, 1997, "Measuring mutual fund performance with characteristic-based benchmarks," *The Journal of finance*, 52(3), 1035–1058.

DellaVigna, S., and M. Gentzkow, 2009, "Persuasion: empirical evidence," working paper, National Bureau of Economic Research.

DellaVigna, S., and J. M. Pollet, 2009, "Investor inattention and Friday earnings announcements," *The Journal of Finance*, 64(2), 709–749.

Fama, E. F., 1998, "Market efficiency, long-term returns, and behavioral finance," *Journal of financial economics*, 49(3), 283–306.

Hirshleifer, D., S. S. Lim, and S. H. Teoh, 2009, "Driven to distraction: Extraneous events and underreaction to earnings news," *The Journal of Finance*, 64(5), 2289–2325.

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

Hong, H., J. D. Kubik, and A. Solomon, 2000, "Security Analysts' Career Concerns and Herding of Earnings Forecasts," *RAND Journal of Economics*, 31(1), 121–144.

Iskoz, S., 2003, "Bias in underwriter analyst recommendations: does it matter?," *Available at SSRN 481442*.

Keane, M. P., and D. E. Runkle, 1998, "Are Financial Analysts' Forecasts of Corporate Profits Rational?," *Journal of Political Economy*, 106(4), 768–805.

Kothari, S., and J. B. Warner, 2004, "The econometrics of event studies," *Available at SSRN 608601*.

Lim, T., 2001, "Rationality and Analysts' Forecast Bias," *The Journal of Finance*, 56(1), 369–385.

Lin, H.-w., and M. F. McNichols, 1998, "Underwriting relationships, analysts' earnings forecasts and investment recommendations," *Journal of Accounting and Economics*, 25(1), 101–127.

Lyon, J. D., B. M. Barber, and C.-L. Tsai, 1999, "Improved methods for tests of long-run abnormal stock returns," *The Journal of Finance*, 54(1), 165–201.

MacKinlay, A. C., 1997, "Event studies in economics and finance," *Journal of economic literature*, pp. 13–39.

Malmendier, U., and D. Shanthikumar, 2007, "Are small investors naive about incentives?," *Journal of Financial Economics*, 85(2), 457–489.

McNichols, M., and P. C. O'Brien, 1997, "Self-Selection and Analyst Coverage," *Journal of Accounting Research*, 35, 167–199.

Michaely, R., and K. L. Womack, 1999, "Conflict of interest and the credibility of underwriter analyst recommendations," *Review of financial studies*, 12(4), 653–686.

Mikhail, M. B., B. R. Walther, and R. H. Willis, 2007, "When security analysts talk, who listens?," *The Accounting Review*, 82(5), 1227–1253.

6 Appendix

Proof of Proposition 1:

We directly compute the correlation of CE with the market surprise S:

$$\begin{aligned}
\text{Cor}[CE, S] &= \frac{\text{Cov}[CE, S]}{\sqrt{\text{Var}[CE] \cdot \text{Var}[S]}} \\
&= \frac{\sigma_A^2}{\sqrt{(\sigma_A^2 + \frac{\sigma_F^2}{N}) \cdot \sigma_A^2}} \\
&= \frac{1}{\sqrt{1 + r_F^2/N}},
\end{aligned} \tag{30}$$

where $r_F = \sigma_F/\sigma_A$ is the ratio between the standard deviation of forecasts and the actual.

We then rewrite FOM as:

$$\begin{aligned}
FOM &= \frac{\#\{\epsilon_i < S\} - \#\{\epsilon_i > S\}}{N} \\
&= \frac{1}{N}\sum_{i=1}^{N} M_i,
\end{aligned} \tag{31}$$

where

$$M_i = \begin{cases} 1 & \text{if } \epsilon_i < S \\ -1 & \text{if } \epsilon_i > S \end{cases} \tag{32}$$

If we work out the math,

$$\begin{aligned}
\text{Cov}[FOM, S] &= \text{E}[S(\frac{1}{N}\sum_{i=1}^{N} M_i)] - \text{E}[S] \cdot \text{E}[FOM] \\
&= \frac{1}{N}\sum_{i=1}^{N} \text{E}[S \cdot (I_{\epsilon_i < S} - I_{\epsilon_i > S})] \\
&= \text{E}\left[S \cdot \left(\Phi(\frac{S}{\sigma_F}) - \left(1 - \Phi(\frac{S}{\sigma_F})\right)\right)\right] \\
&= 2\sigma_F \text{E}[X \cdot \Phi(X)],
\end{aligned} \tag{33}$$

where $\Phi(\cdot)$ is the cdf of standard normal and $X \sim \mathcal{N}(0, 1/r_F^2)$. Similarly,

$$\mathrm{Var}[FOM] \;=\; \frac{4}{N}\mathrm{E}[\Phi(X)(1-\Phi(X))] + 4\mathrm{Var}[\Phi(X)]. \tag{34}$$

Combining (33) and (34), we have

$$\mathrm{Cor}[FOM, S] \;=\; \frac{r_F \mathrm{E}[X \cdot \Phi(X)]}{\sqrt{\mathrm{E}[\Phi(X)(1-\Phi(X))]/N + \mathrm{Var}[\Phi(X)]}}, \tag{35}$$

where $X \sim \mathcal{N}(0, 1/r_F^2)$.

To show the second half, first note that (9) goes to $l(r_F) = \frac{r_F \mathrm{E}[X \cdot \Phi(X)]}{\sqrt{\mathrm{Var}[\Phi(X)]}}$ as $N \to \infty$. The limit $l(r_F)$ can be rewritten as $\frac{\mathrm{Cov}(X, \Phi(X))}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(\Phi(X))}}$, which is the correlation between a normal random variable X and its transformation $\Phi(X)$. This takes the value 1 if and only if $\Phi(X) = a + b \cdot X$ for some constants a and b, however using integration by parts

$$\Phi(X) = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} e^{-X^2/2}\left(X + \frac{X^3}{3} + \frac{X^5}{3 \cdot 5} + \cdots + \frac{X^{2n+1}}{3 \cdot 5 \cdots (2n+1)} + \cdots\right),$$

$$\Phi(X) = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} e^{-X^2/2} \sum_{k=0}^{\infty} \frac{X^{2k+1}}{(2k+1)!},$$

which is nonlinear in X. Therefore $l(r_F)$ must be strictly less than 1. In other words, for N large and when there is no bias, *CE* is a better measure of S than *FOM*. QED

Proof of Proposition 2.

Similar calculations to those leading to (8) yields a formula for (10). Similar calculations to those leading to (9) yields a formula for the correlation of *FOM* with S:

$$\mathrm{Cor}[FOM, S] \;=\; \frac{r_F \mathrm{E}[X \cdot \Phi_\omega(X, Y)]}{\sqrt{\mathrm{E}[\Phi_\omega(X, Y)(1-\Phi_\omega(X, Y))]/N + \mathrm{Var}[\Phi_\omega(X, Y)]}}, \tag{36}$$

where $\Phi_\omega(X, Y) = \omega_0 \Phi(X) + \omega_1 \Phi(\tilde{X} - Y)$, $\tilde{X} = \frac{X}{\sqrt{1 + r_b^2/r_F^2}}$, $X \sim \mathcal{N}(0, 1/r_F^2)$, and $Y \sim$

$\mathcal{N}(0, \frac{r_B^2}{r_F^2+r_b^2})$ independent of X. To derive the lower bound, first observe that

$$\text{Cov}[FOM, S] = 2\omega_0\sigma_F \text{E}[X \cdot \Phi(X)] + 2\omega_1\sigma_F \text{E}[X \cdot \Phi(\tilde{X} - Y)], \tag{37}$$

where the first term in ω_0 is positive and the second term is non-negative because X and $\Phi(\tilde{X} - Y)$ are both monotonically increasing in X when given Y and must have positive covariance. Therefore,

$$\begin{aligned}\text{Cov}[FOM, S] &\geq 2\omega_0\sigma_F \text{E}[X \cdot \Phi(X)] \\ &= \omega_0 \cdot \text{Cov}[FOM, S|\omega_1 = 0]. \end{aligned} \tag{38}$$

This means the covariance is at least a fraction of that in the ideal case. The more unbiased forecasts (the larger ω_0), the more positive relationship preserved. Consequently, the correlation between FOM and S is bounded from below

$$\begin{aligned}\text{Cor}[FOM, S] &= \frac{\text{Cov}[FOM, S]}{\sqrt{\text{Var}[FOM] \cdot \text{Var}[S]}} \\ &\geq \frac{2\omega_0\sigma_F \text{E}[X \cdot \Phi(X)]}{\sigma_A \cdot \sqrt{\text{Var}[FOM]}} \\ &\geq 2\omega_0 r_F \text{E}[X \cdot \Phi(X)], \end{aligned} \tag{39}$$

where the last inequality follows from the fact that the variance of any bounded random variable in $[a, b]$ is at most $(b-a)^2/4$ and FOM takes value between -1 and 1. QED

Proof of Proposition 3:

We first derive the following set of relationships:

$$\text{Cor}[f^*(\mathbf{U}), S] \leq \max_h \text{Cor}[S, h(\mathbf{U}, \mathcal{A})] = \sqrt{\frac{\text{E}[\text{E}[S|\mathbf{U}, \mathcal{A}]]^2}{\text{E}[S^2]}}.$$

The calculation of the first inequality follows from the discussion in the main body. The calculation for the second equality is as follows. By definition, where h^* is the optimal functional,

$$\text{Cor}(S, h^*) = \text{Cov}(S, h^*)/\sqrt{\text{Var}(S) * \text{Var}(h^*)}.$$

and

$$\text{Cov}(S, h^*) = \text{E}[S\,h^*] - \text{E}[S]\text{E}[h^*] = \text{E}[S\,\text{E}[S|\mathbf{U}, A]] = \text{E}[\text{E}[S|\mathbf{U}, A]]^2$$

because $E[S] = E[h^*] = 0$. The denominator

$$\sqrt{\text{Var}(S) * \text{Var}(h^*)} = \sqrt{\text{E}[S^2]\text{E}[h^{*2}]}.$$

Note that $\text{E}[h^{*2}] = \text{E}[\text{E}[S|\mathbf{U}, A]]^2$ by definition. Taking the ratio, we get the third term.

It can then be shown that

$$\text{E}[S|\mathbf{U}, \mathcal{A}] = \frac{\sigma_1^2}{\sigma_F^2} \sum_{i \in \mathcal{A}} U_i + \frac{\sigma_1^2}{\sigma_2^2} \sum_{i \notin \mathcal{A}} U_i,$$

where

$$\sigma_1^{-2} = \frac{1}{\sigma_A^2} + \frac{|\mathcal{A}|}{\sigma_F^2} + \frac{1}{\frac{\sigma_F^2 + \sigma_b^2}{N - |\mathcal{A}|} + \sigma_B^2}, \quad \text{and} \quad \sigma_2^{-2} = \frac{\frac{1}{N - |\mathcal{A}|}}{\frac{\sigma_F^2 + \sigma_b^2}{N - |\mathcal{A}|} + \sigma_B^2}.$$

To see this, note that the joint density of $\mathbf{U}$, S and B (where B is the aggregate bias shock) conditional on $\mathcal{A}$ is (ignoring the normalization constant):

$$\exp\{-\frac{S^2}{2\sigma_A^2}\} \cdot \exp\{-\frac{1}{\sigma_F^2} \sum_{i \in \mathcal{A}} (U_i - S)^2\} \cdot \exp\{-\frac{B^2}{2\sigma_B^2}\} \cdot \exp\{-\frac{1}{\sigma_3^2} \sum_{i \notin \mathcal{A}} (U_i - B - S)^2\},$$

where $\sigma_3^2 = \sigma_F^2 + \sigma_b^2$. The last two terms related to B can be written as

$$\exp\left\{-\frac{1}{2}\left[(\frac{1}{\sigma_B^2} + \frac{N - |\mathcal{A}|}{\sigma_3^2})B^2 - \frac{2}{\sigma_3^2} \sum_{i \notin \mathcal{A}} (U_i - S) + \frac{1}{\sigma_3^2} \sum_{i \notin \mathcal{A}} (U_i - S)^2\right]\right\}.$$

Let $\sigma_4^{-2} = \frac{1}{\sigma_B^2} + \frac{N-|\mathcal{A}|}{\sigma_3^2}$. Integrating the above expression with respect to B, we obtain

$$\exp\left\{-\frac{1}{\sigma_3^2}\sum_{i\notin\mathcal{A}}(U_i - S)^2 + \frac{\sigma_4^2}{\sigma_3^4}\left[\sum_{i\notin\mathcal{A}}(U_i - S)\right]^2\right\}.$$

Therefore, the joint density of $\mathbf{U}$ and S conditional on $\mathcal{A}$ is

$$\exp\left\{-\frac{S^2}{2\sigma_A^2} - \frac{1}{2\sigma_F^2}\sum_{i\in\mathcal{A}}(U_i - S)^2 - \frac{1}{2\sigma_3^2}\sum_{i\notin\mathcal{A}}(U_i - S)^2 + \frac{\sigma_4^2}{2\sigma_3^4}\left[\sum_{i\notin\mathcal{A}}(U_i - S)\right]^2\right\}.$$

Combining the terms related to S, we have

$$\begin{aligned}&\exp\left\{-\frac{1}{2}\left[\frac{1}{\sigma_A^2} + \frac{|\mathcal{A}|}{\sigma_F^2} + \frac{N-|\mathcal{A}|}{\sigma_3^2} - \frac{\sigma_4^2}{\sigma_3^4}(N-|\mathcal{A}|)^2\right]S^2\right\}\\ &\cdot\exp\left\{\left[\frac{1}{\sigma_F^2}\sum_{i\in\mathcal{A}}(U_i - S) + \frac{1}{\sigma_3^2}\sum_{i\notin\mathcal{A}}(U_i - S) - \frac{\sigma_4^2}{\sigma_3^4}\sum_{i\notin\mathcal{A}}U_i(N-|\mathcal{A}|)\right]S\right\}.\end{aligned}$$

Therefore, the conditional distribution of S given $\mathbf{U}$ and $\mathcal{A}$ is

$$\mathcal{N}\left(\frac{\sigma_1^2}{\sigma_F^2}\sum_{i\in\mathcal{A}}U_i + \frac{\sigma_1^2}{\sigma_2^2}\sum_{i\notin\mathcal{A}}U_i, \sigma_1^2\right),$$

where

$$\begin{aligned}&\sigma_1^{-2} = \frac{1}{\sigma_A^2} + \frac{|\mathcal{A}|}{\sigma_F^2} + \frac{N-|\mathcal{A}|}{\sigma_3^2} - \frac{\sigma_4^2}{\sigma_3^4}(N-|\mathcal{A}|)^2 = \frac{1}{\sigma_A^2} + \frac{|\mathcal{A}|}{\sigma_F^2} + \frac{1}{\frac{\sigma_F^2+\sigma_b^2}{N-|\mathcal{A}|} + \sigma_B^2},\\ \text{and}\qquad &\sigma_2^{-2} = \frac{1}{\sigma_3^2} - \frac{\sigma_4^2(N-|\mathcal{A}|)}{\sigma_3^4} = \frac{\frac{1}{N-|\mathcal{A}|}}{\frac{\sigma_F^2+\sigma_b^2}{N-|\mathcal{A}|} + \sigma_B^2}.\end{aligned}$$

Since all of the N forecasts depend on a common aggregated bias B, the variance σ_B^2 is not scaled by the number of forecasts. As σ_B^2 grows, we have

$$\lim_{\sigma_B^2\to\infty}\mathrm{E}[S|\mathbf{U},\mathcal{A}] = \frac{1}{r_F^2 + |\mathcal{A}|}\sum_{i\in\mathcal{A}}U_i.$$

Denoting the above as h^*, we can compute its second moment using the law of total variance,

$$\begin{aligned}
\mathrm{E}[h^{*2}] &= \mathrm{E}[\mathrm{Var}[h^*|S,\mathcal{A}]] + \mathrm{Var}[\mathrm{E}[h^*|S,\mathcal{A}]] \\
&= \mathrm{E}\left[\frac{1}{(r_F^2+|\mathcal{A}|)^2}|\mathcal{A}|\sigma_F^2\right] + \mathrm{Var}\left[\frac{1}{r_F^2+|\mathcal{A}|}|\mathcal{A}|S\right] \\
&= \mathrm{E}\left[\frac{1}{1+r_F^2/|\mathcal{A}|}\right]\sigma_A^2.
\end{aligned}$$

Recall that $\mathrm{Cor}[FOM, S] \geq 2\omega_0 \mathrm{E}[Z\Phi(Z/r_F)]$. Taken together we have shown Equation (29).

