Robust Measures of Earnings Surprises

Chin-Han Chiang† Wei Dai‡ Jianqing Fan§ Harrison Hong¶
Jun Tu∥

First Draft: June 20, 2014
This Draft: October, 2017

Abstract

Event studies of market efficiency measure an earnings surprise with the consensus error ($CE$), defined as earnings minus the average of professional forecasts. Even if a subset of forecasts can be biased, the ideal but difficult to estimate parameter-dependent alternative to $CE$ is a nonlinear filter of individual errors that adjusts for bias. We show that $CE$ is a poor parameter-free approximation for this ideal measure. The fraction of misses on the same side ($FOM$), by discarding the magnitude of misses, offers a far-better approximation. $FOM$ performs particularly well against $CE$ in predicting the returns of US stocks, where bias is potentially large, than that of international stocks.

---

*We thank Ken Singleton (Editor) and anonymous referees for many helpful comments. We also 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 Management University, Seoul National University and Emory University. Jun Tu acknowledges support from Sim Kee Boon Institute of Financial Economics. This paper was previously titled "When Everyone Misses on the Same Side: Debiased Earnings Surprises and Stock Returns". Please address correspondences to Jun Tu, Singapore Management University, tujun@smu.edu.sg.

†International Monetary Fund
‡Dimensional Fund Advisors
§Princeton University
¶Columbia University and NBER
∥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 ($\text{CAR}$) or subsequent to the announcement date ($\text{POSTCAR}$) on the consensus error ($\text{CE}$): the more positive the consensus error $\text{CE}$ the higher is the $\text{CAR}$ and also the higher is the $\text{POSTCAR}$ (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. One important reason is analysts’ conflicts of interest. A large literature shows that a number of firms and analysts engage in an earnings-guidance game where analysts make optimistic forecasts at the start of the year and then walk down their estimates to a level the firm can beat by the end of the year (see, e.g., Richardson, Teoh, and Wysocki (2004)).\(^1\) These biases in earnings forecasts tend to be stronger in the US stock market, where firm issuance incentives are more likely to distort earnings forecasts, than in international markets (see, e.g., Chan, Karceski, and Lakonishok

\(^1\)See also Brown (2001), Bartov, Givoly, and Hayn (2002), and Kasznik and McNichols (2002) for studies on biases in earnings surprises and the returns to firms beating analyst expectations. Complementary evidence on the importance of career concerns for analyst forecast bias include Hong and Kubik (2003).
Another rationale 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)). In short, some fraction of individual forecasts can be significantly biased depending on the exposure of analysts to differing incentives.

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 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 to 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 since these statistics are meant to deal with outliers and not systematic bias of forecasts.

To deal with this problem, we first articulate a tractable and empirically sensible model of earnings and forecasts where some fraction of individual forecasts can be biased. This bias has an aggregate component that can have a non-zero mean and variance. We then derive the ideal measure of the true earnings surprise, defined as the filter of individual
forecast errors that is maximally correlated with the true surprise assuming the parameters
governing the individual forecasts are known. It is in general a non-linear filter of individual
forecast errors that accounts for biases and precisions. We show that this non-linear filter
optimally down-weighs extreme misses of individual forecast errors when there is potential
for bias. Even if the bias has a mean zero, the optimal filter will down-weigh extreme misses
as long as there is ex-ante uncertainty or variance in the size of the aggregate bias. But it is
hard to estimate the underlying parameters of individual forecasts to implement this filter
in practice, as we show below. This is no doubt an important reason why the literature
has retained the use of CE, which is parametric-free, in the face of extensive evidence of
individual forecast biases.

We then develop a robustness criterion where we evaluate the robustness of a given
parametric-free measure of earnings surprise, be it CE or some other measure, as defined
by how well it approximates this ideal measure. The relative efficiency of CE to this full
information benchmark is close to zero when the potential bias is large, i.e, CE is not a
robust earnings surprise measure.

We prove that this ideal earnings surprise measure is far better approximated by the
fraction of forecasts that miss on the same side or FOM. 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 \(I_{Actual>\text{All}} = 1\) (0 otherwise). 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). In the case where \( N = 1 \), \( FOM \) either equals 1, 0 or -1.

\( FOM \) is equivalent to taking the average of the signs (either -1, 0, or 1) of individual forecast errors. \( FOM \) discards information on the magnitude of misses much in the same way that the ideal filter would down-weigh large misses when potential bias is significant. We provide a theoretical lower bound on the efficiency of \( FOM \) relative to this full information benchmark. The lower bound on the relative efficiency \( FOM \) can be as high as 50% using plausible parameter values. \( FOM \) shares the strength of \( CE \) in that it is parametric-free but is less sensitive to biased forecasts and hence potentially superior to \( CE \). We use earnings forecasts to frame our model and motivate our empirical analysis but the methodology and idea apply equally to other types of economic forecasts where bias is potentially important.

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. 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. Since \( FOM \) is more robust in that it tracks the market surprises better than \( CE \), we expect that a regression of \( CAR \) or \( POSTCAR \) on \( FOM \) to have superior explanatory power relative to \( CE \) when potential bias is important.

Using the median of the forecasts rather than the mean as the consensus does not help the \( CE \) measure since medians deal with outliers but not when a significant fraction of the forecasts are biased. In practice, event studies are implemented using a transformation of \( CE \) into a cross-sectional decile score from 1 to 10, which we call \( \text{Rank}(CE) \). The \( \text{Rank}(CE) \) measure deals with outliers and offers a better fit for \( CAR \) and \( POSTCAR \) than \( CE \) (see Hirshleifer, Lim, and Teoh (2009)). 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 to have positive or negative surprises.

Using \( CE \) constructed with the average of the recent forecasts for the annual year-end FY1 earnings and the alternative \( FOM \), we test the predictions of our model. Every firm has one
observation per year over the sample period from 1983 to 2015. In a canonical regression of \textit{CAR} (measured using the 3-day firm-size-adjusted return around the announcement date), a one standard deviation increase in \textit{Rank(CE)} increases the \textit{CAR} by 1.2%, a sizeable economic effect. For \textit{POSTCAR}, the portfolio long positive earnings surprise (decile rank score 10) and short negative earnings surprise (decile rank score 1) yields a return of less than 1% over the subsequent six months (126 trading days to be exact) after the announcement date or 2% annualized.

Our \textit{FOM} variable, however, performs better than \textit{CE} or \textit{Rank(CE)}. A one standard deviation increase in \textit{FOM} increases \textit{CAR} by 1.5%. The improvement in $R^2$ is around 20%. When we run a horse race of \textit{FOM} and \textit{Rank(CE)}, the coefficient in front of \textit{FOM} is virtually unchanged whereas the one in front of \textit{Rank(CE)} is no longer significant. For the \textit{POSTCAR}, a portfolio long \textit{FOM} = 1 stocks and short \textit{FOM} = −1 stocks yields a six-month subsequent return of 3% or 6% annualized. Again, in a multiple regression to explain \textit{POSTCAR}, our \textit{FOM} measure remains significant, whereas \textit{Rank(CE)} is insignificant.

Rather than using OLS, we re-estimate our model using maximum likelihood so we can implement Bayesian Information Criterion (BIC) for non-nested model selection, i.e. choosing between \textit{Rank(CE)} and \textit{FOM} to explain \textit{CAR} and \textit{POSTCAR}. The difference in BIC (400 for \textit{CAR} and 50 for \textit{POSTCAR}) far exceeds 10, pointing to the overwhelming preference for \textit{FOM} as the independent variable in a model to explain \textit{CAR} and \textit{POSTCAR}. We also conduct a nested likelihood ratio tests to compare a model of \textit{CAR} and \textit{POSTCAR} using \textit{FOM} versus using both \textit{FOM} and \textit{CE}. We cannot reject that the univariate model with \textit{FOM} is sufficient for the \textit{POSTCAR} regression.

We then conduct these BIC tests for international stocks where potential bias is smaller than in US stocks. Consistent with our model, we find that the incremental improvement of \textit{FOM} compared to \textit{CE} is smaller for international stocks. We also relate our \textit{FOM} measure to the walk-down effect in earnings forecasts and earnings persistence. Finally, we show that a parametric-dependent method involving estimating individual analyst bias parameters based
on their forecast histories and adjusting individual forecasts before calculating CE yields equally poor results as using the parametric-free CE.

Our paper proceeds as follows. We present a model to evaluate the robustness of alternative parametric-free earnings surprise measures 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 Robustness of Earnings Surprise Measures

In this section, we develop a model to evaluate the robustness of parametric-free measures of earnings surprises when analyst forecasts are potentially biased.

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

\[ A = e + \epsilon_A, \]  

(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

\[ S = A - e. \]  

(2)

We then assume that an individual forecast \( F_i \) is the market expectation \( e \) plus some
noise $\epsilon_i$ and a possible bias term $b_i$:

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

where $\epsilon_i \sim N(0, \sigma_F^2)$ and is uncorrelated with $\epsilon_A$. Each forecast is unbiased with probability $\omega_0$, 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 $B \sim N(\mu_B, \sigma_B^2)$ is drawn first, and conditional on this realized $B$ individual bias $b_i$ follows $N(B, \sigma_b^2)$. Note that while $\omega_0$ and $\omega_1$ 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 $\omega_1$ of them are biased by a random magnitude.

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. $\epsilon_A$ is the unexpected shock to earnings which the market cannot know. The bias $b_i$ can be derived in a number of ways as we discussed above.

### 2.2 Ideal Earnings Surprise Measure $f^*$ versus CE and FOM

Since the earning surprise $S \sim N(0, \sigma_A^2)$, the error of the $i^{th}$ forecast is

$$U_i = A - F_i = S + Y_i,$$

where

$$Y_i \sim N(0, \sigma_F^2) \text{ with probability } \omega_0$$

$$\sim N(-b_i, \sigma_F^2) \text{ with probability } \omega_1 = 1 - \omega_0$$
and the $b_i$’s are drawn from $\mathcal{N}(B, \sigma_b^2)$ conditional on the realization of $B \sim \mathcal{N}(\mu_B, \sigma_B^2)$.

We first derive an ideal earnings surprise measure assuming all the parameters in the model are known. The ideal measure should be a function of $U_i$’s that highly correlates with the unobservable surprise $S$:

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

It is well-known that $f^*$ is the linear projection of $S$ onto the space of $U_i$’s:

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

This conditional expectation can be written as

$$f^*(\mathbf{U}) = \mathbb{E}[S|\mathbf{U}] = \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} S g(S, B, \mathbf{U}) dS dB}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(S, B, \mathbf{U}) dS dB},$$

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

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

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

The optimal choice $f^*(\mathbf{U})$ is highly nonlinear and in general can not be analytically derived. It depends on many unknown parameters ($\omega_0$, $\mu_B$, $\sigma_B^2$, $\sigma_b^2$, $\sigma_F^2$ and $\sigma_A^2$). Since these parameters are not known to the econometrician, it is in practice difficult to implement this optimal measure as we demonstrate in Section 4.7 below.

This is no doubt one reason why the literature uses a parametric-free measure. The main measure used by far is the consensus forecast error (CE) which is simply the average of the
forecast errors:

\[ CE = \frac{1}{N} \sum_{i=1}^{N} U_i. \]  

But we now show that \( CE \) is a poor approximation of the ideal filter \( f^* \). We propose instead a better approximation, \( FOM \), which is an average of the signs of the forecast errors:

\[ FOM = \frac{1}{N} \sum_{i=1}^{N} \text{sgn}(U_i), \]  

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

To get an intuition for why this is the case, we focus on the \( N = 1 \) case. Although this simple case contains only one forecast and does not show the interaction among forecasts, it nevertheless provides insights into the robustness of \( FOM \) to bias, since analysts are stochastically exchangeable under our model. It is also empirically relevant since many of the stocks in our sample only have one forecast. In this instance the joint distribution of \( S \), \( B \) and \( U \) is

\[ g(S, B, U) = \phi(S; \sigma^2_A)\phi(-B + \mu_B; \sigma^2_B) \left( \omega_0 \phi(U - S; \sigma^2_F) + \omega_1 \phi(U - S + B; \sigma^2_F + \sigma^2_B) \right). \]

Given the following identities,

\[
\int_{-\infty}^{\infty} \phi(x; \sigma^2_x)\phi(y - x; \sigma^2_y)dx = \phi(y; \sigma^2_x + \sigma^2_y)
\]

\[
\int_{-\infty}^{\infty} x\phi(x; \sigma^2_x)\phi(y - x; \sigma^2_y)dx = \frac{\sigma^2_y}{\sigma^2_x + \sigma^2_y} \phi(y; \sigma^2_x + \sigma^2_y) \cdot y
\]

we have

\[ f^*(U) = \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(S, B, U)SdB dSdB}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(S, B, U)SdB dSdB} \]

\[ = \frac{\omega_0 \frac{\sigma^2}{\sigma^2_F + \sigma^2_A} \phi(U; \sigma^2_F + \sigma^2_A) \cdot U + \omega_1 \frac{\sigma^2}{\sigma^2_F + \sigma^2_A + \sigma^2_B + \sigma^2_B} \phi(U + \mu_B; \sigma^2_F + \sigma^2_A + \sigma^2_B + \sigma^2_B) \cdot (U + \mu_B)}{\omega_0 \phi(U; \sigma^2_F + \sigma^2_A) + \omega_1 \phi(U + \mu_B; \sigma^2_F + \sigma^2_A + \sigma^2_B + \sigma^2_B)}. \]
We plot in Figures 1 and 2 \( f^* = E[S|U] \), \( CE \) and \( FOM \) in the simple case of one analyst \((N = 1)\) where analytical formulas allow for ease of comparison. These figures depict the shape of the optimal transformation \( f^* = E[S|U] \) for a few different configurations (different \( \sigma_B \) in Figure 1 and different \( \mu_B \) in Figure 2) as a function of the one forecast error \( U \). Recall that \( w_0 \) is just the probability that the forecast is unbiased. It is highly nonlinear and asymmetric in \( U \) when \( \mu_B \neq 0 \). When the aggregate bias \( B \) is positive (negative) on average, \( U \) is more likely to be negatively (positively) biased, so the ideal measure discounts negative (positive) values of \( U \) more. Also notice how the optimal measure levels off as \( U \)'s magnitude increases in both directions - because of the existence of biases, the benefit of ignoring the magnitude outweighs the loss of information. Our proposed measure \( FOM \) is motivated by this insight.