QED

Table 1: Summary Statistics

This table presents the summary statistics of the variables used in the regression estimations. Mean (median) consensus is the mean (median) across all qualified individual analyst forecasts in a given fiscal year. Consensus error (*CE*) is the difference between the actual annual earnings and the consensus forecast scaled by the stock price 20 days prior to the earnings announcement. We consider both mean consensus and median consensus in formulating *CE*. Dispersion (*DISP*) is the standard deviation of analysts forecasts provided by I/B/E/S scaled by *price(-20)*. *Rank(CE)* based on mean (median) consensus is the rank score of consensus errors, from 1 to 10 (1 to 6). *Rank(DISP)* is the rank score of forecast dispersion, from 1 to 10). *FOM* is defined as $\frac{K}{N} - \frac{M}{N}$, where K (M) is the number of forecasts strictly smaller (greater) than the actual earnings. N is the total number of analysts. $I_{Actual<All}$ is a dummy variable which equals 1 when all analysts' forecasts are higher than the actual earnings, and $I_{Actual>All}$ is a dummy variable which equals 1 when all analysts' forecasts are lower than the actual earnings. Panel A reports the summary statistics of the variables. Panel B reports the correlation of variables when *CE* is based on mean consensus. Panel C reports the correlation of variables when *CE* is based on median consensus.

Panel A: Summary statistics

	Mean	25th	Median	75th	Std Dev	Skewness	Kurtosis	Correlation
Mean consensus	1.7514	0.7264	1.4843	2.53	2.0925	2.3712	56.6458	
Median consensus	1.7419	0.72	1.48	2.52	2.0590	1.3806	29.4462	
CE (based on mean forecast)	-0.0031	-0.0019	0.00025	0.0019	0.0433	-30.5307	1926.6760	
CE (based on median forecast)	-0.0026	-0.0012	0.00029	0.0018	0.0376	-19.7122	801.0382	
Rank(CE) (based on mean forecast)	5.4970	3	5	8	2.8725	0.0008	1.7755	
Rank(CE) (based on median forecast)	3.4633	2	3	5	1.7064	0.0413	1.6993	
DISP	0.0070	0.0008	0.0022	0.0057	0.0592	127.5928	19814.6800	
Rank(DISP)	5.4975	3	5	8	2.8725	0.0003	1.7758	
FOM	0.1454	-0.5251	0.25	0.8333	0.7178	-0.3085	1.6528	
$I_{Actual<All}$	0.1251	0	0	0	0.3308	2.2670	6.1393	
$I_{Actual>All}$	0.2009	0	0	0	0.4007	1.4930	3.2290	
Correlation between *CE*								0.8447

Table 1 (cont'd): Summary Statistics

Panel B: Correlation matrix (*CE* based on mean consensus)

	CE	*DISP*	*Rank(CE)*	*Rank(DISP)*	*FOM*	$I_{Actual<All}$	$I_{Actual>All}$
CE	1						
DISP	-0.5184	1					
Rank(CE)	0.284	-0.0414	1				
Rank(DISP)	-0.1015	0.138	-0.1175	1			
FOM	0.2414	-0.0402	0.8103	-0.1881	1		
$I_{Actual<All}$	-0.247	0.0195	-0.4774	0.073	-0.6032	1	
$I_{Actual>All}$	0.1305	-0.0291	0.5317	-0.1722	0.597	-0.1896	1

Panel C: Correlation matrix (*CE* based on median consensus)

	CE	*DISP*	*Rank(CE)*	*Rank(DISP)*	*FOM*	$I_{Actual<All}$	$I_{Actual>All}$
CE	1						
DISP	-0.124	1					
Rank(CE)	0.2853	-0.025	1				
Rank(DISP)	-0.1002	0.138	-0.0547	1			
FOM	0.2672	-0.0402	0.8473	-0.1881	1		
$I_{Actual<All}$	-0.2746	0.0195	-0.4709	0.073	-0.6032	1	
$I_{Actual>All}$	0.1357	-0.0291	0.5231	-0.1722	0.597	-0.1896	1

Table 2: Sensitivity of earnings announcement returns to *FOM*, out-of-bound dummies, *CE*, and *Rank(CE)*

This table presents the ordinary least squares estimates of the sensitivity of earnings announcement stock returns (*CAR*) to consensus errors (*CE*) or *Rank(CE)*, *FOM*, $I_{Actual<All}$, and $I_{Actual>All}$. The dependent variable is *CAR* (cumulative abnormal return from trading day -1 to 1 around annual earnings announcement dates). The independent variables are *CE* (consensus errors in raw values), *Rank(CE)* (the rank score of consensus errors, from 1 to 10 for *Rank(CE)* based on mean consensus and 1 to 6 for *Rank(CE)* based on median consensus), *FOM* ($\frac{K}{N} - \frac{M}{N}$, where K (M) is the number of forecasts strictly smaller (greater) than the actual earnings, and N is the total number of analysts), $I_{Actual<All}$ (a dummy variable which equals 1 when all analysts' forecasts are higher than the actual earnings), and $I_{Actual>All}$ (a dummy variable which equals 1 when all analysts' forecasts are lower than the actual earnings). In Panel A, we report regression coefficients of *CE*, *FOM*, and two out-of-bound dummies. *CE* based on both mean and median consensus are considered. Panel B reports regression coefficients of *Rank(CE)*, *FOM*, and two out-of-bound dummies. *Rank(CE)* based on both mean and median consensus are considered. 34,859 observations are in each of the regression models. All standard errors are clustered by stocks. t statistics are in parentheses.

Panel A: ***CE*, *FOM*, and out-of-bound dummies as the independent variables.**

	CE is based on mean forecast					*CE* is based on median forecast				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
CE	0.0761***			-0.00278	0.00857	0.0927***			-0.00883	0.00836
	(4.09)			(-0.22)	(0.71)	(4.79)			(-0.58)	(0.55)
FOM		0.0210***		0.0210***			0.0210***		0.0211***	
		(33.86)		(33.60)			(33.86)		(33.26)	
$I_{Actual<All}$			-0.0211***		-0.0208***			-0.0211***		-0.0208***
			(-15.21)		(-15.06)			(-15.21)		(-14.83)
$I_{Actual>All}$			0.0265***		0.0264***			0.0265***		0.0264***
			(24.37)		(24.21)			(24.37)		(24.19)
Year effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.003	0.041	0.035	0.041	0.035	0.003	0.041	0.035	0.041	0.035

Panel B: ***Rank(CE)*, *FOM*, and out-of-bound dummies as the independent variables.**

	Rank(CE) is based on mean forecast					*Rank(CE)* is based on median forecast				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Rank(CE)	0.00425***			-0.0000758	0.00198***	0.00807***			0.00218***	0.00501***
	(27.15)			(-0.29)	(10.34)	(30.84)			(4.40)	(15.89)
FOM		0.0210***		0.0213***			0.0210***		0.0164***	
		(33.86)		(20.45)			(33.86)		(13.89)	
$I_{Actual<All}$			-0.0211***		-0.0142***			-0.0211***		-0.0107***
			(-15.21)		(-9.39)			(-15.21)		(-7.13)
$I_{Actual>All}$			0.0265***		0.0198***			0.0265***		0.0166***
			(24.37)		(15.98)			(24.37)		(13.58)
Year effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.028	0.041	0.035	0.041	0.038	0.035	0.041	0.035	0.041	0.042

Table 3: Sensitivity of post earnings announcement returns to *FOM*, out-of-bound dummies, *CE*, and *Rank(CE)*

This table presents the ordinary least squares estimates of the sensitivity of post earnings announcement stock returns (*POSTCAR*) to consensus errors (*CE*) or *Rank(CE)*, *FOM*, $I_{Actual<All}$, and $I_{Actual>All}$. The dependent variable is *POSTCAR* (cumulative abnormal return from trading day 2 to 126 post annual earnings announcement dates). The independent variables are *CE* (consensus errors in raw values), *Rank(CE)* (the rank score of consensus errors, from 1 to 10 for *Rank(CE)* based on mean consensus and 1 to 6 for *Rank(CE)* based on median consensus), *FOM* ($\frac{K}{N} - \frac{M}{N}$, where K (M) is the number of forecasts strictly smaller (greater) than the actual earnings, and N is the total number of analysts), $I_{Actual<All}$ (a dummy variable which equals 1 when all analysts' forecasts are higher than the actual earnings), and $I_{Actual>All}$ (a dummy variable which equals 1 when all analysts' forecasts are lower than the actual earnings). In Panel A, we report regression coefficients of *CE*, *FOM*, and two out-of-bound dummies. *CE* based on both mean and median consensus are considered. Panel B reports regression coefficients of *Rank(CE)*, *FOM*, and two out-of-bound dummies. *Rank(CE)* based on both mean and median consensus are considered. 33,644 observations are in each of the regression models. All standard errors are clustered by stocks. t statistics are in parentheses.

Panel A: *CE*, *FOM*, and out-of-bound dummies as the independent variables.

	CE is based on mean forecast					*CE* is based on median forecast				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
CE	0.0606			0.0115	0.0206	0.0613			-0.00313	0.0100
	(0.93)			(0.17)	(0.30)	(0.72)			(-0.04)	(0.11)
FOM		0.0135***		0.0133***			0.0135***		0.0135***	
		(6.27)		(5.97)			(6.27)		(5.81)	
$I_{Actual<All}$			-0.0137**		-0.0131**			-0.0137**		-0.0134**
			(-2.87)		(-2.68)			(-2.87)		(-2.64)
$I_{Actual>All}$			0.0151***		0.0149***			0.0151***		0.0150***
			(3.83)		(3.75)			(3.83)		(3.77)
Year effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2 0.002	0.003	0.003	0.003	0.003	0.002	0.003	0.003	0.003	0.003	

Panel B: *Rank(CE)*, *FOM*, and out of bound as the independent variables.

	Rank(CE) is based on mean forecast					*Rank(CE)* is based on median forecast				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Rank(CE)	0.00188***			-0.00278**	-0.0000184	0.00442***			-0.00177	0.00226
	(3.45)			(-2.87)	(-0.03)	(4.76)			(-0.96)	(1.84)
FOM		0.0135***		0.0229***			0.0135***		0.0172***	
		(6.27)		(5.98)			(6.27)		(4.02)	
$I_{Actual<All}$			-0.0137**		-0.0137*			-0.0137**		-0.00902
			(-2.87)		(-2.54)			(-2.87)		(-1.69)
$I_{Actual>All}$			0.0151***		0.0151**			0.0151***		0.0106*
			(3.83)		(3.25)			(3.83)		(2.30)
Year effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.003	0.003	0.003	0.004	0.003	0.003	0.003	0.003	0.003	0.003

Table 4: Robustness checks–Control for dispersion of analysts' forecasts (*DISP*)

This table checks the robustness of the regression results presented in Table 2 and Table 3 by further controlling for analyst forecast dispersion (*DISP*). Dependent variables are *CAR* and *POSTCAR*, respectively. Independent variables are *Rank(CE)*, *FOM*, $I_{Actual<All}$, and $I_{Actual>All}$ as in Table 3 and 4. Cases in which *Rank(CE)* is based on mean consensus or median consensus are reported. 34,859 (33,644) observations are in regressions with *CAR* (*POSTCAR*). All standard errors are clustered by stocks. t statistics are in parentheses.

	Rank(CE) **is based on mean consensus**				*Rank(CE)* **is based on median consensus**			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	CAR	*POSTCAR*	*CAR*	*POSTCAR*	*CAR*	*POSTCAR*	*CAR*	*POSTCAR*
Rank(CE)	0.00284***	0.000935	0.00558***	0.00553**	0.00703***	0.00299	0.00915***	0.00826**
	(4.52)	(0.42)	(10.62)	(2.93)	(7.47)	(0.89)	(11.68)	(2.92)
FOM	0.0187***	0.0199***			0.0142***	0.0163***		
	(16.03)	(4.54)			(11.33)	(3.56)		
$I_{Actual<All}$			-0.0118***	-0.0101			-0.00967***	-0.00774
			(-7.64)	(-1.79)			(-6.38)	(-1.42)
$I_{Actual>All}$			0.0161***	0.00962			0.0140***	0.00745
			(12.27)	(1.91)			(10.98)	(1.53)
Rank(DISP)	0.00211***	0.00302*	0.00258***	0.00418**	0.00221***	0.00298*	0.00186***	0.00315*
	(5.35)	(2.19)	(6.74)	(3.08)	(5.88)	(2.21)	(4.91)	(2.31)
*CE*DISP*	-0.000339***	-0.000436	-0.000457***	-0.000704**	-0.000619***	-0.000667	-0.000566***	-0.000833*
	(-4.95)	(-1.79)	(-6.97)	(-2.99)	(-5.98)	(-1.78)	(-5.47)	(-2.20)
Year effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.042	0.004	0.040	0.003	0.043	0.004	0.043	0.003

Table 5: Number of Analysts, *CAR*, *FOM*, out-of-bound dummies, and *Rank(CE)*

This table presents the ordinary least squares estimates of the sensitivity of earnings announcement stock returns (*CAR*) to *Rank(CE)*, *FOM*, $I_{Actual<All}$, and $I_{Actual>All}$ by further classifying stocks into 4 groups based on the number of analyst coverage. Group 1 includes stocks with 5 to 9 analysts, group 2 is 10 to 14, group 3 is 15 to 19, and group 4 is stocks with more than 20 analysts. The dependent variable is *CAR*. The independent variables are *Rank(CE)*, *FOM*, $I_{Actual<All}$, and $I_{Actual>All}$, as defined in Table 2. In Panel A, *Rank(CE)* is calculated based on mean consensus. In panel B, *Rank(CE)* is based on median consensus. All standard errors are clustered by stocks. t statistics are in parentheses.

Panel A: ***Rank(CE)*** **is based on mean consensus**

		N = 5 to 9	N = 10 to 14	N = 15 to 19	N $\geq$ 20
(1)	*Rank(CE)*	0.00465***	0.00420***	0.00356***	0.00282***
		(22.03)	(12.54)	(8.28)	(7.92)
	Year effect	Yes	Yes	Yes	Yes
	R^2	0.032	0.026	0.019	0.018
	N	18,405	7,929	4,201	4,324
(2)	*FOM*	0.0226***	0.0213***	0.0187***	0.0153***
		(26.77)	(16.04)	(11.32)	(10.72)
	Year effect	Yes	Yes	Yes	Yes
	R^2	0.044	0.040	0.034	0.032
	N	18,405	7,929	4,201	4,324
(3)	$I_{Actual<All}$	-0.0210***	-0.0218***	-0.0271***	-0.0133***
		(-11.84)	(-6.33)	(-5.74)	(-3.85)
	$I_{Actual>All}$	0.0272***	0.0287***	0.0202***	0.0222***
		(18.54)	(12.88)	(6.80)	(7.60)
	Year effect	Yes	Yes	Yes	Yes
	R^2	0.038	0.036	0.025	0.020
	N	18,405	7,929	4,201	4,324
(4)	*Rank(CE)*	0.000188	-0.000345	-0.000909	-0.000804
		(0.52)	(-0.63)	(-1.13)	(-1.43)
	FOM	0.0219***	0.0224***	0.0216***	0.0179***
		(15.01)	(10.26)	(6.94)	(7.80)
	Year effect	Yes	Yes	Yes	Yes
	R^2	0.044	0.040	0.034	0.032
	N	18,405	7,929	4,201	4,324
(5)	*Rank(CE)*	0.00212***	0.00184***	0.00179***	0.00175***
		(7.56)	(4.56)	(3.54)	(4.30)
	$I_{Actual<All}$	-0.0138***	-0.0152***	-0.0206***	-0.00697
		(-7.04)	(-4.11)	(-4.09)	(-1.91)
	$I_{Actual>All}$	0.0198***	0.0227***	0.0146***	0.0167***
		(11.34)	(8.85)	(4.40)	(5.20)
	Year effect	Yes	Yes	Yes	Yes
	R^2	0.041	0.038	0.028	0.024
	N	18,405	7,929	4,201	4,324

Table 5 (cont'd): Number of Analysts, *CAR*, *FOM*, out-of-bound dummies, and *Rank(CE)*

		N = 5 to 9	N = 10 to 14	N = 15 to 19	N ≥ 20
Panel B: *Rank(CE)* based on median consensus					
(1)	*Rank(CE)*	0.00869***	0.00808***	0.00712***	0.00562***
		(24.57)	(14.59)	(9.84)	(9.56)
	Year effect	Yes	Yes	Yes	Yes
	R^2	0.039	0.034	0.028	0.025
	N	18,405	7,929	4,201	4,324
(2)	*FOM*	0.0226***	0.0213***	0.0187***	0.0153***
		(26.77)	(16.04)	(11.32)	(10.72)
	Year effect	Yes	Yes	Yes	Yes
	R^2	0.044	0.040	0.034	0.032
	N	18,405	7,929	4,201	4,324
(3)	$I_{Actual<All}$	-0.0210***	-0.0218***	-0.0271***	-0.0133***
		(-11.84)	(-6.33)	(-5.74)	(-3.85)
	$I_{Actual>All}$	0.0272***	0.0287***	0.0202***	0.0222***
		(18.54)	(12.88)	(6.80)	(7.60)
	Year effect	Yes	Yes	Yes	Yes
	R^2	0.038	0.036	0.025	0.020
	N	18,405	7,929	4,201	4,324
(4)	*Rank(CE)*	0.00258***	0.00186	0.00147	0.000241
		(3.70)	(1.84)	(1.07)	(0.21)
	FOM	0.0171***	0.0174***	0.0156***	0.0148***
		(10.25)	(7.15)	(5.01)	(5.25)
	Year effect	Yes	Yes	Yes	Yes
	R^2	0.045	0.040	0.034	0.032
	N	18,405	7,929	4,201	4,324
(5)	*Rank(CE)*	0.00532***	0.00482***	0.00480***	0.00423***
		(11.47)	(7.35)	(5.85)	(6.28)
	$I_{Actual<All}$	-0.0104***	-0.0117**	-0.0166***	-0.00402
		(-5.30)	(-3.17)	(-3.36)	(-1.09)
	$I_{Actual>All}$	0.0164***	0.0193***	0.0114***	0.0146***
		(9.53)	(7.61)	(3.50)	(4.57)
	Year effect	Yes	Yes	Yes	Yes
	R^2	0.045	0.042	0.034	0.029
	N	18,405	7,929	4,201	4,324

Table 6: Number of Analysts, *POSTCAR*, *FOM*, out-of-bound dummies, and *Rank(CE)*

This table presents the ordinary least squares estimates of the sensitivity of post earnings announcement stock returns to *Rank(CE)*, *FOM*, $I_{Actual<All}$, and $I_{Actual>All}$ by further classifying stocks into 4 groups based on the number of analyst coverage. Group 1 includes stocks with 5 to 9 analysts, group 2 is 10 to 14, group 3 is 15 to 19, and group 4 is stocks with more than 20 analysts. The dependent variable is *POSTCAR*. The independent variables are *Rank(CE)*, *FOM*, $I_{Actual<All}$, and $I_{Actual>All}$, as defined in Table 3. In Panel A, *Rank(CE)* is calculated based on mean consensus. In panel B, *Rank(CE)* is based on median consensus. All standard errors are clustered by stocks. t statistics are in parentheses.

Panel A: *Rank(CE)* based on mean consensus

		N = 5 to 9	N = 10 to 14	N = 15 to 19	N ≥ 20
(1)	*Rank(CE)*	0.00170*	0.00223	0.00144	0.00282*
		(2.16)	(1.88)	(0.98)	(2.20)
	Year effect	Yes	Yes	Yes	Yes
	R^2	0.002	0.002	0.006	0.020
	N	17,696	7,659	4,075	4,214
(2)	*FOM*	0.0143***	0.0143**	0.0121*	0.0110*
		(4.60)	(3.05)	(2.16)	(2.25)
	Year effect	Yes	Yes	Yes	Yes
	R^2	0.003	0.003	0.007	0.020
	N	17,696	7,659	4,075	4,214
(3)	$I_{Actual<All}$	-0.0125	-0.0210*	-0.0137	-0.00796
		(-1.93)	(-1.98)	(-0.89)	(-0.67)
	$I_{Actual>All}$	0.0183***	0.0132	0.00205	0.00992
		(3.32)	(1.67)	(0.20)	(0.94)
	Year effect	Yes	Yes	Yes	Yes
	R^2	0.003	0.003	0.006	0.019
	N	17,696	7,659	4,075	4,214
(4)	*Rank(CE)*	-0.00401**	-0.00203	-0.00332	0.00173
		(-2.84)	(-1.03)	(-1.29)	(0.79)
	FOM	0.0281***	0.0210**	0.0230*	0.00540
		(5.06)	(2.73)	(2.33)	(0.65)
	Year effect	Yes	Yes	Yes	Yes
	R^2	0.003	0.003	0.007	0.020
	N	17,696	7,659	4,075	4,214
(5)	*Rank(CE)*	-0.00124	0.000426	0.00110	0.00290
		(-1.11)	(0.28)	(0.60)	(1.86)
	$I_{Actual<All}$	-0.0167*	-0.0195	-0.00968	0.00258
		(-2.25)	(-1.63)	(-0.58)	(0.20)
	$I_{Actual>All}$	0.0227***	0.0118	-0.00140	0.000805
		(3.38)	(1.28)	(-0.12)	(0.07)
	Year effect	Yes	Yes	Yes	Yes
	R^2	0.003	0.003	0.006	0.020
	N	17,696	7,659	4,075	4,214

Table 6 (cont'd): Number of Analysts, *POSTCAR*, *FOM*, out-of-bound dummies, and *Rank(CE)*

Panel B: *Rank(CE)* based on median consensus					
		N = 5 to 9	N = 10 to 14	N = 15 to 19	N $\geq$ 20
(1)	*Rank(CE)*	0.00419**	0.00477*	0.00394	0.00550*
		(3.17)	(2.36)	(1.58)	(2.55)
	Year effect	Yes	Yes	Yes	Yes
	R^2	0.002	0.003	0.006	0.021
	N	17,696	7,659	4,075	4,214
(2)	*FOM*	0.0143***	0.0143**	0.0121*	0.0110*
		(4.60)	(3.05)	(2.16)	(2.25)
	Year effect	Yes	Yes	Yes	Yes
	R^2	0.003	0.003	0.007	0.020
	N	17,696	7,659	4,075	4,214
(3)	$I_{Actual<All}$	-0.0125	-0.0210*	-0.0137	-0.00796
		(-1.93)	(-1.98)	(-0.89)	(-0.67)
	$I_{Actual>All}$	0.0183***	0.0132	0.00205	0.00992
		(3.32)	(1.67)	(0.20)	(0.94)
	Year effect	Yes	Yes	Yes	Yes
	R^2	0.003	0.003	0.006	0.019
	N	17,696	7,659	4,075	4,214
(4)	*Rank(CE)*	-0.00392	-0.00138	-0.00178	0.00580
		(-1.49)	(-0.35)	(-0.37)	(1.34)
	FOM	0.0226***	0.0171	0.0158	-0.000834
		(3.69)	(1.90)	(1.44)	(-0.08)
	Year effect	Yes	Yes	Yes	Yes
	R^2	0.003	0.003	0.007	0.020
	N	17,696	7,659	4,075	4,214
(5)	*Rank(CE)*	0.000666	0.00245	0.00415	0.00597*
		(0.37)	(0.93)	(1.34)	(2.28)
	$I_{Actual<All}$	-0.0111	-0.0158	-0.00463	0.00519
		(-1.52)	(-1.34)	(-0.28)	(0.40)
	$I_{Actual>All}$	0.0170**	0.00848	-0.00553	-0.000914
		(2.58)	(0.91)	(-0.48)	(-0.08)
	Year effect	Yes	Yes	Yes	Yes
	R^2	0.003	0.003	0.006	0.020
	N	17,696	7,659	4,075	4,214

Table 7: Sensitivity of forecast revision to *Rank(CE)*, *FOM*, $I_{Actual<All}$ and $I_{Actual>All}$

This table presents the ordinary least squares estimates of the sensitivity of analysts' forecast revision to *Rank(CE)*, *FOM*, $I_{Actual<All}$, and $I_{Actual>All}$. The dependent variable is the forecast revision (the difference in mean (median) consensus between two adjacent fiscal years). The independent variables are *Rank(CE)* (the rank score of consensus errors, from 1 to 10 for *Rank(CE)* based on mean consensus and 1 to 6 for *Rank(CE)* based on median consensus), *FOM* ($\frac{K}{N} - \frac{M}{N}$, where K (M) is the number of forecasts strictly smaller (greater) than the actual earnings, and N is the total number of analysts), $I_{Actual<All}$ (a dummy variable which equals 1 when all analysts' forecasts are higher than the actual earnings), and $I_{Actual>All}$ (a dummy variable which equals 1 when all analysts' forecasts are lower than the actual earnings). 27,701 observations are in each of the regression models.

Panel A: *Rank(CE)* is based on mean consensus

	(1)	(3)	(2)	(5)	(4)
Rank(CE)	0.189***			0.121***	0.165***
	(27.27)			(10.28)	(18.22)
FOM		0.736***		0.334***	
		(29.00)		(7.78)	
$I_{Actual<All}$			-0.676***		-0.0992
			(-11.60)		(-1.53)
$I_{Actual>All}$			0.792***		0.251***
			(17.97)		(4.89)
Year effects	Yes	Yes	Yes	Yes	Yes
R^2	0.035	0.033	0.021	0.037	0.036

Panel B: *Rank(CE)* is based on median consensus

	(1)	(3)	(2)	(5)	(4)
Rank(CE)	0.182***			0.103***	0.155***
	(26.63)			(7.75)	(17.47)
FOM		0.736***		0.369***	
		(29.00)		(7.57)	
$I_{Actual<All}$			-0.676***		-0.125
			(-11.60)		(-1.92)
$I_{Actual>All}$			0.792***		0.283***
			(17.97)		(5.64)
Year effects	Yes	Yes	Yes	Yes	Yes
R^2	0.033	0.033	0.021	0.035	0.034

Table 8: Parametric-dependent (bias and precision adjusted) *CE* and *Rank(CE)* compared to parametric-free *FOM* for explaining earnings announcement returns

This table presents the ordinary least squares estimates of the sensitivity of earnings announcement stock returns (*CAR*) to consensus errors (*CE*) or *Rank(CE)*, *FOM*, $I_{Actual<All}$, and $I_{Actual>All}$, where *CE* and *Rank(CE)* are calculated based on parametric dependent method using a bias and precision adjusted consensus. The dependent variable is *CAR* (cumulative abnormal return from trading day -1 to 1 around annual earnings announcement dates). All other independent variables are as described in Table 2. In Panel A, we report regression coefficients of parametric-dependent-*CE*, *FOM*, and two out-of-bound dummies. Panel B reports regression coefficients of parametric-dependent-*Rank(CE)*, *FOM*, and two out-of-bound dummies. 21,370 observations are in each of the regression models. All standard errors are clustered by stocks. *t* statistics are in parentheses.

Panel A: Parametric-dependent-*CE* and *FOM*

	(1)	(2)	(3)
CE	0.00000344	0.00000563	0.00000354
	(0.41)	(0.65)	(0.45)
FOM		0.0210***	
		(28.67)	
$I_{Actual<All}$			-0.0208***
			(-13.26)
$I_{Actual>All}$			0.0248***
			(20.61)
Year effect	Yes	Yes	Yes
R^2	0.001	0.048	0.039

Panel B: Parametric-dependent-*Rank(CE)* and *FOM*

	(1)	(2)	(3)
Rank(CE)	0.000557**	-0.000324	-0.000139
	(3.20)	(-1.88)	(-0.81)
FOM		0.0212***	
		(28.73)	
$I_{Actual<All}$			-0.0210***
			(-13.28)
$I_{Actual>All}$			0.0249***
			(20.64)
Year effect	Yes	Yes	Yes
R^2	0.001	0.048	0.039

Table 9. Lower bounds of relative efficiency of FOM in comparison with the ideal but infeasible measure (N = 10).