### 2.3 Robustness Criterion

We can quantify the extent to which \( CE \) and \( FOM \) approximate the ideal filter \( f^*(U) \) by comparing the efficiency loss of using these approximations. The relative efficiency of \( CE \) to the optimal but infeasible measure \( f^*(U) \) is defined as

\[
\lim_{\sigma_B^2 \to \infty} \frac{\text{Cor}[CE, S]}{\text{Cor}[f^*(U), S]}.
\]

Similarly, the relative efficiency of \( FOM \) is defined as

\[
\lim_{\sigma_B^2 \to \infty} \frac{\text{Cor}[FOM, S]}{\text{Cor}[f^*(U), S]}.
\]

These relative efficiency calculations provide a metric to quantify the robustness of \( FOM \) versus \( CE \). Notice that these calculations depend on \( \sigma_B \) as opposed to \( \mu_B \). While the discussion above shows that both play a role in shaping the bias of the earnings surprise measures, we focus on \( \sigma_B \) for a few reasons. First, the empirical evidence very much points to a large degree of potential bias both on the positive or negative side depending on incentives.
and horizons of forecasts. Second, from Lemma 1, $\sigma_B$ remains a large source of bias relative to $\mu_B$ even when $N$ is large. As we show below, regardless of $\mu_B$, a large enough $\sigma_B$ will ensure that FOM is more robust than CE.

To facilitate our analysis below, we establish two lemmas concerning the correlation of CE and FOM with $S$.

**Lemma 1.** The correlation of CE with the market surprise $S$ is given by

$$\text{Cor}[CE, S] = \frac{\sigma_A^2}{\sigma_A \sqrt{\frac{1}{N}(\sigma_F^2 + \omega^2 \sigma_b^2 + \omega^2 \omega_1 \sigma_B^2 + \omega^2 \omega_1 \mu_B^2) + \sigma_A^2 + \omega^2 \sigma_B^2}}.$$  \hspace{1cm} (6)

Notice that either a large $\mu_B$ or $\sigma_B$ will lead to a low correlation. In particular, it follows that $\lim_{\sigma_B \to \infty} \text{Cor}[CE, S] = 0$.

**Lemma 2.** The correlation of FOM with the market surprise $S$ is given by

$$\text{Cor}[FOM, S] \geq \frac{2\omega_0}{\sqrt{2\pi(1 + \sigma_F^2/\sigma_A^2)}}.$$  \hspace{1cm} (7)

This lower bound is independent of $\mu_B$ and $\sigma_B$. This suggests that FOM retains at least a fraction of useful information no matter how bad the bias is, so it is expected that FOM will outperform CE as $\sigma_B$ grows.

It follows from Lemma 1 that the relative efficiency of CE can be seriously compromised in the presence of bias

$$\lim_{\sigma_B \to \infty} \frac{\text{Cor}[CE, S]}{\text{Cor}[f^*(U), S]} = 0,$$  \hspace{1cm} (8)

i.e. there is a complete loss of efficiency using CE as an approximation for $f^*(U)$.

In contrast, we show that the efficiency loss of using FOM relative to the ideal $f^*$ is bounded and dependent critically on $w_0$, the fraction of unbiased forecasts. We first derive a lower bound for the case of $N = 1$ forecast in Proposition 1. We then derive in Proposition 2 a looser bound for the general $N$ case.

**Proposition 1.** For $N = 1$ forecast, the relative efficiency of FOM to the optimal but
The infeasible measure $f^*(U)$ is given by

$$
\lim_{\sigma_h^2 \to \infty} \frac{\text{Cor}[\text{FOM}, S]}{\text{Cor}[f^*(U), S]} \geq \sqrt{\frac{2}{\pi}} \cdot \omega_0^{1/2}.
$$

(9)

This lower bound depends on $\omega_0$ only, the probability of drawing an unbiased forecast. Panel A of Table 1 gives the numerical value of this lower bound. In the extreme case when the variance of aggregate biases is infinitely large, FOM still retains a decent efficiency relative to the optimal but infeasible measure - the relative efficiency is 56.4% when $\omega_0 = 0.5$ (the probability of unbiased forecasts is one-half), and goes up to 79.8% as $\omega_0$ increases (higher probability of unbiased forecasts).

In the general case of $N > 1$ forecasts, we can still provide a lower bound of the efficiency of our simple parameter-free measure FOM relative to this optimal measure $f^*(U)$, but it will not be as tight as in the case of $N = 1$. Our derivation assumes that, for a given set of $N$ forecasts, we can condition on both $U$ and additional information on the subset of forecasts that are unbiased. We call this subset of unbiased forecasts $A$ and denote the size of this set by $|A|$, 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 $h(U, A)$ (which also includes $f^*(U)$ 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 $U$ and $A$ should be larger:

$$
\text{Cor}[f^*(U), S] \leq \max_h \text{Cor}[S, h(U, A)].
$$

(10)

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

**Proposition 2.** For an arbitrary number $N$ forecasts, the relative efficiency of FOM to the optimal but infeasible measure $f^*(U)$ is given by

$$
\lim_{\sigma_b^2 \to \infty} \frac{\text{Cor}[\text{FOM}, S]}{\text{Cor}[f^*(U), S]} \geq \lim_{\sigma_b^2 \to \infty} \frac{\text{Cor}[\text{FOM}, S]}{\max_h \text{Cor}[S, h(U, A)]} = \frac{2\omega_0 E[Z \Phi(Z/r_F)]}{\sqrt{E\left[\frac{1}{1+r_F^2/|A|}\right]}},
$$

(11)

where $A$ is the subset of unbiased forecasts in a given draw of $N$ forecasts, $|A|$ is the size of the $A$, and $r_F = \frac{\sigma_F}{\sigma_A}$ is the ratio of the standard deviation of analysts’ forecast errors to the standard deviation of the earnings shock.

The numerator of the third term in (11) can be evaluated analytically.\(^2\) Using this and the fact that $|A| \sim \text{Binomial}(N, \omega_0)$ we can evaluate (11), which only depends on $N$, $r_F$ and $\omega_0$.

Panel B of Table 1 summarizes the numerical values of the lower bound for $N = 10$ 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 decreases 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 0.5 to 2 seem plausible as the standard deviation of analyst forecasts are on par with how much actual uncertainty there is about the stock.

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

13
At these parameter values, the lower bound is 0.5. 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.

### 2.4 Comparative Statics

If we were to compute a crude bound of the improvement of $FOM$ to $CE$, we have

\[
\text{Cor}[FOM, S] - \text{Cor}[CE, S] \geq \frac{2\omega_0}{\sqrt{2\pi(1 + \sigma_F^2/\sigma_A^2)}} - \frac{\sigma_A^2}{\sigma_A \sqrt{\frac{1}{N} \left(\sigma_F^2 + \omega_1\sigma_b^2 + \omega_0\omega_1\sigma_B^2 + \omega_0\omega_1\mu_B^2\right) + \sigma_A^2 + \omega_1^2\sigma_B^2}} - \frac{2\omega_0}{\sqrt{2\pi(1 + \sigma_F^2/\sigma_A^2)}} - \frac{\sigma_A}{\sqrt{\sigma_A^2 + \omega_1^2\sigma_B^2}}.
\]

That is, (12) is positive as long as $\sigma_B^2 \geq \frac{\pi(\sigma_F^2 + \sigma_A^2) - 2\omega_0\omega_1\sigma_A^2}{2\omega_0^2\omega_1}$. Notice that this condition becomes harder to satisfy as $\sigma_F^2$, the variance of individual forecast errors rise. We will use this fact to relate our $FOM$ measure to the walk-down effect of analyst forecast bias below.

### 2.5 Robustness to Outliers

We now add tail events with small probability to the baseline model - this is motivated by the fact that in the real data occasionally most analysts or even everyone miss by a huge margin. An extreme $CE$ has a magnitude multiple times more than the majority of observations (e.g., the 3% on two tails is 30 times of the central 97% in average absolute value). In contrast, the impact of outliers on $FOM$ is very limited because by design the distortion is not magnified by the magnitude. We formally show here that the potential outliers have minimal impact on the efficiency of $FOM$ compared to the baseline case.

We model potential outliers in a straightforward manner: with probability $1 - \theta$, the forecasts follow the baseline model; with probability $\theta$ which is supposed to be very small,
all the forecasts are off by a potentially huge margin:

\[ F_i = e + \tilde{b}_i + \epsilon_i, \quad i = 1, \ldots, N \tag{13} \]

where \( \epsilon_i \sim \mathcal{N}(0, \sigma_F^2) \), and conditional on \( \tilde{B} \sim \mathcal{N}(0, \sigma_B^2) \), we have \( \tilde{b}_i \sim \mathcal{N}(\tilde{B}, \sigma_b^2) \). \( \sigma_B \) should be large relatively to \( \sigma_B \).

We will use the \( N = 1 \) case to show that the conclusion around the robustness of FOM still holds in the presence of extreme outliers. Under the extended model, the optimal measure takes a very similar form as in the baseline model without outliers:

\[ f^*(U) = E[S|U] = \frac{\text{numerator}}{\text{denominator}} \]

where

\[
\text{numerator} = \omega_0 (1 - \theta) \frac{\sigma_A^2}{\sigma_F^2 + \sigma_A^2} \phi(U; \sigma_F^2 + \sigma_A^2) \cdot U \\
+ \omega_1 (1 - \theta) \frac{\sigma_A^2}{\sigma_F^2 + \sigma_A^2 + \sigma_b^2 + \sigma_B^2} \phi(U + \mu_B; \sigma_F^2 + \sigma_A^2 + \sigma_b^2 + \sigma_B^2) \cdot (U + \mu_B) \\
+ \theta \frac{\sigma_A^2}{\sigma_F^2 + \sigma_A^2 + \sigma_b^2 + \sigma_B^2} \phi(U; \sigma_F^2 + \sigma_A^2 + \sigma_b^2 + \sigma_B^2) \cdot U \\
\text{denominator} = \omega_0 (1 - \theta) \phi(U; \sigma_F^2 + \sigma_A^2) + \omega_1 (1 - \theta) \phi(U + \mu_B; \sigma_F^2 + \sigma_A^2 + \sigma_b^2 + \sigma_B^2) \\
+ \theta \phi(U; \sigma_F^2 + \sigma_A^2 + \sigma_b^2 + \sigma_B^2)
\]

Since \( \sigma_B >> \sigma_B \), it can be shown that

\[
\lim_{\sigma_B^2 \to \infty} E[f^*(U)^2] = \omega_0 (1 - \theta) \sigma_A^2 \frac{\sigma_A^2}{\sigma_F^2 + \sigma_A^2}. \tag{14}
\]

and

\[
\lim_{\sigma_B^2 \to \infty} \text{Cor}[f^*(U), S] = \frac{\omega_0^{1/2} (1 - \theta)^{1/2}}{\sqrt{1 + r_F^2}}. \tag{15}
\]
Therefore, the lower bound for the relative efficiency between FOM and \(f^*(U)\) is:

\[
\lim_{\sigma_B^2 \to \infty} \frac{\text{Cor}[\text{FOM}, S]}{\text{Cor}[f^*(U), S]} \geq \sqrt{2/\pi} \cdot \omega_0^{1/2} \cdot (1 - \theta)^{1/2}. \tag{16}
\]

This lower bound depends on \(\theta\) (the probability of drawing outliers) and \(\omega_0\) (the probability of drawing an unbiased forecast in the baseline model). For reasonably small \(\theta\), the decrease in efficiency is minimal - for example, when \(\theta = 0.05\), the additional multiplier is \((1 - \theta)^{1/2} = 0.975\), so the reduction in the minimum relative efficiency is only 2.5%.