ω_0	**0.1**	**0.3**	**0.5**	**0.7**	**0.9**	**1.0**
$r_F = 0$	**0.099**	**0.243**	**0.399**	**0.559**	**0.718**	**0.798**
$r_F = 0.5$	**0.096**	**0.228**	**0.367**	**0.509**	**0.651**	**0.723**
$r_F = 1$	**0.092**	**0.202**	**0.312**	**0.423**	**0.536**	**0.592**
$r_F = 2$	**0.086**	**0.169**	**0.243**	**0.315**	**0.386**	**0.422**
$r_F = \infty$	**0.080**	**0.138**	**0.178**	**0.211**	**0.239**	**0.252**

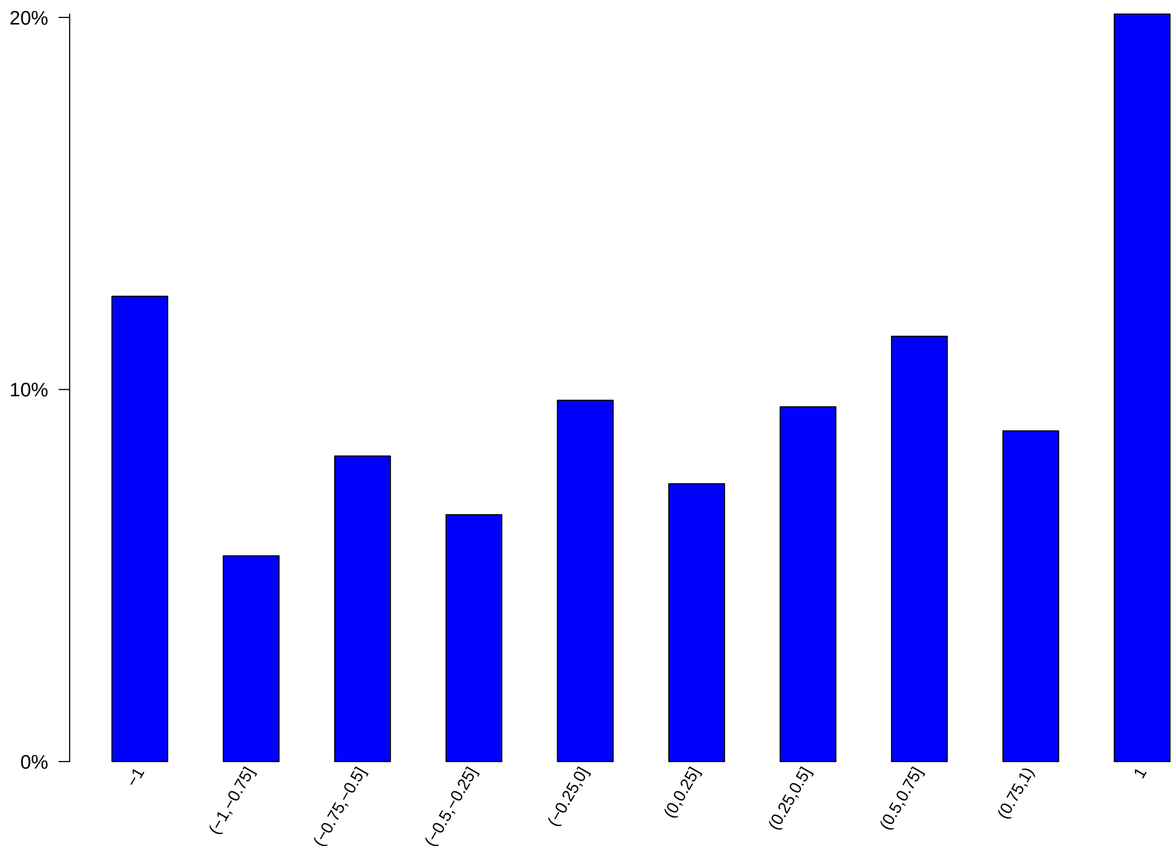


Figure 1: The distribution of *FOM* over the whole sample. *FOM* is the fraction of misses defined as $\frac{K}{N} - \frac{M}{N}$, where $K(M)$ is the number of forecasts strictly smaller (greater) than the actual earnings, and N is the total number of analysts.

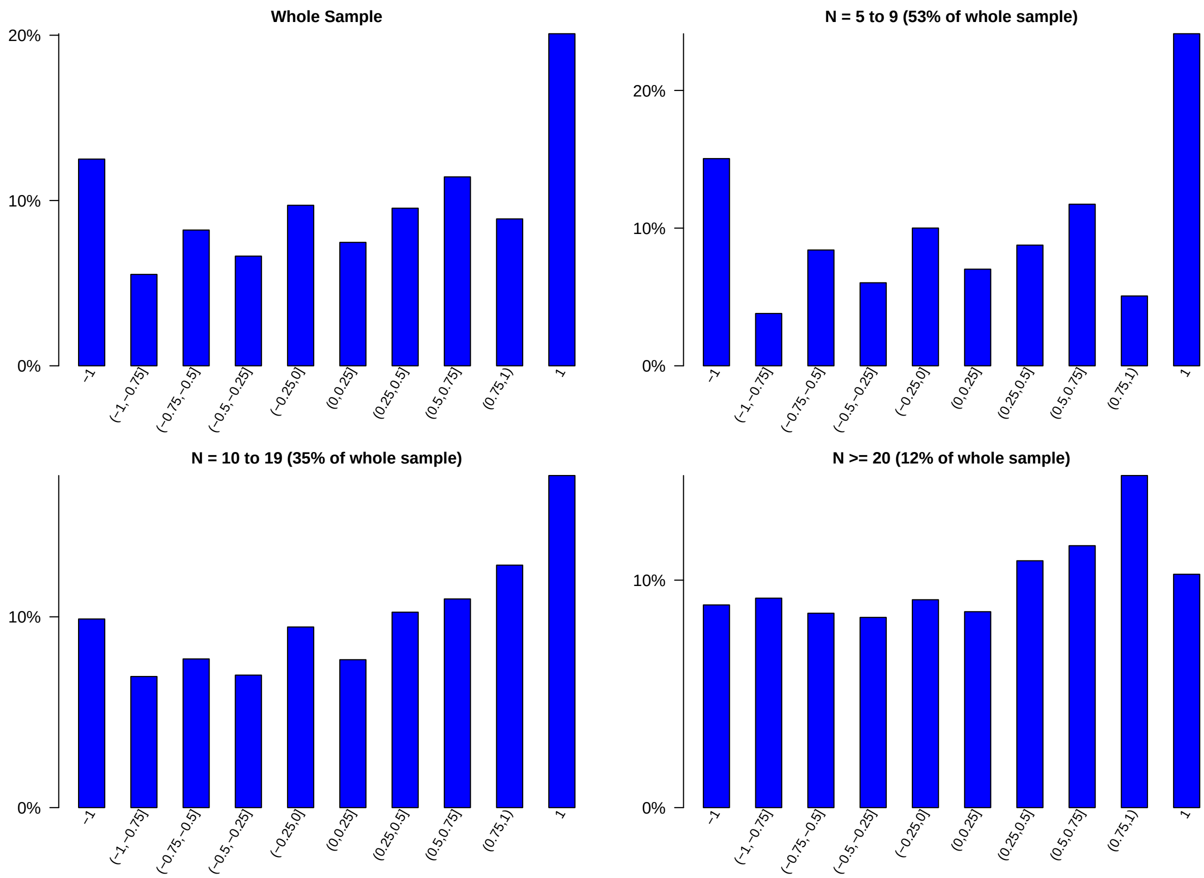


Figure 2: The distribution of *FOM* over the whole sample and conditional on different number of analysts N. *FOM* is the fraction of misses defined as $\frac{K}{N} - \frac{M}{N}$, where $K(M)$ is the number of forecasts strictly smaller (greater) than the actual earnings, and N is the total number of analysts.

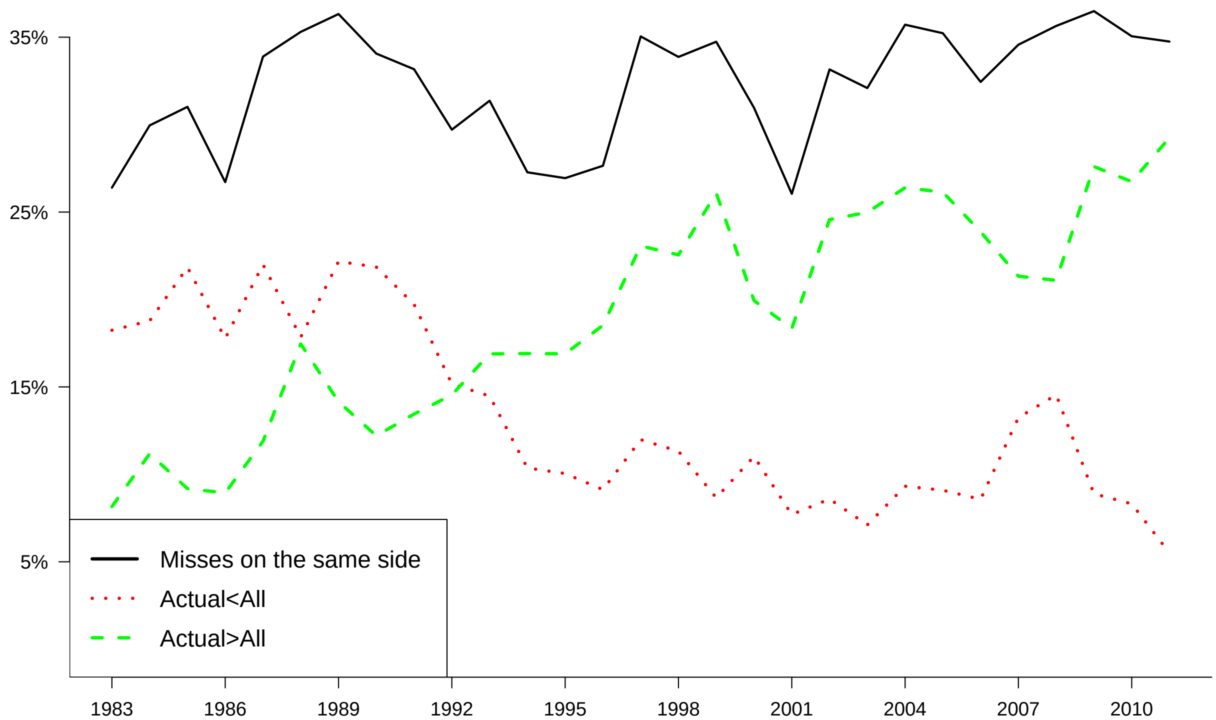


Figure 3: The time series of the percentage of misses on the same side.

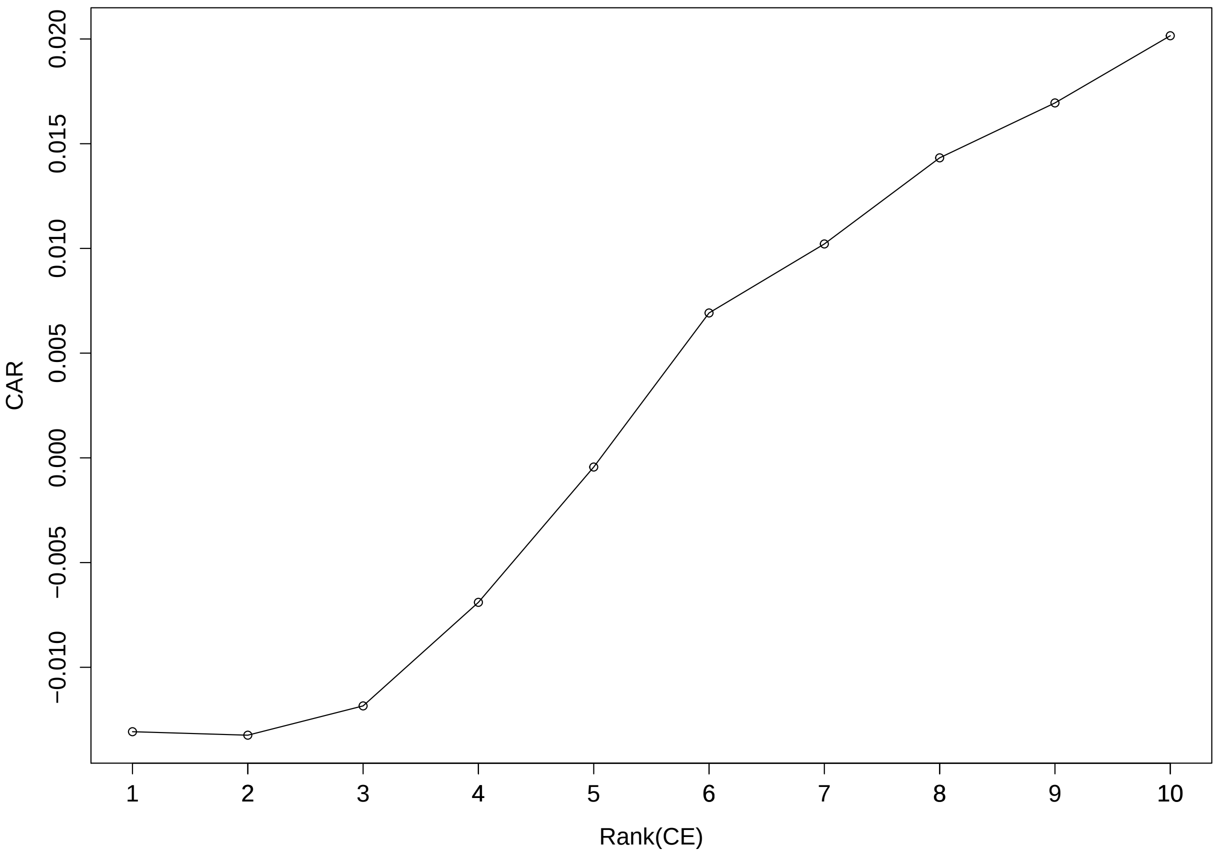


Figure 4: Average *CAR* against *Rank(CE)*. *CAR* is the cumulative abnormal return from trading day −1 to 1 around annual earnings announcement dates, and *Rank(CE)* is the rank score 1 to 10 of consensus errors CE based on mean consensus.

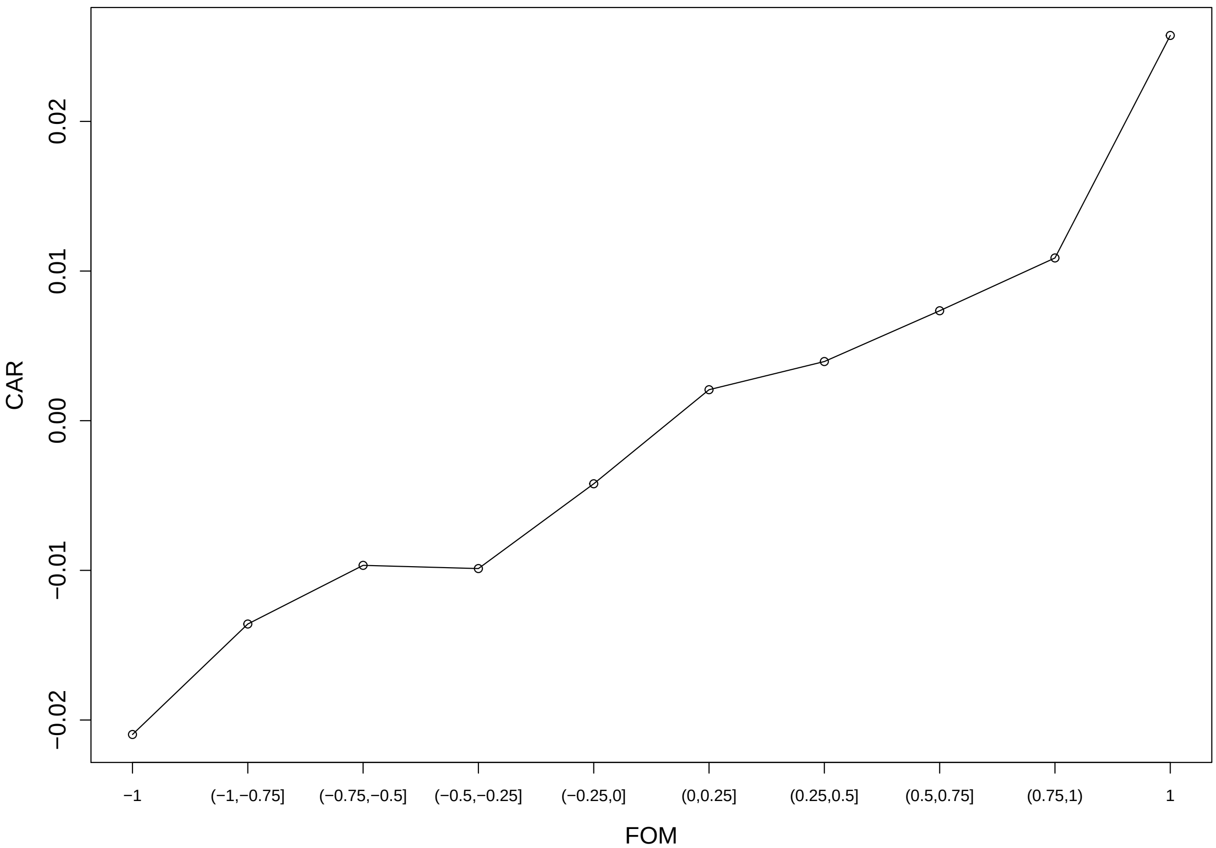


Figure 5: Average *CAR* against *FOM*. *CAR* is the cumulative abnormal return from trading day -1 to 1 around annual earnings announcement dates, and *FOM* is the fraction of misses defined as $\frac{K}{N} - \frac{M}{N}$, where $K(M)$ is the number of forecasts strictly smaller (greater) than the actual earnings, and N is the total number of analysts.

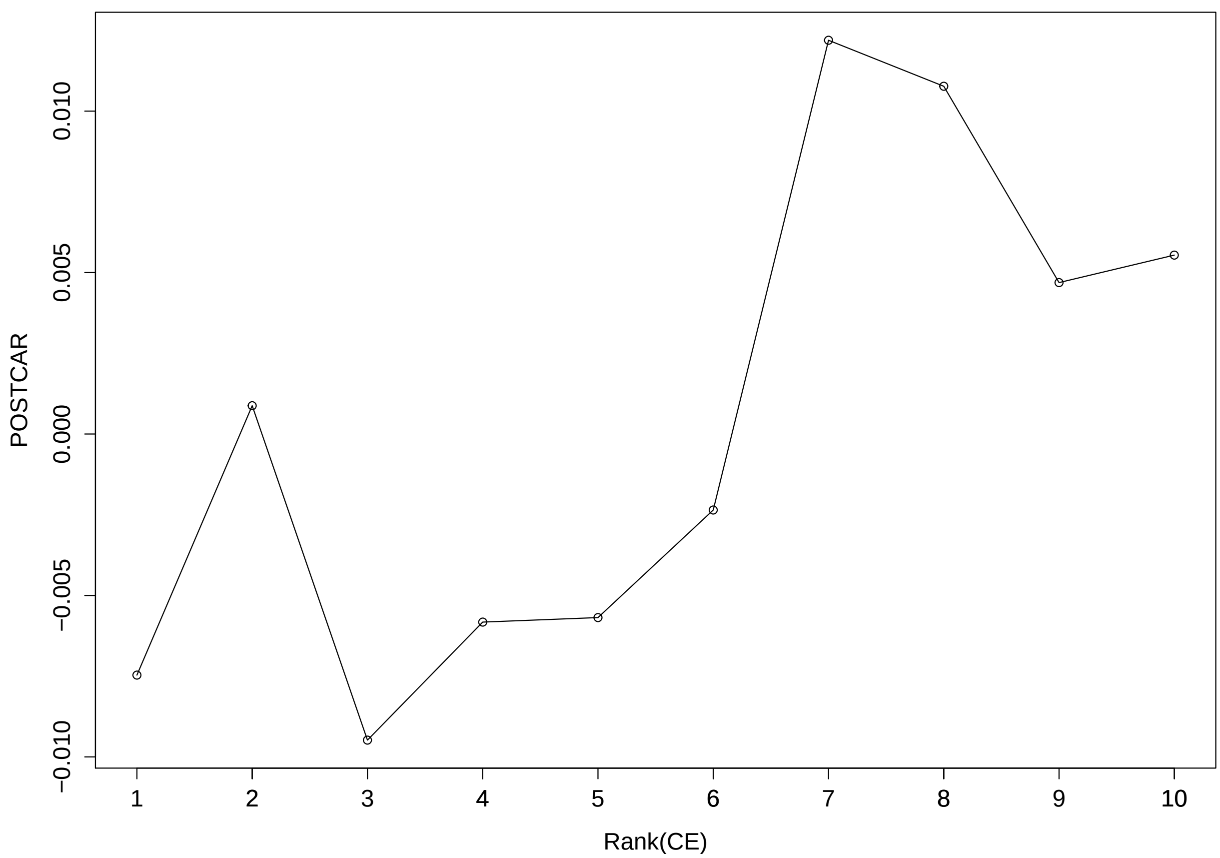


Figure 6: Average *POSTCAR* against *Rank(CE)*. *POSTCAR* is the cumulative abnormal return from trading day 2 to 126 post annual earnings announcement dates, and *Rank(CE)* is the rank score 1 to 10 of consensus errors CE based on mean consensus.

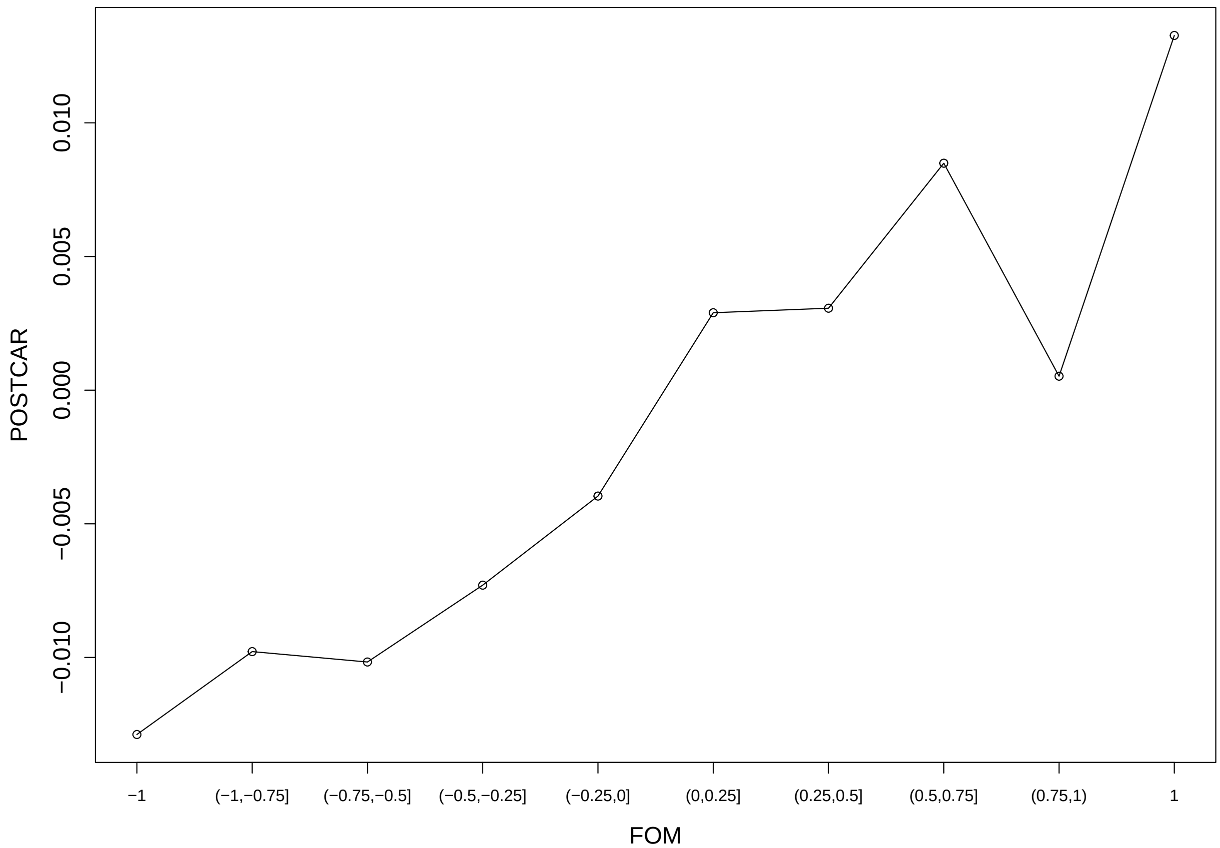


Figure 7: Average *POSTCAR* against *FOM*. *POSTCAR* is the cumulative abnormal return from trading day 2 to 126 post annual earnings announcement dates, and *FOM* is the fraction of misses defined as $\frac{K}{N} - \frac{M}{N}$, where $K(M)$ is the number of forecasts strictly smaller (greater) than the actual earnings, and N is the total number of analysts.

Supplementary Internet Appendix

A Numerical Calculations and Extensions

In this section, we provide more color on how bias affects the relative performance of *CE*, *Rank(CE)* and *FOM* and why *FOM* is a robust measure of surprises S.

A.1 Unbiased Forecasts Benchmark: $\omega_1 = 0$

We start with the unbiased benchmark. In the earlier model section, we had established some results for N large but have also done extensive calculations over wide parameter ranges. To evaluate their relative performance (*CE* compared to *FOM*), we can compute the exact value of (8) and (9) for any given pair of parameters. Appendix Figure 1 shows the contour plot of the correlation between *CE* and S minus the correlation between *FOM* and S (i.e., $\text{Cor}[CE, S] - \text{Cor}[FOM, S]$) as a function of r_F and N. Although we cannot prove it in full generality, we searched over a sufficiently large space with realistic parameter values and the difference stays positive, so we conclude that *CE* is superior than *FOM* for practical use in this ideal case.

The relative performance of *CE* and *FOM* changes with r_F and N in a nonlinear manner. But we can try to get some intuition and a flavor of what drives this difference in performance. If we take a horizontal slice of this contour by fixing N, the difference is the smallest at around $r_F = 1$ and when r_F is large (see the bottom right corner). The intuition behind the first observation is that *FOM* tries to gauge one realization of S by using N realized noise as a benchmark, i.e., counting how many ϵ_i's are above or below it. If S and ϵ_i's have roughly the same distribution, it gives the most accurate account for the location of S in its own unobserved distribution. This in our case leads to $r_F = \sigma_F/\sigma_A \sim 1$ (the exact maximal point depends on N). On the other hand, as r_F increases, the correlation of both measures drop and they become equally bad. Appendix Figure 2 shows the pattern when $N = 10$, which is rather representative of different N's.

A.2 Biased Forecasts: $\omega_1 > 0$

While *CE* can be large simply due to the existence of one very negative F_i, *FOM* is much less affected because each observation only contributes as 1 or -1 in the sum (31) regardless of its magnitude. One consequence is that *CE* and *FOM* are no longer highly correlated. While we observe a rather low correlation (around 0.28) in earnings data, which is also due to other reasons as we argue in Section A.5, here we use simulations to reveal part of the dynamic caused by biased forecasts. We simulate data according to the model and calculate the correlations using $50,000$ samples, where the key parameters ω_1 and r_B vary over their range, and the others fixed at $N = 20$, $r_F = 1/2$ and $r_b = r_B/5$. Appendix Figure 3 shows how the correlation decreases with r_B, the relative uncertainty level of the bias component B. In terms of ω_1, recall it is the proportion of biased forecasts, so the correlation first decreases with the introduction of biased forecasts as soon as ω_1 becomes nonzero, and then picks up when both measures get equally bad.

Along with the lower correlation between these two measures, the discrepancy between their performance measuring market surprise also widens, mainly due to their different resistance to bias. We have shown earlier (Proposition 2) that *FOM* will eventually outperform *CE* as bias becomes more significant, because *FOM*'s correlation with S has a positive lower bound whereas $\text{Cor}(CE, S)$ can be reduced to zero quickly. Indeed this is what we observe in simulation studies. As an illustration, again let the key parameters ω_1 and r_B vary over their range, with the others fixed at $r_F = 1/2$, $r_b = r_B/5$ and $N = 20$. We directly compute the correlation between *CE* and S in (10) and simulate $100,000$ samples of X and Y to compute the correlation between *FOM* and S in (36). Appendix Figure 4 shows a representative pattern of their relative performance as a function of ω_1 and r_B, where the difference between $\text{Cor}[CE, S]$ and $\text{Cor}[FOM, S]$ becomes negative (i.e., *FOM* outperforms) as the relative dispersion of bias $r_B = \sigma_B/\sigma_A$ increases.

A.2.1 *CE* and *Rank(CE)*

In practice, people use *Rank(CE)*, i.e., sort *CE* into 10 deciles in order to be robust to outliers. However, this global adjustment may not work in the presence of bias. For example, one single large biased forecast can still move *CE* from decile 10 down to decile 1 and distort the ordering. Appendix Figure 5 shows a representative pattern of the difference in the performance of *Rank(CE)* and *FOM* (i.e., $\text{Cor}[\textit{Rank(CE)}, S] - \text{Cor}[\textit{FOM}, S]$) as a function of ω_1 and r_B with the same set of parameters as in Section A.2, where each $\text{Cor}[\textit{Rank(CE)}, S]$ is computed using 50,000 simulated samples. Comparing with Appendix Figure 4, there is some improvement when r_B is not too large. However, the essence of the analysis on *CE* carries over to *Rank(CE)* because when *CE* is greatly contaminated, the coding of *Rank(CE)* does not help much: the damage is already done. In this sense, *FOM* measure does the robustness adjustment on a local level, so the impact from bias is alleviated when aggregating N forecasts, instead of afterwards. Therefore, *FOM* improves over *Rank(CE)* for the same reason as it does over *CE*, the reason being their sensitivity to large bias. That being said, *Rank(CE)* does have better property when treating the few outliers that overthrow *CE*, and Section A.5 develops this aspect of the relationship between *CE* and *Rank(CE)* in an extended model.

A.3 Winsorized Mean and Median

In order to be robust to the noisy forecasts, one may also Winsorize the forecasts. For example, a 5% Winsorization would set all forecasts below the 5th percentile set to the 5th percentile, and data above the 95th percentile set to the 95th percentile. The average of the resulting data is the Winsorized mean of forecasts. Similarly, we can define the Winsorized consensus error as

$$CE_{\lambda}^{win} = A - \bar{F}_{\lambda}^{win},$$

where λ is the percentage of data on each tail being replaced. Note that when $\lambda = 50\%$, the Winsorized mean becomes median:

$$CE^{win}_{50\%} = CE_{med} = A - \text{median}(F_i).$$

However, such measures do not show much, if any, improvement in our regression results of earnings announcement event study. This is not surprising because although Winsorization is designed to remove the two tails in a set of forecasts, it is by no means equivalent to removing the biased ones. Since the realization of bias is unknown in each draw, it is impossible for Winsorization to correctly pick up all the bad forecasts without sacrificing the good ones. In the same spirit as the analysis of consensus errors, the Winsorized measures by definition still strongly depend on the magnitude of forecasts, which inevitably leads to their vulnerability to bias. The more volatile B is, the harder it is for Winsorization to achieve consistent performance. Appendix Figure 6 illustrates how the performance drops with increasing r_B through 5000 simulations, where the other parameters in the model are set as $\omega_1 = 0.3$, $r_F = 1/2$, $r_b = r_B/5$ and $N = 20$.