The numerical value of this lower bound for \(\theta = 0.05\) are shown in Panel C of Table 1, which are similar to those given in Panel A of Table 1 for the baseline case. FOM still retains a decent efficiency relative to the optimal but infeasible measure - the relative efficiency is 55.0% when the probability of unbiased forecasts is one-half, and goes up to 77.8% as the probability of unbiased forecast increases.

### 2.6 Extensions

We show that our results hold even if we had a model where there was heterogeneity in bias, with \(\omega_1\) representing the proportion of analysts with "big" bias and \(\omega_0\) the proportion of analysts with "little" bias but not unbiased per se as in our baseline model. If \(\tilde{\mu}_B\), \(\tilde{\sigma}_B\) and \(\tilde{\sigma}_b\) are the terms related to the "little" bias,

\[
\text{Cov}[\text{FOM}, S] = 2\sigma_F \left[2 \Phi\left(\frac{X - \tilde{\mu}_B/\sigma_F}{\sqrt{1 + (\tilde{\sigma}_b^2 + \tilde{\sigma}_B^2)/\sigma_F^2}}\right) + \omega_1 \Phi\left(\frac{X - \mu_B/\sigma_F}{\sqrt{1 + (\sigma_b^2 + \sigma_B^2)/\sigma_F^2}}\right)\right],
\]

where \(\Phi(\cdot)\) is the cdf of standard normal, and \(X \sim \mathcal{N}(0, \frac{\sigma_A^2}{\sigma_F^2})\).

Even if \(\sigma_B \to \infty\), we still have

\[
\text{Cov}[\text{FOM}, S] \geq 2\omega_0 \sigma_F \left[2 \Phi\left(\frac{X - \tilde{\mu}_B/\sigma_F}{\sqrt{1 + (\tilde{\sigma}_b^2 + \tilde{\sigma}_B^2)/\sigma_F^2}}\right)\right].
\]
The lower bound is a non-negative number because both $X$ and $\Phi\left( \frac{X - \hat{\mu}_B}{\sigma_F} \right)$ are monotonically increasing in $X$, but the value now depends on additional terms related to the “little” bias, and will go to zero if the “little” bias becomes huge (in which case all forecasts are contaminated by huge bias).

3 Data

The data on analysts’ earnings forecasts are taken from the Institutional Brokers Estimate System (I/B/E/S). We gather data both for US stocks as well as international stocks following IBES classifications.\(^3\) We conduct our analysis using the Unadjusted Detailed files. We focus on forecasts of the fiscal year-end earnings (FY1) from 1983 to 2015. 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 is 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

\(^3\)Countries in the international sample include Canada, United Kingdom and a number of markets in Europe and Asia. Among European countries, they include Austria, Belgium, Cyprus, Czech Republic, Germany, Denmark, Spain, Finland, Faroe Islands, France, Greece, Croatia, Hungary, Ireland, Italy, Liechtenstein, Luxembourg, Netherlands, Norway, Poland, Portugal, Romania, Serbia, Russian Federation, Sweden, Slovenia, Turkey, and Ukraine. Among Asian countries, they include United Arab Emirates, China, Hong Kong, Indonesia, Israel, India, Jordan, Japan, Kuwait, Lebanon, Luxembourg, Malaysia, Oman, Philippines, Pakistan, Qatar, Saudi Arabia, Singapore, Thailand, and Taiwan.
period end date. Otherwise, we drop the observations.\footnote{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 announcements. Relevant summary statistics based on qualified quarterly forecasts are then calculated.} We remove penny stocks with a price of less 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 require an earnings announcement to have at least one analyst forecast.

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 six groups. Analyst forecast dispersion ($DISP$) is defined as the standard deviation of analyst forecasts scaled by $price(-20)$. Following the literature, we calculate earnings persistence using an AR(1) model with annual earnings. Each firm only has one persistence measure in our sample since we require all the years to reliably calculate the persistence of annual earnings.

We use two indicator functions, $I_{Actual<All}$ and $I_{Actual>Alt}$, 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, $I_{Actual>Alt}$ equals 1 if all analysts are too pessimistic and the maximum forecast is lower than the actual earnings. By construction, $FOM$ equals 1 if $I_{Actual>Alt}$ is 1 and -1 if $I_{Actual<All}$ is 1. In the instances of $N = 1$ forecast, $FOM$ is either 1, 0 or -1.
Using CRSP, we calculate cumulative abnormal returns (CAR) as follows:

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

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 2 provides the summary statistics of the variables. In Panel A for US stock, notice that the CE (using the mean consensus) has a mean of -0.0034, and a standard deviation of 0.08. The CE using the median consensus has similar magnitudes. \( \text{Rank}(CE) \) using mean consensus has a mean of 5.5 and a standard deviation of 2.87. \( \text{Rank}(CE) \) using median consensus has a mean of 3.5 and a standard deviation of 1.7.\(^5\) Our FOM has a mean of 0.11 and a standard deviation of 0.8. \( I_{\text{Actual}<\text{All}} \) has a mean of 0.23 and \( I_{\text{Actual}>\text{All}} \) has a mean of 0.3. Moreover, notice that CE based on either median or mean consensus have a correlation of 0.96 with each other. In Panel B for international stocks, we report the analogous summary statistics. They are overall very similar to the US. There is a lower mean value for FOM and lower mean values for misses on the same side.

In Figure 3, we plot the distribution of FOM across the entire sample. On the x-axis are the bins for various values of FOM. Around 23% of our sample is in the -1 bin (which denotes \( I_{\text{Actual}<\text{All}} \)) and 30% in the 1 bin (which denotes \( I_{\text{Actual}>\text{All}} \)). For the bins in the middle, we

\(^5\)The standard deviation of the \( \text{Rank}(CE) \) 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.
have a bin width of 0.25. The bins with positive \textit{FOM}'s each have around 5 to 10\% of the observations. The bins with negative \textit{FOM}'s have a somewhat similar representation. In Figure 4, we show the \textit{FOM} distribution conditional on the number of analysts \textit{N}. In Figure 5, 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 2, we report the correlation matrix for our variables of interest in which \textit{CE} is based on the median consensus. The correlation of \textit{CE} with \textit{Rank(CE)} is 0.22 and the correlation of \textit{FOM} and \textit{CE} is 0.18. As we discussed above in the extension of our model with outliers, a \textit{CE} that has fat-tails can drive down these correlations. The correlation of \textit{FOM} with \textit{Rank(CE)} is higher at 0.83 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. In any event, \textit{FOM} will have different information about market surprises than \textit{CE} and \textit{Rank(CE)}. Results in Panel C using the international data are similar. One thing to note is that the outliers make \textit{CE} not as effective a measure of market surprises as \textit{Rank(CE)}. But \textit{FOM} does better than both as we show below. We winsorize these variables of interest at the 1\% and 99\% in our regression analysis below.

It is also useful to do a decomposition of \textit{FOM} on firm and time characteristics. In a table which we omit for brevity, we report the \textit{R}^2 of \textit{FOM} regressions on firm and year dummies. We consider three models: (1) \textit{FOM} on firm dummies only, (2) \textit{FOM} on year dummies only, and (3) \textit{FOM} on firm and year dummies. The \textit{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 \textit{FOM}. \textit{FOM} is mostly driven by idiosyncratic events, consistent with the premise of our model.
4 Empirical Findings

4.1 FOM and CAR

In Table 3, 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>Alt} 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 0.002. In column (2) FOM attracts a coefficient of 0.0183 with a t-statistic of 23.96 and an $R^2$ of 0.045. A one standard deviation increase in FOM increases CAR by 1.5%, which is a sizeable 3-day move in stock returns. In column (3), we find that the coefficient in front of I_{Actual<All} is as expected negative with a coefficient of -0.0191 and a t-statistic of -12. For I_{Actual>Alt}, it is positive at 0.0187 with a t-statistic of 13.01. 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 1.9% decrease in stock prices over 3 days when all analysts miss too high and a 1.8% 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 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 0.0044 with a t-statistic of 19.87. The $R^2$ is 0.033. 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 0.045, which is higher than that of \( \text{Rank}(CE) \). 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 of \( \text{FOM} \) is largely unchanged. The coefficient for \( \text{Rank}(CE) \) is close to zero and is no longer significant. Moreover, the \( R^2 \) remains the same as when \( \text{FOM} \) is by itself in the regression.

In column (5), we find that adding in a horserace of \( \text{Rank}(CE) \) with the everyone-misses-on-the-same-side indicators, \( \text{Rank}(CE) \) retains more explanatory power. This indicates that \( \text{FOM} \) does better than \( \text{Rank}(CE) \) not only when everyone misses on the same side. There is also information not captured in \( \text{Rank}(CE) \) in intermediate values of \( \text{FOM} \).

In columns (6)-(10), we consider \( \text{Rank}(CE) \) but using the median forecast as the consensus forecast. The coefficient in front of \( \text{Rank}(CE) \) is 0.008 with a t-statistic of 21.25 and an \( R^2 \) of 0.038, which is better than \( \text{Rank}(CE) \) using mean forecasts. Nonetheless, the improvement in \( R^2 \) going from \( \text{Rank}(CE) \) to \( \text{FOM} \) is around 20% in this instance. When we combine \( \text{Rank}(CE) \) with \( \text{FOM} \), we see again that the coefficient in front of \( \text{Rank}(CE) \) falls to 0.0022 with a t-statistic of 3.4, while the coefficient in front of \( \text{FOM} \) is 0.0144 with a t-statistic of 10.52.

One way to compare the economic magnitudes is to ask how a one standard deviation increase in \( \text{Rank}(CE) \) or \( \text{FOM} \) increases the \( \text{CAR} \) from the point estimates in column (9). For \( \text{Rank}(CE) \), its standard deviation is 1.7, while for \( \text{FOM} \), it is 0.8. The implied \( \text{CAR} \) effect of \( \text{Rank}(CE) \) is just 0.0034 compared to the implied \( \text{CAR} \) effect for \( \text{FOM} \), which is 0.012. The \( \text{FOM} \) effect is 3 to 4 times as large as the \( \text{Rank}(CE) \) effect using the median forecast for the consensus. It is not surprising that the \( R^2 \) does not change much from the univariate \( \text{FOM} \) specification when we add \( \text{Rank}(CE) \). In column (10), we combine the \( \text{Rank}(CE) \) 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 \( \text{CE} \) improves performance, \( \text{FOM} \) is still the best univariate measure by a substantial margin. This will become even more apparent when we consider \( \text{POSTCAR} \).
4.2 FOM and POSTCAR

In Table 4, 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 .025 but is not statistically significant. In column (2), the coefficient in front of FOM is 0.0147 with a t-statistic of 5.12. In column (3), we see that the coefficients in front of the indicators when everyone misses on the same side are -0.0184 with a t-statistic of -3.24 and 0.0127 with a t-statistic of 2.34. 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 0.0028 and a t-statistic of 3.21. 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 0.0199 with a t-statistic of 4.18. So here moving from an FOM of -1 to 1 would lead to an increase in the POSTCAR of nearly 4% per six months or nearly 8% 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 $\text{Rank(CE)}$ instead of the mean forecast and find very similar results.

### 4.3 International Firms

In Table 5, we conduct our analysis from Tables 3 and 4 but for an international sample of stocks. Our purpose for this analysis is two-fold. First, the international sample gives us an out-of-sample test of our robustness theory. Second, our theory also predicts that the incremental improvement of $FOM$ over $CE$ should be lower in markets where bias is potentially unimportant, i.e. where $\sigma_B$ is not large. The literature on earnings-guidance games points out that the issuance incentives which lead to these games and bias are much less prevalent in international markets.

Panel A reports the results for $\text{CAR}$ and Panel B for $\text{POSTCAR}$ as the dependent variables of interest, respectively. We break down the international sample into three groups: UK and Canada (columns (1)-(3)), Asia (columns (4)-(6)) and Europe (columns (7)-(9)). It is easy to see that for overall $FOM$ still does better than $\text{Rank(CE)}$ but the difference is much smaller than in the US sample for the $\text{CAR}$. This can be most easily seen from the multiple regressions where we have both $FOM$ and $CE$. Indeed, for Europe, $\text{Rank(CE)}$ performs better than $FOM$ in this horserace. For $\text{POSTCAR}$, $FOM$ is still more robust but again as we will formally show next, the differences are less pronounced, consistent with our theory.