Furthermore, the performance also depends on the fraction of biased forecasts and the choice of λ for Winsorization. Unfortunately, the fraction of biased forecasts ω_1 is usually unknown in practice and may even be varying, so it is hard if not impossible to set λ, the single important parameter for Winsorization, and an inappproriate choice might result in undesirable performance. This is illustrated in Appendix Figure 7, where the relative performance of different Winsorized measures changes with the fraction of biased forecasts ω_1, and the other parameters in the model are set as $r_B = 10$, $r_F = 1/2$, $r_b = r_B/5$ and $N = 20$.

A.4 Remark on the Model

A key assumption in our model is that for each stock a fraction of analysts are biased. Recall that under our modelling, the forecasts come from a mixture composed of two normal distributions, one centered around the unknown market expectation e and the other biased by a magnitude of the realized B. While the aggregated bias magnitude B can be huge or moderate, ω_1 the weight of the biased distribution in the mixture is with respect to N so the number of biased analysts scales with the total number and makes the law of large numbers fail. In this normal mixture framework, the bias component is essential and we have shown how it drives the behaviour of different measures that is consistent with our observations. If we remove the bias part of the modelling and instead introduce bad forecasts by having large variance in one of the distributions, it will fail to represent some important features in the real data. More specifically, suppose the forecasts are given by

$$F_i = e + \epsilon_i, \tag{40}$$

where ϵ_i's follow a mixture of two normal distributions: $\mathcal{N}(0, \sigma_0^2)$ with probability ω_0 and $\mathcal{N}(0, \sigma_1^2)$ with probability $\omega_1 = 1 - \omega_0$, and $\sigma_1^2 > \sigma_0^2$. Notice that this is actually a limiting case of our specification (3) by setting $\sigma_B = 0$, which means B is always 0 so that its impact disappears. Under this alternative modelling, even though individual forecasts can be very volatile, the variance of the average forecast error is given by:

$$\mathrm{Var}[\frac{1}{N}\sum_{i=1}^{N}\epsilon_i] = \frac{1}{N}(\omega_0\sigma_0^2 + \omega_1\sigma_1^2), \tag{41}$$

so *CE* still converges to S by the law of large numbers. That is, although σ_1^2 can be large, the distortion from fat-tails is greatly discounted and the variance decreases linearly in N, unlike in the original model the variance of the average noise never vanishes no matter how big N is. This implies that *CE* or *Rank(CE)* should be better for larger N under the alternative model, which does not quite match what we see in the real data (recall Table 5).

Furthermore, in the absence of random bias all the forecasts are centered around the real market expectation e, so it is much easier for Winsorisation to filter the bad forecasts. As a comparative example to Appendix Figure 6, Appendix Figure 8 illustrates the much stronger performance of Winsorized mean and median through 5000 simulations, which is again different from what we see in the empirical study and undermines the validity of this alternative modelling.

A.5 Extension Allowing for Outliers in *CE*

Although by comparing Appendix Figure 4 and Appendix 5 we show that *Rank(CE)* has slight improvement over *CE*, so far in our analysis they play a very similar role. Consequently, when our model generates a low correlation between *CE* and *FOM*, *Rank(CE)* and *FOM* are also much less correlated. However in real data we find the correlation between *CE* and *Rank(CE)* is merely over 0.28, which leads to a low Cor(*CE*, *FOM*) around 0.24 and a rather high Cor(*Rank(CE)*, *FOM*) over 0.81. As we further delve into data, we find rare events when most analysts or even everyone miss by quite a margin, which produces huge *CE* that has a magnitude multiple times more than the regular majority (e.g., the 3% on two tails is 30 times of the central 97% in average absolute value). Note that our *CE* is scaled by stock price and controlled for split, so this is not an issue about firm heterogeneity. These large values are able to drive the correlation between *CE* and other measures down. For example, Appendix Figure 9 shows how the extreme tails of *CE* diminish its covariation with *Rank(CE)* in the regular region, that is, when we zoom in and conditional on *Rank(CE)* being 2 to 9 only, the correlation bounces back to 0.72.

Recall that in our model ω_1 is a constant and B follows a normal distribution, we clearly are not able to generate the tail events of huge *CE* within a reasonable range of parameter values. In order to close this gap with the real data, we introduce a tail event scenario with a small probability. That is, with probability $1-\theta$ the forecasts follow the original model as in (3); with probability θ which is supposed to be very small, all the forecasts are off by a magnitude possibly huge:

$$F_i = e + \tilde{b}_i + \epsilon_i, \quad i = 1, \cdots, N \tag{42}$$

where $\epsilon_i \sim \mathcal{N}(0, \sigma_F^2)$, and conditional on $\tilde{B} \sim \mathcal{N}(0, \sigma_{\tilde{B}}^2)$, we have $\tilde{b}_i \sim \mathcal{N}(\tilde{B}, \sigma_b^2)$. $\sigma_{\tilde{B}}$ should be large relatively to σ_B, and we use $\sigma_{\tilde{B}} = 30 \cdot \sigma_B$ which seems a reasonable scale to represent the real data. As argued above, this formulation helps to explain the low correlation between *CE* and *Rank(CE)* as well as *FOM*, and its poor performance as a proxy of the market surprise S. On the other hand, the impact on *Rank(CE)* and *FOM* is very limited as long as θ is small. Since huge values of *CE* only translate to the boundary points in *Rank(CE)* and *FOM*, their distortion is not magnified by the magnitude. By a similar argument as in the case of *FOM* with respect to biased forecasts, the behaviour of *Rank(CE)* and *FOM* should not deviate too much from their respective $\theta = 0$ case.

We now confirm our hypothesis through simulation studies. Throughout this section, the correlations are computed using $100,000$ simulated samples for each pair of parameters θ and r_B, with the others fixed at $\omega_1 = 0.3$, $r_F = 1/2$, $r_b = r_B/5$ and $N = 20$. Appendix Figure 10 and Appendix 11 show how the correlation between *CE* and other measures decreases dramatically with the introduction of θ. On the other hand, the relationship between *Rank(CE)* and *FOM* are rather stable, indicated by the horizontal stripes in Appendix Figure 12. In terms of the performance as a proxy of market surprise, the gap between *FOM* and *CE* widens because of the tail scenario that undermines *CE* (Appendix Figure 13), while the improvement of *FOM* over *Rank(CE)* remains as in $\theta = 0$ case (Appendix Figure 14).

A.6 Extended Model

Our model above assumes that the market's expectation conditions on information outside the set of analyst forecasts. But we can model the market's expectation as dependent just on the set of analysts' forecasts and obtain the same results.

Suppose now that $A \sim \mathcal{N}(0, \sigma_A^2)$ for simplicity. There are $i = 1, ..., N$ forecasts. We then assume that individual forecasts i is given by

$$F_i = \begin{cases} A + \epsilon_i & \text{with prob. } \omega_0 \\ A + b_i + \epsilon_i & \text{with prob. } \omega_1 = 1 - \omega_0 \end{cases} \tag{43}$$

where $\epsilon_i \sim \mathcal{N}(0, \sigma_F^2)$ and is uncorrelated with the randomness in A. Each forecast is unbiased with probability ω_0, and is contaminated by an individual bias term b_i with probability $\omega_1 = 1 - \omega_0$. We model the bias in the same manner as before. For each set of N forecasts an aggregated bias level $B \sim \mathcal{N}(0, \sigma_B^2)$ is drawn first, and conditional on this realized B individual bias b_i follows $\mathcal{N}(B, \sigma_b^2)$.

We assume that investors are able to de-bias whereas the econometrician cannot. Hence, the market's posterior of A is given by

$$\hat{A} = \frac{1}{N} \sum_{i=1}^{N} F_i^*, \tag{44}$$

where $F_i^* = A + \epsilon_i$ is the debiased forecasts. This follows from the usual Kalman Filtering results in linear-normal models where each forecast can be interpreted as a linear signal of the actual A. Since each signal has equal precision, there is then equal weighting of the signals in forming the posterior $\hat{A}$. The market surprise then is given by

$$S = A - \hat{A} \tag{45}$$

Notice that CE is now given by

$$CE = A - \frac{1}{N} \sum_{i=1}^{N} F_i \tag{46}$$

and FOM is now given by

$$FOM = \frac{1}{N}\sum_{i=1}^{N}(I_{F_i<A} - I_{F_i>A}) \tag{47}$$

We want to compare again the correlation of CE and FOM with the market surprise S, respectively,

We can calculate that

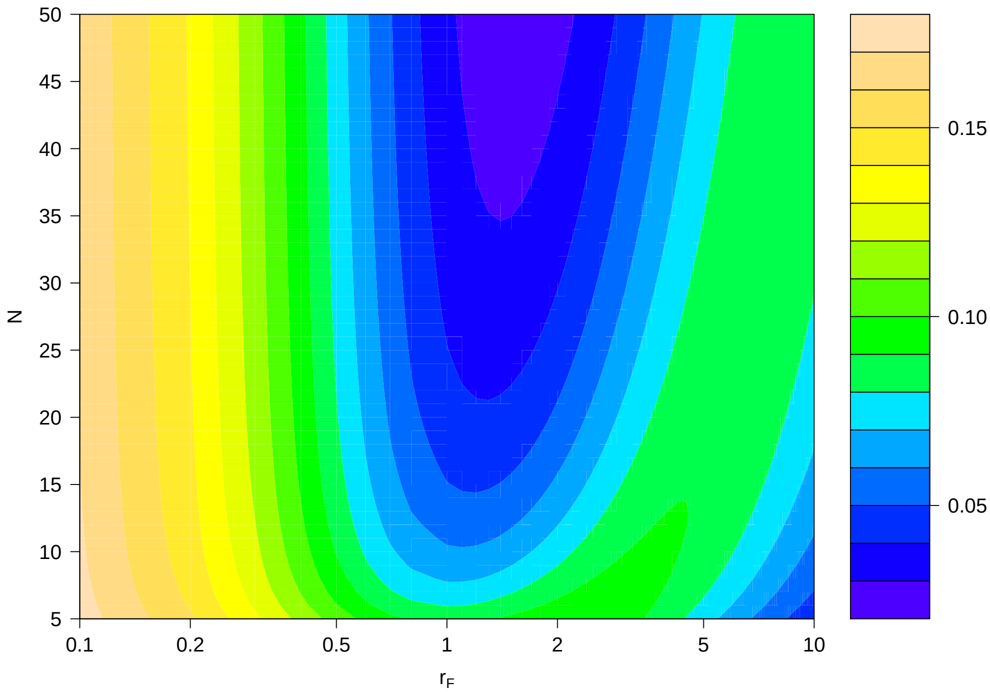
$$\mathrm{Cor}(CE, S) = \frac{1}{\sqrt{1 + \omega_0\omega_1 r_B^2 + \omega_1 r_b^2 + \omega_1^2 r_B^2 N}} \tag{48}$$

where $r_B = \sigma_B/\sigma_F$ and $r_b = \sigma_b/\sigma_F$. We can also show that

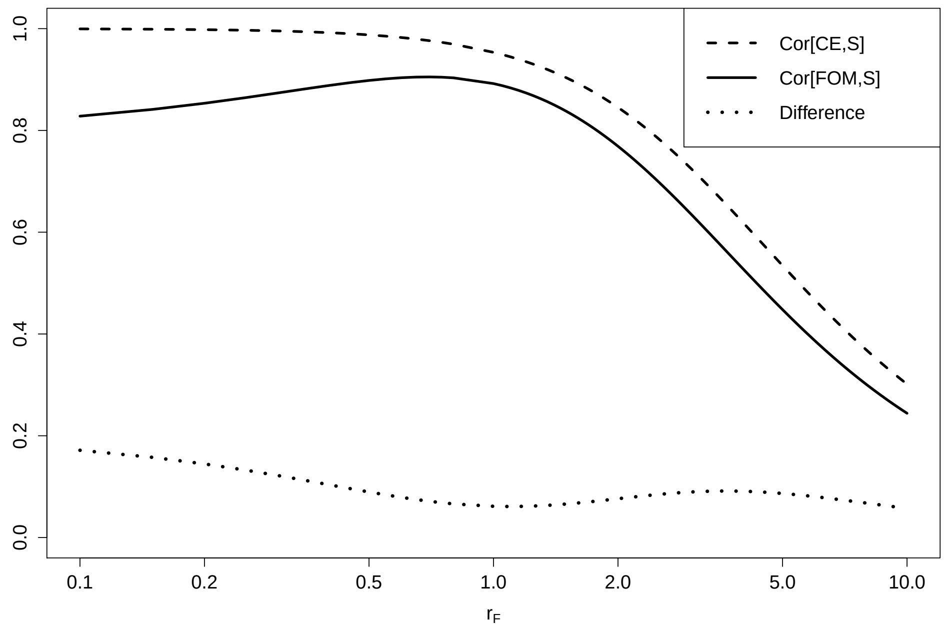
$$\mathrm{Cor}(FOM, S) = \frac{\omega_0\frac{1}{\sqrt{2\pi}} + \omega_1 \mathrm{E}[X\Phi(\tilde{X} - Y)]}{\sqrt{\frac{\omega_0}{2}(1 - \frac{\omega_0}{2}) + \omega_1^2\mathrm{E}[\Phi(\tilde{X} - Y)(1 - \Phi(\tilde{X} - Y))] + N\omega_1^2\mathrm{Var}[\Phi(\tilde{X} - Y)]}} \tag{49}$$

where $X \sim \mathcal{N}(0, 1)$ and $\tilde{X} = X/r_b$ which is orthogonal to $Y \sim \mathcal{N}(0, \frac{r_B^2}{r_b^2})$.

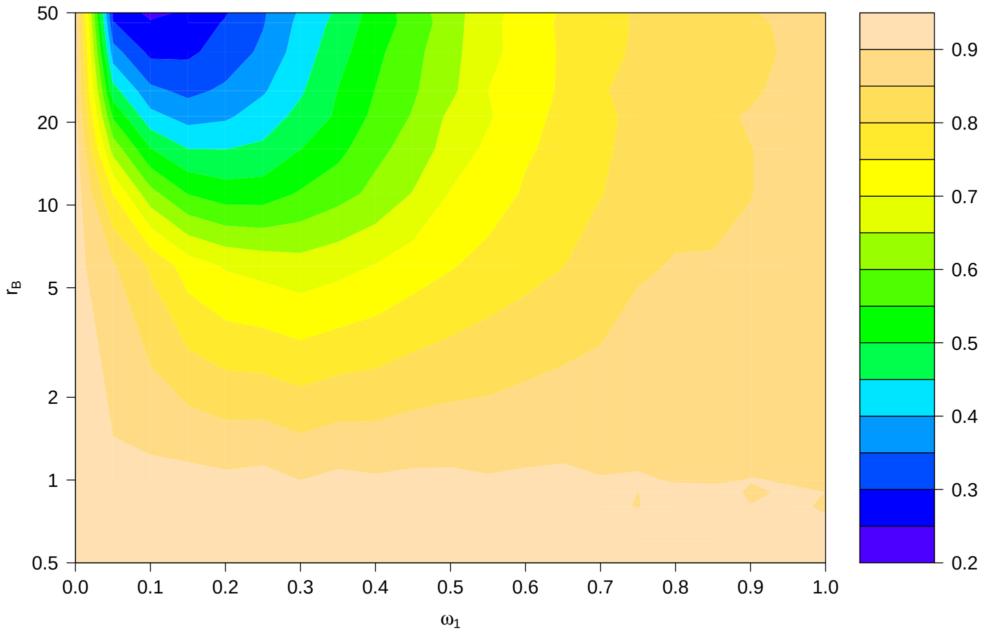
Since $\mathrm{Cor}(FOM, S) \geq \frac{w_0\sqrt{2/\pi}}{\sqrt{1+\omega_1^2 N}}$, it follows then that if r_B gets large, then $\mathrm{Cor}(CE, S)$ drops below $\mathrm{Cor}(FOM, S)$. This then confirms our results in our baseline model.



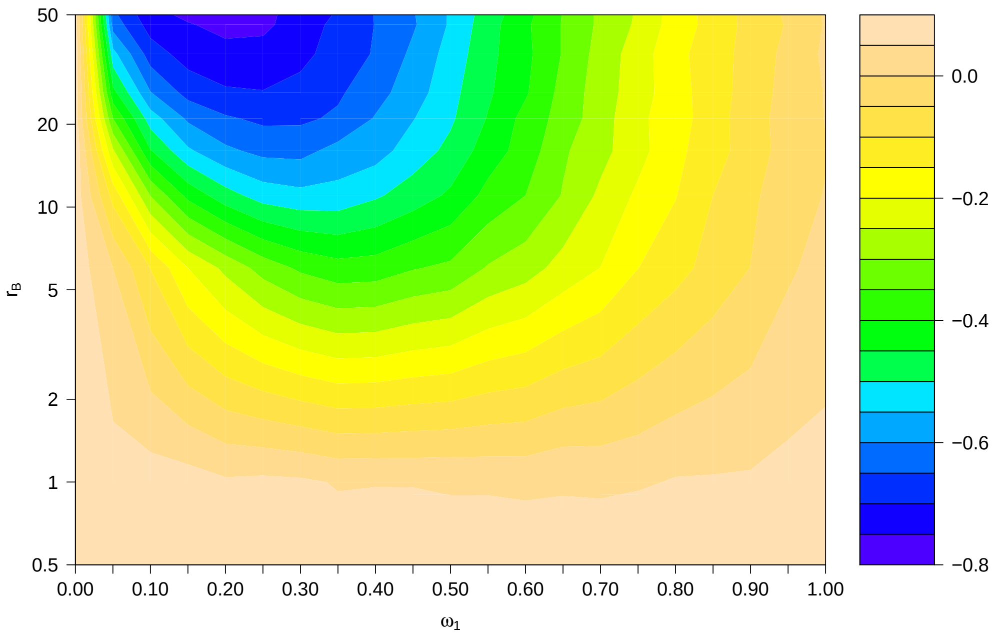
Appendix Figure 1: The contour plot of $\text{Cor}[CE, S] - \text{Cor}[FOM, S]$ as a function of r_F and N in unbiased forecasts benchmark case. The contour value is the difference between the correlations of consensus errors CE and fraction of misses FOM to S the market surprise, the y-axis is N the number of analysts, and the x-axis is $r_F = \sigma_F / \sigma_A$ the ratio between the standard deviation of forecasts and the actual (shown in log-scale).



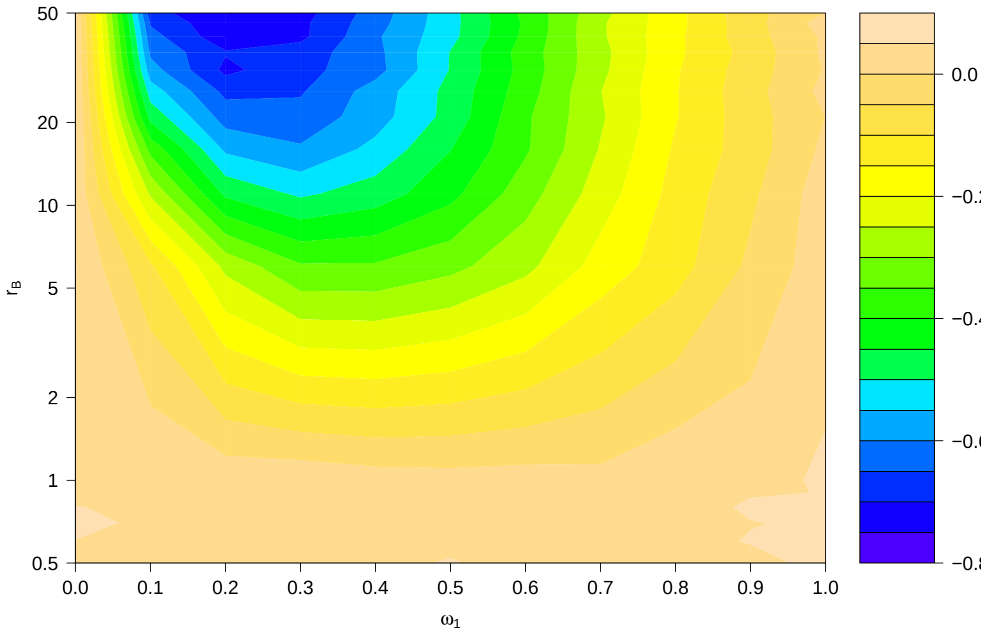
Appendix Figure 2: The comparison between the correlations of consensus errors CE and fraction of misses FOM to S the market surprise as a function of r_F for fixed number of analysts $N = 10$, where $r_F = \sigma_F/\sigma_A$ is the ratio between the standard deviation of forecasts and the actual (shown in log-scale).



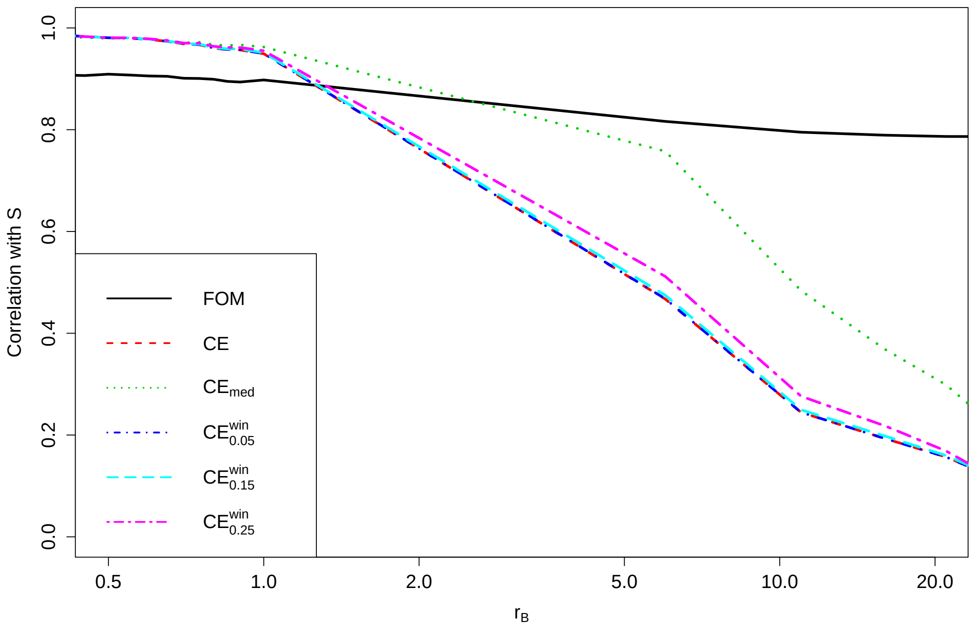
Appendix Figure 3: The contour plot of Cor$[CE, FOM]$ as a function of the key parameters ω_1 and r_B in biased forecasts case. The contour value is the correlation between consensus errors CE and fraction of misses FOM, the y-axis is ω_1 the proportion of biased forecasts, and the x-axis is $r_B = \sigma_B/\sigma_A$ the ratio between the standard deviation of aggregated bias and the actual (shown in log-scale). The other parameters in the model are set as $r_F = 1/2$, $r_b = r_B/5$ and $N = 20$.



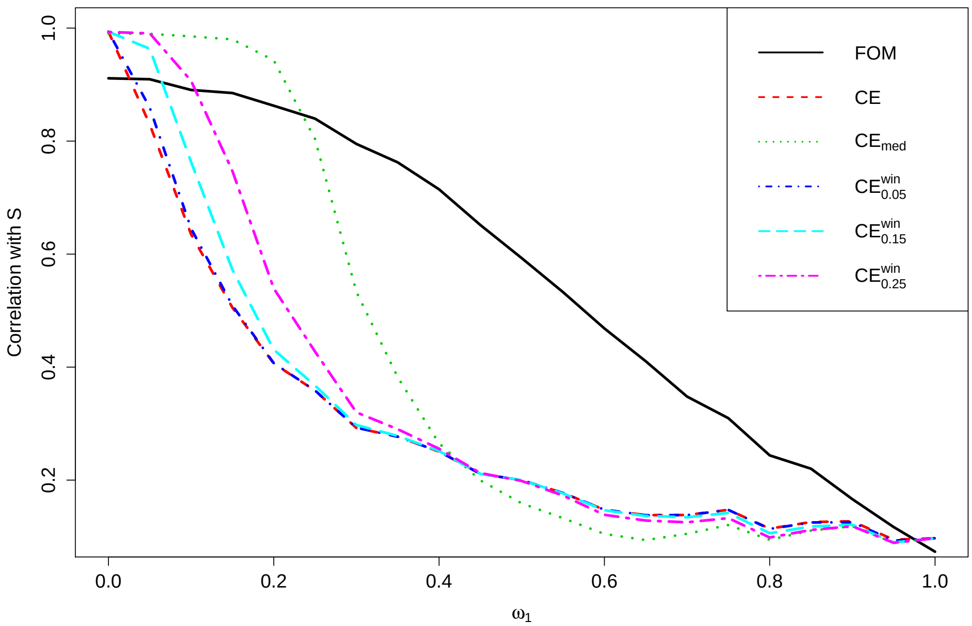
Appendix Figure 4: The contour plot of $\text{Cor}[CE, S] - \text{Cor}[FOM, S]$ as a function of the key parameters ω_1 and r_B in biased forecasts case. The contour value is the difference between the correlations of consensus errors CE and fraction of misses FOM to S the market surprise, the y-axis is ω_1 the proportion of biased forecasts, and the x-axis is $r_B = \sigma_B/\sigma_A$ the ratio between the standard deviation of aggregated bias and the actual (shown in log-scale). The other parameters in the model are set as $r_F = 1/2$, $r_b = r_B/5$ and $N = 20$.



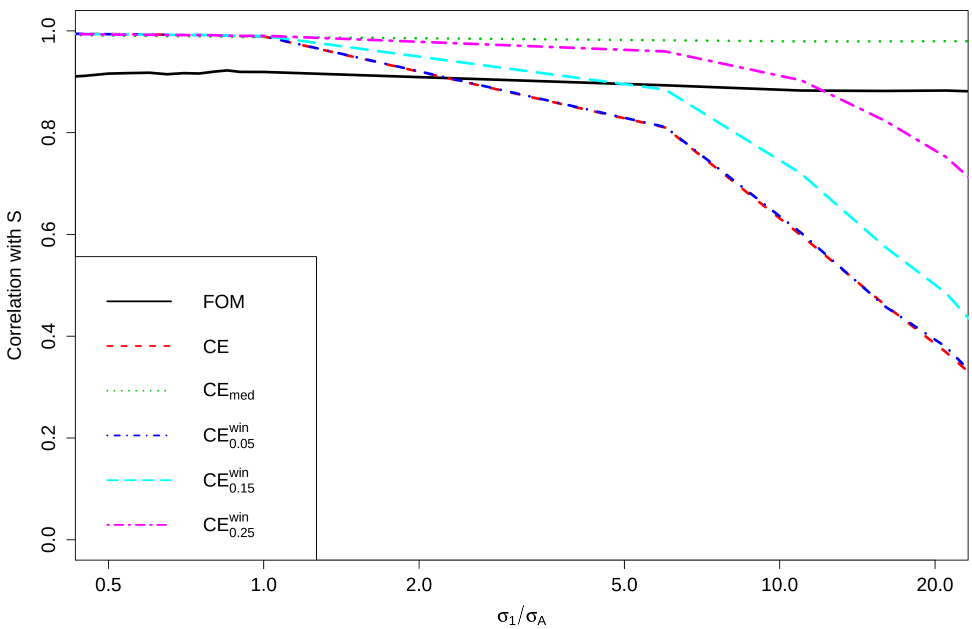
Appendix Figure 5: The contour plot of Cor[*Rank(CE)*, S] − Cor[*FOM*, S] as a function of the key parameters ω_1 and r_B in biased forecasts case. The contour value is the difference between the correlations of the rank score of consensus errors *Rank(CE)* and fraction of misses *FOM* to S the market surprise, the y-axis is ω_1 the proportion of biased forecasts, and the x-axis is $r_B = \sigma_B/\sigma_A$ the ratio between the standard deviation of aggregated bias and the actual (shown in log-scale). The other parameters in the model are set as $r_F = 1/2$, $r_b = r_B/5$ and $N = 20$.



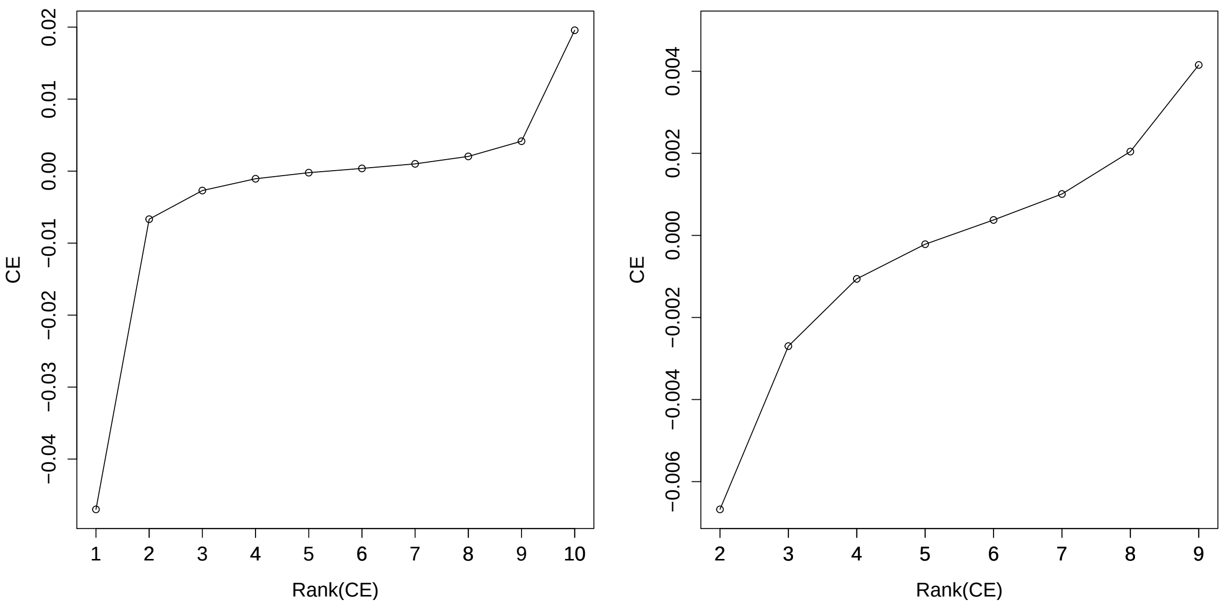
Appendix Figure 6: The comparison between the correlations of fraction of misses *FOM* and different Winsorized measures $CE_{\lambda}^{\text{win}}$ to S the market surprise as a function of r_B in biased forecasts case, where $r_B = \sigma_B/\sigma_A$ is the ratio between the standard deviation of aggregated bias and the actual (shown in log-scale). The other parameters in the model are set as $\omega_1 = 0.3$, $r_F = 1/2$, $r_b = r_B/5$ and $N = 20$.