### 4.4 Model Selection

In Table 6, we calculate BIC difference statistics ($\Delta \text{BIC}$) to evaluate various models of earnings surprises in explaining $\text{CAR}$ and $\text{POSTCAR}$. We also calculate likelihood ratio tests to compare nested model comparisons in which we evaluate a multi-variate model with both $FOM$ and $CE$ against the univariate counterparts.

The first three columns are for the $\text{CAR}$ regression and the second three columns for the $\text{POSTCAR}$ regression. Panel A reports the results for the US sample. Let’s first consider
the BIC difference test between \( \text{Rank(CE)} \) and \( \text{FOM} \). The BIC of a model is a negative number. The better the fit of the model, the more negative is the BIC. The difference test is simply the difference in the BIC of the two non-nested models. A positive number means that \( \text{Rank(CE)} \) is a worse model to explain \( \text{CAR} \) than \( \text{FOM} \). BIC has natural cut-offs to judge how severe is the lack of fit comparing these two models. A BIC difference bigger than 10 is considered very strong evidence against \( \text{Rank(CE)} \). The BIC difference is 443. For \( \text{POSTCAR} \), it is 49.75. So we can very strongly choose \( \text{FOM} \) as the superior model for predicting stock returns when compared to \( \text{Rank(CE)} \).

We next turn to comparisons of \( \text{Rank(CE)} \) versus a model with both \( \text{Rank(CE)} \) and \( \text{FOM} \). Here we can easily conclude that we would never have a univariate model with just \( \text{Rank(CE)} \). This is true both in the BIC difference test and also the likelihood ratio test with associated p-values. The likelihood ratio test is an additional test we can run since \( \text{Rank(CE)} \) is nested in the multiple regression model. In the third row of Panel A, we find that at the same time \( \text{FOM} \) is inferior to a multiple regression model with itself and \( \text{Rank(CE)} \) as explanatory variables. However, for \( \text{POSTCAR} \), the BIC difference is -7.87, which means that \( \text{FOM} \) is better than the more elaborate model. Panels B-D report the results for the international countries. Here, we can see that the BIC difference numbers are all much smaller than for the US, consistent with the point that \( \text{FOM} \) is not far superior to \( \text{Rank(CE)} \).

### 4.5 Robustness Exercises, Walk-down Effects and Earnings Persistence

Our Table 7 is motivated by the discussion in our theory on whether \( \text{FOM} \) does better than \( \text{Rank(CE)} \) for long-horizon versus short-horizon forecasts. As we derived in the model, whether or not this is the case depends on a few parameters. One is of course the extent of the bias. We know that the mean bias is more positive for long-horizon forecasts from the walk-down of earnings forecasts literature (i.e. \( \mu_B \) more positive). But at the same time, it
is well-known that long-horizon forecasts are noisier than short-horizon forecasts (i.e. $\sigma_F$ is higher). And there can be potentially large biases for near-term forecasts (i.e. $\sigma_B$ might be large). As such, there is not an unambiguous prediction on how the FOM correction should perform when we consider forecasts that are issued further back in time.

In Table 7, Panel A (B) shows the results where we require the forecasts to be issued at least three (six) months before the announcement date. We find that FOM is still better than CE for these forecasts issued further out. But the incremental improvement performance is less stark then for the the most recent forecasts featured in our baseline cases of Table 3. To the extent one is interested in explaining stock returns, the most recent forecasts would be the most relevant ones to consider and FOM here is a much better model as we have established.

In Table 8, we break down our regressions by subgroups of stocks based on in Panel A, earnings persistence of a stock, and in Panel B, the dispersion of analyst forecasts. The motivation for the earnings persistence is that earnings have to be persistent enough for market surprises to matter for stock returns. Annual earnings persistence is moderately persistent with an AR(1) coefficient of 0.67 on average. We break our sample by the median of earnings persistence. The improvement in $R^2$ from using FOM rather than CE is higher for the higher earnings-persistence sample. A higher persistence of earnings means that surprises would then matter more for stock returns. As such, a better market surprise measure would matter more for this sub-sample of stocks. In Panel B, we split the sample by the dispersion of analyst forecasts ($DISP$) and find similar results across the two sub-samples. In total, these results point to the importance of FOM for US stocks.

In our results up to this point, 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.
4.6 \textit{FOM and Revision of Consensus}

In Table 9, we compare the relative performance of \textit{Rank(CE)} and \textit{FOM} in explaining revisions of analysts expectations (between two adjacent fiscal years) in the same direction as market returns. In other words, if both \textit{Rank(CE)} and \textit{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 \textit{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 \textit{Rank(CE)} and \textit{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 \textit{FOM} is stronger than \textit{Rank(CE)} in both Panels A and B. But the difference is far smaller when it comes to predicting stock returns. In sum, it is comforting that \textit{FOM} and the everyone-misses-on-the-same-side indicators are picking up revisions of the consensus.

4.7 \textit{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

and can be obtained from the authors.
breaking down square consensus error into bias and variance:

\[
SCE = \left(\sum_{i=1}^{N} \omega_i b_i\right)^2 + \sum_{i=1}^{N} \omega_i^2 \sigma_i^2 = \mathbf{wbb'} + \mathbf{w'Dw} \tag{18}
\]

where \(SCE\) is the square consensus error, \(\omega_i\) is the optimal weight of individual forecast \(i\), \(b_i\) is the bias of individual analyst \(i\), and \(\sigma_i^2\) is the square debiased forecast error of analyst \(i\), and \(D = \text{diag}\{\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2\}\). 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:

\[
L(\lambda) = SCE - \lambda \mathbf{w'1}. \tag{19}
\]

By taking the first order derivative with respect to \(w\), we get

\[
2(\mathbf{bb'} + \mathbf{D})\mathbf{w} - \lambda \mathbf{1} = 0, \tag{20}
\]

which we can then use to solve for \(w\) as

\[
\mathbf{w} = \frac{\lambda}{2}(\mathbf{bb'} + \mathbf{D})^{-1}\mathbf{1}. \tag{21}
\]

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

\[
(\mathbf{bb'} + \mathbf{D})^{-1} = \mathbf{D}^{-1} - \frac{\mathbf{D}^{-1}\mathbf{bb'D}^{-1}}{1 + \mathbf{b'D}^{-1}\mathbf{b}}, \tag{22}
\]

and \(w\) then becomes

\[
\mathbf{w} = \frac{\lambda}{2}\mathbf{D}^{-1}\mathbf{1} - \frac{\lambda}{2 + 2\mathbf{b'D}^{-1}\mathbf{b}}\mathbf{D}^{-1}\mathbf{bb'D}^{-1}\mathbf{1}. \tag{23}
\]
Therefore, for each $\omega_i$ we have

$$\omega_i = \frac{\lambda}{2} \left( 1 - \frac{b_i \sum_{k=1}^{N} b_k \sigma_k^{-2}}{1 + \sum_{k=1}^{N} b_k^2 \sigma_k^{-2}} \right) \sigma_i^{-2}, i = 1, 2, ..., N. \tag{24}$$

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,

$$b_i = \frac{1}{T} \sum_{t=-T}^{-1} (F_{i,t} - Actual_{i,t}). \tag{25}$$

We require each analyst to have at least two forecasts in the past five years. If an analyst does not have a forecast history long enough (less than two observations over the past five 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 $i$ ($\sigma_i^2$) is the individual square forecast error:

$$\sigma_i^2 = (F_i - Actual_i - b_i)^2. \tag{26}$$

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 10 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 3 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.

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 better job of explaining stock returns than the consensus error because individual forecasts can be biased. The ideal but difficult to estimate parameter-dependent measure of market surprise is a nonlinear filter of individual errors that adjusts for individual forecast bias. We show that $CE$ is a poor parameter-free approximation for this ideal measure. The fraction of misses on the same side ($FOM$), by discarding the magnitude of misses, offers a far-better approximation. We demonstrate the empirical validity of our theoretical analysis using earnings forecasts data from a global sample of markets. 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


Proof of Lemma 1:

First we have

\[ \text{Cov}(CE, S) = \text{Cov}(S + \frac{1}{N} \sum_{i=1}^{N} Y_i, S) = \sigma_A^2 \]

To compute \( \text{Var}(CE) \), we condition on \( b_i \)'s, and then take expectations.

\[
\begin{align*}
\text{Var}[CE] &= \text{Var} \left[ S + \frac{1}{N} \sum_{i=1}^{N} Y_i \right] \\
&= \text{Var}[S] + \text{Var} \left[ \frac{1}{N} \sum_{i=1}^{N} (-b_i - \epsilon_i) \right] \\
&= \text{Var}[S] + \mathbb{E} \left[ \text{Var} \left[ \frac{1}{N} \sum_{i=1}^{N} (b_i + \epsilon_i) \Big| b_1, \ldots, b_N \right] \right] \\
& \quad + \text{Var} \left[ \mathbb{E} \left[ \frac{1}{N} \sum_{i=1}^{N} (b_i + \epsilon_i) \Big| b_1, \ldots, b_N \right] \right] \\
&= \sigma_A^2 + \frac{1}{N} \sigma_b^2 + \text{Var} \left[ \frac{1}{N} \sum_{i=1}^{N} b_i \right] \quad (27)
\end{align*}
\]

The \( b_i \)'s have probability \( \omega_0 \) of being 0, and probability \( \omega_1 \) of being drawn from \( \mathcal{N}(B, \sigma_b^2) \).

They are correlated because of the same \( B \).

\[
\begin{align*}
\text{Var} \left[ \frac{1}{N} \sum_{i=1}^{N} b_i \right] &= \mathbb{E} \left[ \text{Var} \left[ \frac{1}{N} \sum_{i=1}^{N} b_i \Big| B \right] \right] + \text{Var} \left[ \mathbb{E} \left[ \frac{1}{N} \sum_{i=1}^{N} b_i \Big| B \right] \right] \quad (28)
\end{align*}
\]

Conditional on \( B \), \( b_i \)'s are effectively i.i.d., so

\[
\begin{align*}
\mathbb{E}[b_i|B] &= \omega_0 \cdot 0 + \omega_1 \cdot B \\
&= \omega_1 B \\
\mathbb{E}[b_i|B] &= \omega_0 \cdot ((0 - \omega_1 B)^2) + \omega_1 \cdot ((B - \omega_1 B)^2 + \sigma_b^2) \\
&= \omega_1 \sigma_b^2 + \omega_0 \omega_1 B^2 \quad (30)
\end{align*}
\]
Therefore,

\[
E \left[ \text{Var} \left( \frac{1}{N} \sum_{i=1}^{N} b_i | B \right) \right] = E \left[ \frac{1}{N^2} \sum_{i=1}^{N} \text{Var}[b_i | B] \right] \\
= \frac{1}{N} E[\omega_1 \sigma_b^2 + \omega_0 \omega_1 B^2] \\
= \frac{1}{N} (\omega_1 \sigma_b^2 + \omega_0 \omega_1 \sigma_B^2 + \omega_0 \omega_1 \mu_B^2) \tag{31}
\]

\[
\text{Var} \left[ E \left( \frac{1}{N} \sum_{i=1}^{N} b_i | B \right) \right] = \text{Var} \left[ \frac{1}{N} \sum_{i=1}^{N} E[b_i | B] \right] \\
= \text{Var}[\omega_1 B] \\
= \omega_1^2 \sigma_B^2 \tag{32}
\]

Substitute Eq (31) and (32) into Eq (28) and then Eq (27),

\[
\text{Var}[CE] = \sigma_A^2 + \frac{1}{N} \left( \sigma_F^2 + \omega_1 \sigma_b^2 + \omega_0 \omega_1 \sigma_B^2 + \omega_0 \omega_1 \mu_B^2 \right) + \omega_1^2 \sigma_B^2 \tag{33}
\]

Taken together,

\[
\text{Cor}[CE, S] = \frac{\text{Cov}[CE, S]}{\sqrt{\text{Var}[CE] \cdot \text{Var}[S]}} \\
= \frac{\sigma_A^2}{\sigma_A \sqrt{\frac{1}{N} (\sigma_F^2 + \omega_1 \sigma_b^2 + \omega_0 \omega_1 \sigma_B^2 + \omega_0 \omega_1 \mu_B^2) + \sigma_A^2 + \omega_1^2 \sigma_B^2}} \tag{34}
\]

QED

**Proof of Lemma 2.**

We can rewrite FOM as:

\[
FOM = \frac{\# \{ \epsilon_i + b_i < S \} - \# \{ \epsilon_i + b_i > S \}}{N} \\
= \frac{1}{N} \sum_{i=1}^{N} M_i, \tag{35}
\]
where
\begin{equation}
M_i = \begin{cases} 
1 & \text{if } \epsilon_i + b_i < S \\
-1 & \text{if } \epsilon_i + b_i > S
\end{cases}
\end{equation}

and
\begin{equation}
b_i = \begin{cases} 
0 & \text{with probability } \omega_0 \\
\sim \mathcal{N}(B, \sigma_b^2) & \text{with probability } \omega_1 = 1 - \omega_0 \text{ where } B \sim \mathcal{N}(\mu_B, \sigma_B^2)
\end{cases}
\end{equation}

If we work out the math,
\begin{align*}
\text{Cov}[FOM, S] &= \text{E}[S \left( \frac{1}{N} \sum_{i=1}^{N} M_i \right)] - \text{E}[S] \cdot \text{E}[FOM] \\
&= \frac{1}{N} \sum_{i=1}^{N} \text{E}[S \cdot (I_{\epsilon_i+b_i<S} - I_{\epsilon_i+b_i>S})] \\
&= \frac{1}{N} \sum_{i=1}^{N} \text{E}[E[S \cdot (P(\epsilon_i + b_i < S) - P(\epsilon_i + b_i > S)) | S]] \\
&= \frac{1}{N} \sum_{i=1}^{N} \text{E}[E[S \cdot (P(\epsilon_i + b_i < S) - (1 - P(\epsilon_i + b_i < S))) | S]] \\
&= 2E \left[ S \cdot \left( \omega_0 \Phi(\frac{S}{\sigma_F}) + \omega_1 \Phi(\frac{S - \mu_B}{\sqrt{\sigma_F^2 + \sigma_b^2 + \sigma_B^2}}) \right) \right] \\
&= 2\sigma_F E \left[ X \cdot (\omega_0 \Phi(X) + \omega_1 \Phi(\tilde{X})) \right],
\end{align*}

where \( \Phi(\cdot) \) is the cdf of standard normal, \( X \sim \mathcal{N}(0, \frac{\sigma^2}{\sigma_F^2}) \), and \( \tilde{X} = \frac{X - \mu_B/\sigma_F}{\sqrt{1+(\sigma_b^2+\sigma_B^2)/\sigma_F^2}} \).

Note that the second term in Eq (37) is non-negative because \( X \) and \( \Phi(\tilde{X}) \) are both monotonically increasing in \( X \) and must have positive covariance. Therefore,
\begin{align*}
\text{Cov}[FOM, S] &\geq 2\omega_0 \sigma_F E[X \cdot \Phi(X)] \\
&= \omega_0 \cdot \text{Cov}[FOM, S | \omega_1 = 0].
\end{align*}

This means the covariance is at least a fraction of what we would get in the absence of bias. The more unbiased forecasts (the larger \( \omega_0 \)), the more positive relationship preserved.
Consequently, the correlation between FOM and $S$ is bounded from below

$$
\text{Cor}[\text{FOM}, S] = \frac{\text{Cov}[\text{FOM}, S]}{\sqrt{\text{Var}[\text{FOM}] \cdot \text{Var}[S]}} \\
\geq \frac{2\omega_0 \sigma_F E[X \cdot \Phi(X)]}{\sigma_A \cdot \sqrt{\text{Var}[\text{FOM}]}} \\
\geq 2\omega_0 \sigma_F E[X \cdot \Phi(X)]/\sigma_A
$$

(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 $1$. We can analytically evaluate Eq (39) by denoting $\mathcal{N}(0,1)$ as $Z$ and taking derivative w.r.t. $r_F = \sigma_F/\sigma_A$

$$
\frac{d}{dr_F} E[r_F X \Phi(X)] = \frac{d}{dr_F} E[Z \Phi(Z/r_F)] \\
= - \frac{1}{r_F^2} E[Z^2 \phi(Z/r_F)] \\
= - \frac{1}{\sqrt{2\pi}} \cdot \frac{r_F}{(1 + r_F^2)^{3/2}}.
$$

Integrating the derivative back, we have

$$
E[r_F X \Phi(X)] = \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{1 + r_F^2}}.
$$

Therefore, the lower bound of Cor$[\text{FOM}, S]$ is given by

$$
\text{Cor}[\text{FOM}, S] \geq \frac{2\omega_0}{\sqrt{2\pi} (1 + \sigma_F^2/\sigma_A^2)}
$$

(40)

QED

**Proof of Proposition 1:** We first calculate the correlation between the ideal measure of earning surprise with earning surprise itself. Note that by the double expectation formula,

$$
\text{Cov}[f^*(U), S] = E[f^*(U)S] = E[f^*(U)^2].
$$
Since $E[f^*(U)] = E[E[S|U]] = 0$, we conclude that $\text{Var}[f^*(U)] = E[f^*(U)]^2$, and

$$\text{Cor}[f^*(U), S] = \sqrt{E[f^*(U)]^2}/\sigma_A.$$

Let $f_U(u) = \omega_0 \phi(u; \sigma_F^2 + \sigma_A^2) + \omega_1 \phi(u + \mu_B; \sigma_F^2 + \sigma_A^2 + \sigma_B^2 + \sigma_A^2)$ be the marginal density of $U$. It can be shown that

$$\lim_{\sigma_B^2 \to \infty} E[f^*(U)^2] = \omega_0 \sigma_A^2 \sigma_F^2 / \sigma_A^2.$$

Therefore, we conclude that

$$\lim_{\sigma_B^2 \to \infty} \text{Cor}[f^*(U), S] = \frac{\omega_1}{\sqrt{1 + \sigma_F^2/\sigma_A^2}}.$$

This combined with Lemma 2 implies leads to a lower bound for the relative efficiency between FOM and $f^*(U)$.

**Proof of Proposition 2:**

We first derive the following set of relationships:

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

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,

$$\text{Cor}(S, h^*) = \text{Cov}(S, h^*)/\sqrt{\text{Var}(S) \ast \text{Var}(h^*)}.$$
and
\[
\]
because \(E[S] = E[h^*] = 0\). The denominator
\[
\sqrt{\text{Var}(S) \ast \text{Var}(h^*)} = \sqrt{E[S^2]E[h^{*2}]}.
\]
Note that \(E[h^{*2}] = E[E[S|U, A]^2]\) by definition. Taking the ratio, we get the third term.

It can then be shown that
\[
E[S|U, A] = \frac{\sigma_1^2}{\sigma_F^2} \sum_{i \in A} U_i + \frac{\sigma_2^2}{\sigma_F^2} \sum_{i \notin A} U_i,
\]
where
\[
\sigma_1^{-2} = \frac{1}{\sigma_A^2} + \frac{|A|}{\sigma_F^2} + \frac{1}{\frac{\sigma_F^2 + \sigma_B^2}{N-|A|} + \sigma_B^2}, \quad \text{and} \quad \sigma_2^{-2} = \frac{1}{\frac{\sigma_F^2 + \sigma_B^2}{N-|A|} + \sigma_B^2}.
\]

To see this, note that the joint density of \(U, S\) and \(B\) (where \(B\) is the aggregate bias shock) conditional on \(A\) is (ignoring the normalization constant):
\[
\exp\left\{-\frac{S^2}{2\sigma_A^2}\right\} \cdot \exp\left\{-\frac{1}{\sigma_F^2} \sum_{i \in A} (U_i - S)^2\right\} \cdot \exp\left\{-\frac{B^2}{2\sigma_B^2}\right\} \cdot \exp\left\{-\frac{1}{\sigma_3^2} \sum_{i \notin A} (U_i - B - S)^2\right\},
\]
where \(\sigma_3^2 = \sigma_F^2 + \sigma_B^2\). The last two terms related to \(B\) can be written as
\[
\exp\left\{-\frac{1}{2} \left[ \frac{1}{\sigma_B^2} + \frac{N-|A|}{\sigma_3^2} \right] B^2 - \frac{2}{\sigma_3^2} \sum_{i \notin A} (U_i - S) + \frac{1}{\sigma_3^2} \sum_{i \notin A} (U_i - S)^2 \right\}.
\]
Let \(\sigma_4^{-2} = \frac{1}{\sigma_B^2} + \frac{N-|A|}{\sigma_3^2}\). Integrating the above expression with respect to \(B\), we obtain
\[
\exp\left\{-\frac{1}{\sigma_3^2} \sum_{i \notin A} (U_i - S)^2 + \frac{\sigma_4^2}{\sigma_3^2} \left[ \sum_{i \notin A} (U_i - S) \right]^2 \right\}.
\]
Therefore, the joint density of $U$ and $S$ conditional on $A$ is
\[
\exp \left\{ -\frac{S^2}{2\sigma_A^2} - \frac{1}{2\sigma_F^2} \sum_{i \in A} (U_i - S)^2 - \frac{1}{2\sigma_3^2} \sum_{i \notin A} (U_i - S)^2 + \frac{\sigma_2^2}{2\sigma_3^2} \left[ \sum_{i \notin A} (U_i - S) \right]^2 \right\}.
\]

Combining the terms related to $S$, we have
\[
\exp \left\{ -\frac{1}{2} \left[ \frac{1}{\sigma_A^2} + \frac{|A|}{\sigma_F^2} + \frac{N - |A|}{\sigma_3^2} \right] - \frac{\sigma_2^2}{\sigma_3^2} (N - |A|)^2 \right\} \cdot \exp \left\{ \frac{1}{\sigma_F^2} \sum_{i \in A} (U_i - S) + \frac{1}{\sigma_3^2} \sum_{i \notin A} (U_i - S) - \frac{\sigma_2^2}{\sigma_3^2} \sum_{i \notin A} U_i (N - |A|) \right\} S \right\}.
\]

Therefore, the conditional distribution of $S$ given $U$ and $A$ is
\[
\mathcal{N} \left( \frac{\sigma_1^2}{\sigma_F^2} \sum_{i \in A} U_i + \frac{\sigma_2^2}{\sigma_3^2} \sum_{i \notin A} U_i, \sigma_1^2 \right),
\]

where
\[
\sigma_1^2 = \frac{1}{\sigma_A^2} + \frac{|A|}{\sigma_F^2} + \frac{N - |A|}{\sigma_3^2} - \frac{\sigma_2^2}{\sigma_3^2} (N - |A|)^2 = \frac{1}{\sigma_A^2} + \frac{|A|}{\sigma_F^2} + \frac{1}{\sigma_3^2 + \sigma_F^2 + \sigma_B^2},
\]

and
\[
\sigma_2^2 = \frac{1}{\sigma_3^2} - \frac{\sigma_1^2 (N - |A|)}{\sigma_3^2} = \frac{1}{N - |A|} \frac{\sigma_1^2 + \sigma_3^2}{\sigma_3^2 + \sigma_B^2}.
\]

Since all of the $N$ forecasts depend on a common aggregated bias $B$, the variance $\sigma_B^2$ is not scaled by the number of forecasts. As $\sigma_B^2$ grows, we have
\[
\lim_{\sigma_B^2 \to \infty} \mathbb{E}[S|U, A] = \frac{1}{\sigma_F^2 + |A|} \sum_{i \in A} U_i.
\]
Denoting the above as $h^*$, we can compute its second moment using the law of total variance,

\[
E[h^*^2] = E[\text{Var}[h^* | S, A]] + \text{Var}[E[h^* | S, A]]
\]

\[
= E \left[ \frac{1}{(r_F^2 + |A|)^2} |A| \sigma_F^2 \right] + \text{Var} \left[ \frac{1}{r_F^2 + |A|} |A| S \right]
\]

\[
= E \left[ \frac{1}{1 + r_F^2/|A|} \right] \sigma_A^2.
\]

Recall that $\text{Cor}[\text{FOM}, S] \geq 2\omega_0 E[Z \Phi(Z/r_F)]$. Taken together we have shown Equation (11).

QED
Table 1: Lower bounds of relative efficiency of FOM in comparison with the ideal but infeasible measure.

Panel A: $N = 1$ forecast. Various values of $\omega_0$ (the fraction of unbiased forecasts).

<table>
<thead>
<tr>
<th>$\omega_0$</th>
<th>0.1</th>
<th>0.3</th>
<th>0.5</th>
<th>0.7</th>
<th>0.9</th>
<th>1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.252</td>
<td>0.437</td>
<td>0.564</td>
<td>0.668</td>
<td>0.757</td>
<td>0.798</td>
</tr>
</tbody>
</table>

Panel B. $N = 10$ forecasts. Various values of $\omega_0$ (the fraction of unbiased forecasts) and $r_F$ (the standard deviation of analysts’ forecast errors divided by standard deviation of earnings shocks).