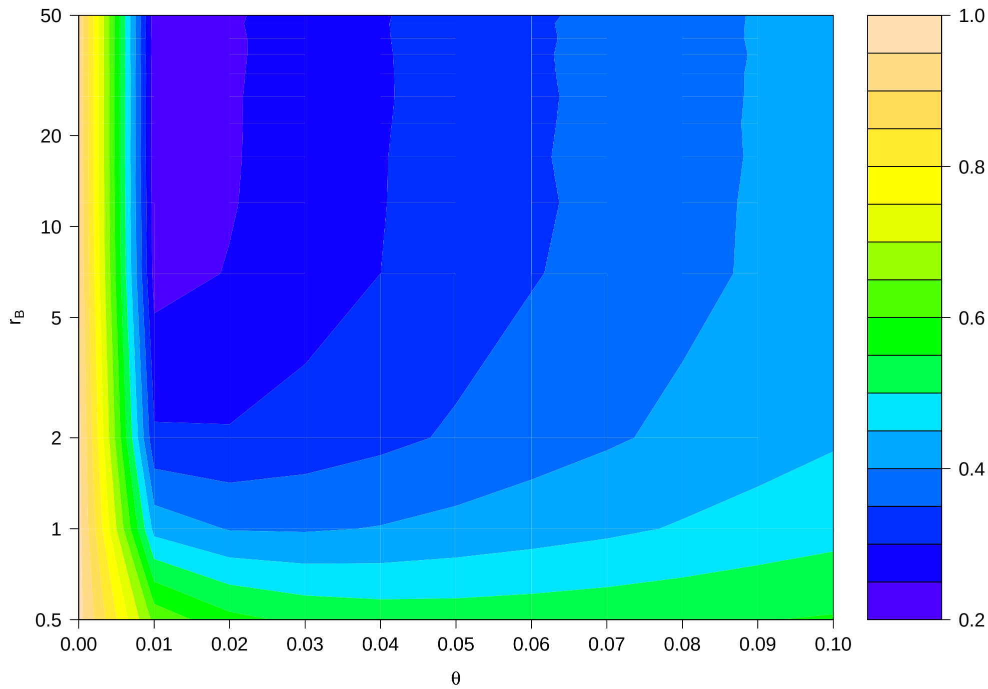
Appendix Figure 7: The comparison between the correlations of fraction of misses *FOM* and different winsorized measures $CE_{\lambda}^{\text{win}}$ to S the market surprise as a function of ω_1 in biased forecasts case, where ω_1 is the proportion of biased forecasts. The other parameters in the model are set as $r_B = 10$, $r_F = 1/2$, $r_b = r_B/5$ and $N = 20$.



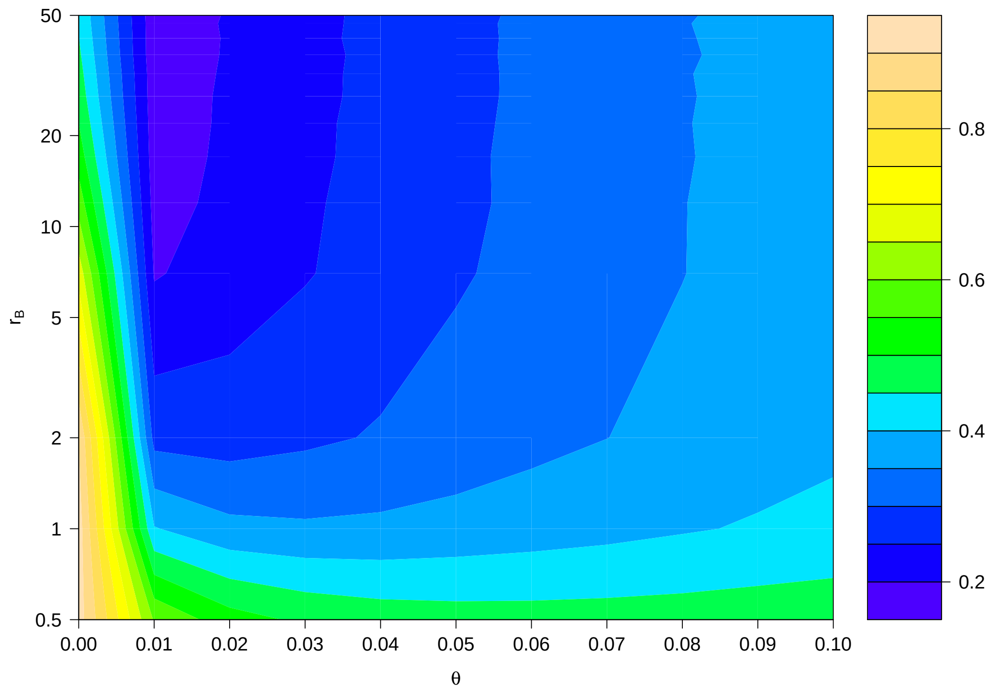
Appendix Figure 8: The comparison between the correlations of fraction of misses FOM and different winsorized measures $CE_{\lambda}^{\text{win}}$ to S the market surprise as a function of σ_1/σ_A (shown in log-scale) under the alternative modelling without introducing bias, where σ_1 is the variance of bad forecasts. The other parameters in the model are set as $\omega_1 = 0.3$, $\sigma_0/\sigma_A = 1/2$ and $N = 20$.



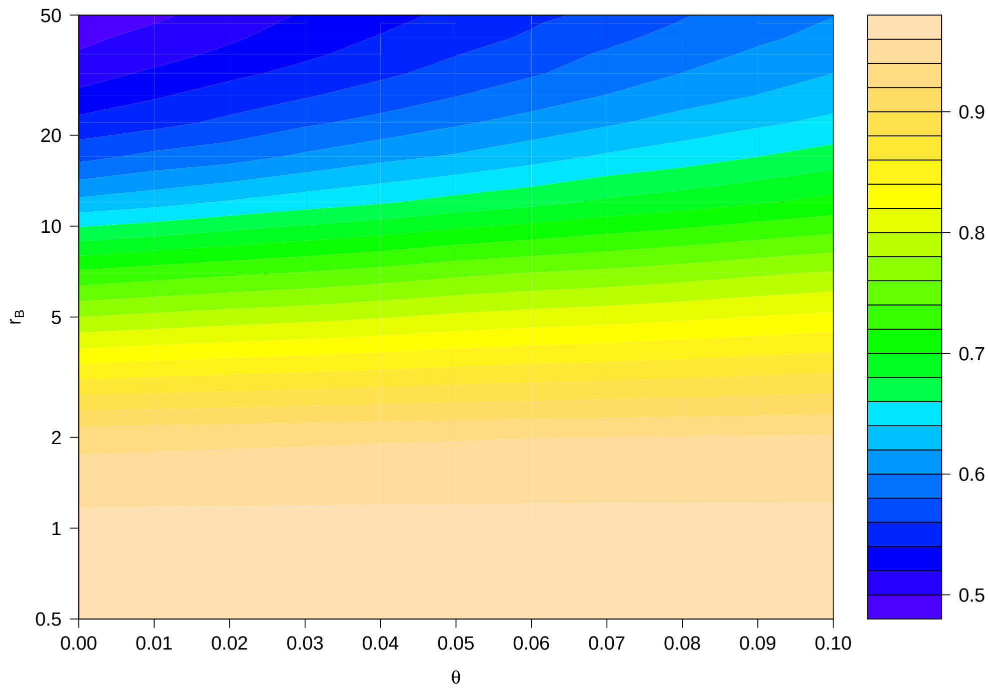
Appendix Figure 9: Average *CE* against *Rank(CE)* in earnings data. Left: over the whole sample; Right: conditional on *Rank(CE)* not being in the top or bottom decile.



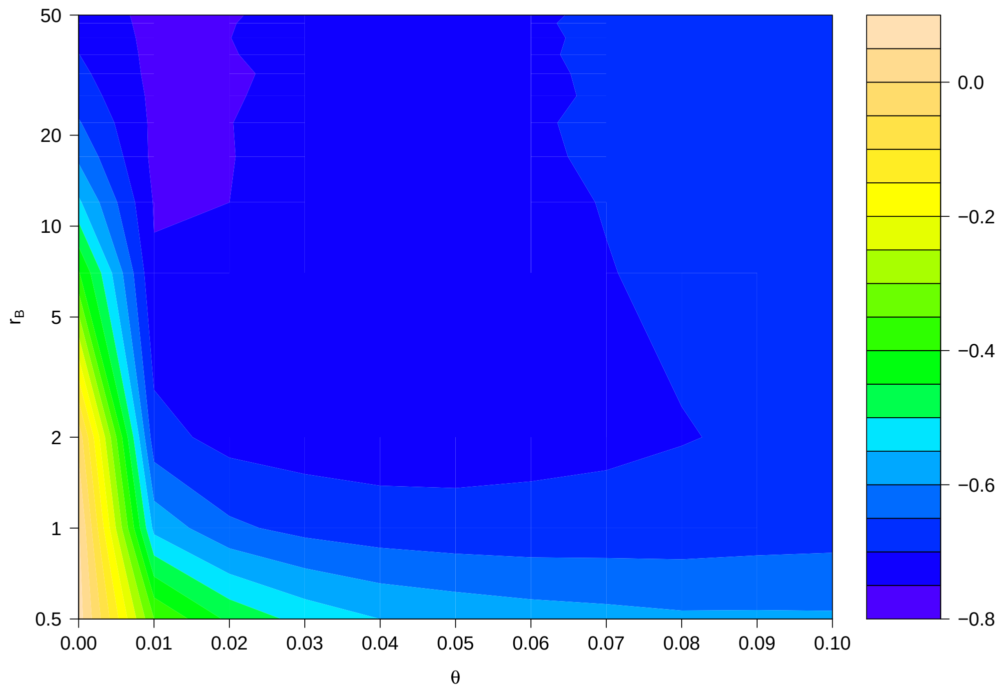
Appendix Figure 10: The contour plot of Cor[$CE, Rank(CE)$] as a function of the key parameters θ and r_B under the extended model allowing for outliers in CE. The contour value is the correlation between consensus errors CE and its rank score $Rank(CE)$, the x-axis is θ the probability of tail events as defined in Section A.5, and the y-axis is $r_B = \sigma_B/\sigma_A$ the ratio between the standard deviation of aggregated bias and the actual (shown in log-scale). The other parameters in the model are set as $\omega_1 = 0.3$, $r_F = 1/2$, $r_b = r_B/5$ and $N = 20$.



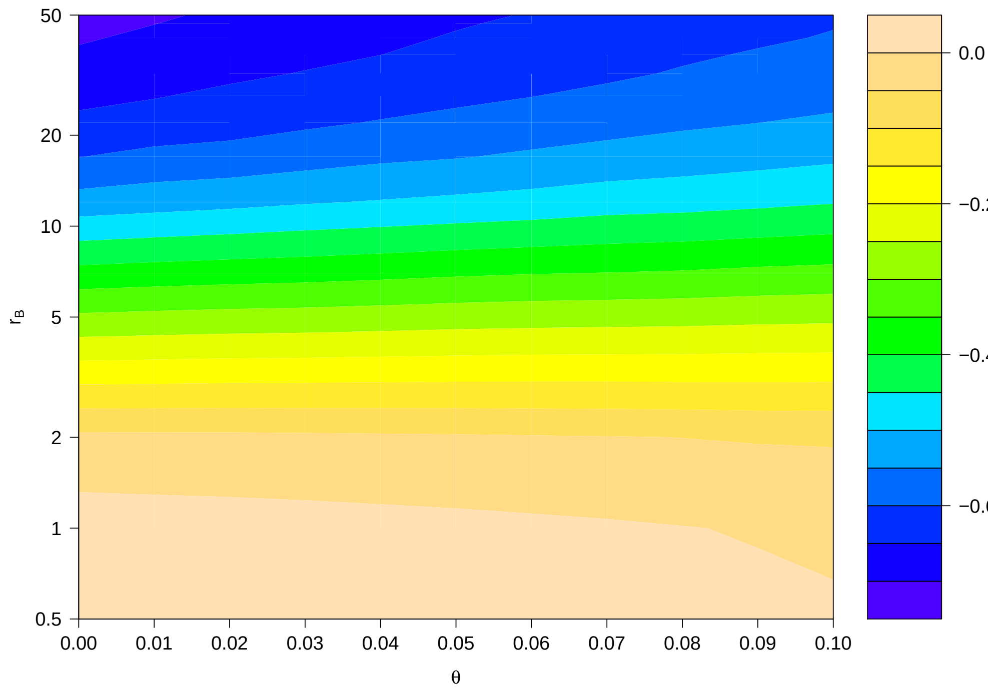
Appendix Figure 11: The contour plot of Cor[CE, FOM] as a function of the key parameters θ and r_B under the extended model allowing for outliers in CE. The contour value is the correlation between consensus errors CE and fraction of misses FOM, the x-axis is θ the probability of tail events as defined in Section A.5, and the y-axis is $r_B = \sigma_B/\sigma_A$ the ratio between the standard deviation of aggregated bias and the actual (shown in log-scale). The other parameters in the model are set as $\omega_1 = 0.3$, $r_F = 1/2$, $r_b = r_B/5$ and $N = 20$.



Appendix Figure 12: The contour plot of Cor$[FOM, Rank(CE)]$ as a function of the key parameters θ and r_B under the extended model allowing for outliers in *CE*. The contour value is the correlation between fraction of misses *FOM* and the rank score of consensus errors *Rank(CE)*, the x-axis is θ the probability of tail events as defined in Section A.5, and the y-axis is $r_B = \sigma_B/\sigma_A$ the ratio between the standard deviation of aggregated bias and the actual (shown in log-scale). The other parameters in the model are set as $\omega_1 = 0.3$, $r_F = 1/2$, $r_b = r_B/5$ and $N = 20$.



Appendix Figure 13: The contour plot of $\text{Cor}[CE, S] - \text{Cor}[FOM, S]$ as a function of the key parameters θ and r_B under the extended model allowing for outliers in CE. The contour value is the difference between the correlations of consensus errors CE and fraction of misses FOM to S the market surprise, the x-axis is θ the probability of tail events as defined in Section A.5, and the y-axis is $r_B = \sigma_B/\sigma_A$ the ratio between the standard deviation of aggregated bias and the actual (shown in log-scale). The other parameters in the model are set as $\omega_1 = 0.3$, $r_F = 1/2$, $r_b = r_B/5$ and $N = 20$.



Appendix Figure 14: The contour plot of Cor[*Rank(CE)*, S] − Cor[*FOM*, S] as a function of the key parameters θ and r_B under the extended model allowing for outliers in *CE*. The contour value is the difference between the correlations of the rank score of consensus errors *Rank(CE)* and fraction of misses *FOM* to S the market surprise, the x-axis is θ the probability of tail events as defined in Section A.5, and the y-axis is $r_B = \sigma_B/\sigma_A$ the ratio between the standard deviation of aggregated bias and the actual (shown in log-scale). The other parameters in the model are set as $\omega_1 = 0.3$, $r_F = 1/2$, $r_b = r_B/5$ and $N = 20$.