<table>
<thead>
<tr>
<th>$\omega_0$</th>
<th>0.1</th>
<th>0.3</th>
<th>0.5</th>
<th>0.7</th>
<th>0.9</th>
<th>1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_F = 0$</td>
<td>0.099</td>
<td>0.243</td>
<td>0.399</td>
<td>0.559</td>
<td>0.718</td>
<td>0.798</td>
</tr>
<tr>
<td>$r_F = 0.5$</td>
<td>0.096</td>
<td>0.228</td>
<td>0.367</td>
<td>0.509</td>
<td>0.651</td>
<td>0.723</td>
</tr>
<tr>
<td>$r_F = 1$</td>
<td>0.092</td>
<td>0.202</td>
<td>0.312</td>
<td>0.423</td>
<td>0.536</td>
<td>0.592</td>
</tr>
<tr>
<td>$r_F = 2$</td>
<td>0.086</td>
<td>0.169</td>
<td>0.243</td>
<td>0.315</td>
<td>0.386</td>
<td>0.422</td>
</tr>
<tr>
<td>$r_F = \infty$</td>
<td>0.080</td>
<td>0.138</td>
<td>0.178</td>
<td>0.211</td>
<td>0.239</td>
<td>0.252</td>
</tr>
</tbody>
</table>

Panel C. $N = 1$ forecast. Various values of $\omega_0$ (the fraction of unbiased forecasts) and probability of outlier $\theta = 0.05$.

<table>
<thead>
<tr>
<th>$\omega_0$</th>
<th>0.1</th>
<th>0.3</th>
<th>0.5</th>
<th>0.7</th>
<th>0.9</th>
<th>1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.246</td>
<td>0.426</td>
<td>0.550</td>
<td>0.651</td>
<td>0.738</td>
<td>0.778</td>
</tr>
</tbody>
</table>
Table 2: 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). $FOM$ is defined as $\frac{K - M}{N}$, 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. Earnings Persistence is the coefficient from an AR(1) model for annual earnings. Panel A and Panel B report the summary statistics of the US and international samples, respectively. Panel C and Panel D report the correlation of variables for the US and international samples, respectively.

### Panel A: Summary statistics for US data

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>25th</th>
<th>Median</th>
<th>75th</th>
<th>Std Dev</th>
<th>Skewness</th>
<th>Kurtosis</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean consensus</td>
<td>1.2030</td>
<td>0.36</td>
<td>0.88</td>
<td>1.6689</td>
<td>2.2095</td>
<td>11.5785</td>
<td>327.6362</td>
<td></td>
</tr>
<tr>
<td>Median consensus</td>
<td>1.2077</td>
<td>0.36</td>
<td>0.88</td>
<td>1.67</td>
<td>2.1972</td>
<td>11.6030</td>
<td>328.9491</td>
<td></td>
</tr>
<tr>
<td>$CE(\text{based on mean forecast})$</td>
<td>-0.0034</td>
<td>-0.0018</td>
<td>0.0002</td>
<td>0.002</td>
<td>0.0807</td>
<td>-69.3071</td>
<td>10474.3359</td>
<td></td>
</tr>
<tr>
<td>$CE(\text{based on median forecast})$</td>
<td>-0.0034</td>
<td>-0.0016</td>
<td>0.0002</td>
<td>0.0019</td>
<td>0.0786</td>
<td>-74.1088</td>
<td>11581.6894</td>
<td></td>
</tr>
<tr>
<td>$Rank(CE)(\text{based on mean forecast})$</td>
<td>5.5001</td>
<td>3</td>
<td>5.5</td>
<td>8</td>
<td>2.8711</td>
<td>0.0001</td>
<td>1.7761</td>
<td></td>
</tr>
<tr>
<td>$Rank(CE)(\text{based on median forecast})$</td>
<td>3.5063</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>1.7050</td>
<td>-0.0082</td>
<td>1.7391</td>
<td></td>
</tr>
<tr>
<td>$DISP$</td>
<td>0.0069</td>
<td>0.0005</td>
<td>0.0014</td>
<td>0.0042</td>
<td>0.0688</td>
<td>87.7531</td>
<td>10357.6580</td>
<td></td>
</tr>
<tr>
<td>$FOM$</td>
<td>0.1145</td>
<td>-0.8</td>
<td>0.25</td>
<td>1</td>
<td>0.8071</td>
<td>-0.2476</td>
<td>1.4407</td>
<td></td>
</tr>
<tr>
<td>$I_{Actual&lt;All}$</td>
<td>0.2333</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.4229</td>
<td>1.2615</td>
<td>2.5914</td>
<td></td>
</tr>
<tr>
<td>$I_{Actual&gt;all}$</td>
<td>0.3050</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.4604</td>
<td>0.8470</td>
<td>1.7173</td>
<td></td>
</tr>
<tr>
<td>Earnings Persistence</td>
<td>0.6762</td>
<td>0.3816</td>
<td>0.5841</td>
<td>0.808</td>
<td>2.0772</td>
<td>66.3240</td>
<td>5417.5335</td>
<td></td>
</tr>
<tr>
<td>Correlation between $CE$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.9645</td>
<td></td>
</tr>
</tbody>
</table>

### Panel B: Summary statistics for International data

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>25th</th>
<th>Median</th>
<th>75th</th>
<th>Std Dev</th>
<th>Skewness</th>
<th>Kurtosis</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean consensus</td>
<td>1.4433</td>
<td>0.387</td>
<td>0.7168</td>
<td>1.5049</td>
<td>4.455</td>
<td>7.2102</td>
<td>143.1036</td>
<td></td>
</tr>
<tr>
<td>Median consensus</td>
<td>1.4343</td>
<td>0.371</td>
<td>0.7121</td>
<td>1.4954</td>
<td>4.8210</td>
<td>10.7571</td>
<td>267.5562</td>
<td></td>
</tr>
<tr>
<td>$CE(\text{based on mean forecast})$</td>
<td>-0.0111</td>
<td>-0.0041</td>
<td>0.0001</td>
<td>0.004</td>
<td>0.1761</td>
<td>-19.1933</td>
<td>645.2286</td>
<td></td>
</tr>
<tr>
<td>$CE(\text{based on median forecast})$</td>
<td>-0.0104</td>
<td>-0.0034</td>
<td>0.0002</td>
<td>0.0038</td>
<td>0.1994</td>
<td>-25.0162</td>
<td>1166.1518</td>
<td></td>
</tr>
<tr>
<td>$Rank(CE)(\text{based on mean forecast})$</td>
<td>5.5388</td>
<td>3</td>
<td>6</td>
<td>8</td>
<td>2.8551</td>
<td>-0.0164</td>
<td>1.7833</td>
<td></td>
</tr>
<tr>
<td>$Rank(CE)(\text{based on median forecast})$</td>
<td>3.5159</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>1.7008</td>
<td>-0.0107</td>
<td>1.7387</td>
<td></td>
</tr>
<tr>
<td>$DISP$</td>
<td>0.0220</td>
<td>0.002</td>
<td>0.0043</td>
<td>0.01</td>
<td>0.1630</td>
<td>23.3455</td>
<td>799.0245</td>
<td></td>
</tr>
<tr>
<td>$FOM$</td>
<td>0.0543</td>
<td>-0.714</td>
<td>0.0588</td>
<td>0.8667</td>
<td>0.7653</td>
<td>-0.1171</td>
<td>1.5023</td>
<td></td>
</tr>
<tr>
<td>$I_{Actual&lt;All}$</td>
<td>0.2072</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.4053</td>
<td>1.4447</td>
<td>3.0872</td>
<td></td>
</tr>
<tr>
<td>$I_{Actual&gt;all}$</td>
<td>0.2382</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.4260</td>
<td>1.2291</td>
<td>2.5107</td>
<td></td>
</tr>
<tr>
<td>Correlation between $CE$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.9290</td>
<td></td>
</tr>
</tbody>
</table>
### Panel C: Correlation matrix for US data (CE based on median consensus)

<table>
<thead>
<tr>
<th></th>
<th>CE</th>
<th>DISP</th>
<th>Rank(CE)</th>
<th>FOM</th>
<th>I_{Actual&lt;All}</th>
<th>I_{Actual&gt;All}</th>
<th>Earnings Persistence</th>
</tr>
</thead>
<tbody>
<tr>
<td>CE</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DISP</td>
<td>-0.107</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rank(CE)</td>
<td>0.2229</td>
<td>-0.0339</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FOM</td>
<td>0.1824</td>
<td>-0.0492</td>
<td>0.8284</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I_{Actual&lt;All}</td>
<td>-0.1871</td>
<td>0.0284</td>
<td>-0.5642</td>
<td>-0.7025</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I_{Actual&gt;All}</td>
<td>0.1098</td>
<td>-0.0406</td>
<td>0.5714</td>
<td>0.6769</td>
<td>-0.2846</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Earnings Persistence</td>
<td>-0.0265</td>
<td>0.0186</td>
<td>-0.0164</td>
<td>-0.0173</td>
<td>0.0191</td>
<td>-0.0032</td>
<td>1</td>
</tr>
</tbody>
</table>

### Panel D: Correlation matrix for International data (CE based on median consensus)

<table>
<thead>
<tr>
<th></th>
<th>CE</th>
<th>DISP</th>
<th>Rank(CE)</th>
<th>FOM</th>
<th>I_{Actual&lt;All}</th>
<th>I_{Actual&gt;All}</th>
</tr>
</thead>
<tbody>
<tr>
<td>CE</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DISP</td>
<td>-0.4916</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rank(CE)</td>
<td>0.1465</td>
<td>-0.073</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FOM</td>
<td>0.114</td>
<td>-0.0479</td>
<td>0.8581</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I_{Actual&lt;All}</td>
<td>-0.0884</td>
<td>-0.0001</td>
<td>-0.506</td>
<td>-0.6573</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>I_{Actual&gt;All}</td>
<td>0.0611</td>
<td>-0.0474</td>
<td>0.5219</td>
<td>0.6458</td>
<td>-0.2253</td>
<td>1</td>
</tr>
</tbody>
</table>
Table 3: 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 (\(\frac{K}{M} - \frac{M}{N}\)), where \(K\) 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. There are 58,477 observations. 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.

<table>
<thead>
<tr>
<th></th>
<th>CE is based on mean forecast</th>
<th></th>
<th>CE is based on median forecast</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>CE</td>
<td>0.0322*** (3.03)</td>
<td>0.00579 (1.06)</td>
<td>0.00733 (1.44)</td>
<td>0.0321*** (2.81)</td>
</tr>
<tr>
<td>FOM</td>
<td>0.0183*** (23.96)</td>
<td>0.0182*** (23.94)</td>
<td>0.0183*** (23.96)</td>
<td>0.0182*** (24)</td>
</tr>
<tr>
<td>I_{Actual&lt;All}</td>
<td>-0.0191*** (-12)</td>
<td>-0.0189*** (-12.04)</td>
<td>-0.0191*** (-12)</td>
<td>-0.019*** (-12.09)</td>
</tr>
<tr>
<td>I_{Actual&gt;All}</td>
<td>0.0187*** (13.01)</td>
<td>0.0186*** (12.94)</td>
<td>0.0187*** (13.01)</td>
<td>0.0186*** (12.95)</td>
</tr>
<tr>
<td>Year effect</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.002</td>
<td>0.045</td>
<td>0.040</td>
<td>0.040</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>Rank(CE) is based on mean forecast</th>
<th></th>
<th>Rank(CE) is based on median forecast</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>Rank(CE)</td>
<td>0.0044*** (19.87)</td>
<td>0.000585 (1.53)</td>
<td>0.00182*** (5.89)</td>
<td>0.00789*** (21.25)</td>
</tr>
<tr>
<td>FOM</td>
<td>0.0183*** (23.96)</td>
<td>0.0166*** (12.36)</td>
<td>0.0183*** (23.96)</td>
<td>0.0144*** (10.52)</td>
</tr>
<tr>
<td>I_{Actual&lt;All}</td>
<td>-0.0191*** (-12)</td>
<td>-0.0137*** (-7.16)</td>
<td>-0.0191*** (-12)</td>
<td>-0.0116*** (-6.21)</td>
</tr>
<tr>
<td>I_{Actual&gt;All}</td>
<td>0.0187*** (13.01)</td>
<td>0.0134*** (8.51)</td>
<td>0.0187*** (13.01)</td>
<td>0.0116*** (7.24)</td>
</tr>
<tr>
<td>Year effect</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.033</td>
<td>0.045</td>
<td>0.040</td>
<td>0.045</td>
</tr>
</tbody>
</table>
Table 4: 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 scores 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 (K - MN), 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. There are 56,919 observations. 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.

<table>
<thead>
<tr>
<th></th>
<th>CE based on mean forecast</th>
<th>CE based on median forecast</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>CE</td>
<td>0.0254</td>
<td>0.00427</td>
</tr>
<tr>
<td></td>
<td>(0.94)</td>
<td>(0.16)</td>
</tr>
<tr>
<td>FOM</td>
<td>0.0147***</td>
<td>0.0146***</td>
</tr>
<tr>
<td></td>
<td>(5.12)</td>
<td>(5.19)</td>
</tr>
<tr>
<td>I_{Actual&lt;All}</td>
<td>-0.0184***</td>
<td>-0.0183***</td>
</tr>
<tr>
<td></td>
<td>(-3.24)</td>
<td>(-3.32)</td>
</tr>
<tr>
<td>I_{Actual&gt; All}</td>
<td>0.0127**</td>
<td>0.0126**</td>
</tr>
<tr>
<td></td>
<td>(2.34)</td>
<td>(2.33)</td>
</tr>
<tr>
<td>Year effect</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.004</td>
<td>0.006</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>Rank(CE) based on mean forecast</th>
<th>Rank(CE) based on median forecast</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Rank(CE)</td>
<td>0.00278***</td>
<td>-0.00179</td>
</tr>
<tr>
<td></td>
<td>(3.21)</td>
<td>(-1.24)</td>
</tr>
<tr>
<td>FOM</td>
<td>0.0147***</td>
<td>0.0199***</td>
</tr>
<tr>
<td></td>
<td>(5.12)</td>
<td>(4.18)</td>
</tr>
<tr>
<td>I_{Actual&lt;All}</td>
<td>-0.0184***</td>
<td>-0.0193***</td>
</tr>
<tr>
<td></td>
<td>(-3.24)</td>
<td>(-2.89)</td>
</tr>
<tr>
<td>I_{Actual&gt; All}</td>
<td>0.0127**</td>
<td>0.0135**</td>
</tr>
<tr>
<td></td>
<td>(2.34)</td>
<td>(2.01)</td>
</tr>
<tr>
<td>Year effect</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.005</td>
<td>0.006</td>
</tr>
</tbody>
</table>

45
Table 5: Sensitivity of earnings announcement returns and post earnings announcement to FOM and Rank(CE) for international firms

This table presents the ordinary least squares estimates of the sensitivity of earnings announcement stock returns (CAR) or post earnings announcement stock returns (POSTCAR) to Rank(CE) and FOM for international firms. The dependent variable are CAR (cumulative abnormal return from trading day -1 to 1 around annual earnings announcement dates), and POSTCAR (cumulative abnormal return from trading day 2 to 126 post annual earnings announcement dates). The independent variables are Rank(CE) based on median consensus, from 1 to 6, and FOM($\frac{K}{N} - \frac{M}{N}$), where K(M) is the number of forecasts strictly smaller (greater) than the actual earnings. In Panel A, we report regression coefficients with dependent variable of CAR for international firms. Panel B reports regression coefficients with dependent variable of POSTCAR for international firms. UK and Canada results are reported in column (1) to (3), Asia results are reported in column (4) to (6), and Continental Europe results are reported in column (7) to (9). All standard errors are clustered by stocks. t-statistics are in parentheses.

### Panel A: Dependent variable of CAR

<table>
<thead>
<tr>
<th></th>
<th>UK and Canada</th>
<th>Asia</th>
<th>Continental Europe</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Rank(CE)</td>
<td>0.00263*** (3.01)</td>
<td>0.000921 (0.58)</td>
<td>0.00329*** (6.28)</td>
</tr>
<tr>
<td>FOM</td>
<td>0.00607*** (3.45)</td>
<td>0.00432 (1.38)</td>
<td>0.00685*** (5.95)</td>
</tr>
<tr>
<td>Year effect</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.019</td>
<td>0.02</td>
<td>0.02</td>
</tr>
</tbody>
</table>

### Panel B: Dependent variable of POSTCAR

<table>
<thead>
<tr>
<th></th>
<th>UK and Canada</th>
<th>Asia</th>
<th>Continental Europe</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Rank(CE)</td>
<td>0.00414 (1.45)</td>
<td>-0.0105* (-1.69)</td>
<td>0.00422* (1.66)</td>
</tr>
<tr>
<td>FOM</td>
<td>0.017*** (2.73)</td>
<td>0.0369*** (2.69)</td>
<td>0.00749 (1.37)</td>
</tr>
<tr>
<td>Year effect</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.079</td>
<td>0.083</td>
<td>0.084</td>
</tr>
</tbody>
</table>
This table compares models of earnings announcement stock returns (CAR) or post earnings announcement stock returns (POSTCAR) using Rank(CE) based on median consensus and FOM for US and international firms. The dependent variables are CAR (cumulative abnormal return from trading day -1 to 1 around annual earnings announcement dates), and POSTCAR (cumulative abnormal return from trading day 2 to 126 post annual earnings announcement dates). The independent variables are Rank(CE) based on median consensus, from 1 to 6, and FOM \((K/M)\), where \(K(M)\) is the number of forecasts strictly smaller (greater) than the actual earnings. There are three models to be compared pairwise, with independent variables of (1) Rank(CE) (2)FOM and (3)Rank(CE) and FOM. Year fixed effects are applied to all three models. We show the difference between the Bayesian information criteria (BIC) of each pair of models. \(BIC = \ln(n)k - 2\ln(\hat{L})\), where \(n\) is the number of observation, \(k\) is the number of free parameters to be estimated, and \(\hat{L}\) is the estimated maximum likelihood. The lower the BIC the better the model. We also conduct a likelihood ratio test for the two pairs of nested-models. We show Chi squares \((\chi^2 = -2 \times [\ln(\hat{L}_{\text{restricted}}) - \ln(\hat{L}_{\text{unrestricted}})])\), where \(\hat{L}\) are the estimated maximum likelihood for the restricted and unrestricted models, respectively) and p-values for the likelihood ratio test. All standard errors are clustered by stocks. t-statistics are in parentheses.

### Panel A: US

<table>
<thead>
<tr>
<th></th>
<th>CAR BIC difference</th>
<th>likelihood ratio test</th>
<th>POSTCAR BIC difference</th>
<th>likelihood ratio test</th>
<th>p value</th>
<th>Chi square p value</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rank(CE) vs FOM</td>
<td>443.64</td>
<td></td>
<td></td>
<td>49.75</td>
<td></td>
<td>49.75</td>
<td></td>
</tr>
<tr>
<td>Rank(CE) vs Rank(CE)+FOM</td>
<td>484.9</td>
<td>496</td>
<td>7.5×10^{-110}</td>
<td>41.88</td>
<td>52.8</td>
<td>3.6×10^{-13}</td>
<td>41.88</td>
</tr>
<tr>
<td>FOM vs Rank(CE)+FOM</td>
<td>41.26</td>
<td>52</td>
<td>4.9×10^{-13}</td>
<td>-7.87</td>
<td>3.1</td>
<td>3.6×10^{-13}</td>
<td>3.1</td>
</tr>
</tbody>
</table>

### Panel B: UK and Canada

<table>
<thead>
<tr>
<th></th>
<th>CAR BIC difference</th>
<th>likelihood ratio test</th>
<th>POSTCAR BIC difference</th>
<th>likelihood ratio test</th>
<th>p value</th>
<th>Chi square p value</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rank(CE) vs FOM</td>
<td>3.43</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>11.61</td>
<td></td>
</tr>
<tr>
<td>Rank(CE) vs Rank(CE)+FOM</td>
<td>-3.78</td>
<td>4.39</td>
<td>0.036</td>
<td>10.74</td>
<td>18.9</td>
<td>1.4×10^{-05}</td>
<td>18.9</td>
</tr>
<tr>
<td>FOM vs Rank(CE)+FOM</td>
<td>-7.21</td>
<td>0.96</td>
<td>0.327</td>
<td>-0.87</td>
<td>7.3</td>
<td>0.0069</td>
<td>7.3</td>
</tr>
</tbody>
</table>

### Panel C: Asia

<table>
<thead>
<tr>
<th></th>
<th>CAR BIC difference</th>
<th>likelihood ratio test</th>
<th>POSTCAR BIC difference</th>
<th>likelihood ratio test</th>
<th>p value</th>
<th>Chi square p value</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rank(CE) vs FOM</td>
<td>6.31</td>
<td>-2.36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rank(CE) vs Rank(CE)+FOM</td>
<td>3.97</td>
<td>13.2</td>
<td>0.00027</td>
<td>-9.24</td>
<td>0.024</td>
<td>0.88</td>
<td>0.88</td>
</tr>
<tr>
<td>FOM vs Rank(CE)+FOM</td>
<td>-2.34</td>
<td>6.9</td>
<td>0.00849</td>
<td>-6.88</td>
<td>2.38</td>
<td>0.12</td>
<td>0.12</td>
</tr>
</tbody>
</table>

### Panel D: Continental Europe

<table>
<thead>
<tr>
<th></th>
<th>CAR BIC difference</th>
<th>likelihood ratio test</th>
<th>POSTCAR BIC difference</th>
<th>likelihood ratio test</th>
<th>p value</th>
<th>Chi square p value</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rank(CE) vs FOM</td>
<td>-15.14</td>
<td>3.19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rank(CE) vs Rank(CE)+FOM</td>
<td>-5.10</td>
<td>3.8</td>
<td>0.051</td>
<td>-5.61</td>
<td>3.274</td>
<td>0.07</td>
<td>0.07</td>
</tr>
<tr>
<td>FOM vs Rank(CE)+FOM</td>
<td>10.05</td>
<td>18.9</td>
<td>1.3×10^{-05}</td>
<td>-8.81</td>
<td>0.079</td>
<td>0.78</td>
<td>0.78</td>
</tr>
</tbody>
</table>
Table 7: Sensitivity of earnings announcement returns to $FOM$, out-of-bound dummies, and $Rank(CE)$, based on forecasts in earlier months

This table presents the ordinary least squares estimates of the sensitivity of earnings announcement stock returns ($CAR$) to consensus errors ($Rank(CE)$), $FOM$, $I_{Actual<All}$, 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 $Rank(CE)$ based on median consensus, from 1 to 6, $FOM(\frac{K}{N} - \frac{M}{N})$, 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, the independent variables are calculated based on forecasts within 90 days prior to the date of 3 months prior to the announcement. In Panel B, the independent variables are calculated based on forecasts within 90 days prior to the date of 6 months prior to the announcement. Observations with larger than 100 absolute median consensus value are excluded. All standard errors are clustered by stocks. t-statistics are in parentheses.

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Rank(CE)$</td>
<td>0.00605***</td>
<td>0.00284***</td>
<td>0.00311***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(15.42)</td>
<td>(3.47)</td>
<td>(4.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$FOM$</td>
<td>0.0127***</td>
<td>0.00781***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(17.34)</td>
<td>(5.01)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$I_{Actual&lt;All}$</td>
<td>-0.0121***</td>
<td></td>
<td>-0.00666***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-8.17)</td>
<td></td>
<td>(-3.66)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$I_{Actual&gt; All}$</td>
<td>0.0156***</td>
<td></td>
<td>0.0106***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(9.83)</td>
<td></td>
<td>(5.55)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year effect</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.0224</td>
<td>0.0235</td>
<td>0.0239</td>
<td>0.0248</td>
<td>0.0263</td>
</tr>
</tbody>
</table>

Panel A: forecasts are within 90 days prior to the day of 3 months prior to the announcement

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Rank(CE)$</td>
<td>0.00389***</td>
<td>0.00224***</td>
<td>0.00233***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(9.11)</td>
<td>(2.66)</td>
<td>(3.34)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$FOM$</td>
<td>0.00757***</td>
<td>0.00385***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(10.25)</td>
<td>(2.62)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$I_{Actual&lt;All}$</td>
<td>-0.00884***</td>
<td></td>
<td>-0.00472***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-6.12)</td>
<td></td>
<td>(-2.64)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$I_{Actual&gt; All}$</td>
<td>0.00704***</td>
<td></td>
<td>0.00344*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.24)</td>
<td></td>
<td>(1.76)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year effect</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.0099</td>
<td>0.0097</td>
<td>0.0095</td>
<td>0.0105</td>
<td>0.0107</td>
</tr>
</tbody>
</table>

Panel B: forecasts are within 90 days prior to the day of 6 months prior to the announcement
Table 8: Sensitivity of earnings announcement returns to \textit{FOM}, out-of-bound dummies, and \textit{CE} for subsamples split based on earnings persistence and \textit{DISP}

This table presents the ordinary least squares estimates of the sensitivity of earnings announcement stock returns (\textit{CAR}) to consensus errors (\text{Rank(CE)}), \textit{FOM}, \text{I}_{\text{Actual < All}}\text{, and } \text{I}_{\text{Actual > All}}. The dependent variable is \textit{CAR} (cumulative abnormal return from trading day -1 to 1 around annual earnings announcement dates). The independent variables are \text{Rank(CE)} based on median consensus, from 1 to 6, \textit{FOM} (\(\frac{K - M}{N}\), where \(K(M)\) is the number of forecasts strictly smaller (greater) than the actual earnings, and \(N\) is the total number of analysts, \text{I}_{\text{Actual < All}} (a dummy variable which equals 1 when all analysts’ forecasts are higher than the actual earnings), and \text{I}_{\text{Actual > All}} (a dummy variable which equals 1 when all analysts’ forecasts are lower than the actual earnings). Panel A splits samples based on the median of earnings persistence. Panel B splits samples based the median of dispersion (\textit{DISP}). All standard errors are clustered by stocks. t-statistics are in parentheses.

### Panel A: Subsamples grouped by earnings persistence

<table>
<thead>
<tr>
<th></th>
<th>High Persistence</th>
<th>Low Persistence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)  (2)  (3)  (4)  (5)  (6)  (7)  (8)  (9)  (10)</td>
<td></td>
</tr>
<tr>
<td>Rank(CE)</td>
<td>0.00885***</td>
<td>0.00212*</td>
</tr>
<tr>
<td></td>
<td>(14.55)</td>
<td>(1.81)</td>
</tr>
<tr>
<td>FOM</td>
<td>0.02***</td>
<td>0.0165***</td>
</tr>
<tr>
<td></td>
<td>(16.99)</td>
<td>(7.12)</td>
</tr>
<tr>
<td>I_{Actual &lt; All}</td>
<td>-0.0198***</td>
<td>-0.0112***</td>
</tr>
<tr>
<td></td>
<td>(-8.73)</td>
<td>(-4.01)</td>
</tr>
<tr>
<td>I_{Actual &gt; All}</td>
<td>0.0203***</td>
<td>0.0124***</td>
</tr>
<tr>
<td></td>
<td>(8.57)</td>
<td>(4.48)</td>
</tr>
<tr>
<td>Year effect</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Adjusted (R^2)</td>
<td>0.043</td>
<td>0.053</td>
</tr>
</tbody>
</table>

### Panel B: Subsamples grouped by \textit{DISP}

<table>
<thead>
<tr>
<th></th>
<th>High DISP</th>
<th>Low DISP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)  (2)  (3)  (4)  (5)  (6)  (7)  (8)  (9)  (10)</td>
<td></td>
</tr>
<tr>
<td>Rank(CE)</td>
<td>0.00776***</td>
<td>0.000347</td>
</tr>
<tr>
<td></td>
<td>(15.17)</td>
<td>(0.28)</td>
</tr>
<tr>
<td>FOM</td>
<td>0.0222***</td>
<td>0.0214***</td>
</tr>
<tr>
<td></td>
<td>(16.62)</td>
<td>(6.41)</td>
</tr>
<tr>
<td>I_{Actual &lt; All}</td>
<td>-0.0235***</td>
<td>-0.0141***</td>
</tr>
<tr>
<td></td>
<td>(-8.62)</td>
<td>(-4.43)</td>
</tr>
<tr>
<td>I_{Actual &gt; All}</td>
<td>0.0249***</td>
<td>0.0154***</td>
</tr>
<tr>
<td></td>
<td>(9.29)</td>
<td>(5.3)</td>
</tr>
<tr>
<td>Year effect</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Adjusted (R^2)</td>
<td>0.047</td>
<td>0.055</td>
</tr>
</tbody>
</table>
Table 9: Sensitivity of forecast revision to $FOM$, out-of-bound dummies, $CE$, and $Rank(CE)$

This table presents the ordinary least squares estimates of the sensitivity of analysts’ forecast revision to $Rank(CE)$, $FOM$, $I_{Actual<All}$, and $I_{Actual>\text{All}}$. 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 scores 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), $FOM$ ($\frac{K}{M} - \frac{M}{N}$, 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 $I_{Actual>\text{All}}$ (a dummy variable which equals 1 when all analysts’ forecasts are lower than the actual earnings). All standard errors are clustered by stocks. t-statistics are in parentheses.

<table>
<thead>
<tr>
<th></th>
<th>Panel A: $Rank(CE)$ is based on mean consensus</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>$Rank(CE)$</td>
<td>0.0274*** (7.14)</td>
<td>0.00677 (0.97)</td>
<td>0.0187*** (2.85)</td>
<td></td>
</tr>
<tr>
<td>$FOM$</td>
<td>0.109*** (9.26)</td>
<td>0.0893*** (4.11)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$I_{Actual&lt;All}$</td>
<td>-0.128*** (-5.22)</td>
<td>-0.0718** (-2.25)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$I_{Actual&gt;\text{All}}$</td>
<td>0.0785*** (4.06)</td>
<td>0.0255 (1.03)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year effect</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.046</td>
<td>0.048</td>
<td>0.045</td>
<td>0.048</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Panel B: $Rank(CE)$ is based on median consensus</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>$Rank(CE)$</td>
<td>0.0349*** (5.59)</td>
<td>-0.0219* (-1.85)</td>
<td>0.0109 (1.02)</td>
<td></td>
</tr>
<tr>
<td>$FOM$</td>
<td>0.104*** (9.24)</td>
<td>0.143*** (6.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$I_{Actual&lt;All}$</td>
<td>-0.119*** (-5.28)</td>
<td>-0.0987*** (-3.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$I_{Actual&gt;\text{All}}$</td>
<td>0.0781*** (4.04)</td>
<td>0.06** (2.29)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year effect</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.042</td>
<td>0.047</td>
<td>0.044</td>
<td>0.047</td>
</tr>
</tbody>
</table>
Table 10: 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. All standard errors are clustered by stocks. t-statistics are in parentheses.

<table>
<thead>
<tr>
<th>Panel A: Parametric-dependent-( CE ) and ( FOM )</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( CE )</td>
<td>0.00453***</td>
<td>-0.000924</td>
<td>-0.000396</td>
</tr>
<tr>
<td></td>
<td>(2.86)</td>
<td>(-0.75)</td>
<td>(-0.35)</td>
</tr>
<tr>
<td>( FOM )</td>
<td>0.0219***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(19.55)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( I_{Actual&lt;All} )</td>
<td>-0.0236***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-10.18)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( I_{Actual&gt;All} )</td>
<td>0.0232***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(11.54)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year effect</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Adjusted ( R^2 )</td>
<td>0.005</td>
<td>0.061</td>
<td>0.053</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: Parametric-dependent-( Rank(CE) ) and ( FOM )</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Rank(CE) )</td>
<td>0.00495***</td>
<td>0.00051</td>
<td>0.00229***</td>
</tr>
<tr>
<td></td>
<td>(17.16)</td>
<td>(1.02)</td>
<td>(7.05)</td>
</tr>
<tr>
<td>( FOM )</td>
<td>0.0201***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(10.44)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( I_{Actual&lt;All} )</td>
<td>-0.0161***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-5.77)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( I_{Actual&gt;All} )</td>
<td>0.0159***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(8.29)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year effect</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Adjusted ( R^2 )</td>
<td>0.046</td>
<td>0.061</td>
<td>0.057</td>
</tr>
</tbody>
</table>
Figure 1: $E[S|U]$ for different $\sigma_B$’s. The parameters used are $\omega_0 = 0.7$, $\sigma_A = 1$, $\sigma_F = 1/2$, $\sigma_b = 1/5$ and $\mu_B = 2$.

Figure 2: $E[S|U]$ for different $\mu_B$’s. The parameters used are $\omega_0 = 0.7$, $\sigma_A = 1$, $\sigma_F = 1/2$, $\sigma_b = 1/5$ and $\sigma_B = 2$. 
Figure 3: 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 4: 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 5: The time series of the percentage of misses on the same side.
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 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 $-1$ regardless of its magnitude. One consequence is that $CE$ and FOM are no longer highly correlated. While we observe a rather low correlation in earnings data, which is also due to outliers, 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 $\omega_1$ and $r_B = \frac{\sigma_B}{\sigma_A}$ vary over their range, and the others fixed at $N = 20$, $r_F = 1/2$ and $r_b = r_B/5$. Appendix Figure 1 shows how the correlation decreases with $r_B$, the relative uncertainty level of the bias component $B$. In terms of $\omega_1$, recall it is the proportion of biased forecasts, so the correlation first decreases with the introduction of biased forecasts as soon as $\omega_1$ 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 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 $\omega_1$ 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 $S$ from our formula and simulate 100,000 samples of $X$ and $Y$ to compute the correlation between FOM and $S$. Appendix Figure 2 shows a representative pattern of their relative performance as a function of $\omega_1$ and $r_B$, where the difference be-
tween $\text{Cor}[CE, S]$ 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.1.1 $CE$ and $\text{Rank}(CE)$

In practice, people use $\text{Rank}(CE)$, 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 3 shows a representative pattern of the difference in the performance of $\text{Rank}(CE)$ and $FOM$ (i.e., $\text{Cor}[\text{Rank}(CE), S] - \text{Cor}[FOM, S]$) as a function of $\omega_1$ and $r_B$ with the same set of parameters as in Section A.1, where each $\text{Cor}[\text{Rank}(CE), S]$ is computed using 50,000 simulated samples. Comparing with Appendix Figure 2, there is some improvement when $r_B$ is not too large. However, the essence of the analysis on $CE$ carries over to $\text{Rank}(CE)$ because when $CE$ is greatly contaminated, the coding of $\text{Rank}(CE)$ 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 $\text{Rank}(CE)$ for the same reason as it does over $CE$, the reason being their sensitivity to large bias. That being said, $\text{Rank}(CE)$ does have better property when treating the few outliers that overthrow $CE$.

A.2 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

$$CE_{\lambda}^{\text{win}} = A - \bar{F}_\lambda^{\text{win}},$$
where $\lambda$ is the percentage of data on each tail being replaced. Note that when $\lambda = 50\%$, the Winsorized mean becomes median:

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

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 4 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 $\lambda$ for Winsorization. Unfortunately, the fraction of biased forecasts $\omega_1$ is usually unknown in practice and may even be varying, so it is hard if not impossible to set $\lambda$, the single important parameter for Winsorization, and an inappropriate choice might result in undesirable performance. This is illustrated in Appendix Figure 5, where the relative performance of different Winsorized measures changes with the fraction of biased forecasts $\omega_1$, 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.3 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, $\omega_1$ 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

$$F_i = e + \epsilon_i,$$

where $\epsilon_i$'s follow a mixture of two normal distributions: $\mathcal{N}(0, \sigma_0^2)$ with probability $\omega_0$ 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:

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

so $CE$ still converges to $S$ by the law of large numbers. That is, although $\sigma_1^2$ 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 $\text{Rank}(CE)$ should be better for larger $N$ under the alternative model, which does not quite match what we see in the real data.

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 4, Appendix Figure 6 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.

59
A.4 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

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

where $\epsilon_i \sim \mathcal{N}(0, \sigma^2_F)$ and is uncorrelated with the randomness in $A$. Each forecast is unbiased with probability $\omega_0$, 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 $B \sim \mathcal{N}(0, \sigma^2_B)$ is drawn first, and conditional on this realized $B$ individual bias $b_i$ follows $\mathcal{N}(B, \sigma^2_b)$.

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

$$\hat{A} = \frac{1}{N} \sum_{i=1}^{N} F_i^* \quad (44)$$

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

$$S = A - \hat{A} \quad (45)$$

Notice that $CE$ is now given by

$$CE = A - \frac{1}{N} \sum_{i=1}^{N} F_i \quad (46)$$
and $FOM$ is now given by

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

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

We can calculate that

$$ \text{Cor}(CE, S) = \frac{1}{\sqrt{1 + \omega_0 \omega_1 r_B^2 + \omega_1 r_b^2 + \omega_2^2 r_B^2 N}} \quad (48) $$

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

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

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

Since $\text{Cor}(FOM, S) \geq \frac{\omega_0 \sqrt{2/\pi}}{\sqrt{1 + \omega_1 N}}$, it follows then that if $r_B$ gets large, then $\text{Cor}(CE, S)$ drops below $\text{Cor}(FOM, S)$. This then confirms our results in our baseline model.
Appendix Figure 1: The contour plot of Cor[CE, FOM] as a function of the key parameters \( \omega_1 \) 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 \( \omega_1 \) 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 2: The contour plot of $\text{Cor}(CE, S) - \text{Cor}(FOM, S)$ as a function of the key parameters $\omega_1$ 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 $\omega_1$ 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 3: The contour plot of $\text{Cor}[\text{Rank}(CE), S] - \text{Cor}[\text{FOM}, S]$ as a function of the key parameters $\omega_1$ and $r_B$ in biased forecasts case. The contour value is the difference between the correlations of the rank score of consensus errors $\text{Rank}(CE)$ and fraction of misses $\text{FOM}$ to $S$ the market surprise, the $y$-axis is $\omega_1$ 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 comparison between the correlations of fraction of misses FOM and different Winsorized measures $CE_{\text{win}}^\lambda$ 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 5: 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 $\omega_1$ in biased forecasts case, where $\omega_1$ 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 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 $\sigma_1/\sigma_A$ (shown in log-scale) under the alternative modelling without introducing bias, where $\sigma_1$ 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 $N = 20$. 